Working Paper

Free-Steering Relaxation Methods for Problems with Strictly Convex Costs and Linear Constraints

Krzysztof C. Kiwiel

WP-94-89
September 1994

IIASA
International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 71521 □ Telex: 079 137 iiasa a □ Telefax: +43 2236 71313

Free-Steering Relaxation Methods for Problems with Strictly Convex Costs and Linear Constraints

Krzysztof C. Kiwiel

WP-94-89
September 1994

IIASA International Institute for Applied Systems Analysis ◻ A-2361 Laxenburg ◻ Austria
Telephone: +43 2236 71521 ◻ Telex: 079 137 iiasa a ◻ Telefax: +43 2236 71313

Free-steering relaxation methods for problems with strictly convex costs and linear constraints*

Krzysztof C. Kiwiel[†]

August 12, 1994

Abstract

We consider dual coordinate ascent methods for minimizing a strictly convex (possibly nondifferentiable) function subject to linear constraints. Such methods are useful in large-scale applications (e.g., entropy maximization, quadratic programming, network flows), because they are simple, can exploit sparsity and in certain cases are highly parallelizable. We establish their global convergence under weak conditions and a free-steering order of relaxation. Previous comparable results were restricted to special problems with separable costs and equality constraints. Our convergence framework unifies to a certain extent the approaches of Bregman, Censor and Lent, De Pierro and Iusem, and Luo and Tseng, and complements that of Bertsekas and Tseng.

Key words. Convex programming, entropy maximization, nondifferentiable optimization, relaxation methods, dual coordinate ascent, B-functions.

1 Introduction

We study algorithms for the following convex programming problem

$$\text{minimize} \quad f(x), \tag{1.1a}$$
$$\text{subject to} \quad Ax \le b, \tag{1.1b}$$

where $f : \mathbb{R}^n \to (-\infty, \infty]$ is a (possibly nondifferentiable) strictly convex function that has some properties of differentiable Bregman functions [Bre67, CeL81] (cf. §2), A is a given $m \times n$ matrix and b is a given m-vector. (Equality constraints are discussed later.)

This problem arises in many applications, e.g., linear programming [Bre67, Erl81, Man84], quadratic programming [Hil57, LeC80, LiP87], image reconstruction [Cen81, Cen88, CeH87, Elf89, HeL78, ZeC91b], matrix balancing [CeZ91, Elf80, Kru37, LaS81], "$x \log x$" entropy optimization [DaR72, CDPE90, Elf80], "$\log x$" entropy optimization

*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)

[CDPI91, CeL87], and network flow programming [BHT87, BeT89, NiZ92, NiZ93a, Roc84, ZeC91a, Zen91]. Further references can be found, e.g., in [LuT92b, LuT92c, Tse90, Tse91, TsB91].

The usual dual problem of (1.1) consists in maximizing a concave differentiable dual functional subject to nonnegativity constraints. This motivates coordinate ascent methods for solving the dual problem which, at each iteration, increase the dual functional by adjusting one coordinate of the dual vector. Such methods are simple, use little storage and can exploit problem sparsity. They are among the most popular (and sometimes the only practical) methods for large-scale optimization. Also such methods may be used as subroutines in the proximal minimization algorithms with D-functions [CeZ92, Eck93, Teb92], giving rise to massively parallel methods for problems with huge numbers of variables and constraints [CeZ91, NiZ92, NiZ93a, NiZ93b, Zen91, ZeC91a, ZeC91b]. Other examples include methods for specific problems quoted above, and methods for more general problems [Bre67, CeL81, DPI86, LuT92b, LuT92c, Tse90, Tse91, TsB87, TsB91].

At least three general approaches to convergence analysis of such methods can be distinguished. Because different assumptions on the problem are employed, each approach covers many applications, but not all. First, the approach based on Bregman functions [Bre67, CeL81, DPI86] imposes some smoothness assumptions on f and so-called zone consistency conditions that may be difficult to ensure. Second, the approach of [LuT92b, Tse91] assumes that f is essentially smooth. (Our terminology follows [Roc70]; see below for a review.) Third, the approach of [Tse90, TsB87, TsB91] requires that f be cofinite. Usually it is assumed that the relaxed coordinates are chosen in an almost (essentially) cyclic order [CeL81, DPI86, LuT92b, Tse90, Tse91, TsB87, TsB91] (i.e., each coordinate is chosen at least once every i_{cyc} iterations, for some fixed $i_{\text{cyc}} \geq m$), by a Gauss-Southwell max-type rule [Bre67, LuT92b, Tse90, Tse91, TsB91], or—for strongly convex costs only—in a quasi-cyclic order [Tse90, TsB87, TsB91] (in which the lengths of the cycles, i.e., i_{cyc}, are allowed to grow, but not too fast). Convergence under the weakest assumption of *free-steering* relaxation (in which each coordinate is merely chosen infinitely many times) has so far been established only for network flow problems with separable costs and equality constraints [BHT87], [BeT89, §5.5] and for special cases of iterative scaling [BCP93].

In this paper we establish global convergence of a general dual ascent method under free-steering relaxation (for both equality and inequality constraints), weak assumptions on (1.1) and inexact line searches. Our assumptions on problem (1.1) (cf. §2) are weaker than those of [Bre67, CeL81, DPI86] and [LuT92b, Tse91]; thus we generalize those approaches. We show that inexact line searches are implementable because the dual functional, being essentially smooth, may act like a barrier to keep iterates within the region where it is differentiable.

In particular, our results imply global convergence under free-steering relaxation of Hildreth's method [Hil57] for quadratic programming. We note that for the related problem of finding a point in the intersection of a finite family of closed convex sets, convergence of "inexact" free-steering versions of the successive projection method [GPR67] has been established quite recently [ABC83, Kiw94, FlZ90]; see [Ott88, Tse92] for results under "exact" projections.

Attempting to capture objective features essential to optimization, we introduce the class of B-functions (cf. Definition 2.1) which generalizes that of Bregman functions [CeL81] and covers more applications. The usefulness of our B-functions is not limited to linearly constrained minimization; this will be shown elsewhere.

We concentrate on global convergence under general conditions, whereas the recent results on linear rate of convergence of relaxation methods [Ius91, LuT91, LuT92a, LuT92b, LuT92c, LuT93] require additional regularity assumptions.

The paper is organized as follows. In §2 we introduce the class of B-functions, highlight some of its properties and present our method. Its global convergence under free-steering relaxation control is established in §3. In §4 we relate Bregman projections [CeL81] to exact linesearches and give conditions for overrelaxation that supplement those in [DPI86, Tse90]. Convergence under conditions similar to those in [LuT92b, Tse91] and under another regularity condition is established in §§5 and 6 respectively. Some additional remarks are given in §7. In §8 we discuss block coordinate relaxation. The Appendix contains proofs of certain technical results.

Our notation and terminology mostly follow [Roc70]. $\langle\cdot,\cdot\rangle$ and $|\cdot|$ are the Euclidean inner product and norm respectively, a^i is column i of the transpose A^T of A, b_i is component i of b, $\mathbb{R}^m_+$ and $\mathbb{R}^m_>$ are the nonnegative and positive orthants of $\mathbb{R}^m$ respectively, $[\cdot]_+$ denotes the orthogonal projection onto $\mathbb{R}^m_+$, i.e., $([p]_+)_i = \max\{p_i, 0\}$ for $p \in \mathbb{R}^m$ and $i = 1{:}\,m$, where $1{:}\,m$ denotes $1, 2, \cdots, m$, and e^i is the ith coordinate vector in $\mathbb{R}^m$. 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. $\sigma_C(\cdot) = \sup_{x\in C}\langle\cdot, x\rangle$ is the *support* function of C. For any closed proper convex function f on $\mathbb{R}^n$ and x in its *effective domain* $C_f = \{x : f(x) < \infty\}$, $\partial f(x)$ denotes the *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

$$f'(x; d) \geq \sigma_{\partial f(x)}(d) = \sup\{\langle g, d\rangle : g \in \partial f(x)\}. \tag{1.2}$$

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 differentiable at x iff $\partial f(x) = \{\nabla f(x)\}$, where ∇f is the gradient of f [Roc70, Thm 25.1]. f is called *essentially strictly convex* if f is strictly convex on every convex subset of $C_{\partial f}$. 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 $f'(x + t(y - x); y - x) \downarrow -\infty$ as $t \downarrow 0$ for any $y \in \mathring{C}_f$ and $x \in \operatorname{bd} C_f$ (equivalently $|\nabla f(x^k)| \to \infty$ if $\{x^k\} \subset \mathring{C}_f$, $x^k \to x \in \operatorname{bd} C_f$ [Roc70, Lem. 26.2]); then $f'(y + t(x - y); x - y) \uparrow \infty$ as $t \uparrow 1$ (cf. $f'(\cdot; d) \geq -f'(\cdot; -d)\ \forall d$).

2 B-functions and the algorithm

We first define our B-functions and Bregman functions [CeL81].

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

$$D_f^\flat(x, y) = f(x) - f(y) - \sigma_{\partial f(y)}(x - y) \qquad \forall x, y \in C_f, \tag{2.1a}$$

$$D_f^\sharp(x, y) = f(x) - f(y) + \sigma_{\partial f(y)}(y - x) \qquad \forall x, y \in C_f. \tag{2.1b}$$

By convexity (cf. (1.2)), $f(x) \geq f(y) + \sigma_{\partial f(y)}(x-y)$ and

$$0 \leq D_f^\flat(x,y) \leq f(x) - f(y) - \langle g, x-y\rangle \leq D_f^\sharp(x,y) \quad \forall x,y \in C_f, g \in \partial f(y). \tag{2.2}$$

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

$$D_f(x,y) = f(x) - f(y) - \langle \nabla f(y), x-y\rangle \quad \forall x \in C_f, y \in C_{\nabla f}, \tag{2.3}$$

since

$$D_f(x,y) = D_f^\flat(x,y) = D_f^\sharp(x,y) \quad \forall x \in C_f, y \in C_{\nabla f}. \tag{2.4}$$

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

(a) f is strictly convex on C_f.
(b) f is continuous on C_f.
(c) For every $\alpha \in \mathbb{R}$ and $x \in C_f$, the set $\mathcal{L}_f^1(x,\alpha) = \{y \in C_{\partial f} : D_f^\flat(x,y) \leq \alpha\}$ is bounded.
(d) For every $\alpha \in \mathbb{R}$ and $x \in C_f$, if $\{y^k\} \subset \mathcal{L}_f^1(x,\alpha)$ is a convergent sequence with limit $y^* \in C_f \setminus \{x\}$, then $D_f^\sharp(y^*, y^k) \to 0$.

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 $\mathcal{L}_h^2(\tilde{y},\alpha) = \{x \in \bar{S} : D_h(x,\tilde{y}) \leq \alpha\}$ and $\mathcal{L}_h^3(\tilde{x},\alpha) = \{y \in S : D_h(\tilde{x},y) \leq \alpha\}$ 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 f of h to $\mathbb{R}^n$, defined by $f(x) = h(x)$ if $x \in \bar{S}$, $f(x) = \infty$ otherwise, is a B-function with $C_f = \bar{S}$, $\operatorname{ri} C_f = S$ and $D_f^\flat(\cdot,y) = D_f^\sharp(\cdot,y) = D_f(\cdot,y)$ $\forall y \in S$.)

$D_f^\flat$ and $D_f^\sharp$ are used like distances, because for $x,y \in C_f$, $0 \leq D_f^\flat(x,y) \leq D_f^\sharp(x,y)$, and $D_f^\flat(x,y) = 0 \iff D_f^\sharp(x,y) = 0 \iff x = y$ 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 f essential for algorithmic purposes. We show in §7 that condition (b) implies (c) if f is cofinite. Sometimes one may verify the following stronger version of condition (d)

$$C_{\partial f} \supset \{y^k\} \to y^* \in C_f \Rightarrow D_f^\sharp(y^*, y^k) \to 0 \tag{2.5}$$

by using the following lemmas. Their proofs are given in the Appendix.

Lemma 2.3. (a) *Let f be a closed proper convex function on $\mathbb{R}^n$, and let $S \neq \emptyset$ be a compact subset of* $\operatorname{ri} C_f$. *Then there exists $\alpha \in \mathbb{R}$ such that $|\sigma_{\partial f(y)}(x-z)| \leq \alpha|x-z|$, $|f(x) - f(y)| \leq \alpha|x-y|$ and $|D_f^\sharp(x,y)| \leq 2\alpha|x-y|$ 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* $\mathbb{R}^n$, *i.e.,* $\delta_S(x) = 0$ *if* $x \in S$, $\delta_S(x) = \infty$ *if* $x \notin S$. *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* f *be a closed proper convex function on* $\mathbb{R}$. *Then* f *is continuous on* C_f, *and* $D_f^\sharp(y^*, y^k) \to 0$ *if* $y^k \to y^* \in C_f$, $\{y^k\} \subset C_f$.

Lemma 2.4. (a) *Let* $f = \sum_{i=1}^k f_i$, *where* $f_1, \ldots, f_k$ *are closed proper convex functions such that* $f_{j+1}, \ldots, f_k$ $(j \geq 0)$ *are polyhedral and* $\cap_{i=1}^j \operatorname{ri}(C_{f_i}) \cap_{i=j+1}^k C_{f_i} \neq \emptyset$. *If* f_1 *satisfies condition* (c) *of Definition* 2.1, *then so does* f. *If* $f_1, \ldots, f_j$ *satisfy condition* (d) *of Definition* 2.1 *or* (2.5), *then so does* f. *If* f_1 *is a* B*-function,* $f_2, \ldots, f_j$ *are continuous on* $C_f = \cap_{i=1}^k C_{f_i}$ *and satisfy condition* (d) *of Definition* 2.1, *then* f *is a* B*-function. In particular,* f *is a* B*-function if so are* $f_1, \ldots, f_j$.

