December 17, 1988 LIDS-P-1840

Coordinate Ascent for Maximizing Nondifferentiable Concave Functions*

by

Paul Tseng†

Abstract

We present a coordinate ascent method for maximizing concave (not necessarily differentiable) functions possessing a certain separable structure. This method, when applied to the dual of a linearly constrained convex program, includes as special cases a successive projection algorithm of Han [11], the method of multipliers [10, 12, 20], and a number of dual coordinate ascent methods [5-7, 13, 15, 17-18, 25, 30]. We also generalize the results of Auslender [1, §6] and of Bertsekas-Tsitsiklis [3, §3.3.5] on the convergence of this method. One primal application of this method is the proximal minimization algorithm [23].

KEY WORDS: separable convex programming, coordinate ascent, method of multipliers

---

* This research is partially supported by the U.S. Army Research Office, contract DAAL03-86-K-0171 (Center for Intelligent Control Systems), and by the National Science Foundation under grant NSF-ECS-8519058.

† Laboratory for Information and Decision Systems, Room 35-205, Massachusetts Institute of Technology, Cambridge, MA 02139.

Acknowledgement: Thanks are due to Professor D. P. Bertsekas for his numerous helpful comments and for bringing to my attention the work of Auslender.

1. **Introduction**

Consider linearly constrained convex programs of the form

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{subject to} & Ex = 0, \end{array}$$

where $f:\Re^m \to (-\infty,+\infty]$ is a convex function and E is an $n \times m$ matrix. A classical method for solving this problem, the dual coordinate ascent method, assigns a Lagrange multiplier vector $p \in \Re^n$ to the constraints $Ex = 0$ and, at each iteration, adjusts a subset of the components of p to maximize (perhaps approximately) the dual functional

$$q(p) = \min_x \{ f(x) - \langle p, Ex \rangle \}$$

while the other components of p are held fixed. This method has a number of advantages: it is easy to implement, uses little storage, can readily exploit special structures in either the cost or the constraint matrix, and is highly parallelizable on sparse problems (see [4, 15, 17, 28, 30-31] for computational tests). Its convergence properties have been studied extensively [4-9, 13-16, 18-19, 21, 24-29], but the analyses have required the dual functional q to be <u>differentiable</u>, which is often not the case when, for example, a problem is transformed in a way to bring about a structure that is favorable for decomposition. A solution to this difficulty has been suggested recently by a successive projection algorithm of Han [11]. Han's algorithm, as it turns out, is effectively a coordinate ascent algorithm for maximizing functions q having the special form

$$q(p_1,\ldots,p_K) = \phi(p_1,\ldots,p_K) + \sum_{k=1}^{K} \phi_k(p_k), \tag{1.1}$$

where each $\phi_k:\Re^{n_k} \to [-\infty,+\infty)$ is a concave function and $\phi:\Re^{n_1+\cdots+n_K} \to \Re$ is a differentiable concave function (see §4 for details). The remarkable fact here is that, although q may not be differentiable or have bounded level sets, the separable structure of q nonetheless enables it to be maximized by coordinate ascent. Coordinate ascent for maximizing concave functions of the form (1.1) has been considered by Auslender [1, pp. 94], but convergence requires q to be strictly convex and to have bounded level sets.

In this paper, we generalize the results in [1] and [11]. More specifically, we propose a coordinate ascent method for maximizing concave functions of the form (1.1), give conditions under which it converges, and show that it contains as special cases the method of multipliers, Auslender's algorithm, the proximal minimization algorithm, some dual coordinate ascent methods, as well as Han's algorithm.

[Hence our results unify a number of seemingly unrelated methods.]

This paper is organized as follows: in §2 we present a coordinate ascent method for maximizing concave functions of the form (1.1). In §3 we generalize the results of Auslender and Bertsekas-Tsitsiklis on the convergence of this method and give an application to the proximal minimization algorithm. In §4 we apply this method to solve certain separable linearly constrained convex programs. In §5 to §6 we show that this method contains as special cases the algorithm of Han and the method of multipliers.

All actions will take place in the Euclidean space. All vectors are column vectors, and superscript T denotes transpose. We let $\langle\cdot,\cdot\rangle$ denote the usual Euclidean inner product and let $\|\cdot\|$ denote its induced norm. For any set S in $\Re^m$ $(m \geq 1)$, we denote by int(S), ri(S) and cl(S) respectively the interior, the relative interior, and the closure of S. For any convex function $h:\Re^m \to (-\infty,+\infty]$, we denote by dom(h) the <u>effective domain</u> of h, i.e.,

$$\mathrm{dom}(h) = \{\, x \in \Re^m \mid h(x) < +\infty \,\}.$$

For any x and d in $\Re^m$, we denote by $\partial h(x)$ the <u>subdifferential</u> of h at x, and by $h'(x;d)$ the <u>directional derivative</u> of h at x in the direction d [22, pp. 213 and 217], i.e.,

$$h'(x;d) = \max\{\langle d,\eta\rangle \mid \eta \in \partial h(x)\}. \tag{1.2}$$

The effective domain and the directional derivative of concave functions are defined analogously.

2. A Coordinate Ascent Method for Concave Programming

Consider the following concave program

$$\text{Maximize} \quad \phi(p_1,\ldots,p_K) + \sum_{k=1}^{K} \phi_k(p_k) \tag{2.1}$$
subject to no constraint on the p_k's,

where $\phi:\Re^{n_1+\cdots+n_K} \to \Re$ is a concave, continuously differentiable function and each $\phi_k:\Re^{n_k} \to [-\infty,+\infty)$ is a concave, upper semicontinuous function that is continuous in its effective domain.

Let $\phi:\Re^{n_1+\dots+n_K}\to[-\infty,+\infty)$ denote the objective function for (2.1), i.e.

$$\Phi(p_1,\dots,p_K) = \phi(p_1,\dots,p_K) + \sum_{k=1}^{K} \phi_k(p_k).$$

Consider the following coordinate ascent algorithm, which we call the block coordinate relaxation (BCR) algorithm, for solving (2.1):

The BCR Algorithm

Begin with any multiplier vector $(p_1,\dots,p_K)\in \mathrm{dom}(\Phi)$.

At each iteration, choose an $s\in\{1,\dots,K\}$ and set

$$p_s := \mathrm{argmax}_{\Delta}\, \Phi(p_1,\dots,p_{s-1},\Delta,p_{s+1},\dots,p_K).$$

We will use the following rule to choose the index s at each iteration:

Essentially Cyclic Rule: There exists a positive constant T for which every $k\in\{1,\dots,K\}$ is chosen at least once between iterations r and r+T ($r = 0,1,\dots$).

