Mathematical Programming Study 17 (1982) 139–159.
North-Holland Publlishing Company

PROJECTION METHODS FOR VARIATIONAL INEQUALITIES WITH APPLICATION TO THE TRAFFIC ASSIGNMENT PROBLEM*

Dimitri P. BERTSEKAS and Eli M. GAFNI

Laboratory for Information and Decision Systems, Massachussetts Institute of Technology, Cambridge, MA 02139, U.S.A.

Received 1 August 1980

It is well known [2, 3, 16] that if $\bar{T}:\mathbf{R}^n \to \mathbf{R}^n$ is a Lipschitz continuous, strongly monotone operator and X is a closed convex set, then a solution $x^* \in X$ of the variational inequality $(x-x^*)'\bar{T}(x^*) \geq 0$, $\forall x \in X$ can be found iteratively by means of the projection method $x_{k+1} = P_X[x_k - \alpha\bar{T}(x_k)]$, $x_0 \in X$, provided the stepsize α is sufficiently small. We show that the same is true if $\bar{T}$ is of the form $\bar{T} = A'TA$, where $A:\mathbf{R}^n \to \mathbf{R}^m$ is a linear mapping, provided $T:\mathbf{R}^m \to \mathbf{R}^m$ is Lipschitz continuous and strongly monotone, and the set X is polyhedral. This fact is used to construct an effective algorithm for finding a network flow which satisfies given demand constraints and is positive only on paths of minimum marginal delay or travel time.

Key words: Projection Methods, Variational Inequalities, Traffic Assignment, Network Routing, Multicommodity Network Flows.

1. Introduction

We consider the problem of finding $x^* \in X$ satisfying the variational inequality

$$(x-x^*)'A'T(Ax^*) \geq 0, \quad \forall x \in X \tag{1}$$

where X is a nonempty subset of $\mathbf{R}^n$, A is a given $m \times n$ matrix and $T:\mathbf{R}^m \to \mathbf{R}^m$ is a nonlinear operator which is Lipschitz continuous and strongly monotone in the sense that there exist positive scalars L and λ such that for all y_1, y_2 in the set $Y = \{y \mid y = Ax, x \in X\}$ we have

$$|T(y_1) - T(y_2)| \leq L|y_1 - y_2|, \tag{2}$$

$$[T(y_1) - T(y_2)]'(y_1 - y_2) \geq \lambda|y_1 - y_2|^2. \tag{3}$$

In the relations above and throughout the paper all vectors are considered to be column vectors, and a prime denotes transposition. The standard norm in $\mathbf{R}^n$ is denoted $|\cdot|$, i.e., $|x| = (x'x)^{1/2}$ for all $x \in R^n$.

We are interested in the projection algorithm

$$x_{k+1} = P_X^S[x_k - \alpha S^{-1}A'T(Ax_k)], \quad x_0 \in X \tag{4}$$

* Work supported by Grants ONR–N00014-75-C-1183 and NSF ENG-7906332.

where $\alpha > 0$ is a stepsize parameter, S is a fixed positive definite symmetric matrix, and, for any $z \in \mathbf{R}^n$, $P_X^S(z)$ denotes the unique projection of z on the set X with respect to the norm $\|\cdot\|$ corresponding to S

$$\|w\| = (w'Sw)^{1/2}, \quad \forall\, w \in R^n. \tag{5}$$

The variational inequality (1) arises from the variational inequality

$$(y - y^*)'T(y^*) \geq 0, \quad \forall\, y \in Y \tag{6}$$

through the transformation

$$y = Ax, \qquad Y = AX = \{y \mid y = Ax, x \in X\}. \tag{7}$$

It is possible to employ the projection algorithm

$$y_{k+1} = P_Y^Q[y_k - \alpha Q^{-1}T(y_k)], \quad y_0 \in Y \tag{8}$$

for solving (6) where α is a positive stepsize parameter and Q is positive definite symmetric. It has been shown by Sibony [16] that if T is Lipshitz continuous and strongly monotone, and α is chosen sufficiently small, then the sequence $\{y_k\}$ generated by (8) converges to the unique solution y^* of (6). The rate of convergence is typically linear although a superlinear convergence rate is possible in exceptional cases [9]. Strong monotonicity of the mapping T is an essential assumption for these results to hold. Our motivation for considering iteration (4) stems from the fact that in some cases projection on the set Y is very difficult computationally while projection on the set X through the transformation $y = Ax$ [cf. (7)] may be very easy. Under these circumstances if all other factors are equal, the projection method (4) is much more efficient than the method (8). This situation occurs for example in the application discussed in Section 3.

A potential difficulty with the transformation idea described above is that the mapping $A'TA$ is not strongly monotone unless the matrix $A'A$ is invertible. Thus convergence of iteration (4) is not guaranteed by the existing theory [2, 3, 16]. One of the contributions of this paper is to show that the convergence and rate of convergence properties of iteration (4) are satisfactory and comparable with those of iteration (8). These results hinge on the assumption that X is polyhedral, and it is unclear whether and in what form they hold if X is a general convex set.

