RUTCOR
RESEARCH
REPORT

A Practical General Approximation Criterion for Methods of Multipliers Based on Bregman Distances

Jonathan Eckstein[a]

RRR 61-2000, December 2000, revised July 2002.

RUTCOR
Rutgers Center for
Operations Research
Rutgers University
640 Bartholomew Road
Piscataway, New Jersey
08854-8003
Telephone: 732-445-3804
Telefax: 732-445-5472
Email: rrr@rutcor.rutgers.edu
http://rutcor.rutgers.edu/~rrr

[a]Rutgers Business School and RUTCOR, 640 Bartholomew Road, Busch Campus, Rutgers University, Piscataway, NJ 08854 USA(jeckstei@rutcor.rutgers.edu).

Rutcor Research Report
RRR 61-2000, December 2000, revised July 2002.

A Practical General Approximation Criterion for Methods of Multipliers Based on Bregman Distances

Jonathan Eckstein

Abstract. This paper demonstrates that for generalized methods of multipliers for convex programming based on Bregman distance kernels — including the classical quadratic method of multipliers — the minimization of the augmented Lagrangian can be truncated using a simple, generally implementable stopping criterion based only on the norms of the primal iterate and the gradient (or a subgradient) of the augmented Lagrangian at that iterate. Previous results in this and related areas have required conditions that are much harder to verify, such as ϵ-optimality with respect to the augmented Lagrangian, or strong conditions on the convex program to be solved. Here, only existence of a KKT pair is required, and the convergence properties of the exact form of the method are preserved. The key new element in the analysis is the use of a full conjugate duality framework, as opposed to mainly examining the action of the method on the standard dual function of the convex program. An existence result for the iterates, stronger than those possible for the exact form of the algorithm, is also included.

Acknowledgements: I would like to thank Teemu Pennanen, Paulo Silva, and Bob Vanderbei for their helpful comments when I was developing these results. Paulo Silva and an anonymous referee of [25] also deserve credit for stimulating me to work on this problem. One of the anonymous referees of an earlier version of this paper also made numerous helpful suggestions. I would also like to thank the ORFE department of Princeton University for making office facilities available to me during my sabbatical.

1 Introduction

This paper concerns solution of convex programming problems of the form:

$$\begin{array}{ll} \min & f(x) \\ \text{S.T.} & g(x) \le 0, \end{array} \tag{1}$$

where $g(x)$ denotes the vector $(g_1(x), \ldots, g_m(x))$ and the functions $f, g_1, \ldots, g_m : \mathbb{R}^n \to (-\infty, +\infty]$ are closed proper convex.

Specifically, we consider solving (1) by a generalized *method of multipliers* or *augmented Lagrangian* method. Such methods include the classical quadratic augmented Lagrangian method [22]

$$x^k \in \operatorname*{Arg\,min}_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{1}{2c_k} \sum_{i=1}^{m} \max\{0, p_i^{k-1} + c_k g_i(x)\}^2 \right\} \tag{2}$$

$$p_i^k = \max\{0, p_i^{k-1} + c_k g_i(x^k)\} \qquad i = 1, \ldots, m, \tag{3}$$

and the *exponential method of multipliers* (see for example [30])

$$x^k \in \operatorname*{Arg\,min}_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{1}{c_k} \sum_{i=1}^{m} p_i^{k-1} e^{c_k g_i(x)} \right\} \tag{4}$$

$$p_i^k = p_i^{k-1} e^{c_k g_i(x^k)} \qquad i = 1, \ldots, m. \tag{5}$$

Here, $\{c_k\}$ is a sequence of positive scalars bounded away from zero, and the $p^k \in \mathbb{R}^m$ are Lagrange multiplier estimates for the constraints $g(x) \le 0$. Both these methods are examples of the following general form, as shown, for example, in [25, Appendix A]: let $D(q, p)$ be a function that measures the "distance", in some generalized sense, between two points $q, p \in \mathbb{R}^m$. Define $D^{\oplus_1}(u, p) = \sup_{q \ge 0}\{\langle q, u\rangle - D(q, p)\}$,[1] and consider the method

$$x^k \in \operatorname*{Arg\,min}_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{1}{c_k} D^{\oplus_1}(c_k g(x), p^{k-1}) \right\} \tag{6}$$

$$p^k = \nabla_1 D^{\oplus_1}(c_k g(x^k), p^{k-1}), \tag{7}$$

where ∇_1 denotes differentiation with respect to the first argument.

Methods of this form all involve minimization of an *augmented Lagrangian*, as in (2), (4), or (6), followed by a *multiplier update*, as in (3), (5), or (7). In practice, the augmented Lagrangian minimization may itself involve a lengthy iterative calculation, and one may not want to compute the minimum exactly, but only approximate it. Generally speaking, this approximation should be quite rough near the outset of the algorithm, when one may be far from the solution, and gradually become more exact with each iteration. Let ϕ_k denote the function to be minimized at iteration k, for example, $\phi_k(x) =$

[1] This paper will use the convention that $0 \cdot \infty = \infty$, so that if $u_i = \infty$ for any i, then $D^{\oplus_1}(u, p) = \infty$.

$f(x)+(1/2c_k)\sum_{i=1}^m \max\{0, p_i^{k-1}+c_k g_i(x)\}^2$ for (2), or $\phi_k(x) = f(x)+(1/c_k)\sum_{i=1}^m p_i^{k-1}e^{c_k g_i(x)}$ for (4). Further, let $\{\epsilon_k\}_{k=1}^\infty$ denote some sequence of nonnegative tolerances, with $\epsilon_k \to 0$. Ideally, one would want to truncate the augmented Lagrangian minimization when the gradient of ϕ_k is sufficiently small, that is, when $\|\nabla\phi_k(x^k)\| \le \epsilon_k$, or if ϕ_k is not smooth, when $\mathrm{dist}(\partial\phi_k(x^k), 0) \le \epsilon_k$, "$\partial$" denoting the set of subgradients.

Unfortunately, the theoretical basis for such truncated minimizations of augmented Lagrangians is not entirely satisfactory. If we let $Q_0 : \mathbb{R}^m \to [-\infty, +\infty)$ denote the dual function of problem (1) (see Section 2.1 below), then it can be shown, for example in [25, Appendix A], that the sequence $\{p^k\}$ computed by (6)-(7) is the same as would be generated by the generalized proximal minimization algorithm

$$p^k = \arg\min_{p\in\mathbb{R}^m}\left\{-Q_0(p) + \frac{1}{c_k}D(p, p^{k-1})\right\}, \tag{8}$$

where D is the same generalized distance measure generating $D^{\oplus_1}(\cdot,\cdot)$. Of course, (8) may be only a conceptual method, since no explicit formula for Q_0 may exist. Now, while a wealth of different criteria are known for approximately computing the minimum in (8) for various forms of the distance kernel D, as for example in [1, 11, 15, 21, 25, 26, 29], all have proven awkward or impossible to translate into tractable conditions for the minimization in the equivalent algorithm (6)-(7). Some approaches, such as [1, 15, 23], use ϵ-subgradients, which may be difficult to identify in general practice; other approaches, such as [1, 15, 22], require ϵ_k-optimality of x^k, that is,

$$\phi_k(x^k) \le \inf_{x\in\mathbb{R}^n}\{\phi_k(x)\} + \epsilon_k. \tag{9}$$

In practice, $\inf_{x\in\mathbb{R}^n}\{\phi_k(x)\}$ is generally unknown, and such a condition may be difficult or impossible to verify without making additional, stringent assumptions on f, g, and/or D. Other criteria are known, as for example in [25, Appendix B], but, again, highly restrictive assumptions are required on the problem (1), such as strong convexity of the objective function f.

This paper shows that, for a broad class of distance functions D derived from *Bregman functions* (see Section 2.2 for a definition), it is possible to preserve the theoretical convergence properties of the algorithm (6)-(7) while truncating the minimization (6) using a simple, generally implementable stopping criterion that requires knowing only the norms of x^k and some subgradient $y^k \in \partial\phi_k(x^k)$ — see (16)-(17) below. No extraordinary assumptions on the original problem (1) are required: one only needs the existence of a primal-dual solution pair satisfying the usual Karush-Kuhn-Tucker (KKT) conditions.

The analytical methods used here resemble those applied to various methods related to (8) in [11], but with an important difference: instead of just analyzing the action of the method on the dual functional Q_0, as has been traditional, the algorithm is placed in the context of a full conjugate duality structure for (1); Section 2.1 will review conjugate duality. There is also a connection between the key inequality (27) below and the notion of *ϵ-enlarging* a monotone operator [6], although the concept cannot be used directly.

Historically, there are other kinds of nonlinear optimization methods in which the gradient of the augmented Lagrangian figures prominently in the convergence criteria; see for example [13] and references therein. However, such methods do not appear to fit the simple pattern (2)-(3) or (6)-(7) of (approximately) minimizing the augmented Lagrangian and then updating the multipliers, and have very different convergence analyses.

Section 2 describes some notation and reviews necessary theoretical background material related to conjugate duality and Bregman functions. Section 3 then presents the algorithm and its basic dual convergence analysis, showing that the sequence of multiplier estimates $\{p^k\}$ converges to an optimal dual solution.

Section 4 then analyzes the behavior of the primal iterates $\{x^k\}$. Although the analysis becomes somewhat involved at this point, the key point is that nothing appears to be lost by using the new, more tractable approximation criterion — the behavior of $\{x^k\}$ appears identical to the corresponding algorithms with exact minimization of the augmented Lagrangian.

While the analysis deals only with algorithms for problems in the form (1), it should be noted that the results extend in a straightforward way to the case where the constraints $g(x) \leq 0$ are either augmented or replaced by affine equalities of the form $Ax = b$, where A is some $\overline{m} \times n$ real matrix and $b \in \mathbb{R}^{\overline{m}}$. This extension is described, briefly and without proof, in Section 5.

Finally, Section 6 gives results indicating when sequences meeting the recursion conditions of Section 3's algorithm are guaranteed to exist. These results include cases where the recursions of the exact version of algorithm have no solution.

2 Theoretical Background

Let $\mathbb{R}_+$ denote the set of all nonnegative real numbers, and $\mathbb{R}_{++}$ the set of all positive real numbers.

For any closed convex set $C \subseteq \mathbb{R}^q$, we denote its convex indicator function by δ_C, that is,

$$\delta_C(x) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise.} \end{cases}$$

The subgradient map of this function is the *normal cone* map of C, denoted by

$$N_C(x) \stackrel{\text{def}}{=} \partial\delta_C(x) = \begin{cases} \{d \in \mathbb{R}^q \mid \langle d, x' - x \rangle \leq 0 \ \forall\, x' \in C\} & \text{if } x \in C \\ \emptyset & \text{otherwise.} \end{cases}$$

2.1 Conjugate Duality Framework

The analysis will require a conjugate duality framework, as described in [19, Chapters 28-30], [20], [16], or [17]. Proofs of the results stated here may be found in these references.

We define the *perturbed objective function* $F : \mathbb{R}^n \times \mathbb{R}^m \to (-\infty, +\infty]$ of (1) via

$$F(x, u) \stackrel{\text{def}}{=} \begin{cases} f(x) & \text{if } g(x) + u \leq 0 \\ +\infty & \text{otherwise.} \end{cases}$$

The *essential objective* $F_0 : \mathbb{R}^n \to (-\infty, +\infty]$ of (1) is

$$F_0(x) \stackrel{\text{def}}{=} F(x,0) = \begin{cases} f(x) & \text{if } g(x) \le 0 \\ +\infty & \text{otherwise.} \end{cases}$$

Solving (1) is equivalent to minimizing F_0 over $\mathbb{R}^n$. We let ∂F denote the subgradient mapping of F. A *(primal) feasible solution* is any $x \in \mathbb{R}^n$ such that $F_0(x) < \infty$, and a *primal solution* is any feasible solution x^* such that $F_0(x^*)$ is minimized.