(b) *Let* $f_1, \ldots, f_j$ *be* B*-functions such that* $\cap_{i=1}^j \operatorname{ri} C_{f_i} \neq \emptyset$. *Then* $f = \max_{i=1:j} f_i$ *is a* B*-function.*

(c) *Let* f_1 *be a* B*-function and let* f_2 *be a closed proper convex function such that* $C_{f_1} \subset \operatorname{ri} C_{f_2}$. *Then* $f = f_1 + f_2$ *is a* B*-function.*

(d) *Let* $f_1, \ldots, f_n$ *be closed proper strictly convex functions on* $\mathbb{R}$ *such that* $\mathcal{L}_{f_i}^1(t, \alpha)$ *is bounded for any* $t, \alpha \in \mathbb{R}$, $i = 1\!:\!n$. *Then* $f(x) = \sum_{i=1}^n f_i(x_i)$ *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* $C_{h^*} = \mathring{C}_{h^*}$.

Examples 2.6. Let $\psi : \mathbb{R} \to (-\infty, \infty]$ and $f(x) = \sum_{i=1}^n \psi(x_i)$. In each of the examples, it can be verified that f is an essentially smooth B-function.

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

2. $\psi(t) = -t^\alpha/\alpha$ if $t \geq 0$ and $\alpha \in (0, 1)$, $\psi(t) = \infty$ if $t < 0$, i.e., $f(x) = -\|x\|_\alpha^\alpha/\alpha$ if $x \geq 0$. Then $f^*(y) = -\|y\|_\beta^\beta/\beta$ if $y < 0$ and $\alpha + \beta = \alpha\beta$, $f^*(y) = \infty$ 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 $f^*(y) = \sum_{i=1}^n \exp(y_i - 1)$ [Roc70, p. 105] and $D_f(x, y) = \sum_{i=1}^n x_i \ln(x_i/y_i) + y_i - x_i$ (the Kullback-Liebler entropy).

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

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

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

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 $f^*(y) = \sum_{i=1}^n (1 - y_i/\beta)^{-\beta}$ for $y \in C_f^* = (-\infty, \beta)^n$, where $\beta - \alpha = \alpha\beta$. For $\alpha = \frac{1}{2}$, $D_f(x, y) = \sum_{i=1}^n (x_i^{1/2} - y_i^{1/2})^2/y_i^{1/2}$.

Examples 2.7. In both examples, it can be verified that f is a cofinite B-function.

1. $f(x) = \sum_{i=1}^n x_i \ln x_i$ if $x \geq 0$ and $\sum_{i=1}^n x_i = 1$, $f(x) = \infty$ otherwise (cf. Lemmas 2.3(b) and 2.4(a)). Then $f^*(y) = \ln(\sum_{i=1}^n \exp(y_i))$ [Roc70, pp. 148–149].

2. $f(x) = -(\alpha^2 - |x|^2)^{1/2}$ if $|x| \leq \alpha$, $\alpha \geq 0$, $f(x) = \infty$ if $|x| > \alpha$. (Here (2.5) fails if $n > 1$ and $\alpha > 0$.) Then $f^*(y) = \alpha(1 + |y|^2)^{1/2}$ [Roc70, p. 106].

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

Assumption 2.8. (i) f is a (possibly nonsmooth) B-function.
(ii) The *feasible set* $X = \{x \in C_f : Ax \leq b\}$ of (1.1) is nonempty.
(iii) $(-A^T P) \cap \operatorname{im} \partial f \neq \emptyset$, where $P = \mathbb{R}_+^m$.

This assumption is required in [Bre67, CeL81, DPI86], where the condition $(-A^T P) \cap \operatorname{im} \partial f \neq \emptyset$ is only used to start algorithms. We now exhibit important implications of this condition for the usual dual problem of (1.1) (missing in [Bre67, CeL81, DPI86]). The *dual problem*, obtained by assigning a multiplier vector p to the constraints $Ax \leq b$, is

$$\text{maximize} \quad q(p), \tag{2.6a}$$
$$\text{subject to} \quad p \geq 0, \tag{2.6b}$$

where $q : \mathbb{R}^m \to [-\infty, \infty)$ is the concave *dual functional* given by

$$q(p) = \inf_x \{f(x) + \langle p, Ax - b \rangle\} = -f^*(-A^T p) - \langle p, b \rangle . \tag{2.7}$$

The dual problem (2.6) is a concave program with simple bounds. *Weak duality* means $\sup_{p \in P} q(p) \leq \inf_{x \in X} f(x)$. The following lemma is proven in the Appendix.

Lemma 2.9. *Let f be a closed proper essentially strictly convex function, $(-A^T)^{-1} \operatorname{im} \partial f \neq \emptyset$ and $C_q = \{p : q(p) > -\infty\}$. Then q is closed proper concave and continuously differentiable on*

$$\mathring{C}_q = \{p : -A^T p \in \operatorname{im} \partial f\} = \{p : \operatorname*{Arg\,min}_x [f(x) + \langle p, Ax \rangle] \neq \emptyset\},$$

$\operatorname{im} \partial f = \mathring{C}_{f^*}$ *and $\nabla q(p) = Ax(p) - b$ for any $p \in \mathring{C}_q$, where*

$$x(p) = \nabla f^*(-A^T p) = \arg\min_x \{f(x) + \langle p, Ax \rangle\} = (\partial f)^{-1}(-A^T p) \tag{2.8}$$

is continuous on $\mathring{C}_q$. Further, $-q$ is essentially smooth, so that $q'(p + t(\bar{p} - p), \bar{p} - p) \downarrow -\infty$ as $t \uparrow 1$ for any $p \in \mathring{C}_q$ and $\bar{p} \in \operatorname{bd} C_q$.

The first assertion of Lemma 2.9 is well known (cf. [Fal67]). The final assertion will be used to keep our algorithm within $\mathring{C}_q$, where q is smooth. For each $p \in \mathring{C}_q \cap P$, we let

$$r(p) = \nabla q(p) = Ax(p) - b. \tag{2.9}$$

Note that $x(p)$ and p solve (1.1) and (2.6) if

$$p = [p + r(p)]_+. \tag{2.10}$$

Indeed, then $r(p) \le 0$, $p \ge 0$, $\langle p, r(p)\rangle = 0$ and $-A^T p \in \partial f(x(p))$ by (2.8).

In the kth iteration of our method for solving (2.6), given $p^k \in \mathring{C}_q \cap P$, a coordinate i_k such that $r_{i_k}(p^k) > 0$ (or < 0) is chosen and p_{i_k} is increased (or decreased, respectively) to increase the value of q, using the fact that $r_{i_k}(p)$ is continuous around p^k and nonincreasing in p_{i_k}, since q is concave. We let (cf. (2.8) and (2.1)-(2.2)) $x^k = x(p^k)$,

$$g^k := -A^T p^k \in \partial f(x^k), \tag{2.11}$$

$$D_f^k(x, x^k) = f(x) - f(x^k) - \left\langle g^k, x - x^k \right\rangle \quad \forall x. \tag{2.12}$$

Algorithm 2.10.

Step 0 (*Initiation*). Select an initial $p^1 \in P \cap \mathring{C}_q$, *relaxation bounds* $\omega_{\min} \in (0,1)$ and $\omega_{\max} \in [1,2)$ and a *relaxation tolerance* $\kappa_D \in (0,1]$. Set $x^1 = x(p^1)$ by (2.8). Set $k = 1$.

Step 1 (*Coordinate selection*). Choose $i_k \in \{1\colon m\}$.

Step 2 (*Direction finding*). Find the derivative $q_k'(0)$ of the reduced objective

$$q_k(t) = q(p^k(t)) \quad \text{with} \quad p^k(t) = p^k + te^{i_k} \quad \forall t \in \mathbb{R}. \tag{2.13}$$

Step 3 (*Trivial step*). If $q_k'(0) = 0$, or $q_k'(0) < 0$ and $p_{i_k}^k = 0$, set $t_k = 0$ and go to Step 5.

Step 4 (*Linesearch*). Find $t_k \ge -p_{i_k}^k$ such that $p^k(t_k) \in \mathring{C}_q$ and
(i) if $q_k'(0) > 0$ then $\omega_k \in [\omega_{\min}, \omega_{\max}]$;
(ii) if $q_k'(0) < 0$ then either $\omega_k \in [\omega_{\min}, \omega_{\max}]$, or $\omega_k \in [0, \omega_{\min})$ and $t_k = -p_{i_k}^k$;
and

$$q(p^k(t_k)) - q(p^k) \ge \kappa_D D_f^k(x(p^k(t_k)), x^k) \quad \text{if} \quad \omega_k > 1, \tag{2.14}$$

where

$$\omega_k = [q_k'(0) - q_k'(t_k)]/q_k'(0). \tag{2.15}$$

Step 5 (*Dual step*). Set $p^{k+1} = p^k(t_k)$, $x^{k+1} = x(p^{k+1})$, increase k by 1 and go to Step 1.

A few remarks on the algorithm are in order.

Step 0 is well defined, since $P \cap \mathring{C}_q \ne \emptyset$ by Assumption 2.8 and Lemma 2.9. Suppose $p^k \in P \cap \mathring{C}_q$ at Step 1. By (2.13) and Lemma 2.9, $q_k'(\cdot) = r_{i_k}(p^k(\cdot))$ is *continuous and nonincreasing* on the nonempty open interval $T^0 = \{t : p^k(t) \in \mathring{C}_q\}$. Step 4 chooses $p^k(t_k) \ge 0$, since $-p_{i_k}^k = \inf_{p^k(t) \ge 0} t$. To see that Step 4 is well defined, suppose $q_k'(0) > 0$ (the case of $q_k'(0) < 0$ is similar). Let $t_k^0 = \sup_{t \in T^0} t$ and $T' = \{t \in T^0 : 0 \le q_k'(t) \le (1-\omega_{\min})q_k'(0)\}$. It suffices to show that T' is a nontrivial interval. If t_k^0 is finite then $q_k'(t) \downarrow -\infty$ as $t \uparrow t_k^0$ by Lemma 2.9, whereas if $t_k^0 = \infty$ then, if we had $q_k'(t) > \epsilon$ for some fixed $\epsilon \in (0, (1-\omega_{\min})q_k'(0)]$ and all $t \ge 0$, $q(p^k(t)) = q(p^k) + \int_0^t q_k'(\theta)d\theta \to \infty$ as $t \to \infty$ would contradict the weak duality relation $\sup_P q \le \inf_X f$. Hence, using the continuity and monotonicity of q_k' on T^0, the required t_k can be found, e.g., via bisection. Note that $t_k q_k'(t_k) \ge 0$ iff $\omega_k \le 1$ (q_k is monotone), where $\omega_k = 1$ if $t_k = 0$. To sum up, by induction we have for all k

$$p^k \in P \cap \mathring{C}_q, \tag{2.16}$$

$$t_k q_k'(t_k) \ge 0 \quad \text{if} \quad \omega_k \le 1. \tag{2.17}$$

We make the following standing assumption on the order of relaxation.

Assumption 2.11. Every element of $\{1\colon m\}$ appears in $\{i_k\}$ infinitely many times.

3 Convergence

We shall show that $\{x^k\}$ converges to the solution of (1.1). Because the proof of convergence is quite complex, it is broken into a series of lemmas.

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

Lemma 3.1 ([TsB91, Lemma 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 $\epsilon > 0$ such that $\{x \in C_h : |x - y| \leq \epsilon\}$ is closed.*

(b) *For any $y \in C_h$ and z such that $y + z \in C_h$, and any sequences $y^k \to y$ and $z^k \to z$ such that $y^k \in C_h$ and $y^k + z^k \in C_h$ for all k, we have $\limsup_{k\to\infty} h'(y^k; z^k) \leq h'(y; z)$.*

Lemma 3.2. *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 such that, 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, Lemma 2]. □

Lemmas 3.1–3.2 could be expressed in terms of the following analogue of (2.1)

$$D'_f(x, y) = f(x) - f(y) - f'(y; x - y) \quad \forall x, y \in C_f. \tag{3.1}$$

Lemma 3.3. *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 such that $D'_h(y^*, y^k) \to 0$ 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 Lemma 3.2 and $h(y^k) \xrightarrow{K} h(y^\infty)$ by continuity of h on C_h. Then by Lemma 3.1(b), $0 = \liminf_{k\in K} D'_h(y^*, y^k) \geq h(y^*) - h(y^\infty) - h'(y^\infty; y^* - y^\infty)$ yields $y^\infty = y^*$ by strict convexity of h. Hence $y^k \to y^*$. □

Using (2.9) and (2.16), we let

$$r^k = \nabla q(p^k) = Ax^k - b. \tag{3.2}$$

By (2.7), (2.8), (2.13), (1.2), (2.11), (2.2), (2.12) and (3.1), for all k

$$q(p^k) = f(x^k) + \left\langle p^k, Ax^k - b \right\rangle, \tag{3.3}$$

$$x^k \in C_{\partial f} \subset C_f, \tag{3.4}$$

$$q'_k(0) = r^k_{i_k} = \left\langle a^{i_k}, x^k \right\rangle - b_{i_k}, \tag{3.5}$$

$$q'_k(t_k) = r^{k+1}_{i_k} = \left\langle a^{i_k}, x^{k+1} \right\rangle - b_{i_k}, \tag{3.6}$$

$$0 \leq D'_f(x, x^k) \leq D^\flat_f(x, x^k) \leq D^k_f(x, x^k) \leq D^\sharp_f(x, x^k) \quad \forall x. \tag{3.7}$$

