LIDS-P-1717 November, 1987

Relaxation Methods for Problems withStrictly Convex Costs and Linear Inequality Constraints*

by

Paul Tseng†

and Dimitri P. Bertsekas†

Abstract

In this paper we consider the problem of minimizing a strictly convex, possibly nondifferentiable cost subject to linear inequality constraints. We consider a dual coordinate ascent method that uses inexact line search and essentially cyclic order of coordinate relaxation. We show, under very weak conditions, that this method generates a sequence of primal vectors converging to the optimal primal solution. Under an additional regularity assumption, and assuming the domain of the cost function is polyhedral, we show that a related sequence of dual vectors converges in cost to the optimal cost. Alternately, we show that if the constraint set has an interior point in the domain of the cost function then this sequence of dual vectors is bounded and each of its limit point(s) is an optimal dual solution. We also show, for the special case where the cost function is strongly convex, that the order in which the coordinates are relaxed may become gradually asynchronous. These results extend those in [9] for separable cost functions but the extension is nontrivial.

Key words: convex program, dual functional, coordinate ascent.

*Work supported by the National Science Foundation under grant NSF-ECS-8519058 and by the Army Research Office under grant DAAL03-86-K-0171.

†The authors are with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139.

1. Introduction

Consider the following minimization problem

$$\text{Minimize} \quad f(x) \tag{P}$$

$$\text{subject to} \quad Ex \geq b, \tag{1}$$

where f: $\mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$, E is a given $n \times m$ real matrix and b is a vector in $\mathbb{R}^n$. In our notation all vectors are column vectors, and superscript T denotes transpose. We will denote the (i,j)th entry of E by e_{ij}, the effective domain of f by S, i.e.

$$S = \{ x \mid f(x) < +\infty \}$$

and the constraint set by Q, i.e. $Q = \{ x \mid Ex \geq b \}$. Note that S may be any convex set (not necessarily closed), so arbitrary convex inequality constraints can be embedded into f. We make the following standing assumptions:

Assumption A: f is strictly convex, lower semicontinuous and continuous within S. Furthermore the conjugate function of f defined by

$$g(t) = \sup\{ t^T\xi - f(\xi) \mid \xi \in \mathbb{R}^m \}. \tag{2}$$

is real valued, i.e. $-\infty < g(t) < +\infty$ for all $t \in \mathbb{R}^m$.

Assumption B: There exists at least one feasible solution for **(P)**, i.e. $S \cap Q \neq \emptyset$.

Assumption A implies (based on Corollary 13.3.1 of [7]) that for every t there is some x with $f(x) < +\infty$ attaining the supremum in (2), and furthermore $f(x) \to +\infty$ as $\|x\| \to +\infty$ ($\|\cdot\|$ denotes the L_2 norm). It follows that the cost function f has bounded level sets. Because f is lower semicontinuous the level sets are compact. This together with Assumption B and the strict convexity of f within S imply that there exists a unique optimal solution to (P) which we denote x^*.

A dual problem for (P), obtained by assigning Lagrange multiplier p_i to the ith constraint of $Ex \geq b$, is

$$\begin{array}{ll} \textit{Maximize} & q(p) \\ \textit{subject to} & p \geq 0, \end{array} \tag{D}$$

where

$$q(p) = \min\{ f(x) + p^T(b-Ex) \mid x \in \mathbb{R}^m \} = p^T b - g(E^T p). \tag{3}$$

(D) is a concave program with simple positive orthant constraints. Furthermore, strong duality holds for (P) and (D), i.e. the optimal value in (P) equals the optimal value in (D). To see this we denote by F the convex bifunction associated with (P) ([7], pp. 293), i.e.

$$(Fu)(x) = \begin{cases} f(x) & \textit{if } b-Ex \leq u, \\ +\infty & \textit{otherwise}. \end{cases}$$

According to Theorem 30.4 (i) in [7], (P) and (D) have the same optimal value if F is closed and convex and (P) has a unique optimal solution. Thus it suffices to show that F is closed, or equivalently, that the set

$$\{ (x,u,z) \mid (Fu)(x) \leq z \} = \{ (x,u,z) \mid f(x) \leq z \textit{ and } b-Ex \leq u \}$$

is closed. This set is the intersection of $\{ (x,u,z) \mid f(x) \leq z \}$ with $\{ (x,u,z) \mid b-Ex \leq u \}$, each of which is closed, so it is also closed ($\{ (x,u,z) \mid f(x) \leq z \}$ is closed because f is lower semicontinuous).

Since g is real valued and f is strictly convex, g and q are continuously differentiable (cf. [7] Theorem 26.3). Using the chain rule, we obtain the gradient of q at p to be

$$\nabla q(p) = b - Ex,$$

where

$$x = \nabla g(E^T p) = \arg\sup \{ p^T E \xi - f(\xi) \mid \xi \in \mathbb{R}^m \}. \tag{4a}$$

Note from (4a) that x is also the unique vector satisfying

$$E^T p \in \partial f(x), \tag{4b}$$

where $\partial f(x)$ is the subdifferential of f at x. For convenience we will denote $\nabla q(p)$ by d(p) and its ith coordinate by $d_i(p)$. It can be seen that $d_i(p)$ is continuous, real valued and nonincreasing in p_i.

