MATHEMATICS OF OPERATIONS RESEARCH
Vol. 16, No 3, August 1991
Printed in U S A

RELAXATION METHODS FOR PROBLEMS WITH STRICTLY CONVEX COSTS AND LINEAR CONSTRAINTS*†

PAUL TSENG AND DIMITRI P. BERTSEKAS

Consider the problem of minimizing a strictly convex (possibly nondifferentiable and nonseparable) cost subject to linear constraints. We propose a dual coordinate ascent method for this problem that uses inexact line search and either essentially cyclic or Gauss-Southwell 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 that the effective domain of the cost function is polyhedral, we show that a related sequence of dual vectors converges in cost to the optimal cost. If the constraint set has an interior point in the effective 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. When the cost function is strongly convex, we show that the order of coordinate relaxation can become progressively more chaotic. These results significantly improve upon those obtained previously.

1. Introduction. In this paper, we consider the problem of minimizing a strictly convex, possibly nondifferentiable and nonseparable, cost subject to linear constraints. This is an important problem, containing as special cases nonlinear cost network flow problems [5], [6], [13], [21], [30], [32], [42], matrix balancing [1], [9], [22], [23], "$x \log x$" entropy optimization [12], [19], [20], [24], and strictly convex quadratic programming [2], [18], [26], [28], [29]. The problem of computing the orthogonal projection of a point onto the intersection of a finite collection of convex sets [16] can also be formulated as such a problem.

A popular approach for solving this problem is based on dualizing the linear constraints to obtain a dual problem of maximizing a concave *differentiable* dual functional. A coordinate ascent method is then applied to solve the dual problem whereby, at each iteration, one of the coordinates of the dual vector is chosen and the dual functional is maximized along this coordinate direction. The resulting method is simple, uses little storage, can exploit problem sparsity, and in certain cases is highly parallelizable [4], [6], [43]. The first such method is a matrix balancing method of Kruithof [22]. Subsequently, other dual coordinate ascent methods were proposed for network flow [4], [5], [30], [38], [42], matrix balancing [1], [9], [23], entropy optimization [24], and strictly convex quadratic programming [11], [18], [29]. Methods for solving more general problems are described in [7], [8], [32], [39]. (See [23] for additional references on methods for matrix balancing and related problems.)

A fundamental question concerning these dual coordinate ascent methods is their convergence. Existing convergence results regarding coordinate ascent methods for maximizing general concave differentiable functions (see [6], [14], [27], [33], [34], [37],

*Received December 18, 1987; revised October 13, 1989.
AMS 1980 subject classification. Primary: 90C25.
IAOR 1973 subject classification. Main: Programming: convex.
OR/MS Index 1978 subject classification. Primary: 656 Programming/nonlinear/convex.
Key words. Convex program, dual functional, coordinate ascent, strict convexity.
†This research is 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.

0364-765X/91/1603/0462/$01.25

[41]) require compactness of the level sets and some form of strict concavity of the objective function, neither of which necessarily holds for the dual functional that we consider here. For this reason, convergence for our methods has been quite difficult to establish. Only for the special case of *separable* costs and linear equality constraints is the issue of convergence fairly well resolved (see [5], [39]). The strength of the available results depends on the problem's structure. For example, for separable, strictly convex, single commodity network flow problems, dual coordinate ascent methods can be shown to be convergent with essentially no restriction in the order of coordinate relaxation [5], [6], and even using an *asynchronous* implementation [4], [6], but there are no results of comparable strength for other types of separable problems.

For problems with nonseparable cost, previous convergence results [7], [8], [32] have all required restrictive assumptions. Thus, [7] and [8] require differentiability of the cost function within its effective domain, and place additional restrictions on the form of the effective domain. Reference [32] requires differentiability and uniform convexity of the cost function, and assumes that its effective domain is polyhedral. Furthermore, in [7], [8], [32], each coordinate ascent iteration uses exact line search.

In this paper, we propose a dual coordinate ascent method for solving the general nonseparable case under much weaker regularity assumptions than in previous works; for example we do not require differentiability of the primal cost function and we place very mild and easily verifiable restrictions on its effective domain. We show that the method is convergent even when the line search is inexact and the coordinates are relaxed either in an essentially cyclic manner or in a Gauss-Southwell manner. Moreover, we show that when the cost function is in some sense strongly convex, the coordinates can be relaxed in a manner that is progressively more chaotic. This latter result extends those obtained in [39] for the separable cost case.

The paper proceeds as follows: In §2, we introduce the problem. In §3 we propose a dual coordinate ascent method for its solution, In §4 we develop some technical results concerning convex functions and in §5 we use these results to establish convergence of our method for the general problem. In §6 we strengthen our results for the special case where the cost is strongly convex.

2. Problem description. Consider the following minimization problem

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

$$(1.1) \qquad \text{subject to} \quad Ax \geqslant b,$$

where $f\colon \Re^m \to (-\infty, \infty]$, A is a given $n \times m$ matrix, and b is a vector in $\Re^n$.

We pause to explain our notation: Throughout the paper, $\Re^n$ is the n-dimensional Euclidean space. All vectors in $\Re^n$ are viewed as column vectors and superscript T denotes transpose. For any vector x, we denote by x_i the ith coordinate of x and, for any two vectors x and y of the same dimension, we denote by $\langle x, y \rangle$ the Euclidean inner product of x with y, i.e., $\langle x, y \rangle = \sum_i x_i y_i$. We denote by $\|x\|$ the standard Euclidean norm of x, i.e., $\|x\| = \langle x, x \rangle^{1/2}$. For any set D in $\Re^m$, we denote by cl(D) and ri(D), respectively, the closure and the relative interior of D. We denote the (i, j)th entry of A by a_{ij}, the *effective domain* of f by C_f, i.e.,

$$C_f = \{x \mid f(x) < \infty\},$$

and the constraint set by X, i.e.,

$$X = \{x \mid Ax \geqslant b\}.$$

For any $x \in C_f$ and $z \in \Re^m$, let $f'(x; z)$ denote the *directional derivative* of f at x in the direction z, i.e.,

$$f'(x;z) = \lim_{\theta \downarrow \infty} \frac{f(x+\theta z) - f(z)}{\theta}.$$

It is well known [35, Theorem 23.4] that $f'(x; z)$ can also be given in terms of $\partial f(x)$, the *subdifferential* of f at x, by

$$f'(x;z) = \sup_{\eta \in \partial f(x)} \langle z, \eta \rangle. \tag{1.2}$$

Notice that our problem formulation admits quite general constraints. In particular, the effective domain C_f may be any convex set (not necessarily closed), so arbitrary convex inequality constraints can be embedded into f. However, the linear constraints $Ax \geqslant b$ are the only ones that will be "dualized" in the subsequent development. Note also that our analysis extends in a straightforward manner to problems with linear *equality* constraints (it in fact simplifies in this case) or even problems with mixed linear equality/inequality constraints, but for simplicity we will not treat such problems here.

We make the following standing assumptions about (**P**):

ASSUMPTION A. *f is strictly convex, lower semicontinuous, and continuous within C_f. Furthermore, the conjugate function of f* [35], *defined by*

$$f^*(t) = \sup_x \{\langle t, x \rangle - f(x)\},$$

is real-valued, i.e., $-\infty < f^*(t) < \infty$ for all $t \in \Re^m$. (By f is continuous within C_f, we mean that for any $x \in C_f$ and any sequence of vectors $\{x^1, x^2, \dots\}$ in C_f converging to x, there holds $\{f(x^k)\} \to f(x)$.)

ASSUMPTION B. *There exist convex sets P and Q in $\Re^m$ such that* $\mathrm{cl}(P)$ *is a polyhedral set, $C_f = P \cap Q$, and $P \cap \mathrm{ri}(Q) \cap X \neq \varnothing$.*

The restrictions that Assumption A places on f, with the exception of strict convexity, are fairly mild. For instance, the assumption that f is continuous within C_f holds if C_f is an open set or if f is the pointwise sum of a real-valued convex function with the indicator function of a closed convex set. The assumption that f^* is real-valued roughly means that the rate of growth of f along any one direction tends to ∞ as $\|x\|$ tends to ∞, that is, f is *cofinite* in the terminology of [35, Corollary 13.3.1]. Most of the strictly convex functions encountered in practice satisfy this assumption; for example, uniformly convex functions [31] or functions with bounded effective domain. Note that because f is cofinite and lower semicontinuous, its level sets are compact. Thus, for every t there is some x with $f(x) < \infty$ attaining the supremum in the definition of the conjugate $f^*(t)$. Furthermore, using also the strict convexity of f and the feasibility assumption for (**P**) (implied by Assumption B), it follows that *there exists a unique optimal solution* to (**P**), which we denote by x^*.

Assumption B may be viewed as a feasibility assumption for (**P**) plus a constraint qualification (if we replace "ri(Q)" by "Q" in Assumption B, then we obtain exactly $C_f \cap X \neq \varnothing$). The constraint qualification, which is necessary for establishing the convergence of our algorithm (see §§5 and 6), is almost always satisfied. For example, it is satisfied when the usual regularity condition $\mathrm{ri}(C_f) \cap X \neq \varnothing$ holds or when $C_f \cap X \neq \varnothing$ and f is *separable* (in which case C_f is a box, possibly with some faces missing).

