Mathematical Programming 59 (1993) 231-247
North-Holland

Dual coordinate ascent methods for non-strictly convex minimization

Paul Tseng
Department of Mathematics, University of Washington, Seattle, USA

Received 16 January 1989
Revised manuscript received 13 June 1991

We consider a dual method for solving non-strictly convex programs possessing a certain separable structure. This method may be viewed as a dual version of a block coordinate ascent method studied by Auslender [1, Section 6]. We show that the decomposition methods of Han [6, 7] and the method of multipliers may be viewed as special cases of this method. We also prove a convergence result for this method which can be applied to sharpen the available convergence results for Han's methods.

Key words: Separable convex programming, dual block coordinate ascent, decomposition, method of multipliers.

1. Introduction

Recently, Han [6] proposed a decomposition method for the problem of minimizing a strictly convex quadratic function over the intersection of a collection of closed convex sets. This method has the advantage that it reduces the original problem to a sequence of problems each of which involves the projection onto only one of the convex sets. Subsequently, this method was extended to problems in which indicator functions for the convex sets are replaced by arbitrary proper closed convex functions [7]. Interestingly, although Han's methods have features reminiscent of augmented Lagrangian methods, they do not obviously fall into any of the existing algorithmic frameworks. Has a new algorithmic framework being found? In this paper, we answer this and related questions raised by our investigation.

Our first main result is the demonstration that Han's methods are in fact special cases of a (block) *coordinate ascent* method for maximizing a dual functional q of

Correspondence to: Prof. P. Tseng, Department of Mathematics, GN-50, University of Washington, Seattle, WA 98195, USA.

The main part of this research was conducted while the author was with the Laboratory for Information and Decision Systems, M.I.T., Cambridge, with support by the U.S. Army Research Office, Contract No. DAAL03-86-K-0171 (Center for Intelligent Control Systems) and by the National Science Foundation under Grant ECS-8519058.

the form

$$q(p_1, \dots, p_K) = g(p_1, \dots, p_K) - \sum_{k=1}^{K} g_k(p_k), \tag{1.1}$$

where g is a concave *differentiable* function and each g_k is a proper closed convex function defined on some Euclidean space (see Section 4). Dual coordinate ascent methods have been studied extensively (see [2–4, 9, 13–15] and references therein), but in all cases the dual functional is assumed to be *smooth*, which is not the case in the problems studied by Han (or, in general, when a problem is transformed in a way to bring about structure that is favourable for decomposition). The only exception, to our knowledge, is a work of Auslender [1, p. 94] which studies a coordinate ascent method for maximizing concave functions of precisely the form (1.1). However, Auslender's study considers only the primal settting and makes the additional assumption that g is *strictly* concave and q has bounded level sets, neither of which holds in our dual setting. In addition, we show that the method of multipliers is also a special case of this dual coordinate ascent method (see Section 5) and suggest some new applications of this method. Thus, coordinate ascent provides a unifying framework for a number of seemingly unrelated methods.

Our second main result is a convergence proof for the above dual (block) coordinate ascent method. We show that, under mild constraint qualification conditions on the problem, this method generates a bounded sequence every cluster point of which is an optimal solution of the primal problem (see Theorem 3.1). Under additional assmptions on the problem, a similar conclusion is shown for the dual. Upon applying this result to Han's method, we are able to sharpen the convergence results presented in [6, 7].

This paper is organized as follows: in Section 2 we describe the problem. In Section 3 we describe the dual method and in Section 6 we prove a convergence result for this method. In Sections 4 and 5 we show that the methods of Han and the method of multipliers may be viewed as special cases of this method.

In our notation, every vector is a column vector in some Euclidean space and superscript T denotes transpose. We denote by $\langle \cdot, \cdot \rangle$ the usual Euclidean inner product and let $\|\cdot\|$ denote the L_2-norm induced by it. For any set $S \subseteq \mathbb{R}^m$, we denote by $\mathrm{ri}(S)$ and $\mathrm{cl}(S)$, respectively, the relative interior and the closure of S. The *indicator* function of S is the function that is 0 everywhere on S and ∞ everywhere else. For any proper convex function h in $\mathbb{R}^m$, we denote by $\mathrm{dom}\, h$ the *effective domain* of h, i.e., $\mathrm{dom}\, h = \{x \mid h(x) < \infty\}$, and by h^* the *conjugate* function of h, i.e.,

$$h^*(y) = \sup_x \langle y, x \rangle - h(x).$$

When h^* is real-valued, h is said to be *co-finite.* For any $x \in \mathrm{dom}\, h$, we denote by $\partial h(x)$ the *subdifferential* of h at x (possibly empty) and, for any $d \in \mathbb{R}^m$, by $h'(x; d)$ the *directional derivative* of h at x in the direction d. A relation between the subdifferential and the directional derivative which we will use frequently is the following (see [11, Theorem 23.2])

$$h'(x; d) \geq \langle d, \eta \rangle \quad \forall \eta \in \partial h(x). \tag{1.2}$$

2. Problem description

Consider a convex program with the following separable structure

$$\text{minimize} \quad f_0(x_0)+f_1(x_1)+f_2(x_2)+\cdots+f_K(x_K) \tag{2.1a}$$

$$\begin{array}{llllll} \text{subject to} & A_1x_0+B_1x_1 & & & & =b_1, \\ & A_2x_0 & +B_2x_2 & & & =b_2, \\ & \vdots & & \ddots & & \vdots \\ & A_Kx_0 & & & +B_Kx_K & =b_K, \end{array} \tag{2.1b}$$

where, for each $k \in \{0, 1, \ldots, K\}$ $(K \geq 2)$, f_k is a proper closed convex function in $\mathbb{R}^{m_k}$ $(m_k \geq 1)$, A_k is an $n_k \times m_0$ matrix $(n_k \geq 1)$, B_k is an $n_k \times m_k$ matrix, and b_k is an n_k-vector. We can think of x_0 as a set of "coupling" variables, without which the problem decomposes into K independent subproblems. We remark that we can also allow inequality constraints in (2.1), but for simplicity we will consider only the equality constrained case.

We make the following standing assumptions on the f_k's and the B_k's:

Assumption A.

(a) For each $k \in \{0, 1, \ldots, K\}$, f_k is continuous on $\operatorname{dom} f_k$.

(b) f_0 is co-finite and, for each $k \in \{1, \ldots, K\}$, either f_k is a co-finite or B_k has rank m_k.

(c) f_0 is strictly convex.

Part (a) of Assumption A ensures that the f_k's are well behaved, and part (b) ensures that a certain dual functional associated with (2.1) (to be defined below) is proper closed. Part (c) ensures that the problem has some curvature (which translates to a certain differentiability condition on the dual functional).