Differentiability of q motivates a coordinate ascent method for solving (D) whereby, given a dual vector p, a coordinate p_i such that $\partial q(p)/\partial p_i > 0$ (<0) is chosen and p_i is increased (decreased) in order to increase the dual functional. An important advantage of such a coordinate relaxation method is its suitability for parallel implementation on problems where f and E have special structures. As an example, suppose E is sparse and f is quadratic of the form $x^TQx + r^Tx$, where r is a vector in $\Re^m$ and Q is a $m \times m$ sparse, symmetric, positive definite matrix, then two coordinates p_i and p_j are uncoupled, and can be iterated upon (relaxed) simultaneously if the (i,j)th entry of $EQ^{-1}E^T$ is zero (another example is if f is separable and the (i,j)th entry of EE^T is zero).

Convergence of coordinate ascent methods for maximizing general concave functions are well studied (see [3], [4], [6], [8], [13], [14]) but convergence typically requires compactness of the level sets and some form of strict concavity of the objective function, neither of which necessarilly holds for q(p). We show in Section 3 that, under the standing assumptions, the coordinate ascent method applied to (D), and using inexact line search and essentially cyclic order of coordinate relaxation, generates a sequence of primal vectors converging to x^*. If S is a polyhedral set and (P) satisfies a "regular feasibility" condition, then a related sequence of dual costs converges to the optimal cost. Alternately, if S intersects the interior of Q then the method generates a bounded sequence of dual vectors, each of whose limit point(s) is an optimal dual solution. We also show that if f is strongly convex then increasing asynchrony in the order of relaxation is tolerable. These results represent significant improvements over previous works on convergence of dual ascent methods for (P). In [1], [2] and [5], f is required to be differentiable, line search is exact, and only convergence to the optimal primal solution are shown (no result on convergence to an optimal dual solution). Furthermore, additional restrictive assumptions such as "strong zone consistency" are required in [1] and [2], while in [5] f is required to be strongly convex and S is required to be a polyhedral set. Another paper [16] treats the more special case where f is quadratic and E is the node-arc incidence matrix for a given directed graph. A convergence result similar to ours is obtained in [9] for the special case where f is separable and the constraints are linear equalities, however the analysis used in [9] does not readily extend to our problem.

2. Relaxation Algorithm

Consider the following coordinate ascent algorithm for (D): We fix a scalar δ in the interval (0,1) and, beginning with a nonnegative p, we repeatedly apply the following iteration:

Relaxation Iteration

If $d_i(p) \leq 0$ and $p_i d_i(p) = 0$ for all i then STOP.

Else

Choose any coordinate p_s. Set $\beta = d_s(p)$.

If $\beta = 0$ or if $\beta < 0$ and $p_s = 0$ then do nothing.

If $\beta > 0$ then increase p_s until $0 \leq d_s(p) \leq \delta\beta$.

If $\beta < 0$ and $p_s > 0$ then decrease p_s until either $0 \geq d_s(p) \geq \delta\beta$, $p_s \geq 0$ or $d_s(p) < \delta\beta$, $p_s = 0$.

We first show that each relaxation iteration is well defined. Denote, for each p and $u \in \mathbb{R}^n$, the directional derivative of q at p in the direction u by $q'(p;u)$, and let e^s be the s-th coordinate vector in $\mathbb{R}^n$. The relaxation iteration is not well defined if for some p and some s we have $d_s(p) > 0$ and there does not exist a $\Delta \geq 0$ such that $d_s(p + \Delta e^s) \leq \delta d_s(p)$. Then we must have

$$\lim_{\Delta \to \infty} \inf q'(p+\Delta e^s; e^s) \;\geq\; \delta d_s(p) \;>\; 0.$$

Therefore

$$\lim_{\Delta \to \infty} q(p+\Delta e^s) \;=\; +\infty,$$

which in view of the strong duality condition

$$max\{\, q(p) \mid p \geq 0 \,\} \;=\; min\{\, f(x) \mid Ex \geq b \,\}$$

contradicts feasibility of (P) [cf. Assumption B]. It is easily seen that if the iteration stops then $\nabla g(E^T p)$ and p satisfy the Kuhn-Tucker conditions, and are therefore optimal for (P) and (D) respectively.

We will consider the following assumption regarding the order in which the coordinates are chosen for relaxation:

Assumption C: There exists constant T for which every coordinate is chosen at least once for relaxation between iterations r and r + T, for r = 0, 1, 2, ...

Assumption C says that the coordinates must be relaxed in an essentially cyclic order. We will weaken this assumption later.

3. Convergence analysis

Let us denote the price vector generated by the relaxation algorithm at the rth iteration (r = 0,1,2,...) by p^r, and the index of the coordinate relaxed at the rth iteration by s^r. To simplify notation we also denote

$$t^r = E^T p^r, \qquad d^r = d(p^r) \qquad and \qquad x^r = \nabla g(E^T p^r).$$

For each $x \in S$ and $z \in \mathbb{R}^m$ we denote $f'(x;z)$ the directional derivative of f at x in the direction z. We will first show that under Assumption C the sequence $\{x^r\}$ converges to x^*. Because the proof is quite complex, we have broken it into two parts. The first part comprises of the following four technical lemmas, each of which holds independently of the order of relaxation (these lemmas will be used again later when we relax Assumption C). The first lemma lower bounds the amount of dual functional increase per iteration by the corresponding change in the primal vector.