By assigning a nonnegative Lagrange multiplier p_i to the ith constraint of $Ax \geqslant b$ we obtain the following dual problem of **(P)**:

$$\textbf{(D)} \qquad \begin{array}{ll} \text{Maximize} & q(p) \\ \text{subject to} & p \geqslant 0, \end{array}$$

where $q\colon \Re^n \to (-\infty, \infty]$ is the concave *dual functional* given by

$$q(p) = \min\{f(x) + \langle p, b - Ax\rangle\} = \langle p, b\rangle - f^*(A^{\mathrm{T}}p). \tag{1.3}$$

We will call p the *dual vector*. The dual problem **(D)** is a concave program with simple nonnegativity constraints. Furthermore, *strong duality holds for* **(P)** *and* **(D)**, *i.e., the optimal value of* **(P)** *equals the optimal value of* **(D)**; this is proved in Appendix A. Notice, however, that **(D)** may not have an optimal solution (consider minimizing the function $f\colon \Re \to (-\infty, \infty]$ given by $f(x) = x^2 - (x)^{1/2}$ if $x \geqslant 0$ and ∞ otherwise, subject to $x \leqslant 0$).

Since f^* is real-valued and f is strictly convex, f^* and q are continuously differentiable (cf. [35, Theorem 26.3]). For convenience, we will denote the gradient of q at p by $d(p)$ (i.e., $d(p) = \nabla q(p)$) and its ith coordinate by $d_i(p)$. Since q is continuously differentiable, $d_i(p)$ is continuous, and since q is concave, $d_i(p)$ is nonincreasing in p_i (with the other coordinates of p held fixed). By differentiating (1.3) and by using the chain rule, we obtain the dual cost gradient

$$d(p) = b - Ax, \tag{1.4}$$

where

$$x = \nabla f^*(A^{\mathrm{T}}p) = \arg\sup_{\xi}\{\langle p, A\xi\rangle - f(\xi)\} \tag{1.5}$$

(the last equality above can be justified using Danskin's Theorem; see [6, p. 649] and [35, Theorem 23.5]). Note from (1.5) that x is also the unique vector in $\Re^m$ satisfying

$$A^{\mathrm{T}}p \in \partial f(x). \tag{1.6}$$

From the optimality conditions for **(D)**, it can be seen that a dual vector p is an optimal solution of **(D)** if and only if the orthogonal projection of $p + d(p)$ on the positive orthant of $\Re^n$ is p itself, i.e.,

$$p = [p + d(p)]^+, \tag{1.7}$$

where, for any $y \in \Re^n$, $[y]^+$ denotes the orthogonal projection of y onto the positive orthant, i.e., $[y]^+ = (\max\{0, y_1\}, \ldots, \max\{0, y_n\})$. Given an optimal dual solution p, we may obtain an optimal primal solution from the equation $x = \nabla f^*(A^{\mathrm{T}}p)$. To see this, note that we have (cf. (1.4)–(1.5))

$$d(p) = b - Ax, \tag{1.8}$$

so the optimality condition (1.7) implies that

$$Ax \geqslant b \quad \text{and} \quad p_i = 0 \quad \text{for all } i \text{ such that } \sum_{j=1}^{m} a_{ij}x_j > b_i. \tag{1.9}$$

On the other hand, from (1.8) we obtain (cf. (1.6))

$$A^{\mathrm{T}}p \in \partial f(x). \tag{1.10}$$

From (1.9) and (1.10), we see that x and p satisfy the Kuhn-Tucker optimality conditions for **(P)**, so x is also an optimal solution for **(P)**.

Differentiability of q motivates a coordinate ascent method for solving **(D)** whereby, given a dual vector p, a coordinate p_i such that $d_i(p) > 0$ (or < 0) is chosen and p_i is increased (or decreased, respectively) in order to increase the dual function value. An important advantage of such a coordinate relaxation method is its suitability for parallel implementation on problems where f and A have special structure. As an example, suppose that $f(x)$ is quadratic of the form $\langle x, Qx\rangle/2 + \langle c, x\rangle$, where Q is an $m \times m$ symmetric positive definite matrix and c is a vector in $\Re^m$. Then, two coordinates p_i and p_j are uncoupled, and can be iterated upon simultaneously if the (i, j)th entry of $AQ^{-1}A^{\mathrm{T}}$ is zero (see [6, p. 228]); another example arises when f is separable and the (i, j)th entry of AA^{T} is zero. In the next section, we describe the dual coordinate ascent method and in subsequent sections, we develop its convergence properties.

3. Relaxation algorithm. We now describe our dual coordinate ascent algorithm for solving **(P)** and **(D)**: We fix a scalar δ in the interval (0, 1) and, beginning with a nonnegative p in $\Re^n$, we repeatedly apply the following iteration:

RELAXATION ITERATION.

If $d_i(p) \leqslant 0$ for all i, and $d_i(p) = 0$ for all i with $p_i > 0$, then STOP; p is an optimal solution of **(D)** (cf. (1.7)).

Else

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

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

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

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

We first show that each relaxation iteration is well defined, in the sense that the approximate search along the coordinate p_s can be executed as specified. Indeed, let e^s denote the sth coordinate vector in $\Re^n$. If the relaxation iteration is not well defined, then, for some $p \geqslant 0$ and some index s, we have (a) $d_s(p) > 0$ and there does not exist a $\theta \geqslant 0$ such that $d_s(p + \theta e^s) \leqslant \delta d_s(p)$ or (b) $d_s(p) < 0$ and there does not exist a $\theta \leqslant 0$ such that $d_s(p + \theta e^s) \geqslant \delta d_s(p)$. Assume case (a) holds. Then, since $d_s(p + \theta e^s)$ is nonincreasing in θ, we have

$$\liminf_{\theta \to \infty} d_s(p + \theta e^s) \geqslant \delta d_s(p) > 0,$$

which together with the fact $q(p + \lambda e^s) = q(p) + \int_0^\lambda d_s(p + \theta e^s)\, d\theta$, yields

$$\lim_{\lambda \to \infty} q(p + \lambda e^s) = \infty.$$

This, in view of the strong duality condition $\max_{p \geqslant 0} q(p) = \min_{Ax \geqslant b} f(x)$, contradicts the feasibility of **(P)** (cf. Assumption B). Using a similar argument, we also reach a contradiction in case (b). We remark that, under the additional regularity assumption $\mathrm{ri}(C_f) \cap X \neq \varnothing$, it can be shown that the relaxation iteration with $\delta = 0$ is also well defined. The proof of this, given in Appendix B, is constructive and is based on showing, following [32], that the result of the line search is the unique solution of a certain optimization problem.

We will consider one of the following two assumptions regarding the order in which the coordinates are chosen for relaxation:

ASSUMPTION C1 (Essentially Cyclic Order). *There exists a constant T for which every coordinate is chosen at least once for relaxation between iterations r and $r + T$, for $r = 0, 1, \ldots$.*

ASSUMPTION C2. *The coordinate p_s chosen for relaxation satisfies $|p_s - [p_s + d_s(p)]^+| \geqslant \omega \max_t |p_t - [p_t + d_t(p)]^+|$, where ω is a fixed scalar in $(0, 1]$.*

A classical special case of Assumption C1 is when the coordinates are chosen according to a fixed cyclic order. We will weaken Assumption C1 in §6 to allow the constant T to become unbounded in the course of the algorithm. Assumption C2, which is reminiscent of the *Gauss-Southwell method* (see [27, §7.8]), says that the coordinate chosen for relaxation must be one along which the rate of ascent is in some sense the largest. Most of the coordinate ascent methods proposed previously use the cyclic order of relaxation. In some cases, however, using a noncyclic order of relaxation improves the efficiency of the method (see for example [1]).

Our main convergence result is the following:

PROPOSITION 1. *If either Assumption* C1 *or Assumption* C2 *is satisfied, then the following hold*:

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

(b) *If* $\mathrm{cl}(C_f)$ *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 $x, y \in B \cap C_f$, then $\{q(p^r)\} \to f(x^*)$.*

(c) *If (interior of X) $\cap C_f \neq \varnothing$, then $\{p^r\}$ is bounded and every one of its limit points is an optimal solution of* (**D**).

We prove this result in the next two sections and we extend it under a strong convexity assumption in §6.

Notice that the conditions in part (b) of Proposition 1 hold for many important problem instances. For example, they hold when f is separable and (**P**) is *regularly feasible* [36, Chapter 11]. (However, in the separable case, the conclusion of part (b) holds even without the regular feasibility assumption. This can be shown by using a different proof that exploits the separable structure [39].) Also notice that the condition that $f'(x; (y - x)/\|y - x\|)$ is bounded for all $x, y \in C_f$ near x^* holds automatically if x^* is in $\mathrm{ri}(C_f)$. Finally, we note that, without a regularity condition like the one in part (c) of Proposition 1 to ensure boundedness of the set of optimal dual solutions, convergence to an optimal dual solution is likely very difficult to prove (if at all possible). Only for the special cases of network flow problems and strictly convex quadratic programs has such convergence been shown without assuming boundedness of the optimal dual solution set [5], [28], and the proofs for these special cases are quite complicated.