Note that a special case of Essentially Cyclic rule is the classical cyclic rule for relaxation.

3. Primal Application

Let $p^r = (p_1{}^r,\dots,p_K{}^r)$ denote the iterate generated by the BCR algorithm at the rth iteration ($r = 0,1,\dots$) and let s^r denote the choice of $s\in\{1,\dots,K\}$ at the rth iteration. By modifying an argument given in [3, §3.3.5], we have the following general result:

Proposition 1 If the level sets of Φ are bounded and, for each k, Φ is strictly concave in p_k when the value of p_j, $j \neq k$, are fixed, then, under the Essentially Cyclic rule, the sequence $\{p^r\}$ is bounded and each of its limit points is an optimal solution of (2.1).

Proof: First, because $\Phi(p^0) > -\infty$ and the value of Φ is increased at each iteration, we have that

$$-\infty < \Phi(p^0) \leq \Phi(p^1) \leq \Phi(p^2) \leq \ \ldots .$$

Since Φ has bounded level sets, this implies that $\{p^r\}$ is bounded and $\Phi(p^r)$ converges to some limit. Next we claim that

$$p^r - p^{r+1} \to 0.$$

Suppose that this were not the case. Then there exist subsequence $\{p^r\}_{r\in R}$ and scalar $\varepsilon \geq 0$ such that $\|p^r - p^{r-1}\| \geq \varepsilon$ for all $r \in R$. Since $\{p^r\}$ is bounded, by further passing into a subsequence if necessary, we can assume that

$$\{p^r\}_{r\in R} \to \text{some } p', \qquad \{p^{r+1}\}_{r\in R} \to \text{some } p'', \tag{3.1}$$

and s^r = some s, for all $r \in R$. Hence, for each $r \in R$, p^{r+1} is obtained from p^r by maximizing Φ with respect to p_s, while the other components are fixed. Since $(p^r+p^{r+1})/2$ differs from p^r only in the component p_s, this implies that

$$\Phi((p^r+p^{r+1})/2) \leq \Phi(p^{r+1}) \ \forall\, r \in R. \tag{3.2}$$

Since Φ is continuous within its effective domain and $\Phi(p^r)$ converges to a limit, (3.1) and (3.2) imply that

$$\Phi(p') = \Phi(p''), \qquad \Phi((p'+p'')/2) \leq \Phi(p''),$$

which contradicts the assumption that Φ is strictly concave in p_s (p'' and p' differ only in the component p_s).

Let p^∞ be any limit point of the sequence $\{p^r\}$ and consider a subsequence $\{p^r\}_{r\in R}$ converging to p^∞. Consider any $k \in \{1,\ldots,K\}$. Since $p^r - p^{r+1} \to 0$, the Essentially Cyclic rule implies that there exists another subsequence $\{p^r\}_{r\in R'}$ converging to p^∞ such that $s^{r-1} = k$ for all $r\in R'$. Fix any $r\in R'$. Since $p_k{}^r$ maximizes $\phi(p_1{}^{r-1},\ldots,p_{k-1}{}^{r-1},\Delta,p_{k+1}{}^{r-1},\ldots,p_m{}^{r-1}) + \phi_k(\Delta)$ over all Δ, the maximality conditions yield

$$\langle\, y_k - p_k{}^r \,,\, \partial\phi(p^r)/\partial p_k \,\rangle + \phi_k'(p_k{}^r; y_k - p_k{}^r) \;\le\; 0, \;\; \forall\; y_k \in \mathrm{dom}(\phi_k).$$

Taking the limit as r tends to infinity, $r\in R'$, and using the continuity of $\nabla\phi$ and Lemma 3 in [27], we obtain that

$$\langle\, y_k - p_k{}^\infty \,,\, \partial\phi(p^\infty)/\partial p_k \,\rangle + \phi_k'(p_k{}^\infty; y_k - p_k{}^\infty) \;\le\; 0, \;\; \forall\; y_k \in \mathrm{dom}(\phi_k).$$

Adding the above inequality over all k yields

$$\Phi'(p^\infty; y - p^\infty) \;\le\; 0, \;\; \forall\; y \in \mathrm{dom}(\Phi).$$

Hence p^∞ is an optimal solution of (2.1). Q.E.D.

Proposition 1 improves upon both Proposition 3.9 in [3, §3.3.5] (which further assumes that Φ is differentiable and that cyclic order of relaxation is used) and Theorem 1.2 in [1, pp. 95] (which further assumes that Φ is strictly convex in <u>every</u> coordinate and that cyclic order of relaxation is used). [See also [8, 16, 19, 24, 29, 33-34] for related results.] We remark that Proposition 1 can be further extended to the case where ϕ is quasiconcave (instead of concave) and Φ is <u>hemivariate</u> [32] (instead of strictly concave) in each p_k.

Proposition 1 has the following interesting application (cf. [3, §3.4.3]): Consider a convex, lower semicontinuous function $\psi:\Re^n \to (-\infty,+\infty]$ that is continuous in its effective domain. Then the BCR algorithm applied to maximize the function $-c\|p_2 - p_1\|^2 - \psi(p_1)$ $(c > 0)$ is equivalent to the algorithm:

$$p_1 \;:=\; \mathrm{argmin}_{\,\Delta}\{\; c\|p_1 - \Delta\|^2 + \psi(\Delta) \;\},$$

which we recognize to be the <u>proximal minimization algorithm</u> for minimizing ψ [23]. Proposition 1 implies that this algorithm converges if ψ has bounded level sets. [This however is not the sharpest result available for this algorithm.]

4. Dual Application

Consider the following convex program

$$\text{Minimize} \quad f_0(x_0) + \ldots + f_K(x_K) \tag{4.1}$$
$$\text{subject to} \quad A_k x_0 + B_k x_k = b_k, \ k = 1,\ldots,K,$$

where $f_k:\Re^{m_k}\to(-\infty,+\infty]$, A_k is an $n_k\times m_0$ real matrix, B_k is an $n_k\times m_k$ real matrix, and each b_k is an n_k-vector. We can also have inequality constraints in (4.1), but for simplicity we will consider only equality constraints throughout.