The problem (2.1) contains a number of interesting special cases. For example, the problem

$$\begin{array}{ll} \text{minimize} & h(x)+h_1(x)+\cdots+h_K(x) \\ \text{subject to} & x \in C_1 \cap \cdots \cap C_K, \end{array}$$

where h is a strongly convex continuous function and the h_k's are convex continuous functions and the C_k's are closed convex sets, can be casted in the form (2.1) with $f_0 = h$ and, for $k = 1, \ldots, K$, $f_k = h_k +$ (indicator function of C_k) and $A_k = -B_k = I$, $b_k = 0$. For another example, if $A_k = I$ for all k in (2.1), we would model a situation in which it is required that the 'slacks' $b_k - B_kx_k$ be the same for all k. (See Sections 4 and 5 for additional examples.)

By assigning a Lagrange multiplier vector $p_k \in \mathbb{R}^{n_k}$ to the constraints $A_k x_0 + B_k x_k = b_k$, we obtain the following dual functional q in $\mathbb{R}^{n_1+\cdots+n_K}$ given by

$$q(p_1,\ldots,p_K) = \min_{x_0,\ldots,x_K} \sum_{k=0}^{K} f_k(x_k) + \sum_{k=1}^{K} \langle p_k, b_k - A_k x_0 - B_k x_k \rangle \tag{2.2}$$

$$= \sum_{k=1}^{K} \langle p_k, b_k \rangle - f_0^*\left(\sum_{k=1}^{K} A_k^{\mathrm{T}} p_k \right) - \sum_{k=1}^{K} f_k^*(B_k^{\mathrm{T}} p_k), \tag{2.3}$$

where f_k^* denotes the conjugate function of f_k. The fual problem associated with (2.1) is

$$\max_{p_1,\ldots,p_K} q(p_1,\ldots,p_K). \tag{2.4}$$

Since f_0 is strictly convex and co-finite (cf. Assumption A(b), (c)), then f_0^* is convex differentiable (see [11, Theorem 26.3]). In addition, for each $k \in \{1,\ldots,K\}$, either f_k^* is real-valued or B_k^{T} has full row rank (cf. Assumption A(b)), so the function $p_k \mapsto f_k^*(B_k^{\mathrm{T}} p_k)$ is proper closed convex. The, by (2.3), $-q$ is the sum of a convex differential function and a proper closed convex separable function. Whence $-q$ is proper closed convex.

For notational convenience, let X denote the constraint set of (2.1), i.e.,

$$X = \{(x_0,\ldots,x_K) \mid A_k x_0 + B_k x_k = b_k,\ k = 1,\ldots,K\}. \tag{2.5}$$

To ensure that (2.1) has an optimal solution and that strong duality holds between (2.1) and (2.4) (that is, the optimal value of (2.1) equals the optimal value of (2.4)), we make the following standing assumption on (2.1):

Assumption B. There is a $\bar{k} \in \{0, 1, \ldots, K\}$ such that $f_{\bar{k}+1},\ldots,f_K$ are polyhedral and $\mathrm{ri}(\mathrm{dom}\, f_0) \times \cdots \times \mathrm{ri}(\mathrm{dom}\, f_{\bar{k}}) \times \mathrm{dom}\, f_{\bar{k}+1} \times \cdots \times \mathrm{dom}\, f_K$ makes a nonempty intersection with X.

Assumption B is the usual feasibility assumption for (2.1), i.e., $\mathrm{dom}\, f \cap X \neq \emptyset$, where $f : \mathbb{R}^{m_0+\cdots+m_K} \mapsto (-\infty, \infty]$ denotes the objective function of (2.1), i.e.,

$$f(x_0,\ldots,x_K) = \sum_{k=0}^{K} f_k(x_k), \tag{2.6}$$

plus a mild constraint qualification condition. The constraint qualification condition, which is required to ensure that the proposed method is well defined and has the desired convergence properties (see Theorem 3.1), is satisfied in most applications. For example, it is satisfied when each f_k is polyhedral or when f_0 is real-valued and each A_k has full row rank (an example of which is given in Section 5).

We have the following duality result relating (2.1) and (2.4):

Lemma 2.1. *The optimal solution set for* (2.1) *is nonempty and compact. Moreover, strong duality holds for* (2.1) *and* (2.4).

Proof. Since f_0 is co-finite and, for each k, either f_k is co-finite or B_k has full column rank (cf. Assumption A(b), (c)), then f has bounded level sets on X. Since both f and X are closed, these level sets are in fact compact, which together with the hypothesis that (2.1) is feasible (cf. Assumption B) implies that the set of optimal solutions for (2.1) is nonempty and compact. (In fact, because f_0 is strictly convex, the first m_0 coordinates of the optimal solutions are *unique.*)

To see that strong duality holds for (2.1) and (2.4), notice that the convex set

$$\left\{(x_0,\ldots,x_K,u_1,\ldots,u_K,\zeta)\,\middle|\, A_kx_0+B_kx_k=u_k,\ k=1,\ldots,K,\ \sum_{k=0}^{K} f_k(x_k)\le\zeta\right\}$$

is closed. Hence the *convex bifunction* associated with (2.1) [11, p. 293] is closed. Since the optimal solution set for (2.1) is bounded, Theorem 30.4(i) in [11] states that the optimal values of, respectively, (2.1) and (2.4) are equal. □

(In fact, by using the constraint qualification conditions implicit in Assumption B, it can be argued that (2.4) has a maximum point.)

3. Algorithm description

Since strong duality holds for (2.1) and (2.4), we can consider solving (2.1) by ways of maximizing the dual functional q. From (2.3), we see that q has the form (1.1) with

$$g(p_1,\ldots p_K)=\sum_{k=1}^{K}\langle p_k,b_k\rangle-f_0^*\left(\sum_{k=1}^{K}A_k^{\mathrm{T}}p_k\right),$$

$$g(p_k)=f_k^*(B_k^{\mathrm{T}}p_k),\quad k=1,\ldots,K.$$

This suggests a block coordinate ascent method for maximizing q whereby, given an iterate $p\in\operatorname{dom} q$, one of the coordinate blocks, say p_s, is chosen and q is maximized with respect to p_s, while the other coordinate blocks are held fixed. This method, which we henceforth call the *dual block coordinate ascent* (DBCA for short) algorithm, operates formally as follows.