4. Technical preliminaries. In this section we develop two technical lemmas characterizing a convex and lower semicontinuous function which is continuous in its effective domain (the second lemma assumes also that the function is cofinite). These lemmas will be used in §§5 and 6 to analyze the convergence of the relaxation algorithm of §3. These lemmas are also of independent interest, since they can be used to analyze dual ascent methods in general (see [40]).

LEMMA 1. *Let $h: \Re^m \to (-\infty, \infty]$ be any convex lower semicontinuous function that is continuous on its effective domain C_h. Then the following hold*:

(a) *For any $y \in C_h$, there exists a positive scalar ϵ such that $C_h \cap B(y, \epsilon)$ is closed, where $B(y, \epsilon)$ denotes the closed ball around y with radius ϵ.*

(b) *For any $y \in C_h$, any z such that $y + z \in C_h$, and any sequences $\{y^1, y^2, \ldots\} \to y$ and $\{z^1, z^2, \ldots\} \to z$ such that $y^k \in C_h$, $y^k + z^k \in C_h$ for all k, we have*

$$\limsup_{k \to \infty} \{h'(y^k; z^k)\} \leqslant h'(y; z).$$

PROOF. (a) Let $Y = \mathrm{cl}(C_h) \setminus C_h$. It then suffices to show that $Y \cap B(y, \epsilon) = \varnothing$ for some positive ϵ. Suppose the contrary. Then there exists a sequence of vectors $\{y^1, y^2, \ldots\}$ in Y converging to $y \in C_h$. Consider a fixed $k \in \{1, 2, \ldots\}$. Since $y^k \in Y$, there exists a sequence of points $\{w^{k,1}, w^{k,2}, \ldots\}$ in C_h converging to y^k. Since h is lower semicontinuous and $h(y^k) = \infty$, it must be that $h(w^{k,i}) \to \infty$ as $i \to \infty$. Therefore there exists an integer $m(k)$ for which $\|w^{k,m(k)} - y^k\| < 1/k$ and $h(w^{k,m(k)}) > k$. Then $\{w^{k,m(k)}\}_{k=1,2,}$ is a sequence of points in C_h converging to y for which $h(w^{k,m(k)}) \to \infty$, contradicting the continuity of h on C_h since $h(y) < \infty$.

(b) We will use the idea from the proof of Theorem 24.5 in [35]. Fix any $\mu > h'(y; z)$ ($h'(y; z) < \infty$ since $h'(y; z) < h(y + z) - h(y)$). Then, for all $\lambda > 0$ sufficiently small, we have $y + \lambda z \in C_h$ and

$$\frac{h(y + \lambda z) - h(y)}{\lambda} < \mu. \tag{4.1}$$

Since $y^k + z^k \in C_h$ for all k, we can (by taking $\lambda \leqslant 1$) further assume that $y^k + \lambda z^k \in C_h$ for all k. Since $h(y^k + \lambda z^k) \to h(y + \lambda z)$ and $h(y^k) \to h(y)$ by continuity of h on C_h, it follows from (4.1) that, for all k sufficiently large,

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

Since, by convexity of h,

$$h'(y^k; z^k) \leqslant \frac{h(y^k + \lambda z^k) - h(y^k)}{\lambda} \quad \forall\, k,$$

it follows that

$$\limsup_{k \to \infty} \{h'(y^k; z^k)\} \leqslant \mu.$$

This holds for any $\mu > h'(y; z)$, so part (b) follows. Q.E.D.

LEMMA 2. *Let $h: \Re^m \to (-\infty, \infty]$ be any convex lower semicontinuous function that is cofinite (i.e., the conjugate h^* is real-valued) and continuous on its effective domain C_h. If $\{y^1, y^2, \ldots\}$ is any sequence of vectors in C_h such that, for some $y \in C_h$, $\{h(y^k) + h'(y^k; y - y^k)\}$ is bounded from below, then both $\{y^k\}$ and $\{h(y^k)\}$ are bounded, and every limit point of $\{y^k\}$ is in C_h.*

PROOF. First we show that the sequence $\{y^k\}$ is bounded. Suppose that $\{y^k\}$ is not bounded. Then there exists subsequence K of $\{1, 2, \ldots\}$ for which $\{\|y^k\|\}_{k \in K} \to \infty$ and $\|y^k - y\| \geqslant 1$ for all $k \in K$. Fix an arbitrary scalar $\beta \in (0, 1)$ and let

$$\xi^k = y + \beta(y^k - y)/\|y^k - y\| \quad \forall\, k \in K. \tag{4.2}$$

The sequence $\{\xi^k\}_{k \in K}$ has the property that

$$\xi^k \in C_h, \quad \|\xi^k - y\| = \beta \quad \forall\, k \in K,$$

and, for β sufficiently small, $\{h(\xi^k)\}_{k \in K}$ is bounded (otherwise there exists a subsequence K' of K and $\{\beta^k\}_{k \in K'} \downarrow 0$ such that $\{h(y + \beta^k(y^k - y)/\|y^k - y\|)\}_{k \in K'} \to \infty$, thus contradicting the continuity of h within C_h since $y + \beta^k(y^k - y)/\|y^k - y\|$ converges to y as $k \to \infty$, $k \in K'$, and is in C_h for all $k \in K'$ sufficiently large). From the hypothesis that $\{h(y^k) + h'(y^k;\, y - y^k)\}$ is bounded from below, we have

$$h(y^k) \geqslant h(y) - h'(y^k; y - y^k) - \Delta \quad \forall\, k \in K,$$

for some constant Δ. Combining this with the convexity of h yields

$$\begin{aligned} h(\xi^k) &\geqslant h(y^k) + h'(y^k; \xi^k - y^k) \\ &\geqslant h(y) - h'(y^k; y - y^k) - \Delta + h'(y^k; \xi^k - y^k) \\ &= h(y) - h'(y^k; y - \xi^k) - \Delta, \end{aligned}$$

where the equality follows from the positive homogeneous property of $h'(y^k; \cdot\,)$ [35, Theorem 23.1] and the observation that $y - y^k, \xi^k - y^k, y - \xi^k$ are all positive scalar multiples of each other. Therefore

$$\text{(4.3)} \quad \begin{aligned} \big(h(\xi^k) - h(y)\big)/\beta + \Delta/\beta &\geqslant -h'(y^k; y - \xi^k)/\beta \\ &= -h'\big(y^k; (y - \xi^k)/\beta\big) \\ &= -h'\big(y^k; (y - y^k)/\|y - y^k\|\big) \quad \forall\, k \in K. \end{aligned}$$

Since the left-hand side of (4.3) has been shown to be bounded for sufficiently small $\beta > 0$, it follows that there exists a constant ρ for which

$$\begin{aligned} \rho &\geqslant -h'\big(y^k; (y - y^k)/\|y - y^k\|\big) \\ &\geqslant \big(h(y^k) - h(y)\big)/\|y - y^k\| \quad \forall\, k \in K, \end{aligned}$$

where the second inequality follows from the convexity of h. Hence

$$\text{(4.4)} \qquad \rho \geqslant \frac{h(y + \lambda^k z^k) - h(y)}{\lambda^k} \quad \forall\, k \in K,$$

where we define $\lambda^k = \|y - y^k\|$ and $z^k = (y^k - y)/\|y^k - y\|$ for all $k \in K$. Then $\{\lambda^k\}_{k \in K} \to \infty$ and $\{z^k\}_{k \in K}$ lies on the unit circle. By further passing to a subsequence if necessary, we will assume $\{z^k\}_{k \in K} \to z$ for some nonzero vector z. Now consider any fixed $\lambda > 0$. Then, for all $k \in K$ sufficiently large (so $\lambda^k > \lambda$), we have $y + \lambda z^k \in C_h$ and (using the convexity of h and (4.4))

$$\frac{h(y + \lambda z^k) - h(y)}{\lambda} \leqslant \frac{h(y + \lambda^k z^k) - h(y)}{\lambda^k} \leqslant \rho.$$

Upon passing to the limit as $k \to \infty$, $k \in K$, and using the lower semicontinuity of h, we obtain

$$\frac{h(y+\lambda z)-h(y)}{\lambda}\leqslant\rho.$$

Our choice of $\lambda>0$ was arbitrary, so the above inequality holds for all $\lambda>0$. This then implies that the *recession function* of h has a finite value at z (cf. Theorem 8.5 in [35]), a contradiction of the assumption that h is cofinite, i.e., the recession function of h has the value ∞ at all points except the origin [35, p. 116].

