April 1990 LIDS-P-1967

*An Alternating Direction Method for Linear Programming**

by

Jonathan ECKSTEIN
Graduate School of Business, Harvard University
Boston, MA 02163

Dimitri P. BERTSEKAS
Laboratory for Information and Decision Systems, Massachusetts Institute of Technology
Cambridge, MA 02139

Abstract

This paper presents a new, simple, massively parallel algorithm for linear programming, called the *alternating step method.* The algorithm is unusual in that it does not maintain primal feasibility, dual feasibility, or complementary slackness; rather, all these conditions are gradually met as the method proceeds. We derive the algorithm from an extension of the alternating direction method of multipliers for convex programming, giving a new algorithm for monotropic programming in the course of the development. Concentrating on the linear programming case, we give a proof that, under a simple condition on the algorithm parameters, the method converges at a globally linear rate. Finally, we give some preliminary computational results.

Key Words: Linear programming, parallel algorithms, proximal point algorithms, monotropic programming

Abbreviated Title for Running Heads: *An Alternating Direction Method*

---

* This paper consists mainly of refinements of dissertation research results of the first author. The dissertation was performed at M.I.T. under the supervision of the second author, and was supported in part by the Army Research Office under grant number DAAL03-86-K-0171, and by the National Science Foundation under grant number ECS-8519058. Alliant FX/8 computer time was donated by the Mathematics and Computer Science division of Argonne National Laboratories.

1. Introduction

This paper develops the *alternating step method*, a novel, simple algorithm for linear programming that is very amenable to parallel implementation. This algorithm works by applying simple linear updates alternately to the primal and dual variables of the linear program (hence its name), but is unusual in that it does not enforce primal feasibility, dual feasibility, or complementary slackness. Instead, all these properties are gradually satisfied as the algorithm progresses. Simplex and dual relaxation methods for linear programming generally enforce two of these properties, while working to satisfy the third; even interior point methods generally enforce feasibility restrictions.

The alternating step method is original to the authors; however, a preliminary derivation appears in Bertsekas and Tsitsiklis (1989, p. 254). The more thorough analysis presented here is a refinement of that of Eckstein (1989).

We derive the alternating step method by specializing the *generalized alternating direction method of multipliers*, a decomposition algorithm for convex programming (Eckstein and Bertsekas 1989). This algorithm is presented in Section 2, and is an extension of the earlier *alternating direction method of multipliers*, which was first introduced by Glowinski and Marroco (1975), and by Gabay and Mercier (1976). Further contributions to the theory of the alternating direction method of multipliers appeared in Fortin and Glowinski (1983), and Gabay (1983). Glowinski and Le Tallec (1987) give an extensive treatment of the method and its relatives, and a relatively accessible exposition appears in Bertsekas and Tsitsiklis (1989, pp. 253-261). However, an intimate understanding of the workings of these alternating direction methods is not necessary for the majority of this paper.

Section 3 uses the generalized alternating direction method of multipliers to derive the alternating step method for monotropic programming, that is, linearly-constrained convex programming with a separable objective (Rockafellar 1984), while Section 4 then specializes this algorithm to the case of linear programming. Related work in the literature includes Spingarn's (1985b) application of the method of partial inverses to optimization, in particular the algorithm for block-separable convex programming. Gol'shtein's "block" method of convex programming (1985, 1986) are based on a fundamentally identical decomposition strategy. The theoretical underpinnings relating the Spingarn and Gol'shtein convex programming algorithms to the alternating direction method of multipliers are elaborated in Eckstein and Bertsekas (1989) and Eckstein (1989), where additional references are given. A distinguishing feature of our approach is that we apply a dimension-increasing transformation to the original monotropic program, increasing the number of variables from n to nm, where m is the number of constraints. After deriving an algorithm for this expanded problem, we simplify it until it involves only primal variables of dimension n and dual variables of dimension m. We should also note that there have been numerous applications of the conventional (as opposed to alternating direction) method of multipliers to linear programming. For one example, consider De Leone and Mangasarian (1988).

Section 5 gives a proof that, under a simple condition on the algorithm parameters, the alternating step method for linear programs converges at a globally linear rate. The analysis draws on ideas from the convergence analysis of the proximal point algorithm (Luque 1984), and results in linear programming stability theory (Mangasarian and Shiau 1987). Similar preceeding work on convergence rates includes the asymptotic convergence rate result for Spingarn's block-separable convex programming method (1985b, Section 4); however, that analysis makes the rather stringent assumption that the problem meets the strong second-order sufficiency conditions. Here, we make only the assumptions that the linear program in question is feasible with bounded objective value. Our result is closer in spirit to Spingarn's (1985a, 1987) finite termination and asymptotic convergence rate results concerning his partial-inverse-based methods for linear inequality systems. These methods could in principle be applied to linear programming by posing the primal and dual problems as a single system of inequalities. However, we prove convergence at a *global*, as opposed to asymptotic, linear rate. Because of the complexity of the analysis, Section 5 assumes familiarity with the theory of the generalized alternating direction method of multipliers, and the proximal point algorithm that underlies it (Eckstein and Bertsekas 1989, Lawrence and Spingarn 1987, Eckstein 1988, Eckstein 1989). It is the only part of this paper that requires such familiarity.

Section 6 gives preliminary computational results for the method, and describes some simple heuristics that accelerate its convergence in practice. For general linear programming, the method appears to be uncompetitive in a serial computing environmnet. Taking advantage of parallelism and certain specialized constraint structures, however, the method does show some promise. Versions of the method have been implemented on a variety of parallel computers; details of these implementations are reserved for a future paper.

2. *The Generalized Alternating Direction Method of Multipliers*

We now present the generalized alternating direction method of multipliers, upon which our derivation is based. For background material on convex analysis, see Rockafellar (1970). First, consider a general finite-dimensional optimization problem of the form

$$\text{minimize}_{\mathbf{x}\in\mathbb{R}^n} \; f(\mathbf{x}) + g(\mathbf{M}\mathbf{x}) \quad , \tag{P}$$

where $f\colon \mathbb{R}^n \to (-\infty, +\infty]$ and $g\colon \mathbb{R}^s \to (-\infty, +\infty]$ are closed proper convex, and $\mathbf{M}$ is some $s \times n$ matrix. One can also write (P) is the form

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) + g(\mathbf{z}) \\ \text{subject to} & \mathbf{M}\mathbf{x} = \mathbf{z} \;, \end{array} \tag{P'}$$

and attaching a multiplier vector $\mathbf{p} \in \mathbb{R}^s$ to the constraints $\mathbf{M}\mathbf{x} = \mathbf{w}$, one obtains an equivalent dual problem

$$\text{maximize}_{\mathbf{p}\in\mathbb{R}^s} \; -\left(f^*(-\mathbf{M}^{\mathsf{T}}\mathbf{p}) + g^*(\mathbf{p})\right) \quad , \tag{D}$$

where * denotes the convex conjugacy operation. The *alternating direction method of multipliers* for (P) takes the form

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} \{f(\mathbf{x}) + \langle \mathbf{p}^k, \mathbf{M}\mathbf{x}\rangle + \tfrac{\lambda}{2}\| \mathbf{M}\mathbf{x} - \mathbf{z}^k\|^2\}$$
$$\mathbf{z}^{k+1} = \arg\min_{\mathbf{z}} \{g(\mathbf{z}) - \langle \mathbf{p}^k, \mathbf{z}\rangle + \tfrac{\lambda}{2}\| \mathbf{M}\mathbf{x}^{k+1} - \mathbf{z}\|^2\}$$
$$\mathbf{p}^{k+1} = \mathbf{p}^k + \lambda(\mathbf{M}\mathbf{x}^{k+1} - \mathbf{z}^{k+1}) \qquad ,$$

where λ is a given positive scalar. This method resembles the conventional Hestenes-Powell method of multipliers for (P') (*e.g.* Bertsekas 1982), except that it minimizes the augmented Lagrangian function

$$L_\lambda(\mathbf{x}, \mathbf{z}, \mathbf{p}) \;=\; f(\mathbf{x}) + g(\mathbf{z}) + \langle \mathbf{p}^k, \mathbf{M}\mathbf{x} - \mathbf{z}\rangle + \tfrac{\lambda}{2}\| \mathbf{M}\mathbf{x} - \mathbf{z}\|^2$$

first with respect to $\mathbf{x}$, and then with respect to $\mathbf{z}$, rather than with respect to both $\mathbf{x}$ and $\mathbf{z}$ simultaneously.

A pair $(\mathbf{x}, \mathbf{p}) \in \mathbb{R}^n \times \mathbb{R}^s$ is said to be a *Kuhn-Tucker pair* for (P) if $-\mathbf{M}^{\mathrm{T}}\mathbf{p} \in \partial f(\mathbf{x})$ and $\mathbf{p} \in \partial g(\mathbf{M}\mathbf{x})$, where ∂ denotes the subgradient mapping, as in Rockafellar (1970). It is a basic exercise in convex analysis to show that if $(\mathbf{x}, \mathbf{p})$ is a Kuhn-Tucker pair, then $\mathbf{x}$ is optimal for (P), $\mathbf{p}$ is optimal for (D), and the two problems have an identical, finite optimum value. We now present the generalized alternating direction method of multipliers:

Theorem 1 (The generalized alternating direction method of multipliers). Consider a convex program in the form (P), minimize$_{\mathbf{x} \in \mathbb{R}^n}$ $f(\mathbf{x}) + g(\mathbf{M}\mathbf{x})$, where $\mathbf{M}$ has full column rank. Let $\mathbf{p}^0$, $\mathbf{z}^0 \in \mathbb{R}^m$, and suppose we are given some scalar $\lambda > 0$ and

$$\{\mu_k\}_{k=0}^{\infty} \subseteq [0, \infty)\ , \qquad \sum_{k=0}^{\infty} \mu_k \;<\; \infty$$

$$\{\nu_k\}_{k=0}^{\infty} \subseteq [0, \infty)\ , \qquad \sum_{k=0}^{\infty} \nu_k \;<\; \infty$$

$$\{\rho_k\}_{k=0}^{\infty} \subseteq (0, 2)\ , \qquad 0 < \inf_{k \ge 0} \rho_k \le \sup_{k \ge 0} \rho_k < 2 \quad .$$

Suppose $\{\mathbf{x}^k\}_{k=1}^{\infty}$, $\{\mathbf{z}^k\}_{k=0}^{\infty}$, and $\{\mathbf{p}^k\}_{k=0}^{\infty}$ conform, for all k, to

$$\left\| \mathbf{x}^{k+1} - \arg\min_{\mathbf{x}} \left\{ f(\mathbf{x}) + \langle \mathbf{p}^k, \mathbf{M}\mathbf{x} \rangle + \tfrac{\lambda}{2} \| \mathbf{M}\mathbf{x} - \mathbf{z}^k \|^2 \right\} \right\| \leq \mu_k$$
$$\left\| \mathbf{z}^{k+1} - \arg\min_{\mathbf{z}} \left\{ g(\mathbf{z}) - \langle \mathbf{p}^k, \mathbf{z} \rangle + \tfrac{\lambda}{2} \| \rho_k \mathbf{M}\mathbf{x}^{k+1} + (1-\rho_k)\mathbf{z}^k - \mathbf{z} \|^2 \right\} \right\| \leq \nu_k$$
$$\mathbf{p}^{k+1} = \mathbf{p}^k + \lambda(\rho_k \mathbf{M}\mathbf{x}^{k+1} + (1-\rho_k)\mathbf{z}^k - \mathbf{z}^{k+1}) \quad .$$

Then if (P) has a Kuhn-Tucker pair, $\{\mathbf{x}^k\}$ converges to a solution of (P) and $\{\mathbf{p}^k\}$ converges to a solution of the dual problem (D). Furthermore, $\{\mathbf{z}^k\}$ converges to $\mathbf{M}\mathbf{x}^*$, where $\mathbf{x}^*$ is the limit of $\{\mathbf{x}^k\}$. If (D) has no optimal solution, then at least one of the sequences $\{\mathbf{p}^k\}$ or $\{\mathbf{z}^k\}$ is unbounded.

The proof of this theorem is given in Eckstein and Bertsekas (1989), and is based on an equivalence to an extension of the proximal point algorithm. Basically, the generalized alternating direction method of multipliers reduces to the alternating direction method of multipliers is the special case that $\rho_k = 1$ for all $k \geq 0$, but also allows for approximate minimization of the augmented Lagrangian.

3. *Monotropic Programming*

A *monotropic program* (Rockafellar 1984) is an optimization problem taking the canonical form

$$\begin{aligned} \text{minimize} \quad & \sum_{j=1}^{n} h_j(x_j) \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \in \mathbb{R}^n \quad , \end{aligned} \qquad \text{(MP)}$$

where the $h_j\colon \mathbb{R} \to (-\infty, +\infty]$, $j=1,\ldots,n$, are closed proper convex, $\mathbf{A}$ is a real $m \times n$ matrix, and $\mathbf{b} \in \mathbb{R}^m$.

Since the h_j can take on the value $+\infty$, they may contain implicit constraints of the form $x_j \geq l_j \in \mathbb{R}$ and $x_j \leq u_j \in \mathbb{R}$. Thus, by proper choice of the h_j and the familiar device of slack variables, any convex optimization problem with a separable lower semicontinuous objective and subject to a finite number of linear equality and inequality constraints can be converted to the form (MP).

Given a monotropic program (MP), define $d(i)$, the *degree of constraint* i, to be the number of nonzero elements in row i of $\mathbf{A}$. If $\mathbf{A}$ is the node-arc incidence matrix of a network or graph, then this definition agrees with the usual notion of the degree of node i. Let $\mathbf{a}_j$ denote column j of $\mathbf{A}$, $1 \leq j \leq n$, and let $\mathbf{A}_i$ denote row i of $\mathbf{A}$, $1 \leq i \leq m$.

The *surplus* or *residual* $r_i(\mathbf{x})$ of constraint i of the system $\mathbf{Ax} = \mathbf{b}$, with respect to the primal variables $\mathbf{x}$, is $b_i - \langle \mathbf{A}_i, \mathbf{x} \rangle$. Intuitively, $r_i(\mathbf{x})$ is the amount by which the i^{th} constraint of the system $\mathbf{Ax} = \mathbf{b}$ is violated. Let $\mathbf{r}(\mathbf{x})$ denote the vector $\mathbf{b} - \mathbf{Ax}$ of all m surpluses.