Let X denote the constraint set of (4.1), i.e.,

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

and let $g_k:\Re^{m_k}\to(-\infty,+\infty]$ denote the <u>conjugate function</u> of f_k [22, pp. 104], i.e.

$$g_k(t) = \sup_{\xi}\{ \langle t,\xi\rangle - f_k(\xi) \}. \tag{4.2}$$

We make the following assumptions about the problem (4.1):

Assumption A:

(a) f_0 is strictly convex, lower semicontinuous and continuous within $\mathrm{dom}(f_0)$. g_0 is real valued.

(b) For each $k \neq 0$, f_k is convex lower semicontinuous and continuous within $\mathrm{dom}(f_k)$. Either g_k is real valued or B_k has rank m_k and $\mathrm{dom}(f_k)$ is closed.

(c) For each k, $\mathrm{dom}(f_k) = S_k'\cap S_k''$, where S_k'' is a convex set and $\mathrm{cl}(S_k')$ is a polyhedral set. Moreover $(S_1'\times\ldots\times S_K')\cap\mathrm{ri}(S_1''\times\ldots\times S_K'')\cap X \neq \varnothing$.

Assumption A essentially ensures that the f_k's and the g_k's are well-behaved and that (4.1) is feasible. Assumption A (c) is actually slightly stronger than the assumption that (4.1) is feasible, i.e. $(S_1'\times\ldots\times S_K')\cap(S_1''\times\ldots\times S_K'')\cap X \neq \varnothing$, but is still very general. For example, if f_k is <u>separable</u> [26], then $\mathrm{cl}(\mathrm{dom}(f_k))$ (but not necessarily $\mathrm{dom}(f_k)$) is a polyhedral set (it is in fact a box).

Let $f:\Re^{m_0+\cdots+m_K}\to(-\infty,+\infty]$ denote the objective function of (4.1), i.e.

$$f(x_0,\ldots,x_K) = \sum_k f_k(x_k). \tag{4.3}$$

By Assumptions A (a)-(b), f has compact level sets on X. This, together with the assumption (cf. Assumption A (c)) that (4.1) is feasible, implies that the set of optimal solutions for (4.1) is nonempty and bounded. Furthermore, because f_0 is strictly convex, the first m_0 components of the optimal solutions are unique.

By assigning a Lagrange multiplier vector p_k to the constraints $A_kx_0 + B_kx_k = b_k$, we obtain the following dual of (4.1):

$$\begin{aligned}&\text{Maximize} \quad q(p_1,\ldots,p_K)\\ &\text{subject to no constraint on the } p_k\text{'s},\end{aligned} \tag{4.4}$$

where (cf. (4.2)) $q:\Re^{n_1+\cdots+n_K}\to(-\infty,+\infty]$ is the dual functional

$$\begin{aligned} q(p_1,\ldots,p_K) &= \min\{ \textstyle\sum_k f_k(x_k) + \sum_{k\neq 0} \langle p_k, b_k - A_kx_0 - B_kx_k\rangle \} \\ &= \textstyle\sum_{k\neq 0} \langle p_k,b_k\rangle - g_0(\sum_{k\neq 0} A_k{}^T p_k) - \sum_{k\neq 0} g_k(B_k{}^T p_k). \end{aligned} \tag{4.5}$$

From (4.5) we see that q is a concave function (since each g_k is a convex function) and has the form (1.1) (since g_0 is differentiable by Theorem 26.3 of [22]). Furthermore, strong duality holds for (4.1) and (4.4), i.e. the optimal value in (4.1) equals the optimal value in (4.4). [To see this, note that the set

$$\{ (x_0,\ldots,x_K,u_1,\ldots,u_K,z) \mid A_kx_0 + B_kx_k = u_k,\ k = 1,\ldots,K,\ \textstyle\sum_k f_k(x_k) \le z \}$$

is closed. Hence the <u>convex bifunction associated with (4.1)</u> [22, pp. 293] is closed. Since the optimal solution set for (4.1) is bounded, Theorem 30.4 (i) in [22] states that (4.1) and (4.4) have the same optimal value.] Note, however, that (4.4) may or may not have an optimal solution.

Since q has the form (1.1), let us apply the BCR algorithm to solve (4.4). To ensure that the algorithm is well defined, we make the following technical assumption:

Assumption B: For each $k \neq 0$, there exists an $x_0 \in \mathrm{ri}(\mathrm{dom}(f_0))$ and an $x_k \in \mathrm{ri}(\mathrm{dom}(f_k))$ satisfying $A_k x_0 + B_k x_k = b_k$.

[Note that Assumption B holds if f_0 is real-valued and each A_k has full row rank (an example of this is given in §5 and §6).] It can be seen that, for any multiplier vectors $p_1, \ldots, p_K$ and any $s \in \{1,\ldots,K\}$, there exists a p_s' that maximizes $q(p_1,\ldots,p_{s-1},\Delta,p_{s+1},\ldots,p_K)$ with respect to Δ if and only if p_s' is a Kuhn-Tucker vector of the problem

$$\begin{aligned} &\text{Minimize} && f_0(x_0) + f_s(x_s) - \textstyle\sum_{k\neq s}\langle p_k, A_k x_0\rangle \\ &\text{subject to} && A_s x_0 + B_s x_s = b_s. \end{aligned} \tag{4.6}$$

Since (cf. Assumption A) (4.6) has an optimal primal solution, Theorem 28.2 in [22], together with Assumption B, implies that (4.6) has a Kuhn-Tucker vector.

Let $p^r = (p_1{}^r,\ldots,p_K{}^r)$ denote the iterate generated by the BCR algorithm (applied to maximize q) at the rth iteration ($r = 0,1,\ldots$) and let s^r denote the choice of $s \in \{1,\ldots,K\}$ at the rth iteration. Also let

$$x_0{}^r = \nabla g_0(\textstyle\sum_k A_k{}^T p_k{}^r) \tag{4.7}$$

and, for each $s \neq 0$, let

$$x_s{}^r = \begin{cases} \operatorname{argmin}\{\, f_s(x_s) \mid B_s x_s = b_s - A_s x_0{}^r \,\} & \text{if } s^r = s, \\ x_s{}^{r-1} & \text{if } s^r \neq s, \end{cases} \tag{4.8}$$

where $x_s{}^0$ denotes any element of $\partial g_s(B_s{}^T p_s{}^0)$. [For simplicity we assume that $\partial g_s(B_s{}^T p_s{}^0)$ is nonempty, although it suffices that $g_s(B_s{}^T p_s{}^0)$ is finite.] Because q may not have bounded level sets and may not be strictly concave in each p_k, Proposition 1 is not applicable here. Nonetheless we can show the following result:

Proposition 2 Under the Essentially Cyclic rule, the following hold:

(a) $\{(x_0{}^r,\ldots, x_K{}^r)\}$ is bounded and each of its limit points is an optimal solution of (4.1).