The following standard regularity assumption will apply throughout the rest of this paper:

Assumption 2.1 *For $i = 1, \ldots, m$, $\operatorname{ri}\operatorname{dom} g_i \supseteq \operatorname{ri}\operatorname{dom} f$ and $\operatorname{dom} g_i \supseteq \operatorname{dom} f$, where "$\operatorname{ri} S$" denotes the relative interior of the set S (see for example [19, p. 44]).*

Lemma 2.2 *Under Assumption 2.1, F and F_0 are closed proper convex functions.*

One obtains the *Lagrangian* $L : \mathbb{R}^n \times \mathbb{R}^m \to [-\infty, +\infty)$ of (1) by taking the concave conjugate [19, p. 111] of F with respect to the second argument u:

$$\begin{aligned} L(x,p) &\stackrel{\text{def}}{=} \inf_{u \in \mathbb{R}^m} \{F(x,u) - \langle u, p\rangle\} \\ &= \inf_{u \le -g(x)} \{f(x) - \langle u, p\rangle\} \\ &= \begin{cases} f(x) + \langle p, g(x)\rangle & \text{if } p \ge 0 \\ -\infty & \text{otherwise.} \end{cases} \end{aligned}$$

To simplify the notation, we adopt the convention (mentioned in footnote 1) that $0 \cdot \infty = \infty$: if for some i one has $p_i = 0$ and $g_i(x) = \infty$, then we take $p_i g_i(x) = \infty$ and hence $\langle p, g(x)\rangle = \infty$.

$L(x,p)$ is convex with respect to x and concave (and in fact linear) with respect to p, and we let ∂L denote its convex-concave subgradient map, that is, $\partial L(x,p)$ is the set consisting of all $(y,u) \in \mathbb{R}^n \times \mathbb{R}^m$ such that

$$\begin{aligned} L(x',p) &\ge L(x,p) + \langle y, x' - x\rangle \quad \forall\, x' \in \mathbb{R}^n \\ L(x,p') &\le L(x,p) - \langle u, p' - p\rangle \quad \forall\, p' \in \mathbb{R}^m. \end{aligned}$$

The *perturbed dual function* $Q : \mathbb{R}^n \times \mathbb{R}^m \to [-\infty, +\infty)$ of (1) can be defined two ways: either as the concave conjugate of F jointly with respect to both arguments x and u, or as the concave conjugate of L with respect to x. Thus,

$$\begin{aligned} Q(y,p) &\stackrel{\text{def}}{=} \inf_{\substack{x \in \mathbb{R}^n \\ u \in \mathbb{R}^m}} \{F(x,u) - \langle x, y\rangle - \langle u, p\rangle\} \\ &= \inf_{x \in \mathbb{R}^n} \{L(x,p) - \langle x, y\rangle\} \\ &= \begin{cases} \inf_{x \in \mathbb{R}^n} \{f(x) + \langle p, g(x)\rangle - \langle x, y\rangle\} & \text{if } p \ge 0 \\ -\infty & \text{otherwise.} \end{cases} \end{aligned}$$

By the elementary properties of concave conjugates, Q is necessarily a closed concave function. Let ∂Q denote its subgradient map, or equivalently the subgradient map of the convex function $-Q$. Thus, $\partial Q(y,p)$ consists of all $(x,u) \in \mathbb{R}^n \times \mathbb{R}^m$ such that

$$Q(y',p') \le Q(y,p) - \langle x, y'-y\rangle - \langle u, p'-p\rangle \quad \forall\, y' \in \mathbb{R}^n \;\forall\, p' \in \mathbb{R}^m. \tag{10}$$

The *dual function* $Q_0 : \mathbb{R}^m \to [-\infty, +\infty)$ of (1) is the perturbed dual function evaluated at $y = 0$, that is,

$$Q_0(p) \overset{\text{def}}{=} Q(0,p) = \begin{cases} \inf_{x\in\mathbb{R}^n} \{f(x) + \langle p, g(x)\rangle\} & \text{if } p \ge 0 \\ -\infty & \text{otherwise.} \end{cases}$$

The *dual problem* of (1) is that of maximizing Q_0 over $p \in \mathbb{R}^m$, or equivalently over $p \in \mathbb{R}^m_+$. A *dual feasible solution* is any $p \in \mathbb{R}^m_+$ such that $Q_0(p) > -\infty$, and a *dual solution* is any $p^* \in \mathbb{R}^m_+$ such that $Q_0(p^*)$ is maximized.

We also let $V(y,u)$ denote the concave conjugate of F with respect to x, instead of u, that is,

$$\begin{aligned} V(y,u) &\overset{\text{def}}{=} \inf_{x\in\mathbb{R}^n} \{F(x,u) - \langle x, y\rangle\} \\ &= \inf\{f(x) - \langle x,y\rangle \mid x \in \mathbb{R}^n : g(x) + u \le 0\}. \end{aligned}$$

This function is known as the *dual Lagrangian*, and is concave in y and convex in u. We let ∂V denote its concave-convex subgradient mapping. We further let

$$V_0(u) \overset{\text{def}}{=} V(0,u) = \inf\{f(x) \mid x \in \mathbb{R}^n : g(x) + u \le 0\}, \tag{11}$$

which must be convex, and call it the *parametric value function* of (1); it is also sometimes referred to as the the *primal functional* of (1).

Proposition 2.3 *(Weak duality and subgradient properties) Let Assumption 2.1 hold. For any $x \in \mathbb{R}^n$ and $p \in \mathbb{R}^m$, $F_0(x) \ge Q_0(p)$. The point-to-set mappings ∂F, ∂L, ∂V, and ∂Q are all maximal monotone [19, p. 240], and for all $x, y \in \mathbb{R}^n$ and $u, p \in \mathbb{R}^m$,*

$$\begin{aligned} & (y,p) \in \partial F(x,u) \;\Leftrightarrow\; (y,u) \in \partial L(x,p) \\ \Leftrightarrow\; & (x,p) \in \partial V(y,u) \;\Leftrightarrow\; (x,u) \in \partial Q(y,p). \end{aligned} \tag{12}$$

Finally,

$$\partial L(x,p) = \partial_x[f(x) + \langle p, g(x)\rangle] \times \left(\{-g(x)\} + N_{\mathbb{R}^m_+}(p)\right). \tag{13}$$

We will assume that problem (1) is well-behaved:

Assumption 2.4 *We assume that there exists at least one point $(x^*, p^*) \in \mathbb{R}^n \times \mathbb{R}^m$ such that $(0,0) \in \partial L(x^*, p^*)$.*

Assumption 2.4 can be guaranteed under a number of standard constraint qualification conditions, such as the optimal value of (1) being finite and existence of some $\tilde{x} \in \operatorname{ri}\operatorname{dom} f$ satisfying the Slater condition $g(\tilde{x}) < 0$ [19, Theorem 28.2]. We call points (x^*, p^*) satisfying the conditions of Assumption 2.4 *KKT pairs*, and denote the set of all such pairs by $Z = (\partial L)^{-1}(0,0)$. The assumption $Z \neq \emptyset$ guarantees that problem (1) and its dual are well-behaved:

Proposition 2.5 *Suppose Assumptions 2.1 and 2.4 hold. Then*

(i) F_0 *attains a minimum value at precisely the set of points* x^* *for which* $(x^*, p^*) \in Z$ *for some* $p^* \in \mathbb{R}^m$. *These points are the optimal solutions of (1).*

(ii) Q_0 *attains a maximum value at precisely the set of points* p^* *for which* $(x^*, p^*) \in Z$ *for some* $x^* \in \mathbb{R}^n$.

(iii) $\min_{x \in \mathbb{R}^n} F_0(x) = \max_{p \in \mathbb{R}^m} Q_0(p)$.

(iv) If x^* *is any minimizer of* F_0, *and* p^* *is any maximizer of* Q_0, *then* $(x^*, p^*) \in Z$.

We let f^* denote the common value of $\min_{x \in \mathbb{R}^n} F_0(x)$ and $\max_{p \in \mathbb{R}^m} Q_0(p)$ in (iii) above, that is, the optimal primal and dual objective value. We also let X^* denote the set of primal solutions and P^* the set of dual solutions. From the last proposition, $Z = X^* \times P^*$.

Finally, the following result will be useful in Section 6:

Proposition 2.6 Q_0 *is the concave conjugate of the convex function* V_0. *In particular, if* p^* *is an optimal dual solution, then* $p^* \in \partial V_0(0)$.

2.2 Bregman Functions and Distances

To construct generalized multiplier methods for (1), we will use an auxiliary strictly convex function $h : \mathbb{R}^m \to (-\infty, +\infty]$. We then define an associated *Bregman distance* $D_h : \operatorname{dom} h \times \operatorname{dom} \nabla h \to (-\infty, +\infty]$ via [5]

$$D_h(u, v) = h(u) - h(v) - \langle \nabla h(v), u - v \rangle. \tag{14}$$

$D_h(u, v)$ is easily seen to be nonnegative, and zero only when $u = v$.

We will choose h to be a *Bregman function* [5, 7, 26] whose "zone" S contains $\mathbb{R}^m_{++}$, that is, to conform to the following:

Assumption 2.7 *The function* $h : \mathbb{R}^m \to (-\infty, +\infty]$ *and set* $S \subseteq \mathbb{R}^m$ *have the following properties:*

(i) S *is open and convex, with* $S \supseteq \mathbb{R}^m_{++}$.

(ii) h *is finite, continuous, and strictly convex on* $\operatorname{cl} S$, *and* $h(x) = +\infty$ *for* $x \notin \operatorname{cl} S$.

(iii) *h is essentially smooth [19, Chapter 26], that is, h is differentiable throughout S, but the derivative becomes infinite as one approaches* $\operatorname{bd} S$.

(iv) *For all $p \in \operatorname{cl} S$ and scalars $\alpha \geq 0$, the level set $\{q \in S \mid D_h(p,q) \leq \alpha\}$ is bounded.*

(v) *For all sequences $\{p^k\} \subset S$ that converge to a finite limit p^∞, one has*

$$\lim_{k\to\infty} D_h(p^\infty, p^k) = 0.$$

We now state a few results that are useful in proving the convergence of proximal algorithms based on Bregman distances.

