Mathematical Programming 38 (1987) 303-321
North-Holland

RELAXATION METHODS FOR PROBLEMS WITH STRICTLY CONVEX SEPARABLE COSTS AND LINEAR CONSTRAINTS

Paul TSENG and Dimitri P. BERTSEKAS

Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Received 16 June 1986
Revised manuscript received 17 February 1987

We consider the minimization problem with strictly convex, possibly nondifferentiable, separable cost and linear constraints. The dual of this problem is an unconstrained minimization problem with differentiable cost which is well suited for solution by parallel methods based on Gauss-Seidel relaxation. We show that these methods yield the optimal primal solution and, under additional assumptions, an optimal dual solution. To do this it is necessary to extend the classical Gauss-Seidel convergence results because the dual cost may not be strictly convex, and may have unbounded level sets.

Key words: Gauss-Seidel relaxation, Fenchel duality, strict convexity, strong convexity.

1. Introduction

We consider the problem

$$\text{minimize} \quad f(x)=\sum_{j=1}^{m} f_j(x_j) \tag{1}$$

$$\text{subject to} \quad Ex=0$$

where x is the vector in R^m with coordinates denoted x_j, $j=1,2,\ldots,m$, $f_j: R\to(-\infty,\infty]$, and E is a $n\times m$ matrix with elements denoted e_{ij}, $i=1,\ldots,n, j=1,\ldots,m$. We make the following standing assumptions on f_j:

Assumption A. Each f_j is strictly convex, lower semicontinuous, and there exists at least one feasible solution for (1), i.e. the set

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

and the constraint subspace

$$C=\{x \mid Ex=0\}. \tag{2}$$

have a nonempty intersection.

Work supported by the National Science Foundation under grant NSF-ECS-3217668.

Assumption B. The conjugate convex function of f_j defined by

$$g_j(t_j) = \sup_{x_j}\{t_j x_j - f_j(x_j)\} \tag{3}$$

is real valued, i.e. $-\infty < g_j(t_j) < +\infty$ for all $t_j \in R$.

It is easily seen that Assumptions A and B imply that for every t_j there is some x_j with $f_j(x_j) < \infty$ attaining the supremum in (3), and furthermore

$$\lim_{|x_j| \to +\infty} f_j(x_j) = +\infty.$$

It follows that the cost function of (1) has bounded level sets, and therefore (using also the lower semicontinuity and strict convexity of f) *there exists a unique optimal solution to* (1).

Note that, because f_j is extended real valued, upper and lower bound constraints on the variables x_j can be incorporated into f_j by letting $f_j(x_j) = +\infty$ whenever x_j lies outside these bounds. We denote by

$$l_j = \inf\{\xi \mid f_j(\xi) < +\infty\},$$

$$c_j = \sup\{\xi \mid f_j(\xi) < +\infty\}$$

the lower and upper bounds on x_j implied by f_j. Note also that by introducing additional variables it is possible to convert linear manifold constraints of the form $Ax = b$ into a subspace constraint such as the one of (1). We assume a subspace rather than a linear manifold constraint because this simplifies notation and leads to a symmetric duality theory [11].

A dual problem for (1) is

$$\begin{aligned} &\text{minimize} \quad q(p) \\ &\text{subject to} \quad \text{no constraint on } p \end{aligned} \tag{4}$$

where q is the dual functional given by

$$q(p) = \sum_{j=1}^{m} g_j(E_j^{\mathrm{T}} p).$$

E_j denotes the jth column of E, and T denotes transpose. We refer to p as the price vector and to its coordinates p_i as *prices*. The duality between problems (1) and (4) can be developed either by viewing p_i as the Lagrange multiplier associated with the ith equation of the system $Ex = 0$, or via Fenchel's duality theorem. It is explored extensively in [11], where it is shown that, under Assumption A, there is no duality gap in the sense that the primal and dual optimal costs are opposites of each other. It is shown in [10, p. 337–338] that a vector $x = \{x_j \mid j = 1, \ldots, m\}$ satisfying $Ex = 0$ is optimal for (1) and a price vector $p = \{p_i \mid i = 1, \ldots, n\}$ is optimal for (4) if and only if

$$f_j^-(x_j) \le E_j^{\mathrm{T}} p \le f_j^+(x_j), \qquad j = 1, \quad , m \tag{6}$$

where $f_j^-(x_j)$ and $f_j^+(x_j)$ denote the left and right derivatives of f_j at x_j (see Fig. 1). These derivatives are defined in the usual way for x_j belonging to (l_j, c_j). When $-\infty < l_j < c_j$ we define

$$f_j^+(l_j) = \lim_{\xi \downarrow l_j} f_j^+(\xi), \qquad f_j^-(l_j) = -\infty.$$

When $l_j < c_j < +\infty$ we define

$$f_j^-(c_j) = \lim_{\xi \uparrow c_j} f_j^-(\xi), \qquad f_j^+(c_j) = +\infty.$$

Finally when $l_j = c_j$ we define $f_j^-(l_j) = -\infty, f_j^+(c_j) = +\infty$. Because of the strict convexity assumed in Assumption A, the conjugate function g_j is continuously differentiable and its gradient denoted $\nabla g_j(t_j)$ is the unique x_j attaining the supremum in (3) (see [10], p. 218, 253]), i.e.

$$\nabla g_j(t_j) = \arg \sup_{x_j} \{t_j x_j - f_j(x_j)\}. \tag{7}$$

Note that $\nabla g_j(t_j)$, being the gradient of a differentiable convex function, is continuous and monotonically nondecreasing. Since (6) is equivalent to $E_j^{\mathrm{T}} p$ being a subgradient of f_j at x_j, it follows in view of (7), that (6) is equivalent to

$$x_j = \nabla g_j(E_j^{\mathrm{T}} p) \quad \forall j = 1, 2, \ldots, m. \tag{8}$$

Any one of the two equivalent relations (6) and (8) is referred to as the *Complementary Slackness condition.*

The differentiability of q [cf. (5)] motivates a coordinate descent method of the Gauss–Seidel relaxation type for solving (4) whereby, given a price vector p, a coordinate p_i such that $\partial q(p)/\partial p_i > 0$ (<0) is chosen and p_i is decreased (increased) in order to decrease the dual cost. One then repeats the procedure iteratively. One important advantage of such a coordinate relaxation method is its suitability for parallel implementation on problems where E has special structure. To see this note, from (5), that two prices p_i and p_k are uncoupled, and can be iterated upon (relaxed) simultaneously if there is no column index j such that $e_{ij} \neq 0$ and $e_{kj} \neq 0$. For example when E is the node-arc incidence matrix of a directed network this