DBCA algorithm. We choose arbitrarily the initial iterate $(p_1^0,\ldots,p_K^0)\in\operatorname{dom} q$. At the rth iteration ($r\ge1$), we have on hand a current iterate $(p_1^{r-1},\ldots,p_K^{r-1})\in\operatorname{dom} q$, and we compute a new iterate $(p_1^r,\ldots,p_K^r)\in\operatorname{dom} q$ by choosing an index $s^r\in\{1,\ldots,K\}$ and setting

$$p_k^r\in\arg\max_{p_k} q(p_1^{r-1},\ldots,p_{k-1}^{r-1},p_k,p_{k+1}^{r-1},\ldots,p_K^{r-1})\quad\text{for } k=s^r,\tag{3.1}$$

$$p_k^r=p_k^{r-1}\quad\text{for } k\ne s^r.\tag{3.2}$$

The lemma below shows that the DBCA algorithm is indeed well defined.

Lemma 3.1. *The maximization in* (3.1) *is attained for every r.*

Proof. It can be seen from (2.2) that, for any r and any $k \in \{1, \ldots, K\}$, the function $p_k \mapsto q(p_1^{r-1}, \ldots, p_{k-1}^{r-1}, p_k, p_{k+1}^{r-1}, \ldots, p_K^{r-1})$ attains its maximum at some point p_k if and only if p_k is a Kuhn-Tucker vector of the problem

$$\begin{aligned} &\text{minimize} \quad f_0(x_0) + f_k(x_k) - \sum_{j \neq k} \langle p_j^{r-1}, A_j x_0 \rangle \\ &\text{subject to} \quad A_k x_0 + B_k x_k = b_k. \end{aligned} \tag{3.3}$$

Since (3.3) has an optimal solution (by an argument analogous to that in the proof of Lemma 2.1) and, by Assumption B, $\mathrm{ri}(\mathrm{dom}\, f_0) \times \mathrm{ri}(\mathrm{dom}\, f_k)$ (or, if f_k is polyhedral, the second "ri" can be dropped) makes a nonempty intersection with $\{(x_0, x_k) \mid A_k x_0 + B_k x_k = b_k\}$, Theorem 28.2 in [11] shows that (3.3) indeed has a Kuhn-Tucker vector. □

The proof of Lemma 3.1 shows that, at each iteration r, the maximization (3.1) can in fact be achieved by solving the subproblem (3.3) (with k set to s^r) which is much simpler than the original problem (2.1). Let us denote by $(x_0^r, x_{s^r}^r)$ the optimal solution to this subproblem and, for each $k \in \{1, \ldots, K\}$ not equal to s^r, let $x_k^r = x_k^{r-1}$. (The choice of $x_1^0, \ldots, x_K^0$ is immaterial.) It can be seen that $(x_0^r, x_1^r, \ldots, x_K^r)$, $r = 1, 2, \ldots$, are equivalently given by the formula

$$x_0^r = \nabla f_0^* \left(\sum_{k=1}^K A_k^{\mathrm{T}} p_k^r \right), \tag{3.4}$$

and, for $k \in \{1, \ldots, K\}$,

$$x_k^r \in \begin{cases} \{x_k \mid B_k x_k = b_k - A_k x_0^r,\ B_k^{\mathrm{T}} p_k^r \in \partial f_k(x_k)\}, & \text{if } s^r = k, \\ \{x_k^{r-1}\}, & \text{if } s^r \neq k. \end{cases} \tag{3.5}$$

We are interested in the convergence properties of both the dual sequence $\{(p_1^r, \ldots, p_K^r)\}$ and the associated primal sequence $\{(x_0^r, x_1^r, \ldots, x_K^r)\}$. To ensure convergence, some restriction on the order in which the coordinate blocks are iterated upon must be imposed. We will consider the following well known restriction which states that each coordinate block is iterated upon at least once every T iterations, for some fixed T.

Essentially cyclic order. There exists an integer $T \geq K$ such that $\{1, \ldots, K\} \subseteq \{s^{r+1}, \ldots, s^{r+T}\}$, for all r.

The above order of iteration is a direct generalization of the classical cyclic order of iteration (corresponding to $s^r = (r \bmod K) + 1$). On some problems, its employment can improve the performance of the algorithm over that obtained by using the cyclic order.

Below we give the main result of this section, showing that the DBCA algorithm using the essentially cyclic order of iteration converges in the sense that the associated primal sequence given by (3.4) and (3.5) is bounded and every one of its cluster points solves (2.1). We remark that, because q is not strictly concave in any sense nor differentiable and its level sets are not necessarily bounded, the conventional approaches to proving the convergence of coordinate ascent methods break down and new proof techniques must be developed.

Theorem 3.1. *Let* $\{p^r = (p_1^r, \ldots, p_K^r)\}$ *be a sequence generated by the DBCA algorithm* (3.1)–(3.2) *using the essentially cyclic order of iteration, and let* $\{(x_0^r, x_1^r, \ldots, x_K^r)\}$ *be given by* (3.4)–(3.5). *The following hold*:

(a) $\{(x_0^r, x_1^r, \ldots, x_K^r)\}$ *is bounded and every one of its cluster points is an optimal solution of* (2.1).

(b) *If the set of maximum points for* q *is bounded, then* $\{p^r\}$ *is bounded and every one of its cluster points is a maximum point of* q.

The proof of Theorem 3.1 is rather long and is given in Section 6. We remark that, in the DBCA algorithm, we can maximize q with respect to more than one coordinate blocks at each iteration. Also, if q is differentiable in p_s, for some s, then inexact maximization with respect to some of the coordinates of p_s, as discussed in [13–15], is permisslble.

4. Relation with Han's methods

In this section we show that the decomposition methods of Han given in [6] and [7] are special cases of the DBCA algorithm. This result is interesting in that the connection between Han's methods and dual coordinate ascent is far from obvious. We also sharpen, by applying Theorem 3.1, the convergence results obtained by Han.

Consider the following convex program treated in [7]:

$$\min_x \; h(x) + h_1(x) + \cdots + h_K(x), \tag{4.1}$$

where $h_1, \ldots, h_K$ are proper closed convex functions in $\mathbb{R}^m$ and h is a strictly convex quadratic function in $\mathbb{R}^m$ given by

$$h(x) = \tfrac{1}{2}\langle x - d, Q(x - d)\rangle,$$

with Q an $m \times m$ symmetric positive definite matrix and d an m-vector. For expositional simplicity, we will assume that Q is the identity matrix in what follows, though our analysis readily extends to the case of general Q.

We make the following standing assumptions about (4.1) (cf. [7]):

Assumption C.

(a) For each $k \in \{1, \ldots, K\}$, h_k is continuous on dom h_k.

(b) There exists a $k\in\{0,1,\ldots,K\}$ such that $h_{k+1},\ldots,h_K$ are polyhedral and $\mathrm{ri}(\mathrm{dom}\, h_1)\cap\cdots\cap\mathrm{ri}(\mathrm{dom}\, h_k)\cap\mathrm{dom}\, h_{k+1}\cap\cdots\cap\mathrm{dom}\, h_K$ is nonempty.