Proposition 2.8 *[26, Theorem 2.4] Suppose h and S satisfy Assumption 2.7. For any pair of sequences $\{p^k\} \subset \operatorname{cl} S$ and $\{q^k\} \subset S$, either one of which is convergent, with the property that $\lim_{k\to\infty} D_h(p^k, q^k) = 0$, the other sequence also converges to the same limit.*

Lemma 2.9 *[9, Lemma 3.1] Given $h : \mathbb{R}^m \to (-\infty, +\infty]$, D_h as defined in (14), $p, q \in \operatorname{dom} \nabla h$, and $r \in \operatorname{dom} h$,*

$$D_h(r,p) = D_h(q,p) + D_h(r,q) + \langle \nabla h(q) - \nabla h(p), r - q \rangle. \tag{15}$$

Lemma 2.10 *[18, Section 2.2] Suppose $\{a_k\}$, $\{\gamma_k\} \subset \mathbb{R}$ are sequences such that $\{a_k\}$ is bounded below, $\sum_{k=0}^{\infty} \gamma_k$ exists and is finite, and the recursion $a_{k+1} \leq a_k + \gamma_k$ holds for all $k \geq 1$. Then $\{a_k\}$ converges to a finite limit.*

Finally, we define the *monotone conjugate* [19, p. 111] $h^\oplus$ of h to be the convex conjugate [19, Ch. 12] of $h + \delta_{\mathbb{R}^m_+}$, that is,

$$h^\oplus(v) \stackrel{\text{def}}{=} \sup_{p\in\mathbb{R}^m} \left\{ \langle v, p \rangle - h(p) - \delta_{\mathbb{R}^m_+}(p) \right\} = \sup_{p\geq 0} \left\{ \langle v, p \rangle - h(p) \right\}$$

From the strict convexity of h and the definition of $h^\oplus$, the following result is immediate:

Lemma 2.11 *Under Assumption 2.7, $h^\oplus$ is an essentially smooth [19, Ch. 26] closed proper convex function, and is nondecreasing in the sense that, for $v, v' \in \mathbb{R}^m$, $v' \geq v$ implies $h^\oplus(v') \geq h^\oplus(v)$.*

Some common examples of functions meeting Assumption 2.7 are $h(u) = \frac{1}{2}\|u\|^2$ and $S = \mathbb{R}^m$, in which case $h^\oplus(v) = \frac{1}{2}\|[v]_+\|^2$, where $[v]_+$ denotes the positive part of v, or $h(u) = \sum_{i=1}^m u_i \log u_i$ and $S = \mathbb{R}^m_{++}$, in which case $h^\oplus(v) = \sum_{i=1}^m e^{v_i}$.

3 Algorithm and Convergence Analysis for the Dual Iterates

We now consider solving (1) via the following approximate method of multipliers:

Algorithm 3.1 *Suppose h and S are as specified in Assumption 2.7, and choose:*

(i) An arbitrary initial multiplier estimate $p^0 \in S$.

(ii) A positive scalar sequence of stepsizes $\{c_k\}_{k=1}^{\infty} \subset \mathbb{R}_{++}$ *with* $\inf_{k\geq 1}\{c_k\} > 0$.

(iii) A nonnegative scalar sequence of allowable errors $\{\epsilon_k\}_{k=1}^{\infty} \subset \mathbb{R}_+$ *with the property that* $\sum_{k=1}^{\infty} c_k \epsilon_k < \infty$.

(iv) An arbitrary positive scalar $\beta > 0$.

Furthermore, let the sequences $\{x^k\}_{k=1}^{\infty}, \{y^k\}_{k=1}^{\infty} \subset \mathbb{R}^n$ and $\{p^k\}_{k=0}^{\infty} \subset \mathbb{R}^m$ conform to the following recursions for all $k \geq 1$:

$$y^k \in \partial_x \Big[f(x) + \frac{1}{c_k} h^{\oplus}\big(\nabla h(p^{k-1}) + c_k g(x)\big) \Big]_{x=x^k} \tag{16}$$

$$\|y^k\| \leq \frac{\epsilon_k}{\max\{\beta, \|x^k\|\}} \tag{17}$$

$$p^k = \nabla h^{\oplus}\big(\nabla h(p^{k-1}) + c_k g(x^k)\big). \tag{18}$$

Here, the notation "$[\,\cdot\,]_{x=x^k}$" denotes evaluation at $x = x^k$. Below, when we refer to Algorithm 3.1, we implicitly assume that h and S conform to Assumption 2.7.

Note that it is possible to choose $\epsilon_k \equiv 0$, in which case $y^k \equiv 0$, and the recursions (16)-(18) reduce to

$$x^k \in \operatorname*{Arg\,min}_{x \in \mathbb{R}^n} \Big\{ f(x) + \frac{1}{c_k} h^{\oplus}\big(\nabla h(p^{k-1}) + c_k g(x)\big) \Big\} \tag{19}$$

$$p^k = \nabla h^{\oplus}\big(\nabla h(p^{k-1}) + c_k g(x^k)\big), \tag{20}$$

the generalized multiplier method suggested in [10, 27]. When $\epsilon_k > 0$, then the effect of (16)-(17) is to solve (19) approximately. However, the only information needed for this approximation is a subgradient (or equivalently in the smooth case, the gradient) of the augmented Lagrangian function

$$\phi_k(x) \stackrel{\text{def}}{=} f(x) + \frac{1}{c_k} h^{\oplus}\big(\nabla h(p^{k-1}) + c_k g(x)\big). \tag{21}$$

There is no requirement to ascertain ϵ-optimality, as in (9), or identify ϵ-subgradients.

Note that when there is some *a priori* bound B on $\|x^k\|$, for example, when $\operatorname{dom} f$ is bounded, we can set $\beta = B$ and define $\epsilon'_k = \epsilon_k / B$. Then, we can rewrite (16) as $\|y^k\| \leq \epsilon'_k$, where $\{\epsilon'_k\}$ is some summable nonnegative sequence. This type of condition has appeared before, for example in [15, eq. (7.18)], but was only used to verify a condition like (9) in the presence of strong convexity of f. The advantage of the analysis here is that no such assumption on f is required; however, if $\{x^k\}$ is not known to be bounded, the condition $\|y^k\| \le \epsilon'_k$, with $\{\epsilon'_k\}$ summable, must be strengthened to (17), with $\{\epsilon_k\}$ summable.

Note that the choice of algorithm for the approximate unconstrained minimization (16)-(17) is application-dependent, and this article leaves it unspecified. In general, however, one would expect to use some subsequentially convergent iterative unconstrained convex optimization method. One common choice would be if $f, g_1, \ldots, g_m$ are all twice differentiable, and h is chosen so that $h^\oplus$ is twice differentiable. Then ϕ_k will be twice differentiable, and a standard Newton method with an Armijo line search would be applicable to (16)-(17); see for example [4, Sections 1.2-1.4]. Examples of h that lead to twice-differentiable $h^\oplus$ include the aforementioned $h(p) = \sum_{i=1}^m p_i \log p_i$ with $S = \mathbb{R}^m_{++}$, and $h(p) = (2/3)\sum_{i=1}^m |p_i|^{3/2}$ with $S = \mathbb{R}^m$ [12, 14].

The approximation criterion (17) may seem somewhat nonconstructive, in that the bound on the norm of y^k cannot in general be determined in advance, but depends on the accompanying primal iterate x^k. However, in most practical situations, the condition should be easily attainable and verifiable. For instance, suppose that ϕ_k is differentiable and we apply some subsequentially convergent iterative procedure — such as the just-mentioned Newton method — to solving the subproblem (19). Let $\{x^{k,j}\}_{j=0}^\infty$ be the sequence of points generated by this procedure. Then there should be an infinite set $\mathcal{J} \subseteq \mathbb{N}$ such that $\lim_{j\to\infty, j\in\mathcal{J}} \nabla\phi_k(x^{k,j}) = 0$ and $\lim_{j\to\infty, j\in\mathcal{J}} x^{k,j}$ exists. Thus, we should eventually have $\|\nabla\phi_k(x^{k,j})\| \le \epsilon_k / \max\{\beta, \|x^{k,j}\|\}$ for large enough $j \in \mathcal{J}$. At this point, one may take $x^k = x^{k,j}$ and $y^k = \nabla\phi_k(x^{k,j})$.

Due to the somewhat technical nature of the arguments required, the question of the existence of a sequence $\{(x^k, y^k, p^k)\}$ satisfying the recursions (16)-(18) is deferred to Section 6. For the moment, we simply assume that such a sequence exists, and analyze its convergence properties.

In the past, the algorithm (19)-(20) has been analyzed by showing that it produces the same sequence $\{p^k\}$ as applying the generalized proximal point algorithm [8, 9, 10]

$$p^k = \arg\min_{p\in\mathbb{R}^m} \left\{ -Q_0(p) + \frac{1}{c_k} D_h(p, p^{k-1}) \right\} \tag{22}$$

to the negative of the dual function, $-Q_0(\cdot)$. This method is just (8) with $D(\cdot,\cdot)$ specialized to the form $D_h(\cdot,\cdot)$ of (14). Specifically, letting "$\oplus_1$" denote the monotone conjugate with respect to the first argument, one has

$$\begin{aligned} D^{\oplus_1}(u,p) &= D_h^{\oplus_1}(u,p) \\ &= \sup_{q\ge 0} \left\{ \langle q, u\rangle - (h(q) - h(p) - \langle \nabla h(p), q - p\rangle) \right\} \\ &= h^\oplus\Big(\nabla h(p) + u\Big) + h(p) + \langle \nabla h(p), p\rangle. \end{aligned}$$

Substituting this form into the minimand of (6), the resulting constant terms $h(p^{k-1}) + \langle \nabla h(p^{k-1}), p^{k-1} \rangle$ may be dropped, yielding the minimand of (16) and (19). The convergence properties of $\{p^k\}$ in (19)-(20) can then be deduced from the behavior of the generalized proximal algorithm (22).

However, this analytic approach has yielded at best unwieldy approximation criteria for (19). Here, we take a different approach, analyzing the action of the algorithm on the Lagrangian function $L(x, p)$ and the perturbed dual function $Q(y, p)$. That said, the structure of the convergence proof for the dual iterates $\{p^k\}$ is similar in structure to most convergence proofs for proximal algorithms, dating back at least to [21]:

- First, one shows that the sequence $\{p^k\}$ must be bounded (and, in this case, some additional related results).
- Second, one shows that any limit point of $\{p^k\}$ must be a dual solution.
- Finally, one shows that the $\{p^k\}$ has exactly one limit point.

We start by establishing some basic properties of the algorithm.

Lemma 3.2 *In Algorithm 3.1,* $\lim_{k\to\infty} y^k = 0$ *and* $\lim_{k\to\infty}\langle x^k, y^k\rangle = 0$.

Proof. Since $\sum_{k=1}^{\infty} c_k \epsilon_k < \infty$ and $\{c_k\}$ is bounded below, we must also have $\epsilon_k \to 0$. Combining this observation with (17), we have $\|y_k\| \le \epsilon_k/\beta \to 0$. Using (17) again,

$$\left|\langle x^k, y^k\rangle\right| \le \left\|x^k\right\| \left\|y^k\right\| \le \left\|x^k\right\| \frac{\epsilon_k}{\|x^k\|} = \epsilon_k \to 0.$$

□

Lemma 3.3 *Suppose Assumption 2.1 holds. In Algorithm 3.1, define*

$$u^k \stackrel{\text{def}}{=} \frac{1}{c_k}\left(\nabla h(p^{k-1}) - \nabla h(p^k)\right) \tag{23}$$

for all $k \ge 1$. Then, for all $k \ge 1$,

$$\begin{aligned} u^k &\in \left\{-g(x^k)\right\} + N_{\mathbb{R}^m_+}(p^k) && (24)\\ g(x^k) &\le -u^k && (25)\\ (y^k, u^k) &\in \partial L(x^k, p^k). && (26)\end{aligned}$$

Proof. We first note from (18) and the definition of $h^{\oplus}$ that, where "*" denotes the convex conjugate,

$$\begin{aligned} & p^k = \nabla\Big((h + \delta_{\mathbb{R}^m_+})^*\Big)\left(\nabla h(p^{k-1}) + c_k g(x^k)\right) \\ \Leftrightarrow\quad & \partial\Big(h + \delta_{\mathbb{R}^m_+}\Big)(p^k) \ni \nabla h(p^{k-1}) + c_k g(x^k). \end{aligned}$$

Since $\operatorname{ri}\operatorname{dom} h = S \supseteq \mathbb{R}^m_{++}$ and $\operatorname{ri}\operatorname{dom} \delta_{\mathbb{R}^m_+} = \mathbb{R}^m_{++}$ intersect, the sum rule for subgradients [19, Theorem 23.8] implies that $\partial\Big(h + \delta_{\mathbb{R}^m_+}\Big) = \nabla h + N_{\mathbb{R}^m_+}$. Thus, an equivalent condition is

$$
\begin{aligned}
& \left\{\nabla h(p^k)\right\} + N_{\mathbb{R}^m_+}(p^k) \ni \nabla h(p^{k-1}) + c_k g(x^k) \\
\Leftrightarrow \quad & \left\{-c_k g(x^k)\right\} + N_{\mathbb{R}^m_+}(p^k) \ni \nabla h(p^{k-1}) - \nabla h(p^k) \\
\Leftrightarrow \quad & \frac{1}{c_k}\left(\nabla h(p^{k-1}) - \nabla h(p^k)\right) \in \left\{-g(x^k)\right\} + N_{\mathbb{R}^m_+}(p^k),
\end{aligned}
$$

establishing (24). Since every element of $N_{\mathbb{R}^m_+}(p^k)$ is nonpositive, (25) follows immediately.