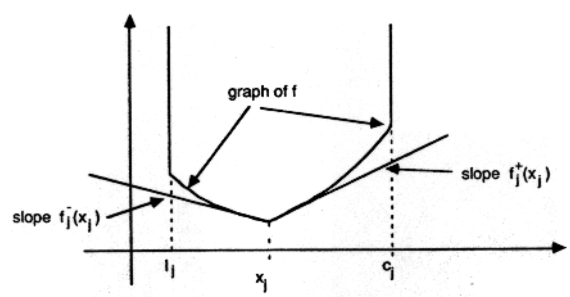

Fig. 1. The left and right derivatives of f_j.

translates to the condition that nodes i and k are not joined by an arc j. Computational testing conducted by Zenios and Mulvey [16] on network problems showed that such a synchronous parallelization scheme can improve the solution time many-fold.

Convergence of the Gauss-Seidel method for differentiable optimization has been well studied [6, 8, 12, 14, 15]. However it has typically been assumed that the cost function is strictly convex and has compact level sets, that exact line search is done during each descent, and that the coordinates are relaxed in an essentially cyclical manner. The strict convexity assumption is relaxed in [14] but the proof used there assumes that the algorithmic map associated with exact line search over the interval $(-\infty, \infty)$ is closed. Powell [9] gave an example of nonconvergence for a particular implementation of the Gauss-Seidel method, which is effectively a counterexample to the closure assertion, and shows that strict convexity is in general a required assumption. For our problem (4) the dual functional q is not strictly convex and it does not necessarily have bounded level sets. Indeed the dual problem (4) need not have an optimal solution. One of the contributions of this paper is to show that, under quite weak assumptions, the Gauss-Seidel method applied to (4) generates a sequence of primal vectors converging to the optimal solution for (1) and a sequence of dual costs that converges to the optimal cost for (4). The assumptions permit the line search to be done approximately and require that either (i) the coordinates are relaxed in an essentially cyclical manner or (ii) the primal cost is strongly convex. For case (ii) a certain mild restriction regarding the order of relaxation is also required. The result on convergence to the optimal primal solution (regardless of convergence to an optimal dual solution) is similar in flavor to that obtained by Pang [7] for problems whose primal cost is not necessarily separable. However his result further requires that the primal cost is differentiable and strongly (rather than strictly) convex, that the coordinates are relaxed in a cyclical manner, and that each line search is done exactly. The results of this paper extend also those obtained for separable strictly convex network flow problems in [2], where convergence to optimal primal and dual solutions is shown without any assumption on the order of relaxation. References [2] and [16] contain computational results with the relaxation method of this paper applied to network problems. Reference [1] explores convergence for network problems in a distributed asynchronous framework.

2. Algorithm description

The ith partial derivative of the dual cost (5) is denoted by $d_i(p)$. We have

$$d_i(p) = \frac{\partial q(p)}{\partial p_i} = \sum_{j=1}^{m} e_{ij} \nabla g_j(E_j^{\mathrm{T}} p), \quad i = 1, 2, \quad , n. \tag{9}$$

Since $d_i(p)$ is a partial derivative of a differentiable convex function we have that *$d_i(p)$ is continuous and monotonically nondecreasing in the ith coordinate.* Note from

(8), (9) that if x and p satisfy Complementary Slackness then

$$d(p) = \nabla q(p) = Ex. \tag{10}$$

We now define a Gauss-Seidel type of method whereby at each iteration a coordinate p_s with positive (negative) $d_s(p)$ is chosen and p_s is decreased (increased) in order to decrease the dual cost $q(p)$. We initially choose a fixed scalar δ in the interval (0, 1) which controls the accuracy of line search. Then we execute repeatedly the relaxation iteration described below.

Relaxation Iteration

If $d_i(p) = 0 \, \forall i$ then STOP.

Else

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

If $\beta = 0$, do nothing.

If $\beta > 0$, then decrease p_s so that $0 \leqslant d_s(p) \leqslant \delta\beta$.

If $\beta < 0$, then increase p so that $0 \geqslant d_s(p) \geqslant \delta\beta$.

Each relaxation iteration is well defined, in the sense that every step in the iteration is executable. To see this note that if $d_s(p) > 0$ and there does not exist a Δ ($\Delta > 0$) such that $d_s(p - \Delta e_s) \leqslant \delta\beta$, where e_s denotes the s-th coordinate vector, then using the definition of d and the fact that

$$\lim_{\eta \to \infty} \nabla g_j(\eta) = c_j, \quad \lim_{\eta \to -\infty} \nabla g_j(\eta) = l_j, \qquad j = 1, 2, \quad , m,$$

we have (cf. (8), (9))

$$\lim_{\Delta \to \infty} d_s(p - \Delta e_s) = \sum_{e_{sj} > 0} e_{sj} l_j + \sum_{e_{sj} < 0} e_{sj} c_j \geqslant \delta\beta > 0.$$

On the other hand for every x satisfying the constraint $Ex = 0$ we have

$$0 = \sum_{e_{sj} > 0} e_{sj} x_j + \sum_{e_{sj} < 0} e_{sj} x_j \geqslant \sum_{e_{sj} > 0} e_{sj} l_j + \sum_{e_{sj} < 0} e_{sj} c_j$$

which contradicts the previous relation. An analogous argument can be made for the case where $d_s(p) < 0$. The appendix provides an implementation of the approximation line search of the relaxation iteration.

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

Assumption C. There exists a positive integer T such that every coordinate is chosen at least once for relaxation between iterations r and $r + T$, for $r = 0, 1, 2, \ldots$

Assumption C is more general than the usual assumption that the order in which the coordinates are relaxed is cyclical. We will weaken this assumption later.

3. Convergence analysis

We will first show under Assumption C that by successively executing the relaxation iteration we generate a sequence of primal vectors that converges to the optimal primal solution, and a sequence of dual costs that converges to the optimal dual cost.