Lemma 1

$$q(p^{r+1}) - q(p^r) \;\geq\; f(x^{r+1}) - f(x^r) - f'(x^r; x^{r+1} - x^r) \qquad r = 0,1,2,\ldots \tag{5}$$

and

$$f(x^*) - q(p^r) \;\geq\; f(x^*) - f(x^r) - f'(x^r; x^* - x^r) \qquad r = 0,1,2,\ldots \tag{6}$$

Proof : Consider a fixed index r and denote $s = s^r$, $\alpha = p_s^{r+1} - p_s^r$. From (2), (3) and (4) we have

$$q(p^r) \;=\; f(x^r) + (p^r)^T b - (t^r)^T x^r \qquad r = 0,1,2,\ldots$$

and therefore

$$q(p^{r+1}) - q(p^r) \;=\; (p^r)^T E x^r - (p^r)^T b - f(x^r) - (p^{r+1})^T E x^{r+1} + (p^{r+1})^T b + f(x^{r+1})$$

$$= (p^r)^T E x^r - f(x^r) - (p^r + \alpha e^s)^T E x^{r+1} + f(x^{r+1}) + (p^{r+1} - p^r)^T b$$

$$= (p^r)^T E (x^r - x^{r+1}) - \alpha (e^s)^T E x^{r+1} - f(x^r) + f(x^{r+1}) + \alpha (e^s)^T b$$

$$= f(x^{r+1}) - f(x^r) - (t^r)^T (x^{r+1} - x^r) + \alpha (e^s)^T (b - E x^{r+1})$$

$$= f(x^{r+1}) - f(x^r) - (t^r)^T (x^{r+1} - x^r) + |\alpha d_s^{r+1}|.$$

Thus we obtain

$$q(p^{r+1}) - q(p^r) \;\geq\; f(x^{r+1}) - f(x^r) - (t^r)^T (x^{r+1} - x^r).$$

Since [cf. (4b)] $t^r \in \partial f(x^r)$ and by definition

$$f'(x^r; x^{r+1} - x^r) \;=\; \sup \{\, \eta^T (x^{r+1} - x^r) \mid \eta \in \partial f(x^r) \,\}$$

we obtain

$$q(p^{r+1}) - q(p^r) \;\geq\; f(x^{r+1}) - f(x^r) - f'(x^r; x^{r+1} - x^r).$$

Similarly we have (using the fact $p^r \geq 0$ and x^* is feasible for (P))

$$f(x^*) - q(p^r) \;\geq\; f(x^*) - q(p^r) + (p^r)^T (b - E x^*)$$

$$= f(x^*) - f(x^r) - (t^r)^T (x^* - x^r) \;\geq\; f(x^*) - f(x^r) - f'(x^r; x^* - x^r).$$

Q.E.D.

Lemma 1 allows us to prove the next lemma:

Lemma 2 The sequences $\{x^r\}$ and $\{f(x^r)\}$ are bounded, and every limit point of $\{x^r\}$ is in S.

Proof: We first show that (6) implies that the sequence $\{x^r\}$ is bounded. Suppose $\{x^r\}$ is not bounded. Then there exists subsequence R for which $\{\|x^r\|\}_{r\in R} \to +\infty$ and $\|x^r-x^*\| \geq 1$ for all $r\in R$. Fix an arbitrary scalar $\beta\in(0,1)$ and let

$$\xi^r = x^* + \beta(x^r-x^*)/\|x^r-x^*\| \quad \forall\, r\in R.$$

The subsequence $\{\xi^r\}_{r\in R}$ has the property that

$$\xi^r\in S \quad and \quad \|\xi^r-x^*\| = \beta \quad \forall\, r\in R,$$

and for β sufficiently small $f(\xi^r)$ is bounded for all $r\in R$ (otherwise there exists subsequence R′ of R and $\{\beta^k\}_{k\in R'} \downarrow 0$ such that $\{f(x^* + \beta^k(x^k-x^*)/\|x^k-x^*\|)\}_{k\in R'} \to +\infty$, thus contradicting the continuity of f within S since $x^* + \beta^k(x^k-x^*)/\|x^k-x^*\|$ converges to x^* and is in S for all k such that $\beta^k \leq 1$). Using (6) and the fact that $q(p^r)$ is nondecreasing with r we obtain

$$f(x^r) \;\geq\; f(x^*) - f'(x^r;x^*-x^r) - \Delta \quad \forall\, r\in R,$$

where we define $\Delta = f(x^*) - q(p^0)$. Combining this with the convexity of f yields

$$f(\xi^r) \geq f(x^r) + f'(x^r;\xi^r-x^r) \geq f(x^*)-f'(x^r;x^*-x^r)-\Delta+f'(x^r;\xi^r-x^r) = f(x^*)- f'(x^r;x^*-\xi^r)-\Delta.$$

Therefore

$$(f(\xi^r) - f(x^*))/\beta + \Delta/\beta \;\geq\; -f'(x^r;(x^*-\xi^r)/\beta) \;=\; -f'(x^r;(x^*-x^r)/\|x^r-x^*\|) \quad \forall\, r\in R. \tag{7}$$

Since the left hand side of (7) has been shown to be bounded for sufficiently small $\beta>0$ it follows there exists constant $K > 0$ for which

$$K \;\geq\; -f'(x^r;(x^*-x^r)/\|x^r-x^*\|) \;\geq\; (f(x^r)-f(x^*))/\|x^r-x^*\| \quad \forall\, r\in R,$$

implying

$$K \;\geq\; (f(x^*+\lambda^r z^r)-f(x^*))/\lambda^r \quad \forall\, r\in R,$$

where we define $\lambda^r = \|x^r-x^*\|$ and $z^r = (x^r-x^*)/\|x^r-x^*\|$ for all $r\in R$. Passing into a subsequence if necessary we will assume $\{z^r\}_{r\in R} \to z$ for some z. Now consider a fixed $\lambda > 0$ and let $\zeta^r = x^* + \lambda z^r$