From $y^k \in \partial\phi_k(x^k)$ and the subdifferential chain rule of [10, Lemma A4],

$$
\begin{aligned}
y^k &= z^{0,k} + \sum_{i=1}^m \left[\nabla h^{\oplus}(\nabla h(p^{k-1}) + c_k g(x^k))\right]_i z^{i,k} + \nu^k \\
&= z^{0,k} + \sum_{i=1}^m p_i^k z^{i,k} + \nu^k
\end{aligned}
$$

for some $z^{0,k} \in \partial f(x^k)$, $z^{i,k} \in \partial g_i(x^k)$, $i = 1, \ldots, m$, and $\nu^k \in N_{\operatorname{dom}\phi_k}(x^k)$. Since Assumption 2.1 guarantees $\operatorname{dom}\phi_k = \operatorname{dom} f$, we can set

$$
\overline{z}^k = z^{0,k} + \nu^k \in \partial f(x^k) + N_{\operatorname{dom} f}(x^k) = \partial f(x^k),
$$

and we obtain

$$
y^k = \overline{z}^k + \sum_{i=1}^m p_i^k z^{i,k}.
$$

Thus, from [19, Theorem 23.8] and Assumption 2.1,

$$
y^k \in \partial f(x^k) + \sum_{i=1}^m p_i^k \partial g_i(x^k) = \partial\left[f + \sum_{i=1}^m p_i^k g_i\right](x^k).
$$

Then (26) follows from (13) and (24). □

Note that when f and the g_i are finite and differentiable, the second part of the proof can be simplified by using the classical chain rule to observe that

$$
\begin{aligned}
y^k &= \nabla f(x^k) + \nabla g(x^k)^\top \nabla h^{\oplus}\big(\nabla h(p^{k-1}) + c_k g(x^k)\big) \\
&= \nabla f(x^k) + \nabla g(x^k)^\top p^k \\
&= \nabla_1 L(x^k, p^k).
\end{aligned}
$$

Lemma 3.4 *Suppose Assumptions 2.1 and 2.4 hold, and let (x^*, p^*) be any KKT pair for (1). In Algorithm 3.1, we have for all $k \geq 1$ that*

$$
\langle p^k - p^*, u^k \rangle \geq -\left(1 + \frac{\|x^*\|}{\beta}\right)\epsilon_k. \tag{27}
$$

Proof. From the previous lemma, we have $(y^k, u^k) \in \partial L(x^k, p^k)$ for all $k \geq 1$. We also have $(0,0) \in \partial L(x^*, p^*)$, and by the monotonicity of ∂L, we conclude that

$$\langle x^k - x^*, y^k - 0\rangle + \langle p^k - p^*, u^k - 0\rangle \geq 0.$$

Rearranging this inequality and using (17), we obtain

$$\begin{aligned}
\langle p^k - p^*, u^k\rangle &\geq -\langle x^k - x^*, y^k\rangle \\
&\geq -\left\|x^k - x^*\right\| \left\|y^k\right\| \\
&\geq -\left(\left\|x^k\right\| \left\|y^k\right\| + \|x^*\| \left\|y^k\right\|\right) \\
&\geq -\left(\left\|x^k\right\| \frac{\epsilon_k}{\|x^k\|} + \|x^*\| \frac{\epsilon_k}{\beta}\right) \\
&= -\left(1 + \frac{\|x^*\|}{\beta}\right)\epsilon_k.
\end{aligned}$$

□

Lemma 3.5 *In Algorithm 3.1, for all $k \geq 1$ and any $\overline{p} \in \operatorname{cl} S$,*

$$D_h(\overline{p}, p^k) = D_h(\overline{p}, p^{k-1}) - D_h(p^k, p^{k-1}) - c_k\langle p^k - \overline{p}, u^k\rangle. \tag{28}$$

Proof. Apply Lemma 2.9 (the "three point" lemma) with $p = p^{k-1}$, $q = p^k$, and $r = \overline{p}$ to obtain

$$D_h(\overline{p}, p^{k-1}) = D_h(p^k, p^{k-1}) + D_h(\overline{p}, p^k) + \langle \nabla h(p^k) - \nabla h(p^{k-1}), \overline{p} - p^k\rangle.$$

Rearranging and using the definition of u^k,

$$\begin{aligned}
D_h(\overline{p}, p^k) &= D_h(\overline{p}, p^{k-1}) - D_h(p^k, p^{k-1}) - \langle \nabla h(p^k) - \nabla h(p^{k-1}), \overline{p} - p^k\rangle \\
&= D_h(\overline{p}, p^{k-1}) - D_h(p^k, p^{k-1}) + \langle c_k u^k, \overline{p} - p^k\rangle,
\end{aligned}$$

which is equivalent to (28). □

We are now in a position to prove that $\{p^k\}$ is bounded, along with several other useful facts.

Proposition 3.6 *Suppose Assumptions 2.1 and 2.4 hold, and let p^* be any dual solution to (1). In Algorithm 3.1, define*

$$w_k(p^*) \stackrel{\text{def}}{=} c_k\langle p^k - p^*, u^k\rangle.$$

Then the following hold:

(i) $\{p^k\}$ is a bounded sequence.

(ii) $\{D_h(p^*, p^k)\}$ *converges.*

(iii) $D_h(p^k, p^{k-1}) \to 0$, *and furthermore* $\sum_{k=1}^{\infty} D_h(p^k, p^{k-1}) < \infty$.

(iv) $\lim_{k\to\infty} w_k(p^*) = 0$, *and furthermore* $\sum_{k=1}^{\infty} w_k(p^*)$ *exists and is finite.*

(v) $\lim_{k\to\infty}\langle p^k - p^*, u^k\rangle = 0$.

Proof. From Assumption 2.4 and Proposition 2.5, there exists some $x^* \in \mathbb{R}^n$ such that (x^*, p^*) is a KKT pair. Furthermore, from Lemma 3.5 with $\overline{p} = p^*$, we have

$$D_h(p^*, p^k) = D_h(p^*, p^{k-1}) - D_h(p^k, p^{k-1}) - c_k\langle p^k - p^*, u^k\rangle. \tag{29}$$

By induction, we may then conclude for all $k \geq 1$ that

$$\begin{aligned} D_h(p^*, p^k) &= D_h(p^*, p^0) - \sum_{j=1}^{k}\left[D_h(p^j, p^{j-1}) + c_j\langle p^j - p^*, u^j\rangle\right] \\ &= D_h(p^*, p^0) - \sum_{j=1}^{k} D_h(p^j, p^{j-1}) - \sum_{j=1}^{k} w_j(p^*). \end{aligned} \tag{30}$$

From Lemma 3.4, we have

$$w_j(p^*) \geq -c_j\left(1 + \frac{\|x^*\|}{\beta}\right)\epsilon_j,$$

and therefore

$$\sum_{j=1}^{k} w_j(p^*) \geq -\left(1 + \frac{\|x^*\|}{\beta}\right)\sum_{j=1}^{k} c_j\epsilon_j \geq -\left(1 + \frac{\|x^*\|}{\beta}\right)\sum_{j=1}^{\infty} c_j\epsilon_j.$$

Let $E = \sum_{k=1}^{\infty} c_k\epsilon_k < \infty$. We then have from (30) that, for all $k \geq 1$,

$$D_h(p^*, p^k) \leq D_h(p^*, p^0) + \left(1 + \frac{\|x^*\|}{\beta}\right)E - \sum_{j=1}^{k} D_h(p^j, p^{j-1}). \tag{31}$$

Applying induction to this relation and using that $D_h(p^*, p^k)$ and $D_h(p^k, p^{k-1})$ must both be nonnegative for all k, we conclude that

$$\sum_{k=1}^{\infty} D_h(p^k, p^{k-1}) \leq D_h(p^*, p^0) + \left(1 + \frac{\|x^*\|}{\beta}\right)E < \infty,$$

which proves (iii). Using the nonnegativity of $D_h(p^k, p^{k-1})$ in (31), we also have

$$D_h(p^*, p^k) \leq D_h(p^*, p^0) + \left(1 + \frac{\|x^*\|}{\beta}\right)E,$$

for all $k \geq 1$, and hence that $\{p^k\}$ must be bounded by the level set properties of D_h; thus, (i) also holds.

We next establish (iv). To this end, note that from the nonnegativity of the D_h function, (30) implies $\sum_{j=1}^{k} w_j(p^*) \leq D_h(p^*, p^0)$ for all $k \geq 1$. Then, for all $k \geq 1$, let

$$\begin{aligned} a_k &\stackrel{\text{def}}{=} -\sum_{j=1}^{k} w_j(p^*) \\ \gamma_k &\stackrel{\text{def}}{=} \left(1 + \frac{\|x^*\|}{\beta}\right) c_{k+1}\epsilon_{k+1}. \end{aligned}$$

Then, a_k is bounded below, $\sum_{k=1}^{\infty} \gamma_k$ exists and is finite, and

$$\begin{aligned} a_{k+1} &= -\sum_{j=1}^{k+1} w_j(p^*) \\ &= -\sum_{j=1}^{k} w_j(p^*) + (-w_{k+1}(p^*)) \\ &\leq -\sum_{j=1}^{k} w_j(p^*) + \left(1 + \frac{\|x^*\|}{\beta}\right) c_{k+1}\epsilon_{k+1} \\ &= a_k + \gamma_k, \end{aligned}$$

so the hypotheses of Lemma 2.10 hold. From that lemma, we may conclude that

$$\sum_{k=1}^{\infty} w_k(p^*) = -\lim_{k\to\infty} a_k$$

exists and is finite, and consequently $w_k(p^*) \to 0$, completing the proof of (iv).

Since $\{c_k\}$ is bounded below, (v) follows directly from (iv).