Let Q_i, for $1 \leq i \leq m$, be any set such that

$$\{\, j \mid 1 \leq j \leq n,\ a_{ij} \neq 0 \,\} \subseteq Q_i \subseteq \{1, \ldots, n\} \quad .$$

Let $q_i = |Q_i|$ for all i, hence $d(i) \leq q_i \leq n$. Furthermore, let

$$Q = \{(i, j) \mid 1 \leq i \leq m, j \in Q_i\} \quad .$$

We now convert (MP) to the form (P), minimize $f(\mathbf{x}) + g(\mathbf{Mx})$. Here, f will be defined on $\mathbb{R}^n$ and g on $\mathbb{R}^{mn}$. Index the components of vectors $\mathbf{z} \in \mathbb{R}^{mn}$ as z_{ij}, where $1 \leq i \leq m$ and $1 \leq j \leq n$. Then let

$$f(\mathbf{x}) = \sum_{j=1}^{n} h_j(x_j)$$

$$C = \{\, \mathbf{z} \in \mathbb{R}^{mn} \mid \sum_{j=1}^{n} z_{ij} = b_i \;\; \forall\, 1 \le i \le m,\;\; z_{ij} = 0 \;\; \forall\, (i,j) \notin Q \}$$

$$g(\mathbf{z}) \;=\; \delta_C(\mathbf{z}) \;\triangleq\; \begin{cases} 0, & \mathbf{z} \in C \\ +\infty, & \mathbf{z} \notin C \end{cases}$$

$$\mathbf{M} = \begin{bmatrix} \begin{bmatrix} \mathrm{diag}(\mathbf{A}_1) \end{bmatrix} \\ \begin{bmatrix} \mathrm{diag}(\mathbf{A}_2) \end{bmatrix} \\ \vdots \\ \begin{bmatrix} \mathrm{diag}(\mathbf{A}_m) \end{bmatrix} \end{bmatrix} ,$$

where, for any vector $\mathbf{v} \in \mathbb{R}^n$, $\mathrm{diag}(\mathbf{v})$ denotes the $n \times n$ matrix $\mathbf{D}$ with entries $d_{jj} = v_j$, $j=1,\ldots,n,$ along the diagonal, and zeroes elsewhere.

Lemma 1. With the above choices of f, g, and $\mathbf{M}$, the problem (P),

$$\text{minimize}_{\;\mathbf{x}\in\mathbb{R}^n} \;\; f(\mathbf{x}) + g(\mathbf{M}\mathbf{x}) \quad ,$$

is equivalent to (MP) in the sense that $\mathbf{x}$ is feasible for (MP) if and only if $f(\mathbf{x}) + g(\mathbf{M}\mathbf{x}) < \infty$, and for all feasible $\mathbf{x}$, $f(\mathbf{x}) + g(\mathbf{M}\mathbf{x}) = \sum_{j=1}^{n} h_j(x_j)$.

Proof. A vector $\mathbf{x}$ is feasible for (MP) if and only if $h_j(x_j) < \infty$ for all j, and $\mathbf{A}\mathbf{x} = \mathbf{b}$. Now, $[\mathbf{M}\mathbf{x}]_{ij} = a_{ij}x_j$, and $a_{ij} = 0$ for all $(i, j) \notin Q$, so

$$g(\mathbf{Mx}) = \begin{cases} 0\ , & \sum_{j=1}^{n} a_{ij}x_j = b_i\ \forall\ i,\quad a_{ij}x_j = 0\ \ \forall\ (i,j) \notin Q \\ +\infty\ , & \text{otherwise} \end{cases}$$

$$= \begin{cases} 0\ , & \mathbf{Ax} = \mathbf{b} \\ +\infty\ , & \text{otherwise} \end{cases} .$$

Thus, $f(\mathbf{x}) + g(\mathbf{Mx})$ is finite if and only if $\mathbf{x}$ is feasible for (MP), establishing the first claim. For the second claim, note that $\mathbf{x}$ being feasible implies $g(\mathbf{Mx}) = 0$, hence $f(\mathbf{x}) + g(\mathbf{Mx}) = f(\mathbf{x}) = \sum_{j=1}^{n} h_j(x_j)$. ■

Lemma 2. Unless $\mathbf{A}$ has a column that consists entirely of zeroes, the matrix $\mathbf{M}$ defined above has full column rank.

Proof. Let $\mathbf{x} \in \mathbb{R}^n$ be such that $\mathbf{Mx} = \mathbf{0}$. Then $a_{ij}x_j = 0$ for all i and j. As $\mathbf{A}$ is supposed to have no zero columns, there exists, for every j, an $i(j) \in \{1,\ldots, m\}$ such that $a_{i(j)j} \neq 0$. It follows that $x_j = 0$ for all j, and that $\mathbf{x}$ is the zero vector. ■

For any $\mathbf{z} \in \mathbb{R}^{mn}$, and $1 \leq i \leq m$, let $\mathbf{z}_i \in \mathbb{R}^n$ be the subvector of $\mathbf{z}$ with components z_{ij}, $j=1,\ldots,n$. Let

$$C_i = \{\ \mathbf{z}_i \in \mathbb{R}^n \ \mid\ \sum_{j=1}^{n} z_{ij} = b_i,\ z_{ij} = 0\ \ \forall\ \ j \notin Q_i\}\ ,$$

so that $C = C_1 \times \ldots \times C_m$.

We now apply the genernalize alternating direction method of multipliers to this formulation. We let $v_k = 0$ for all k, so that the method reduce to

$$\text{Choose } \mathbf{x}^{k+1}\text{: } \left\| \mathbf{x}^{k+1} - \arg\min_{\mathbf{x}} \{f(\mathbf{x}) + \langle \mathbf{p}^k, \mathbf{M}\mathbf{x}\rangle + \tfrac{\lambda}{2}\| \mathbf{M}\mathbf{x} - \mathbf{z}^k\|^2\} \right\| \le \mu_k$$

$$\mathbf{z}^{k+1} = \arg\min_{\mathbf{z}} \{g(\mathbf{z}) - \langle \mathbf{p}^k, \mathbf{z}\rangle + \tfrac{\lambda}{2}\| \rho_k\mathbf{M}\mathbf{x}^{k+1} + (1-\rho_k)\mathbf{z}^k - \mathbf{z}\|^2\}$$

$$\mathbf{p}^{k+1} = \mathbf{p}^k + \lambda(\rho_k\mathbf{M}\mathbf{x}^{k+1} + (1-\rho_k)\mathbf{z}^k - \mathbf{z}^{k+1}) \quad .$$

Consider first the task of choosing $\mathbf{x}^{k+1}$. By the separability of f and the form of $\mathbf{M}$, this problem can be decomposed into n independent tasks of the form

$$\text{Choose } x_j^{k+1}\text{: } \left\| x_j^{k+1} - \arg\min_{x_j} \left\{ h_j(x_j) + \left(\sum_{i=1}^{m} p_{ij}^k a_{ij} \right) x_j + \frac{\lambda}{2} \sum_{i:j\in Q_i} (a_{ij}x_j - z_{ij}^k)^2 \right\} \right\| \le \varepsilon_k \quad ,$$

where we let $\varepsilon_k = n^{(-1/2)}\mu_k$. Now consider the computation of $\mathbf{z}^{k+1}$, which decomposes into the m independent calculations

$$\mathbf{z}_i^{k+1} = \arg\min_{\mathbf{z}_i\in C_i} \left\{ -\langle \mathbf{p}_i^k, \mathbf{z}_i\rangle + \frac{\lambda}{2} \sum_{j=1}^{n} (\rho_k a_{ij} x_j^{k+1} + (1-\rho_k) z_{ij}^k - z_{ij})^2 \right\}, \quad i=1,\ldots,m \quad .$$

To solve this problem for each i, we attach a single Lagrange multiplier π_i^{k+1} to the constraint $\sum_{j=1}^{n} z_{ij} = b_i$ which defines C_i. At the optimum, the Karush/Kuhn-Tucker conditions give, for all i and j such that $j \notin Q_i$, that

$$\frac{\partial}{\partial z_{ij}} \left\{ -\langle \mathbf{p}_i^k, \mathbf{z}_i\rangle + \pi_i^{k+1}(b_i - \sum_{j=1}^{n} z_{ij}) + \frac{\lambda}{2} \sum_{j=1}^{n} (\rho_k a_{ij} x_j^{k+1} + (1-\rho_k) z_{ij}^k - z_{ij})^2 \right\} = 0 \quad .$$

Equivalently,

$$-p_{ij}{}^k - \pi_i{}^{k+1} + \lambda(z_{ij} - \rho_k a_{ij} x_j{}^{k+1} - (1-\rho_k) z_{ij}{}^k) = 0 \qquad \forall\, j \in Q_i \ ,$$

and so we obtain that

$$z_{ij}^{k+1} = \begin{cases} \rho_k a_{ij} x_j^{k+1} + (1-\rho_k) z_{ij}^k + \dfrac{p_{ij}^k + \pi_i^{k+1}}{\lambda} , & j \in Q_i \\ 0 , & j \notin Q_i \ . \end{cases}$$

We determine the values of the π_i^{k+1} by setting, for each i,

$$b_i = \sum_{j=1}^{n} z_{ij}^{k+1} = \rho_k\left(\sum_{j \in Q_i} a_{ij}x_j^{k+1}\right) + (1-\rho_k)\left(\sum_{j \in Q_i} z_{ij}^k\right) + \frac{1}{\lambda}\left(\sum_{j \in Q_i} p_{ij}^k + \pi_i^{k+1}\right) .$$

For all $k \geq 1$, $\sum_{j \in Q_i} z_{ij}^k = b_i$ by construction. Assume that $\mathbf{z}^0$ is chosen so that $\sum_{j \in Q_i} z_{ij}^0 = b_i$ for all i. Then, for all k, we can simplify the above equation to

$$\rho_k b_i = \rho_k \sum_{j \in Q_i} a_{ij}x_j^{k+1} + \frac{1}{\lambda}\sum_{j \in Q_i} \left(p_{ij}^k + \pi_i^{k+1}\right) .$$

Solving for π_i^{k+1} and using the definition of $r_i(\mathbf{x})$,

$$\pi_i^{k+1} = \frac{1}{q_i}\left(\lambda \rho_k r_i(\mathbf{x}^{k+1}) - \sum_{j \in Q_i} p_{ij}^k\right) .$$

The final step in the generalized alternating direction method of multipliers iteration is the update of the multipliers $\mathbf{p}$, which takes the form

$$\mathbf{p}^{k+1} = \mathbf{p}^k + \lambda(\rho_k \mathbf{M}\mathbf{x}^{k+1} + (1-\rho_k)\mathbf{z}^k - \mathbf{z}^{k+1}) .$$

For $(i, j) \notin Q$, this multiplier update formula implies $p_{ij}^{k+1} = p_{ij}^k$. For $(i, j) \in Q$, we have

$$\begin{aligned} p_{ij}^{k+1} &= p_{ij}^k + \lambda(\rho_k a_{ij}x_j^{k+1} + (1-\rho_k)z_{ij}^k - z_{ij}^{k+1}) \\ &= p_{ij}^k + \lambda(-\tfrac{1}{\lambda}[p_{ij}^k + \pi_i^{k+1}]) \\ &= -\pi_i^{k+1} . \end{aligned}$$

Thus the $p_{ij}{}^{k+1}$, $(i, j) \in Q$, do not vary with j. This is no coincidence, but is a natural consequence of the structure of g and the monotone operator splitting method underlying the generalized alternating direction method of multipliers (Eckstein 1989).

Now assume that $\mathbf{p}^0$ is such that for some $\boldsymbol{\pi}^0 \in \mathbb{R}^m$, $p_{ij}{}^0 = -\pi_i{}^0$ for all $(i, j) \in Q$. It then follows $p_{ij}{}^k = -\pi_i{}^k$ for all $k \geq 0$ and $(i, j) \in Q$. Noting that the values of the $p_{ij}{}^k$, $(i, j) \notin Q$, are inconsequential, one can then substitute $-\pi_i^k$ for p_{ij}^k everywhere, and obtain the overall method

$$\text{Choose } x_j^{k+1}: \left\| x_j^{k+1} - \arg\min_{x_j} \left\{ h_j(x_j) - \langle \mathbf{a}_j, \boldsymbol{\pi}^k \rangle x_j + \frac{\lambda}{2} \sum_{i:j \in Q_i} (a_{ij}x_j - z_{ij}^k)^2 \right\} \right\| \leq \varepsilon_k$$

$$\pi_i^{k+1} = \pi_i^k + \frac{\lambda \rho_k}{q_i} r_i(\mathbf{x}^{k+1})$$

$$z_{ij}^{k+1} = \begin{cases} \rho_k a_{ij} x_j^{k+1} + (1-\rho_k) z_{ij}^k - \frac{1}{q_i} r_i(\mathbf{x}^{k+1}), & (i,j) \in Q \\ 0, & (i,j) \notin Q \end{cases} .$$

Just as the sequence of mn-dimensional multipliers $\{\mathbf{p}^k\}$ can be replaced by the m-dimensional sequence $\{\boldsymbol{\pi}^k\}$, the variables $\{\mathbf{z}^k\} \subseteq \mathbb{R}^{mn}$ can be replaced by a sequence $\{\mathbf{y}^k\} \subseteq \mathbb{R}^n$, as follows:

Lemma 3. Assume that $\mathbf{z}^0$ is of the form

$$z_{ij}^0 = \begin{cases} a_{ij} y_j^0 - \frac{1}{q_i} r_i(\mathbf{y}^0), & (i,j) \in Q \\ 0, & (i,j) \notin Q \end{cases}$$

for some $\mathbf{y}^0 \in \mathbb{R}^n$. Define the sequence $\{\mathbf{y}^k\}$ via the recursion

$$\mathbf{y}^{k+1} = (1-\rho_k)\mathbf{y}^k + \rho_k \mathbf{x}^{k+1} \qquad \forall\, k \geq 0 .$$

Then

$$z_{ij}^k = a_{ij} y_j^k - \frac{1}{q_i} r_i(\mathbf{y}^k) \qquad \forall\, (i,j) \in Q, \ \ \forall\, k \geq 0 .$$

Proof. The statement holds by hypothesis for $k = 0$. Assuming it holds for k, we have, for all $(i,j) \in Q$,

$$\begin{aligned} & a_{ij} y_j^{k+1} - \frac{1}{q_i} r_i(\mathbf{y}^{k+1}) \\ = \ & a_{ij}((1-\rho_k) y_j^k + \rho_k x_j^{k+1}) - \frac{1}{q_i}(b_i - \langle \mathbf{A}_i, \mathbf{y}^{k+1} \rangle) \\ = \ & (1-\rho_k) a_{ij} y_j^k + \rho_k a_{ij} x_j^{k+1} - \frac{1}{q_i}((1-\rho_k)(b_i - \langle \mathbf{A}_i, \mathbf{y}^k \rangle) + \rho_k(b_i - \langle \mathbf{A}_i, \mathbf{x}^{k+1} \rangle)) \end{aligned}$$

$$= (1-\rho_k)[a_{ij}y_j^k - \tfrac{1}{q_i} r_i(\mathbf{y}^k)] \ + \ \rho_k a_{ij} x_j^{k+1} \ - \ \tfrac{\rho_k}{q_i} r_i(\mathbf{x}^{k+1})$$

$$= (1-\rho_k) z_{ij}^k \ + \ \rho_k a_{ij} x_j^{k+1} \ - \ \tfrac{\rho_k}{q_i} r_i(\mathbf{x}^{k+1})$$

$$= z_{ij}^{k+1} \quad .$$

So, the claim holds by induction. ■

Under the assumptions of Lemma 3, one can eliminate the z_{ij}^k and use that $(i, j) \notin Q$ implies $a_{ij} = 0$, obtaining

$$\left\| x_j^{k+1} - \arg\min_{x_j} \left\{ h_j(x_j) - \langle \mathbf{a}_j, \boldsymbol{\pi}^k \rangle x_j + \frac{\lambda}{2} \sum_{i:j \in Q_i} \left(a_{ij}x_j - \left(a_{ij}y_j^k + \frac{1}{q_i} r_i(\mathbf{y}^k)\right)\right)^2 \right\} \right\| \le \varepsilon_k$$

$$\pi_i^{k+1} = \pi_i^k + \frac{\lambda}{q_i} r_i(\mathbf{x}^{k+1})$$

$$y_j^{k+1} = (1-\rho_k) y_j^{k+1} + \rho_k x_j^{k+1} \qquad .$$

Collecting terms in the squared expressions, we obtain the *alternating step method* for monotropic programming:

$$\begin{aligned} &\left\| x_j^{k+1} - \arg\min_{x_j} \left\{ h_j(x_j) - \langle \mathbf{a}_j, \boldsymbol{\pi}^k \rangle x_j + \frac{\lambda \|\mathbf{a}_j\|^2}{2} \left(x_j - \left[y_j^k + \frac{1}{\|\mathbf{a}_j\|^2} \sum_{i=1}^{m} \frac{a_{ij} r_i(\mathbf{y}^k)}{q_i} \right] \right)^2 \right\} \right\| \le \varepsilon_k \\ &\pi_i^{k+1} = \pi_i^k + \frac{\lambda \rho_k}{q_i} r_i(\mathbf{x}^{k+1}) \\ &\mathbf{y}^{k+1} = (1-\rho_k)\mathbf{y}^k + \rho_k \mathbf{x}^{k+1} \quad , \end{aligned} \qquad \text{(ASMP)}$$

where $\mathbf{y}^0 \in \mathbb{R}^n$ and $\boldsymbol{\pi}^0 \in \mathbb{R}^m$ are completely arbitrary. Note that there is no longer any direct reference to the sets Q or Q_i. We merely allow q_i to be any integer between $d(i)$ and n, inclusive. In the degenerate case that $d(i)$=0 for some i, we must require that q_i be at least 1 so that (ASMP) remains well-defined. In practice, any such constraints of degree zero can be immediately eliminated from the problem.