The methods of Han, proposed in [6] for solving the case of (4.1) where the h_k's are indicator functions of closed convex sets and subsequently extended in [7] to arbitrary convex functions, operate as follows. We choose $y_1^0,\ldots,y_K^0$ to be any vectors in, respectively, $\mathrm{dom}\, h_1^*,\ldots,\mathrm{dom}\, h_K^*$ and set

$$x_K^0=d-y_1^0-\cdots-y_K^0. \tag{4.2}$$

(Recall h_k^* is the conjugate function of h_k.) Then, we generate two sequences $\{(x_0^t,\ldots,x_K^t)\}_{t\geq 1}$ and $\{(y_1^t,\ldots,y_K^t)\}_{t\geq 1}$ according to the formulas: $x_0^t=x_K^{t-1}$ and

$$x_K^t=\arg\min_x\|x_{k-1}^t+y_k^{t-1}-x\|^2+h_k(x), \tag{4.3}$$

$$y_k^t=y_k^{t-1}+x_{k-1}^t-x_k^t, \tag{4.4}$$

for $k=1,\ldots,K$. Thus, Han's methods reduce the original problem to a sequence of problems each of which involves minimization of only one of the h_k's. This is a major advantage of these methods.

To relate Han's methods to the DBCA algorithm, we introduce artificial variables and write (4.1) in the following equivalent form:

$$\begin{aligned}&\text{minimize}\quad h(x_0)+\sum_{k=1}^K h_k(x_k)\\ &\text{subject to}\quad x_0-x_k=0,\quad k=1,\ldots,K.\end{aligned} \tag{4.5}$$

Clearly (4.5) is of the form (2.1) with $f_0=h$ and, for $k=1,\ldots,K$, $f_k=h_k$ and $A_k=-B_k=I$ and $b_k=0$. Moreover, by Assumption C, both Assumptions A and B hold for this problem. (The latter holds because each A_k is the identity matrix (so it has full row rank) and h is real-valued.) Thus, we conclude from Lemma 2.1 that (4.5) has an optimal solution. This implies (4.1) has an optimal solution which, by the strong convexity of h, is unique.

By assigning a Lagrange multiplier vector y_k to the constraints $x_0-x_k=0$ for each k, we obtain the dual functional associated with (4.5) (cf. (2.3)) as

$$\begin{aligned}q(y_1,\ldots,y_K)&=\min_{x_0,\ldots,x_K} h(x_0)+\sum_{k=1}^K h_k(x_k)+\sum_{k=1}^K\langle y_k,x_0-x_k\rangle\\ &=\tfrac12\|d\|^2-h\Big(\sum_{k=1}^K y_j\Big)-\sum_{k=1}^K h_k^*(y_k).\end{aligned} \tag{4.6}$$

Below is the main result of this section.

Proposition 5.1. *Let $\{(y_1^t,\ldots,y_K^t)\}$ and $\{(x_0^t,\ldots,x_K^t)\}$ be the sequences generated by (4.2)-(4.4). Then,*

$$y_k^t=\arg\max_{y_k} q(y_1^t,\ldots,y_{k-1}^t,y_k,y_{k+1}^{t-1},\ldots,y_K^{t-1}), \tag{4.7}$$

$$x_k^t=\nabla h^*(y_1^t+\cdots+y_k^t+y_{k+1}^{t-1}+\cdots+y_K^{t-1}), \tag{4.8}$$

for $k=1,\ldots,K$ and all positive integers t.

Proof. It is easily seen from (4.2), (4.4) and $x_0^t = x_K^{t-1}$ that

$$y_1^t + \cdots + y_{k-1}^t + y_k^{t-1} + \cdots + y_K^{t-1} = d - x_{k-1}^t, \quad k = 1, \ldots, K, \ \forall t \geq 1. \tag{4.9}$$

Fix any $k \in \{1, \ldots, K\}$ and any positive integer t. From (4.3) and (4.4) we have

$$\begin{aligned} -y_k^t + x_{k-1}^t + y_k^{t-1} &= \arg\min_x \tfrac{1}{2}\|x_{k-1}^t + y_k^{t-1} - x\|^2 + h_k(x) \\ &= \arg\min_x -\langle y_k^t, x\rangle + \tfrac{1}{2}\| -y_k^t + x_{k-1}^t + y_k^{t-1} - x\|^2 + h_k(x) \\ &= \arg\min_x -\langle y_k^t, x\rangle + h_k(x), \end{aligned}$$

where the last equality follows from the simple observation that, for any proper closed convex function g in $\mathbb{R}^m$, if $\bar{z} = \arg\min_z g(z) + \frac{1}{2}\|\bar{z} - z\|^2$, then $\bar{z} = \arg\min_z g(z)$. Hence, by Theorem 23.5 in [6],

$$-y_k^t + x_{k-1}^t + y_k^{t-1} \in \partial h_k^*(y_k^t).$$

This in turn implies

$$\begin{aligned} y_k^t &= \arg\min_{y_k} \tfrac{1}{2}\|y_k - x_{k-1}^t - y_k^{t-1}\|^2 + h_k^*(y_k) \\ &= \arg\min_{y_k} h(d - x_{k-1}^t - y_k^{t-1} + y_k) + h_k^*(y_k) \\ &= \arg\min_{y_k} h(y_1^t + \cdots + y_{k-1}^t + y_k + y_{k+1}^{t-1} + \cdots + y_K^{t-1}) + h_k^*(y_k), \end{aligned} \tag{4.10}$$

where the second equality follows from the definition of h (with $Q = I$) and the third equality follows from (4.9). Equation (4.10) together with (4.6) then proves (4.7). Equation (4.8) follows from (4.9) and the definition of h. $\square$

Upon comparing (4.7) with (3.1)-(3.2), we see that $(y_1^t, \ldots, y_K^t)$ is precisely the tKth term in the sequence generated by the DBCA algorithm (using cyclic order of iteration) applied to (4.5). Moreover, upon comparing (4.8) with (3.4), we see that x_k^t is the $((t-1)K + k)$th term in the associated primal sequence given by (3.4). Since Assumption A and B hold for (4.5) and the above method is an instance of the DBCA algorithm, we can apply our general convergence result for the latter (Theorem 3.1) to conclude convergence for the former.

Corollary 5.1. *Let* $\{(y_1^t, \ldots, y_K^t)\}$ *and* $\{(x_0^t, \ldots, x_K^t)\}$ *be the sequences generated by Han's algorithm* (4.2)-(4.4). *Then* $\{x_k^t\}$ *converges to the unique optimal solution of* (4.1) *for every* k *and* $\{\sum_{k=1}^K y_k^t\}$ *converges. If in addition* q *has bounded level sets, then* $\{(y_1^t, \ldots, y_K^t)\}$ *is bounded and every one of its cluster points maximizes the function* q.