The line of argument that we will use is as follows: We first show through a rather technical argument that the sequence of primal vectors is bounded. Then we show that if the sequence of primal vectors does not approach the constraint subspace C, we can bound from below the amount of improvement in the dual functional q per iteration by a positive quantity whose sum over all iterations tends to infinity. It follows that the optimal dual cost has a value of $-\infty$, a contradiction of Assumption A. Thus each limit point of the primal vector sequence by the above argument must be primal feasible which together with the fact that Complementary Slackness is maintained at all iterations imply that each limit point is necessarily optimal. Convergence to the optimal primal solution then follows from the uniqueness of the solution.

We will denote the price vector generated by the method at the rth iteration by p^r, $r=0,1,2,\ldots$ (p^0 is the initial price vector) and the index of the coordinate relaxed at the rth iteration by s^r, $r=0,1,2,\ldots$ To simplify the presentation we denote

$$t_j^r = E_j^{\mathrm{T}} p^r,$$

$$x_j^r = \nabla g_j(t_j^r),$$

and.by t^r and x^r the vectors with coordinates t_j^r and x_j^r respectively. Note that from (9) and (10) we have

$$\nabla q(p^r) = Ex^r$$

so that the dual gradient sequence $\nabla q(p^r)$ approaches zero if and only if the primal vector sequence x^r approaches primal feasibility. We develop our convergence result through a sequence of lemmas the first of which provides a lower bound to the dual cost improvement at each iteration. (Note from (6) that t_j^r is a subgradient of f_j at x_j^r, so the right side of (11) below is nonnegative.)

Lemma 1. *We have, for all r,*

$$q(p^r) - q(p^{r+1}) \geq \sum_{j=1}^{m} [f_j(x_j^{r+1}) - f_j(x_j^r) - (x_j^{r+1} - x_j^r)t_j^r], \quad r=0,1,2, \tag{11}$$

with equality holding if exact line minimization is used ($d_{s^r}(p^r)=0$).

Proof. From (3), (5), and (7) we have

$$q(p^r) = \sum_{j=1}^{m} [x_j^r t_j^r - f_j(x_j^r)] \qquad r=0,1,2,$$

Consider a fixed index $r \geq 0$. Denote $s = s^r$ and $\Delta = p_s^{r+1} - p_s^r$. Then

$$\begin{aligned}
q(p^r) - q(p^{r+1}) &= \sum_{j=1}^{m} [x_j^r t_j^r - f_j(x_j^r)] - \sum_{j=1}^{m} [t_j^{r+1} x_j^{r+1} - f_j(x_j^{r+1})] \\
&= \sum_{j=1}^{m} [x_j^r t_j^r - f_j(x_j^r)] - \sum_{j=1}^{m} [(t_j^r + e_{sj}\Delta) x_j^{r+1} - f_j(x_j^{r+1})] \\
&= \sum_{j=1}^{m} [f_j(x_j^{r+1}) - f_j(x_j^r) - (x_j^{r+1} - x_j^r) t_j^r - e_{sj}\Delta x_j^{r+1}] \\
&= \sum_{j=1}^{m} [f_j(x_j^{r+1}) - f_j(x_j^r) - (x_j^{r+1} - x_j^r) t_j^r] - \Delta \sum_{j=1}^{m} e_{sj} x_j^{r+1} \\
&= \sum_{j=1}^{m} [f_j(x_j^{r+1}) - f_j(x_j^r) - (x_j^{r+1} - x_j^r) t_j^r] - \Delta d_s(p^{r+1}).
\end{aligned}$$

Since $\Delta d_s(p^{r+1}) \leq 0$ (and $d_s(p^{r+1}) = 0$ if we use exact line minimization) (11) follows. $\square$

For notational simplicity let us denote

$$d_i^r = d_i(p^r) = \sum_{j=1}^{m} e_{ij} \nabla g_j(t_j^r)$$

and denote by d^r the vector with coordinates d_i^r. Also we denote the orthogonal complement of C by $C^\perp$, i.e.

$$C^\perp = \{t \mid t = E^{\mathrm{T}} p \text{ for some } p\}.$$

For each x and z in R^m, we denote the directional derivative of f at x in the direction z by $f'(x; z)$, i.e.,

$$f'(x; z) = \lim_{\mu \downarrow 0} \frac{f(x + \mu z) - f(x)}{\mu}.$$

Similarly, for each p and u in R^n, we denote

$$q'(p; u) = \lim_{\lambda \downarrow 0} \frac{q(p + \lambda u) - q(p)}{\lambda}.$$

We will next show that the sequence $\{d^r\}$ is bounded. For this we will require the following lemma:

Lemma 2. *If each coordinate of t^r either tends to ∞, or tends to $-\infty$, or is bounded, then there exists a vector v in $C^\perp$ such that*

$$\begin{aligned}
&v_j > 0 \quad \forall j \text{ such that } t_j^r \to \infty, \\
&v_j < 0 \quad \forall j \text{ such that } t_j^r \to -\infty, \\
&v_j = 0 \quad \forall j \text{ such that } t_j^r \text{ is bounded.}
\end{aligned}$$

Proof. If each coordinate of t^r is bounded as r tends to ∞ then we can trivially take $v=0$. If each coordinate of t^r either tends to ∞ or tends to $-\infty$ then we can take v to be any t^r with r sufficiently large. Otherwise there exists an index j such that t_j^r tends to either ∞ or $-\infty$ and an index j such that t_j^r is bounded. Let J denote the nonempty set of j's such that t_j^r is bounded. For each fixed r consider the solution of the following system of linear equations in π and τ

$$\tau = E^{\mathrm{T}}\pi, \qquad \tau_j = t_j^r$$

This system is clearly consistent since (p^r, t^r), where p^r is some n-vector satisfying $t^r = E^{\mathrm{T}}p^r$, is a solution. Furthermore, if for each r we can find a solution (π^r, τ^r) to it such that the sequence $\{\tau^r\}$ is bounded, then it follows that we can take $v = t^r - \tau^r$ for any r sufficiently large. To find such a sequence $\{\tau^r\}$, we consider, for each r, the following reduced system of linear equations

$$t_j^r = \sum_{i=1}^{n} e_{ij}\pi_i, \quad j \in J',$$

where J' is a subset of J such that the columns of E whose index belongs to J' are linearly independent and span the same space as the columns of E whose index belongs to J. Then we partition the above reduced system into

$$t_{J'}^r = B\pi_B + N\pi_N, \qquad \pi = (\pi_B, \pi_N)$$

where B is an invertible matrix and $t_{J'}^r$ denotes the vector with coordinates $t_j^r, j \in J'$ and set

$$\pi^r = (\pi_B^r, \pi_N^r) = (B^{-1}t_{J'}^r, 0), \qquad \tau^r = E^{\mathrm{T}}\pi^r. \qquad \square$$