for all $r \in R$. Then for all $r \in R$ sufficiently large (so $\lambda^r > \lambda$) we have

$$\zeta^r \in S, \quad \frac{f(\zeta^r) - f(x^*)}{\lambda} = \frac{f(x^* + \lambda z^r) - f(x^*)}{\lambda} \leq \frac{f(x^* + \lambda^r z^r) - f(x^*)}{\lambda^r} \leq K,$$

and

$$\{\zeta^r\}_{r \in R} \to x^* + \lambda z.$$

This together with continuity of f on S imply

$$\frac{f(x^* + \lambda z) - f(x^*)}{\lambda} \leq K.$$

Our choice of $\lambda > 0$ was arbitrary, so the above inequality holds for all $\lambda > 0$. This then implies, in the terminology of [7], that f is not co-finite and therefore (cf. Corollary 13.3.1 in [7]) g is not real valued – a contradiction.

We next show that $\{f(x^r)\}$ is bounded. If $\{f(x^r)\}$ is not bounded then there exists subsequence R for which $\{f(x^r)\}_{r \in R} \to +\infty$. Since $\{x^r\}_{r \in R}$ is bounded we will assume, passing into a subsequence if necessary, that $\{x^r\}_{r \in R}$ converges to some x' which must be different from x^* (otherwise because $x^* \in S$ it must be $\{f(x^r)\}_{r \in R} \to f(x^*) < +\infty$). Fix a scalar $\beta > 0$ such that $\beta < \|x^r - x^*\|$ for all $r \in R$ and define ξ^r, $r \in R$, as above. For each $r \in R$ we have, using the convexity of f,

$$f(\xi^r) \geq f(x^r) + (1 - \beta/\|x^r - x^*\|) f'(x^r; x^* - x^r),$$

which together with (6) imply

$$f(\xi^r) \geq f(x^r) + (1 - \beta/\|x^r - x^*\|)(q(p^r) - f(x^r)) \geq \beta f(x^r)/\|x^r - x^*\| + (1 - \beta/\|x^r - x^*\|)q(p^0).$$

It follows $\{f(\xi^r)\}_{r \in R} \to +\infty$. This then again contradicts the fact that for β sufficiently small, $\{f(\xi^r)\}_{r \in R}$ is bounded. Finally, since $\{x^r\}$ and $\{f(x^r)\}$ are both bounded and the level sets of f are closed (since f is lower semicontinuous), it follows that every limit point of $\{x^r\}$ is in S. Q.E.D.

Lemma 3 For any $y \in S$, any z such that $y + z \in S$, and any sequences $\{y^k\} \to y$ and $\{z^k\} \to z$ such that $y^k \in S$ and $y^k + z^k \in S$ for all k, we have

$$\lim_{k \to \infty} sup \{ f'(y^k; z^k) \} \leq f'(y; z). \tag{8}$$

Proof: To prove (8) we use the proof idea for Theorem 24.5 in [7]. Given any $\mu > f'(y;z)$ ($f'(y;z) < +\infty$ since $f'(y;z) < f(y+z) - f(y)$) we have $y + \lambda z \in S$ for each $\lambda > 0$ sufficiently small and

$$\frac{f(y+\lambda z) - f(y)}{\lambda} < \mu. \tag{9}$$

Since $y^k + z^k \in S$ for all k, we can (by taking $\lambda \leq 1$) further assume $y^k + \lambda z^k \in S$ for all k. Since $f(y^k + \lambda z^k) \to f(y + \lambda z)$ and $f(y^k) \to f(y)$ by continuity of f on S, it follows from (9) that, for all k sufficiently large,

$$\frac{f(y^k + \lambda z^k) - f(y^k)}{\lambda} < \mu.$$

Since

$$f'(y^k;z^r) \leq \frac{f(y^k + \lambda z^k) - f(y^k)}{\lambda} \quad \forall k,$$

it follows

$$\lim_{k \to +\infty} sup\{f'(y^k;z^k)\} \leq \mu.$$

This holds for any $\mu > f'(y;z)$ so (8) follows. Q.E.D.

Lemma 4

$$\lim_{r \to +\infty} sup\{d_{s^r}{}^r\} \leq 0. \tag{12}$$

Proof: If (12) does not hold then there exist scalar $\varepsilon > 0$, coordinate p_s and subsequence R for which

$$d_s{}^r \geq \varepsilon \quad \forall r \in R.$$

Therefore we have

$$E(x^r - x^{r+1}) = d_s{}^r - d_s{}^{r+1} \geq (1-\delta)\varepsilon \quad \forall r \in R,$$

so that

$$\|x^r - x^{r+1}\| \geq (1-\delta)\varepsilon/\|E\| \quad \forall r \in R.$$

Since by Lemma 2 $\{x^r\}_{r \in R}$ and $\{x^{r+1}\}_{r \in R}$ are bounded, further passing into a subsequence if necessary we will assume that $\{x^r\}_{r\in R}$ converges to some point x' and $\{x^{r+1}-x^r\}_{r\in R}$ converges to some nonzero vector z. By Lemma 2, both x' and $x'+z$ are in S. Then using (5), continuity of f on S and (8) we obtain

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

Since $q(p^r)$ is nondecreasing with r and f is strictly convex (so the right hand side of above is a positive scalar) it follows that

$$q(p^r) \to +\infty,$$

contradicting the feasibility of (**P**). Q.E.D.