(b) If cl(dom(f)) is a polyhedral set and there exists a closed ball B around x^* such that

$f'(x;(y-x)/\|y-x\|)$ is bounded for all x, y in $B \cap dom(f)$, where x^* denotes any optimal solution of (4.1), then $q(p^r) \to f(x^*)$.

(c) If the optimal solution set of (4.4) is nonempty and bounded, then $\{p^r\}$ is bounded and every one of its limit points is an optimal solution for (4.4).

The proof of Proposition 2 is rather long and is given in Appendix A. Proposition 2 generalizes the results in [5-7, 13, 17-18, 25] on the convergence of dual (single) coordinate ascent methods using exact line search (see [13] for additional references).

Extensions: Note that we can maximize q with respect to more than one p_s simultaneously. Also, if q is differentiable in p_s, for some s, then inexact maximization with respect to only some of the components of p_s is permissible – as is done in [4, 14, 26-28]. In fact, it is possible to write down a method that generalizes the BCR algorithm, the dual gradient algorithms [15, 18], and the dual coordinate ascent algorithms [28], but such a method would be somewhat cumbersome.

5. Han's Algorithm is a Special Case

In this section we show that the successive projection algorithm of Han [11] is a special case of the BCR algorithm. We also strengthen the results in [11] by applying Proposition 2.

Consider the following convex program treated in [11]

$$\begin{aligned} &\text{Minimize} \quad h(x) = \langle x-d, Q(x-d)\rangle/2 \\ &\text{subject to} \quad x \in C_1 \cap \ldots \cap C_K, \end{aligned} \tag{5.1}$$

where Q is an m×m symmetric, positive definite matrix and each C_k is a closed convex set in $\Re^m$. For simplicity, we will assume that Q is the identity matrix in what follows. The general case can be treated by making the transformation $y = Q^{1/2}x$.

We can write (5.1) in the following equivalent form:

$$\text{Minimize} \quad f(x_0,x_1,\ldots,x_K) = h(x_0) + \textstyle\sum_k \delta(x_k|C_k) \tag{5.2}$$
$$\text{subject to} \quad x_0 - x_k = 0, \quad k = 1,\ldots K,$$

where $\delta(\cdot|C_k)$ is the indicator function for C_k. We make the following assumption regarding (5.1):

Assumption D: There exists $k\in\{0,\ldots,K\}$ such that $C_1\cap\ldots\cap C_k$ is a polyhedral set, $C_{k+1}\cap\ldots\cap C_K$ is a closed convex set, and $C_1\cap\ldots\cap C_k\cap ri(C_{k+1})\cap\ldots\cap ri(C_K)$ is nonempty.

Note that (5.2) is a special case of (4.1). Moreover, Assumption D implies that Assumptions A and B hold for this problem.

By assigning a Lagrange multiplier vector p_k to the constraints $x_0-x_k = 0$, we obtain the following dual program of (5.2) (cf. (4.4) and (4.5)):

$$\text{Maximize} \quad q(p_1,\ldots,p_K)$$
$$\text{subject to} \quad p_k\in \Re^m,\ k = 1,\ldots,K,$$

where we denote

$$q(p_1,\ldots,p_K) = \min\{\ h(x) + \textstyle\sum_k \delta(p_k\ |C_k) + \sum_k \langle p_k,x-x_k\rangle\ \}$$
$$= \|d\|^2/2 - h(\textstyle\sum_k p_k) - \sum_k \delta^*(p_k|C_k), \tag{5.3}$$

and $\delta^*(\cdot|C_k)$ denotes the conjugate function of $\delta(\cdot|C_k)$, i.e., $\delta^*(p_k\ |C_k) = \sup\{\ \langle p_k,x\rangle \mid x\in C_k\ \}$.

Consider the following iterative algorithm for solving (5.2):

$$y_k^t = \text{argmax}_\Delta\ q(y_1^t,\ldots,y_{k-1}^t,\Delta,\ y_{k+1}^{t-1},\ldots,y_K^{t-1}),\ \ k = 1,\ldots,K, \tag{5.4a}$$
$$y_1^0 = \ldots = y_K^0 = 0. \tag{5.4b}$$

The algorithm (5.4a)-(5.4b) is clearly a special case of the BCR algorithm using cyclic order of relaxation (and with zero starting multiplier vector). Let $\{(x_1^t,\ldots,x_K^t)\}$ be the sequence of primal vectors associated with $\{(y_1^t,\ldots, y_K^t)\}$ given by:

$$x_K^0 = d, \tag{5.5a}$$

$$x_0^t = x_K^{t-1},\ \ x_{k-1}^t - x_k^t = y_k^t - y_k^{t-1},\ \ k = 1,\ldots,K,\ t \geq 1. \tag{5.5b}$$

We have the following main result of this section:

Proposition 3 The sequences $\{(y_1^t,\ldots,y_K^t)\}$ and $\{(x_1^t,\ldots,x_K^t)\}$ generated by (5.4a)-(5.4b) and (5.5a)-(5.5b) are identical to those generated by Han's algorithm.

Proof: Han's algorithm is given by (5.4b), (5.5a)-(5.5b), and

$$x_k^t = \operatorname{argmin}\{\ \|x_{k-1}^t + y_k^{t-1} - x\|^2 \mid x \in C_k\ \},\ \ k = 1,\ldots,K,\ t \geq 1. \tag{5.6}$$

Hence it suffices to prove (5.6).