We next show that $\{h(y^k)\}$ is bounded. If $\{h(y^k)\}$ is not bounded, then there exists a subsequence K of $\{1,2,\ldots\}$ for which $\{h(y^k)\}_{k\in K}\to\infty$. Since $\{y^k\}$ is bounded, we will assume, by passing to a subsequence if necessary, that $\{y^k\}_{k\in K}$ converges to some y^∞. Then, $h(y^\infty)=\infty$ (cf. lower semicontinuity of h), so that y^∞ is different from y. By further passing to a subsequence if necessary, we can assume that there exists a scalar $\delta>0$ such that $\|y^k-y\|\geqslant\delta$ for all $k\in K$. Fix any $\beta\in(0,\delta)$ (so that $\|y^k-y\|>\beta$ for all $k\in K$) and define ξ^k, for all $k\in K$, as in (4.2). For each $k\in K$ we have, by using the convexity of h,

$$\begin{aligned}h(\xi^k)&\geqslant h(y^k)+(1-\beta/\|y-y^k\|)h'(y^k;y-y^k)\\&=(\beta/\|y-y^k\|)h(y^k)+(1-\beta/\|y-y^k\|)\big(h(y^k)+h'(y^k;y-y^k)\big),\end{aligned}$$

which together with the hypothesis that $\{h(y^k)+h'(y^k;y-y^k)\}$ is bounded from below implies that $\{h(\xi^k)\}_{k\in K}\to\infty$. This again contradicts the fact that, for β sufficiently small, $\{h(\xi^k)\}_{k\in K}$ is bounded. Finally, since $\{y^k\}$ and $\{h(y^k)\}$ are both bounded and the level sets of h are closed (since h is lower semicontinuous), every limit point of $\{y^k\}$ is in C_h. Q.E.D.

5. Convergence analysis for the strictly convex cost case. Let p^r denote the dual vector generated by the relaxation algorithm at the rth iteration ($r=0,1,2,\ldots$) and denote

(5.1) $$x^r=\nabla f^*(A^{\mathrm{T}}p^r).$$

We will show that, under either Assumption C1 or Assumption C2, the sequence $\{x^r\}$ converges to x^*, the optimal solution of (**P**). We will also consider conditions under which $\{p^r\}$ converges.

In what follows, let s^r denote the index of the coordinate relaxed at the rth iteration ($r=0,1,2,\ldots$) and let d^r denote the gradient of q at p^r. Thus, using (1.4)–(1.5) and (5.1), we have

(5.2) $$d^r=d(p^r)=b-Ax^r.$$

By (1.3), (1.5)–(1.6), and (5.1), we also have that, for $r=0,1,\ldots,$

(5.3) $$q(p^r)=f(x^r)+\langle p^r,b-Ax^r\rangle$$

and

(5.4) $$A^{\mathrm{T}}p^r\in\partial f(x^r).$$

Because the proof of convergence is quite complex, we have broken it into two parts. The first part comprises the following three technical lemmas, each of which holds independently of the order of coordinate relaxation (these lemmas will be used again in §6 when we modify Assumption C1).

LEMMA 3. *For $r = 0, 1, 2, \ldots,$ the following hold*:

$$q(p^{r+1}) - q(p^r) \geqslant f(x^{r+1}) - f(x^r) - f'(x^r; x^{r+1} - x^r), \tag{5.5}$$

$$f(x^r) + f'(x^r; x^* - x^r) \geqslant q(p^r). \tag{5.6}$$

PROOF. Consider a fixed index $r \in \{0, 1, \ldots\}$ and denote $s = s^r$. By using (5.3), we have

$$\begin{aligned} q(p^{r+1}) - q(p^r) &= f(x^{r+1}) + \langle p^{r+1}, b - Ax^{r+1} \rangle - f(x^r) - \langle p^r, b - Ax^r \rangle \\ &= f(x^{r+1}) - f(x^r) - \langle p^r, Ax^{r+1} - Ax^r \rangle \\ &\quad + \langle p^{r+1} - p^r, b - Ax^{r+1} \rangle \\ &= f(x^{r+1}) - f(x^r) - \langle A^{\mathrm{T}} p^r, x^{r+1} - x^r \rangle + \langle p^{r+1} - p^r, d^{r+1} \rangle \\ &= f(x^{r+1}) - f(x^r) - \langle A^{\mathrm{T}} p^r, x^{r+1} - x^r \rangle + \left(p_s^{r+1} - p_s^r \right) d_s^{r+1}, \end{aligned}$$

where the last equality follows from the fact that p^{r+1} and p^r differ only in their sth coordinate. From the description of the relaxation iteration in §3, we see that $(p_s^{r+1} - p_s^r) d_s^{r+1} \geqslant 0$ and therefore

$$q(p^{r+1}) - q(p^r) \geqslant f(x^{r+1}) - f(x^r) - \langle A^{\mathrm{T}} p^r, x^{r+1} - x^r \rangle.$$

Since $A^{\mathrm{T}} p^r \in \partial f(x^r)$ (cf. (5.4)), by (1.2) we have

$$f'(x^r; x^{r+1} - x^r) \geqslant \langle A^{\mathrm{T}} p^r, x^{r+1} - x^r \rangle,$$

and combining the last two inequalities we obtain

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

Similarly, we have (also using the fact $p^r \geqslant 0$ and the feasibility of x^* for **(P)**)

$$\begin{aligned} q(p^r) &= f(x^r) + \langle p^r, b - Ax^r \rangle \leqslant f(x^r) + \langle p^r, Ax^* - Ax^r \rangle \\ &= f(x^r) + \langle A^{\mathrm{T}} p^r, x^* - x^r \rangle \leqslant f(x^r) + f'(x^r; x^* - x^r). \end{aligned}$$ Q.E.D.

Since $\{q(p^r)\}$ is clearly nondecreasing, (5.6) shows that $\{f(x^r) + f'(x^r; x^* - x^r)\}$ is bounded from below. Then, by Lemma 2, we immediately have the following result:

LEMMA 4. *Both $\{x^r\}$ and $\{f(x^r)\}$ are bounded and every limit point of $\{x^r\}$ is in C_f.*

The next lemma follows from Lemmas 1 and 4.

LEMMA 5. *The following hold*:
(a) $\{x^{r+1} - x^r\} \to 0$.
(b) $\{[p_{s^r}^r + d_{s^r}^r]^+ - p_{s^r}^r\} \to 0$.

PROOF. (a) Suppose that $\{x^{r+1} - x^r\}$ does not tend to zero. Then there exist a scalar $\epsilon > 0$ and an infinite subsequence R of $\{0, 1, \ldots\}$ such that $\|x^{r+1} - x^r\| \geqslant \epsilon$ for all $r \in R$. Since $\{x^r\}$ and $\{x^{r+1}\}$ are bounded (cf. Lemma 4), by further passing to 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 4, both x^∞ and

$x^\infty + z$ are in C_f. This together with (5.5), Lemma 1(b), and the continuity of f on C_f yields

$$\liminf_{r\to\infty,\, r\in R} \{q(p^{r+1}) - q(p^r)\} \geqslant f(x^\infty + z) - f(x^\infty) - f(x^\infty; z).$$

Since $\{q(p^r)\}$ is nondecreasing and is bounded above (by the optimal dual value), $\{q(p^r)\}$ must converge to some real number, implying that the left side of the above inequality is zero. On the other hand, f is strictly convex, so the right-hand side of the above inequality is a positive scalar, a contradiction.

(b) Suppose that the claim does not hold. Then there exist a scalar $\epsilon > 0$, an $s \in \{1,\dots,n\}$, and an infinite subsequence R of $\{0,1,\dots\}$ such that

$$s^r = s \quad \text{and} \quad \left|[p_s^r + d_s^r]^+ - p_s^r\right| \geqslant \epsilon \quad \forall\, r \in R.$$

The latter condition implies that, for each $r \in R$, either (i) $d_s^r \geqslant \epsilon$ or (ii) $d_s^r \leqslant -\epsilon$, $p_s^r \geqslant \epsilon$. By the description of the relaxation iteration in §3, we see that, in case (i), $d_s^{r+1} \leqslant \delta d_s^r$ and, in case (ii), either $d_s^{r+1} \geqslant \delta d_s^r$ or $d_s^{r+1} < \delta d_s^r$, $p_s^{r+1} = 0$. If case (ii) holds and $d_s^{r+1} < \delta d_s^r$, $p_s^{r+1} = 0$, then (since $d_s(p)$ is nonincreasing in p_s) $d_s(p^r - \theta e^s) < \delta d_s^r \leqslant -\delta\epsilon$ for all $\theta \in [0, p_s^r]$, so that

$$q(p^{r+1}) - q(p^r) = -\int_0^{p_s^r} d_s(p^r - \theta e^s)\, d\theta \geqslant \delta\epsilon p_s^r \geqslant \delta(\epsilon)^2.$$

Otherwise, we have

$$\left|\sum_{j=1}^m a_{sj}\left(x_j^{r+1} - x_j^r\right)\right| = \left|d_s^r - d_s^{r+1}\right| \geqslant (1-\delta)|d_s^r| \geqslant (1-\delta)\epsilon,$$

so that

$$\|x^{r+1} - x^r\| \geqslant (1-\delta)\epsilon/\|A\|,$$

where $\|A\|$ denotes the L_2-norm of A, i.e., $\|A\| = \max_{\|x\|=1} \|Ax\|$. Since $\{x^{r+1} - x^r\} \to 0$ (cf. part (a)), the case $\|x^{r+1} - x^r\| \geqslant (1-\delta)\epsilon/\|A\|$ can occur for only a finite number of $r \in R$, so we must have $q(p^{r+1}) - q(p^r) \geqslant \delta(\epsilon)^2$ for all $r \in R$ sufficiently large. Since $q(p^r)$ is nondecreasing with r, this implies that $\{q(p^r)\} \to \infty$, which is a contradiction since the optimal dual value is finite. Q.E.D.