Finally, we address (ii). Using the nonnegativy of $D_h(p^k, p^{k-1})$ in (29), we have for all $k \geq 1$ that

$$D_h(p^*, p^k) \leq D_h(p^*, p^{k-1}) - w_k(p^*).$$

Again applying Lemma 2.10, this time with $a_k = D_h(p^*, p^k)$ (bounded below by 0) and $\gamma_k = -w_k(p^*)$ (summable by (iv)), we conclude that $\{D_h(p^*, p^k)\}$ is convergent. □

We now know that $\{p^k\}$ must have limit points, since it is bounded. Following the standard pattern, but again using conjugate duality, we now establish the dual optimality of these limit points:

Proposition 3.7 *Let the convex program (1) satisfy Assumptions 2.1 and 2.4. Then every limit point of the sequence $\{p^k\}$ generated by Algorithm 3.1 is a dual solution.*

Proof. Let (x^*, p^*) be some KKT pair for (1), which must exist by hypothesis. From Lemma 3.3, $(y^k, u^k) \in \partial L(x^k, p^k)$, so it then follows from Proposition 2.3 that $(x^k, u^k) \in \partial Q(y^k, p^k)$. Applying (10) with $(y', p') = (0, p^*)$ and $(x, u, y, p) = (x^k, u^k, y^k, p^k)$, we have

$$\begin{aligned} Q(0, p^*) &\leq Q(y^k, p^k) - \langle x^k, 0 - y^k \rangle - \langle u^k, p^* - p^k \rangle & (32) \\ &= Q(y^k, p^k) + \langle x^k, y^k \rangle + \langle p^k - p^*, u^k \rangle. & (33) \end{aligned}$$

Now consider any limit point p^∞ of $\{p^k\}$, and let $\mathcal{K} \subseteq \mathbb{N}$ be an infinite sequence of indices with $p^k \to_{\mathcal{K}} p^\infty$. Next, we would like to take limits in (32)/(33). From Lemma 3.2, one has $y^k \to 0$ and $\langle x^k, y^k \rangle \to 0$. From Proposition 3.6(v), we also have $\langle p^k - p^*, u^k \rangle \to 0$.

Thus, taking limits over $\mathcal{K}$ and using from that Q is a closed (upper semicontinuous) concave function, it follows from (32)/(33) that

$$\begin{aligned} f^* = Q(0, p^*) &\leq \liminf_{\substack{k \to \infty \\ k \in \mathcal{K}}} Q(y^k, p^k) & (34) \\ &\leq \limsup_{\substack{k \to \infty \\ k \in \mathcal{K}}} Q(y^k, p^k) & (35) \\ &\leq Q(\lim_{\substack{k \to \infty \\ k \in \mathcal{K}}} y^k, \lim_{\substack{k \to \infty \\ k \in \mathcal{K}}} p^k) & (36) \\ &= Q(0, p^\infty) \leq f^*, & (37) \end{aligned}$$

where the last inequality is a consequence of f^* being the maximum value of Q_0. Thus, $Q_0(p^\infty) = Q_0(p^*) = f^*$. □

At this point, it is now possible to assemble the above results and show that $\{p^k\}$ has a unique limit:

Proposition 3.8 *Let the convex program (1) satisfy Assumptions 2.1 and 2.4. Then the sequence $\{p^k\}$ generated by Algorithm 3.1 converges to a dual solution. Furthermore, one also has $\lim_{k\to\infty} Q(y^k, p^k) = f^*$, the optimal value of (1).*

Proof. From Assumption 2.4 and Proposition 3.6, we know that $\{p^k\}$ is bounded. Let p^∞ be any of its limit points, with associated subsequence $\mathcal{K}$. By Proposition 3.7, we know p^∞ is a dual solution. From Assumption 2.7(v), $D_h(p^\infty, p^k) \to_{\mathcal{K}} 0$. From Proposition 3.6(ii) with $p^* = p^\infty$, however, we know that the entire sequence $\{D_h(p^\infty, p^k)\}$ is convergent. Therefore, the limit of the entire sequence $\{D_h(p^\infty, p^k)\}$ must be 0. From Proposition 2.8, it then follows that $p^k \to p^\infty$. Since this is the case and p^∞ is a dual solution, one may set $\mathcal{K} = \mathbb{N}$ in (34)-(37) to obtain $\liminf_{k\to\infty} Q(y^k, p^k) = \limsup_{k\to\infty} Q(y^k, p^k) = \lim_{k\to\infty} Q(y^k, p^k) = f^*$.
□

4 Behavior of the Primal Iterates

In this section, we will analyze the behavior of the primal sequence $\{x^k\}$ generated by Algorithm 3.1. Although simplified and reorganized, most of the ideas largely follow those presented in [15, Sections 8 and 9], which treats less practical approximation criteria, but more general forms of the Bregman function h, distance measure D_h, and resulting augmented Lagrangians ϕ_k. The exception is Proposition 4.7, whose analysis is patterned more after [25].

We call a sequence $\{z^k\}_{k=1}^{\infty} \subset \mathbb{R}^n$ *asymptotically optimal* for the convex program (1), with optimal value f^*, if

$$\begin{aligned} \limsup_{k\to\infty} g_i(z^k) &\leq 0 \quad \forall\, i = 1, \ldots, m \\ \limsup_{k\to\infty} f(z^k) &\leq f^*. \end{aligned}$$

Such a sequence, even if it does not converge, can provide arbitrarily good approximate solutions to (1). We summarize the properties of such sequences in the proposition below; related results may be found in [19, Chapter 27] and [3, Section 5.3].

Proposition 4.1 *Let $\{z^k\}_{k=1}^{\infty}$ be asymptotically optimal for the convex program (1), which has finite optimal value f^*. Then*

(i) All limit points of $\{z^k\}$ are solutions of (1).

(ii) If $\{z^k\}$ is bounded, then $\lim_{k\to\infty} f(z^k) = f^*$.

(iii) If the set X^* of optimal primal solutions to (1) is bounded, then $\{z^k\}$ is bounded, and $\lim_{k\to\infty} f(z^k) = f^*$.

(iv) If the primal optimal solution is unique, that is, X^* is a singleton set $\{x^*\}$, then $\lim_{k\to\infty} z^k = x^*$.

Proof. To establish (i), suppose that z^∞ is any limit point of $\{z^k\}$, and $\mathcal{K} \subseteq \mathbb{N}$ is some corresponding infinite set such that $z^k \to_{\mathcal{K}} z^\infty$. From the definition of asymptotic optimality and the lower semicontinuity of $f, g_1, \ldots, g_m$, we have

$$\begin{aligned} g_i(z^\infty) &\leq \limsup_{\substack{k\to\infty \\ k\in\mathcal{K}}} g_i(z^k) \leq 0 \qquad i = 1, \ldots, m, \\ f(z^\infty) &\leq \limsup_{\substack{k\to\infty \\ k\in\mathcal{K}}} f(z^k) \leq f^*. \end{aligned}$$

Thus, z^∞ is feasible for (1) and has $f(z^\infty) \leq f^*$. It follows that z^∞ is an optimal solution and $f(z^\infty) = f^*$, and so (i) holds.

To prove (ii), let f^∞ be any extended-real-valued limit point of the scalar sequence $\{f(z^k)\}_{k=1}^{\infty}$. Let $\mathcal{K} \subseteq \mathbb{N}$ be some infinite set such that $f(z^k) \to_{\mathcal{K}} f^\infty$. If $\{z^k\}$ is bounded, there exists an infinite set $\mathcal{K}' \subseteq \mathcal{K}$ such that $\{z^k\}_{k\in\mathcal{K}'}$ converges to some limit z^∞. By (i), z^∞ is an optimal solution, and thus $f(z^\infty) = f^*$. By the lower semicontinuity of f, taking limits over $k \in \mathcal{K}'$ gives $f^* = f(z^\infty) \le \lim_{k\to\infty, k\in\mathcal{K}'} f(z^k) = f^\infty$. But asymptotic optimality gives $f^\infty \le \limsup_{k\to\infty} f(z^k) \le f^*$, so $f^\infty = f^*$. By the arbitrary choice of f^∞, $f(z^k) \to f^*$.

Next, consider (iii), and suppose the primal optimal solution set X^* is bounded. Define $\psi : \mathbb{R}^n \to (-\infty, +\infty]$ via

$$\psi(x) = \max\{f(x) - f^*, g_1(x), g_2(x), \ldots, g_m(x)\},$$

and note that X^* is equal to the 0-level set of ψ, that is,

$$L_\psi(0) \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid \psi(x) \le 0\} = X^* \ne \emptyset.$$

Now, ψ is a closed convex function because $f, g_1, \ldots, g_m$ are closed, and proper since it is finite on $X^* \ne \emptyset$. Next, pick any $\epsilon > 0$ and consider

$$L_\psi(\epsilon) \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid \psi(x) \le \epsilon\} \supseteq L_\psi(0) = X^* \ne \emptyset.$$

Since all nonempty level sets of a closed proper convex function have the same recesion cone [19, Theorem 8.7], the boundedness of $L_\psi(0)$ implies that $L_\psi(\epsilon)$ is also bounded. Now, asymptotic optimality implies that for all sufficiently large k, $z^k \in L_\psi(\epsilon)$, so $\{z^k\}$ must be bounded. In view of (ii), the proof of (iii) is complete.

Only (iv) remains. Suppose that $X^* = \{x^*\}$ is a singleton. Then X^* is bounded, so (iii) implies that $\{z^k\}$ is bounded, while (i) implies that its only possible limit point is x^*. Therefore, $z^k \to x^*$. □

The goal of the remainder of this section is to show, under various assumptions, that the sequence $\{x^k\}$ generated by Algorithm 3.1, or a related sequence $\{\breve{x}^\ell\}_{\ell=1}^\infty$ defined below, is asymptotically optimal. To this end, the following identity is used in several places:

Lemma 4.2 *In Algorithm 3.1, one has, for all $k \ge 1$, that*

$$f(x^k) = Q(y^k, p^k) + \langle x^k, y^k\rangle + \langle u^k, p^k\rangle. \tag{38}$$

Proof. The perturbed dual objective Q is minus the convex conjugate of the perturbed primal objective F, and $(y^k, p^k) \in \partial F(x^k, u^k)$ by (26) and Proposition 2.3. So, [19, Theorem 23.5] gives

$$F(x^k, u^k) - Q(y^k, p^k) = \left\langle (x^k, u^k), (y^k, p^k)\right\rangle. \tag{39}$$

From (25), Lemma 3.3 also gives $g(x^k) + u^k \le 0$ for all $k \ge 1$, implying $F(x^k, u^k) = f(x^k)$. Substituting this equality into (39) and solving for $f(x^k)$ yields (38). □

It is now possible to derive a strong result for the $S \supset \mathbb{R}^m_+$ case of Algorithm 3.1, which includes the classic quadratic case.

Proposition 4.3 *Suppose the convex program (1) satisfies Assumptions 2.1 and 2.4, and $S \supset \mathbb{R}^m_+$ in Algorithm 3.1. Then the primal sequence $\{x^k\}$ generated by Algorithm 3.1 is asymptotically optimal, and one always has $\lim_{k\to\infty} f(x^k) = f^*$.*

Proof. The dual sequence $\{p^k\}$ must be convergent from Proposition 3.8, so let $p^\infty \geq 0$ be its limit. Since $\{c_k\}$ is bounded below and ∇h must be continuous at $p^\infty \in \mathbb{R}^m_+ \subset S$,

$$u^k = \frac{1}{c_k}\left(\nabla h(p^{k-1}) - \nabla h(p^k)\right) \to 0.$$

So, taking limits in (25), $\limsup_{k\to\infty} g_i(x^k) \leq 0$ for $i = 1, \ldots, m$.

Now consider the three terms on the right-hand side of (38). $Q(y^k, p^k) \to f^*$ from Proposition 3.8, and $\langle x^k, y^k\rangle \to 0$ from Lemma 3.2. Since $u^k \to 0$ and $p^k \to p^\infty$, one also has $\langle u^k, p^k\rangle \to \langle 0, p^\infty\rangle = 0$. Thus, taking limits in (38) yields $\lim_{k\to\infty} f(x^k) = f^*$, and the proof is complete. □

Without the assumption $S \supset \mathbb{R}^m_+$, the analysis becomes much more delicate: unless all the constraints $g(x) \leq 0$ are binding, the limit p^∞ of $\{p^k\}$ will typically lie in $\operatorname{bd}\mathbb{R}^m_+$, where ∇h may not exist, making the properties of the sequence $\{u^k\}$ hard to control.