In the special case that $\rho_k = 1$ for all k, the sequences $\{\mathbf{x}^k\}$ and $\{\mathbf{y}^k\}$, are identical, and the method can be expressed

$$x_j^{k+1} = \arg\min_{x_j} \left\{ h_j(x_j) - \langle \mathbf{a}_j, \boldsymbol{\pi}^k \rangle x_j + \frac{\lambda \|\mathbf{a}_j\|^2}{2} \left(x_j - \left[x_j^k + \frac{1}{\|\mathbf{a}_j\|^2} \sum_{i=1}^{m} \frac{a_{ij} r_i(\mathbf{x}^k)}{q_i} \right] \right)^2 \right\}$$

$$\pi_i^{k+1} = \pi_i^k + \frac{\lambda}{q_i} r_i(\mathbf{x}^{k+1}) \quad .$$

We use the term *alternating step method* because, in this form, the algorithm makes alternate updates to the variables $\mathbf{x}$ of (MP) and those of its dual (Rockafellar 1984, Chapter 11A)

$$\text{minimize}_{\boldsymbol{\pi} \in \mathbb{R}^m} \sum_{j=1}^{n} h_j^*(\mathbf{a}_j^\top \boldsymbol{\pi}) - \mathbf{b}^\top \boldsymbol{\pi} \quad ,$$

where h_j^* denotes the convex conjugate of h_j.

The updates of the primal variables x_j and y_j are completely independent over j, and the updates of the dual variables π_i are likewise independent over i. Thus, the method has the potential for massive parallelism.

We now seek to apply Theorem 1 to show that (ASMP) converges. Theorem 1 requires the existence of a Kuhn-Tucker pair, so we address this issue first. For any j, we say that h_j is *pseudopolyhedral* if

$$h_j(x) = \begin{cases} \bar{h}_j(x) , & x \in V_j \\ +\infty , & x \notin V_j \end{cases} ,$$

where $\bar{h}_j$ is finite and convex throughout $\mathbb{R}$, and each $V_j \subseteq \mathbb{R}$ is a closed (but not necessarily bounded) interval. For an explanation of this terminology, see Eckstein (1989), Section 3.5.1.

Lemma 4. If the h_j are pseudopolyhedral for all j, then for the above definition of f, g, and $\mathbf{M}$, (P) has a Kuhn-Tucker pair if and only if (MP) has a finite-valued optimal solution.

Proof. If (P) has a Kuhn-Tucker pair, then it has a finite optimal solution $\mathbf{x}^*$, and, by Lemma 1, $\mathbf{x}^*$ must also be optimal for (MP). Conversely, suppose, that (MP) has a finite-valued optimal solution $\mathbf{x}^*$. Then $\mathbf{x}^*$ must be optimal for (P), and hence $\mathbf{0} \in \partial[f + g \circ \mathbf{M}](\mathbf{x}^*)$. Writing f as $\overline{f} + \delta_V$, where

$$\overline{f}(\mathbf{x}) = \sum_{j=1}^{n} \overline{h}_j(x_j)$$

and

$$\delta_V(\mathbf{x}) = \begin{cases} 0 , & x_j \in V_j,\ j=1,\dots,n \\ +\infty , & \text{otherwise} , \end{cases}$$

we have that δ_V and $g \circ \mathbf{M}$ are polyhedral functions, and, as (MP) is assumed to be feasible,

$$\begin{aligned} & \operatorname{ri}(\operatorname{dom} f) \cap \operatorname{dom} \delta_V \cap \operatorname{dom}(g \circ \mathbf{M}) \\ = \ & \mathbb{R}^n \cap (V_1 \times \dots \times V_n) \cap \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} = \mathbf{b}\} \\ \neq \ & \varnothing . \end{aligned}$$

Hence, (Rockafellar 1970, Theorem 23.8)

$$\begin{aligned} & \partial[f + g \circ \mathbf{M}](\mathbf{x}^*) \\ = \ & \partial[\overline{f} + \delta_V + g \circ \mathbf{M}](\mathbf{x}^*) \\ = \ & \partial \overline{f}(\mathbf{x}^*) + \partial \delta_V(\mathbf{x}^*) + \partial[g \circ \mathbf{M}](\mathbf{x}^*) \\ = \ & \partial f(\mathbf{x}^*) + \partial[g \circ \mathbf{M}](\mathbf{x}^*) . \end{aligned}$$

As $g \circ \mathbf{M}$ is polyhedral, $\partial[g \circ \mathbf{M}](\mathbf{x}^*) = \mathbf{M}^{\mathrm{T}} \partial g(\mathbf{x}^*)$ (Rockafellar 1970, Theorem 23.9), and so $\mathbf{0} \in \partial f(\mathbf{x}^*) + \mathbf{M}^{\mathrm{T}} \partial g(\mathbf{x}^*)$. This inclusion implies that there exists some $\mathbf{p}^*$ such that $(\mathbf{x}^*, \mathbf{p}^*)$ is a Kuhn-Tucker pair. ■

Theorem 2. Suppose one is given a monotropic program in the form (MP), where $\mathbf{A}$ has no column consisting entirely of zeroes, and that h_j is pseudopolyhedral for all j. Let $\{\varepsilon_k\}_{k=0}^{\infty} \subseteq [0, \infty)$ and $\{\rho_k\}_{k=0}^{\infty} \subseteq (0, 2)$ be sequences such that

$$\sum_{k=0}^{\infty} \varepsilon_k < \infty \qquad 0 < \inf_{k \geq 0} \rho_k \leq \sup_{k \geq 0} \rho_k < 2 \ .$$

Then, if (MP) has a finite-valued optimal solution, the sequence $\{\mathbf{x}^k\}$ generated by the alternating step method converges to a solution $\mathbf{x}^*$ of (MP), $\{\boldsymbol{\pi}^k\}$ converges to an optimal solution $\boldsymbol{\pi}^*$ of the dual problem (DMP), and $\mathbf{a}_j^{\mathrm{T}}\boldsymbol{\pi}^* \in \partial h_j(x_j^*)$ for all j. If (MP) does not have a finite-valued optimal soltution, and we also have that the conjugate functions h_j^* are pseudopolyhedral, at least one of the sequences $\{\mathbf{y}^k\}$ and $\{\boldsymbol{\pi}^k\}$ is unbounded.

Proof. Given the integers q_i, where $d(i) \leq q_i \leq n$ for $i = 1,\ldots,m$, choose the sets Q_i in the definition of C and g above such that $|Q_i| = q_i$ for all i. Lemma 1 then assures the equivalence of (P) and (MP), with the above choice of f and $\mathbf{M}$. Since $\mathbf{A}$ has no zero column, Lemma 2 gives that $\mathbf{M}$ has full column rank. Furthermore, by the analysis above, if we choose $p_{ij}{}^0 = -\pi_i{}^0$ for all i and j, and

$$z_{ij}^0 = \begin{cases} a_{ij}y_j^0 - \dfrac{1}{q_i}\, r_i(\mathbf{y}^0), & (i,j) \in Q \\ 0, & (i,j) \notin Q \ , \end{cases}$$

then the sequences $\{\mathbf{x}^k\}$ evolved by the generalized alternating direction method of multipliers and (ASMP) are identical, and for the sequences $\{\mathbf{p}^k\}$ and $\{\mathbf{z}^k\}$, we have $p_{ij}{}^k = -\pi_i{}^k$ for all $k \geq 0$ and $(i, j) \in Q$, while for all $k \geq 0$,

$$z_{ij}^k = \begin{cases} a_{ij}y_j^k - \dfrac{1}{q_i}\, r_i(\mathbf{y}^k), & (i,j) \in Q \\ 0, & (i,j) \notin Q \ . \end{cases}$$

Now consider the case that (MP) has a finite-valued optimal solution. By Lemma 4, (P) has a Kuhn-Tucker pair, so Theorem 1 implies that $\{\mathbf{x}^k\}$, $\{\mathbf{p}^k\}$, and $\{\mathbf{z}^k\}$ converge. In particular, $\{\mathbf{x}^k\}$ converges to a solution of (P), hence to a solution of (MP). The convergence of $\{\mathbf{p}^k\}$ implies the convergence of $\{\boldsymbol{\pi}^k\}$. Let

$$\mathbf{x}^* = \lim_{k \to \infty} \mathbf{x}^k \qquad \boldsymbol{\pi}^* = \lim_{k \to \infty} \boldsymbol{\pi}^k \qquad .$$

We now prove the dual optimality of $\boldsymbol{\pi}^*$. Letting

$$\mathbf{s}^k = \mathbf{x}^{k+1} - \arg\min_{\mathbf{x}} \{f(\mathbf{x}) + \langle \mathbf{p}^k, \mathbf{Mx}\rangle + \tfrac{\lambda}{2}\| \mathbf{Mx} - \mathbf{z}^k \|^2\} \quad \forall\, k \geq 0 ,$$

whence $\mathbf{s}^k \to \mathbf{0}$, we have, for all k and j,

$$0 \in \partial h_j(x_j^{k+1} - s_j^k) - \mathbf{a}_j^{\mathrm{T}}\boldsymbol{\pi}^k + \lambda \sum_{i=1}^{n} a_{ij}\left(a_{ij}(x_j^{k+1} - s_j^k) - z_{ij}^k\right) ,$$

that is,

$$\mathbf{a}_j^{\mathrm{T}}\boldsymbol{\pi}^k + \lambda \sum_{i=1}^{n} a_{ij}\left(a_{ij}(x_j^{k+1} - s_j^k) - z_{ij}^k\right) \in \partial h_j(x_j^{k+1} - s_j^k) .$$

Since the h_j are closed convex functions, the graphs of their subgradient mappings ∂h_j are closed sets in $\mathbb{R}^2$, and we may take the limits as $k \to \infty$, yielding $\mathbf{a}_j^{\mathrm{T}}\boldsymbol{\pi}^* \in \partial h_j(x_j^*)$ for all j, where we have also used that $\mathbf{z}^k \to \mathbf{Mx}^*$. From the definition of the conjugacy operation, one has that

$$f^*(\mathbf{x}) = \sum_{j=1}^{n} h^*_j(x_j) ,$$

and that the dual problem (DMP) is that of minimizing $f^*(\mathbf{A}^{\mathrm{T}}\boldsymbol{\pi}) - \mathbf{b}^{\mathrm{T}}\boldsymbol{\pi}$. From $\mathbf{a}_j^{\mathrm{T}}\boldsymbol{\pi}^* \in \partial h_j(x_j^*)$ for all j, it follows that $x_j^* \in \partial h_j^*(\mathbf{a}_j^{\mathrm{T}}\boldsymbol{\pi}^*)$ for all j (Rockafellar 1970, 1984), and thus that $\mathbf{x}^* \in \partial f^*(\mathbf{A}^{\mathrm{T}}\boldsymbol{\pi})$. Rockafellar (1970), Theorem 23.9, gives $\mathbf{Ax}^* \in \partial[f^* \circ \mathbf{A}^{\mathrm{T}}](\boldsymbol{\pi})$. Thus,

$$\partial_{\boldsymbol{\pi}}\left[f^*(\mathbf{A}^{\mathrm{T}}\boldsymbol{\pi}) - \mathbf{b}^{\mathrm{T}}\boldsymbol{\pi}\right]_{\boldsymbol{\pi} = \boldsymbol{\pi}^*} \ni \mathbf{Ax}^* - \mathbf{b} = \mathbf{0} ,$$

and $\boldsymbol{\pi}^*$ is optimal for (DMP). Here, the notation "$\partial_{\boldsymbol{\pi}}$" denotes the subgradient with respect to the variable $\boldsymbol{\pi}$. This concludes the proof in the case that (MP) has a finite-valued optimal solution.

Now suppose that (MP) has no finite-valued optimal solution, and also that the h_j^* are pseudopolyhedral. Then it follows by an argument almost identical to that of Lemma 4, but applied to (D) instead of (P), that (D) has no solution, for otherwise a Kuhn-Tucker pair would exist, and (P) would have a solution. Theorem 1 then implies that at least one of the sequences $\{\mathbf{p}^k\}$ or $\{\mathbf{z}^k\}$ is unbounded. If $\{\mathbf{p}^k\}$ is unbounded, then $\{\boldsymbol{\pi}^k\}$ is unbounded. If $\{\mathbf{z}^k\}$ is unbounded, we have from

$$z_{ij}^k = a_{ij}y_j^k - \frac{1}{q_i} r_i(\mathbf{y}^k) \ , \quad k \geq 0, \ (i,j) \in Q,$$

that $\{\mathbf{y}^k\}$ must be unbounded. ■

Note that in the case where $\rho_k = 1$ for all sufficiently large k, the unboundedness of $\{\mathbf{y}^k\}$ implies the unboundedness of $\{\mathbf{x}^k\}$.

The theorem does not cover the degenerate case in which $\mathbf{A}$ contains entirely zero columns. The variables x_j associated with such columns may be set to their optimal values by a simple one-dimensional minimization of the corresponding h_j, and then removed from the problem.