The next lemma, which is a direct consequence of Lemma 5, states that, under either Assumption C1 or Assumption C2, there holds $\{p^r - [p^r + d^r]^+\} \to 0$, thus showing that the necessary and sufficient optimality condition (1.7) is satisfied asymptotically.

LEMMA 6. *Under either Assumption* C1 *or Assumption* C2, $\{[p^r + d^r]^+ - p^r\} \to 0$.

PROOF. Suppose that Assumption C1 holds. Fix any $i \in \{1,\dots,n\}$ and any $k \in \{0,1,\dots\}$. Let r be the smallest integer greater than or equal to k such that $s^r = i$. Then $p_i^k = p_i^r$ and

$$|d_i^k - d_i^r| = \left|\sum_{h=k}^{r-1}\sum_{j=1}^m a_{ij}\left(x_j^{h+1} - x_j^h\right)\right|$$

$$\leqslant \|A\| \sum_{h=k}^{r-1} \|x^{h+1} - x^h\| \leqslant \|A\| \sum_{h=k}^{k+T-1} \|x^{h+1} - x^h\|,$$

where the last inequality follows from the fact $r - k \leq T$ (cf. Assumption C1). Hence

$$\left|[p_i^k + d_i^k]^+ - p_i^k\right| \leq \left|[p_i^k + d_i^k]^+ - [p_i^k + d_i^r]^+\right| + \left|[p_i^k + d_i^r]^+ - p_i^k\right|$$

$$\leq \left|d_i^k - d_i^r\right| + \left|[p_i^k + d_i^r]^+ - p_i^k\right|$$

$$\leq \|A\| \sum_{h=k}^{k+T-1} \|x^{h+1} - x^h\| + \left|[p_{s'}^r + d_{s'}^r]^+ - p_{s'}^r\right|,$$

where the first inequality follows from the triangle inequality and the second inequality follows from the nonexpansive property of the projection mapping $[\cdot]^+$. Since the choice of i and k was arbitrary, the above inequality holds for all $i \in \{1, \ldots, n\}$ and all $k \in \{0, 1, \ldots\}$. Then, by Lemma 5, we have

$$[p_i^k + d_i^k]^+ - p_i^k \to 0 \quad \text{as } k \to \infty, \forall\, i \in \{1, \ldots, n\}.$$

Now suppose that Assumption C2 holds. Then we have

$$\max_i \left|[p_i^r + d_i^r]^+ - p_i^r\right| \leq \left|[p_{s'}^r + d_{s'}^r]^+ - p_{s'}^r\right|/\omega \quad \forall\, r = 0, 1, \ldots,$$

and it follows from Lemma 5(b) that

$$\max_i \left|[p_i^r + d_i^r]^+ - p_i^r\right| \to 0 \quad \text{as } r \to \infty. \quad \text{Q.E.D.}$$

By using Lemmas 1, 4, and 6, we can now prove Proposition 1, which was stated at the end of §3.

Proof of Proposition 1. (a) Let x^∞ be any limit point of $\{x^r\}$ and let R be an infinite subsequence of $\{0, 1, \ldots\}$ for which $\{x^r\}_{r \in R} \to x^\infty$. By Lemma 4, $x^\infty \in C_f$. Denote $d^\infty = b - Ax^\infty$. Then, $\{d^r\}_{r \in R} \to d^\infty$, which, together with the facts $\{[p^r + d^r]^+ - p^r\}_{r \in R} \to 0$ (cf. Lemma 6) and $p^r \geq 0$ for all $r \in R$, implies

$$(5.7) \qquad d^\infty \leq 0 \quad \text{and} \quad \{p_i^r\}_{r \in R} \to 0 \quad \forall\, i \in I,$$

where $I = \{i | d_i^\infty < 0\}$. Therefore $x^\infty \in X$. Suppose that $x^\infty \neq x^*$ and let y be an element of $P \cap \mathrm{ri}(Q) \cap X$ (such a y exists by Assumption B). Fix any $\lambda \in (0, 1)$ and denote $y(\lambda) = \lambda y + (1 - \lambda)x^*$. Then $y(\lambda) \in P \cap \mathrm{ri}(Q) \cap X$ and $y(\lambda) \neq x^\infty$. By Lemma 1(a), there exists an $\epsilon > 0$ such that $C_f \cap B(x^\infty, \epsilon) = \mathrm{cl}(P) \cap Q \cap B(x^\infty, \epsilon)$. Since $\mathrm{cl}(P)$ is a polyhedral set and $y(\lambda) - x^\infty$ belongs to the tangent cone of $\mathrm{cl}(P)$ at x^∞, this implies that, for any fixed $\delta \in (0, \epsilon/\|z\|)$,

$$x^r + \delta z \in \mathrm{cl}(P) \quad \forall\, r \in R \text{ sufficiently large},$$

where $z = y(\lambda) - x^\infty$. Suppose that δ is further taken to be smaller than 1. Then, since $y(\lambda) \in \mathrm{ri}(Q)$, we have $x^\infty + \delta z \in \mathrm{ri}(Q)$, which together with the facts $x^r \in Q$, for all r, and $\{x^r\}_{r \in R} \to x^\infty$ yields

$$x^r + \delta z \in Q \quad \forall\, r \in R \text{ sufficiently large}.$$

By combining the above two expressions with the observation that, for all $r \in R$

sufficiently large, $x^r + \delta z \in B(x^\infty, \epsilon)$, we obtain

$$x^r + \delta z \in \mathrm{cl}(P) \cap Q \cap B(x^\infty, \epsilon) = C_f \cap B(x^\infty, \epsilon) \quad \forall\, r \in R \text{ sufficiently large.}$$

Since $x^\infty \in C_f$ and $x^\infty + \delta z \in C_f$, this together with Lemma 1(b) implies

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

Since $A^{\mathrm{T}}p^r \in \partial f(x^r)$ for all r (cf. (5.4)), we have from (1.2) that $f'(x^r; z) \geqslant \langle p^r, Az\rangle$, for all r, so that

$$f'(x^\infty; z) \geqslant \liminf_{r\to\infty,\, r\in R} \langle p^r, Az\rangle.$$

Now since $Ay(\lambda) \geqslant b$, we also have $\sum_{j=1}^m a_{ij} z_j \geqslant 0$ for all $i \notin I$, which together with the facts $p^r \geqslant 0$, for all $r \in R$, and $\{p_i^r\}_{r\in R} \to 0$, for all $i \in I$ (cf. (5.7)), implies

$$\liminf_{r\to\infty,\, r\in R} \langle p^r, Az\rangle \geqslant 0.$$

Hence $f'(x^\infty; z) \geqslant 0$ so 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 within C_f), we obtain that $f(x^\infty) \leqslant f(x^*)$. Since $x^\infty \in X$, this implies x^∞ is an optimal solution of **(P)**. Since x^* is the unique optimal solution of **(P)**, we have $x^\infty = x^*$, a contradiction.

(b) By Lemma 1(a), there exists some $\epsilon > 0$ for which $B(x^*, \epsilon) \cap C_f$ is closed, where $B(x^*, \epsilon)$ is the closed ball around x^* with radius ϵ. Since $\mathrm{cl}(C_f)$ is a polyhedral set and $\{x^r\} \to x^*$, there must exist some $\mu \in (0, \epsilon]$, such that $x^r - x^* \in K \cap B(0, \mu)$ for all r sufficiently large, where K denotes the tangent cone of C_f at x^*. K is closed since C_f is closed locally near x^* and K is a polyhedral set since $\mathrm{cl}(C_f)$ is a polyhedral set. 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 [35, §19], i.e.,

$$x^r - x^* = \sum_{k=1}^h \lambda_k^r w^k \quad \forall\, r \text{ sufficiently large,}$$

where $\lambda_1, \ldots, \lambda_h$ are nonnegative scalars tending to zero and $\{w^1, \ldots, w^h\}$ is some subset of the generators of K. By passing to a subsequence R if necessary we can assume that $\lambda_k^r > 0$ for all $r \in R$ and all k. Let W denote the $m \times h$ matrix whose kth column is w^k, let λ^r denote the vector in $\Re^h$ whose kth coordinate is λ_k^r, and consider the substitution

$$x^r = W\lambda^r + x^*, \qquad \tau^r = W^{\mathrm{T}}A^{\mathrm{T}}p^r. \tag{5.8}$$