Lemma 3.4. *Let $\Delta_q^k = q(p^{k+1}) - q(p^k)$ for all k. Then:*

$$\Delta_q^k = D_f^k(x^{k+1}, x^k) + t_k q_k'(t_k) \geq \kappa_D D_f^k(x^{k+1}, x^k) \geq 0 \quad \forall k, \tag{3.8}$$

$$\sum_{k=1}^{\infty} D_f^k(x^{k+1}, x^k) \leq \sum_{k=1}^{\infty} \Delta_q^k / \kappa_D < \infty, \tag{3.9}$$

$$\sum_{k=1}^{\infty} |t_k q_k'(t_k)| \leq (1 + 1/\kappa_D) \sum_{k=1}^{\infty} \Delta_q^k < \infty, \tag{3.10}$$

$$q(p^k) \leq f(x) - D_f^k(x, x^k) \leq f(x) - D_f^\flat(x, x^k) \leq f(x) \quad \forall x \in X, \ \forall k. \tag{3.11}$$

Proof. Using (3.3), (3.2), (2.11), (2.12), (3.6) and $p^{k+1} = p^k + t_k e^{i_k}$, we have

$$\begin{aligned}
\Delta_q^k &= f(x^{k+1}) - f(x^k) + \left\langle p^{k+1}, Ax^{k+1} - b \right\rangle - \left\langle p^k, Ax^k - b \right\rangle \\
&= f(x^{k+1}) - f(x^k) + \left\langle p^k, Ax^{k+1} - Ax^k \right\rangle + \left\langle p^{k+1} - p^k, Ax^{k+1} - b \right\rangle \\
&= f(x^{k+1}) - f(x^k) + \left\langle A^T p^k, x^{k+1} - x^k \right\rangle + \left\langle p^{k+1} - p^k, r^{k+1} \right\rangle \\
&= f(x^{k+1}) - [f(x^k) + \left\langle g^k, x^{k+1} - x^k \right\rangle] + t_k \left\langle e^{i_k}, r^{k+1} \right\rangle \\
&= D_f^k(x^{k+1}, x^k) + t_k q_k'(t_k),
\end{aligned}$$

so (3.8) follows from (2.14) and (2.17), since $\kappa_D \in (0, 1]$ and $D_f^k(x^{k+1}, x^k) \geq 0 \ \forall k$ (cf. (3.7)). Then by summing (3.8), we get (3.9)–(3.10) from $\sum_{k=1}^{\infty} \Delta_q^k = \lim_{k \to \infty} q(p^k) - q(p^1) \leq \inf_X f - q(p^1)$, using $\sup_P q \leq \inf_X f$. In fact for $x \in X$, $\left\langle p^k, Ax^k - b \right\rangle \leq \left\langle p^k, Ax^k - Ax \right\rangle = -\left\langle A^T p^k, x - x^k \right\rangle$, since $Ax \leq b$ and $p^k \geq 0$ (cf. (2.16)), so by (3.3), (2.11), (2.12) and (3.7),

$$q(p^k) \leq f(x^k) + \left\langle -A^T p^k, x - x^k \right\rangle = f(x) - D_f^k(x, x^k) \leq f(x). \qquad \square$$

Lemma 3.5. *$\{x^k\}$ is bounded and $\{x^k\} \subset \mathcal{L}_f^1(x, f(x) - q(p^1)) \ \forall x \in X$.*

Proof. Let $x \in X$. Since $\{q(p^k)\}$ is nondecreasing, (3.11) yields $D_f^\flat(x, x^k) \leq f(x) - q(p^1)$ for all k, so by (3.4), $x^k \in \mathcal{L}_f^1(x, f(x) - q(p^1))$, a bounded set by Definition 2.1(c). $\square$

Lemma 3.6. *$\{f(x^k)\}$ is bounded and every limit point of $\{x^k\}$ is in C_f.*

Proof. Let $x \in X$. By (3.1), (3.7) and (3.11), $f(x^k) + f'(x^k, x - x^k) \geq f(x) - D_f^k(x, x^k) \geq q(p^k) \geq q(p^1)$ for all k, so the desired conclusion follows from the continuity of f on C_f (cf. Definition 2.1(b)), $\{x^k\} \subset C_f$ (cf. (3.4)) and Lemmas 3.5 and 3.2. $\square$

Lemma 3.7. $x^{k+1} - x^k \to 0$.

Proof. If the assertion does not hold, then (since $\{x^k\}$ is bounded; cf. Lemma 3.5) there exists a subsequence K such that $\{x^k\}_{k \in K}$ and $\{x^{k+1}\}_{k \in K}$ converge to some x^∞ and $x^\infty + z$ respectively with $z \neq 0$. By Lemma 3.6, $x^\infty \in C_f$ and $x^\infty + z \in C_f$. By (3.1), (3.7) and (3.8), $\Delta_q^k \geq \kappa_D[f(x^{k+1}) - f(x^k) - f'(x^k; x^{k+1} - x^k)]$, so from (3.9), the continuity of f on C_f (cf. Definition 2.1(b)) and Lemma 3.1(b), we get $0 = \liminf_{k \in K} \Delta_q^k \geq \kappa_D[f(x^\infty + z) - f(x^\infty) - f'(x^\infty; x^\infty + z)]$, contradicting the strict convexity of f. $\square$

Lemma 3.8. $r^{k+1} - r^k \to 0$ *and* $q_k'(t_k) - q_k'(0) \to 0$.

Proof. We have $r^{k+1} - r^k = A(x^{k+1} - x^k)$ by (3.2), and $q_k'(t_k) - q_k'(0) = r_{i_k}^{k+1} - r_{i_k}^k$ by (3.5)-(3.6), so the desired conclusion follows from $x^{k+1} - x^k \to 0$ (cf. Lemma 3.7). $\square$

Lemma 3.9. $[p_{i_k}^k + r_{i_k}^k]_+ - p_{i_k}^k \to 0$.

Proof. If the claim does not hold, there exist $\epsilon > 0$ and an infinite $K \subset \{1,2,\ldots\}$ such that $|[p_{i_k}^k + r_{i_k}^k]_+ - p_{i_k}^k| \geq \epsilon\ \forall k \in K$. Thus for each $k \in K$, either (a) $r_{i_k}^k \geq \epsilon$ or (b) $r_{i_k}^k \leq -\epsilon$ and $p_{i_k}^k \geq \epsilon$, where $r_{i_k}^k = q_k'(0)$ by (3.5). Using (3.9) and Lemma 3.8, pick $\hat{k}$ such $\Delta_q^k < (1-\omega_{\min})\epsilon^2$ and $|q_k'(0) - q_k'(t_k)| < \omega_{\min}\epsilon\ \forall k \geq \hat{k}$. Let $k \in K$, $k \geq \hat{k}$. Since $|q_k'(0)| \geq \epsilon$, $\omega_k < \omega_{\min}$ (cf. (2.15)). Hence case (a) cannot occur, and for case (b) Step 4(ii) sets $t_k = -p_{i_k}^k$. Thus $t_k \leq -\epsilon$ and $q_k'(t_k) < (1-\omega_{\min})q_k'(0) \leq -(1-\omega_{\min})\epsilon$, so, since $q_k'(\cdot)$ is nonincreasing, $q(p^{k+1}) - q(p^k) = \int_0^{t_k} q_k'(\tau)d\tau \geq (1-\omega_{\min})\epsilon^2$, a contradiction. $\square$

Lemma 3.10. $\{x^k\}$ *converges to some* $x^\infty \in C_f$.

Proof. We first show that for all k,

$$D_f^{k+1}(x, x^{k+1}) + D_f^k(x^{k+1}, x^k) - D_f^k(x, x^k) = t_k \left\langle a^{i_k}, x - x^{k+1} \right\rangle \quad \forall x \in C_f. \tag{3.12}$$

By (2.12), the left side equals $\left\langle g^k - g^{k+1}, x - x^{k+1} \right\rangle$, where $g^k - g^{k+1} = A^T(p^{k+1} - p^k) = t_k a^{i_k}$ (cf. (2.11)), since $p^{k+1} = p^k + t_k e^{i_k}$.

Since $\{x^k\}$ is bounded (Lemma 3.5), a subsequence $\{x^{k_j}\}$ converges to some $x^\infty \in C_f$ (cf. Lemma 3.6). Let $I_< = \{i : \langle a^i, x^\infty \rangle < b_i\}$, $I_= = \{i : \langle a^i, x^\infty \rangle = b_i\}$ and $I_> = \{i : \langle a^i, x^\infty \rangle > b_i\}$. Pick $\epsilon > 0$ for $B(x^\infty, \epsilon) = \{x : |x - x^\infty| \leq \epsilon\}$ such that

$$\left\langle a^i, x \right\rangle - b_i < -\epsilon \quad \forall i \in I_<,\ x \in B(x^\infty, \epsilon), \tag{3.13}$$

$$\left\langle a^i, x \right\rangle - b_i > \epsilon \quad \forall i \in I_>,\ x \in B(x^\infty, \epsilon). \tag{3.14}$$

Suppose $\{x^k\}$ does not converge. Then there exists $\epsilon_\infty > 0$ such that, for each j, $x^k \notin B(x^\infty, \epsilon_\infty)$ for some $k > k_j$. Replacing ϵ by $\min\{\epsilon, \epsilon_\infty\}$, for each j such that $x^{k_j} \in B(x^\infty, \epsilon)$ let $k_j' = \min\{k \geq k_j : x^{k+1} \notin B(x^\infty, \epsilon)\}$, so that $x^k \in B(x^\infty, \epsilon)$ for $k \in K^j = [k_j, k_j']$. Summing (3.12) for $x = x^\infty$ and using $D_f^k(x^{k+1}, x^k) \geq 0$ (cf. (3.7)) gives

$$D_f^{k_j'}(x^\infty, x^{k_j'}) \leq D_f^{k_j}(x^\infty, x^{k_j}) + \sum_{k=k_j}^{k_j'-1} t_k \left\langle a^{i_k}, x^\infty - x^{k+1} \right\rangle \quad \forall j. \tag{3.15}$$

We need to show that the sum above vanishes. Let $K_<^j = \{k \in K^j : i_k \in I_<\}$, $K_=^j = \{k \in K^j : i_k \in I_=\}$ and $K_>^j = \{k \in K^j : i_k \in I_>\}$. Since $p_{i_k}^k \geq 0\ \forall k$, Lemma 3.9 yields $\limsup_{k\to\infty} r_{i_k}^k \leq 0$, where $r_{i_k}^k = \left\langle a^{i_k}, x^k \right\rangle - b_{i_k}$ (cf. (3.5)), so there exists $j_>$ such that, for all $j \geq j_>$ and $k \geq k_j$, $r_{i_k}^k < \epsilon$ and $K_>^j = \emptyset$ (otherwise $i_k \in I_>$ and $x^k \in B(x^\infty, \epsilon)$ would give $r_{i_k}^k > \epsilon$ by (3.14), a contradiction). Since $q_k'(t_k) - q_k'(0) \to 0$ (Lemma 3.8), there exists $j_< \geq j_>$ such that $q_k'(t_k) < q_k'(0) + \epsilon/2$ for all $k \geq k_j$,

$j \geq j_<$. Then for $j \geq j_<$ and $k = k_j : k_j' - 1$, $q_k'(0) = \left\langle a^{i_k}, x^k \right\rangle - b_{i_k} < -\epsilon$ (cf. (3.5) and (3.13)) and $q_k'(t_k) \leq -\epsilon/2$ yield $|t_k \left\langle a^{i_k}, x^\infty - x^{k+1} \right\rangle| \leq |t_k||a^{i_k}|\epsilon \leq 2|a^{i_k}||t_k q_k'(t_k)|$, using $x^{k+1} \in B(x^\infty, \epsilon)$. For each $k \in K_=^j$, $\langle a^{i_k}, x^\infty \rangle = b_{i_k}$, so $q_k'(t_k) = \left\langle a^{i_k}, x^{k+1} - x^\infty \right\rangle$ (cf. (3.6)) and $t_k \left\langle a^{i_k}, x^\infty - x^{k+1} \right\rangle = -t_k q_k'(t_k)$. Combining the preceding relations and using $K_j = K_<^j \cup K_=^j \cup K_>^j$ and $\sum_{k=1}^\infty |t_k q_k'(t_k)| < \infty$ (cf. (3.10)) yields

$$\sum_{k=k_j}^{k_j'-1} t_k \left\langle a^{i_k}, x^\infty - x^{k+1} \right\rangle \leq (1 + 2\epsilon \max_i |a^i|) \sum_{k=k_j}^{\infty} |t_k q_k'(t_k)| \to 0 \quad \text{as} \quad j \to \infty. \tag{3.16}$$

Suppose $x^\infty \neq x$ for some $x \in X$. Using $x^{k_j} \to x^\infty \in C_f$, Definition 2.1(d) and Lemma 3.5, we get $D_f^\sharp(x^\infty, x^{k_j}) \to 0$ and $D_f^{k_j}(x^\infty, x^{k_j}) \to 0$ from (3.7). Then (3.15)–(3.16) yield $D_f^{k_j'}(x^\infty, x^{k_j'}) \to 0$. Since $D_f'(x^\infty, x^{k_j'}) \to 0$ (cf. (3.7)), $x^\infty \in C_f$, $\{x^{k_j'}\}$ is bounded in C_f, and f is strictly convex and continuous on C_f (cf. Definition 2.1(a,b)), Lemma 3.3 yields $x^{k_j'} \to x^\infty$. Then $x^{k+1} - x^k \to 0$ (cf. Lemma 3.7) implies $x^{k_j'+1} \to x^\infty$, contradicting the fact $x^{k_j'+1} \notin B(x^\infty, \epsilon)$ for all j. Next, suppose $X = \{x^\infty\}$. Since $\{x^{k_j'}\}$ is bounded, we may assume without loosing generality that it converges to some x'. But $x' \neq x^\infty$ (since $x^{k_j'+1} \notin B(x^\infty, \epsilon)\ \forall j$), so by replacing $\{x^{k_j}\}$ and x^∞ with $\{x^{k_j'}\}$ and x' respectively and using $x = x^\infty \neq x'$ in the preceding argument, we again get a contradiction. Hence $x^k \to x^\infty$. □