Lemma 3. *$\{d^r\}$ is bounded.*

Proof. Suppose that $\{d^r\}$ is not bounded. Then in view of (12), there exist a $j^* \in \{1, 2, \ldots, m\}$ and a subsequence R such that either $c_{j^*} = \infty$, $\{t_{j^*}^r\}_R \to \infty$ or $l_{j^*} = -\infty$, $\{t_{j^*}^r\}_R \to -\infty$. Without loss of generality we will assume that $\{t_{j^*}^r\}_R \to -\infty$. Passing to a subsequence if necessary we assume that, for each j, $\{t_j^r\}_R$ is either bounded, or tends to ∞, or tends to $-\infty$. From Lemma 2 we have that there exists $v \in C^{\perp}$ such that v satisfies (13). Let u be such that $v = E^{\mathrm{T}}u$. Then for any nonnegative Δ we have

$$q'(p^r - \Delta u; -u) = -\sum_{t_j^k \to \infty, k \in R} \nabla g_j(t_j^r - \Delta v_j)v_j - \sum_{t_j^k \to -\infty, k \in R} \nabla g_j(t_j^r - \Delta v_j)v_j \quad \forall r \in R$$

and since

$$\lim_{\eta \to \infty} \nabla g_j(\eta_j) = c_j \quad \text{and} \quad \lim_{\eta \to -\infty} \nabla g_j(\eta_j) = l_j, \qquad j = 1, 2, \ldots, m,$$

it follows that

$$\lim_{r \to \infty, r \in R} q'(p^r - \Delta u; -u) = -\sum_{v_j > 0} c_j v_j - \sum_{v_j < 0} l_j v_j. \tag{14}$$

By construction each term on the right hand side of (14) is less than $+\infty$ and at least one (namely the one which is indexed by j^*) has the value of $-\infty$ we obtain that

$$\lim_{r\to\infty, r\in R} q'(p^r - \Delta u; -u) = -\infty. \tag{15}$$

Also by integrating from 0 to Δ and using the convexity of q we obtain that

$$q(p^r - \Delta u) \le q(p^r) + \Delta q'(p^r - \Delta u; -u) \quad \forall r \in R \text{ sufficiently large.}$$

This result, together with (15), implies that the dual cost can be decreased by any arbitrary amount by taking $r \in R$ sufficiently large. Since $q(p^r)$ is nonincreasing, this implies that $\inf q(p) = -\infty$, contradicting Assumption A. □

The following lemma is an intermediate step toward showing that $\{x^r\}$ is bounded.

Lemma 4. *If, for each j, $\{x_j^r\}$ either tends to ∞, or tends to $-\infty$, or is bounded then, for each r, x^r can be decomposed into $x^r = y^r + z^r$ such that $\{y^r\}$ is bounded and $\{z^r\}$ satisfies $Ez^r = 0$ for all r and, for each j,*

$$\begin{aligned} &z_j^r \to \infty && \text{if } x_j^r \to \infty \\ &z_j^r \to -\infty && \text{if } x_j^r \to -\infty \\ &z_j^r = 0\ \forall r && \text{if } x_j^r \text{ is bounded.} \end{aligned}$$

Proof (by construction). Let J denote the set of j for which $\{x_j^r\}$ is bounded. For each r, consider the solution to the following system of linear equations in ξ

$$E\xi = d^r, \qquad \xi_j = x_j^r \quad \forall j \in J.$$

This system is consistent since x^r is a solution to it. Its solution set can be expressed as

$$\left\{ L\begin{bmatrix} d^r \\ x_J^r \end{bmatrix} + \eta \;\middle|\; E\eta = 0,\ \eta_j = 0 \quad \forall j \in J \right\}$$

where x_J^r is a vector with coordinates x_j^r, $j \in J$ and L is some linear operator that depends on E and J only. Let y^r denote the element of the above solution set with minimum L_2 norm. Since each of the sequences $\{d^r\}$ and $\{x_j^r\}$, $j \in J$, is bounded it follows that the sequence $\{y^r\}$ is bounded. It is easily verified that $\{y^r\}$ and $\{z^r\}$, where $z^r = x^r - y^r$ for all r, give the desired decomposition. □

Lemma 5. *$\{x^r\}$ is bounded.*

Proof. We will argue by contradiction. Suppose that $\{x^r\}$ is not bounded. Then passing to a subsequence if necessary we can assume that each x_j^r either tends to ∞, or tends to $-\infty$, or is bounded. Using Lemma 4 we decompose x^r into the sum of a bounded part and an unbounded part:

$x^r = w^r + z^r$ where w^r is bounded, $Ez^r = 0$, and for each j, $z_j^r \to \infty$ if $x_j^r \to \infty$, $z_j^r \to -\infty$ if $x_j^r \to -\infty$, $z_j^r = 0\ \forall r$ if x_j^r is bounded. Since, for all r,

$$f_j^-(x_j^r) \le t_j^r \le f_j^+(x_j^r), \quad j = 1, 2, \ldots, m,$$

it follows that for r sufficiently large

$$\sum_{j \ni z_j^r \to \infty} f_j^-(x_j^r) z_j^r + \sum_{j \ni z_j^r \to -\infty} f_j^+(x_j^r) z_j^r \leq \sum_{j=1}^m t_j^r z_j^r = 0. \tag{16}$$

From Assumption B and the boundedness of w^r we have

$$z_j^r \to \infty \Rightarrow f_j^-(x_j^r) \to \infty \quad \text{and} \quad z_j^r \to -\infty \Rightarrow f_j^+(x_j^r) \to -\infty$$

implying that the quantity on the left hand side of (16) tends to ∞ thus contradicting (16). □

Using Lemmas 3 and 5 we obtain:

Lemma 6. *$d_{s^r}^r \to 0$ as $r \to \infty$.*

Proof. Consider a fixed r and let $s = s^r$. Since the decrease in the magnitude of $d_s(p)$ during the rth iteration is at least $|d_s^r|(1-\delta)$ we obtain

$$|d_s^r|(1-\delta) \leq |d_s^r - d_s^{r+1}| \leq \sum_{j=1}^m |e_{sj}| |x_j^r - x_j^{r+1}| \leq \sum_{j=1}^m |e_{sj}| \max_j |x_j^r - x_j^{r+1}|.$$

This implies that

$$\max_j |x_j^r - x_j^{r+1}| \geq \frac{|d_s^r|(1-\delta)}{\sum_{j=1}^m |e_{sj}|} \quad \forall r.$$

Suppose that $d_{s^r}^r$ does not tend to zero, then there exist $\varepsilon > 0$, subsequence R, and an index s such that $s^r = s$, $|d_s^r| \geq \varepsilon$ for all $r \in R$. It follows from (17) that for each $r \in R$ there exists some j such that x_j^r must change by at least

$$\frac{\varepsilon(1-\delta)}{\sum_{j=1}^m |e_{sj}|}.$$

We will assume without loss of generality that x_j^r increases and that it is the same j for all $r \in R$.

Let θ denote the scalar in (18). Since x^r is bounded it has a limit point χ. Passing to a subsequence if necessary we will assume that $\{x^r\}_R \to \chi$. Since $t_j^r \leq f_j^+(x_j^r)$ we have that, for each $r \in R$,

$$\begin{aligned} f_j(x_j^{r+1}) - f_j(x_j^r) - t_j^r(x_j^{r+1} - x_j^r) &\geq f_j(x_j^{r+1}) - f_j(x_j^r) - f_j^+(x_j^r)(x_j^{r+1} - x_j^r) \\ &\geq f_j(x_j^r + \theta) - f_j(x_j^r) - f_j^+(x_j^r)\theta. \end{aligned}$$

Using the fact that $x^r \to \chi$ and the upper semicontinuity of f_j^+ we obtain

$$\lim_{\substack{r \to \infty \\ r \in R}} f_j^+(x_j^r) \leq f_j^+(\chi_j)$$

so that (using the lower semicontinuity of f_j)

$$\lim_{\substack{r\to\infty\\ r\in R}} [f_j(x_j^r+\theta)-f_j(x_j^r)-f_j^+(x_j^r)\theta] \geq f_j(\chi_j+\theta)-f_j(\chi_j)-f_j^+(\chi_j)\theta.$$

Using Lemma 1 we obtain that

$$\lim_{\substack{r\to\infty\\ r\in R}} \inf [q(p^r)-q(p^{r+1})] \geq f_j(\chi_j+\theta)-f_j(\chi_j)-f_j^+(\chi_j)\theta$$

and since the right hand side of the relation above is a positive quantity (due to the strict convexity of f_j), we have that $q(p^r)\to-\infty$, contradicting Assumption A. □

Using Lemma 6 we obtain our first convergence result:

Proposition 1. *Under Assumption C, $x^r\to x^*$ and $q(p^r)\to -f(x^*)$, where x^* denotes the optimal primal solution.*

Proof. We first derive an upper bound on the change

$$|d_i(p^r)-d_i(p^{r+1})|, \quad i=1,2,\ldots,n.$$

We have

$$|d_i(p^r)-d_i(p^{r+1})| = \left|\sum_{j=1}^m e_{ij}(x_j^r-x_j^{r+1})\right| \leq \sum_{j=1}^m |e_{ij}| \max_j \Delta_j^r \tag{19}$$

where

$$\Delta_j^r = |x_j^r-x_j^{r+1}|.$$

Let us denote for notational convenience

$$s=s^r$$

If $d_s^r>0$ then $p_s^{r+1}-p_s^r<0$ while $p_i^{r+1}=p_i^r$ for $i\neq s$. Since

$$t_j^{r+1}-t_j^r = \sum_{i=1}^n e_{ij}(p_i^{r+1}-p_i^r)$$

we see that

$$\begin{aligned} t_j^{r+1}-t_j^r<0 \quad \text{if } e_{sj}>0,\\ t_j^{r+1}-t_j^r>0 \quad \text{if } e_{sj}<0. \end{aligned} \tag{20a}$$

If $d_s^r<0$ then similarly

$$\begin{aligned} t_j^{r+1}-t_j^r>0 \quad \text{if } e_{sj}>0,\\ t_j^{r+1}-t_j^r<0 \quad \text{if } e_{sj}<0. \end{aligned} \tag{20b}$$

We also have

$$x_j^{r+1}-x_j^r = \nabla g_j(t_j^{r+1})-\nabla g_j(t_j^r),$$

and the gradient ∇g_j is monotonically nondecreasing since g_j is convex. Using this fact together with (20) we obtain

$$d_s^r > 0 \Rightarrow e_{sj}(x_j^{r+1} - x_j^r) \leq 0 \quad \forall j$$

$$d_s^r < 0 \Rightarrow e_{sj}(x_j^{r+1} - x_j^r) \geq 0 \quad \forall j.$$

After the rth relaxation iteration d_s^{r+1} will be smaller in absolute value and will have the same sign as d_s^r, so we have using the relations above

$$|d_s^r| \geq |d_s^r - d_s^{r+1}| = \left| \sum_{j=1}^m e_{sj}(x_j^r - x_j^{r+1}) \right| = \sum_{j=1}^m |e_{sj}| |x_j^r - x_j^{r+1}|$$
$$\geq [\min_{e_{sj} \neq 0} |e_{sj}|][\max_j \Delta_j^r].$$

Therefore

$$\max_j \Delta_j^r \leq \frac{|d_s^r|}{\min_{e_{sj} \neq 0} |e_{sj}|}.$$

Combining this relation with (19) we have, for all i,

$$|d_i(p^r) - d_i(p^{r+1})| \leq \frac{|d_s^r| \sum_{j=1}^m |e_{ij}|}{\min_{e_{sj} \neq 0} |e_{sj}|}$$
$$\leq |d_s^r| L, \tag{21}$$

where

$$L = \frac{\max_i \sum_{j=1}^m |e_{ij}|}{\min_{e_{sj} \neq 0} |e_{sj}|}.$$

For a fixed s, if $s = s^r$ for some index r then for $k \in \{r+1, \ldots, r+T\}$ we have (using (21))

$$|d_s^k| \leq |d_s^r| + L \sum_{h=r+1}^{r+T} |d_{s^h}^h|,$$

where T is the upper bound in Assumption C. By Lemma 6 we obtain that

$$\lim_{k \to \infty} |d_s^k| = 0.$$

Since the choice of s was arbitrary, we have that $d^r \to 0$. Therefore, since $d^r = Ex^r$, every limit point of the sequence $\{x^r\}$ is primal feasible.