Then, for all $r \in R$, we have

$$\langle p^r, A(x^r - x^*)\rangle = \langle \tau^r, \lambda^r\rangle = \sum_{k=1}^h \tau_k^r \lambda_k^r. \tag{5.9}$$

We claim that the above quantity tends to zero as $r \to \infty$, $r \in R$. To see this, consider the function $g\colon \Re^h \to (-\infty, \infty]$ given by $g(\lambda) = f(W\lambda + x^*)$. This function is convex, its tangent cone at the origin is the positive orthant in $\Re^h$, and its subdifferential ∂g is given by (cf. [35, Theorem 23.9])

$$\partial g(\lambda) = \{W^{\mathrm{T}}\eta \mid \eta \in \partial f(W\lambda + x^*)\}. \tag{5.10}$$

Since $A^Tp^r \in \partial f(x^r)$ for all r (cf. (5.4)), equations (5.8) and (5.10) imply $\tau^r \in \partial g(\lambda^r)$, for all $r \in R$, so that (using (1.2))

$$(5.11)\qquad -g'(\lambda^r; -e^k) \leqslant \tau_k^r \leqslant g'(\lambda^r; e^k), \qquad k = 1, \ldots, h,$$

where e^k denotes the kth coordinate vector in $\Re^h$. Fix any $k \in \{1, \ldots, h\}$ and let $y^r = x^r - \lambda_k^r w^k$ for all $r \in R$. Then, $y^r = \Sigma_{j \neq k} w^j \lambda_j^r + x^*$ (cf. (5.8)), so that $y^r \in B(x^*, \mu) \cap C_f$ for all $r \in R$ sufficiently large. Moreover, since $\lambda_k^r > 0$, we have $f'(x^r; -w^k/\|w^k\|) = f'(x^r; (y^r - x^r)/\|y^r - x^r\|)$ for all $r \in R$ sufficiently large. Then, from our assumption on f, we obtain that $f'(x^r; -w^k/\|w^k\|)$ is bounded for all $r \in R$ sufficiently large. Since (by a direct calculation using the definition of g) $g'(\lambda^r; -e^k) = f'(x^r; -w^k)$ for all $r \in R$, this implies that $g'(\lambda^r; -e^k)$ is bounded for all $r \in R$ sufficiently large. By an analogous argument, we also obtain that $g'(\lambda^r; e^k)$ is bounded for all $r \in R$ sufficiently large. Since the choice of k was arbitrary, this implies that both $-g'(\lambda^r; -e^k)$ and $g'(\lambda^r; e^k)$ are bounded, for $k = 1, \ldots, h$, as $r \to \infty$, $r \in R$. Then (5.11) (also using the fact $\{\lambda^r\}_{r \in R} \to 0$) shows that

$$\lim_{r \to \infty,\, r \in R} \sum_{k=1}^{h} \tau_k^r \lambda_k^r = 0,$$

which together with (5.9) yields

$$(5.12a)\qquad \lim_{r \to \infty,\, r \in R} \langle p^r, A(x^r - x^*) \rangle = 0.$$

Also, since $\{d^r\} \to b - Ax^*$ (cf. $\{x^r\} \to x^*$) and $\{[p^r + d^r]^+ - p^r\} \to 0$ (cf. Lemma 6), we have $\{p_i^r\} \to 0$ if $\Sigma_{j=1}^m a_{ij} x_j^* > b_i$, so that

$$(5.12b)\qquad \{\langle p^r, b - Ax^* \rangle\} \to 0.$$

Moreover, since $\{x^r\} \to x^*$ and f is continuous on C_f, we have

$$(5.12c)\qquad \{f(x^r)\} \to f(x^*).$$

Combining (5.12a)–(5.12c) with the fact (cf. (5.3))

$$\begin{aligned} q(p^r) &= f(x^r) + \langle p^r, b - Ax^r \rangle \\ &= f(x^r) + \langle p^r, b - Ax^* \rangle + \langle p^r, A(x^* - x^r) \rangle \quad \forall\, r, \end{aligned}$$

yields $\{q(p^r)\}_{r \in R} \to f(x^*)$. Since $q(p^r)$ is nondecreasing with r, this implies that $\{q(p^r)\} \to f(x^*)$ and part (b) is proven.

(c) We claim that the assumption of part (c) shows that the convex program **(P)** is *strictly consistent* (see [35, p. 300]). To show this, it suffices to verify that the effective domain of the convex bifunction associated with **(P)**, as derived in Appendix A, has a nonempty interior or, equivalently, the set

$$S = \{u \in \Re^n | b - Ax \leqslant u,\, f(x) < \infty \text{ for some } x\}$$

has a nonempty interior. By assumption, there exists an $\bar{x}$ such that $f(\bar{x}) < \infty$ and $A\bar{x} > b$. Since $A\bar{x} > b$, there exists a scalar $\delta > 0$ such that $b - A\bar{x} \leqslant u$ for all u satisfying $\|u\| \leqslant \delta$, so that S contains these u's and hence has a nonempty interior. Thus, **(P)** is strictly consistent and it follows from Corollary 29.1.5 of [35] that the level sets of the dual functional q are compact. Since $q(p^r)$ is nondecreasing with r, $\{p^r\}$ is bounded. By Lemma 6, every limit point p^∞ of $\{p^r\}$ satisfies the optimality conditions for (**D**), i.e., $p^\infty = [p^\infty + d(p^\infty)]^+$. Q.E.D.

6. Convergence analysis for the strongly convex cost case. In this section we further assume that f is strongly convex, in the sense that there exist scalars $\sigma > 0$ and $\gamma > 1$ such that

$$(6.1) \qquad f(y) - f(x) - f'(x; y - x) \geqslant \sigma \|y - x\|^\gamma \quad \forall\, x, y \in C_f.$$

For $\gamma = 2$, this notion of strong convexity reduces to the notion of *uniform convexity* considered in [31]. As an example, the function $f\colon \Re \to (-\infty, \infty]$ given by $f(x) = x^4$ if $x \geqslant 0$, and $f(x) = \infty$ otherwise, is strongly convex (with $\gamma = 4$, $\sigma = 1/4$), but not uniformly convex.

We consider the following assumption regarding the order of coordinate relaxation that is weaker than Assumption C1. Let $\{\tau_1, \tau_2, \dots\}$ be a sequence of integers given by the recursion:

$$(6.2\text{a}) \qquad \tau_1 = 0 \quad \text{and} \quad \tau_{k+1} = \tau_k + b_k, \qquad k = 1, 2, \dots,$$

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

$$(6.2\text{b}) \qquad b_k \geqslant n, \quad k = 1, 2, \dots, \quad \text{and} \quad \sum_{k=1}^{\infty} (b_k)^{1-\gamma} = \infty.$$

The assumption is as follows:

Assumption C3. *For $k = 1, 2, \dots$, every coordinate is chosen at least once for relaxation between iterations τ_k and $\tau_{k+1} - 1$.*

The condition $b_k \geqslant n$ for all k is required to allow each coordinate to be relaxed at least once between iterations τ_k and $\tau_{k+1} - 1$ so that Assumption C3 can be satisfied. Notice that if we choose b_k to be a fixed constant for all k, then Assumption C3 reduces to Assumption C1. On the other hand, if we choose b_k such that $\{b_k\} \to \infty$, then the length of a "cycle" $\{\tau_k, \tau_k + 1, \dots, \tau_{k+1} - 1\}$ tends to ∞ with k ($b_k = k^{1/(\gamma-1)}n$ is one such sequence).

Assumption C3 allows the time between successive relaxations of each coordinate to grow, but not too fast. Such an order of relaxation in some cases can be quite useful. For example, if the relaxation of a certain coordinate is much more expensive than that of the rest, then Assumption C3 allows this coordinate to be left out of the computation increasingly more often. We will show that, under (6.1) and Assumption C3, most of the conclusions of Proposition 1 hold. These results are of interest since they show that, for a broad class of problems, essentially cyclic is not the weakest possible order of relaxation for the dual coordinate ascent method to be convergent. To the best of our knowledge, the only other (dual) coordinate ascent methods that allow order of relaxation weaker than essentially cyclic are the methods considered in [4], [5], and [39] for separable cost problems.

Let p^r denote the dual vector generated by the relaxation algorithm at the rth iteration ($r = 0, 1, 2, \dots$) and let x^r and $d^r = b - Ax^r$ be defined by, respectively, (5.1) and (5.2). Notice that Lemmas 3 to 5 hold for the sequences $\{p^r\}$, $\{x^r\}$, and $\{d^r\}$. We will show that if (6.1) and Assumption C3 (in addition to Assumptions A and B) hold, then $\{x^r\}$ contains an infinite subsequence converging to the optimal solution of

(P) and, under additional regularity assumptions, $\{x^r\}$ and $\{p^r\}$ are themselves convergent. To prove this main result, we first need the following analog of Lemma 6:

LEMMA 7. *If* (6.1) *and Assumption* C3 *hold, then there exists an infinite subsequence* R *of* $\{0, 1, \dots\}$ *such that* $\{[p^r + d^r]^+ - p^r\}_{r \in R} \to 0$.