(That $\{\sum_{k=1}^{K} y_k^t\}$ converges follows from the convergence of $\{x_1^t\}$ and (4.9).)

Corollary 5.1 sharpens the convergence results given in [6] (see Theorems 4.8 and 4.10 therein) and in [7] (see Theorems 4.6 and 4.7 therein) which require additional nonempty interior assumptions on the problem to assert convergence. Moreover, we see from Theorem 3.1 that neither the cyclic order of iteration nor the quadratic nature of h is required for convergence. (On the other hand, [7] does not require each h_k to be continuous on dom h_k.)

The connection between Han's methods and the DBCA algorithm readily suggests new methods for solving (4.1). For example, we can replace each h_k in (4.5) by $h_k + h$ and apply the DBCA algorithm to solve this modified problem. The resulting algorithm preserves the main features of Han's methods, but its subproblems are "closer" to the original so it might converge faster.

5. Relation with the method of multipliers

Consider the following convex program with linear constraints

$$\begin{aligned} &\text{minimize} && h(x) \\ &\text{subject to} && Bx = b, \end{aligned} \tag{5.1}$$

where h is a proper closed convex function in $\mathbb{R}^m$, B is an $n \times m$ matrix, and b is an n-vector. We assume that h is co-finite and continuous on dom h and that ri(dom h) $\cap \{x \mid Bx = b\}$ is nonempty.

By introducing the auxiliary vector x_0, we can write (5.1) equivalently as

$$\begin{aligned} &\text{minimize} && \tfrac{1}{2}c\|x_0\|^2 + h(x_1) \\ &\text{subject to} && x_0 = 0, \ \ x_0 + Bx_1 = b, \end{aligned} \tag{5.2}$$

where c is any fixed positive scalar. Clearly this problem is in the form of (2.1) with $f_0(x_0) = \frac{1}{2}c\|x_0\|^2$ and $f_1 = h$. Furthermore, it is easily seen that Assumptions A and B hold for (5.2). By assigning Lagrange multiplier vectors p_1 and p_2 to, respectively, the constraints $x_0 + Bx_1 = b$ and $x_0 = 0$, we obtain the dual functional q in $\mathbb{R}^{2n}$ associated with (5.2) (cf. (2.3)) as

$$q(p_1, p_2) = \langle b, p_1 \rangle - h^*(B^{\mathrm{T}} p_1) - \frac{1}{2c}\|p_1 - p_2\|^2. \tag{5.3}$$

Applying the DBCA algorithm to this special case of (2.1) yields an algorithm that alternately maximizes q with respect to p_1 and with respect to p_2:

$$p_1^r = \arg\max_{p_1} q(p_1, p_2^{r-1}), \tag{5.4}$$

$$p_2^r = \arg\max_{p_2} q(p_1^r, p_2), \tag{5.5}$$

for $r=0, 1, \ldots$. Using the special structure of q (cf. (5.3)), we see from (5.5) that $p_2^r = p_1^r$, so p_2^r is a Kuhn–Tucker vector of the problem (cf. (3.3) and (5.4))

$$\begin{aligned} &\text{minimize} \quad \tfrac{1}{2}c\|x_0\|^2 + h(x_1) + \langle p_2^{r-1}, x_0\rangle \\ &\text{subject to} \quad x_0 + Bx_1 = b. \end{aligned} \tag{5.6}$$

Let (x_0^r, x_1^r) be any optimal solution of the above problem. The Kuhn–Tucker conditions for (5.6) yields

$$x_1^r = \arg\min_{x_1} \tfrac{1}{2}c\|Bx_1 - b\|^2 + h(x_1) + \langle p_2^{r-1}, b - Bx_1\rangle,$$

$$p_2^r = p_2^{r-1} + c(b - Bx_1^r),$$

which we recognize to be the *method of multipliers* [5, 8, 10] (also see [2, 12]).

The above connection between the method of multipliers and dual coordinate descent can also be inferred from the duality relation between the method of multipliers and the proximal minimization algorithm (see [12]) together with the connection between the latter algorithm and coordinate ascent (see [2]).

Since Assumptions A and B hold for (5.2), it follows from Theorem 3.1 that the method of multipliers converges in the sense that the generated sequence $\{x_1^r\}$ is bounded and any of its cluster points is an optimal solution of (5.1). This result is however weaker than those available [12, Theorem 4] since, in particular, it requires f to be co-finite and does not asert convergence of the dual sequence $\{p_2^r\}$.

6. Convergence proof for the DBCA algorithm

In this section, we prove Theorem 3.1. Our proof is patterned after the arguments given in [15, Section 3], but due to the presence of functions that are not strictly convex, a number of new proof ideas must be introduced. We prove part (a) below. Part (b) follows readily from part (a).

First we state a technical lemma from [14] (see Lemmas 1 and 2 therein; also see [15, Lemma 1]).

Lemma 6.1. *Let h be any proper closed convex function in $\mathbb{R}^m$ ($m \geq 1$) that is continuous on its effective domain, which we denote by S. The following hold:*

(a) *For any $y \in S$, any $z \in \mathbb{R}^m$ such that $y + z \in S$, and any sequences $\{y^1, y^2, \ldots\} \to y$ and $\{z^1, z^2, \ldots\} \to z$ such that $y^t \in S$ and $y^t + z^t \in S$ for all t,*

$$h'(y; z) \geq \limsup_{t\to\infty} h'(y^t, z^t).$$

(b) *If in addition h is co-finite, then for any $y \in S$ and any sequence of vectors $\{y^1, y^2, \ldots\}$ in S such that $\{h(y^t) + h'(y^t; y - y^t)\}$ is bounded from below, both $\{y^t\}$ and $\{h(y^t)\}$ are bounded, and every cluster point of $\{y^t\}$ is in S.* $\square$

Let $p^r=(p_1^r,\ldots,p_K^r)$, $r=0,1,2,\ldots$, be a dual sequence generated by the DBCA algorithm (3.1)-(3.2) using the essentially cyclic order of iteration. Let $\{(x_0^r,x_1^r,\ldots,x_K^r)\}$ be the associated primal sequence given by (3.4)-(3.5).