Consider any $k \in \{1,\ldots,K\}$ and any positive integer t. It can be seen from (5.4b) and (5.5a)-(5.5b) that

$$y_1^t + \ldots + y_{k-1}^t + y_k^{t-1} + \ldots + y_K^{t-1} = d - x_{k-1}^t. \tag{5.7}$$

Hence (5.3) and (5.4b) imply that

$$\begin{aligned} y_k^t &= \operatorname{argmax}_\Delta \{-h(y_1^t + \ldots + y_{k-1}^t + \Delta + y_{k+1}^{t-1} + \ldots + y_K^{t-1}) - \delta^*(\Delta \mid C_k)\} \\ &= \operatorname{argmax}_\Delta \{-h(d - x_{k-1}^t - y_k^{t-1} + \Delta) - \delta^*(\Delta \mid C_k)\} \\ &= \operatorname{argmax}_\Delta \{-\|\Delta - x_{k-1}^t - y_k^{t-1}\|^2/2 - \delta^*(\Delta \mid C_k)\}. \end{aligned} \tag{5.8}$$

The optimality condition for the right hand side of (5.8) gives

$$-y_k^t + x_{k-1}^t + y_k^{t-1} \in \partial\delta^*(y_k^t \mid C_k)$$

and hence (using Theorem 23.5 in [22])

$$-y_k^t+x_{k-1}^t+y_k^{t-1} = \text{argmin}\{ -\langle y_k^t,x\rangle \mid x\in C_k\}$$

$$= \text{argmin}\{ -\langle y_k^t,x\rangle + \|-y_k^t+x_{k-1}^t+y_k^{t-1}-x\|^2/2 \mid x\in C_k\}$$

$$= \text{argmin}\{ \|x_{k-1}^t+y_k^{t-1}-x\|^2/2 \mid x\in C_k \},$$

where the second equality follows from the fact that, for any function $g:\Re^m\to(-\infty,\infty]$, if $z' = \text{argmin}_z\, g(z)$, then $z' = \text{argmin}_z \{g(z) + \|z'-z\|^2/2\}$. This, together with (5.5b), proves (5.6). Q.E.D.

Corollary 3 $\{x_k^t\}$ converges to the optimal solution of (5.1) and $\{\sum_k y_k^t\}$ also converges.

Proof: From (5.7) we have that $x_k^t = \nabla h^*(y_1^t+\ldots+y_k^t+y_{k+1}^{t-1}+\ldots+y_K^{t-1})$. Since $(y_1^t,\ldots,y_k^t,y_{k+1}^{t-1},\ldots,y_K^{t-1})$ is the Lagrange multiplier vector generated by the BCR algorithm at the $(t-1)*K+k$-th iteration, under cyclic order of relaxation (cf. (5.4a)-(5.4b)), it follows that x_k^t is the corresponding primal vector given by (4.7). Hence, by Proposition 2, $\{x_k^t\}$ converges to the optimal solution of (5.1) for all k. By (5.7), the sequence $\{\sum_k y_k^t\}$ also converges. Q.E.D.

Corollary 4 extends the results given in Theorems 4.8 and 4.10 of [11], which in addition assume that a certain (relative) interior intersection condition holds. Moreover, from Proposition 2 we see that neither cyclic relaxation nor zero starting multiplier vector is necessary for convergence. Also the nonempty interior assumptions made in Lemma 4.9 and Theorem 4.10 of [11] ensure that the dual functional has bounded level sets, in which case convergence of the sequences $\{y_k^t\}$ follows from Proposition 2.

6. The Method of Multipliers is a Special Case

Consider the following convex program

$$\begin{array}{ll} \text{Minimize} & h(x) \\ \text{subject to} & Bx = b, \end{array} \tag{6.1}$$

where $h:\Re^m\to(-\infty,+\infty]$, B is an n×m real matrix, and b is an n-vector. We make the following assumption regarding (6.1):

Assumption E:

(a) h is convex, lower semicontinuous and continuous within dom(h). The conjugate function of h, denoted by h^*, is real valued.

(b) Dom(h) = $S' \cap S''$, where S'' is a convex set and cl(S') is a polyhedral set. Moreover, there exists $x \in S' \cap \mathrm{ri}(S'')$ satisfying $Bx = b$.

By introducing the auxiliary variable x_0, we can write (6.1) as the following problem

$$\begin{array}{ll} \text{Minimize} & c\|x_0 - b\|^2/2 + h(x_1) \\ \text{subject to} & x_0 = b, \;\; x_0 = Bx_1, \end{array} \tag{6.2}$$

where c is any positive scalar. This problem can be seen to be a special case of (4.1) with $f_0(x_0) = c\|x_0 - b\|^2/2$ and $f_1 = h$. Furthermore, Assumption E implies that Assumptions A and B hold for this problem. The dual functional associated with this problem (cf. (4.5)) is

$$q(p_1,p_2) = \langle b,p_1\rangle - h^*(B_1{}^T p_1) - \|p_2 - p_1\|^2/2c, \tag{6.3}$$

where p_2 and p_1 are the Lagrange multiplier vectors assigned to, respectively, the constraints $x_0 = b$ and $x_0 = Bx_1$.

Consider the following algorithm for maximizing q, whereby a maximization with respect to p_2 alternates with a maximization with respect to p_1:

$$p_2{}^{r+1} \;=\; \mathrm{argmax}_\Delta\, q(\Delta, p_1{}^r). \tag{6.4a}$$

$$p_1{}^{r+1} \;=\; \mathrm{argmax}_\Delta\, q(p_2{}^{r+1}, \Delta), \tag{6.4b}$$

This algorithm is well defined and is a special case of the BCR algorithm under the cyclic rule of relaxation. By using the special structure of q (cf. (6.3)), we see from (6.4a) that $p_2{}^{r+1} = p_1{}^r$. Hence $p_1{}^{r+1}$ given by (6.4b) is a Kuhn-Tucker vector of the problem (cf. (6.2) and (4.6))

$$\begin{array}{ll} \text{Minimize} & c\|x_0 - b\|^2/2 + h(x_1) - \langle p_1{}^r, x_0\rangle \\ \text{subject to} & x_0 = Bx_1. \end{array} \tag{6.5}$$

Let $(x_0{}^r, x_1{}^r)$ be any optimal solution of the above problem. We then obtain from the Kuhn-Tucker conditions for (6.5) that

$$x_1{}^r = \operatorname{argmin}_{x_1} \{ c\|Bx_1 - b\|^2/2 + h(x_1) - \langle p_1{}^r, Bx_1 \rangle \},$$

$$p_1{}^{r+1} = p_1{}^r + c(b - Bx_1{}^r),$$

which we recognize to be the method of multipliers [10, 12, 20] (also see [2], [3, §3.4.4], [23]). By Proposition 2, the sequence $\{x_1{}^r\}$ is bounded and any of its limit points is an optimal solution of (6.1). [However, by exploiting the special form that q possesses, sharper results on the convergence of the method of multipliers can be obtained [23].] It has been shown [23] that the method of multipliers is the proximal minimization algorithm applied to maximize the dual functional $\langle b,p\rangle - h^*(B^Tp)$. Our results from this section and §3 provide an alternative view: these methods are all coordinate ascent methods, differing in only whether each is applied in the primal space or in the dual space.