We also need the following preliminary result:

Lemma 5 Under Assumption C, if x' is any limit point of $\{x^r\}$ then $x'\in S\cap Q$ and there exists a subsequence R for which

$$\{x^r\}_{r\in R} \to x' \quad and \quad b_i - \sum_{j=1}^{m} e_{ij}x_j' < 0 \;\Rightarrow\; p_i^r = 0 \;\; \forall\, r\in R. \tag{13}$$

Proof: First we prove that $x'\in S\cap Q$. From the proof of Lemma 4 it can be seen that

$$x^r - x^{r+1} \to 0 \quad as\;\; r\to+\infty. \tag{14}$$

For a fixed coordinate p_i, if $i = s^r$ then for all $k \in \{r+1, \ldots, r+T\}$ we have

$$d_i^k = d_{s^r}^r + \sum_{h=r}^{k-1}\sum_{j=1}^{m} e_{ij}(x_j^{h+1}-x_j^h) \leq d_{s^r}^r + M\sum_{h=r}^{r+T-1}\|x^{h+1}-x^h\|,$$

where M is some constant depending on E only. Since T is a constant by Assumption C, from (12) and (14) we obtain

$$\lim_{k\to\infty} sup\{d_i^k\} \leq 0.$$

Since the choice of i was arbitrary the above inequality holds for all i. Thus every limit point of $\{x^r\}$ is in Q, which together with Lemma 2 imply that $x'\in S\cap Q$.

Now we prove (13). Let $d = b - Ex'$. By Lemma 4, we have $d_i \leq 0$ for all i. Let $I = \{ i \mid d_i < 0 \}$ and assume that I is nonempty (otherwise we are done). Since x' is a limit point of $\{x^r\}$ then there exists subsequence R' for which $\{x^r\}_{r \in R'} \to x'$, implying

$$\{d_i^r\}_{r \in R'} \to d_i, \quad \forall\, i \in I.$$

Further passing into a subsequence if necessary we will assume [cf. (14) so that $\{d_i^{k+1} - d_i^k\} \to 0$]

$$d_i^{r+l+1} < \delta d_i^{r+l} \quad \forall\, l = 0,1,\ldots,T-1, \forall\, r \in R', \forall\, i \in I. \tag{15}$$

Consider a fixed $r \in R'$ and any $i \in I$. If $s^k = i$ for some $k \in [r, r+T-1]$ then by (15) we have

$$d_i^{k+1} < \delta d_i^k ,$$

which together with the statement of the relaxation iteration implies

$$p_i^{k+1} = 0.$$

Since $p_i^{k+1} = p_i^k$ if $s^k \neq i$ for all k then necessarilly $p_i^{r+T} = 0$. By Assumption C, for each $i \in I$, $s^k = i$ for at least one $k \in [r, r+T-1]$ so it follows that

$$p_i^{r+T} = 0 \quad \forall\, i \in I.$$

Since the choice of $r \in R$ was arbitrary we can choose $R = \{ r + T \mid r \in R' \}$ (note that by (14) we have $\{x^r\}_{r \in R} \to x'$). Q.E.D.

The following is our main result for the strictly convex case.

<u>Proposition 1</u> If Assumption C is satisfied then the following hold:

(a) $\{x^r\} \to x^*$.

(b) If the closure of S 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 $x, y \in B \cap S$ then $\{q(p^r)\} \to f(x^*)$.

(c) If $\text{int}(Q) \cap S \neq \emptyset$ then $\{p^r\}$ is bounded and every one of its limit points is an optimal dual solution.

Proof: We first prove (a). Let x' be a limit point of $\{x^r\}$ and let R be a subsequence satisfying (13). Also denote $d = b-Ex'$ and $I = \{ i \mid d_i < 0 \}$. By Lemma 5, $x' \in S \cap Q$. Suppose $x' \neq x^*$ and let $z = x^*-x'$. Clearly $x' + z \in S$ and $\sum_j e_{ij}z_j \geq 0$ for all $i \in I$. The latter implies [cf. (13)]

$$(p^r)^T(Ez) = \sum_{i \in I} p_i^r \left(\sum_{j=1}^{m} e_{ij} z_j \right) \geq 0 \quad \forall\, r \in R. \tag{16}$$

Using this inquality and the fact $t^r \in \partial f(x^r)$, for all r, we have

$$f'(x^r;z) = max\{ \xi^T z \mid \xi \in \partial f(x^r) \} \geq (t^r)^T z = (p^r)^T(Ez) \geq 0 \quad \forall\, r \in R. \tag{17}$$

Equation (17) and Lemma 3 imply $f'(x';z) \geq 0$ from which $f(x') \leq f(x^*)$, contradicting the hypothesis $x' \neq x^*$ and showing that $x' = x^*$.

Now we prove (b). First we claim that there exists some positive ε for which $S \cap B(x^*,\varepsilon)$ is closed, where $B(x^*,\varepsilon)$ denotes the closed ball around x^* with radius ε. To see this let Y denote the (closure of S)\S. It then suffices to show that $Y \cap B(x^*,\varepsilon) = \emptyset$ for some positive ε. Suppose the contrary. Then there exists a sequence of points $\{y^k\}$ in Y converging to x^*. Consider a fixed k. Since $y^k \in Y$ there exists a sequence of points $\{w^i(k)\}$ in S converging to y^k. Since f is lower semicontinuous it must be that $f(w^i(k)) \to +\infty$. Therefore there exists integer m_k for which $\|w^{m_k}(k)-y^k\| < 1/k$ and $f(w^{m_k}(k)) > k$. Then $\{w^{m_k}(k)\}$ is a sequence of points in S converging to x^* for which $f(w^{m_k}(k)) \to +\infty$, contradicting the continuity of f on S since $f(x^*) < +\infty$.