We have from (3.4) and [11, Theorem 23.5] that

$$\sum_{k=1}^{K} A_k^{\mathrm{T}} p_k^r \in \partial f_0(x_0^r) \quad \forall r \geq 1, \tag{6.1}$$

Also, we have from (3.5) that, for every $k\in\{1,\ldots,K\}$,

$$B_k^{\mathrm{T}} p_k^r \in \partial f_k(x_k^r) \quad \forall r \geq 1 \text{ with } s^r=k, \tag{6.2}$$

$$A_k x_0^r + B_k x_k^r = b_k \quad \forall r \geq 1 \text{ with } s^r=k, \tag{6.3}$$

$$x_k^r = x_k^{r-1} \quad \forall r \geq 1 \text{ with } s^r \neq k. \tag{6.4}$$

By the essentially cyclic order of iteration, for any $k\in\{1,\ldots,K\}$ and any $r\geq T$, there is an index j not exceeding r with $s^j=k$. Choose j to be the largest such index. Then, by (3.2), $p_k^r=p_k^j$ and, by (6.4), $x_k^r=x_k^j$, so (6.2) yields $B_k^{\mathrm{T}}p_k^r\in\partial f_k(x_k^r)$. The above choice of k and r was arbitrary, so

$$B_k^{\mathrm{T}} p_k^r \in \partial f_k(x_k^r) \quad \forall r \geq T,\ k=1,\ldots,K. \tag{6.5}$$

To simplify the presentation, let

$$\mathcal{K}=\{k\in\{0,1,\ldots,K\}\,|\,f_k \text{ is co-finite}\}.$$

(By Assumption A(b), we have $0\in\mathcal{K}$.) Let

$$\bar{x}=(\bar{x}_0,\ldots,\bar{x}_K)$$

be any optimal solution of (2.1). The next lemma provides lower bounds on the per iteration improvement in the dual functional value and on the duality gap.

Lemma 6.2. *For each $r>T$,*

$$q(p^r)-q(p^{r-1}) \geq f_0(x_0^r)-f_0(x_0^{r-1})-f_0'(x_0^{r-1};\, x_0^r-x_0^{r-1}), \tag{6.6}$$

$$f(\bar{x})-q(p^r) \geq f_k(\bar{x}_k)-f_k(x_k^r)-f_k'(x_k^r;\, \bar{x}_k-x_k^r), \quad k=0,1,\ldots,K. \tag{6.7}$$

(*Recall that f is the function given by* (2.6).)

Proof. We first prove (6.6). For each $r\geq T$, we see from (6.1) and (6.5) that $(x_0^r,\ldots,x_K^r)$ is a solution of the minimization in (2.2) with $(p_1,\ldots,p_K)$ set to $(p_1^r,\ldots,p_K^r)$, so (2.2) yields

$$q(p^r)=\sum_{k=0}^{K} f_k(x_k^r)+\sum_{k=1}^{K}\langle p_k^r,\, b_k-A_k x_0^r-B_k x_k^r\rangle. \tag{6.8}$$

Fix any $r > T$. We have from the above relation that

$$q(p^r) - q(p^{r-1}) = f_0(x_0^r) - f_0(x_0^{r-1}) - \left\langle \sum_{k=1}^{K} A_k^T p_k^{r-1}, x_0^r - x_0^{r-1} \right\rangle$$

$$+ \sum_{k=1}^{K} [f_k(x_k^r) - f_k(x_k^{r-1}) - \langle B_k^T p_k^{r-1}, x_k^r - x_k^{r-1} \rangle$$

$$+ \langle p_k^r - p_k^{r-1}, b_k - A_k x_0^r - B_k x_k^r \rangle]$$

$$= f_0(x_0^r) - f_0(x_0^{r-1}) - \left\langle \sum_{k=1}^{K} A_k^T p_k^{r-1}, x_0^r - x_0^{r-1} \right\rangle$$

$$+ \sum_{k=1}^{K} [f_k(x_k^r) - f_k(x_k^{r-1}) - \langle B_k^T p_k^{r-1}, x_k^r - x_k^{r-1} \rangle],$$

where the last equality follows from $\langle p_k^r - p_k^{r-1}, b_k - A_k x_0^r - B_k x_k^r \rangle = 0$ for all $k \neq 0$ (cf. (3.2), (6.3)). Since $B_k^T p_k^{r-1} \in \partial f_k(x_k^{r-1})$ for all $k \neq 0$ (cf. (6.5)) and the f_k's are convex, then every quantity enclosed by a pair of brackets above is nonnegative, so

$$q(p^r) - q(p^{r-1}) \geqslant f_0(x_0^r) - f_0(x_0^{r-1}) - \left\langle \sum_{k=1}^{K} A_k^T p_k^{r-1}, x_0^r - x_0^{r-1} \right\rangle.$$

This together with (6.1) and the relation between the directional derivative and the subdifferential (i.e., (1.2)) proves (6.6).

The proof of (6.7) is similar to that of (6.6). Fix any $r \geqslant T$. Since $\bar{x}$ is feasible for (2.1), we have

$$f(\bar{x}) - q(p^r) = f(\bar{x}) - q(p^r) + \sum_{k=1}^{K} \langle p_k^r, b_k - A_k \bar{x}_0 - B_k \bar{x}_k \rangle$$

$$= \left[f_0(\bar{x}_0) - f_0(x_0^r) - \left\langle \sum_{k=1}^{K} A_k^T p_k^r, \bar{x}_0 - x_0^r \right\rangle \right]$$

$$+ \sum_{k=1}^{K} [f_k(\bar{x}_k) - f_k(x_k^r) - \langle B_k^T p_k^r, \bar{x}_k - x_k^r \rangle]$$

$$\geqslant [f_0(\bar{x}_0) - f_0(x_0^r) - f_0'(x_0^r; \bar{x}_0 - x_0^r)]$$

$$+ \sum_{k=1}^{K} [f_k(\bar{x}_k) - f_k(x_k^r) - f_k'(x_k^r; \bar{x}_k - x_k^r)],$$

where the second equality follows from (2.6) and (6.8) and the inequality follows from $\sum_{k=1}^{K} A_k^T p_k^r \in \partial f_0(x_0^r)$ (cf. (6.1)), $B_k^T p_k^r \in \partial f_k(x_k^r)$ for $k = 1, \ldots, K$ (cf. (6.5)), and (1.2). Since the f_k's are convex, every quantity enclosed by a pair of brackets is nonnegative. This proves (6.7). □

For notational convenience, let

$$S_k = \operatorname{dom} f_k, \quad k = 0, \ldots, K.$$

Lemmas 6.1 and 6.2 yield the following two lemmas:

Lemma 6.3.

(a) *For every $k \in \mathcal{K}$, both $\{x_k^r\}$ and $\{f_k(x_k^r)\}$ are bounded, and every cluster point of $\{x_k^r\}$ is in S_k.*

(b) $x_0^r - x_0^{r-1} \to 0$.