Lemma 3.11. *$\{x^k\}$ converges to some $x^\infty \in X$, $[p^k + r^k]_+ - p^k \to 0$, $r^k \to r^\infty = Ax^\infty - b$, $\left\langle p^k, r^\infty \right\rangle \to 0$ and*

$$\left\langle p^k, Ax^k - b \right\rangle = \left\langle A^T p^k, x^k - x^\infty \right\rangle + \left\langle p^k, r^\infty \right\rangle \quad \forall k. \tag{3.17}$$

Proof. By Lemma 3.10, $x^k \to x^\infty \in C_f$. By (3.2), $r^k \to r^\infty = Ax^\infty - b$. For any $i \in \{1{:}m\}$, $K = \{k : i_k = i\}$ is infinite by Assumption 2.11. Since $\lim_{k \in K} r_i^k \leq \limsup_{k \to \infty} r_{i_k}^k \leq 0$ (cf. Lemma 3.9), $r_i^\infty \leq 0$. If $r_i^\infty < 0$ then $[p_{i_k}^k + r_{i_k}^k]_+ - p_{i_k}^k \to 0$ (Lemma 3.9) and $p^k \geq 0\ \forall k$ yield $p_i^k \xrightarrow{K} 0$, so in fact $p_i^k \to 0$ because $p_i^{k+1} = p_i^k$ if $i_k \neq i$, and hence $[p_i^k + r_i^k]_+ - p_i^k \to 0$. Similarly, $[p_i^k + r_i^k]_+ - p_i^k \to 0$ if $r_i^\infty = 0$. Since i was arbitrary, $r^\infty \leq 0$ (i.e., $Ax^\infty \leq b$ and $x^\infty \in X$), $[p^k + r^k]_+ - p^k \to 0$, $p_i^k \to 0$ if $r_i^\infty < 0$ and $\left\langle p^k, r^\infty \right\rangle = \sum_{r_i^\infty < 0} r_i^\infty p_i^k \to 0$. Since

$$\left\langle p^k, Ax^k - b \right\rangle = \left\langle p^k, Ax^k - Ax^\infty \right\rangle + \left\langle p^k, Ax^\infty - b \right\rangle = \left\langle A^T p^k, x^k - x^\infty \right\rangle + \left\langle p^k, r^\infty \right\rangle,$$

we have (3.17). □

We may now prove our main convergence result.

Theorem 3.12. (a) *Problem* (1.1) *has a unique solution, say x^*, and $x^k \to x^*$.*

(b) $q(p^k) \uparrow \sup_P q = \min_X f$, *i.e.,* strong duality *holds, under either of the following conditions:*

(i) $X \neq \{x^*\}$;

(ii) *Condition* (2.5) *holds, i.e.,* $C_{\partial f} \supset \{y^k\} \to y^* \in C_f \Rightarrow D_f^\sharp(y^*, y^k) \to 0$;

(iii) $\mathrm{cl}\, C_f$ *is a polyhedral set and there exist* $\epsilon > 0$ *and* $\alpha \in \mathbb{R}$ *such that* $\sigma_{\partial f(y)}((x - y)/|x-y|) \le \alpha$ *for all* x, y *in* $C_f \cap B(x^*, \epsilon)$, *where* $B(x^*, \epsilon) = \{x : |x - x^*| \le \epsilon\}$;

(iv) $\mathrm{cl}\, C_f$ *is a polyhedral set and there exist* $\epsilon > 0$ *and* $\alpha \in \mathbb{R}$ *such that* $|f'(y; (x - y)/|x-y|)| \le \alpha\ \forall x, y \in C_f \cap B(x^*, \epsilon)$.

(c) *Every limit point of* $\{p^k\}$ *(if any) solves the dual problem* (2.6). *In particular, if Slater's condition holds, i.e.,* $A\hat{x} < b$ *for some* $\hat{x} \in C_f$, *then* $\{p^k\}$ *is bounded and* $q(p^k) \uparrow \max_P q = \min_X f$.

Proof. (a) By Lemma 3.11, $\{x^k\}$ converges to some $x^\infty \in X$ and $\langle p^k, r^\infty \rangle \to 0$ in (3.17). Hence (3.3), (2.11), (2.12), $x^\infty \in C_f$ and (2.1b)–(2.2) imply

$$\begin{aligned} q(p^k) &= f(x^k) + \langle -A^T p^k, x^\infty - x^k \rangle + \langle p^k, r^\infty \rangle \\ &= f(x^\infty) - D_f^k(x^\infty; x^k) + \langle p^k, r^\infty \rangle \\ &\ge f(x^\infty) - D_f^\sharp(x^\infty; x^k) + \langle p^k, r^\infty \rangle \to f(x^\infty) \end{aligned} \tag{3.18}$$

if $X \neq \{x^\infty\}$, since $D_f^\sharp(x^\infty; x^k) \xrightarrow{K} 0$ from Definition 2.1(d) and Lemma 3.5 with $x \in X \setminus \{x^\infty\}$. Then $f(x^\infty) \le \lim_{k\to\infty} q(p^k) \le \sup_P q \le \inf_X f$ and $x^\infty \in X$ yield $x^\infty \in \operatorname{Arg\,min}_X f$ and $q(p^k) \uparrow \sup_P q = \min_X f$. Otherwise $X = \{x^\infty\}$. Because the solution of (1.1) is unique by the strict convexity of f (Definition 2.1(a)), $x^\infty = x^*$.

(b) As shown above, condition (i) yields (3.18), which also holds under condition (ii). Condition (iv) implies (iii) (cf. (1.2)). Hence (3.18) will yield the desired conclusion as in part (a) if we show that $D_f^\sharp(x^\infty; x^k) \to 0$ under condition (iii). Let $y^k = 2x^k - x^\infty$, so that $y^k - x^k = x^k - x^\infty$. By Lemma 3.1(a), we may shrink ϵ to ensure that $C_f \cap B(x^\infty, \epsilon)$ is closed. Since $\mathrm{cl}\, C_f$ is a polyhedral set and $x^k \to x^\infty$ in C_f, there exists $\mu \in (0, \epsilon]$ such that for all large k, $y^k = 2(x^k - x^\infty) + x^\infty \in T \cap B(0, \mu) + x^\infty \subset C_f \cap B(x^\infty, \epsilon)$, where T denotes the tangent cone of C_f at x^∞. Then $\sigma_{\partial f(x^k)}((x^k - x^\infty)/|x^k - x^\infty|) = \sigma_{\partial f(x^k)}((y^k - x^k)/|y^k - x^k|) \le \alpha$ and $\sigma_{\partial f(x^k)}(x^k - x^\infty) \le \alpha |x^k - x^\infty|$ yield $\limsup_{k\to\infty} D_f^\sharp(x^\infty, x^k) \le f(x^\infty) - \liminf_{k\to\infty} f(x^k) \le 0$ (cf. (2.1b), $x^k \to x^\infty$ and closedness of f), so (cf. (3.7)) $D_f^\sharp(x^\infty; x^k) \to 0$ as desired.

(c) Suppose a subsequence $\{p^k\}_{k\in K}$ converges to some p^∞. By (2.16), $p^\infty \in P$. Then $-A^T p^\infty \in \partial f(x^\infty)$ from $-A^T p^k \in \partial f(x^k)$ (cf. (2.11)), i.e., $f(x) \ge f(x^k) - \langle A^T p^k, x - x^k \rangle$ $\forall x$ with $x^k \to x^\infty$, f being closed and $A^T p^k \to A^T p^\infty$. Thus $p^\infty \in \mathring{C}_q$ by Lemma 2.9. Since $[p^k + r^k]_+ - p^k \to 0$ (cf. Lemma 3.11), p^∞ satisfies the optimality condition (2.10) for (2.6). Under Slater's condition, the set P^* of Kuhn-Tucker multipliers of (1.1) is nonempty and bounded, $P^* = \operatorname{Arg\,max}_P q$ and $\max_P q = \min_X f$ (cf. [GoT89, Thm 1.3.5] or [Roc70, Cor. 29.1.5]). Hence $\{p \in P : q(p) \ge q(p^1)\}$ is bounded (cf. [Roc70, Cor. 8.7.1]), and so is $\{p^k\}$, since $\{q(p^k)\}$ is nondecreasing. Thus $q(p^k) \uparrow f(x^*)$. □

4 Bregman's projections and overrelaxation

We now relate Bregman's projections [Bre67, CeL81] with exact linesearches. Let $H^k = \{x : \langle a^{i_k}, x\rangle = b_{i_k}\}$. By (2.13) and (2.9),

$$q_k'(t) = \left\langle a^{i_k}, x(p^k(t))\right\rangle - b_{i_k} \quad \text{if} \quad p^k(t) \in \mathring{C}_q. \tag{4.1}$$

We say that $\tilde{x}^{k+1}$ is the *$D_f^k(\cdot, x^k)$-projection* (cf. (2.12)) of x^k on H^k with *parameter* $\tilde{t}_k$ if

$$\tilde{x}^{k+1} = \arg\min\{D_f^k(x, x^k) : x \in H^k\} = \arg\min\{f(x) + \left\langle A^T p^k, x\right\rangle : \left\langle a^{i_k}, x\right\rangle = b_{i_k}\} \tag{4.2}$$

and $\tilde{t}_k$ is the Kuhn-Tucker multiplier of (4.2). In other words, since $A^T p^k + \tilde{t}_k a^{i_k} = A^T p^k(\tilde{t}_k)$,

$$\tilde{x}^{k+1} = \arg\min_x\{f(x) + \left\langle A^T p^k(\tilde{t}_k), x\right\rangle\} \quad \text{and} \quad \left\langle a^{i_k}, \tilde{x}^{k+1}\right\rangle = b_{i_k}. \tag{4.3}$$

Lemma 4.1. (a) *If* (4.3) *holds then* $p^k(\tilde{t}_k) \in \mathring{C}_q$ *and* $q_k'(\tilde{t}_k) = 0$, *i.e.,* $\tilde{t}_k$ *maximizes* q_k.

(b) *If* $q_k'(\tilde{t}) = 0$ *for some* $\tilde{t}$ *then* (4.3) *holds with* $\tilde{t}_k = \tilde{t}$ *and* $\tilde{x}^{k+1} = x(p^k(\tilde{t}_k))$.

(c) *If* $q_k'(0) > 0$ (< 0) *and* $q_k'(t) < 0$ (> 0) *for some* t *then* (4.3) *is well defined for some* $\tilde{t}_k > 0$ $(< 0$ respectively$)$.

(d) *For any* $\alpha \in \mathbb{R}$, *the level set* $\{x : D_f^k(x, x^k) \le \alpha\}$ *is bounded.*

(e) *If* $H^k \cap C_f \neq \emptyset$ *then* $\tilde{x}^{k+1}$ *is well defined by* (4.2).

(f) *If* $H^k \cap \operatorname{ri} C_f \neq \emptyset$ *then* $\tilde{x}^{k+1}$ *is well defined by* (4.2) *and* (4.3) *holds for some* $\tilde{t}_k$.

(g) *If* $H^k \cap \operatorname{ri} C_f \neq \emptyset$ *and* $C_{\partial f} = \operatorname{ri} C_f$ *(e.g., f is essentially smooth) then* $\tilde{x}^{k+1} \in \operatorname{ri} C_f$.

Proof. If (4.3) holds then $-A^T p^k(\tilde{t}_k) \in \partial f(\tilde{x}^{k+1})$ [Roc70, Thm 23.5], so $p^k(\tilde{t}_k) \in \mathring{C}_q$ and $\tilde{x}^{k+1} = x(p^k(\tilde{t}_k))$ by Lemma 2.9 and $q_k'(\tilde{t}_k) = 0$ by (4.1). Similarly, (b) and (c) follow from Lemma 2.9, (4.1), the monotonicity and continuity of q_k' and the strict convexity of f. As for (d), by (2.11), (2.12) and the strict convexity of f, $\{x : D_f^k(x, x^k) \le 0\} = \{x^k\}$, so $D_f^k(\cdot, x^k)$ has bounded level sets by [Roc70, Cor. 8.7.1]. Then (e) follows from the lower semicontinuity of f and $D_f^k(\cdot, x^k)$, (f) from [Roc70, Thm 28.2] and (g) from [Roc70, Thm 28.3]. □

Remark 4.2. The proof of (d) above shows that the requirement on $\mathcal{L}_h^2(\tilde{y}, \alpha)$ in condition (iv) of Definition 2.2 is a *consequence* of conditions (i)–(iii) (since $\mathcal{L}_h^2(\tilde{y}, 0) = \{\tilde{y}\}$). This fact, implicit in Bregman's original work [Bre67], has been ignored in its follow-up [CeL81].