For all r and all column indexes j we have that the Complementary Slackness condition

$$f_j^-(x_j^r) \leq t_j^r \leq f_j^+(x_j^r)$$

holds. Let z be any vector in the constraint subspace C. Then

$$\sum_{j=1}^m t_j^r z_j = 0 \quad \forall r,$$

so using (22) and (23) we obtain that

$$\sum_{z_j>0} f_j^-(x_j^r)z_j + \sum_{z_j<0} f_j^+(x_j^r)z_j \leq 0 \leq \sum_{z_j>0} f_j^+(x_j^r)z_j + \sum_{z_j<0} f_j^-(x_j^r)z_j \quad \forall r. \tag{24}$$

Let $\{x^r\}_{r\in R}$ be a subsequence converging to [cf. Lemma 5] some limit point χ. Then from (24) and using the lower semicontinuity of f_j^- and the upper semicontinuity of f_j^+ we have, for all z belonging to the constraint subspace C, that

$$\sum_{z_j>0} f_j^-(\chi_j)z_j + \sum_{z_j<0} f_j^+(\chi_j)z_j \leq 0 \leq \sum_{z_j>0} f_j^+(\chi_j)z_j + \sum_{z_j<0} f_j^-(\chi_j)z_j.$$

Therefore the directional derivative $f'(\chi, z)$ is nonnegative for each $z \in C$. Since χ is primal feasible, this implies that χ is an optimal primal solution. Since the optimal primal solution x^* is unique, the entire sequence $\{x^r\}$ converges to x^*.

Now we will prove that $q(p^r) \to -f(x^*)$. We first have, using (3) and (5), the weak duality result

$$0 \leq f(x^*) + q(p^r) \quad \forall r. \tag{25}$$

To obtain a bound on the right hand side of (25) we observe that $(t^r)^{\mathrm{T}}x^* = 0$ so that

$$f(x^*) + q(p^r) = f(x^*) - (t^r)^{\mathrm{T}}x^* + (t^r)^{\mathrm{T}}x^r - f(x^r) \quad \forall r. \tag{26}$$

Using (22) and the lower semicontinuity of f_j^- and the upper semicontinuity of f_j^+ we obtain: For all j such that $l_j < x_j^* < c_j$,

$$-\infty < f_j^-(x_j^*) \leq \lim_{r\to\infty} \inf\{t_j^r\} \quad \text{and} \quad \lim_{r\to\infty} \sup\{t_j^r\} \leq f_j^+(x_j^*) < \infty$$

and therefore $|t_j^r|$ is bounded by some positive scalar M.

For all j such that $l_j = x_j^* < c_j$,

$$\lim_{r\to\infty} t_j^r(x_j^* - x_j^r) \geq \lim_{r\to\infty} f_j^+(l_j)(x_j^* - x_j^r) = 0.$$

For all j such that $l_j < x_j^* = c_j$,

$$\lim_{r\to\infty} t_j^r(x_j^* - x_j^r) \geq \lim_{r\to\infty} f_j^-(c_j)(x_j^* - x_j^r) = 0.$$

For all j such that $l_j = c_j$,

$$t_j^r(x_j^* - x_j^r) = 0 \quad \forall r.$$

Combining (26) with (27) and (30) yields

$$\begin{aligned} f(x^*) + q(p^r) &= \sum_{j=1}^m [f_j(x_j^*) - f_j(x_j^r) - t_j^r(x_j^* - x_j^r)] \\ &\leq \sum_{j=1}^m [f_j(x_j^*) - f_j(x_j^r)] + \sum_{l_j<x_j^*<c_j} M|x_j^* - x_j^r| \\ &\quad - \sum_{x_j^*=l_j<c_j} t_j^r(x_j^* - x_j^r) - \sum_{l_j<c_j=x_j^*} t_j^r(x_j^* - x_j^r). \end{aligned}$$

Since $x^r \to x^*$ it follows from (25), (28), and (29) that $f(x^*) + q(p^r) \to 0$. $\square$

As a consequence of Proposition 1 we obtain that every limit of the dual price sequence $\{p^r\}$ is an optimal dual solution. However the existence and number of limit points of $\{p^r\}$ are unresolved issues at present. For the case of network problems it was shown (under an additional mild condition on the line search in the relaxation iteration) that the entire sequence $\{p^r\}$ converges to some optimal price vector assuming the dual problem has at least one solution [2]. (For network problems the dual optimal solution set is unbounded when it is nonempty [2] but it is possible that no optimal solution exists.) The best that we have been able to show is that the distance of p^r to the optimal dual solution set converges to zero when the dual solution set is nonempty. Since this result is not as strong as the one obtained for network problems in [2] we will not give it here.

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

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

where $\{b_k\}$ is any sequence of scalars such that for some positive scalar ρ

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

The assumption is as follows:

Assumption C′. 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 ∞ 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 the conclusions of Proposition 1 hold, under Assumption C′, if in addition the cost function f is strongly convex. These convergence results are of interest in that they show that, for a large class of problems, cyclical relaxation is not essential for the Gauss-Seidel method to be convergent. To the best of our knowledge, the only other works treating convergence of the Gauss-Seidel method that do not require cyclical relaxation are [1] and [2] dealing with the special case of network flow problems.

Proposition 2. *If f is strongly convex in the sense that there exist scalars $\sigma > 0$ and $\gamma > 1$ such that*

$$f(y) - f(x) - f'(x; y - x) \geq \sigma \|y - x\|^{\gamma} \quad \forall x, y \text{ such that } f(x) < \infty, f(y) < \infty, \tag{31}$$

where $\|\cdot\|$ denotes the L_2 norm, and Assumption C′ holds with $\rho = \gamma - 1$, then $x^r \to x^*$ and $q(p^r) \to -f(x^*)$, where x^* denotes the optimal primal solution.