PROOF. Since $x^r \in C_f$ for all r, then (5.5) and (6.1) imply

$$q(p^{r+1}) - q(p^r) \geqslant \sigma \|x^{r+1} - x^r\|^\gamma, \qquad r = 0, 1, 2, \dots,$$

which, when summed over all r, yields

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

Hence

$$\sum_{r=0}^{\infty} \|x^{r+1} - x^r\|^\gamma < \infty. \tag{6.3}$$

(Note that (5.6) and (6.1) imply

(6.4)
$$f(x^*) - q(p^r) \geqslant f(x^*) - f(x^r) - f'(x^r; x^* - x^r) \geqslant \sigma \|x^* - x^r\|^\gamma \quad \forall\, r,$$

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

Next we show that there exists an infinite subsequence H of $\{1, 2, \dots\}$ for which

$$\sum_{r=\tau_h}^{\tau_{h+1}-1} \|x^{r+1} - x^r\| \to 0 \quad \text{as } h \to \infty,\ h \in H. \tag{6.5}$$

We will argue by contradiction. Suppose that such a subsequence does not exist. Then there must exist a positive scalar ϵ and an integer h' such that

$$\sum_{r=\tau_h}^{\tau_{h+1}-1} \|x^{r+1} - x^r\| \geqslant \epsilon \quad \forall\, h \geqslant h'. \tag{6.6}$$

Consider the Hölder inequality [15], which says that for any integer $N \geqslant 1$ and any $x, y \in \Re^N$, there holds $|\langle x, y\rangle| \leqslant \|x\|_\lambda \|y\|_\mu$, for any scalars $\lambda > 1$, $\mu > 1$ satisfying $1/\lambda + 1/\mu = 1$. ($\|\cdot\|_\lambda$ and $\|\cdot\|_\mu$ denote, respectively, the L_λ-norm and the L_μ-norm in $\Re^N$.) If $x \geqslant 0$ and if we let y be the vector in $\Re^N$ with entries all 1, we obtain

$$\sum_{i=1}^{N} x_i \leqslant \left[\sum_{i=1}^{N} (x_i)^\lambda\right]^{1/\lambda} (N)^{1/\mu}. \tag{6.7}$$

Applying (6.7) to the left-hand side of (6.6) with $\lambda = \gamma$ and $\mu = \gamma/(\gamma - 1)$ yields

$$\epsilon^\gamma \leqslant \left[\sum_{r=\tau_h}^{\tau_{h+1}-1} \|x^{r+1} - x^r\|^\gamma\right] (\tau_{h+1} - \tau_h)^{\gamma-1} \quad \forall\, h \geqslant h',$$

which implies

$$(6.8)\quad \epsilon^{\gamma} \sum_{h=h'}^{\infty} (\tau_{h+1} - \tau_h)^{1-\gamma} \leqslant \sum_{h=h'}^{\infty} \left[\sum_{r=\tau_h}^{\tau_{h+1}-1} \|x^{r+1} - x^r\|^{\gamma} \right] = \sum_{r=\tau_{h'}}^{\infty} \|x^{r+1} - x^r\|^{\gamma}.$$

The leftmost quantity of (6.8), according to (6.2a)–(6.2b), is ∞ while the rightmost quantity of (6.8), according to (6.3), is bounded—a clear contradiction.

Now, let H be an infinite subsequence of $\{1, 2, \dots\}$ for which (6.5) holds. Fix any coordinate index $i \in \{1, \dots, n\}$ and any $h \in H$. Let r be the smallest integer greater than or equal to τ_h such that $s^r = i$. Then $p_i^{\tau_h} = p_i^r$ and

$$|d_i^{\tau_h} - d_i^r| = \left| \sum_{k=\tau_h}^{r-1} \sum_{j=1}^{m} a_{ij} \left(x_j^{k+1} - x_j^k \right) \right|$$

$$\leqslant \|A\| \sum_{k=\tau_h}^{r-1} \|x^{k+1} - x^k\| \leqslant \|A\| \sum_{k=\tau_h}^{\tau_{h+1}-1} \|x^{k+1} - x^k\|,$$

where the last inequality follows from the fact $r \leqslant \tau_{h+1} - 1$ (cf. Assumption C3). Hence

$$\left| [p_i^{\tau_h} + d_i^{\tau_h}]^+ - p_i^{\tau_h} \right| \leqslant \left| [p_i^{\tau_h} + d_i^{\tau_h}]^+ - [p_i^{\tau_h} + d_i^r]^+ \right| + \left| [p_i^{\tau_h} + d_i^r]^+ - p_i^{\tau_h} \right|$$

$$\leqslant \left| d_i^{\tau_h} - d_i^r \right| + \left| [p_i^{\tau_h} + d_i^r]^+ - p_i^{\tau_h} \right|$$

$$\leqslant \|A\| \sum_{k=\tau_h}^{\tau_{h+1}-1} \|x^{k+1} - x^k\| + \left| [p_{s^r}^r + d_{s^r}^r]^+ - p_{s^r}^r \right|,$$

where the first inequality follows from the triangle inequality and the second inequality follows from the nonexpansive property of the projection mapping $[\cdot]^+$. Since the choice of i and k was arbitrary, the above inequality holds for all $i \in \{1, \dots, n\}$ and all $k \in H$. Then, by (6.5) and the fact $\{[p_{s^r}^r + d_{s^r}^r]^+ - p_{s^r}^r\} \to 0$ (cf. Lemma 5(b)), we have

$$\left\{ [p_i^{\tau_h} + d_i^{\tau_h}]^+ - p_i^{\tau_h} \right\}_{h \in H} \to 0 \quad \forall\, i \in \{1, \dots, n\}.$$

Therefore, $\{[p^{\tau_h} + d^{\tau_h}]^+ - p^{\tau_h}\}_{h \in H} \to 0$. Q.E.D.

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

PROPOSITION 2. *If the strong convexity condition* (6.1) *and Assumption* C3 *are satisfied, then the following hold*:

(a) $\{x^r\}_{r \in R} \to x^*$ *for some infinite subsequence* R *of* $\{0, 1, \dots\}$.

(b) *If* $\text{cl}(C_f)$ *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* $x, y \in B \cap C_f$, *then* $\{q(p^r)\} \to f(x^*)$ *and* $\{x^r\} \to x^*$.

(c) *If* (*interior of* X) $\cap\, C_f \neq \varnothing$, *then* $\{q(p^r)\} \to f(x^*)$, $\{x^r\} \to x^*$, *and* $\{p^r\}$ *is bounded and every one of its limit points is an optimal solution of* (**D**).

PROOF. (a) By Lemma 7, there exists an infinite subsequence R of $\{0, 1, \dots\}$ such that $\{[p^r + d^r]^+ - p^r\}_{r \in R} \to 0$. Let x^∞ be any limit point of the subsequence $\{x^r\}_{r \in R}$. Then, by using a proof analogous to that for part (a) of Proposition 1, we obtain $x^\infty = x^*$.

(b) We use part (a) and an argument analogous to that for part (b) of Proposition 1 to conclude that $\{q(p^r)\} \to f(x^*)$. Then (6.4) yields $\{x^r\} \to x^*$.

(c) We first apply the argument for part (c) of Proposition 1 to conclude that $\{p^r\}$ is bounded. Let R be the subsequence constructed in part (a). Then, $\{p^r\}_{r \in R}$ has a limit point, say p^∞, which is an optimal dual solution (since $p^r \geq 0$ for all $r \in R$ and, by Lemma 7, $\{[p^r + d^r]^+ - p^r\}_{r \in R} \to 0$). Therefore, $q(p^\infty) = f(x^*)$ 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 solution. 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 $\{x^r\}$. Only for the special case where f is separable has it been shown that $\{x^r\} \to x^*$, assuming only that (6.1) and Assumption C3 hold [39].

Appendix A. In this appendix, we show that (under Assumptions A and B) the optimal value of **(P)** equals the optimal value of **(D)**.

We denote by F the *convex bifunction associated with* **(P)** [35, p. 293], i.e.,

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

Then F^*, the *adjoint* of F [35, p. 309], is given by

$$(F^*y)(p) = \inf_{u,x} \{(Fu)(x) - \langle y, x\rangle + \langle p, u\rangle\}$$

$$= \begin{cases} \inf_x \{f(x) - \langle y, x\rangle + \langle p, b - Ax\rangle\} & \text{if } p \geq 0, \\ -\infty & \text{otherwise}, \end{cases}$$

so that the *concave program associated with* F^*, i.e., $\max_p (F^*0)(p)$, is exactly **(D)**. Then, in the terminology of [35, p. 311], **(D)** is the program *dual* to **(P)** and, according to Theorem 30.4(i) in [35], **(P)** and **(D)** have the same optimal value if F is lower semicontinuous as a function of (x, u) (or *closed* in the terminology of [35]) and **(P)** has a unique optimal solution. Thus it suffices to show that F is lower semicontinuous or, equivalently, that the set

$$\{(x,u,z) | (Fu)(x) \leq z\} = \{(x,u,z) | f(x) \leq z, b - Ax \leq u\}$$

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

Appendix B. In this appendix we show that, under the regularity assumption $\mathrm{ri}(C_f) \cap X \neq \varnothing$ (in addition to Assumptions A and B), the relaxation iteration with $\delta = 0$ (i.e., exact line search) is well defined.