We now turn to overrelaxation. It follows from (3.8) that

$$\begin{aligned}\Delta_q^k \ge \kappa_D D_f^k(x^{k+1}, x^k) \quad &\Longleftrightarrow \quad t_k q_k'(t_k) \ge (\kappa_D - 1) D_f^k(x^{k+1}, x^k)\\ &\Longleftrightarrow \quad (\kappa_D - 1)\Delta_q^k \le \kappa_D t_k q_k'(t_k).\end{aligned} \tag{4.4}$$

Depending on which quantities are computed, any of these conditions can be used at Step 4 (cf. (2.14)). The third condition occured in the quite abstract framework of [Tse90], where the case of a quadratic f required considerable additional effort. Generalizing an idea from [DPI86], we now give another useful condition based on the following

Lemma 4.3. *For all k, $\omega_k\Delta_q^k = (1-\omega_k)D_f^{k+1}(x^k, x^{k+1}) + D_f^k(x^{k+1}, x^k)$.*

Proof. By (2.15), $\omega_k q_k'(t_k) = (1-\omega_k)[q_k'(0) - q_k'(t_k)]$, and by (3.5), (3.6), (2.11) and (2.12),

$$\begin{aligned} t_k[q_k'(0) - q_k'(t_k)] &= t_k \left\langle a^{i_k}, x^k - x^{k+1} \right\rangle = \left\langle A^T p^{k+1} - A^T p^k, x^k - x^{k+1} \right\rangle \\ &= D_f^{k+1}(x^k, x^{k+1}) + D_f^k(x^{k+1}, x^k), \end{aligned}$$

so

$$\begin{aligned} \omega_k \Delta_q^k &= \omega_k t_k q_k'(t_k) + \omega_k D_f^k(x^{k+1}, x^k) = (1-\omega_k) t_k [q_k'(0) - q_k'(t_k)] + \omega_k D_f^k(x^{k+1}, x^k) \\ &= (1-\omega_k)[D_f^{k+1}(x^k, x^{k+1}) + D_f^k(x^{k+1}, x^k)] + \omega_k D_f^k(x^{k+1}, x^k), \end{aligned}$$

where the first equality follows from (3.8). □

Let $\omega_{\max} \in (1,2)$, $\epsilon_D \in [0, \frac{2-\omega_{\max}}{\omega_{\max}-1})$ and $\epsilon_d = 1 + (1+\epsilon_D)(1-\omega_{\max})$, so that $\epsilon_d > 0$. Condition (2.14) may be replaced by

$$D_f^{k+1}(x^k, x^{k+1}) \le (1+\epsilon_D) D_f^k(x^{k+1}, x^k) \quad \text{if} \quad \omega_k > 1, \tag{4.5}$$

since

$$2\Delta_q^k \ge \omega_k \Delta_q^k \ge [(1-\omega_k)(1+\epsilon_D) + 1] D_f^k(x^{k+1}, x^k) \ge \epsilon_d D_f^k(x^{k+1}, x^k) \quad \text{if} \quad \omega_k > 1 \tag{4.6}$$

by Lemma 4.3 and the choice of ϵ_d, so (2.14) holds with $\kappa_D = \epsilon_d/2$, as required for convergence. If f is a strictly convex quadratic function then $D_f(x,y) = D_f(y,x)$ (cf. (2.3)) for all x and y [DPI86], whereas by (2.11) and (2.12), $D_f^k(\cdot, x^k) = D_f(\cdot, x^k)$. Thus in the quadratic case conditions (4.5) and (4.4) hold *automatically* (for some $\epsilon_d, \kappa_D > 0$).

5 Convergence for essentially smooth objectives

Generalizing the analysis in [Tse91, LuT92b], let us now replace Assumption 2.8 by

Assumption 5.1. (i) f is closed, proper and essentially strictly convex.
(ii) $C_{\partial f} = \mathring{C}_f$.
(iii) $\mathring{C}_{f^*} \supset \{y^k\} \to y^* \in \operatorname{bd} \mathring{C}_{f^*} \Rightarrow f^*(y^k) \to \infty$.
(iv) The feasible set $X = \{x \in C_f : Ax \le b\}$ of (1.1) is nonempty.
(v) $p^1 \in P$ is such that $-A^T p^1 \in \operatorname{im} \partial f$ and the set $A^T \mathcal{L}_q$ is bounded, where $\mathcal{L}_q = \{p \in P \cap \mathring{C}_q : q(p) \ge q(p^1)\}$.

If f is essentially smooth then (i) yields (ii) (cf. [Roc70, Thm 26.1]). In general, (i) implies that f^* is essentially smooth (cf. [Roc70, Thm 26.3]), so $C_{\partial f^*} = \mathring{C}_{f^*}$ (cf. [Roc70, Thm 26.1]) and ∇f^* is continuous on $\mathring{C}_{f^*}$ (cf. [Roc70, Thm 25.5]). Since f^* is lower semicontinuous, (iii) holds if $C_{f^*} = \mathring{C}_{f^*}$, and conversly (iii) yields $C_{f^*} = \mathring{C}_{f^*}$ (otherwise let $y^* \in C_{f^*} \setminus \mathring{C}_{f^*}$ and $y \in \mathring{C}_{f^*}$ to get (cf. [Roc70, Thm 6.1 and Cor. 7.5.1]) $\lim_{t\uparrow 1} f^*((1-t)y + ty^*) = f^*(y^*) < \infty$, contradicting (iii)). Thus under Assumption 5.1, Lemma 2.9 holds with $C_q = \mathring{C}_q$ and conditions (ii)-(iii) of Assumption 2.8 hold, but f need not satisfy conditions (a)-(d) of Definition 2.1. Yet the proofs of Lemmas 3.5–3.7 and 3.10 and Theorem 3.12 can be modified by using the following results.

Lemma 5.2 ([Tse91, Lemma 8.1]). *Let h be a proper convex function on $\mathbb{R}^m$, E be an $n \times m$ matrix, c be an m-vector, and $\mathcal{P}$ be a convex polyhedral set in $\mathbb{R}^m$. Let $\tilde{q}(p) = h(Ep) + \langle c, p\rangle \ \forall p \in \mathbb{R}^m$. Suppose $\inf_{\mathcal{P}} \tilde{q} > -\infty$ and the set $\{Ep : p \in \mathcal{P}, \tilde{q}(p) \leq \xi\}$ is bounded for each $\xi \in \mathbb{R}$. Then for any $\xi \in \mathbb{R}$ such that the set $\{p \in \mathcal{P} : \tilde{q}(p) \leq \xi\}$ is nonempty, the functions $p \rightarrow h(Ep)$ and $p \rightarrow \langle c, p\rangle$ are bounded on this set.*

Proof. This follows from the proof of [Tse91, Lem. 8.1]. □

Lemma 5.3. *If Assumption 5.1 holds, then:*

(a) $p^k \in \mathcal{L}_q$, $-A^T p^k \in (-A^T \mathcal{L}_q) \subset \mathring{C}_{f^*}$ *and* $x^k = \nabla f^*(-A^T p^k) \in \mathring{C}_f \ \forall k$.

(b) $\{A^T p^k\}$, $\{f^*(-A^T p^k)\}$ *and* $\{\langle b, p^k\rangle\}$ *are bounded.*

(c) $\{x^k\}$ *is bounded.*

(d) $\{f(x^k)\}$ *is bounded and every limit point of* $\{x^k\}$ *is in* $\mathring{C}_f$. *Moreover, if a subsequence* $\{x^k\}_{k\in L}$ *converges to some* x^∞ *then* $x^\infty \in \mathring{C}_f$, $f(x^k) \xrightarrow{L} f(x^\infty)$ *and* $D_f^\sharp(x^\infty, x^k) \xrightarrow{L} 0$.

(e) $x^{k+1} - x^k \rightarrow 0$.

(f) *The set* $P^* = \operatorname{Arg\,max}_P q$ *is nonempty and* $P^* \subset \mathring{C}_q$.

Proof. (a) Since $\{q(p^k)\}$ is nondecreasing, this follows from $p^k \in P \cap \mathring{C}_q$ (cf. (2.16)), Lemma 2.9 with $-A^T p^k \in \partial f(x^k)$ for all k and Assumption 5.1(ii).

(b) Apply Lemma 5.2 to $-q$, using part (a) and $\sup_P q \leq \inf_X f < \infty$.

(c) Let $y^k = -A^T p^k$, so that $x^k = \nabla f^*(y^k) \ \forall k$. Suppose $|x^k| \xrightarrow{L} \infty$ for a subsequence L of $\{1, 2, \ldots\}$. Since $\{y^k\}$ is bounded (cf. part (b)), there exist y^∞ and a subsequence K of L such that $y^k \xrightarrow{K} y^\infty$. Since $\{y^k\} \subset \mathring{C}_{f^*}$ (cf. part (a)) and $\{f^*(y^k)\}$ is bounded (cf. part (b)), $y^\infty \in \mathring{C}_{f^*}$ (cf. Assumption 5.1(iii)). But ∇f^* is continuous on $\mathring{C}_{f^*}$, so $x^k \xrightarrow{K} x^\infty = \nabla f^*(y^\infty)$, contradicting $|x^k| \xrightarrow{K} \infty$. Hence $\{x^k\}$ is bounded.

(d) Let $x^k \xrightarrow{L} x^\infty$. As in part (c), we deduce that $x^\infty = \nabla f^*(y^\infty)$ with $y^\infty \in \mathring{C}_{f^*}$. Thus $y^\infty \in \partial f(x^\infty)$ (cf. [Roc70, Thm 23.5]), so $x^\infty \in \mathring{C}_f$ (cf. Assumption 5.1(ii)). Invoking Lemma 2.3(a), we get $f(x^k) \xrightarrow{L} f(x^\infty)$ and $D_f^\sharp(x^\infty, x^k) \xrightarrow{L} 0$. Hence, since $\{x^k\}$ is bounded, so is $\{f(x^k)\}$ (otherwise we could get a contradiction as in part (c)).

(e) If the assertion does not hold, then (cf. parts (c)–(d)) there exists a subsequence K such that $\{x^k\}_{k\in K}$ and $\{x^{k+1}\}_{k\in K}$ converge to some $x^\infty \in \mathring{C}_f$ and $x^\infty + z \in \mathring{C}_f$ respectively with $z \neq 0$, $f(x^k) \xrightarrow{K} f(x^\infty)$ and $f(x^{k+1}) \xrightarrow{K} f(x^\infty + z)$. Then $x^{k+1} - x^k \xrightarrow{K} z$ yields $\limsup_{k\in K} f'(x^k; x^{k+1} - x^k) \leq f'(x^\infty; z)$ (cf. [Roc70, Thm 24.5]). But (cf. (3.1), (3.7) and (3.8)) $\Delta_q^k \geq \kappa_D[f(x^{k+1}) - f(x^k) - f'(x^k; x^{k+1} - x^k)]$, so (cf. (3.9)) $0 = \liminf_{k\in K} \Delta_q^k \geq \kappa_D[f(x^\infty + z) - f(x^\infty) - f'(x^\infty; x^\infty + z)]$ contradicts strict convexity of f on $\mathring{C}_f \subset C_{\partial f}$ (cf. Assumption 5.1(a)).

(f) This follows from [Tse91, p. 429]. □

Parts (c)-(e) of Lemma 5.3 subsume Lemmas 3.5–3.7.

Proof of Lemma 3.10 under Assumption 5.1. We only modify the argument following (3.16). By Lemma 5.3(d), $x^{k_j} \rightarrow x^\infty \in \mathring{C}_f$ yields $D_f^\sharp(x^\infty, x^{k_j}) \rightarrow 0$. Then $D_f'(x^\infty, x^{k_j'}) \rightarrow 0$

as before. Suppose a subsequence $\{x^{k'_j}\}_{j\in J}$ converges to some $x' \neq x^\infty$. By Lemma 5.3(d), $f(x^{k'_j}) \xrightarrow{J} f(x')$ and $x' \in \mathring{C}_f$. Then $\limsup_{j\in J} f'(x^{k'_j}; x^\infty - x^{k'_j}) \leq f'(x'; x^\infty - x')$ (cf. [Roc70, Thm 24.5]), so (cf. (3.1)) $0 = \liminf_{j\in J} D'_f(x^\infty, x^{k'_j}) \geq f(x^\infty) - f(x') - f'(x'; x^\infty - x')$ with $x^\infty \in \mathring{C}_f$ and $x' \in \mathring{C}_f$ contradict strict convexity of f on $\mathring{C}_f \subset C_{\partial f}$ (cf. Assumption 5.1(a)). Therefore, $x^{k'_j} \to x^\infty$. The rest of the proof goes as before. □

Theorem 5.4. *If Assumption* 5.1 *holds, then:*

(a) *Problem* (1.1) *has a unique solution, say* x^*, *in* $\mathring{C}_f$ *and* $x^k \to x^*$.

(b) $q(p^k) \uparrow \max_P q = \min_X f$.

(c) *Every limit point of* $\{p^k\}$ *(if any) solves the dual problem* (2.6). *In particular, if Slater's condition holds then* $\{p^k\}$ *is bounded and* $q(p^k) \uparrow \max_P q = \min_X f$.

Proof. (a) Apply Lemma 5.3(d) to (3.18) in the proof of Theorem 3.12(a).

(b) Proceed as for part (a) and invoke Lemma 5.3(f).

(c) Use the proof of Theorem 3.12(c). □

Remark 5.5. Assumption 5.1 holds if f is closed proper essentially strictly convex and essentially smooth, $C_{f^*} = \mathring{C}_{f^*}$, $\mathring{C}_f \cap \hat{X} \neq \emptyset$, where $\hat{X} = \{x : Ax \leq b\}$, and the set $\{x \in \hat{X} : f(x) \leq \alpha\}$ is bounded for all $\alpha \in \mathbb{R}$. This can be shown as in [Tse91, p. 440]. Also Assumption 5.1 holds if $C_{f^*} = \mathring{C}_{f^*}$, f^* is strictly convex differentiable on C_{f^*}, $\operatorname{Arg\,min}_X f \neq \emptyset$ and $\hat{X} \cap \operatorname{ri} C_f \neq \emptyset$. This follows from the analysis in [LuT92b] (use Lemma 5.2 instead of [LuT92b, Lem. 3.3] and observe that $C_{\partial f} = \mathring{C}_f$).

6 Convergence under a regularity condition

Let us now replace Assumption 2.8 by the following

Assumption 6.1. (i) f satisfies conditions (a)–(c) of Definition 2.1.
(ii) $\hat{X} \cap \operatorname{ri} C_f \neq \emptyset$, where $\hat{X} = \{x : Ax \leq b\}$.
(iii) $(-A^T P) \cap \operatorname{im} \partial f \neq \emptyset$, where $P = \mathbb{R}^m_+$.

Condition (ii) is stronger than Assumption 2.8(ii), but f need not satisfy condition (d) of Definition 2.1. To modify the proofs of Lemma 3.10 and Theorem 3.12, we shall need

Lemma 6.2. *Let* $h : \mathbb{R}^n \to (-\infty, \infty]$ *be a closed proper convex function continuous on* C_h. *If* $\{y^k\}$ *is a sequence in* C_h *convergent to some* $y^* \in C_h$ *such that for some* $x \in \operatorname{ri} C_h$, $\alpha \in \mathbb{R}$ *and* $\tilde{g}^k \in \partial h(y^k)$, $h(x) - h(y^k) - \left\langle \tilde{g}^k, x - y^k \right\rangle \leq \alpha$ *for all* k, *then* $h(y^*) - h(y^k) - \left\langle \tilde{g}^k, y^* - y^k \right\rangle \to 0$.