The alternating step method is based on an alternating direction iteration in which some variables, namely $\{\mathbf{p}^k\}$ and $\{\mathbf{z}^k\}$, have dimension mn. However, both methods reduce to a form in which all the variables have dimension either m or n. A convergence proof involving only the lower-dimensional sequences $\{\mathbf{x}^k\} \in \mathbb{R}^n$, $\{\mathbf{y}^k\} \in \mathbb{R}^n$, $\{\boldsymbol{\pi}^k\} \in \mathbb{R}^m$, and $\{\mathbf{r}(\mathbf{x}^k)\} = \{(r_1(\mathbf{x}^k),....r_m(\mathbf{x}^k))\} \in \mathbb{R}^m$, and also allowing the q_i to be non-integer, would be very appealing. At present, it does not appear that any such alternate proof exists, and the topic must be left open for future research. In practice, the question of non-integer q_i appears to be moot, since it seems optimal to set q_i to the minimum possible value, $d(i)$, thus allowing the method to take the largest possible steps (see below).

4. Linear Programming

We now specialize the method of the previous section to linear programs of the form

$$\begin{aligned} \text{minimize} \quad & \mathbf{c}^{\mathsf{T}}\mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{l} \le \mathbf{x} \le \mathbf{u} \quad , \end{aligned} \tag{LP}$$

where $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{l} \in [-\infty, \infty)^n$, $\mathbf{u} \in (-\infty, \infty]^n$, and $\mathbf{A}$ is an $m \times n$ real matrix. This formulation is completely general, as it subsumes the standard primal form for which $l_j = 0$ and $u_j = \infty$ for j=1,...,n.

Assume that $\mathbf{A}$ has no all-zero rows or columns. Neither of these assumptions is truly restrictive: all-zero rows either make the linear program infeasible (if the corresponding b_i is non-zero), or can be immediately discarded (if the corresponding b_i is zero). Much as in general monotropic programming, all-zero columns can quickly be removed from the problem by setting the corresponding variables to their upper or lower bounds, depending on the signs of the corresponding cost coefficients.

For each j=1,...,n, define $V(j)$ to be the (possibly unbounded) real interval $[l_j, u_j] \cap \mathbb{R}$, the range of permissible values for x_j. Then (LP) can be put into the standard monotropic programming form (MP) by setting

$$h_j(x_j) = \begin{cases} c_j x_j, & x_j \in V(j) \\ +\infty, & x_j \notin V(j) \end{cases}, \quad j=1,\ldots,n \quad .$$

Let $\bar{c}_j(\boldsymbol{\pi})$ denote the reduced cost coefficient of the variable x_j with respect to the dual variables $\boldsymbol{\pi}$, that is,

$$\bar{c}_j(\boldsymbol{\pi}) = c_j - \boldsymbol{\pi}^{\mathsf{T}}\mathbf{a}_j \quad ,$$

and let $\bar{\mathbf{c}}(\boldsymbol{\pi})$ denote the vector $\mathbf{c} - \mathbf{A}^{\mathsf{T}}\boldsymbol{\pi}$ of these coefficients.

In applying the alternating step method to (MP) with this choice of objective function, the calculation of each x_j^{k+1} becomes a one-dimensional quadratic problem, which can easily be done exactly. Thus we set the tolerance parameter ε_k to zero for all k. The calculation of x_j^{k+1} then takes the form

$$x_j^{k+1} = \arg\min_{x_j \in V(j)} \left\{ \bar{c}_j(\boldsymbol{\pi}^k)x_j + \frac{\lambda\|\mathbf{a}_j\|^2}{2}\left(x_j - \left[y_j^k + \frac{1}{\|\mathbf{a}_j\|^2}\sum_{i=1}^{m}\frac{a_{ij}r_i(\mathbf{y}^k)}{q_i}\right]\right)^2\right\} \quad .$$

Setting the derivative of the minimand to zero, and projecting onto the interval $V(j)$, we obtain the *alternating step method for linear programming*,

$$\boxed{\begin{aligned}
x_j^{k+1} &= P_{V(j)}\left[y_j^k + \frac{1}{\|\mathbf{a}_j\|^2}\left(\left[\sum_{i=1}^{m}\frac{a_{ij}r_i(\mathbf{y}^k)}{q_i}\right] - \frac{\bar{c}_j(\boldsymbol{\pi}^k)}{\lambda}\right)\right] && j=1,\ldots,n \\
y_j^{k+1} &= (1-\rho_k)y_j^k + \rho_k x_j^{k+1} && j=1,\ldots,n \\
\pi_i^{k+1} &= \pi_i^k + \frac{\lambda\rho_k}{q_i}r_i(\mathbf{x}^{k+1}) && i=1,\ldots,m \quad ,
\end{aligned}} \qquad \text{(ASLP)}$$

where the function $P_{V(j)}$ denotes projection onto $V(j)$, that is

$$P_{V(j)}(t) \;=\; \max\{l_j, \min\{t, u_j\}\} \quad .$$

As before, $\mathbf{y}^0 \in \mathbb{R}^n$ and $\boldsymbol{\pi}^0 \in \mathbb{R}^m$ are arbitrary, and the $\{q_i\}_{k=1}^{m}$ are integers such that $d(i) \le q_i \le n$ for all i. Setting $\rho_k = 1$ for all k yields the simpler method

$$\boxed{\begin{aligned}
x_j^{k+1} &= P_{V(j)}\left[x_j^k + \frac{1}{\|\mathbf{a}_j\|^2}\left(\left[\sum_{i=1}^{m}\frac{a_{ij}r_i(\mathbf{x}^k)}{q_i}\right] - \frac{\bar{c}_j(\boldsymbol{\pi}^k)}{\lambda}\right)\right] && j=1,\ldots,n \\
\pi_i^{k+1} &= \pi_i^k + \frac{\lambda}{q_i}r_i(\mathbf{x}^{k+1}) && i=1,\ldots,m \quad .
\end{aligned}}$$

Theorem 3. Suppose that the matrix **A** of problem (LP) has no all-zero rows or columns. Then for any starting point $\mathbf{x}^0 \in \mathbb{R}^n$ and $\boldsymbol{\pi}^0 \in \mathbb{R}^m$, any integers $\{q_i\}_{k=1}^{m}$, where $d(i) \leq q_i \leq n$ for all i, and sequence $\{\rho_k\}_{k=0}^{\infty} \subseteq (0, 2)$ such that

$$0 < \inf_{k \geq 0} \rho_k \leq \sup_{k \geq 0} \rho_k < 2 \quad ,$$

the method (ASLP) produces a sequence $\{\mathbf{x}^k\}$ converging to an optimal solution $\mathbf{x}^*$ of (LP), if such a solution exists. In this case, the sequence $\{\boldsymbol{\pi}^k\}$ converges to a vector $\boldsymbol{\pi}^*$ of optimal simplex multipliers for (LP), that is

$$\begin{aligned} x_j^* > l_j \quad &\Rightarrow \quad \bar{c}_j(\boldsymbol{\pi}^*) \leq 0 \\ x_j^* < u_j \quad &\Rightarrow \quad \bar{c}_j(\boldsymbol{\pi}^*) \geq 0 \quad . \end{aligned}$$

If, on the other hand, (LP) is infeasible or unbounded, at least one of the sequences $\{\mathbf{x}^k\}$ or $\{\boldsymbol{\pi}^k\}$ is unbounded.

Proof. The choice of h_j above meets the conditions of Theorem 2, so that if (LP) has an optimal solution, $\mathbf{x}^k \to \mathbf{x}^*$ and $\boldsymbol{\pi}^k \to \boldsymbol{\pi}^*$, where $\mathbf{x}^*$ is optimal for (LP) and $\mathbf{a}_j^{\mathsf{T}}\boldsymbol{\pi} \in \partial h_j(x_j^*)$ for all j. It remains to show that $\boldsymbol{\pi}^*$ is a vector of optimal multipliers. Now, the set

$$K_j = \{ (x_j, \delta_j) \mid \delta_j \in \partial h_j(x_j) \} \subseteq \mathbb{R}^2$$

is equal to

$$\Big[\big((l_j, u_j) \times \{c_j\}\big) \cup \big(\{l_j\} \times (-\infty, c_j)\big) \cup \big(\{u_j\} \times (c_j, +\infty)\big) \Big] \cap \mathbb{R}^2 \quad .$$

Figure 1 depicts the set K_j, which constitutes the classical "kilter diagram" for x_j (see also Rockafellar 1984). From this description of ∂h_j, we see that if $x_j^* > l_j$, then $\mathbf{a}_j^{\mathsf{T}}\boldsymbol{\pi} \geq c_j$, that is, $\bar{c}_j(\boldsymbol{\pi}^*) \leq 0$. Likewise, if $x_j^* < u_j$, then $\mathbf{a}_j^{\mathsf{T}}\boldsymbol{\pi} \leq c_j$, that is, $\bar{c}_j(\boldsymbol{\pi}^*) \geq 0$. Now consider the case that (LP) is infeasible or unbounded. The conjugate functions h_j^* are easily seen from their definitions to be pseudopolyhedral (in fact, they are piecewise linear on certain closed intervals, and $+\infty$ outside them; see Rockafellar 1984). Thus, Theorem 2 implies that at least one of $\{\mathbf{y}^k\}$ or $\{\boldsymbol{\pi}^k\}$ is unbounded. ■

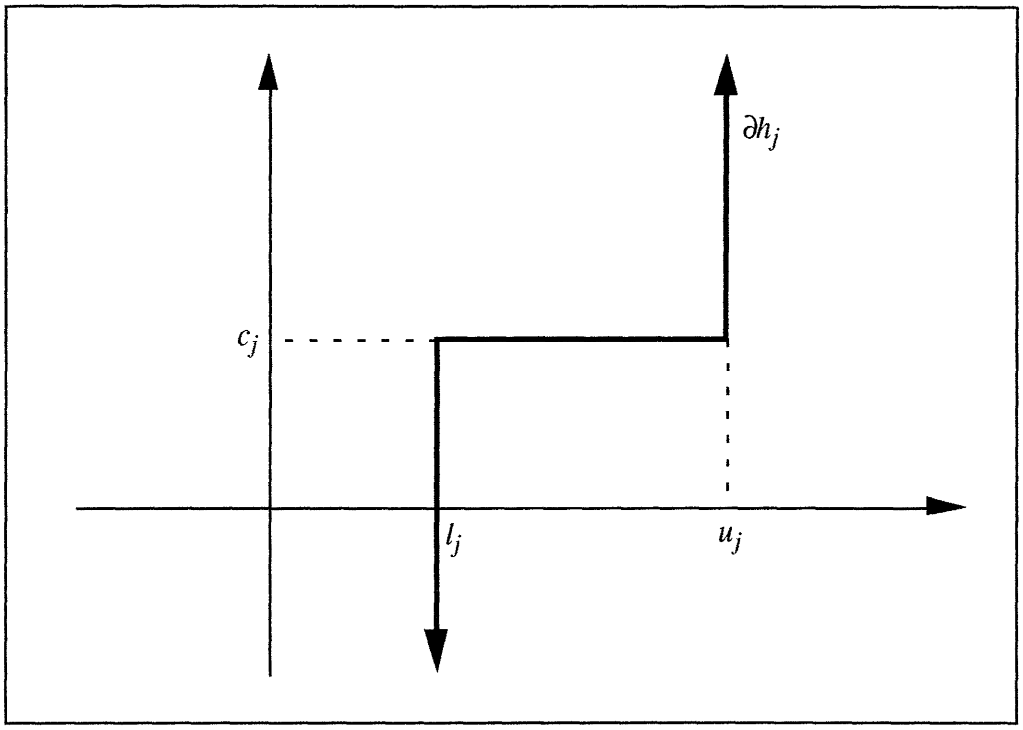

Figure 1. Graph of the "kilter diagram" K_j of ∂h_j for the case in which l_j and u_j are both finite.

We observe, as promised, that the method does not maintain any of the three customary invariants — primal feasibility, dual feasibility, and complementary slackness. It does not even maintain approximate forms of complementary slackness or dual feasibility, as do the ε-relaxation and related algorithms for network optimization (Bertsekas 1985, Bertsekas 1986, Bertsekas and Eckstein 1988, Bertsekas and Tsitsiklis 1989, Goldberg 1987, Goldberg and Tarjan 1987).

The alternating step method bears some resemblance to ε-relaxation and to strictly convex network flow relaxation methods (Bertsekas and El Baz 1987, Bertsekas, Hossein, and Tseng 1987), in that the dual variables π_i are individually updated in response to the amount $r_i(\mathbf{x})$ to which the corresponding constraint is being violated. However, the dual update

$$\pi_i^{k+1} = \pi_i^k + \frac{\rho_k \lambda}{q_i} r_i(\mathbf{x}^{k+1})$$

is unique in that the dual adjustment is *proportional* to the surplus $r_i(\mathbf{x}^{k+1})$, and the dual variable may either rise or fall, depending on the sign of the surplus. In the ε-relaxation and related algorithms, dual variables can be adjusted in only one direction.

To further illuminate the comparison between the alternating step method and recent proposed parallel methods for network optimization, consider the alternating step method specialized to minimum cost network flow problems, with ρ_k fixed at 1. If one assumes that $\mathbf{A}$ is the node-arc incidence matrix of a network and uses the index vw, rather than j, to indicate arc (v, w), one obtains the method

$$x_{vw}^{k+1} = P_{V(vw)}\left[x_{vw}^{k} + \frac{1}{2}\left(\frac{r_v(\mathbf{x}^k)}{d(v)} - \frac{r_w(\mathbf{x}^k)}{d(w)} - \frac{\bar{c}_{vw}(\boldsymbol{\pi}^k)}{\lambda}\right)\right] \qquad \forall \text{ arcs } (v, w)$$

$$\pi_v^{k+1} = \pi_v^k + \frac{\lambda}{d(v)} r_v(\mathbf{x}^{k+1}) \qquad . \qquad \forall \text{ nodes } v \quad .$$

Here, x_{vw} denotes the flow on arc (v, w), π_v is the dual variable associated with node v, and $r_v(\mathbf{x})$ is the amount of flow imbalance at node v. We offer the following interpretation: each node v attempts to "push" away its surplus $r_v(\mathbf{x}^k)$, spreading it evenly among all its incident arcs. If the surplus is negative, the node instead "pulls" flow evenly on all its incident arcs. Each arc (v, w) then takes the two flows suggested by its "start" node v and "end" node w, averages them, makes a further adjustment based on its reduced cost, and then projects the result onto its feasible capacity range $V(vw)$. Each node then makes a price adjustment proportional to the surplus resulting from the new flows.

In summary, while there is some similarity to the alternating "pushing" and price adjustment processes of ε-relaxation, the rules governing these steps are quite different. Furthermore, ε-relaxation has not been generalized to constraint structures more complicated than networks without gains, whereas the alternating step method may be applied to any monotropic program.

Another interesting feature of the alternating step method is that it does not solve a linear system of equations at each iteration, nor does it update a solution to such a system, as does the simplex method. However, by setting $c_j = 0$, $l_j = -\infty$, and $u_j = +\infty$ for all j, it could be used to solve the system of equations $\mathbf{Ax} = \mathbf{b}$. Thus, it is an iterative method that appears to be equally suited to optimizing linear programs or solving linear systems.