References

[1] Auslender, A., *Optimisation Méthodes Numériques*, Masson, Paris (1976).

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

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

[4] Bertsekas, D.P., Hosein, P.A. and Tseng, P., "Relaxation Methods for Network Flow Problems with Convex Arc Costs," *SIAM J. Control and Optimization*, 25 (1987), 1219-1243.

[5] Bregman, L.M., "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.

[6] Censor, Y. and Lent, A., "An Iterative Row-Action Method for Interval Convex Programming," *J. Optimization Theory and Application*, 34 (1981), 321-352.

[7] Cottle, R.W., Duvall, S.G. and Zikan, K., "A Lagrangean Relaxation Algorithm for the Constrained Matrix Problem," *Naval Research Logistics Quarterly*, 33 (1986), 55-76.

[8] D'Esopo, D.A., "A Convex Programming Procedure," *Naval Research Logistics Quarterly*, 6 (1959), 33-42.

[9] Gordon, R., Bender, R. and Herman, G.T., "Algebraic Reconstruction Techniques (ART) for Three-Dimensional Electron Microscopy and X-ray Photography, *J. Theoret. Biol.*, 29 (1970), 471-481.

[10] Haarhoff, P. C. and Buys, J. D., "A New Method for the Optimization of a Nonlinear Function Subject to Nonlinear Constraints," *The Computer Journal*, 13 (1970), 178-184.

[11] Han, S.P., "A Successive Projection Method," *Mathematical Programming*, 40 (1988), 1-14.

[12] Hestenes, M. R., "Multiplier and Gradient Methods," *J. Optimization Theory and Applications*, 4 (1969), 303-320.

[13] Lamond, B. and Stewart, N.F., "Bregman's Balancing Method," *Transportation Research-B*, 15B

(1981), 239-248.

[14] Lent, A., "A Convergent Algorithm for Maximum Entropy Image Restoration with a Medical X-ray Application," in *Image Analysis and Evaluation*, (R. Shaw, Ed.) Society of Photographic Scientists and Engineers (SPSE), Washington, D.C. (1977), 249-257.

[15] Lin, Y.Y. and Pang, J.S., "Iterative Methods for Large Convex Quadratic Programs: A Survey," *SIAM J. Control and Optimization*, 25 (1987), 383-411.

[16] Luenberger, D.G., *Linear and Nonlinear Programming*, Addison-Wesley, MA (1973).

[17] Ohuchi, A. and Kaji, I., "Lagrangian Dual Coordinatewise Maximization Algorithm for Network Transportation Problems with Quadratic Costs," *Networks*, 14 (1984), 515-530.

[18] Pang, J.S., "On the Convergence of Dual Ascent Methods for Large-Scale Linearly Constrained Optimization Problems," Unpublished manuscript, The University of Texas at Dallas (1984).

[19] Polak, E., *Computational Methods in Optimization: A Unified Approach*, Academic Press, N.Y. (1971).

[20] Powell, M. J. D., "A Method for Nonlinear Constraints in Minimization Problems," In *Optimization* (R. Fletcher, ed.), Academic Press, New York, NY (1969), 283-298.

[21] Powell, M.J.D., "On Search Directions for Minimization Algorithms," *Mathematical Programming*, 4 (1973), 193-201.

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

[23] Rockafellar, R. T., "Augmented Lagrangians and Applications of the Proximal Point Algorithm in Convex Programming," *Mathematics of Operations Research*, 1 (1976), 97-116.

[24] Sargent, R.W.H. and Sebastian, D.J., "On the Convergence of Sequential Minimization Algorithms," *J. of Optimization Theory and Applications*, 12 (1973), 567-575.

[25] Stern, T.E., "A Class of Decentralized Routing Algorithms Using Relaxation," *IEEE Transactions on Communications*, COM-25 (1977), 1092-1102.

[26] Tseng, P. and Bertsekas, D.P., "Relaxation Methods for Problems with Strictly Convex Separable Costs and Linear Constraints," *Mathematical Programming*, 38 (1987), 303-321.

[27] Tseng, P. and Bertsekas, D.P., "Relaxation Methods for Problems with Strictly Convex Costs and Linear Inequality Constraints," LIDS Report P-1717, M.I.T., Cambridge, MA (1987).

[28] Tseng, P., "Dual Coordinate Ascent for Problems with Strictly Convex Costs and Linear Constraints: a Unified Approach," LIDS Report P-1792, M.I.T., Cambridge, MA (August 1988).

[29] Zangwill, W.I., *Nonlinear Programming*, Prentice-Hall, N.J. (1969).

[30] Zenios, S.A. and Mulvey, J.M., "Relaxation Techniques for Strictly Convex Network Problems," *Annals of Operations Research 5: Algorithms and Software for Optimization* (Monma, C.L. ed.), Scientific Publishing Company, Switzerland (1986).

[31] Zenios, S.A. and Lasken, R.A., "Nonlinear Network Optimization on a Massively Parallel Connection Machine," *Annals of Operations Research 14* (1988), 147-165.

[32] Ortega, J.M. and Rheinboldt, W.C., *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, NY (1970).

[33] Arimoto, S., "An Algorithm for Computing the Capacity of Arbitrary DMCs," *IEEE Trans. on Information Theory*, IT-18 (1972), 14-20.

[34] Blahut, R., "Computation of Channel Capacity and Rate Distortion Functions," *IEEE Trans. on Information Theory*, IT-18 (1972), 460-473.