Since, by assumption, the closure of S is a convex poyhedron and $\{x^r\} \to x^*$, there must exist some $\mu \in (0,\varepsilon]$ for which $x^r-x^* \in K \cap B(x^*,\mu)$ for all r sufficiently large, where K denotes the convex tangent cone of S at x^*. K is closed since S is closed locally near x^* and K is polyhedral since closure of S is polyhedral. Since $x^r-x^* \in K$ for all r sufficiently large, we can express x^r-x^* as a nonnegative combination of the generators of K ([7], §19), i.e.

$$x^r - x^* = \sum_{k \in I} \lambda_k^r w^k \quad \forall\, r \textit{ sufficiently large},$$

where the λ_k^r's are nonnegative scalars tending to zero and $\{w^k\}_{k \in I}$ is some subset of the generators of K. By passing into a subsequence R if necessary we can assume that $\{w^k\}_{k \in I}$ form a linearly independent set and, for each $k \in I$, $\lambda_k^r > 0$ for all $r \in R$. Let W denote the matrix whose columns are w^k, $k \in I$ and consider the substitution

$$x^r = W\lambda^r + x^*, \qquad \tau^r = W^T t^r, \qquad h(\lambda) = f(W\lambda + x^*).$$

Then h is strictly convex (since W has full column rank), h has as its tangent cone at 0 the positive orthant in $\mathbb{R}^{|I|}$, and for all $r \in R$

$$(t^r)^T(x^r - x^*) = (\tau^r)^T(\lambda^r - \lambda^*) = \sum_{k \in I} \tau_k^r \lambda_k^r \tag{18}$$

and (cf. [7], Theorem 23.9)

$$\tau^r \in \partial h(\lambda^r). \tag{19}$$

It follows from (19) that

$$-h'(\lambda^r; -e^k) \le \tau_k^r \le h'(\lambda^r; e^k) \qquad \forall\, k \in I, \tag{20}$$

where e^k denotes the kth coordinate vector in $\mathbb{R}^{|I|}$. Since $\lambda_k^r > 0$ for all $r \in R$, $k \in I$, and by assumption

$$-h'(\lambda^r; -e^k) \;\; and \;\; h'(\lambda^r; e^k) \;\; are\ bounded \qquad \forall\, k \in I,$$

we obtain from (20) (also using the fact that $\{\lambda^r\}_{r \in R} \to 0$)

$$\lim_{r \to \infty,\, r \in R} \sum_{k \in I} \tau_k^r \lambda_k^r = 0.$$

This together with (18) imply

$$\lim_{r \to \infty,\, r \in R} (t^r)^T(x^r - x^*) = 0. \tag{21}$$

Since $\{x^r\}_{r \in R} \to x^*$ it follows from (21) and the observation

$$f(x^*) - q(p^r) = f(x^*) - (t^r)^T x^* + (t^r)^T x^r - f(x^r) = (t^r)^T(x^r - x^*) + f(x^*) - f(x^r) \;\; \forall\, r,$$

that $\{f(x^*) - q(p^r)\}_{r \in R} \to 0$. Since $q(p^r)$ is nondecreasing with r it follows $f(x^*) - q(p^r) \to 0$ and (b) is proven.

Finally we prove (c). The assumption of part (c) shows that the convex program (P) is strictly consistent (see [7], pp. 300). It follows from Corollary 29.1.5 of [7] that the level sets of the dual functional q are compact. Since $q(p^r)$ is nondecreasing with r this implies $\{p^r\}$ is bounded. Then every limit point of $\{p^r\}$ satisfies the Kuhn-Tucker conditions with x^* (since $p^r \geq 0$ for all r and $(p^r)^T(b-Ex^r) \to 0$ by (13)). Q.E.D.

Note that the condition in part (b) holds if f is separable and (P) is regularly feasible ([15], Ch. 11).

In the remainder of this section we will further assume f is strongly convex, in the sense that there exist scalars $\sigma > 0$ and $\gamma > 1$ such that

$$f(y) - f(x) - f'(x; y-x) \geq \sigma \|y - x\|^{\gamma} \qquad \forall\, x, y \in S. \tag{22}$$

We consider another assumption regarding the order of relaxation that is weaker than Assumption C. Let $\{\tau_k\}$ be a sequence satisfying the following condition:

$$\tau_1 = 0 \qquad and \qquad \tau_{k+1} = \tau_k + b_k \ , \ k = 1, 2, \ldots,$$

where $\{b_k\}$ is any sequence of scalars satisfying

$$b_k \geq n, \ k = 1, 2, \ldots, \quad and \quad \sum_{k=1}^{\infty} \left\{ \frac{1}{b_k} \right\}^{\rho} = \infty, \tag{23}$$

and $\rho = \gamma - 1$. The assumption is as follows:

<u>Assumption C'</u>: For every positive integer k, every coordinate is chosen at least once for relaxation between iterations $\tau_k + 1$ and τ_{k+1}.

The condition $b_k \geq n$ for all k is required to allow each coordinate to be relaxed at least once between iterations $\tau_k + 1$ and τ_{k+1} so that Assumption C' can be satisfied. Note that if $b_k \to +\infty$ then the length of the interval $[\tau_k + 1, \tau_{k+1}]$ tends to $+\infty$ with k. For example, $b_k = k^{1/\rho} n$ gives one such sequence.