5. *Linear Convergence Rate*

We now show that, so long as $\rho_k \le 1$ for all k, the alternating step method for linear programming converges at a globally linear rate, in the sense that the distance of $\{\mathbf{x}^k\}$ to the set of optimal solutions and the distance of $\{\boldsymbol{\pi}^k\}$ to the set of optimal simplex multipliers are bounded above by convergent geometric sequences. Because of the complexity of the analysis, this section assumes familiarity with the proximal point theory developed in Eckstein and Bertsekas (1989).

We have already seen that the alternating step method is equivalent to the generalized alternating direction method of multipliers under certain choices of f, g, and $\mathbf{M}$. From Eckstein and Bertsekas (1989), we also know that this application of the generalized alternating direction method of multipliers is equivalent to the *generalized proximal point algorithm*

$$\mathbf{t}^{k+1} = (1-\rho_k)\mathbf{t}^k + \rho_k(I + c_k S)^{-1}(\mathbf{t}^k) \quad ,$$

where $c_k = 1$ for all $k \ge 0$, and S is a certain set-valued maximal monotone operator. The key to the analysis is that the sequence $\{\mathbf{t}^k\}$ produced by this recursion converges to some *zero* of S, that is, a point $\mathbf{t}^*$ such that $\mathbf{0} \in S(\mathbf{t}^*)$. The set of all such zeroes is denoted $S^{-1}(\mathbf{0})$.

Figure 2 is intended to clarify the role of the relaxation factor ρ_k in the convergence of the method. For $\rho \in (0, 2)$, let

$$\mathbf{t}^{k+1}(\rho) = (1-\rho)\mathbf{t}^k + \rho(I + c_k T)^{-1}(\mathbf{t}^k) \quad .$$

From the figure, it is clear that

$$\| t^{k+1}(1) - t^* \| \leq \| t^{k+1}(\rho) - t^* \|$$

for all $\rho < 1$ and $t^* \in S^{-1}(0)$, so choosing $\rho < 1$ is unlikely to be beneficial. On the other hand, there may be an interval $(1, \bar{\rho}) \subseteq (1, 2)$ on which $\rho \in (1, \bar{\rho})$ implies

$$\text{dist}(t^{k+1}(\rho), S^{-1}(0)) < \text{dist}(t^{k+1}(1), S^{-1}(0)) .$$

Thus, it should be possible for over-relaxation to accelerate convergence.

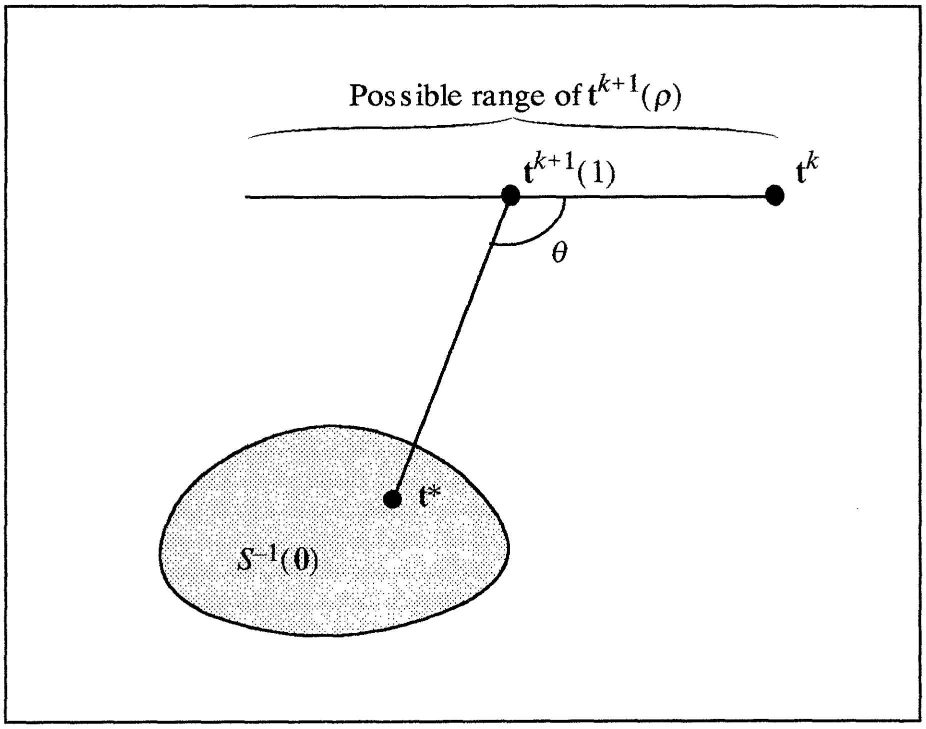

Figure 2. Illustration of the use of the relaxation parameters ρ_k. Here, $t^{k+1}(1) = (I + S)^{-1}(t^k)$, and t^* is an arbitrary member of $S^{-1}(0)$. Because of the maximal monotonicity of S, the angle θ must be at least 90 degrees and the set $S^{-1}(0)$ must be closed and convex.

Referring to Eckstein and Bertsekas (1989), and adopting that paper's notational convention that $S = \{ (t, w) \mid w \in S(t)\}$, that is, that there is no distinction between an operator and its graph, the expression for S is

$$S = \{(v + \lambda b, u - v) \mid (u, b) \in B, (v, a) \in A, v + \lambda a = u - \lambda b\} ,$$

where A and B are the maximal monotone operators

$$A \;=\; \partial[f^* \circ (-\mathbf{M}^\top)] \;=\; -\mathbf{M} \circ (\partial f)^{-1} \circ (-\mathbf{M}^\top)$$
$$B \;=\; \partial g^* = \;(\partial g)^{-1}$$

on $\mathbb{R}^{nm}$. The equivalence of the two expressions for A is guaranteed because the column space of $-\mathbf{M}^\top$ is all of $\mathbb{R}^n$ (Rockafellar 1970, Theorem 23.9). The basic approach to obtaining a linear convergence rate is based on that of Luque (1984), and hinges on the following question: is there a constant α such that $\mathbf{w} \in S(\mathbf{t})$ implies $\text{dist}(\mathbf{t}, S^{-1}(\mathbf{0})) \le \alpha \|\mathbf{w}\|$? To answer this question, we must examine the structure of the operators A, B, and S.

First,

$$\partial f(\mathbf{x}) = \begin{cases} \{\mathbf{c}+\boldsymbol{\xi} \in \mathbb{R}^n \mid x_j < u_j \Rightarrow \xi_j \le 0,\; x_j > l_j \Rightarrow \xi_j \ge 0\}, & \mathbf{l} \le \mathbf{x} \le \mathbf{u} \\ \emptyset, & \text{otherwise}. \end{cases}$$

Thus,

$$(\partial f)^{-1}(\mathbf{y}) = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{l} \le \mathbf{x} \le \mathbf{u},\; y_j < c_j \Rightarrow x_j = l_j,\; y_j > c_j \Rightarrow x_j = u_j \} ,$$

and so for any $\mathbf{q} \in \mathbb{R}^{mn}$,

$$\begin{aligned} A\mathbf{q} \;&=\; -\mathbf{M}(\partial f)^{-1}(-\mathbf{M}^\top\mathbf{q}) \\ &= \Big\{ [-a_{ij}x_j] \in \mathbb{R}^{mn} \,\Big|\, \mathbf{x} \in \mathbb{R}^n, \mathbf{l} \le \mathbf{x} \le \mathbf{u},\; x_j = l_j \;\; \forall j\!: c_j + \textstyle\sum_{i=1}^m a_{ij}q_{ij} > 0, \\ &\qquad\qquad x_j = u_j \;\; \forall j\!: c_j + \textstyle\sum_{i=1}^m a_{ij}q_{ij} < 0 \Big\} . \end{aligned}$$

As for $B = (\partial g)^{-1}$, we have

$$\partial g(\mathbf{q}) \;=\; \begin{cases} V^\perp, & \mathbf{q} \in C \\ \emptyset, & \mathbf{q} \notin C \end{cases} ,$$

where $V^\perp$ is the orthogonal complement of the linear subspace V parallel to C, namely

$$V^\perp \;=\; \{\, \mathbf{p} \in \mathbb{R}^{mn} \mid p_{ij} = p_{il} \;\; \forall\, (i,j), (i,l) \in Q \} \quad .$$

It then follows that for any $\mathbf{p} \in V^\perp$, $B\mathbf{p} = C$, and otherwise $B\mathbf{p}=\emptyset$.

We will now draw a connection with linear programming stability theory. We define $\mathbf{x} \in \mathbb{R}^n$ to be *δ-balanced* if $\mathbf{l} \le \mathbf{x} \le \mathbf{u}$ and $\|\mathbf{r}(\mathbf{x})\|_\infty \le \delta$, that is, if $r_i(\mathbf{x}^k) \le \delta$ for $i = 1,\ldots,m$. We also say that $(\mathbf{x}, \boldsymbol{\pi}) \in \mathbb{R}^n \times \mathbb{R}^m$ obey *ε-complementary slackness* (Tseng and Bertsekas 1987) if $\mathbf{l} \le \mathbf{x} \le \mathbf{u}$ and

$$\begin{aligned} x_j > l_j \quad &\Rightarrow \quad \bar{c}_j(\boldsymbol{\pi}) \le \varepsilon \\ x_j < u_j \quad &\Rightarrow \quad \bar{c}_j(\boldsymbol{\pi}) \ge -\varepsilon \ . \end{aligned}$$

Alternatively, $\mathbf{x}$ is δ-balanced if and only if it is feasible for some problem obtained from (LP) by replacing $\mathbf{b}$ by some $\mathbf{b}'$, where $\| \mathbf{b} - \mathbf{b}' \|_\infty \le \delta$, and $(\mathbf{x}, \boldsymbol{\pi})$ obey ε-complementary slackness if and only if they obey conventional complementary slackness with respect to some costs $\mathbf{c}'$, where $\| \mathbf{c} - \mathbf{c}' \|_\infty \le \varepsilon$.

Choose the sets Q_i as in the proof of Theorem 2. We now state a technical lemma:

Lemma 5. Suppose $(\mathbf{t}, \mathbf{w}) \in S$. Then there exist $\mathbf{p}, \mathbf{q}, \mathbf{s}, \mathbf{z} \in \mathbb{R}^{mn}$, $\mathbf{x} \in \mathbb{R}^n$, and $\boldsymbol{\pi} \in \mathbb{R}^m$ such that

$$\begin{aligned} &(\mathbf{p}, \mathbf{z}) \in B \\ &(\mathbf{q}, \mathbf{s}) \in A \\ &\mathbf{q} + \lambda \mathbf{s} = \mathbf{p} - \lambda \mathbf{z} \\ &\mathbf{q} + \lambda \mathbf{z} = \mathbf{t} \\ &\mathbf{p} - \mathbf{q} = \mathbf{w} \\ &s_{ij} = -a_{ij} x_j \quad \text{for } i = 1,\ldots,m,\ j = 1,\ldots,n \\ &\pi_i = -p_{ij} \quad \text{for all } (i,j) \in Q \ , \end{aligned}$$

and for any $\mathbf{p}, \mathbf{q}, \mathbf{s}, \mathbf{z}, \mathbf{x}$ and $\boldsymbol{\pi}$ satisfying these conditions,

$$\begin{aligned} &\mathbf{x} \quad \text{is } d_{\max}\|\mathbf{w}\|/\lambda\text{-balanced} \\ &(\mathbf{x}, \boldsymbol{\pi}) \quad \text{obeys } a_{\max}\|\mathbf{w}\|\text{-complementary slackness,} \end{aligned}$$

where $a_{\max} \triangleq \max_{j=1,\ldots,n} \{\|\mathbf{a}_j\|\}$ and $d_{\max} \triangleq \max_{i=1,\ldots,m} \{d(i)\}$.

Proof. From the definition of S,

$$S = \{(\mathbf{q} + \lambda \mathbf{z}, \mathbf{p} - \mathbf{q}) \mid (\mathbf{p}, \mathbf{z}) \in B, (\mathbf{q}, \mathbf{s}) \in A, \mathbf{q} + \lambda \mathbf{s} = \mathbf{p} - \lambda \mathbf{z}\} \ ,$$

we immediately have the existence of $\mathbf{p}$, $\mathbf{q}$, s, $\mathbf{z} \in \mathbb{R}^{mn}$ satisfying the first five conditions,

$$\begin{aligned}
&(\mathbf{p}, \mathbf{z}) \in B \\
&(\mathbf{q}, \mathrm{s}) \in A \\
&\mathbf{q}+\lambda \mathrm{s} = \mathbf{p} - \lambda \mathbf{z} \\
&\mathbf{q}+\lambda \mathbf{z} = \mathbf{t} \\
&\mathbf{p} - \mathbf{q} = \mathbf{w} \ .
\end{aligned}$$

In accordance with $(\mathbf{q}, \mathrm{s}) \in A$ and the form of A, let $\mathbf{x}' \in \mathbb{R}^n$ be such that $\mathbf{x}'$ obeys the bounds $\mathbf{l} \leq \mathbf{x}' \leq \mathbf{u}$, $x_j' = l_j$ for all j such that $c_j + \sum_{i=1}^m a_{ij}q_{ij} > 0$, $x_j' = u_j$ for all j such that $c_j + \sum_{i=1}^m a_{ij}q_{ij} < 0$, and $s_{ij} = -\, a_{ij}x_j'$ for all i and j. Given any $\mathbf{x}$ with the property $s_{ij} = -\, a_{ij}x_j$ for all i and j, the fact that $\mathbf{A}$ has no zero columns gives $\mathbf{x}' = \mathbf{x}$. Note that $\mathbf{q}+\lambda\mathrm{s} = \mathbf{p} - \lambda\mathbf{z}$ implies $\lambda(\mathrm{s} + \mathbf{z}) = \mathbf{p} - \mathbf{q} = \mathbf{w}$. We have

$$[\mathrm{s} + \mathbf{z}]_{ij} = z_{ij} - a_{ij}x_j \quad \forall\ i, j\ ,$$

and, as $\|\mathrm{s} + \mathbf{z}\| = \|\mathbf{w}\|/\lambda$,

$$|\, z_{ij} - a_{ij}x_j \,| \ \leq\ \|\mathbf{w}\|/\lambda \quad \forall\ i, j\ ,$$

whence

$$\Bigl|\ \sum_{j:\, a_{ij} \neq 0} (z_{ij} - a_{ij}x_j)\ \Bigr| \ \leq\ d(i)\|\mathbf{w}\|/\lambda \quad \forall\ i\ .$$

Since $(\mathbf{p}, \mathbf{z}) \in B = V^{\perp} \times C$, we must have $\mathbf{z} \in C$, and so $\sum_{j=1}^n z_{ij} = b_i$ for all i. Thus, the last inequality is equivalent to

$$|\, b_i - \mathbf{A}_i^{\mathsf{T}}\mathbf{x} \,| \ =\ |\, r_i(\mathbf{x}) \,| \ \leq\ d(i)\|\mathbf{w}\|/\lambda \quad \forall\ i\ .$$

Thus, using the definition of $d_{\max}$, we conclude that the primal solution $\mathbf{x}$ is $d_{\max}\|\mathbf{w}\|/\lambda$-balanced.