To show this, it suffices to show that, for any dual vector $p \geq 0$ and any coordinate p_s, we can, by adjusting p_s (with the other coordinates held fixed), obtain a dual vector $p' \geq 0$ for which either $d_s(p') = 0$ or $p'_s = 0$ and $d_s(p') < 0$. Choose any such p and p_s and consider the following minimization problem in the vector x

$$\text{(B.1)} \qquad \begin{array}{ll} \text{Minimize} & f(x) - \sum_{t \neq s} p_t \langle a^t, x\rangle \\ \text{subject to} & \langle a^s, x\rangle \geq b_s, \end{array}$$

where a^i denotes the ith column of A^T. Since the cost function of (B.1) is the sum of f and a linear function, then, like f, it has compact level sets. Moreover, its effective domain equals that of f, so that, by using $\text{ri}(C_f) \cap X \neq \emptyset$ and the observation that the constraint set for (B.1), i.e., $\{x | \langle a^s, x\rangle \geq b_s\}$, is contained in X, we have that the relative interior of its effective domain makes a nonempty intersection with the constraint set of (B.1). It then follows that (B.1) has an optimal solution (which, by strict convexity of f, is unique) and there exists an optimal Lagrange multiplier associated with the constraint $\langle a^s, x\rangle \geq b_s$ (cf. Theorem 28.2 in [35]). Let x' denote this optimal solution and let λ denote the optimal Lagrange multiplier. Then x' and λ satisfy the Kuhn-Tucker conditions for (B.1), i.e.,

$$\lambda a^s \in \partial f(x') - \sum_{i \neq s} p_i a^i, \quad \langle a^s, x'\rangle \geq b_s, \quad \lambda \geq 0, \quad \lambda(\langle a^s, x'\rangle - b_s) = 0.$$

Let p' be the dual vector whose ith coordinate is λ if $i = s$ and is p_i otherwise. Then we have

$$A^T p' \in \partial f(x'), \tag{B.2a}$$

$$\langle a^s, x'\rangle \geq b_s, \quad p' \geq 0, \quad p'_s(\langle a^s, x'\rangle - b_s) = 0. \tag{B.2b}$$

By using (1.5)–(1.6), we obtain from (B.2a) that $x' = \nabla f^*(A^T p')$, so that, by (1.4), $d_s(p') = b_s - \langle a^s, x'\rangle$. This together with (B.2b) yields $d_s(p') \leq 0$, $p' \geq 0$, and $p'_s = 0$ if $d_s(p') < 0$.

Acknowledgements. We gratefully acknowledge the comments made by two referees, which led to significant improvement in the presentation of this paper.

References

[1] Bachem, A. and Korte, B. (1979). On the RAS-Algorithm. *Computing* **23** 189–198.

[2] ______. (1980). Minimum Norm Problems over Transportation Polytopes. *Linear Algebra Appl.* **31** 103–118.

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

[4] ______ and El Baz, D. (1987). Distributed Asynchronous Relaxation Methods for Convex Network Flow Problems. *SIAM J. Control Optim.* **25** 74–85.

[5] ______, Hosein, P. A. and Tseng, P. (1987). Relaxation Methods for Network Flow Problems with Convex Arc Costs. *SIAM J. Control Optim.* **25** 1219–1243.

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

[7] Bregman, L. M. (1967). The Relaxation Method of Finding the Common Point Convex Sets and its Application to the Solution of Problems in Convex Programming. *U.S.S.R. Comput. Math. and Math. Phys.* **7** 200–217.

[8] Censor, Y. and Lent, A. (1981). An Iterative Row-Action Method for Interval Convex Programming. *J. Optim. Theory Appl.* **34** 321–352.

[9] Cottle, R. W., Duvall, S. G. and Zikan, K. (1986). A Lagrangean Relaxation Algorithm for the Constrained Matrix Problem. *Naval Res. Logist. Quart.* **33** 55–76.

[10] ______ and Pang, J.-S. (1982). On the Convergence of a Block Successive Over-Relaxation Method for a Class of Linear Complementarity Problems. *Math. Programming Stud.* 17, North-Holland, Amsterdam, 126–138.

[11] Cryer, C. W. (1971). The Solution of a Quadratic Programming Problem Using Systematic Overrelaxation. *SIAM J. Control Optim.* **9** 385–392.

[12] Darroch, J. N. and Ratcliff, D. (1972). Generalized Iterative Scaling for Log-Linear Models. *Ann. Math. Statist.* **43** 1470–1480.

[13] Dembo, R. S., Mulvey, J. M. and Zenios, S. A. (1989). Large-Scale Nonlinear Network Models and Their Application. *Oper. Res.* **37** 353–372.

[14] D'Esopo, D. A. (1959). A Convex Programming Procedure. *Naval Res. Logist. Quart.* **6** 33–42.

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

[16] Han, S. P. (1988). A Successive Projection Method. *Math. Programming* **40** 1–14.

[17] Herman, G. T. and Lent, A. (1978) A Family Of Iterative Quadratic Optimization Algorithms for Pairs of Inequalities, with Application in Diagnostic Radiology. *Math Programming Study* 9, North-Holland, Amsterdam, 15–29.

[18] Hildreth, C. (1957). A Quadratic Programming Procedure. *Naval Res. Logist. Quart.* **4** 79–85; see also (1957). Erratum, *Naval Res. Logist. Quart.* **4** 361.

[19] Jaynes, E. T. (1982). On the Rationale of Maximum Entropy Methods, *Proc. IEEE* **70** 939–952.

[20] Johnson, R. and Shore, J. E. (1984) Which is the Better Entropy Expression for Speech Processing: $-S \log S$ or $\log S$? *IEEE Trans. Acoust. Speech Signal Process.* **ASSP-32** 129–136.

[21] Klincewicz, J. G (1983). A Newton Method for Convex Separable Network Flow Problems *Networks* **13** 427–442.

[22] Kruithof, J. (1937) Calculation of Telephone Traffic, *De Ingenieur (E. Elektrotechnik 3)*, **52** E15–E25.

[23] Lamond, B. and Stewart, N. F. (1981). Bregman's Balancing Method. *Transportation Res. Part B* **15B** 239–248.

[24] Lent, A. (1977). 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., 249–257.

[25] ______ and Censor, Y. (1980). Extensions of Hildreth's Row-Action Method for Quadratic Programming. *SIAM J. Control Optim.* **18** 444–454.

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

[27] Luenberger, D. G. (1984). *Linear and Nonlinear Programming*. Addison-Wesley, Reading, MA.

[28] Luo, Z.-Q. and Tseng, P. (1989). On the Convergence of a Matrix Splitting Algorithm for the Symmetric Linear Complementarity Problem. LIDS Report P-1884, M.I.T., Cambridge, MA: *SIAM J. Control Optim.* (to appear).

[29] Mangasarian, O. L. (1984). Sparsity-Preserving SOR Algorithms for Separable Quadratic and Linear Programming. *Comput. Oper. Res.* **11** 105–112.

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

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

[32] Pang, J -S. (1984). On the Convergence of Dual Ascent Methods for Large-Scale Linearly Constrained Optimization Problems. School of Management, The University of Texas at Dallas, TX.

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

[34] Powell, M. J. D. (1973). On Search Directions for Minimization Algorithms, *Math. Programming* **4** 193–201.

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

[36] ______. (1984). *Network Flows and Monotropic Optimization*. Wiley-Interscience, New York.

[37] Sargent, R. W. H and Sebastian, D. J. (1973). On the Convergence of Sequential Minimization Algorithms. *J. Optim. Theory Appl.* **12** 567–575.

[38] Stern, T. E. (1977). A Class of Decentralized Routing Algorithms Using Relaxation *IEEE Trans. Comm* **COM-25** 1092–1102.

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

[40] ______. (1990). Dual Ascent Methods for Problems with Strictly Convex Costs and Linear Constraints. A Unified Approach. *SIAM J. Control Optim.* **28** 214–242.

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

[42] Zenios, S. A. and Mulvey, J. M. (1986). Relaxation Techniques for Strictly Convex Network Problems. In C L. Monma (Ed.), *Ann. Oper. Res.* 5: *Algorithms and Software for Optimization* Baltzer, Switzerland, 517–538.

[43] ______ and Lasken, R (1988). Nonlinear Network Optimization on a Massively Parallel Connection Machine. In R. R. Meyer and S. A. Zenios (Eds.), *Ann. Oper. Res.* 14: *Parallel Optimization on Novel Computer Architectures*. Baltzer, Switzerland, 147–165.

LABORATORY FOR INFORMATION AND DECISION SYSTEMS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE, MASSACHUSETTS 02139