The analysis is greatly simplified and expedited in the $S \not\supset \mathbb{R}^m_+$ case if one assumes the common situation that h is separable, that is,

$$h(p) = \sum_{i=1}^m h_i(p_i),$$

where $h_1, \ldots, h_m$ are Bregman functions on $\mathbb{R}$ with respectives zones $S_1, \ldots, S_m \supseteq (0, \infty)$. This assumption is met by all the standard choices of h that have $S = \mathbb{R}^m_{++}$. In this case, the sequence $\{f(x^k)\}$ is quickly seen to be well-behaved, as follows:

Proposition 4.4 *Let the convex program (1) obey Assumptions 2.1 and 2.4, and suppose h is separable in Algorithm 3.1. Then $\lim_{k\to\infty} f(x^k) = f^*$, the optimal value of (1).*

Proof. Consider, via Proposition 3.8, the dual solution $p^\infty = \lim_{k\to\infty} p^k$. Adding and subtracting $\langle u^k, p^\infty\rangle$ in (38), one obtains

$$f(x^k) = Q(y^k, p^k) + \langle x^k, y^k\rangle + \langle u^k, p^k - p^\infty\rangle + \langle u^k, p^\infty\rangle.$$

Consider the four terms on the right-hand side of this equation. As before, Proposition 3.8 asserts that $Q(y^k, p^k) \to f^*$, while Lemma 3.2 gives $\langle x^k, y^k\rangle \to 0$. From the dual optimality of p^∞ and Proposition 3.6(v), the third term, $\langle u^k, p^k - p^\infty\rangle$, also converges to zero. It remains to consider the last term,

$$\langle u^k, p^\infty\rangle = \sum_{i=1}^m u_i^k p_i^\infty.$$

For each $i \in \{1, \ldots, m\}$, either $p_i^\infty = 0$ or $p_i^\infty > 0$. In the $p_i^\infty = 0$ case, $u_i^k p_i^\infty = 0$ for all $k \geq 1$. On the other hand, when $p_i^\infty > 0$, h_i' is continuous at p_i^∞, so $u_i^k = (1/c_k)(h_i'(p_i^{k-1}) - h_i'(p_i^k)) \to 0$, where we again use that $\{c_k\}$ is bounded below, and thus $u_i^k p_i^\infty \to 0$. Thus, $\langle u^k, p^\infty\rangle \to 0$, and $f(x^k) \to f^* + 0 + 0 + 0 = f^*$. □

Thus, once one has assumed h to be separable, the main challenge is to show that $\{x^k\}$ is asymptotically feasible, that is, $\limsup_{k\to\infty} g_i(x^k) \le 0$ for $i = 1, \ldots, m$. As with the exact form of the algorithm, it does not appear that asymptotic feasibility can be established without further assumptions.

At this point, we examine two alternative approaches: in the first, one considers, instead of $\{x^k\}$, a related sequence $\{\breve{x}^\ell\}_{\ell=1}^\infty$ defined via

$$s_\ell \stackrel{\text{def}}{=} \sum_{k=1}^{\ell} c_k \qquad\qquad \breve{x}^\ell \stackrel{\text{def}}{=} \frac{1}{s_\ell}\sum_{k=1}^{\ell} c_k x^k. \tag{40}$$

The consideration of $\{\breve{x}^\ell\}$, known as the *ergodic* sequence, was introduced to the analysis the exponential method of multipliers (4)-(5) in [30], and then generalized in [15, Lemma 8.10].

Proposition 4.5 *Suppose the convex program (1) meets Assumptions 2.1 and 2.4, and suppose h is separable. Then the sequence $\{\breve{x}^\ell\}_{\ell=1}^\infty$ derived from Algorithm 3.1 by (40) is asymptotically optimal.*

Proof. First, we claim that $\{\nabla h(p^k)\}_{k=0}^\infty$ is bounded above, that is,

$$\limsup_{k\to\infty} h_i'(p_i^k) < \infty \qquad \forall\, i = 1, \ldots, m.$$

To see this, take any $i \in \{1, \ldots, m\}$ and note that convexity of h_i yields

$$h_i(p_i^k + 1) \ge h_i(p_i^k) + h_i'(p_i^k)\cdot 1,$$

and thus

$$h_i'(p_i^k) \le h_i(p_i^k + 1) - h_i(p_i^k).$$

By the continuity and finiteness of h on $\operatorname{cl} S \supseteq \mathbb{R}^m_+$, taking limits produces

$$\limsup_{k\to\infty} h_i'(p_i^k) \le h_i(p_i^\infty + 1) - h_i(p_i^\infty) < \infty.$$

For $k = 1, \ldots, \ell$, (25) gives $c_k g(x^k) \le -c_k u^k = \nabla h(p^k) - \nabla h(p^{k-1})$. Adding this relation for $k = 1, \ldots, \ell$, one obtains

$$\sum_{k=1}^{\ell} c_k g(x^k) \le \nabla h(p^\ell) - \nabla h(p^0).$$

Dividing by s_ℓ and using the convexity of $g_1, \ldots, g_m$ then produces

$$g(\breve{x}^\ell) = g\!\left(\sum_{k=1}^{\ell} \frac{c_k}{s_k} x^k\right) \le \sum_{k=1}^{\ell} \frac{c_k}{s_k} g(x^k) \le \frac{1}{s_k}\left(\nabla h(p^\ell) - \nabla h(p^0)\right).$$

Since $\{c_k\}$ is bounded below, $\lim_{\ell\to\infty} s_\ell = \infty$. So, using that $\{\nabla h(p^\ell\}$ is bounded above, taking limits for $i = 1,\ldots,m$ gives $\limsup_{\ell\to\infty} g_i(\breve{x}^\ell) \le 0$, and $\{\breve{x}^\ell\}$ is asymptotically feasible.

Finally, using the convexity of f,

$$f(\breve{x}^\ell) = f\left(\sum_{k=1}^{\ell} \frac{c_k}{s_k} x^k\right) \le \sum_{k=1}^{\ell} \frac{c_k}{s_k} f(x^k).$$

Since $f(x^k) \to f^*$ by Proposition 4.4, one must also have $\sum_{k=1}^{\ell}(c_k/s_k)f(x^k) \to f^*$ as $\ell \to \infty$. Thus, taking limits above yields $\limsup_{\ell\to\infty} f(\breve{x}^\ell) \le f^*$, and $\{\breve{x}^\ell\}$ is asymptotically optimal. □

Corollary 4.6 *Suppose the convex program (1) meets Assumptions 2.1 and 2.4, h is separable, and the sequence $\{x^k\}$ produced by Algorithm 3.1 converges to some limit x^∞. Then x^∞ is a solution of (1).*

Proof. If $x^k \to x^\infty$, then we must also have $\lim_{\ell\to\infty} \breve{x}^\ell = x^\infty$. By the preceding proposition, $\{\breve{x}^\ell\}$ is asymptotically optimal, so x^∞ must solve (1) by Proposition 4.1(i). □

Note that, by generalizing the proof slightly, the asymptotic feasibility portion of Proposition 4.5 can be shown to hold even if h is not separable; in this case, however, Proposition 4.4 would not apply. By another minor modification to the analysis, Proposition 4.4, Proposition 4.5, and Corollary 4.6 can also be shown to hold if only the "coercive part" of h is separable, that is, $h(p) = h^\diamond(p) + \sum_{i=1}^m h_i^\triangleright(p_i)$, where $h^\diamond(p)$ is a potentially non-separable function continuously differentiable throughout $\mathbb{R}^m_+$, and $h_1^\triangleright,\ldots,h_m^\triangleright : \mathbb{R} \to (-\infty,+\infty]$ need only be finite and continously differentiable on $(0,+\infty)$. These slightly extended results are not presented here since their practical applications are unclear.

Corollary 4.6 should in fact apply to the most common cases encountered in practice, in which one might apply — for example — a Newton method to satisfy (16)-(17), using x^{k-1} as a starting point. Since $\{p^k\}$ is convergent, the augmented Lagrangian function ϕ_k will eventually change little from iteration to iteration, the starting point x^{k-1} of the subproblem algorithm will be nearly optimal, and thus the subproblem ending point x^k should lie nearby. Thus, one would typically expect $\{x^k\}$ to converge, allowing application of Corollary 4.6. Making these observations rigorous would entail further assumptions about (1) and (at least partial) specification of the algorithm used for the subproblem (19), and is thus outside the scope of this article.

Situations in which $S \not\supset \mathbb{R}^m_+$ and $\{x^k\}$ is not convergent, however, might require one to resort to the sequence $\{\breve{x}^\ell\}$ to construct an approximate solution to (1), which could be undesirable since $\{\breve{x}^\ell\}$ may converge slowly. Thus, we also present an alternative approach that forces $\{x^k\}$ itself to be asymptotically optimal through additional assumptions on h and $\{c_k\}$. The techniques here are adapted from [25].

Proposition 4.7 *Suppose the convex program (1) satisfies Assumptions 2.1 and 2.4, and consider Algorithm 3.1 with h separable. Let $H \subseteq \{1,\ldots,m\}$ denote the set of indices i for which h_i is continuously differentiable at 0, and let $\overline{H} = \{1,\ldots,m\} \setminus H$. Suppose that h and $\{c^k\}$ satisfy the following additional assumptions:*

(a) h_i *is twice continuously differentiable on* $\mathbb{R}^m_{++}$ *for all* $i \in \overline{H}$.

(b) There exists an $\epsilon > 0$ *such that, for all* $i = 1, \ldots, m$, h''_i *is nonincreasing on the interval* $(0, \epsilon)$ *for all* $i \in \overline{H}$.

(c) The stepsize c_k *at each iteration* k *is chosen to satisfy, for two fixed positive scalars* α_1, α_2,

$$c_k \geq \alpha_1 \qquad\qquad c_k \geq \alpha_2 \max_{i \in \overline{H}} \left\{ h''_i(p_i^{k-1}) \right\}. \tag{41}$$

Then $\{x^k\}$ *is asymptotically optimal for (1)*

Proof. Since h is separable, Proposition 4.4 asserts that $f(x^k) \to f^*$, so it remains only to show asymptotic feasibility. Because $g(x^k) \leq -u^k$ from (25), it is sufficient to show that $\liminf_{k\to\infty} u_i^k \geq 0$ for all $i \in \{1, \ldots, m\}$.

Consider first indices i for which $i \in H$ or $p_i^\infty > 0$. For such i, as in the $p_i^\infty > 0$ case in the proof of Proposition 4.4, and we have $u_i^k = (1/c_k)\left(h'_i(p_i^{k-1}) - h'_i(p_i^k)\right) \to 0$, since $\{c_k\}$ is bounded below by α_1 and h'_i must be continuous at $p_i^\infty = \lim_{k\to\infty} p_i^k$. Clearly, then, $\liminf_{k\to\infty} u_i^k \geq 0$ holds for such i.