From $(\mathbf{p}, \mathbf{z}) \in B = V^{\perp} \times C$, we also have $\mathbf{p} \in V^{\perp}$, hence $p_{ij} = p_{il}$ for all (i, j), $(i, l) \in Q$. In particular, $p_{ij} = p_{il}$ for all (i, j), (i, l) such that $a_{ij}, a_{il} \neq 0$. Then it is possible to uniquely define $\boldsymbol{\pi} \in \mathbb{R}^m$ via $\pi_i' = -\, p_{ij}$ for all (i, j) such that $a_{ij} \neq 0$. We now consider $\mathbf{p}$ and $\mathbf{q}$. For all j,

$$| \sum_{i=1}^{m} a_{ij}(p_{ij} - q_{ij}) | = | \mathbf{a}_j^{\mathrm{T}}(\boldsymbol{\pi} + \mathbf{q}_j) | \leq \|\mathbf{a}_j\| \cdot \|\boldsymbol{\pi} + \mathbf{q}_j\| ,$$

where $\mathbf{q}_j$ is the m-vector consisting of the $\{q_{ij}\}_{i=1}^{m}$. Since $\|\mathbf{p} - \mathbf{q}\| = \|\mathbf{w}\|$, $\|\boldsymbol{\pi} + \mathbf{q}_j\| \leq \|\mathbf{w}\|$, and thus

$$| \sum_{i=1}^{m} a_{ij}(p_{ij} - q_{ij}) | \leq \|\mathbf{a}_j\| \cdot \|\mathbf{w}\| .$$

Using that $(\mathbf{q}, \mathrm{s}) \in A$, we have

$$x_j < u_j \Rightarrow c_j + \sum_{i=1}^{m} a_{ij}q_{ij} \geq 0 \Leftrightarrow \left(c_j - \sum_{i=1}^{m} a_{ij}(p_{ij} - q_{ij})\right) - \sum_{i=1}^{m} a_{ij}\pi_i \geq 0$$

$$x_j > l_j \Rightarrow c_j + \sum_{i=1}^{m} a_{ij}q_{ij} \leq 0 \Leftrightarrow \left(c_j - \sum_{i=1}^{m} a_{ij}(p_{ij} - q_{ij})\right) - \sum_{i=1}^{m} a_{ij}\pi_i \leq 0 .$$

Thus, $(\mathbf{x}, \boldsymbol{\pi})$ obey complementary slackness if we permit a perturbation of each cost coefficient c_j by an amount $\sum_{i=1}^{m} a_{ij}(p_{ij} - q_{ij})$, whose magnitude cannot exceed $\|\mathbf{a}_j\| \cdot \|\mathbf{w}\|$. Using that $a_{\max} = \max_j \{\|\mathbf{a}_j\|\}$, we conclude $(\mathbf{x}, \boldsymbol{\pi})$ obeys $a_{\max}\|\mathbf{w}\|$-complementary slackness.

■

We now precisely characterize the zeroes of the operator S.

Lemma 6. For the choices of the maximal monotone operators A and B used in the alternating step method for linear programming,

$$S^{-1}(0) = \Big\{ \boldsymbol{\zeta} \in \mathbb{R}^{mn} \ \Big| \ \begin{array}{l} \zeta_{ij} = -\pi_i + \lambda a_{ij}x_j \ \forall\ (i, j) \in Q, \\ \mathbf{x} \text{ is an optimum of (LP)} \\ \boldsymbol{\pi} \text{ is a vector of optimal simplex multipliers for (LP)} \end{array} \Big\} .$$

Proof. Using Theorem 5 of Eckstein and Bertsekas (1989) and the form of A and B,

$$\begin{aligned} S^{-1}(0) &= \{\mathbf{p}+\lambda\mathbf{z} \mid (\mathbf{p}, \mathbf{z}) \in B, (\mathbf{p}, -\mathbf{z}) \in A\} \\ &= \Big\{ [p_{ij} + \lambda a_{ij}x_j] \ \Big| \ \begin{array}{l} p_{ij} = p_{il} \ \forall\ (i, j), (i, l) \in Q, \\ \mathbf{x} \in \mathbb{R}^n, \mathbf{l} \leq \mathbf{x} \leq \mathbf{u}, \\ \sum_{j=1}^{n} a_{ij}x_j = b_i \ \ \forall i, \end{array} \end{aligned}$$

$$x_j = l_j \ \forall j\colon c_j + \sum_{i=1}^{m} a_{ij}p_{ij} > 0,$$
$$x_j = u_j \ \forall j\colon c_j + \sum_{i=1}^{m} a_{ij}p_{ij} < 0 \Big\}$$

$$= \Big\{ \zeta \in \mathbb{R}^{mn} \ \Big| \ \begin{array}{l} \zeta_{ij} = -\pi_i + \lambda a_{ij}x_j \ \forall \ (i,j) \in Q, \\ \boldsymbol{\pi} \in \mathbb{R}^m, \mathbf{x} \in \mathbb{R}^n, \mathbf{l} \le \mathbf{x} \le \mathbf{u}, \\ \sum_{j=1}^{n} a_{ij}x_j = b_i \ \ \forall i, \\ x_j = l_j \ \forall j\colon \ \bar{c}_j(\pi) > 0, \\ x_j = u_j \ \forall j\colon \ \bar{c}_j(\pi) < 0 \end{array} \Big\}$$

$$= \Big\{ \zeta \in \mathbb{R}^{mn} \ \Big| \ \begin{array}{l} \zeta_{ij} = -\pi_i + \lambda a_{ij}x_j \ \forall \ (i,j) \in Q, \\ \mathbf{x} \text{ is an optimum of (LP)} \\ \boldsymbol{\pi} \text{ is a vector of optimal simplex multipliers for (LP)} \end{array} \Big\} \ . \ \blacksquare$$

So, given $(\mathbf{t}, \mathbf{w}) \in S$, the question of how far $\mathbf{t}$ is from $S^{-1}(\mathbf{0})$ may be interpreted as follows: given a primal-dual solution pair $(\mathbf{x}, \boldsymbol{\pi})$ that is $d_{\max}\|\mathbf{w}\|/\lambda$-balanced and meets $a_{\max}\|\mathbf{w}\|$-complementary slackness, how far can it be from the set of optimal solution-simplex multiplier pairs $(\mathbf{x}^*, \boldsymbol{\pi}^*)$ of (LP)?

We now appeal to the stability theory of linear programming. Let X^* be the set of optimal solutions to (LP), and Π^* the set of optimal simplex multipliers.

Lemma 7. There exists a finite nonnegative scalar $\mu(\mathbf{A}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{u})$ such that for any optimal solution-simplex multiplier pair $(\mathbf{x}', \boldsymbol{\pi}')$ for the perturbed linear program

$$\begin{array}{ll} \text{minimize} & (\mathbf{c}')^{\mathrm{T}}\mathbf{x} \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b}' \\ & \mathbf{l} \le \mathbf{x} \le \mathbf{u} \quad , \end{array} \qquad \text{(LP')}$$

there exists $(\mathbf{x}^*, \boldsymbol{\pi}^*) \in X^* \times \Pi^*$ such that

$$\max\{\|\mathbf{x}^* - \mathbf{x}'\|_\infty, \|\boldsymbol{\pi}^* - \boldsymbol{\pi}'\|_\infty\}$$
$$\le \ \mu(\mathbf{A}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{u}) \max\{ \|\mathbf{c} - \mathbf{c}'\|_\infty, \|\mathbf{b} - \mathbf{b}'\|_\infty, \|\mathbf{c} - \mathbf{c}'\|_\infty \|\mathbf{x}'\|_1 + \|\mathbf{b} - \mathbf{b}'\|_\infty \|\boldsymbol{\pi}'\|_1\} \ .$$

Proof. We give a proof for the case that (LP) is in the standard primal form where $l_j = 0$ and $x_j = +\infty$ for all j. The proof for other cases is entirely analogous. If (LP) has this form, then $(\mathbf{x}, \boldsymbol{\pi})$ is an optimal pair if and only if it solves the linear system

$$\begin{array}{rcl} \mathbf{Ax} & & = \mathbf{b} \\ -\mathbf{Ix} & & \leq \mathbf{0} \\ & \mathbf{A}^{\mathsf{T}}\boldsymbol{\pi} & \leq \mathbf{c} \\ \mathbf{c}^{\mathsf{T}}\mathbf{x} & -\ \mathbf{b}^{\mathsf{T}}\boldsymbol{\pi} & \leq 0 \end{array} \quad , \qquad (*)$$

whereas we have

$$\begin{array}{rcl} \mathbf{Ax'} & & = \mathbf{b'} \\ -\mathbf{Ix'} & & \leq \mathbf{0} \\ & \mathbf{A}^{\mathsf{T}}\boldsymbol{\pi}' & \leq \mathbf{c'} \\ \mathbf{c'}^{\mathsf{T}}\mathbf{x'} & -\ \mathbf{b'}^{\mathsf{T}}\boldsymbol{\pi}' & \leq 0 \end{array} \quad .$$

We may rewrite this last inequality as

$$\mathbf{c}^{\mathsf{T}}\mathbf{x'} - \mathbf{b}^{\mathsf{T}}\boldsymbol{\pi}' \leq (\mathbf{c} - \mathbf{c'})^{\mathsf{T}}\mathbf{x'} - (\mathbf{b} - \mathbf{b'})^{\mathsf{T}}\boldsymbol{\pi}' \quad ,$$

we conclude that $(\mathbf{x'}, \boldsymbol{\pi}')$ is one possible solution in $(\mathbf{x}, \boldsymbol{\pi})$ of the system

$$\begin{array}{rcl} \mathbf{Ax} & & = \mathbf{b'} \\ -\mathbf{Ix} & & \leq \mathbf{0} \\ & \mathbf{A}^{\mathsf{T}}\boldsymbol{\pi} & \leq \mathbf{c'} \\ \mathbf{c}^{\mathsf{T}}\mathbf{x} & -\ \mathbf{b}^{\mathsf{T}}\boldsymbol{\pi} & \leq (\mathbf{c} - \mathbf{c'})^{\mathsf{T}}\mathbf{x'} - (\mathbf{b} - \mathbf{b'})^{\mathsf{T}}\boldsymbol{\pi}' \end{array} \quad ,$$

which differs from (*) only in the right-hand side. Then from Mangasarian and Shiau (1987), Theorem 2.2, there exists a finite scalar $\mu \geq 0$ such that for all choices of $\mathbf{b'}$ and $\mathbf{c'}$,

$$\begin{aligned} \| (\mathbf{x'}, \boldsymbol{\pi}') - (\mathbf{x}^*. \boldsymbol{\pi}^*) \|_\infty &\leq \mu \left\| \begin{array}{c} \mathbf{b'} - \mathbf{b} \\ \mathbf{c'} - \mathbf{c} \\ (\mathbf{c} - \mathbf{c'})^{\mathsf{T}}\mathbf{x'} + (\mathbf{b} - \mathbf{b'})^{\mathsf{T}}\boldsymbol{\pi}' \end{array} \right\|_\infty \\ &\leq \mu \max \left\{ \| \mathbf{b'} - \mathbf{b} \|_\infty, \| \mathbf{c'} - \mathbf{c} \|_\infty, \left| (\mathbf{c} - \mathbf{c'})^{\mathsf{T}}\mathbf{x'} + (\mathbf{b} - \mathbf{b'})^{\mathsf{T}}\boldsymbol{\pi}' \right| \right\} \\ &\leq \mu \max \left\{ \| \mathbf{b'} - \mathbf{b} \|_\infty, \| \mathbf{c'} - \mathbf{c} \|_\infty, \| \mathbf{c'} - \mathbf{c} \|_\infty \|\mathbf{x'}\|_1 + \| \mathbf{b'} - \mathbf{b} \|_\infty \|\boldsymbol{\pi}'\|_1 \right\} \quad . \end{aligned}$$

■

Let $\mathbf{t}^k = \mathbf{p}^k + \lambda \mathbf{z}^k$ for all $k \geq 0$. Then, from Eckstein and Bertsekas (1989),

$$\begin{aligned} \mathbf{t}^{k+1} &= (1 - \rho_k)\mathbf{t}^k + \rho_k(I + S)^{-1}(\mathbf{t}^k) \\ &= [I - \rho_k(I - (I + S)^{-1})](\mathbf{t}^k) \\ &= (I - \rho_k K)(\mathbf{t}^k) \quad , \end{aligned}$$

where K is the operator $I - (I + S)^{-1}$.

Lemma 8. Let $\phi = \inf\{\rho_k(2 - \rho_k) \mid k \geq 0\} \in (0, 1]$. Then the following inequalities hold for all $k \geq 0$:

(i) $\quad \| \mathbf{t}^{k+1} - \mathbf{t}^* \|^2 \leq \| \mathbf{t}^k - \mathbf{t}^* \|^2 - \phi \| K(\mathbf{t}^k) \|^2$ for all $\mathbf{t}^* \in S^{-1}(\mathbf{0})$

(ii) $\quad \mathrm{dist}^2(\mathbf{t}^{k+1}, S^{-1}(\mathbf{0})) \leq \mathrm{dist}^2(\mathbf{t}^k, S^{-1}(\mathbf{0})) - \phi \| K(\mathbf{t}^k) \|^2$.

Proof. From the proof of Theorem 3 of Eckstein and Bertsekas (1989), we have for all $k \geq 0$ and $\mathbf{t}^* \in S^{-1}(\mathbf{0})$ that

$$\| \mathbf{t}^{k+1} - \mathbf{t}^* \|^2 \leq \| \mathbf{t}^k - \mathbf{t}^* \|^2 - \rho_k(2 - \rho_k) \| K(\mathbf{t}^k) \|^2 \quad ,$$

which implies (i). One then obtains (ii) by taking $\mathbf{t}^*$ in (i) to be the projection of $\mathbf{t}^k$ onto $S^{-1}(\mathbf{0})$ (which is closed and convex by the maximality of S). ■

We know from Theorem 3 that if (LP) is feasible and bounded, that $\{\mathbf{x}^k\}$ and $\{\boldsymbol{\pi}^k\}$ are bounded sequences. Define

$$D_{\mathrm{x}} = \sup_{k \geq 0} \left\{ \|\mathbf{x}^k\|_1 \right\} < \infty$$

$$D_{\pi} = \sup_{k \geq 0} \left\{ \|\boldsymbol{\pi}^k\|_1 \right\} < \infty$$

$$R_{\mathrm{x}} = \max\{1, D_{\mathrm{x}}\}$$

$$R_{\pi} = \max\{1, D_{\pi}\}$$

$$\gamma = \mu(\mathbf{A}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{u})\left[a_{\max} R_x + \frac{d_{\max}}{\lambda} R_\pi\right]$$

$$\alpha = \left(\sqrt{mn} + \lambda\left[\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2\right]^{1/2}\right)\gamma$$

$$\rho^* = \inf_{k \geq 0}\{\rho_k\} > 0$$

$$\tau = 1 - \rho^*\left(1 - \frac{\alpha}{\sqrt{\alpha^2 + 1}}\right) < 1$$

$$\delta_0 = \mathrm{dist}(\mathbf{t}^k, S^{-1}(\mathbf{0})) \quad .$$

Theorem 4. Suppose one applies the alternating step method to a feasible linear program (LP) with finite optimal objective value. If $\rho_k \leq 1$ for all $k \geq 0$, the following inequalities hold:

(i) $\mathrm{dist}(\boldsymbol{\pi}^k, \Pi^*) \leq \tau^k \delta_0 \qquad \forall\, k \geq 0$,

(ii) $\mathrm{dist}_\infty(\mathbf{x}^k, X^*) \leq \tau^{k-1}\left(\frac{\gamma\delta_0}{\sqrt{\phi}}\right) \qquad \forall\, k \geq 1$,

(iii) $\mathrm{dist}(\mathbf{t}^{k+1}, \mathrm{zer}(S_{\lambda,A,B})) \leq \tau\, \mathrm{dist}(\mathbf{t}^k, S^{-1}(\mathbf{0})) \qquad \forall\, k \geq 0$,

(iv) $\mathrm{dist}(\mathbf{t}^k, S^{-1}(\mathbf{0})) \leq \tau^k \delta_0 \qquad \forall\, k \geq 0$,

where $\mathrm{dist}_\infty(\mathbf{x}^k, X^*)$ denotes $\min\{\|\mathbf{x}^k - \mathbf{x}\|_\infty \mid \mathbf{x} \in X^*\}$.