Appendix A

In this appendix, we prove Proposition 2. A fact that will be frequently used is that (cf. [22, Theorem 23.5]) $\nabla g_0(\sum_k A_k^T p_k)$ is the unique vector $x_0 \in \Re^{m_0}$ satisfying

$$\sum_k A_k^T p_k \in \partial f_0(x_0), \tag{A.1}$$

or equivalently

$$\nabla g_0(\sum_k A_k^T p_k) = \operatorname{argmax}_{x_0} \{ \sum_k \langle p_k, A_k x_0 \rangle - f_0(x_0) \}. \tag{A.2}$$

First note from (A.1) and (4.7) that

$$\sum_k A_k^T p_k^r \in \partial f_0(x_0^r), \qquad \forall\, r \geq 0, \tag{A.3}$$

Also, note from (4.7) and (4.8) that, for any s and any r such that $s^r = s$, (x_0^r, x_s^r) is the unique optimal solution of (4.6) with $p_k = p_k^r$ for all $k \neq s$. Since $p_k^r = p_k^{r-1}$ for all $k \neq s$, p_s^r is an optimal dual solution for the same problem. Hence, from the Kuhn-Tucker conditions for (4.6) (and using the fact that $p_k^r = p_k^{r-1}$, $x_k^r = x_k^{r-1}$ for all $k \neq s$, $k \neq 0$) we obtain that

$$B_k^T p_k^r \in \partial f_k(x_k^r), \qquad \forall\, r \geq 0,\ \forall\, k \neq 0. \tag{A.4}$$

To simplify the presentation, let $\mathcal{K} = \{ k \in \{0,1,\ldots,K\} \mid g_k \text{ is real valued} \}$. Throughout the proof, f will denote the function given by (4.3). Our argument will follow closely that given in [27, §3] and in [28, §3] (in fact, to simplify the presentation, we will borrow some results from [27]).

We precede our proof of convergence with the following four technical lemmas:

Lemma 1 For r = 1, 2, ...,

$$q(p^{r+1}) - q(p^r) \geq f_0(x_0^{r+1}) - f_0(x_0^r) - f_0'(x_0^r; x_0^{r+1} - x_0^r), \tag{A.5}$$

$$f(x^*)-q(p^r) \;\ge\; \textstyle\sum_{k\in\mathcal{K}} \left(f_k(x_k^*) - f_k(x_k^r) - f_k'(x_k^r;x_k^*-x_k^r)\right), \tag{A.6}$$

where $x^* = (x_0^*,\ldots,x_K^*)$ denotes any optimal solution of (4.1).

Proof: We first prove (A.5). From (4.5), (A.2), (4.7) and (A.4) we have

$$q(p^r) \;=\; \textstyle\sum_k f_k(x_k^r) \;+\; \sum_{k\neq 0} \langle p_k^r, b_k - A_k x_0^r - B_k x_k^r\rangle,$$

and hence

$$\begin{aligned}
q(p^{r+1})-q(p^r) \;&=\; \textstyle\sum_k f_k(x_k^{r+1}) \;+\; \sum_{k\neq 0} \langle p_k^{r+1}, b_k - A_k x_0^{r+1} - B_k x_k^{r+1}\rangle \\
&\quad - \textstyle\sum_k f_k(x_k^r) - \sum_{k\neq 0} \langle p_k^r, b_k - A_k x_0^r - B_k x_k^r\rangle \\
&= f_0(x_0^{r+1}) - f_0(x_0^r) - \langle \textstyle\sum_{k\neq 0} A_k^T p_k^r, x_0^{r+1} - x_0^r\rangle \\
&\quad + \textstyle\sum_{k\neq 0} [f_k(x_k^{r+1}) - f_k(x_k^r) - \langle B_k^T p_k^r, x_k^{r+1} - x_k^r\rangle \\
&\qquad\qquad + \langle p_k^{r+1} - p_k^r, b_k - A_k x_0^{r+1} - B_k x_k^{r+1}\rangle].
\end{aligned}$$

Since (cf. (4.8)), for all $k \neq 0$, $\langle p_k^{r+1} - p_k^r, b_k - A_k x_0^{r+1} - B_k x_k^{r+1}\rangle = 0$, this implies that

$$\begin{aligned}
q(p^{r+1})-q(p^r) \;&\ge\; f_0(x_0^{r+1}) - f_0(x_0^r) - \langle \textstyle\sum_{k\neq 0} A_k^T p_k^r, x_0^{r+1} - x_0^r\rangle \\
&\quad + \textstyle\sum_{k\neq 0} [f_k(x_k^{r+1}) - f_k(x_k^r) - \langle B_k^T p_k^r, x_k^{r+1} - x_k^r\rangle].
\end{aligned} \tag{A.7}$$

For any $k \neq 0$, since (cf. (A.4)) $B_k^T p_k^r \in \partial f_k(x_k^r)$ and f_k is convex, we have

$$f_k(x_k^{r+1}) - f_k(x_k^r) - \langle B_k^T p_k^r, x_k^{r+1} - x_k^r\rangle \ge 0,$$

which, together with (A.7), (1.2) and the fact (cf. (A.3)) $\sum_{k\neq 0} A_k^T p_k^r \in \partial f_0(x_0^r)$, proves (A.5).

The proof of (A.6) is similar to that for (A.5). Fix any $r \ge 0$. Since x^* is feasible for (4.1), we have

$$f(x^*)-q(p^r) = f(x^*) - q(p^r) + \sum_{k\neq 0} \langle p_k{}^r, b_k - A_k x_0{}^* - B_k x_k{}^* \rangle$$

$$= f_0(x_0{}^*) - f_0(x_0{}^r) - \langle \sum_{k\neq 0} A_k{}^T p_k{}^r, x_0{}^* - x_0{}^r \rangle$$

$$+ \sum_{k\neq 0} [f_k(x_k{}^*) - f_k(x_k{}^r) - \langle B_k{}^T p_k{}^r, x_k{}^* - x_k{}^r \rangle]. \qquad \text{(A.8)}$$

Again, for any $k \notin \mathcal{K}$, since $B_k{}^T p_k{}^r \in \partial f_k(x_k{}^r)$ and f_k is convex, we have

$$f_k(x_k{}^*) - f_k(x_k{}^r) - \langle B_k{}^T p_k{}^r, x_k{}^* - x_k{}^r \rangle \geq 0,$$

which, together with (A.8), (1.2) and the fact $\sum_{k\neq 0} A_k{}^T p_k{}^r \in \partial f_0(x_0{}^r)$, proves (A.6). Q.E.D.

Let $S_k = \mathrm{dom}(f_k)$, $k = 0,\ldots,K$. Lemma 1 implies the following facts, whose proof is identical to that for Lemmas 2 and 3 in [27]:

Lemma 2

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

(b) Let $h:\Re^m \rightarrow (-\infty,+\infty]$ $(m \geq 1)$ be any convex, lower semicontinuous function that is continuous in its effective domain S. For any $y \in S$, any $z \in \Re^m$ such that $y+z \in S$, and any sequences $\{y^t\} \rightarrow y$ and $\{z^t\} \rightarrow z$ such that $y^t \in S$ and $y^t+z^t \in S$ for all t, $\lim_{t\rightarrow+\infty}\sup\{h'(y^t;z^t)\} \leq h'(y;z)$.