For the remaining indices i, we must have $p_i^k \to 0$ and $i \in \overline{H}$. For the purpose of contradiction, take any such i, and suppose that $\liminf_{k\to\infty} u_i^k < 0$. Then there must exist a $\zeta > 0$ and an infinite set $\mathcal{K} \subseteq \mathbb{N}$ such that $u_i^k < -\zeta$ for all $k \in \mathcal{K}$. Since $u_i^k = (1/c_k)\left(h'_i(p_i^{k-1}) - h'_i(p_i^k)\right)$, $c_k > 0$, and h'_i is nondecreasing from the convexity of h, we have $p_i^{k-1} < p_i^k$ for all $k \in \mathcal{K}$. Since h'_i is also differentiable, the mean value theorem implies the existence for all $k \in \mathcal{K}$ of some $\tilde{p}_i^k \in [p_i^{k-1}, p_i^k]$ such that

$$h'_i(p_i^k) - h'_i(p_i^{k-1}) = h''_i(\tilde{p}_i^k)(p_i^k - p_i^{k-1}).$$

Then, for all $k \in \mathcal{K}$,

$$\begin{aligned}
\zeta \leq -u_i^k &= \frac{1}{c_k}\left(h'_i(p_i^{k-1}) - h'_i(p_i^k)\right) \\
&= \frac{1}{c_k} h''_i(\tilde{p}_i^k)(p_i^k - p_i^{k-1}) \\
&\leq \frac{1}{\alpha_1 \max_{i\in\overline{H}} \left\{h''_i(p_i^{k-1})\right\}} h''_i(\tilde{p}_i^k)(p_i^k - p_i^{k-1}) \\
&\leq \frac{1}{\alpha_1 h''_i(p_i^{k-1})} h''_i(\tilde{p}_i^k)(p_i^k - p_i^{k-1}),
\end{aligned} \tag{42}$$

where the first inequality follows from (41).

Since $p_i^k \to 0$, there exists a $\bar{k} \geq 1$ such that $p_i^k < \epsilon$ for all $k \geq \bar{k}$. For all $k \geq \bar{k}$ in $\mathcal{K}$, we then have $0 < p_i^{k-1} \leq \tilde{p}_i^k \leq p_i^k < \epsilon$, and since h''_i is nonincreasing on $(0, \epsilon)$, it follows that $h''_i(p_i^{k-1}) \geq h''_i(\tilde{p}_i^k)$. Using this relation in (42), one obtains $\zeta \leq (1/\alpha_1)(p_i^k - p_i^{k-1})$ for all $k \geq \bar{k}$ in $\mathcal{K}$, which is a contradiction since $p_i^k \to 0$.

We now have that $\liminf_{k\to\infty} u_i^k \geq 0$ for all $i = 1, \ldots, m$, and the proof is complete. □

Note that hypotheses (a) and (b) of Proposition 4.7 are derived from [25, Assumption 2.1 and Lemma 2.8]. Despite their somewhat restrictive appearance, they are in fact met, along with separability, by all the standard Bregman functions h that have been suggested in practice for the case $S = \mathbb{R}^m_{++}$. Hypothesis (c), which derives from [25, Section 4], is much more restrictive, but a large c_k appears necessary to force the $c_k g(x)$ term to play a sufficiently large role in the augmented Lagrangian ϕ_k that asymptotic feasibility is attained.

In practice, it would be preferable to use a separate c_k for each constraint, that is, to replace c_k by a diagonal matrix

$$C^k = \operatorname{diag}\{c_{k,1}, c_{k,2}, \ldots, c_{k,m}\},$$

where $c_{k,i} \geq \alpha_1 \max\{\alpha_2, h_i''(p_i^{k-1})\}$ for all $k \geq 1$ and $i = 1, \ldots, m$. This kind of approach has been analyzed in [2, 25, 30], and is also discussed in [4, Section 4.2.8]. Generally, however, one can only prove subsequential convergence of $\{p^k\}$, rather than full convergence; also, the analysis does not follow the same basic outline as in Section 3 above. One topic for possible further research is whether the analyses of [2, 25] can be modified to use an approximation criterion like (16)-(17).

5 Extension to Affine Equality Constraints

As mentioned in the introduction, one may extend the results here to the case that (1) has $\overline{m}$ additional equality constraints in the form $Ax = b$, where A is an $\overline{m} \times n$ real matrix and $b \in \mathbb{R}^{\overline{m}}$. To do so, one chooses a second Bregman function $\overline{h}$ whose zone is all of $\mathbb{R}^{\overline{m}}$, that is, $\overline{h}$ satisfies Assumption 2.7 with $S = \operatorname{cl} S = \mathbb{R}^{\overline{m}}$. We introduce an additional sequence of multiplier estimates $\{\overline{p}^k\}_{k=0}^{\infty} \subset \mathbb{R}^{\overline{m}}$, with $\overline{p}^0$ chosen arbitrarily, and consider the algorithm:

$$\begin{aligned}
y^k &\in \partial_x\left[f(x) + \frac{1}{c_k}\left[h^{\oplus}\big(\nabla h(p^{k-1}) + c_k g(x)\big) + \overline{h}^*\big(\nabla\overline{h}(\overline{p}^{k-1}) + c_k(Ax-b)\big)\right]\right]_{x=x^k} \\
\|y^k\| &\leq \frac{\epsilon_k}{\max\{\beta, \|x^k\|\}} \\
p^k &= \nabla h^{\oplus}\big(\nabla h(p^{k-1}) + c_k g(x^k)\big) \\
\overline{p}^k &= \nabla\overline{h}^*\big(\nabla\overline{h}(\overline{p}^{k-1}) + c_k(Ax^k - b)\big).
\end{aligned}$$

Here, "*" again denotes the usual convex conjugate. This algorithm, by minor variations on the analysis above, can be shown to yield a sequence of multipliers $\{(p^k, \overline{p}^k)\}$ converging to a dual optimum, with the primal sequence $\{x^k\}$ satisfying $Ax^k \to b$, and otherwise behaving as analyzed in Section 4 above.

If there are no inequality constraints, one similarly obtains

$$\begin{aligned}
y^k &\in \partial_x\left[f(x) + \frac{1}{c_k}\overline{h}^*\big(\nabla\overline{h}(\overline{p}^{k-1}) + c_k(Ax-b)\big)\right]_{x=x^k} \\
\|y^k\| &\leq \frac{\epsilon_k}{\max\{\beta, \|x^k\|\}} \\
\overline{p}^k &= \nabla\overline{h}^*\big(\nabla\overline{h}(\overline{p}^{k-1}) + c_k(Ax^k - b)\big),
\end{aligned}$$

with $\{\bar{p}^k\}$ converging to a dual solution, and $\{x^k\}$ behaving much as in Proposition 4.3.

6 Existence Results

This section establishes sufficient conditions for a sequence $\{(x^k, y^k, p^k)\}$ satisfying the recursions (16)-(18) of Algorithm 3.1 to exist. Because (18) gives a formula for p^k in terms of x^k and y^k, this question hinges on the existence of $x^k, y^k \in \mathbb{R}^n$ meeting (16)-(17). There are two cases, $\epsilon_k > 0$ and $\epsilon_k = 0$.

The $\epsilon_k = 0$ case forces $y^k = 0$, and requires that x^k be an exact minimizer of ϕ_k, as defined in (21). The existence question for x^k then reduces to that of the already well-studied algorithm (19)-(20) and its generalizations. For completeness, these results are stated below; it is generally necessary to make some additional assumptions on (1) to guarantee the existence of x^k, typically that the optimal solution set X^* is bounded.

Proposition 6.1 *Suppose that Assumptions 2.1 and 2.4 hold, and that the optimal solution set X^* of (1) is nonempty and compact. Then, for any $p^{k-1} \in S$ and $c_k > 0$, there exists an x^k minimizing the function ϕ_k of (21). This x^k, along with $y^k = 0$, satisfies the recursions (16)-(17) of Algorithm 3.1 for any choice of $\epsilon_k \geq 0$.*

Proof. The result is a special case of [24, Lemma 2.2.9] (see also [25, Theorem 3.8] for a similar result for $f, g_1, \ldots, g_m$ finite and differentiable). $\square$

Corollary 6.2 *Suppose that Assumptions 2.1 and 2.4 hold, and that the optimal solution set X^* of (1) is nonempty and compact. Then there exist infinite sequences $\{x^k\}_{k=1}^{\infty} \subset \mathbb{R}^n$, $\{y^k\}_{k=1}^{\infty} \subset \mathbb{R}^n$, and $\{p^k\}_{k=1}^{\infty} \subset \mathbb{R}^m$ conforming to the recursions of Algorithm 3.1.*

Proof. Apply Proposition 6.1 inductively. $\square$

On the other hand, if one does not allow $\epsilon_k = 0$, then new, stronger existence results may be obtained by showing that if there is no solution to the conditions (16)-(17), then the augmented Lagrangian function ϕ_k must be unbounded below. The following proposition provides the necessary analysis for a general convex function $\phi(\cdot)$:

Proposition 6.3 *Let $\beta, \epsilon > 0$ be given positive scalars. Suppose $\phi : \mathbb{R}^n \to (-\infty, +\infty]$ is a closed proper convex function with the property that there is no pair of vectors $x, y \in \mathbb{R}^n$ such that*

$$y \in \partial\phi(x) \qquad \|y\| \leq \frac{\epsilon}{\max\{\beta, \|x\|\}}. \tag{43}$$

Then $\inf_{x\in\mathbb{R}^n}\{\phi(x)\} = -\infty$.

Proof. Consider the function $b : \mathbb{R} \to (-\infty, +\infty]$ given by

$$b(r) = \inf\{\phi(x) \mid \|x\| \leq r\}.$$

This function is convex, as may be seen either from first principles or by noting that $b(r) = V_0(-r)$, where $V_0(\cdot)$ is the parametric value function of the convex program (1) with $f = \phi$, $m = 1$, and $g_1(x) = \|x\|$.

Let $\beta' \geq \beta$ be a scalar such that there exists some point $\tilde{x} \in \operatorname{ri}\operatorname{dom}\phi$ with $\|\tilde{x}\| < \beta'$. Such a β' must exist, or ϕ could not be proper. Now consider, for $r \geq \beta'$, the convex program

$$\begin{array}{ll} \min & \phi(x) \\ \text{S.T.} & \|x\| \leq r, \end{array} \tag{44}$$

whose optimal value is $b(r)$. This problem is of the form (1), where $f = \phi$, $m = 1$, and $g_1(x) = \|x\| - r$. The essential objective function $F_0^{(r)}$ of this problem has the form

$$F_0^{(r)}(x) = \phi(x) + \delta_{\overline{B}(r)}(x),$$

where $\overline{B}(r)$ denotes the closed unit ball of radius r in $\mathbb{R}^n$. Since ϕ and $\delta_{\overline{B}(r)}$ are closed proper convex and $F_0^{(r)}(\tilde{x}) = \phi(\tilde{x}) < \infty$, $F_0^{(r)}$ is closed proper convex [19, Theorem 9.3]. Since its domain is bounded, it has no directions of recession, and it must attain a finite minimum [19, Theorem 27.3]. Let $x(r)$ denote some point attaining this minimum; then $x(r)$ is a primal solution of (44).

The point $\tilde{x}$ satisfies the Slater condition of [19, Corollary 28.2.1] for (44), so a Lagrange multiplier $\lambda(r) \geq 0$ must exist for the constraint $\|x\| \leq r$. There are two possibilities: either $\|x(r)\| < r$ or $\|x(r)\| = r$. Applying the standard complementary slackness conditions [19, Theorem 28.1], we must have in the former case that $\lambda(r) = 0$ and $0 \in \partial\phi(x(r))$. But then setting $x = x(r)$ and $y = 0$ would satisfy the conditions (43), so it must be that $\|x(r)\| = r$. In that case [19, Theorem 28.1],

$$\begin{array}{rl} & 0 \in \partial\phi(x(r)) + \lambda(r)\partial g_1(x(r)) = \partial\phi(x(r)) + \lambda(r)\dfrac{x(r)}{r} \\ \Leftrightarrow & -\lambda(r)\dfrac{x(r)}{r} \in \partial\phi(x(r)). \end{array}$$

Noting that $\|-\lambda(r)x(r)/r\| = \lambda(r)$, it must be that $\lambda(r) > \epsilon/r$, or setting $x = x(r)$ and $y = -\lambda(r)x(r)/r$ would satisfy (43).