Proof. Note that from the construction of the generalized alternating direction method of multipliers (Eckstein and Bertsekas 1989, Theorem 8), we have $(\mathbf{q}^{k+1}, -\mathbf{M}\mathbf{x}^{k+1}) \in A$ for all $k \geq 0$, where

$$\mathbf{q}^{k+1} = \mathbf{p}^k + \lambda(\mathbf{M}\mathbf{x}^{k+1} - \mathbf{z}^k) \quad .$$

We now start the proof by establishing (iii), first in the case $\rho_k = 1$, and then in the case $\rho_k < 1$. Suppose $\rho_k = 1$. Then $\mathbf{t}^{k+1} = (I + S)^{-1}(\mathbf{t}^k)$ and $\mathbf{t}^k - \mathbf{t}^{k+1} \in S(\mathbf{t}^{k+1})$. Because $\rho_k = 1$ and we are doing all minimizations exactly, $\{\mathbf{x}^k\}$, $\{\mathbf{p}^k\}$, and $\{\mathbf{z}^k\}$ evolve according to the conventional alternating direction method of multipliers

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} \left\{ f(\mathbf{x}) + \langle \mathbf{p}^k, \mathbf{M}\mathbf{x} \rangle + \frac{\lambda}{2} \| \mathbf{M}\mathbf{x} - \mathbf{z}^k \|^2 \right\}$$

$$\mathbf{z}^{k+1} = \arg\min_{\mathbf{z}} \left\{ g(\mathbf{z}) - \langle \mathbf{p}^k, \mathbf{z} \rangle + \frac{\lambda}{2} \| \mathbf{M}\mathbf{x}^{k+1} - \mathbf{z} \|^2 \right\}$$

$$\mathbf{p}^{k+1} = \mathbf{p}^k + \lambda(\mathbf{M}\mathbf{x}^{k+1} - \mathbf{z}^{k+1}) \quad .$$

Let $\mathbf{t} = \mathbf{t}^{k+1}$, and

$$\begin{aligned}
\mathbf{w} &= \mathbf{t}^k - \mathbf{t}^{k+1} \\
&= (\mathbf{p}^k + \lambda \mathbf{z}^k) - (\mathbf{p}^{k+1} + \lambda \mathbf{z}^{k+1}) \\
&= (\mathbf{p}^k - \mathbf{p}^{k+1}) + \lambda (\mathbf{z}^k - \mathbf{z}^{k+1}) \\
&= \lambda (\mathbf{z}^{k+1} - \mathbf{M}\mathbf{x}^{k+1}) + \lambda (\mathbf{z}^k - \mathbf{z}^{k+1}) \\
&= \lambda (\mathbf{z}^k - \mathbf{M}\mathbf{x}^{k+1}) \\
&= \mathbf{p}^k - \mathbf{q}^{k+1} \quad .
\end{aligned}$$

Then we have $(\mathbf{t}, \mathbf{w}) \in S$. Let

$$\begin{aligned}
\mathbf{p} &= \mathbf{p}^k & \qquad \mathbf{s} &= -\mathbf{M}\mathbf{x}^{k+1} \\
\mathbf{z} &= \mathbf{z}^k & \qquad \mathbf{x} &= \mathbf{x}^{k+1} \\
\mathbf{q} &= \mathbf{q}^{k+1} & \qquad \boldsymbol{\pi} &= \boldsymbol{\pi}^k \quad .
\end{aligned}$$

Then we have $(\mathbf{p}, \mathbf{z}) \in B$, $(\mathbf{q}, \mathbf{s}) \in A$,

$$\mathbf{q} + \lambda \mathbf{s} = \mathbf{p}^k + \lambda (\mathbf{M}\mathbf{x}^{k+1} - \mathbf{z}^k) - \lambda \mathbf{M}\mathbf{x}^{k+1} = \mathbf{p}^k - \lambda \mathbf{z}^k = \mathbf{p} - \lambda \mathbf{z} \; ,$$

and

$$\mathbf{q} + \lambda \mathbf{z} = \mathbf{p}^k + \lambda (\mathbf{M}\mathbf{x}^{k+1} - \mathbf{z}^k) + \lambda \mathbf{z}^k = \mathbf{p}^k + \lambda \mathbf{M}\mathbf{x}^{k+1} = \mathbf{t}^{k+1} = \mathbf{t} \quad .$$

So, $\mathbf{p}$, $\mathbf{q}$, $\mathbf{s}$, $\mathbf{z}$, $\mathbf{x}$ and $\boldsymbol{\pi}$ meet the conditions of Lemma 6.1. Therefore, we deduce that

$\mathbf{x} = \mathbf{x}^{k+1}$ is $d_{\max} \| \mathbf{t}^{k-1} - \mathbf{t}^k \| / \lambda$-balanced

$(\mathbf{x}, \boldsymbol{\pi}) = (\mathbf{x}^{k+1}, \boldsymbol{\pi}^k)$ obeys $a_{\max} \| \mathbf{t}^{k-1} - \mathbf{t}^k \|$-complementary slackness.

By Lemma 7, there exists $(\mathbf{x}^*, \boldsymbol{\pi}^*) \in X^* \times \Pi^*$ such that

$$\max\{\|\mathbf{x} - \mathbf{x}'\|_\infty, \|\boldsymbol{\pi} - \boldsymbol{\pi}'\|_\infty\}$$

$$\begin{aligned}\le \mu(\mathbf{A}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{u}) \max\{ & a_{\max}\|\mathbf{t}^{k-1} - \mathbf{t}^k\|, \\ & d_{\max}\|\mathbf{t}^{k-1} - \mathbf{t}^k\|/\lambda, \\ & a_{\max}\|\mathbf{t}^{k-1} - \mathbf{t}^k\| \, \|\mathbf{x}'\|_1 + d_{\max}\|\mathbf{t}^{k-1} - \mathbf{t}^k\| \, \|\boldsymbol{\pi}'\|_1/\lambda\}\end{aligned}$$

$$\le \mu(\mathbf{A}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{u})[a_{\max}R_x + \frac{d_{\max}}{\lambda}R_\pi] \, \|\mathbf{t}^{k-1} - \mathbf{t}^k\|$$

$$= \gamma \, \|\mathbf{t}^{k-1} - \mathbf{t}^k\| \quad .$$

Defining $\mathbf{p}^*$ by

$$p_{ij}^* = \begin{cases} -\pi_i^*, & (i,j) \in Q \\ 0, & (i,j) \notin Q \end{cases} ,$$

we have that $\mathbf{t}^* = \mathbf{p}^* + \lambda\mathbf{M}\mathbf{x}^* \in S^{-1}(\mathbf{0})$. Also,

$$\begin{aligned}\|\mathbf{t}^{k+1} - \mathbf{t}^*\| &\le \|\mathbf{p}^k - \mathbf{p}^*\| + \lambda\|\mathbf{M}\mathbf{x}^{k+1} - \mathbf{M}\mathbf{x}^*\| \\ &\le \|\mathbf{p}^k - \mathbf{p}^*\| + \lambda\|\mathbf{M}(\mathbf{x}^{k+1} - \mathbf{x}^*)\| \\ &\le \sqrt{mn}\,\|\mathbf{p}^k - \mathbf{p}^*\|_\infty + \lambda\left[\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2\right]^{1/2} \|\mathbf{x}^{k+1} - \mathbf{x}^*\|_\infty \\ &= \sqrt{mn}\,\|\boldsymbol{\pi}^k - \boldsymbol{\pi}^*\|_\infty + \lambda\left[\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2\right]^{1/2} \|\mathbf{x}^{k+1} - \mathbf{x}^*\|_\infty \\ &\le \sqrt{mn}\gamma\,\|\mathbf{t}^k - \mathbf{t}^{k+1}\| + \lambda\left[\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2\right]^{1/2} \gamma\,\|\mathbf{t}^k - \mathbf{t}^{k+1}\| \\ &= \alpha\,\|\mathbf{t}^k - \mathbf{t}^{k+1}\| \quad .\end{aligned}$$

Thus,

$$\operatorname{dist}(\mathbf{t}^{k+1}, S^{-1}(\mathbf{0})) \le \alpha\,\|\mathbf{t}^k - \mathbf{t}^{k+1}\| \quad .$$

When $\rho_k = 1$, the inequality

$$\| \mathbf{t}^{k+1} - \mathbf{t}^* \|^2 \leq \| \mathbf{t}^k - \mathbf{t}^* \|^2 - \rho_k(2 - \rho_k) \| Q(\mathbf{t}^k) \|^2 \quad \forall\, \mathbf{t}^* \in S^{-1}(\mathbf{0})$$

reduces to

$$\| \mathbf{t}^{k+1} - \mathbf{t}^* \|^2 \leq \| \mathbf{t}^k - \mathbf{t}^* \|^2 - \|\mathbf{t}^k - \mathbf{t}^{k+1}\|^2 \qquad \forall\, \mathbf{t}^* \in S^{-1}(\mathbf{0}),$$

implying

$$\mathrm{dist}^2(\mathbf{t}^{k+1}, S^{-1}(\mathbf{0})) \leq \mathrm{dist}^2(\mathbf{t}^k, S^{-1}(\mathbf{0})) - \|\mathbf{t}^k - \mathbf{t}^{k+1}\|^2 \ .$$

Hence,

$$\mathrm{dist}^2(\mathbf{t}^{k+1}, S^{-1}(\mathbf{0})) \leq \mathrm{dist}^2(\mathbf{t}^k, S^{-1}(\mathbf{0})) - \frac{1}{\alpha^2}\,\mathrm{dist}^2(\mathbf{t}^{k+1}, S^{-1}(\mathbf{0})) \ .$$

Simplifying and taking the square root, we obtain

$$\mathrm{dist}(\mathbf{t}^{k+1}, S^{-1}(\mathbf{0})) \leq \left(\frac{\alpha}{\sqrt{\alpha^2 + 1}}\right)\mathrm{dist}(\mathbf{t}^k, S^{-1}(\mathbf{0})) \leq \tau\,\mathrm{dist}(\mathbf{t}^k, S^{-1}(\mathbf{0})) \quad .$$

This gives (iii) in the case that $\rho_k = 1$. Now consider the case $\rho_k < 1$. Let

$$\mathbf{u}^{k+1} = (I + S)^{-1}(\mathbf{t}^k)$$

be the value that $\mathbf{t}^{k+1}$ would have taken had ρ_k been 1. Then we have

$$\mathbf{t}^{k+1} = (1 - \rho_k)\mathbf{t}^k + \rho_k\mathbf{u}^{k+1} \ ,$$

and, by the convexity of the function $\mathrm{dist}(\,\cdot\,, S^{-1}(\mathbf{0}))$, we have

$$\begin{aligned}
\mathrm{dist}(\mathbf{t}^{k+1}, S^{-1}(\mathbf{0})) &\leq (1 - \rho_k)\mathrm{dist}(\mathbf{t}^k, S^{-1}(\mathbf{0})) + \rho_k\mathrm{dist}(\mathbf{u}^{k+1}, S^{-1}(\mathbf{0})) \\
&\leq \left(1 - \rho_k\left(1 - \frac{\alpha}{\sqrt{\alpha^2 + 1}}\right)\right) \mathrm{dist}(\mathbf{t}^k, S^{-1}(\mathbf{0})) \\
&\leq \tau\,\mathrm{dist}(\mathbf{t}^k, S^{-1}(\mathbf{0})) \quad ,
\end{aligned}$$

which gives (iii). By induction, we obtain (iv).

We now demonstrate (i) and (ii). For all $k \geq 0$, there exists some $\mathbf{t}^{*k} \in S^{-1}(\mathbf{0})$ such that $\|\mathbf{t}^k - \mathbf{t}^{*k}\| \leq \tau^k \delta_0$. Applying the nonexpansive operator $(I + \lambda B)^{-1}$ to $\mathbf{t}^k$ and $\mathbf{t}^{*k}$, one obtains $\|\mathbf{p}^k - \mathbf{p}^{*k}\| \leq \tau^k \delta_0$, where $\mathbf{p}^{*k}$ and $\mathbf{z}^{*k}$ are such that

$$\mathbf{t}^{*k} = \mathbf{p}^{*k} + \lambda \mathbf{z}^{*k}$$

$$(\mathbf{p}^{*k}, \mathbf{z}^{*k}) \in B$$

$$(\mathbf{p}^{*k}, -\mathbf{z}^{*k}) \in A \ .$$

It follows that $\|\boldsymbol{\pi}^k - \boldsymbol{\pi}^{*k}\| \leq \tau^k \delta_0$, where $\boldsymbol{\pi}^{*k} \in \Pi^*$ is defined by $\pi_i^{*k} = -p_{ij}^{*k}$ for all $(i, j) \in Q$. This fact implies (i). Finally, for $k \geq 0$, we have from above that there exists $\mathbf{x}^{*(k+1)}$ such that

$$\| \mathbf{x}^{k+1} - \mathbf{x}^{*(k+1)} \|_\infty \ \leq \ \gamma \| \mathbf{t}^k - \mathbf{u}^{k+1} \| \ = \ \gamma \| K(\mathbf{t}^k) \| \ .$$

By Lemma 8,

$$\| K(\mathbf{t}^k) \|^2 \ \leq \ \frac{1}{\phi} \operatorname{dist}^2(\mathbf{t}^k, S^{-1}(\mathbf{0})) \ ,$$

hence

$$\| \mathbf{x}^{k+1} - \mathbf{x}^{*(k+1)} \|_\infty \ \leq \ \left(\frac{\gamma \delta_0}{\sqrt{\phi}} \right) ,$$

which gives (ii). ■

A few comments: it is possible to place explicit bounds on the quantities $\mu(\mathbf{A}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{u})$, D_x, and D_π, and hence on the other constants used in the proof. This analysis was omitted in the interest of brevity, but can be found, in the case that all entries of **A** are integer, in Eckstein (1989). The question of linear convergence in the case that $\rho_k > 1$ for an indefinite number of k remains open. The general convergence methodology of Luque (1984) does not seem to extend easily to the proximal point algorithm with over-relaxation ($\rho_k > 1$), although it does apply to under-relaxation ($\rho_k \leq 1$).