Proof. (a) Since $\{q(p^r)\}$ is nondecreasing, then $\{q(p^r)\}$ is bounded from below. Thus (6.7) implies that $\{f_k(x_k^r) + f'_k(x_k^r; \bar{x}_k - x_k^r)\}$ is bounded from below for all $k \in \mathcal{K}$. Now apply Lemma 6.1(b) to f_k, for each $k \in \mathcal{K}$.

(b) Since $0 \in \mathcal{K}$ (cf. Assumption A(b)), part (a) yields that $\{x_0^r\}$ is bounded. Hence if the claim does not hold, there would exist an infinite subsequence $\mathcal{R}$ of $\{0, 1, \ldots\}$ for which $\{x_0^{r-1}\}_{r \in \mathcal{R}}$ converges to some point $x_0^\infty + z$ with $z \neq 0$. By part (a) and $0 \in \mathcal{K}$, both x_0^∞ and $x_0^\infty + z$ are in S_0. Since (6.6) holds for all $r > T$, we then obtain, upon passing into the limit as $r \to \infty$, $r \in \mathcal{R}$, and using Lemma 6.1(a) and the continuity of f_0 on S_0, that

$$\liminf_{r \to \infty,\ r \in \mathcal{R}} q(p^r) - q(p^{r-1}) \geq f_0(x_0^\infty + z) - f_0(x_0^\infty) - f'_0(x_0^\infty; z).$$

Since f_0 is, by Assumption A(c), strictly convex (so the right-hand side of the above relation is a positive scalar), it follows that $q(p^r) \to \infty$. This, in view of the weak duality condition

$$\max_p q(p) \leq \min_{x \in X} f(x),$$

contradicts the feasibility of (2.1) (cf. Assumption B). $\square$

By using Lemma 6.3 and the assumption that the order of iteration is essentially cyclic, we obtain the following intermediate result.

Lemma 6.4. *The sequence $\{(x_0^r, \ldots, x_K^r)\}$ is bounded and its cluster points are in $(S_0 \times \cdots \times S_K) \cap X$.*

Proof. By Lemma 6.3(a), $\{x_k^r\}$ is bounded for all $k \in \mathcal{K}$. Since $0 \in \mathcal{K}$ so $\{x_0^r\}$ is bounded and B_k has rank m_k for all $k \notin \mathcal{K}$, then (6.3) and (6.4) imply that $\{x_k^r\}$ is bounded for all $k \notin \mathcal{K}$. Thus, $\{(x_0^r, \ldots, x_k^r)\}$ is bounded.

Let $(x_0^\infty, \ldots, x_K^\infty)$ be any cluster point of $\{(x_0^r, \ldots, x_K^r)\}$. Fix any $k \in \{1, \ldots, K\}$ and, for each $r \geq T$, let $\tau(r)$ be the largest integer j not exceeding r such that $s^j = k$. Then, by (6.3) and (6.4),

$$A_k x_0^r + B_k x_k^r = \sum_{j=\tau(r)}^{r-1} A_k(x_0^{j+1} - x_0^j) + b_k \quad \forall r \geq T.$$

Since $r-\tau(r)\leq T$ for all $r\geq T$, then, upon passing to the limit and using Lemma 6.3(b), we obtain

$$A_k x_0^\infty + B_k x_k^\infty = b_k.$$

Since the choice of k above was arbitrary so the above relation holds for all $k\in\{1,\ldots,K\}$, then (cf. (2.5)) $(x_0^\infty,\ldots,x_K^\infty)\in X$.

By Lemma 6.3(a), $x_x^\infty\in S_k$ for all $k\in\mathcal{K}$, so it only remains to show that this holds for all $k\notin\mathcal{K}$. We will argue by contradiction. Suppose that there exists some $s\notin\mathcal{K}$ such that $x_s^\infty\notin S_s$ (so $f_s(x_s^\infty)=\infty$). Let $\mathcal{R}$ be any infinite subsequence of $\{T, T+1,\ldots\}$ such that

$$\{x_s^r\}_{r\in\mathcal{R}}\to x_s^\infty. \tag{6.9}$$

Since f_s is closed and $f_s(x_s^\infty)=\infty$, we immediately have

$$\{f_s(x_s^r)\}_{r\in\mathcal{R}}\to\infty. \tag{6.10}$$

Also, we have from (6.7) and the boundedness of $\{q(p^r)\}$ that $\{f_s(x_s^r)+f_s'(x_s^r;\bar{x}_s-x_s^r)\}$ is bounded below, so there exists some scalar constant $\beta>0$ such that

$$f_s'(x_s^r;\bar{x}_s-x_s^r)\geq -f_s(x_s^r)-\beta\quad\forall r\in\mathcal{R}. \tag{6.11}$$

Since (6.10) holds, by taking $r\in\mathcal{R}$ sufficiently large if necessary, we can assume that $f_s(x_s^r)\geq 1$ for all $r\in\mathcal{R}$. For each $r\in\mathcal{R}$, let $\lambda^r=1/\sqrt{f_s(x_s^r)}$ and let

$$y_s^r=\bar{x}_s(1-\lambda^r)+\lambda^r x_s^r.$$

Then, we have $0<\lambda^r\leq 1$ for all $r\in\mathcal{R}$, so that $y_s^r\in S_s$ for all $r\in\mathcal{R}$ (cf. the convexity of S_s, $\bar{x}_s\in S_s$ and $x_s^r\in S_s$ for all $r\geq T$). Thus, we obtain from the convexity of f_s that, for all $r\in\mathcal{R}$,

$$\begin{aligned}
f_s(y_s^r) &= f_s(\bar{x}_s(1-\lambda^r)+\lambda^r x_s^r)\\
&\geq f_s(x_s^r)+(1-\lambda^r)f_s'(x_s^r;\bar{x}_s-x_s^r)\\
&\geq f_s(x_s^r)+(1-\lambda^r)(-f_s(x_s^r)-\beta)\\
&=\sqrt{f_s(x_s^r)}+(1-1/\sqrt{f_s(x_s^r)})\beta,
\end{aligned}$$

where the second inequality follows from (6.11) and the last equality follows from the definition of λ^r. Hence, by (6.10),

$$\{f_s(y_s^r)\}_{r\in\mathcal{R}}\to\infty. \tag{6.12}$$

On the other hand, (6.9) together with $\{\lambda^r\}_{r\in\mathcal{R}}\to 0$ (cf. (6.10)) and the definition of y_s^r implies that

$$\{y_s^r\}_{r\in\mathcal{R}}\to\bar{x}_s. \tag{6.13}$$

Since $f_s(\bar{x}_s)<\infty$ and $y_s^r\in S_s$ for all $r\in\mathcal{R}$, then (6.12) and (6.13) show that f_s is not closed, a contradiction of our assumptions. □