Proof. For all k, let $D_h^k(y, y^k) = h(y) - h(y^k) - \left\langle \bar{g}^k, y - y^k \right\rangle$, where $\bar{g}^k$ is the orthogonal projection of $\tilde{g}^k$ onto the linear span $\mathcal{L}$ of $C_h - x$, so that $y - y^k \in \mathcal{L}$ and $\left\langle \tilde{g}^k, y - y^k \right\rangle = \left\langle \bar{g}^k, y - y^k \right\rangle$ for all $y \in C_h$. Since $D_h^k(\cdot, y^k) \geq 0$ by convexity of h, we need to show that $\limsup_{k\to\infty} D_h^k(y^*, y^k) = 0$. If this does not hold, there exist $\epsilon > 0$ and a subsequence

K of $\{1,2,\ldots\}$ such that $D_h^k(y^*,y^k) \geq \epsilon\ \forall k \in K$. We have $|\bar{g}^k| \xrightarrow{K} \infty$ (otherwise, for some $\beta \in \mathbb{R}$ and a subsequence K' of K, $|\bar{g}^k| \leq \beta$ for all $k \in K'$ and $h(y^k) \to h(y^*)$ (cf. continuity of h on C_h) would yield $\lim_{k\in K'} D_h^k(y^*,y^k) = 0$, a contradiction). Hence we may assume that $\breve{g}^k = \bar{g}^k/|\bar{g}^k|$ converges to some $\breve{g}^\infty$ as $k \xrightarrow{K} \infty$. Clearly, $|\breve{g}^\infty| = 1$ and $\breve{g}^\infty \in \mathcal{L}$ ($\mathcal{L}$ is closed). For any $y \in C_h$, taking the limit of $h(y) \geq h(y^k) + \left\langle \bar{g}^k, y - y^k \right\rangle$ divided by $|\bar{g}^k|$ yields $\langle \breve{g}^\infty, y - y^* \rangle \leq 0$. Similarly, $h(x) - h(y^k) - \left\langle \bar{g}^k, x - y^k \right\rangle \leq \alpha$ for all k yields $\langle \breve{g}^\infty, x - y^* \rangle \geq 0$. Then $x \in C_h$ and $\langle \breve{g}^\infty, y - y^* \rangle \leq 0$ for all $y \in C_h$ imply $\langle \breve{g}^\infty, x - y^* \rangle = 0$ and (since $|\breve{g}^\infty| = 1$ and $\breve{g}^\infty \in \mathcal{L}$) $x \notin \operatorname{ri} C_h$, a contradiction. □

In the proof of Lemma 3.10 (after (3.16)), letting $x \in \hat{X} \cap \operatorname{ri} C_f$ (cf. Assumption 6.1(ii)), use $x^{k_j} \to x^\infty \in C_f$, continuity of f on C_f (cf. Definition 2.1(b)), the fact $D_f^k(x,x^k) \leq f(x) - q(p^1)\ \forall k$ (cf. (3.11)) and Lemma 6.2 (cf. (2.12)) to get $D_f^k(x^\infty, x^{k_j}) \to 0$.

Theorem 6.3. *If Assumption* 6.1 *holds, then:*

(a) *Problem* (1.1) *has a unique solution, say* x^*, *in* $C_{\partial f}$ *and* $x^k \to x^*$.

(b) $q(p^k) \uparrow \max_P q = \min_X f$.

(c) *Every limit point of* $\{p^k\}$ *(if any) solves the dual problem* (2.6). *In particular, if Slater's condition holds then* $\{p^k\}$ *is bounded and* $q(p^k) \uparrow \max_P q = \min_X f$.

Proof. Using the proof of Theorem 3.12(a) and the argument preceding our theorem, we have $x^k \to x^\infty$ and $D_f^k(x^\infty, x^k) \to 0$, so (cf. (3.18)) $q(p^k) \uparrow f(x^\infty)$ yields $x^\infty = x^*$ as before. Since $\hat{X} \cap \operatorname{ri} C_f \neq \emptyset$ (cf. Assumption 6.1(ii)), $x^* \in C_{\partial f}$ and $\max_P q = f(x^*)$ (cf. [Roc70, Cor. 28.3.1 and Cor. 28.4.1]). This yields parts (a) and (b). For part (c), use the proof of Theorem 3.12(c). □

7 Additional remarks

Equality constraints may be handled directly (instead of converting equalities into pairs of inequalities). Consider problem (1.1) with equality constraints $Ax = b$. Then $X = \{x \in C_f : Ax = b\}$ and $P = \mathbb{R}^m$ in Assumption 2.8, (2.6b) is deleted, and (2.10) becomes $r(p) = 0$. Thus p is no longer constrained at Step 4. It is easy to verify the preceding convergence results. In the proof of (3.11), $Ax = b$ yields $\left\langle p^k, Ax^k - b \right\rangle = -\left\langle A^T p^k, x - x^k \right\rangle$ as before. In Lemma 3.9, $r_{i_k}^k \to 0$ can be shown as before by using $\omega_k \in [\omega_{\min}, \omega_{\max}]$ for all k. Then $s^\infty = 0$ in Lemma 3.11. Extension to the case of mixed equality and inequality constraints is straightforward.

Following [TsB91, Tse90], let us now assume that f is closed, proper, strictly convex, continuous on C_f and cofinite. Then Assumption 2.8(ii,iii) and Lemma 2.9 hold with $\operatorname{im} \partial f = C_{f^*} = \mathring{C}_{f^*} = \mathbb{R}^n$ (cf. [Roc70, Thm 25.5]) and $C_q = \mathring{C}_q = \mathbb{R}^m$. Moreover, f satisfies conditions (a)-(b) of Definition 2.1; we now show that condition (c) holds *automatically* by using the following result of [TsB91].

Lemma 7.1 ([TsB91, Lemma 2]). *Let* $h : \mathbb{R}^n \to (-\infty, \infty]$ *be a closed proper convex function continuous on* C_h *and cofinite. If* $\{y^k\}$ *is a sequence in* C_h *such that, for some* $y \in C_h$, $\{h(y^k) + h'(y^k; y - y^k)\}$ *is bounded from below, then* $\{y^k\}$ *and* $\{h(y^k)\}$ *are bounded and any limit point of* $\{y^k\}$ *is in* C_h.

Lemma 7.2. *If f is closed proper convex, continuous on C_f and cofinite, $x \in C_f$ and $\alpha \in \mathbb{R}$, then the sets $\mathcal{L}_f^4(x,\alpha) = \{y \in C_f : D_f'(x,y) \le \alpha\}$ and $\mathcal{L}_f^1(x,\alpha) = \{y \in C_{\partial f} : D_f^\flat(x,y) \le \alpha\}$ are bounded.*

Proof. Note that $\mathcal{L}_f^1(x,\alpha) \subset \mathcal{L}_f^4(x,\alpha)$ (cf. (1.2), (2.1a) and (3.1)) and invoke Lemma 7.1. □

As a corollary we list a simple result which improves on [DPI86, Thm 5.1].

Lemma 7.3. *If $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex and cofinite then f is a B-function satisfying condition* (2.5). *In particular, f is cofinite if* $\lim_{|x|\to\infty} f(x)/|x| = \infty$.

Proof. Condition (a) of Definition 2.1 holds by assumption. Invoke [Roc70, Thm 10.1] for (b), Lemma 7.2 for (c), and Lemma 2.3(a) for (2.5). If $\lim_{|x|\to\infty} f(x)/|x| = \infty$ then $f(x) - \langle x,y\rangle \ge \{f(x)/|x| - |y|\}|x| \to \infty$ as $|x| \to \infty$ and hence $f^*(y) < \infty$ for all y. □

Lemma 7.3 confirms that if f is a strictly convex quadratic function then Theorem 3.12 holds (also under overrelaxation; cf. §4) with $q(p^k) \uparrow \max_P q$ (cf. Remark 5.5).

Remark 7.4. Suppose $f(x) = \sum_{i=1}^n f_i(x_i)$, where each f_i is closed proper strictly convex on $\mathbb{R}$ with $\mathcal{L}_{f_i}^1(t,\alpha)$ bounded $\forall t,\alpha \in \mathbb{R}$. Then f is a B-function (cf. Lemma 2.4(d)) satisfying condition (2.5) (cf. Lemma 2.3(d), $C_f = \prod_{i=1}^n C_{f_i}$ and $D_f^\sharp(x,y) = \sum_{i=1}^n D_{f_i}^\sharp(x_i,y_i)$). In particular, if each f_i is also cofinite then $\mathcal{L}_{f_i}^1(t,\alpha)$ is bounded $\forall t,\alpha \in \mathbb{R}$ (cf. Lemma 7.2), f is cofinite ($f^*(y) = \sum_{i=1}^n f_i^*(y_i)$ $\forall y$) and Assumption 2.8 merely requires that $X \ne \emptyset$, so Theorem 3.12 holds with $q(p^k) \uparrow \sup_P q = \min_X f$ if $X \ne \emptyset$.

8 Block coordinate relaxation

We now consider the block coordinate relaxation (BCR) algorithm of [Tse90, §5.2].

Given the current $p^k \in P \cap \mathring{C}_q$ and $x^k = x(p^k)$, choose a nonempty set $I^k \subset \{1{:}\,m\}$. Let $I_-^k = \{i \in I^k : r_i^k < 0\}$ and $I_+^k = \{i \in I^k : r_i^k > 0\}$, where $r^k = Ax^k - b$. If $I_-^k = I_+^k = \emptyset$, set $p^{k+1} = p^k$ and $x^{k+1} = x^k$. Otherwise, let x^{k+1} be the solution and π_i^k, $i \in I_-^k \cup I_+^k$, be the associated Lagrange multipliers of the following problem with a parameter $\mu \in (0, 1/2]$

$$\text{minimize} \qquad f(x) + \sum_{i \notin I_-^k} p_i^k \left\langle a^i, x\right\rangle, \tag{8.1a}$$

$$\text{subject to} \qquad \left\langle a^i, x\right\rangle \le b_i \quad \forall i \in I_-^k, \tag{8.1b}$$

$$\left\langle a^i, x\right\rangle \le b_i + \mu r_i^k \quad \forall i \in I_+^k. \tag{8.1c}$$

Let

$$p_i^{k+1} = \begin{cases} \pi_i^k & \text{if } i \in I_-^k, \\ p_i^k + \pi_i^k & \text{if } i \in I_+^k, \\ p_i^k & \text{otherwise.} \end{cases} \tag{8.2}$$

Lemma 8.1. *$p^{k+1} \in P \cap \mathring{C}_q$ and $x^{k+1} = x(p^{k+1})$ are well defined,*

$$p_i^{k+1} = [p_i^{k+1} + r_i^{k+1}]_+ \quad \forall i \in I_-^k, \tag{8.3}$$

$$|[p_i^{k+1} + r_i^{k+1}]_+ - p_i^{k+1}| \le |r_i^{k+1} - r_i^k| \quad \forall i \in I^k, \tag{8.4}$$

$$\left\langle r^{k+1}, p^{k+1} - p^k \right\rangle \ge \sum_{i \in I_-^k} p_i^k (b_i - \left\langle a^i, x^{k+1} \right\rangle) \ge 0, \tag{8.5}$$

$$\Delta_q^k = D_f^k(x^{k+1}, x^k) + \left\langle r^{k+1}, p^{k+1} - p^k \right\rangle \ge D_f^k(x^{k+1}, x^k) \ge 0. \tag{8.6}$$

Proof. Let h and X^k denote the objective and the feasible set of (8.1) respectively. Then $C_h = C_f$ and $C_h \cap \mathring{X}^k \neq \emptyset$, since for any $x \in X$, $\lambda \in (0, \mu)$ and $y = (1-\lambda)x + \lambda x^k$, $y \in C_f$ ($x, x^k \in C_f$; cf. (2.8)), $\langle a^i, y \rangle - b_i \le \lambda r_i^k < 0 \ \forall i \in I_-^k$, $\langle a^i, y \rangle - b_i \le \lambda r_i^k < \mu r_i^k \ \forall i \in I_+^k$. By (2.8) and [Roc70, Thm 28.3], $x^k = \arg\min\{h(x) : \langle a^i, x \rangle \le \left\langle a^i, x^k \right\rangle \ \forall i \in I_-^k\}$, where the feasible set and h have no direction of recession in common (otherwise the minimum would be nonunique). Hence h and X^k have no common direction of recession, so $\operatorname{Arg\,min}_{X^k} h \neq \emptyset$ (cf. [Roc70, Thm 27.3]). Since (8.1) is solvable and strictly consistent ($C_h \cap \mathring{X}^k \neq \emptyset$), it has a solution x^{k+1} and Lagrange multipliers π_i^k (cf. [Roc70, Cor. 29.1.5]) satisfying the Kuhn-Tucker conditions (cf. [Roc70, Thm 28.3]) $\pi_{I_-^k}^k = [\pi_{I_-^k}^k + r_{I_-^k}^{k+1}]_+$, $\pi_{I_+^k}^k = [\pi_{I_+^k}^k + r_{I_+^k}^{k+1} - \mu r_{I_+^k}^k]_+$, $0 \in \partial f(x^{k+1}) + \sum_{i \notin I_-^k} p_i^k a^i + \sum_{i \in I_-^k \cup I_+^k} \pi_i^k a^i$, with $r^{k+1} = Ax^{k+1} - b$. Then by (8.2), $p^{k+1} \in P$ and $-A^T p^{k+1} \in \partial f(x^{k+1})$, so $p^{k+1} \in P \cap \mathring{C}_f$ and $x^{k+1} = x(p^{k+1})$ by Lemma 2.9. By the Kuhn-Tucker conditions and (8.2),

$$\left\langle p^{k+1} - p^k, r^{k+1} \right\rangle = \sum_{i \in I_-^k, \pi_i^k = 0} (\pi_i^k - p_i^k)(\left\langle a^i, x^{k+1} \right\rangle - b_i) + \sum_{i \in I_+^k, \pi_i^k > 0} \pi_i^k \mu r_i^k \ge 0,$$

which yields (8.5). The equality in (8.6) follows from the proof of Lemma 3.4. Finally, to prove (8.4), note that $\pi_{I_-^k}^k = [\pi_{I_-^k}^k + r_{I_-^k}^{k+1}]_+$ and (8.2) yield (8.3). Since $p^{k+1} \ge 0$, we also have $|[p_i^{k+1} + r_i^{k+1}]_+ - p_i^{k+1}| \le |r_i^{k+1}| \ \forall i \in I^k$. But $r_i^k = 0$ and $|r_i^{k+1}| = |r_i^{k+1} - r_i^k|$ $\forall i \in I^k \setminus (I_-^k \cup I_+^k)$, whereas for each $i \in I_+^k$, since $r_i^{k+1} \le \mu r_i^k$ and $\mu \in (0, 1/2]$, we have $|r_i^{k+1}| \le |r_i^{k+1} - r_i^k|$. Combining the preceding relations yields (8.4). □

Let us now establish convergence of the BCR method under Assumption 2.8 and

Assumption 8.2. Every element of $\{1 : m\}$ appears in $\{I^k\}$ infinitely many times.