6. *Preliminary Computational Results*

We have implemented the alternating step method for linear programming on a variety of serial and parallel computer systems. The details of the various parallel implementations are reserved for a possible future paper. This section briefly discusses the general practical convergence properties of the method. For more detailed results, see Eckstein (1989).

In preliminary experiments, the alternating step method seems to converge reasonably quickly on linear programs with an "assignment" constraint structure, that is, on transportation problems with all supplies and demands having magnitude 1. The algorithm works best when one takes $q_i = d(i)$ for all constraints i. A good choice for the parameter λ, after extensive experimentation, appeared to be $(0.1)\|\mathbf{c}\|_\infty$. A useful starting procedure was to initially use a small value of λ in the calculation of $\mathbf{x}^{k+1}$, but a larger value in the computation of $\boldsymbol{\pi}^{k+1}$. The method may fail to converge if different values of λ are used in the two steps, so we gradually brought the two values together, until, after a fixed number of iterations, both equalled $(0.1)\|\mathbf{c}\|_\infty$. With this heuristic, it appeared best to choose $\rho_k = 1$ for all k. Results without this starting heuristic were generally not as good, but could be improved slightly by choosing ρ_k's in the range [1.1, 1.5].

Table 1 lists the number of iterations until convergence for a set of 22 NETGEN (Klingman *et. al.* 1974) problems with assignment structure. We tested the trial solution pair $(\mathbf{x}^k, \boldsymbol{\pi}^k)$ every ten iterations, and halted the algorithm when $\mathbf{x}^k$ was 0.001-balanced and $(\mathbf{x}^k, \boldsymbol{\pi}^k)$ obeyed 0.001-complementary slackness. In fact, for all but problems 4, 5, 6, and 17, the algorithm seemed to "jump" to a solution pair that was exact to within machine precision, and thus terminate finitely. The conditions under which the method is guaranteed to exhibit such finite termination are unclear at present, although Lefebvre and Michelot (1988) have studied a related problem. This is in contrast to applications of the conventional (as opposed to alternating direction) method of multipliers to linear programming, which are known to always converge finitely (Bertsekas 1975, Rockafellar 1976b).

Problem	Sources/ Sinks	Arcs	$\| c \|_\infty$	Iterations	CPU Seconds
1	4	16	982	60	0.02
2	32	100	100	60	0.03
3	32	400	100	110	0.13
4	32	600	100	1020	1.47
5	32	800	100	770	1.40
6	32	900	100	350	0.70
7	32	1024	100	60	0.15
8	32	1024	100	50	0.12
9	32	1024	100	60	0.13
10	1024	5120	100	440	7.13
11	1024	5120	100	770	12.57
12	1024	5120	100	930	15.52
13	1024	5120	100	800	13.23
14	200	1000	100	180	0.58
15	400	2000	100	390	2.48
16	600	3000	100	430	4.07
17	100	500	100	490	0.83
18	105	525	200	140	0.25
19	110	550	400	160	0.32
20	115	575	800	140	0.27
21	129	600	1600	140	0.28
22	125	625	3200	150	0.32

Table 1. Detailed results for a FORTRAN implementation of the alternating step method on linear programs with "assigment" constraint structure. CPU times are for an Alliant FX/8, a "supermini" computer allowing a modest degree of parallelism.

The results of Table 1 are fairly encouraging. With large speedups due to its inherent parallelism, the alternating step method should prove quite competitive for problems of this type.

For more general linear programs, including transportation problems, our early computational experiments have been much less encouraging; the method seems prone to a curious "spiral" convergene behavious, which we now illustrate. For all $k \geq 0$, let

$$i_{\max}(k) = \arg\max_{i=1,\ldots,m} \left\{ \left| r_i(\mathbf{x}^k) \right| \right\}$$

$$r_{\max}(k) = r_{[i_{\max}(k)]}(\mathbf{x}^k)$$

$$s_j(k) = \begin{cases} \bar{c}_j(\boldsymbol{\pi}^k) , & l_j < x_j^k < u_j \\ \max\{\bar{c}_j(\boldsymbol{\pi}^k), 0\} , & x_j^k = u_j \\ \min\{\bar{c}_j(\boldsymbol{\pi}^k), 0\} , & x_j^k = l_j \end{cases} \qquad j = 1,\ldots,n$$

$$j_{\max}(k) = \arg\max_{j=1,\ldots,n} \left\{ \left| s_j(k) \right| \right\}$$

$$\sigma(k) = s_{[j_{\max}(k)]}(k) \quad ,$$

so that $| r_{\max}(k) |$ is an estimate of the feasibility of $\mathbf{x}^k$ and $| \sigma(k) |$ is an estimate of how closely ($\mathbf{x}^k$, $\boldsymbol{\pi}^k$)complementary slackness.

For a small transportation problem, Figure 3 shows a typical plot of $r_{\max}(k)$ and $\sigma(k)$ for large k; the graphs forms a very slowly-converging spiral around (0, 0). At present, we are not sure how to accelerate this convergence behaviour.

So far, we have not tested the alternating step method on nonlinear monotropic programs. It will be interesting, for instance, to compare it to dual relaxation methods for nonlinear network problems (Bertsekas and El Baz 1987, Bertsekas, Hossein, and Tseng 1987, Zenios and Lasken 1988). Unlike such dual relaxation methods, the alternating step method can solve problems with *mixtures* of linear cost and strictly convex cost arcs, and can handle constraint structures more general than pure networks.

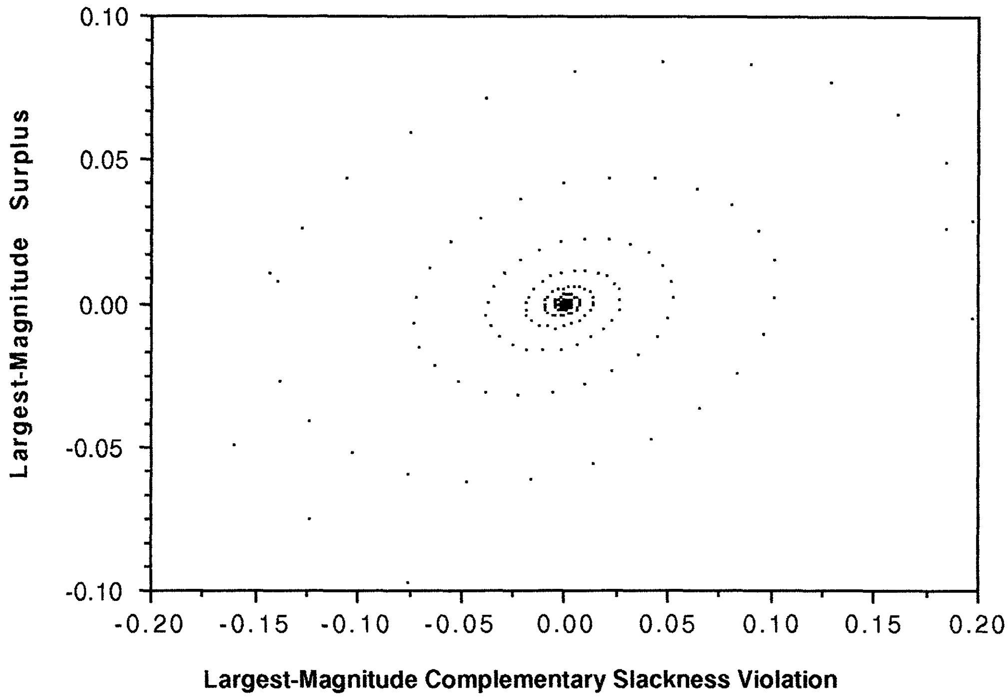

Figure 3: Data from a run of the alternating step method on a small transportation problem, depicting the largest-magnitude surplus $r_{\max}(k)$ versus the largest-magnitude complementary slackness violation $\sigma(k)$ for large k.

References

BERTSEKAS, D. P. (1975), "Necessary and Sufficient Conditions for a Penalty Method to be Exact". *Mathematical Programming* **9**:87-99.

BERTSEKAS, D. P. (1982), *Constrained Optimization and Lagrange Multiplier Methods* (New York: Academic Press).

BERTSEKAS, D. P. (1985), *A Distributed Asynchronous Relaxation Algorithm for the Assignment Problem.* Proceedings of the 24th IEEE Conference on Decision and Control, Fort Lauderdale, Florida, 1703-1704.

BERTSEKAS, D. P. (1986), *Distributed Asynchronous Relaxation Methods for Linear Network Flow Problems.* Report LIDS-P-1606, Laboratory for Information and Decision Sciences, MIT. (Original Report, September 1986, revised version, November 1986.)

BERTSEKAS, D. P., ECKSTEIN, J. (1988), "Dual Coordinate Step Methods for Linear Network Flow Problems". *Mathematical Programming* **42**(2):203-243.

BERTSEKAS, D. P., EL BAZ, D. (1987), "Distributed Asynchronous Methods for Convex Network Flow Problems". *SIAM Journal on Control and Optimization* **25**(1):74-85.

BERTSEKAS, D. P., HOSSEIN, P., TSENG, P. (1987), "Relaxation methods for network flow problems with convex arc costs". *SIAM Journal on Control and Optimization* **25**:1219-1243.

BERTSEKAS, D. P., TSITSIKLIS, J. (1989), *Parallel and Distributed Computation: Numerical Methods* (Englewood Cliffs: Prentice-Hall).

DE LEONE, R., MANGASARIAN, O. L. (1988), "Serial and Parallel Solution of Large Scale Linear Programs by Augmented Lagrangian Successive Overrelaxation". In *Optimization, Parallel Processing, and Applications*, A. KURZHANSKI, K. NEUMANN, and D. PALLASCHKE, editors, *Lecture Notes in Economics and Mathamatical Systems* **304** (Berlin: Springer-Verlag).

ECKSTEIN, J. (1988), *The Lions-Mercier Algorithm and the Alternating Direction Method are Instances of the Proximal Point Algorithm.* Report LIDS-P-1769, Laboratory for Information and Decision Sciences, MIT.

ECKSTEIN, J. (1989), *Splitting Methods for Monotone Operators with Applications to Parallel Optimization.* Doctoral dissertation, department of civil engineering, Massachusetts Institute of Technology. Available as report LIDS-TH-1877, Laboratory for Information and Decision Sciences, MIT, or report CICS-TH-140, Brown/Harvard/MIT Center for Intelligent Control.

ECKSTEIN, J., BERTSEKAS, D. P. (1989). *On the Douglas-Rachford Splitting Method and the Proximal Point Algorithm for Maximal Monotone Operators.* Report CICS-P-167, Brown/Harvard/MIT Center for Intelligent Control Systems, or Harvard Business School working paper 90-033 (Revision conditionally accepted to *Mathematical Programming*).

FORTIN, M., GLOWINSKI, R. (1983), "On Decomposition-Coordination Methods Using an Augmented Lagrangian". In M. FORTIN, R. GLOWINSKI, editors, *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems* (Amsterdam: North-Holland).

GABAY, D. (1983), "Applications of the Method of Multipliers to Variational Inequalities". In M. FORTIN, R. GLOWINSKI, editors, *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems* (Amsterdam: North-Holland).

GABAY, D., MERCIER B. (1976), "A Dual Algorithm for the Solution of Nonlinear Variational Problems via Finite Element Approximations". *Computers and Mathematics with Applications* **2**:17-40.

GLOWINSKI, R., LE TALLEC, P. (1987), *Augmented Lagrangian Methods for the Solution of Variational Problems*. MRC Technical Summary Report #2965, Mathematics Research Center, University of Wisconsin – Madison.

GLOWINSKI, R., LIONS, J.-L., TRÉMOLIERES, R. (1981), *Numerical Analysis of Variational Inequalities* (Amsterdam: North-Holland).

GLOWINSKI, R., MARROCO, A. (1975), "Sur L'Approximation, par Elements Finis d'Ordre Un, et la Resolution, par Penalisation-Dualité, d'une Classe de Problemes de Dirichlet non Lineares". *Revue Française d'Automatique, Informatique et Recherche Opérationelle* **9**(R-2):41-76 (in French).

GOLDBERG, A. V. (1987), *Efficient Graph Algorithms for Sequential and Parallel Computers*. Technical Report TR-374, Laboratory for Computer Science, MIT.

GOLDBERG, A. V., TARJAN, R. E. (1987), *Solving Minimum Cost Flow Problems by Successive Approximation*, Proceedings of the 19th ACM Symposium on the Theory of Computation.

GOL'SHTEIN, E. G. (1985), "Decomposition Methods for Linear and Convex Programming Problems". *Matekon* **22**(4):75-101.

GOL'SHTEIN, E. G. (1986), "The Block Method of Convex Programming". *Soviet Mathematics Doklady* **33**(3):584-587.

KLINGMAN, D., NAPIER, A., STUTZ, J., (1974), "NETGEN — A Program for Generating Large-Scale (Un)capacitated Assignment, Transportation, and Minimum Cost Network Problems". *Management Science* **20**:814-822.

LEFEBVRE O., MICHELOT, C. (1988), "About the Finite Convergence of the Proximal Point Algorithm". *Trends in Mathematical Optimization: 4^{th} French-German Conference on Optimization, International Series of Numerical Mathematics* **84**, K.-H. HOFFMANN ET. AL., editors (Basel: Birkhäuser-Verlag).

LIONS, P.-L., MERCIER, B. (1979), "Splitting Algorithms for the Sum of Two Nonlinear Operators". *SIAM Journal on Numerical Analysis* **16**(6):964-979.

LUQUE, F. J. (1984), "Asymptotic Convergence Analysis of the Proximal Point Algorithm". *SIAM Journal on Control and Optimization* **22**(2):277-293.

MANGASARIAN, O. L., SHIAU T.-H. (1987), "Lipschitz Continuity of Solutions of Linear Inequalities, Programs and Complementarity Problems". *SIAM Journal on Control and Optimization* **25**(3):583-595.

ROCKAFELLAR, R. T. (1970), *Convex Analysis* (Princeton: Princeton University Press).

ROCKAFELLAR, R. T. (1976A), "Monotone Operators and the Proximal Point Algorithm". *SIAM Journal on Control and Optimization* **14**(5):877-898.

ROCKAFELLAR, R. T. (1976B), "Augmented Lagrangians and Applications of the Proximal Point Algorithm in Convex Programming". *Mathematics of Operations Research* **1**(2):97-116.

ROCKAFELLAR, R. T. (1984), *Network Flows and Monotropic Optimization* (New York: John Wiley).

SPINGARN, J. E. (1985A), "A Primal-Dual Projection Method for Solving Systems of Linear Inequalities". *Linear Algebra and Its Applications* **65**:45-62.

SPINGARN, J. E. (1985B), "Application of the Method of Partial Inverses to Convex Programming: Decomposition". *Mathematical Programming* **32**:199-233.

SPINGARN, J. E. (1987), "A Projection Method for Least-Squares Solutions to Overdetermined Systems of Linear Inequalities". *Linear Algebra and Its Applications* **86**:211-236.

TSENG, P., BERTSEKAS, D. P. (1987), "Relaxation Methods for Linear Programs". *Mathematics of Operations Research* **12**(4):569-596.

ZENIOS, S. A., LASKEN, R. A. (1988), "Nonlinear Network Optimization on a Massively Parallel Connection Machince". *Annals of Operations Research* **14**:147-165