Proof. By Lemma 1 and (31) we have that

$$q(p^r)-q(p^{r+1})\geq \sigma\|x^{r+1}-x^r\|^{\gamma} \quad \forall r,$$

which together with (17) implies that there exists a positive scalar K depending only on δ, σ, γ, and the problem data such that

$$q(p^r)-q(p^{r+1})\geq K|d^r_{s^r}|^{\gamma} \quad \forall r.$$

Summing the above inequality over all r, we obtain

$$q(p^0)-\lim_{r\to\infty} q(p^r)\geq K\sum_{r=0}^{\infty}|d^r_{s^r}|^{\gamma}.$$

Since the left hand side of the relation above is real valued it follows that

$$\sum_{r=1}^{\infty}|d^r_{s^r}|^{\gamma}<\infty. \tag{32}$$

We show next that there exists a subsequence R such that

$$\lim_{r\to\infty, r\in R}|d^r_s|=0 \quad \text{for } s=1,2,\ldots,n. \tag{33}$$

Consider a fixed $s\in\{1,2,\ldots,n\}$. By Assumption C′, coordinate p_s is relaxed in at least one iteration, which we denote by $r(h)$, between τ_h+1 and τ_{h+1} for $h=1,2,\ldots$ (for a given h, if more than one choice of value for $r(h)$ is possible then an arbitrary choice is made). We have

$$d_s^{\tau_{h+1}}=d^{r(h)}_{s^{r(h)}}+\sum_{r=r(h)}^{\tau_{h+1}-1}(d_s^{r+1}-d_s^r), \quad h=1,2,\ldots,$$

which together with (21) implies that there exists a scalar L depending only on the problem data such that

$$|d_s^{\tau_{h+1}}|\leq \max_{r\in\{\tau_h+1,\ldots,\tau_{h+1}\}}|d^r_{s^r}|+L\sum_{r=\tau_h+1}^{\tau_{h+1}}|d^r_{s^r}|, \quad h=1,2,\ldots. \tag{34}$$

The choice of s was arbitrary and therefore (34) holds for all s. To prove (33) it is sufficient that we show that there exists some subsequence H of $\{1,2,\ldots\}$ such that the right hand side of (34) tends to zero as $h\to\infty$, $h\in H$, since this will imply that

$$|d_s^{\tau_{h+1}}|\to 0 \quad \text{as } h\to\infty, h\in H$$

for all s.

By Lemma 6 the first term on the right hand side of (34) tends to zero as $h\to\infty$ and therefore we only have to prove that there exists some subsequence H of $\{1,2,\ldots\}$ such that

$$\sum_{r=\tau_h+1}^{\tau_{h+1}}|d^r_{s^r}|\to 0 \quad \text{as } h\to\infty, h\in H.$$

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

$$\varepsilon \leq \sum_{r=\tau_h+1}^{\tau_{h+1}} |d_{s^r}^r| \quad \forall h \geq h^*.$$

We will use the Hölder inequality [5] which says that for any positive integer N and two vectors x and y in R^N

$$|x^{\mathrm{T}} y| \leq \|x\|_v \|y\|_\eta$$

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

$$\sum_{i=1}^{N} x_i \leq \left[\sum_{i=1}^{N} (x_i)^v\right]^{1/v} (N)^{1/\eta}.$$

Applying the above identity to the right hand side of (36) with $v = \gamma$ and $N = \tau_{h+1} - \tau_h$ yields

$$\varepsilon^\gamma \leq \left[\sum_{r=\tau_h+1}^{\tau_{h+1}} |d_{s^r}^r|\right]^\gamma \leq \left[\sum_{r=\tau_h+1}^{\tau_{h+1}} |d_{s^r}^r|^\gamma\right] (\tau_{h+1} - \tau_h)^{\gamma-1} \quad \forall h \geq h^*$$

which implies that

$$\varepsilon^\gamma \sum_{h=h^*}^{\infty} \frac{1}{(\tau_{h+1} - \tau_h)^{\gamma-1}} \leq \sum_{h=h^*}^{\infty} \left[\sum_{r=\tau_h+1}^{\tau_{h+1}} |d_{s^r}^r|^\gamma\right] = \sum_{r=\tau_{h^*}+1}^{\infty} |d_{s^r}^r|^\gamma. \tag{37}$$

The leftmost quantity of (37) by construction of the sequence $\{\tau_k\}$ has value of $+\infty$ while the rightmost quantity of (37) according to (32) has finite value thereby reaching a contradiction. This establishes (33).

By (33) there exists a subsequence R such that $d^r \to 0$ as $r \to \infty$, $r \in R$. It thus follows from Lemma 5 that the subsequence $\{x^r\}_{r \in R}$ has at least one limit point and that each limit point of $\{x^r\}_{r \in R}$ is primal feasible. Then following an argument identical to that used in the second half of the proof of Proposition 1 we obtain that $\{x^r\}_{r \in R}$ converges to the optimal primal solution x^* and that $\{q(p^r)\}_{r \in R} \to -f(x^*)$. Since $q(p^r)$ is monotonically decreasing in r it then follows that

$$q(p^r) \to -f(x^*) \quad \text{as } r \to \infty \tag{38}$$

and the second part of Proposition 2 is proven.

To prove the first part of Proposition 2 we first note that if f satisfies (31) then every primal feasible solution is regularly feasible (in the terminology of [11, Chapter 11]), and guarantees (together with Assumption A) that the dual problem (4) has an optimal price vector [11, Chapter 11]. Let p^* denote one such optimal price vector. Then using (31) and an argument similar to that used in proving Lemma 1 we obtain that

$$q(p^r) - q(p^*) \geq \sigma \|x^r - x^*\|^\gamma, \quad r = 0, 1, \ldots$$

which together with (38) and the fact that $-f(x^*) = q(p^*)$ yields $x^r \to x^*$. □