In view of (8.6), Lemmas 3.4–3.8 hold with (3.10) replaced by

$$\sum_{k=1}^{\infty} | \left\langle r^{k+1}, p^{k+1} - p^k \right\rangle | \le 2 \sum_{k=1}^{\infty} \Delta_q^k < \infty. \tag{8.7}$$

Lemmas 3.9–3.10 are replaced by the following results.

Lemma 8.3. *$|[p_{I^k}^{k+1} + r_{I^k}^{k+1}]_+ - p_{I^k}^{k+1}| \to 0$ and*

$$\limsup_{k \to \infty} \max_{i \in I^k} r_i^{k+1} = \limsup_{k \to \infty} \max_{i \in I^k} r_i^k \le 0. \tag{8.8}$$

Proof. This follows from (8.4) and $r^{k+1} - r^k \to 0$ (cf. Lemma 3.8). □

Lemma 8.4. *$\{x^k\}$ converges to some $x^\infty \in C_f$.*

Proof. In the proof of Lemma 3.10, (3.12) and (3.15) are replaced by

$$D_f^{k+1}(x, x^{k+1}) + D_f^k(x^{k+1}, x^k) - D_f^k(x, x^k) = \left\langle p^{k+1} - p^k, A(x - x^{k+1})\right\rangle \quad \forall x \in C_f,$$

$$D_f^{k_j'}(x^\infty, x^{k_j'}) \le D_f^{k_j}(x^\infty, x^{k_j}) + \sum_{k=k_j}^{k_j'-1} \left\langle p^{k+1} - p^k, A(x - x^{k+1})\right\rangle \quad \forall j, \tag{8.9}$$

and we need only show that the sum above vanishes. Since $\limsup_{k\to\infty} \max_{i\in I^k} r_i^k \le 0$ by Lemma 8.3, there exists $j_>$ such that, for all $j \ge j_>$ and $k \in [k_j, k_j')$, $\left\langle a^i, x^k\right\rangle - b_i = r_i^k < \epsilon$ for all $i \in I^k$ and $I^k \cap I_> = \emptyset$ (otherwise $i \in I^k \cap I_>$ and $x^k \in B(x^\infty, \epsilon)$ would give $r_i^k > \epsilon$ by (3.14)). Now, $I^k \cap I_> = \emptyset$, $\langle a^i, x^\infty\rangle = b_i \ \forall i \in I_=$ and $Ax^\infty - Ax^{k+1} = Ax^\infty - b - r^{k+1}$ yield

$$\left\langle p^{k+1} - p^k, A(x - x^{k+1})\right\rangle = -\left\langle p^{k+1} - p^k, r^{k+1}\right\rangle + \sum_{i\in I^k\cap I_<} (p_i^{k+1} - p_i^k)(\left\langle a^i, x^\infty\right\rangle - b_i). \tag{8.10}$$

If $i \in I^k \cap I_<$ then (cf. (3.13)) $x^k \in B(x^\infty, \epsilon)$ yields $\left\langle a^i, x^k\right\rangle - b_i < -\epsilon$ and $i \in I_-^k$, whereas $x^{k+1} \in B(x^\infty, \epsilon)$ yields $-r_i^{k+1} = b_i - \left\langle a^i, x^{k+1}\right\rangle > \epsilon$, so $p_i^{k+1} = 0$ by (8.3) and $\sum_{i\in I^k\cap I_<} p_i^k \le \left\langle r^{k+1}, p^{k+1} - p^k\right\rangle / \epsilon$ by (8.5). Hence (8.10) and $\langle a^i, x^\infty\rangle < b_i \ \forall i \in I_<$ imply

$$\sum_{k=k_j}^{k_j'-1} \left\langle p^{k+1} - p^k, A(x - x^{k+1})\right\rangle \le (1 + \max_{i\in I_<}[b_i - \left\langle a^i, x^\infty\right\rangle]/\epsilon) \sum_{k=k_j}^{\infty} |\left\langle r^{k+1}, p^{k+1} - p^k\right\rangle| \to 0$$

as $j \to \infty$, using (8.7). This replaces (3.16) in the rest of the proof. □

Proof of Lemma 3.11 for the BCR method. In the original proof, note that, for any $i \in \{1{:}m\}$, $K = \{k : i \in I^k\}$ is infinite by Assumption 8.2, and $r_i^\infty = \lim_{k\in K} r_i^{k+1} \le \limsup_{k\to\infty} r_i^{k+1} \le 0$ (cf. Lemma 8.3). If $r_i^\infty < 0$ then $[p_i^{k+1} + r_i^{k+1}]_+ - p_i^{k+1} \xrightarrow{K} 0$ (Lemma 8.3) and $p^k \ge 0 \ \forall k$ yield $p_i^{k+1} \xrightarrow{K} 0$, so in fact $p_i^k \to 0$ because $p_i^{k+1} = p_i^k$ if $k \notin K$, and hence $[p_i^k + r_i^k]_+ - p_i^k \to 0$. The rest of the proof goes as before. □

We conclude that for the BCR method under Assumption 8.2, Theorem 3.12 holds under Assumption 2.8, and Theorems 5.4 and 6.3 remain true; this follows from their proofs. Following [Tse90, §5.1], we note that these results extend to *exact coordinate maximization* in which $p^{k+1} \in \operatorname{Arg\,max}\{q(p) : p \ge 0, p_i = p_i^k \ \forall i \notin I^k\}$. Specifically, if one can find the solution x^{k+1} and Lagrange multipliers π_i^k, $i \in I^k$, of the following problem

$$\text{minimize} \qquad f(x) + \sum_{i\notin I^k} p_i^k \left\langle a^i, x\right\rangle, \tag{8.11a}$$

$$\text{subject to} \qquad \left\langle a^i, x\right\rangle \le b_i \quad \forall i \in I^k, \tag{8.11b}$$

one may set $p_i^{k+1} = \pi_i^k$, $i \in I^k$, $p_i^{k+1} = p_i^k$, $i \notin I^k$.

A Appendix

Proof of Lemma 2.3. (a) This follows from [Roc70, Thm 24.7] if $\operatorname{ri} C_f = \mathring{C}_f$, and from [GoT89, Thm 1.2.7] if $\mathring{C}_f = \emptyset$ (briefly, letting L be the linear span of $C_f - \tilde{x}$ for any $\tilde{x} \in C_f$, one defines $f_L(y) = f(\tilde{x} + y)$ with $\partial f_L(y) = \partial f(\tilde{x} + y) \cap L \; \forall y \in L$, and applies [Roc70, Thm 24.7] to f_L with $\mathring{C}_{f_L} \neq \emptyset$).

(b) Let $y^* \in S$. There exist $\epsilon > 0$ and $z^i \in \mathbb{R}^n$, $i = 1{:}j$, such that $S \cap B(y^*, \epsilon) = \{y : \langle z^i, y\rangle \leq \langle z^i, y^*\rangle\}$, where $B(y^*, \epsilon) = \{y : |y - y^*| \leq \epsilon\}$. For any $y \in S \cap B(y^*, \epsilon)$, $\partial h(y) = \{0\} \cup \{\mathbb{R}_+ z^i : \langle z^i, y\rangle = \langle z^i, y^*\rangle\}$, so $\sigma_{\partial h(y)}(y - y^*) = 0$ and $D_h^-(y^*, y) = 0$.

(c) We have $h = h_1 + h_2$, where h_1 is polyhedral with $C_{h_1} = \mathbb{R}^n$ and $h_2 = \delta_{C_h}$. Since $\partial h = \partial h_1 + \partial h_2$ (cf. [Roc70, Thm 23.8]), $D_h^- = D_{h_1}^- + D_{h_2}^-$. Apply (a) to h_1 and (b) to h_2.

(d) By [Roc70, Cor. 7.5.1 and Thm 10.1], f is continuous on C_f. Let $l = \inf_{y \in C_f} y$ and $c = \sup_{y \in C_f} y$. If $y^* \in (l, c)$ then $D_f^\sharp(y^*, y^k) \to 0$ by part (a). Suppose $y^k \downarrow y^* = l$. By [Roc70, pp. 227–230], $\sigma_{\partial f(y^k)}(1) = f'(y^k; 1) \downarrow f'(y^*; 1) \in [-\infty, \infty)$. If $f'(y^*; 1) > -\infty$ then $0 \leq D_f^\sharp(y^*, y^k) = f(y^*) - f(y^k) + (y^k - y^*) f'(y^k; 1) \to 0$. If $f'(y^*; 1) = -\infty$ then $0 \leq D_f^\sharp(y^*, y^k) \leq f(y^*) - f(y^k) \to 0$ when $f'(y^k; 1) \leq 0$. The case $y^k \uparrow y^* = c$ is similar. □

Proof of Lemma 2.4. For each part, checking conditions (a)–(b) of Definition 2.1 is standard, so we only provide relations for verifying conditions (c)–(d).

(a) We have $C_f = \cap_{i=1}^k C_{f_i}$, $\partial f = \sum_{i=1}^k \partial f_i$ [Roc70, Thm 23.8] and $C_{\partial f} = \cap_{i=1}^k C_{\partial f_i}$. Use $D_f^\flat = \sum_{i=1}^k D_{f_i}^+$ and (by nonnegativity of $D_{f_i}^+$) $\mathcal{L}_f^1(x, \alpha) \subset \cap_{i=1}^k \mathcal{L}_{f_i}^1(x, \alpha) \; \forall x, \alpha$ for condition (c), and $D_f^\sharp = \sum_{i=1}^k D_{f_i}^-$ for condition (d).

(b) We have $C_f = \cap_{i=1}^j C_{f_i}$ and $\partial f(x) = \operatorname{co}\{\partial f_i(x) : i \in I(x)\}$ for $x \in C_f$ and $I(x) = \{i : f_i(x) = f(x)\}$ [GoT89, Thm 1.5.5], so $C_{\partial f} \subset \cup_{i=1}^j C_{\partial f_i}$, $\sigma_{\partial f(\cdot)} = \max_{i \in I(\cdot)} \sigma_{\partial f_i(\cdot)}$, $D_f^\flat(x, \cdot) \geq \max_{i \in I(\cdot)} D_{f_i}^+(x, \cdot)$ and $\mathcal{L}_f^1(x, \alpha) \subset \cup_{i=1}^j \mathcal{L}_{f_i}^1(x, \alpha) \; \forall \alpha$. If $\{y^k\} \subset C_{\partial f}$ converges to $y^* \in C_f$ then, for all large k, $I(y^k) \subset I(y^*)$ (by continuity of f_i on C_f) and $D_f^\sharp(y^*, y^k) = D_{f_i}^-(y^*, y^k)$ for some $i \in I(y^k)$, so $D_f^\sharp(y^*, y^k) \to 0$.

(c) As in (a), $C_f = C_{f_1}$, $\partial f = \partial f_1 + \partial f_2$, $C_{\partial f} = C_{\partial f_1}$ ($C_{\partial f_1} \subset C_{f_1} \subset \operatorname{ri} C_{f_2} \subset C_{\partial f_2}$), $D_f^\flat = D_{f_1}^+ + D_{f_2}^+ \geq D_{f_1}^+$ and $\mathcal{L}_f^1(x, \alpha) \subset \mathcal{L}_{f_1}^1(x, \alpha)$. If $\{y^k\} \subset \operatorname{ri} C_{f_2}$ converges to $y^* \in \operatorname{ri} C_{f_2}$ then $D_{f_2}^-(y^*, y^k) \to 0$ by Lemma 2.3(a).