Lemma 2 in turn implies the following two lemmas:

Lemma 3 $x_0{}^{r+1} - x_0{}^r \rightarrow 0$.

Proof: Since (cf. Lemma 2 (a)) $\{x_0{}^r\}$ is bounded, if (a) does not hold, then there exists subsequence R for which $\{x_0{}^r\}_{r\in R}$ converges to some point $x_0{}'$ and $\{x_0{}^{r+1}\}_{r\in R}$ converges to some point $x_0{}'' \neq x_0{}'$. Let $z = x_0{}'' - x_0{}'$ $(z \neq 0)$. By Lemma 2 (a), both $x_0{}'$ and $x_0{}'+z$ are in S_0. Then using (A.5), the continuity of f_0 on S_0, and Lemma 2 (b), we obtain

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

Since $q(p^r)$ is nondecreasing with r and f_0 is strictly convex (so the right hand side of above is a positive scalar), it follows that $q(p^r) \to +\infty$. This, in view of the weak duality condition

$$\max\{\, q(p_1,\ldots,p_K) \mid p_k\in \Re^{n_k}, k = 1,\ldots,K \,\} \;\leq\; \min\{\, f(x) \mid x\in X \,\},$$

contradicts the feasibility of (4.1) (cf. Assumption A (c)). Q.E.D.

Lemma 4 Under the Essentially Cyclic rule, $\{(x_0^r,\ldots,x_K^r)\}$ is bounded and its limit points are in $(S_0\times\ldots\times S_K)\cap X$.

Proof: For each $s\in\{1,\ldots,K\}$, we have (cf. (4.8)) that

$$A_sx_0^r + B_sx_s^r \;=\; b_s, \qquad \forall\, r \geq 1 \text{ such that } s = s^r, \tag{A.9}$$

$$x_s^r \;=\; x_s^{r-1}, \qquad \forall\, r \geq 1 \text{ such that } s \neq s^r. \tag{A.10}$$

Since (cf. Lemma 2 (a)) $\{x_s^r\}$ is bounded for all $s\in\mathcal{K}$ and B_s has rank m_s for all $s\notin\mathcal{K}$, this implies that $\{x_s^r\}$ is bounded for all s. Fix any $s\in\{1,\ldots,K\}$ and, for each $r \geq T$, let $\tau(r)$ be the largest integer h not exceeding r such that $s^h = s$. Then (cf. (A.9) and (A.10))

$$A_sx_0^r + B_sx_s^r \;=\; \sum_{h=\tau(r)}^{r-1} A_s(x_0^{h+1}-x_0^h) + b_s,\ \forall\, r \geq T.$$

Since (cf. Essentially Cyclic rule) $r - \tau(r) \leq T$ for all $r \geq T$, this, together with Lemma 3 (a), implies that every limit point $(x_0',\ldots,x_K')$ of $\{(x_0^r,\ldots,x_K^r)\}$ satisfies

$$A_sx_0' + B_sx_s' \;=\; b_s.$$

Since the choice of $s\in\{1,\ldots,K\}$ was arbitrary, this implies that $(x_0',\ldots,x_K')\in X$. By Lemma 2 (a), $x_s'\in S_s$

for all $s \in \mathcal{K}$. For each $s \notin \mathcal{K}$, since $x_s^r \in S_s$ for all r and S_s is closed, $x_s' \in S_s$. Q.E.D.

Lemmas 2 and 4 allow us to prove Proposition 2. We first prove part (a). To simplify the presentation, let $x^r = (x_0^r, \ldots, x_K^r)$, $S' = S_0' \times \ldots \times S_K'$, $S'' = S_0'' \times \ldots \times S_K''$, and $S = S' \cap S''$. Also let E denote the $n_1 + \ldots + n_K$ by $m_0 + \ldots + m_K$ matrix

$$E = \begin{bmatrix} A_1 & B_1 & 0 & \cdots & 0 \\ A_2 & 0 & B_2 & \cdots & 0 \\ & & & \vdots & \\ A_K & 0 & 0 & \cdots & B_K \end{bmatrix}.$$

Consider any convergent subsequence $\{x^r\}_{r \in R}$ converging to some vector x'. By Lemma 4, $x' \in S \cap X$. Let $x^* = (x_0^*, \ldots, x_K^*)$ denote any optimal solution of (4.1) and suppose that $x' \neq x^*$ (otherwise (a) follows trivially).

Let y be any element of $S' \cap \mathrm{ri}(S'') \cap X$ (y exists by Assumptions A (c)). Fix any $\lambda \in (0,1)$ and denote $y(\lambda) = \lambda y + (1-\lambda)x^*$. Then $y(\lambda) \in S' \cap \mathrm{ri}(S'') \cap X$. It can be shown (see proof of Proposition 1 (b) in [27]) that there exists an $\varepsilon > 0$ such that $\{ x \in S' \mid \|x - x'\| \leq \varepsilon \}$ is closed. Since cl(S') is a polyhedral set and $y(\lambda) - x'$ belongs to the tangent cone of S' at x', this implies that there exists $\delta \in (0,1)$ such that, for all $r \in R$ sufficiently large,

$$x^r + \delta z \in S', \tag{A.11}$$

where $z = y(\lambda) - x'$. On the other hand, since $y(\lambda) \in \mathrm{ri}(S'')$, $x^r \in S''$ for all r, and $\{x^r\}_{r \in R} \to x'$, we have that, for all $r \in R$ sufficiently large,

$$x^r + \delta z \in S''. \tag{A.12}$$

Since $y(\lambda) \in X$ and $x' \in X$, $Ez = 0$; hence

$$\langle p^r, Ez\rangle = 0, \; \forall \; r \in R. \tag{A.13}$$

Since (cf. (A.3), (A.4)) $E^T p^r \in \partial f(x^r)$ for all r, (1.2) implies that $f'(x^r;z) \geq \langle p^r, Ez\rangle$ for all r. Since $x' + \delta z \in S$, this, together with (A.11)-(A.13) and Lemma 2 (b), implies that

$$f'(x';z) \geq 0.$$

Hence $f(x') \leq f(y(\lambda))$. Since the choice of $\lambda \in (0,1)$ was arbitrary, by taking λ arbitrarily small (and using the continuity of f within S), we obtain that $f(x') \leq f(x^*)$. Since $x' \in S \cap X$, x' is an optimal solution of (4.1). This completes the proof of part (a). The proof of parts (b) and (c) is analogous to that of Proposition 1 in [27] and is omitted. Q.E.D.