Assumption C' allows the time between successive relaxation of each coordinate to grow, although not to grow too fast. We will show that most of the conclusions of Proposition 1 hold, under Assumption C' and (22). These convergence results are of interest since they show that, for a large class of problems, near cyclical relaxation is not essential for the Gauss-Seidel method to be convergent. To the best of our knowledge, the only other work treating convergence of the Gauss-Seidel method that do not require cyclical relaxation are [9], [10] and [11] dealing with the special case of separable costs.

Before proceeding to our main result we need the following lemma:

Lemma 6 If (22) and Assumption C′ hold, then there exist a limit point x′ of $\{x^r\}$ belonging to S∩Q and a subsequence R satisfying

$$\{x^r\}_{r\in R} \to x' \quad and \quad b_i - \sum_{j=1}^{m} e_{ij}x_j' < 0 \;\Rightarrow\; p_i^r = 0 \;\forall\, r\in R. \tag{24}$$

Proof: We first show that a limit point $x'\in S\cap Q$ exists. Since $x^r\in S$ for all r, then (5) and (22) imply

$$q(p^{r+1}) - q(p^r) \;\ge\; \sigma\,\|x^r - x^{r+1}\|^{\beta} \qquad r=0,1,2,\ldots,$$

which, when summed over all r, yields

$$\lim_{r\to\infty} q(p^r) - q(p^0) \;\ge \sigma \sum_{r=0}^{\infty} \| x^{r+1} - x^r \|^{\gamma},$$

and so it follows

$$\sum_{r=0}^{\infty} \| x^{r+1} - x^r \|^{\gamma} \;<\; \infty\,. \tag{25}$$

[Note that (6) and (22) imply

$$f(x^*) - q(p^r) \;\ge\; f(x^*) - f(x^r) - f'(x^r;x^* - x^r) \;\ge\; \sigma\,\|x^r - x^*\|^{\gamma} \qquad \forall\ r, \tag{26}$$

which together with the fact $q(p^r)$ is nondecreasing with r offers a simpler proof that $\{x^r\}$ is bounded.]

Next we show that there exists subsequence H of {1,2,...} for which

$$\sum_{r=\tau_h+1}^{\tau_{h+1}} \|x^{r+1} - x^r\| \to 0 \quad as \;\; h\to\infty\,,\, h\in H. \tag{27}$$

We will argue by contradiction. Suppose that such a subsequence does not exist. Then there must exist a positive scalar ε and a h^* for which

$$\sum_{r=\tau_h+1}^{\tau_{h+1}} \|x^{r+1}-x^r\| \;\geq\; \varepsilon \qquad \forall\, h \geq h^*. \tag{28}$$

Consider the Hölder inequality [12], which says that for any positive integer N and two vectors x and y in R^N

$$|x^T y| \;\leq\; \|x\|_p \|y\|_q,$$

where $1/p + 1/q = 1$ and $p > 1$. If $x \geq 0$ and if we let y be the vector with entries all 1 we obtain that

$$\sum_{i=1}^{N} x_i \;\leq\; \left[\sum_{i=1}^{N} (x_i)^p\right]^{1/p} (N)^{1/q}. \tag{29}$$

Applying (29) to the left hand side of (27) with $p = \gamma$ yields

$$\varepsilon^\gamma \;\leq\; \left[\sum_{r=\tau_h+1}^{\tau_{h+1}} \|x^{r+1}-x^r\|^\gamma\right](\tau_{h+1}-\tau_h)^{\gamma-1} \qquad \forall\, h \geq h^*,$$

which implies

$$\varepsilon^\gamma \sum_{h=h^*}^{\infty} \frac{1}{(\tau_{h+1}-\tau_h)^{\gamma-1}} \;\leq\; \sum_{h=h^*}^{\infty}\left[\sum_{r=\tau_h+1}^{\tau_{h+1}} \|x^{r+1}-x^r\|^\gamma\right] \;=\; \sum_{r=\tau_{h^*}+1}^{\infty} \|x^{r+1}-x^r\|^\gamma. \tag{30}$$

The leftmost quantity of (30) according to Assumption C' [cf. (23)] is $+\infty$ while the rightmost quantity of (30) according to (25) is bounded – a clear contradiction.

We also have [cf. Lemma 4]

$$\lim_{r\to+\infty} \sup\{\, d^r_{s^r} \,\} \;\leq\; 0\,. \tag{31}$$

Now, let H be the subsequence for which (27) holds and considered a fixed i. By Assumption C', p_i is chosen to be relaxed in at least one iteration, which we denote by r_h, between $\tau_h + 1$ and τ_{h+1} for all $h \in H$, so it follows

$$d_i^{\tau_h} \;=\; d_i^{r_h} \;-\; \sum_{r=\tau_h}^{r_h-1} \sum_{j=1}^{m} e_{ij}\,(x_j^{r+1} - x_j^r),$$

implying

$$d_i^{\tau_h} \le \max\{ d_{s^r}^{r} \mid r=\tau_h+1,\ldots,\tau_{h+1}\} + M \sum_{r=\tau_h}^{\tau_{h+1}-1} \|x^{r+1}-x^r\| \quad \forall\, h\in H, \tag{32}$$

where M is some constant depending on E only. Then (27) and (31) imply

$$\lim_{h\to+\infty,\, h\in H} \sup\{ d_i^{\tau_h} \} \le 0.$$

Thus every limit point of $\{x^{\tau_h}\}_{h\in H}$ is in Q, which [cf. Lemma 2] is also in S.