(d) Using Lemma 2.3(d) and $\partial f = \prod_{i=1}^n \partial f_i$, argue as in part (a). □

Proof of Lemma 2.5. Suppose $C_{h^*} = \mathring{C}_{h^*}$. If $\mathcal{L}_h^1(x, \alpha/2)$ is unbounded for some $x \in C_h$ and $\alpha > 0$ then (cf. (2.1a)) there exist z^k and $\gamma^k \in \partial h(z^k)$ such that $h(z^k) + \left\langle \gamma^k, x - z^k \right\rangle \geq h(x) - \alpha$, $|z^k - x| \geq 1$ for all k and $|z^k| \uparrow \infty$. Let $\xi^k = x + (z^k - x)/|z^k - x|$ for all k. By convexity of h, $\xi^k \in C_h$ and $h(\xi^k) \geq h(z^k) + \left\langle \gamma^k, \xi^k - z^k \right\rangle$, so $h(\xi^k) \geq h(x) - \alpha + \left\langle \gamma^k, \xi^k - x \right\rangle$ for all k. Suppose $K \subset \{1, 2, \ldots\}$ is such that $\{z^k\}_{k \in K} \uparrow \infty$. Then $\{\gamma^k\}_{k \in K}$ is nondecreasing (cf. [Roc70, pp. 227–230]), $\xi^k = x + 1$ and $\left\langle \gamma^k, \xi^k - x \right\rangle \leq h(\xi^k) - h(x) + \alpha$ for all $k \in K$, so there exists $\gamma^\infty \in \mathbb{R}$ such that $\{\gamma^k\}_{k \in K} \uparrow \gamma^\infty$. But $h(z^k) + \left\langle \gamma^k, x - z^k \right\rangle \geq h(x) - \alpha$ means $h^*(\gamma^k) \leq \alpha - h(x) + \left\langle \gamma^k, x \right\rangle$, so in the limit $h^*(\gamma^\infty) < \infty$ from closedness of h^*. Hence $\gamma^\infty \in C_{h^*} = \mathring{C}_{h^*}$ yields the existence in C_{h^*} of $\bar{\gamma} > \gamma^\infty$. Then $-h^*(\bar{\gamma}) + \left\langle \bar{\gamma}, z^k \right\rangle \leq$

$h(z^k) \le h(x) + \langle \gamma^k, z^k - x \rangle$ implies $\langle \bar\gamma - \gamma^k, z^k \rangle \le h(x) - h^*(\bar\gamma) - \langle \gamma^k, x \rangle$ for all k. But the right side tends to ∞ as $\{k\}_{k\in K} \to \infty$ (since $\gamma_k \le \gamma^\infty < \bar\gamma$), whereas (since $\{\gamma^k\}_{k\in K} \uparrow \gamma^\infty$) the left side is bounded, a contradiction. A similar contradiction is derived if $\{z^k\}_{k\in K} \downarrow \infty$, using $\{\gamma^k\}_{k\in K} \downarrow \gamma^\infty > \bar\gamma \in \mathring{C}_{h^*}$. Hence $\mathcal{L}_h^1(x, \alpha/2)$ must be bounded for all $\alpha > 0$.

Next, suppose $\bar\gamma = \max_{\gamma \in C_{h^*}} \gamma \in C_{h^*}$. Using $\phi(z) = h(z) - \langle \bar\gamma, z \rangle + h^*(\bar\gamma) \ge 0$ and $C_{h^*} = \operatorname{im} \partial h$ (cf. [Roc70, Thm 23.5]), we have (cf. [Roc70, Thm 23.8]) $\partial \phi(z) = \partial h(z) - \bar\gamma \subset -\mathbb{R}_+$ for all z, so ϕ is nonincreasing (cf. [Roc70, pp. 227–230]). Hence there exist $z^k \uparrow \infty$, $\gamma^k \in \partial h(z^k)$, $\gamma^1 \le \gamma^k \le \bar\gamma$, and $x > 0$ such that $h(z^k) + \langle \gamma^k, x - z^k \rangle \ge -h^*(\bar\gamma) + \langle \bar\gamma - \gamma^k, z^k \rangle + \langle \gamma^k, x \rangle \ge h(x) - \alpha$ for all k with $\alpha = h^*(\bar\gamma) + h(x) - \langle \gamma^1, x \rangle < \infty$. Thus $\{z^k\} \subset \mathcal{L}_h^1(x, \alpha)$, so $\mathcal{L}_h^1(x, \alpha)$ is unbounded. The case where $\bar\gamma = \min_{\gamma \in C_{h^*}} \gamma \in C_{h^*}$ is similar. □

We need the following result for proving Lemma 2.9.

Lemma A.1. *Let $\tilde f(\cdot) = f^*(\tilde A \cdot)$, where f is a closed proper essentially strictly convex function on $\mathbb{R}^n$ and $\tilde A$ is a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$. If $\tilde A^{-1} \operatorname{im} \partial f \ne \emptyset$ then f^* and $\tilde f$ are essentially smooth, $\mathring{C}_{f^*} = \operatorname{im} \partial f$, $C_{\tilde f} = \tilde A^{-1} C_{f^*}$, $\nabla f^* \tilde A$ and $\nabla \tilde f = \tilde A^* \nabla f^* \tilde A$ are continuous on $C_{\nabla \tilde f} = C_{\partial \tilde f} = \mathring{C}_{\tilde f} = \tilde A^{-1} \mathring{C}_{f^*}$, i.e., $\nabla \tilde f(p) = \tilde A^* \nabla f^*(\tilde A p)$ for any $p \in \mathring{C}_{\tilde f}$, with $x = \nabla f^*(\tilde A p) \iff \tilde A p \in \partial f(x) \iff x \in \operatorname{Arg\,min} f(\cdot) - \langle \tilde A p, \cdot \rangle \iff \tilde f(p) = \langle \tilde A p, x \rangle - f(x)$. If f is cofinite then $\tilde A^{-1} \operatorname{im} \partial f \ne \emptyset$ and $\tilde f$ is continuously differentiable on $\mathbb{R}^m$.*

Proof. By the assumptions on f and [Roc70, Thm 26.3], f^* is closed proper essentially smooth, so $\partial f^*(y) = \{\nabla f^*(y)\}$ $\forall y \in \mathring{C}_{f^*} = C_{\partial f^*}$ by [Roc70, Thm 26.1] and $\nabla f^* \tilde A$ is continuous on $\tilde A^{-1} \mathring{C}_{f^*}$ by [Roc70, Thm 25.5]. By [Roc70, Thm 23.5], $\partial f^* = (\partial f)^{-1}$, i.e., $x \in \partial f^*(y)$ iff $y \in \partial f(x)$, so $\operatorname{im} \partial f = C_{\partial f^*}$. If $\tilde A \tilde p \in \operatorname{im} \partial f = \mathring{C}_{f^*}$ for some $\tilde p$ then $\mathring{C}_{\tilde f} = \tilde A^{-1} \mathring{C}_{f^*}$ by [Roc70, Thm 6.7], since $C_{\tilde f} = \tilde A^{-1} C_{f^*}$, and $\partial \tilde f = \tilde A^* \partial f^* \tilde A$ by [Roc70, Thm 23.9]. Thus $\partial \tilde f$ is single-valued and $\tilde f$ is closed and proper (so is f^*), so $\tilde f$ is essentially smooth and $\partial \tilde f$ reduces to $\nabla \tilde f$ on $\mathring{C}_{\tilde f} = C_{\partial \tilde f}$ by [Roc70, Thm 26.1], where $\nabla \tilde f$ is continuous by [Roc70, Thm 25.5]. The equivalences follow from [Roc70, Thm 23.5]. If f is cofinite then $\operatorname{im} \partial f = C_{f^*} = \mathbb{R}^n$. □

Proof of Lemma 2.9. By Assumption 2.8 and Lemma A.1 with $\tilde A p = -A^T p$, $\tilde f = f^*(-A^T)$ is essentially smooth with $\mathring{C}_{\tilde f} = \{p : -A^T p \in \operatorname{im} \partial f\}$, and so is $-q(\cdot) = \tilde f(\cdot) + \langle b, \cdot \rangle$. □

References

[ABC83] R. Aharoni, A. Berman and Y. Censor, *An interior points algorithm for the convex feasibility problem*, Adv. in Appl. Math. **4** (1983) 479–489.

[BCP93] J. B. Brown, P. J. Chase and A. O. Pittenger, *Order independence and factor convergence in iterative scaling*, Linear Algebra Appl. **190** (1993) 1–38.

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

[BHT87] D. P. Bertsekas, P. A. Hosein and P. Tseng, *Relaxation methods for network flow problems with convex arc costs*, SIAM J. Control Optim. **25** (1987) 1219–1243.

[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.

[CDPE90] Y. Censor, A. R. De Pierro, T. Elfving, G. T. Herman and A. N. Iusem, *On iterative methods for linearly constrained entropy maximization*, in Numerical Analysis and Mathematical Modelling, A. Wakulicz, ed., Banach Center Publications, vol. 24, PWN—Polish Scientific Publishers, Warsaw, 1990, pp. 145–163.

[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.

[CeH87] Y. Censor and G. T. Herman, *On some optimization techniques in image reconstruction from projections*, Appl. Numer. Math. **3** (1987) 365–391.

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

[CeL87] ———, *Optimization of "$\log x$" entropy over linear equality constraints*, SIAM J. Control Optim. **25** (1987) 921–933.

[Cen81] Y. Censor, *Row-action methods for huge and sparse systems and their applications*, SIAM Rev. **23** (1981) 444–466.

[Cen88] ———, *Parallel application of block-iterative methods in medical imaging and radiation therapy*, Math. Programming **42** (1988) 307–325.

[CeZ91] Y. Censor and S. A. Zenios, *Interval-constrained matrix balancing*, Linear Algebra Appl. **150** (1991) 393–421.

[CeZ92] ———, *Proximal minimization algorithm with D-functions*, J. Optim. Theory Appl. **73** (1992) 451–464.

[DaR72] J. N. Darroch and D. Ratclif, *Generalized iterative scaling for log-linear models*, Ann. Math. Statist. **43** (1972) 1470–1480.

[DPI86] A. R. De Pierro and A. N. Iusem, *A relaxed version of Bregman's method for convex programming*, J. Optim. Theory Appl. **51** (1986) 421–440.

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

[Elf80] T. Elfving, *On some methods for entropy maximization and matrix scaling*, Linear Algebra Appl. **34** (1980) 321–339.

[Elf89] ———, *An algorithm for maximum entropy image reconstruction from noisy data*, Math. Comput. Modelling **12** (1989) 729–745.

[Erl81] S. Erlander, *Entropy in linear programs*, Math. Programming **21** (1981) 137–151.

[Fal67] J. E. Falk, *Lagrange multipliers and nonlinear programming*, J. Math. Anal. Appl. **19** (1967) 141–159.

[FlZ90] S. D. Flåm and J. Zowe, *Relaxed outer projections, weighted averages and convex feasibility*, BIT **30** (1990) 289–300.

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

[GPR67] L. G. Gurin, B. T. Polyak and E. V. Raik, *The method of projections for finding a common point of convex sets*, Zh. Vychisl. Mat. i Mat. Fiz. **7** (1967) 1211–1228 (Russian). English transl. in U.S.S.R. Comput. Math. and Math. Phys. 7 (1967) 1–24.

[HeL78] G. T. Herman and A. Lent, *A family of iterative quadratic optimization algorithms for pairs of inequalities with application in diagnostic radiology*, Math. Programming Stud. **9** (1978) 15–29.

[Hil57] C. Hildreth, *A quadratic programming procedure*, Naval Res. Logist. Quart. **4** (1957) 79–85. Erratum, ibid., p. 361.

[Ius91] A. N. Iusem, *On dual convergence and the rate of primal convergence of Bregman's convex programming method*, SIAM J. Optim. **1** (1991) 401–423.

[Kiw94] K. C. Kiwiel, *Block-iterative surrogate projection methods for convex feasibility problems*, Linear Algebra Appl. **?** (1994). To appear.

[Kru37] J. Kruithof, *Calculation of telephone traffic*, De Ingenieur (E. Elektrotechnik 3) **52** (1937) E15–E25.

[LaS81] B. Lamond and F. Stewart, *Bregman's balancing method*, Transportation Res. B **15** (1981) 239–249.

[LeC80] A. Lent and Y. Censor, *Extensions of Hildreth's row-action method for quadratic programming*, SIAM J. Control Optim. **18** (1980) 444–454.

[LiP87] Y. Y. Lin and J.-S. Pang, *Iterative methods for large convex quadratic programs: A survey*, SIAM J. Control Optim. **25** (1987) 383–411.

[LuT91] Z.-Q. Luo and P. Tseng, *On the convergence of a matrix splitting algorithm for the symmetric monotone linear complementarity problem*, SIAM J. Control Optim. **29** (1991) 1037–1060.

[LuT92a] ———, *Error bound and convergence analysis of matrix splitting algorithms for the affine variational inequality problem*, SIAM J. Optim. **2** (1992) 43–54.

[LuT92b] ———, *On the convergence of the coordinate descent method for for convex differentiable minimization*, J. Optim. Theory Appl. **72** (1992) 7–35.

[LuT92c] ———, *On the linear convergence of descent methods for convex essentially smooth minimization*, SIAM J. Control Optim. **30** (1992) 408–425.

[LuT93] ———, *Error bound and reduced-gradient projection algorithms for convex minimization over a polyhedral set*, SIAM J. Optim. **3** (1993) 43–59.

[Man84] O. L. Mangasarian, *Sparsity-preserving SOR algorithms for separable quadratic and linear programming problems*, Comput. Oper. Res. **11** (1984) 105–112.

[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.

[Ott88] N. Ottavy, *Strong convergence of projection-like methods in Hilbert spaces*, J. Optim. Theory Appl. **56** (1988) 433–461.

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

[Roc84] ———, *Network Flows and Monotropic Optimization*, Wiley, New York, 1984.

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

[TsB87] P. Tseng and D. P. Bertsekas, *Relaxation methods for problems with strictly convex separable costs and linear constraints*, Math. Programming **38** (1987) 303–321.

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

[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.

[Tse91] ———, *Descent methods for convex essentially smooth minimization*, J. Optim. Theory Appl. **71** (1991) 425–463.

[Tse92] ———, *On the convergence of the products of firmly nonexpansive maps*, SIAM J. Optim. **2** (1992) 425–434.

[ZeC91a] S. A. Zenios and Y. Censor, *Massively parallel row-action algorithms for some nonlinear transportation problems*, SIAM J. Optim. **1** (1991) 373–400.

[ZeC91b] ———, *Parallel computing with block-iterative image reconstruction algorithms*, Appl. Numer. Math. **7** (1991) 399–415.

[Zen91] S. A. Zenios, *On the fine-grain decomposition of multicommodity transportation problems*, SIAM J. Optim. **1** (1991) 643–669.