Appendix. Implementation of the inexact line search

The inexact line minimization step in the relaxation iteration requires, for a given set of prices p_i and a coordinate s, the determination of a nonnegative scalar Δ satisfying the following set of inequalities:

$$0 \le \sum_j e_{sj} \nabla g_j(t_j - \Delta e_{sj}) \le \delta\beta \quad \text{if } \beta > 0, \tag{A1}$$

$$\delta\beta \le \sum_j e_{sj} \nabla g_j(t_j + \Delta e_{sj}) \le 0 \quad \text{if } \beta < 0, \tag{A2}$$

where $\beta = d_s(p)$ and $t_j = \sum_i e_{ij} p_i$. For simplicity we will assume that $\beta < 0$. The case where $\beta > 0$ may be treated analogously. Consider a fixed $r \in [\delta\beta, 0]$. Then a scalar Δ satisfies (A2) if for some x_j', $j = 1, 2, \ldots, m$,

$$f_j^-(x_j') - t_j \le \Delta e_{sj} \le f_j^+(x_j') - t_j, \quad j = 1, 2, \ldots, m,$$

$$\sum_j e_{sj} x_j' = r,$$

or equivalently if x_j', $j = 1, 2, \ldots, m$, is the optimal solution to

$$\begin{aligned} &\text{minimize} \sum_j f_j(x_j) - t_j x_j, \\ &\text{subject to} \sum_j e_{sj} x_j = r, \end{aligned} \tag{A3}$$

and Δ is the optimal Lagrange multiplier associated with the equality constraint. Thus we can reduce the inexact line search problem to that of finding a solution to (A3) for some $r \in [\delta\beta, 0]$.

In the special case where ∇g_j can be evaluated pointwise, a Δ satisfying (A2) may be computed more directly by applying any one of many zero finding techniques to the function

$$h(\lambda) = \sum_j e_{sj} \nabla g_j(t_j + \lambda e_{sj}).$$

One such technique is binary search. To implement binary search we need an upper bound on Δ. To do this we will make the assumption that $-\infty < l_j < c_j < +\infty$ and $f_j^-(c_j) < +\infty$, $f_j^+(l_j) > -\infty$, for all j (such an assumption is clearly reasonable for practical computation). With this additional assumption we obtain [cf. (A2)] that Δ must satisfy

$$e_{sj}[\nabla g_j(t_j + \Delta e_{sj}) - \nabla g_j(t_j)] \le -\beta \quad \text{for all } j$$

or equivalently

$$\nabla g_j(t_j + \Delta e_{sj}) \le \nabla g_j(t_j) - \beta / e_{sj} \quad \text{for all } j \text{ such that } e_{sj} > 0,$$

$$\nabla g_j(t_j + \Delta e_{sj}) \ge \nabla g_j(t_j) - \beta / e_{sj} \quad \text{for all } j \text{ such that } e_{sj} < 0.$$

Thus an upper bound Δ' on the inexact linesearch stepsize Δ is

$$\Delta' = \min\{\Delta_1, \Delta_2\},$$

where

$$\Delta_1 = \min\left\{\min_{e_{sj}>0} \frac{f_j^-(y_j)-t_j}{e_{sj}}, \min_{e_{sj}<0} \frac{f_j^+(y_j)-t_j}{e_{sj}}\right\},$$

$$\Delta_2 = \max\left\{\max_{e_{sj}>0} \frac{f_j^-(c_j)-t_j}{e_{sj}}, \max_{e_{sj}<0} \frac{f_j^+(l_j)-t_j}{e_{sj}}\right\},$$

and

$$y_j = \nabla g_j(t_j) - \beta / e_{sj} \quad \text{for all } j \text{ such that } e_{sj} \neq 0.$$

An instance for which ∇g_j can be evaluated pointwise is where each f_j is piecewise differentiable and on each piece $(\nabla f_j)^{-1}$ has closed form. An example is when each f_j is the pointwise maximum of scalar functions of forms such as

$$\begin{aligned} &be^{ax_j} + c, && b > 0, \\ &\text{or } b|x_j - d|^a + c, && a > 1, b > 0, \\ &\text{or } b(x_j - d)^{-1} + c, && b > 0. \end{aligned}$$

In the special case where each f_j is piecewise differentiable, and the number of pieces is relatively small we can reduce the work in the binary search by first sorting the breakpoints of $h(\lambda)$ and then applying binary search on the breakpoints to determine the two neighboring breakpoints between which a Δ satisfying (A2) lies. We can then apply binary search to this smaller interval.

References

[1] D.P. Bertsekas and D. Elbaz, "Distributed asynchronous relaxation methods for convex network flow problems," *SIAM Journal on Control and Optimization* 25 (1987) 74–85.

[2] D.P. Bertsekas, P.A. Hosein and P. Tseng, "Relaxation methods for network flow problems with convex arc costs," LIDS Report P-1523, Mass. Institute of Technology, December 1985, *SIAM Journal on Control and Optimization*, to appear.

[3] D.P. Bertsekas and P. Tseng, "Relaxation Methods for Linear Programs," LIDS Report P-1553, Mass. Institute of Technology, April 1986, to appear in *Mathematics of Operations Research*.

[4] R.W. Cottle and J.S. Pang, "On the convergence of a block successive over-relaxation method for a class of linear complementary problems," *Mathematical Programming Study 17* (1982) 126–138.

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

[6] D.G. Luenberger, *Introduction to Linear and Nonlinear Programming* (Addison-Wesley, Reading, MA, 2nd ed., 1984).

[7] J.S. Pang, "On the convergence of dual ascent methods for large-scale linearly constrained optimization problems," Unpublished manuscript, The University of Texas at Dallas, 1984.

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

[9] M.J.D. Powell, "On search directions for minimization algorithms," *Mathematical Programming* 4 (1973) 193–201.

[10] R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, New Jersey, 1970).

[11] R.T. Rockafellar, *Network Flows and Monotropic Programming* (Wiley-Interscience, New York, 1983).

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

[13] P. Tseng, "Relaxation methods for monotropic programming problems," Ph.D. Thesis, Dept. of Electrical Engineering and Computer Science, Operations Research Center, Mass. Institute of Technology (1986).

[14] N. Zadeh, "A note on the cyclic coordinate ascent method," *Management Science* 16 (1970) 642–644.

[15] W.I. Zangwill, *Nonlinear Programming: A Unified Approach* (Prentice-Hall, Englewood Cliffs, New Jersey, 1969).

[16] S.A. Zenios and J.M. Mulvey, "Simulating a distributed synchronous relaxation method for convex network problems," Working Paper, Department of Civil Engineering, Princeton University, Princeton, New Jersey, January 1985.