Let x' be such a limit point (so $x'\in S\cap Q$) and denote $d = b-Ex'$. Then $d_i \le 0$ for all i. Let $I = \{ i \mid d_i < 0 \}$ and assume that I is nonempty (otherwise we can set $R = \{ \tau_h \mid h\in H \}$ and (24) follows). Since x' is a limit point of $\{x^r\}$, there exists subsequence H' of H for which $\{x^{\tau_h}\}_{h\in H'} \to x'$, implying

$$\{d_i^{\tau_h}\}_{h\in H'} \to d_i,\ \forall\, i\in I.$$

Then taking $h\in H'$ sufficiently large we will assume [cf. (27)]

$$d_i^{r+1} < \delta d_i^r \quad \forall\ r=\tau_h, \tau_h+1,\ldots,\tau_{h+1}-1, \forall\, h\in H', \forall\, i\in I. \tag{33}$$

Consider a fixed $h\in H'$ and any $i\in I$. If $s^k = i$ for some $k\in[\tau_h,\tau_{h+1}-1]$ then by (33) we have

$$d_i^{k+1} < \delta d_i^k ,$$

which together with the statement of the relaxation iteration imply

$$p_i^{k+1} = 0.$$

Since $p_i^{k+1} = p_i^k$ if $s^k \ne i$ for all k then necessarilly $p_i^{\tau_{h+1}} = 0$. By Assumption C', for each $i\in I$, $s^k = i$ for at least one $k\in[\tau_h,\tau_{h+1}-1]$ so it follows

$$p_i^{\tau_{h+1}} = 0 \quad \forall\, i\in I.$$

Since the choice of $h\in H'$ was arbitrary we can choose $R = \{ \tau_{h+1} \mid h\in H' \}$ (note that by (27) we have $\{x^r\}_{r\in R} \to x'$). Q.E.D.

The following is our main result for the strongly convex case:

<u>Proposition 2</u> If (22) and Assumption C' are satisfied, then the following hold:

(a) $\{x^r\}_{r\in R} \to x^*$ for some subsequence R.

(b) If the closure of S 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 $x, y \in B \cap S$ then $q(p^r) \to f(x^*)$ and $x^r \to x^*$.

(c) If $\text{int}(Q) \cap S \neq \emptyset$ then $q(p^r) \to f(x^*)$, $x^r \to x^*$, and $\{p^r\}$ is bounded whose each limit point is an optimal dual solution.

Proof: We prove (a) first. Let x' be a limit point and R be a subsequence satisfying (24). Then by a proof analogous to that for Proposition 1 (a) (using (24) and Lemma 6 instead of (13) and Lemma 5) we obtain $x' = x^*$ (unlike the proof of Proposition 1 (a) however, this does not imply $x^r \to x^*$). To prove (b) we use part (a) and an argument analogous to that for Proposition 1 (b) to conclude that $q(p^r) \to f(x^*)$. Then (26) implies $x^r \to x^*$. To prove (c) we first apply the argument for Proposition 1 (c) to conclude that $\{p^r\}$ is bounded. This implies $\{p^r\}_{r\in R}$ has a limit point p' that is an optimal dual solution, where R is the subsequence defined in part (a) (since $p^r \geq 0$ for all $r \in R$ and $\{(p^r)^T(b-Ex^r)\}_{r\in R} \to 0$ by (24)). Therefore $q(p') = f(x^*)$ and $q(p^r) \to f(x^*)$ (since $q(p^r)$ is nondecreasing with r), implying that every limit point of $\{p^r\}$ is an optimal dual soluton. Q.E.D.

The conclusion of Proposition 2 (a) is weaker than that of Proposition 1 (a) since it does not assert convergence of the entire sequence $\{x^r\}$.

References

[1] 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*, Vol. 7, pp. 200–217 (1967).

[2] Censor, Y. and Lent, A., "An Iterative Row-Action Method for Interval Convex Programming", *J. of Optimization Theory and Applications*, Vol. 34, pp. 321–352 (1981).

[3] D'Esopo, D. A., "A Convex Programming Procedure", *Naval Research Logistics Quarterly*, Vol. 6, pp. 33–42 (1959).

[4] Luenberger, D. L., *Linear and Nonlinear Programming*, Addison Wesley (1984).

[5] 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).

[6] Powell, M. J. D., "On Search Directions for Minimization Algorithms", *Mathematical Programming 4*, North-Holland Publishing, pp. 193–201 (1973).

[7] Rockafellar, R. T., *Convex Analysis*, Princeton Univ. Press (1970).

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

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

[10] Bertsekas, D.P. and Elbaz, D., "Distributed Asynchronous Relaxation Methods for Convex Network Flow Problems," *SIAM Journal on Control and Optimization*, Vol. 25, pp. 74–85 (1987).

[11] Bertsekas, D.P., Hosein, P. A. and Tseng, P., "Relaxation Methods for Network Flow Problems with Convex Arc Costs," *SIAM Journal on Control and Optimization*, Vol. 25, pp. 1219–1243 (1987).

[12] Golub, G. H. and Van Loan, C. F., *Matrix Computations*, Johns Hopkins Univ. Press (1985).

[13] Polak, E., *Computational Methods in Optimization: A Unified Approach*, Academic Press, New York (1971).

[14] Zangwill, W. I., *Nonlinear Programming*, Prentice-Hall, Englewood Cliffs, NJ (1969).

[15] Rockafellar, R. T., *Network Flow and Monotropic Optimization*, John Wiley & Sons (1983).

[16] Cottle, R.W. and Pang, J.S., "On the Convergence of a Block Successive Over-Relaxation Method for a Class of Linear Complementarity Problems," *Mathematical Programming Study 17*, North Holland Publishing, pp. 126-138 (1982).