Now, Proposition 2.6 implies that $\lambda(r) \in \partial V_0^{(r)}(0)$, where $V_0^{(r)}$ is the parametric value function of (44), that is, $V_0^{(r)} : \mathbb{R} \to (-\infty, +\infty]$ is given by

$$V_0^{(r)}(u) = \inf\left\{\phi(x) \;\middle|\; x \in \mathbb{R}^n, \|x\| + u \leq r\right\}.$$

It follows that $-\lambda(r) \in \partial b(r)$ for any $r \geq \beta'$. Therefore, for any $\rho \geq \beta'$,

$$\begin{array}{rcl} b(\rho) & \geq & b(\rho+1) + (-1)(-\lambda(\rho+1)) \\ & = & b(\rho+1) + \lambda(\rho+1) \\ & > & b(\rho+1) + \dfrac{\epsilon}{\rho+1}, \end{array}$$

and so

$$b(\rho + 1) < b(\rho) - \frac{\epsilon}{\rho + 1}.$$

Applying this relation inductively starting from $\rho = \beta'$, one obtains that for any integer $t \geq 1$ that

$$b(\beta' + t) < b(\beta') - \epsilon \sum_{s=1}^{t} \frac{1}{\beta' + s}.$$

Letting $t \to \infty$ and using that $b(\beta') < \infty$, it follows that

$$\lim_{t\to\infty} \phi(x(\beta' + t)) = \lim_{t\to\infty} b(\beta' + t) = b(\beta') - \infty = -\infty,$$

implying that $\inf_{x\in\mathbb{R}^n} \{\phi(x)\} = -\infty$. □

Combining this propostion with prior results for the algorithm (6)-(7) yields an existence result for the recursions (16)-(18):

Proposition 6.4 *Suppose that problem (1) conforms to Assumption 2.1 and possesses both primal feasible and dual feasible solutions. At iteration k of Algorithm 3.1, suppose $\epsilon_k > 0$. Then $x^k, y^k \in \mathbb{R}^n$ satisfying (16)-(17) must exist.*

Proof. From [24, Proposition 2.2.5],

$$\inf_{x\in\mathbb{R}^n} \{\phi_k(x)\} = \sup_{p\geq 0} \left\{Q_0(p) - D_h(p, p^{k-1})\right\} \tag{45}$$

(the proof is similar to that of [25, Lemma A.2], which treats the case that $f, g_1, \ldots, g_m$ are finite and differentiable).

Under Assumption 2.1, it follows from [24, Lemmas 2.2.3 and 2.2.4] that the function $P_k : x \mapsto h^{\oplus}\big(\nabla h(p^{k-1}) + c_k g(x)\big)$ is closed proper convex. Now, since $\phi_k(x) = f(x) + P_k(x)$ and both f and P_k are closed proper convex, [19, Theorem 9.3] implies that either ϕ_k is closed proper convex or is everywhere $+\infty$.

To rule out the latter possibility, consider the value of $\phi_k(\overline{x})$, where $\overline{x}$ is a feasible solution to (1). Since $g(\overline{x}) \leq 0$, we have $\nabla h(p^{k-1}) + c_k g(\overline{x}) \leq \nabla h(p^{k-1})$. Because $h^{\oplus}$ is nondecreasing, it then follows that

$$P_k(\overline{x}) = h^{\oplus}\big(\nabla h(p^{k-1}) + c_k g(\overline{x})\big) \leq h^{\oplus}\big(\nabla h(p^{k-1})\big) < \infty, \tag{46}$$

where the last inequality is a consequence of

$$\begin{aligned}
\operatorname{dom} h^{\oplus} &= \operatorname{rge} \partial\Big(h + \delta_{\mathbb{R}^m_+}\Big) \\
&\supseteq \partial\Big(h + \delta_{\mathbb{R}^m_+}\Big)\,(p^{k+1}) \\
&= \nabla h(p^{k-1}) + \partial\delta_{\mathbb{R}^m_+}(p^{k+1}) \quad \text{[using [19, Theorem 23.8], as in Lemma 3.3]} \\
&\ni \nabla h(p^{k-1}).
\end{aligned}$$

We must also have $f(\overline{x}) < \infty$, which with (46) implies $\phi_k(\overline{x}) = f(\overline{x}) + P_k(\overline{x}) < \infty$, and ϕ_k cannot be everywhere $+\infty$. So, ϕ_k is closed proper convex.

Now suppose there is no solution to (16)-(17). One may then apply Proposition 6.3 to conclude that $\inf_{x\in\mathbb{R}^n}\{\phi_k(x)\} = -\infty$. Note also that $D_h(p, p^{k-1}) < \infty$ for all $p \geq 0$ by Assumption 2.7(ii) and (14). So it would then follow from (45) that $\sup_{p\geq 0}\{Q_0(p)\} = -\infty$, implying that (1) has no dual feasible solutions, contradicting the hypothesis. Thus, a solution to (16)-(17) must exist. □

Corollary 6.5 *Suppose that problem (1) conforms to Assumptions 2.1 and 2.4, and* $\epsilon_k > 0$ *for all* k*. Then there exists an infinite sequence* $\{(x^k, y^k, p^k)\}_{k=1}^{\infty}$ *conforming to the recursions of Algorithm 3.1.*

Proof. Assumption 2.4 implies that (1) has both primal and dual feasible solutions. The result follows by using Proposition 6.4 inductively. □

To illustrate how these existence results can apply when those for the exact form of the algorithm do not, consider the trivial convex program

$$\begin{array}{ll} \min & 0x_1 \\ \text{S.T.} & x_1 \geq 0, \end{array}$$

that is, (1) with $n = 1$, $m = 1$, $f(x) \equiv 0$, and $g_1(x) = -x$, and suppose one applies the exponential method of multipliers (4)-(5). Then $p^k > 0$ for all k, and the augmented Lagrangian function at iteration k is

$$\phi_k(x) = \frac{1}{c_k} p^{k-1} e^{-c_k x},$$

which is bounded below by 0, but has no exact minimizer. Therefore, the recursions for the exact form of the algorithm cannot be satisfied.

However, conditions (16)-(17) may be satisfied by choosing $x^k > 0$ large enough that

$$\begin{aligned} & \|\nabla\phi_k(x)\| = p^{k-1}e^{-c_k x} \leq \frac{\epsilon_k}{\max\{\beta, x^k\}} \\ \Leftrightarrow \qquad & c_k x^k - \log\max\{\beta, x^k\} \geq \log\left(\frac{p^{k-1}}{\epsilon_k}\right). \end{aligned}$$

One then has

$$p^k \;=\; p^{k-1}e^{-c_k x^k} \;\leq\; \frac{\epsilon_k}{\max\{\beta, x^k\}} \;\leq\; \frac{\epsilon_k}{\beta}.$$

Thus, $\{p^k\} \subset \mathbb{R}_{++}$ converges to 0, an optimal dual solution, at a rate at least proportional to that of $\{\epsilon_k\}$, although $\{x^k\}$ may be divergent.

References

[1] A. Auslender, M. Teboulle, and S. Ben-Tiba. Interior proximal and multiplier methods based on second order homogeneous kernels. *Mathematics of Operations Research* 24 (1999) 645-668.

[2] A. Ben-Tal and M. Zibulevsky. Penalty/barrier multiplier methods for convex programming problems. *SIAM Journal on Optimization* 7 (1997) 347-366.

[3] D. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods* (Academic Press, New York, 1982).

[4] D. Bertsekas. *Nonlinear Programming* (Athena Scientific, Belmont, MA, 1995).

[5] 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. *USSR Computational Mathematics and Mathematical Physics* 7 (1967) 200-217.

[6] R. S. Burachik, A. N. Iusem, and B. F. Svaiter. Enlargement of monotone operators with applications to variational inequalities. *Set-Valued Analysis* 5 (1997) 159-180

[7] Y. Censor and A. Lent. An iterative row-action method for interval convex programming. *Journal of Optimization Theory and Applications* 34 (1981) 321-353.

[8] Y. Censor and S. A. Zenios. Proximal minimization algorithm with D-functions. *Journal of Optimization Theory and Applications* 73 (1992) 451-464.

[9] G. Chen and M. Teboulle. A convergence analysis of proximal-like minimization algorithms using Bregman functions. *SIAM Journal on Optimization* 3 (1993) 538-543.

[10] J. Eckstein. Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. *Mathematics of Operations Research* 18 (1993) 202-226.

[11] J. Eckstein. Approximate iterations in Bregman-function-based proximal algorithms. *Mathematical Programming* 83 (1998) 113-124.

[12] J. Eckstein and M. C. Ferris. Smooth methods of multipliers for monotone complementarity problems. *Mathematical Programming* 86 (1999) 65-90.

[13] W. W. Hager. Analysis and implementation of a dual algorithm for constrained optimization. *Journal of Optimization Theory and Applications* 79 (1993) 427-462.

[14] K. C. Kiwiel. On the twice differentiable augmented Lagrangian. *Journal of Optimization Theory and Applications* 88 (1996) 233-236.

[15] K. C. Kiwiel. Proximal minimization methods with generalized Bregman functions. *SIAM Journal on Control and Optimization* 35 (1997) 1142-1168.

[16] T. Pennanen. *Dualization of monotone generalized equations.* Ph.D. Thesis, Department of Mathematics, University of Washington, 1999.

[17] T. Pennanen. Dualization of generalized equations of maximal monotone type. *SIAM Journal on Optimization* 10 (2000) 809-835.

[18] B. Polyak. *Introduction to Optimization* (Optimization Software Inc., New York, 1987).

[19] R. T. Rockafellar. *Convex Analysis* (Princeton University Press, Princeton, NJ, 1970).

[20] R. T. Rockafellar. *Conjugate Duality and Optimization* (SIAM, Philadelphia, 1974).

[21] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization* 14 (1976) 877-898.

[22] R. T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research* 1 (1976) 97-116.

[23] A. Ruszczyński. On convergence of an augmented Lagrangian decomposition method for sparse convex optimization. *Mathematics of Operations Research* 20 (1995) 634-656.

[24] P. J. da Silva e Silva. *Tópicos em Métodos de Ponto Proximal.* Ph.D. thesis, Instituto de Matemática e Estatística, University of São Paulo, 2000.

[25] P. J. da Silva e Silva, J. Eckstein, and C. Humes. Rescaling and stepsize selection in proximal methods using separable generalized distances. *SIAM Journal on Optimization* 12 (2001) 238-261.

[26] M. V. Solodov and B. F. Svaiter. An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions. *Mathematics of Operations Research* 25 (2000) 214-230.

[27] M. Teboulle. Entropic proximal mappings with applications to nonlinear programming. *Mathematics of Operations Research* 17 (1992) 670-690.

[28] M. Teboulle. Convergence of proximal-like algorithms. *SIAM Journal on Optimization* 7 (1997) 1069-1083.

[29] P. Tossings. The perturbed proximal point algorithm and some of its applications. *Applied Mathematics and Optimization* 29 (1994) 125-159.

[30] P. Tseng and D. P. Bertsekas. On the convergence of the exponential multiplier method for convex programming. *Mathematical Programming* 60 (1993) 1-19.