Proof of Theorem 3.1. By using Lemmas 6.1(a) 6.3 and 6.4, we can now prove Theorem 3.1(a). To simplify the notation, let

$$x^r = (x_0^r, \ldots, x_K^r) \quad \forall r,$$

and let E denote the $n_1 + \cdots + n_K$ by $m_0 + \cdots + m_K$ constraint matrix for (2.1), that is,

$$E = \begin{pmatrix} A_1 & B_1 & & & \\ A_2 & & B_2 & & \\ \vdots & & & \ddots & \\ A_K & & & & B_K \end{pmatrix}. \tag{6.14}$$

Also, by Assumption B, there exist a polyhedral set S' and a convex set S'' such that

$$S' \cap S'' = \operatorname{dom} f, \qquad S' \cap \operatorname{ri}(S'') \cap X \neq \emptyset. \tag{6.15}$$

Let x^∞ be any cluster point of $\{x^r\}$ and let $\mathscr{R}$ be any subsequence of $\{T, T+1, \ldots\}$ such that $\{x^r\}_{r \in \mathscr{R}}$ converges to x^∞. By Lemma 6.4, $x^\infty \in \operatorname{dom} f \cap X$. Suppose that $x^\infty \neq \bar{x}$ (otherwise we are done). Let y be any element of $S' \cap \operatorname{ri}(S'') \cap X$ (such an y exists by (6.15)). Fix any $\lambda \in (0, 1)$ and denote

$$y(\lambda) = \lambda y + (1 - \lambda)\bar{x}. \tag{6.16}$$

Then $y(\lambda) \in S' \cap \operatorname{ri}(S'') \cap X$. Let

$$z = y(\lambda) - x^\infty. \tag{6.17}$$

Since S' is polyhedral and z belongs to the tangent cone of S' at x^∞, this implies that, for any fixed $\delta \in (0, \lambda)$, there holds

$$x^r + \delta z \in S' \quad \forall r \in \mathscr{R} \text{ sufficiently large.}$$

On the other hand, since $y(\lambda) \in \operatorname{ri}(S'')$, $x^\infty \in S''$ and $0 < \delta < \lambda$, we have $x^\infty + \delta z \in \operatorname{ri}(S'')$, which together with the facts $x^r \in S''$, for all r, and $\{x^r\}_{r \in \mathscr{R}} \to x^\infty$ yields

$$x^r + \delta z \in S'' \quad \forall r \in \mathscr{R} \text{ sufficiently large.}$$

Combining the above two relations and we obtain

$$x^r + \delta z \in S' \cap S'' = \operatorname{dom} f \quad \forall r \in \mathscr{R} \text{ sufficiently large,} \tag{6.18}$$

where the equality follows from (6.15). Since both $y(\lambda)$ and x^∞ are in $\operatorname{dom} f$ so (by (6.17) and $0 < \delta < 1) x^\infty + \delta z \in \operatorname{dom} f$, (6.18) together with Lemma 6.1(a) implies

$$f'(x^\infty; z) \geq \limsup_{r \to \infty \; r \in \mathscr{R}} f'(x^r; z).$$

Since $E^{\mathrm{T}} p^r \in \partial f(x^r)$ for all r (cf. (2.6), (6.1), (6.5), (6.14)) so, by (1.2), $f'(x^r; z) \geq \langle p^r, Ez \rangle$ for all r, the above relation yields

$$f'(x^\infty; z) \geq \limsup_{r \to \infty \; r \in \mathscr{R}} \langle p^r, Ez \rangle. \tag{6.19}$$

On the other hand, since $y(\lambda) \in X$ and $x^\infty \in X$, there holds $Ez = 0$ (cf. (2.5) and (6.14), (6.17)) so that

$$\langle p^r, Ez \rangle = 0 \quad \forall r \in \mathcal{R},$$

which together with (6.19) yields

$$f'(x^\infty; z) \geqslant 0.$$

Since f is convex, it then follows from (6.17) that $f(x^\infty) \leqslant f(y(\lambda))$. Since the choice of $\lambda \in (0, 1)$ was arbitrary, by taking λ arbitrarily small (and using the continuity of f on its effective domain), we obtain $f(x^\infty) \leqslant f(y(0)) = f(\bar{x})$, where the equality follows from (6.16). Since $\bar{x}$ minimizes f over X and $x^\infty \in X$, this implies that x^∞ minimizes f over X also. □

Acknowledgement

Thanks are due to Professor D.P. Bertsekas for his numerous helpful comments.

References

[1] A. Auslender, *Optimisation Méthodes Numériques* (Masson, Paris, 1976).
[2] D.P. Bertsekas and J.N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods* (Prentice-Hall, Englewood Cliffs, NJ, 1989).
[3] L.M., Bregman, "The relaxation method of finding the common point convex sets and its application to the solution of problems in convex programming," *USSR Computational Mathematics and Mathematical Physics* 7 (1967) 200-217.
[4] Y. Censor and A. Lent, "Optimization of "log x" entropy over linear equality constraints," *SIAM Journal on Control and Optimization* 25 (1987) 921-933.
[5] P.C. Haarhoff and J.D. Buys, "A new method for the optimization of a nonlinear function subject to nonlinear constraints," *The Computer Journal* 13 (1970) 178-184.
[6] S.-P. Han, "A successive projection method," *Mathematical Programming* 40 (1988) 1-14.
[7] S.-P. Han, "A decomposition method and its application to convex programming," *Mathematics of Operations Research* 14 (1989) 237-248.
[8] M.R. Hestenes, "Multiplier and gradient methods," *Journal of Optimization Theory and Applications* 4 (1969) 303-320.
[9] J.-S. Pang, "On the convergence of dual ascent methods for large-scale linearly constrained optimizaton problems," Unpublished manuscript, School of Management, The University of Texas (Dallas, 1984).
[10] M.J.D. Powell, "A Method for Nonlinear Constraints in Minimization Problems," in: R. Fletcher, ed., *Optimization* (Academic Press, New York, 1969) pp. 283-298.
[11] R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, NJ, 1970).
[12] R.T. Rockafellar, "Augmented lagrangians and applications of the proximal point algorithm in convex programming," *Mathematics of Operations Research* 1 (1976) 97-116.
[13] P. Tseng and D.P. Bertsekas, "Relaxation methods for problems with strictly convex separable costs and linear constraints," *Mathematical Programming* 38 (1987) 303-321.
[14] P. Tseng and D.P. Bertsekas, "Relaxation methods for problems with strictly convex costs and linear inequality constraints," *Mathematics of Operations Research* 16 (1991) 462-481.
[15] P. Tseng, "Dual ascent methods for problems with strictly convex costs and linear costraints: a unified approach," *SIAM Journal on Control and Optimization* 28 (1990) 214-242.