In Section 3 we consider a classical traffic equilibrium problem arising in several contexts including communication and transportation networks, which can be modelled in terms of a variational inequality of the form (1). A projection algorithm for solving this problem which is essentially of the form (8) has been given by Dafermos [8]. Her algorithm however operates in the space of link flows, and involves a projection iteration which is very costly for large networks. We consider an alternative algorithm which is basically of the form (4) and operates in the space of path flows. Because the projection iteration can be carried out easily in this space our algorithm is much more efficient. An algorithm which has several similarities with ours has been proposed by Aashtiani [1] and has performed well in computational experiments. However, Aashtiani's algorithm cannot be shown to converge in general. By contrast the results of Section 2 guarantee convergence and linear rate of convergence for our method. There are also other methods [4–7, 11–13] for solving the special case of the traffic assignment problem where T is a gradient mapping and there is an underlying convex programming problem. Some of these methods [4–6, 12, 13] are of the projection type. The algorithm of the present paper, however, seems to be the first that can solve the general problem, is suitable for large networks, and is demonstrably convergent.

In Section 4 we consider the generalized version of iteration (4),

$$x_{k+1} = P_X^{S_k}[x_k - \alpha S_k^{-1} A' T(Ax_k)], \quad x_0 \in X$$

where $\{S_k\}$ is a sequence of positive definite symmetric matrices with eigenvalues bounded above and bounded away for zero. Projection algorithms of this type include Newton's method for constrained minimization [10, 14], and several network flow algorithms [1, 4, 5, 13]. Except for the case where T is the gradient of a convex function, there are no convergence results in the literature for this algorithm, even when A is the identity matrix. We show that if care is taken in the way the matrices S_k are allowed to change, then the resulting algorithm is convergent at a rate which is at least linear. We also provide a computational example involving a traffic assignment problem.

2. Projection methods for variational inequalities

Let X^* be the set of all solutions of the variational inequality (1). We have that X^* is polyhedral and is given by

$$X^* = \{x \in X \mid Ax = y^*\} \tag{9}$$

where y^* is the unique solution of the variational inequality (6). For any $x \in \mathbf{R}^n$ we denote by $p(x)$ its unique projection on X^* with respect to the norm (5), i.e.,

$$p(x) = \arg\min\{\|x - z\| \mid z \in X^*\}. \tag{10}$$

We recall that projection on a convex set is a nonexpansive mapping (see e.g. [9]), so we have

$$\|p(x_1) - p(x_2)\| \le \|x_1 - x_2\|, \quad \forall\, x_1, x_2 \in \mathbf{R}^n. \tag{11}$$

We denote the mapping $A'TA$ by $\bar{T}$, i.e.,

$$\bar{T}(x) = A'T(Ax), \quad \forall\, x \in \mathbf{R}^n. \tag{12}$$

With this notation we have

$$(x - x^*)'\bar{T}(x^*) \geq 0, \quad \forall\, x \in X, x^* \in X^*. \tag{13}$$

We also denote for all $x \in X$, $\alpha > 0$

$$\hat{x}(x, \alpha) = P_X^S[x - \alpha S^{-1}\bar{T}(x)]. \tag{14}$$

With this notation the projection iteration (4) is written as

$$x_{k+1} = \hat{x}(x_k, \alpha), \quad x_0 \in X. \tag{15}$$

It can be seen that we can also obtain x_{k+1} as the unique solution to the problem

$$\begin{aligned} &\text{minimize} \quad (x - x_k)'\bar{T}(x_k) + \frac{1}{2\alpha}(x - x_k)'S(x - x_k), \\ &\text{subject to} \quad x \in X. \end{aligned} \tag{16}$$

In view of (13) it is easy to see that for all solutions $x^* \in X^*$ and $\alpha > 0$ we have

$$\hat{x}(x^*, \alpha) = x^* = p(x^*). \tag{17}$$

Thus if $x_k \in X^*$ for some k the algorithm (15) essentially terminates.

Our main result is given in the following proposition.

Proposition 1. *Assume that T is Lipschitz continuous and strongly monotone, and X is polyhedral.*

(a) *There exist positive scalars $q(S)$ and $r(S)$ depending continuously on S, such that for all $x \in X$ and $\alpha > 0$*

$$\|\hat{x}(x, \alpha) - p[\hat{x}(x, \alpha)]\|^2 \leq [1 - 2\alpha q(S) + \alpha^2 r(S)]\|x - p(x)\|^2. \tag{18}$$

(b) *There exists $\bar{\alpha} > 0$ such that for all $\alpha \in (0, \bar{\alpha}]$ the sequence $\{x_k\}$ generated by iteration (15) converges to a solution x^* of the variational inequality (1). The rate of convergence is at least linear in the sense that for each $\alpha \in (0, \bar{\alpha}]$ and $x_0 \in X$ there exist scalars β (depending on α) and q (depending on α and x_0) such that $q > 0$, $\beta \in (0, 1)$, and*

$$\|x_k - x^*\| \leq q\beta^k, \quad k = 0, 1, \ldots$$

The proof of Proposition 1 relies on the following lemma, the proof of which is relegated to the appendix. The lemma is easy to conjecture in terms of geometrical arguments (see Fig. 1).

Lemma 1. *Assume the conditions of Proposition 1 hold. Let*

$$\tilde{X}^* = \{x \in \mathbf{R}^n \mid Ax = y^*\}, \tag{19}$$

$$\tilde{p}(x) = \arg\min\{\|x - z\| \mid z \in \tilde{X}^*\}. \tag{20}$$

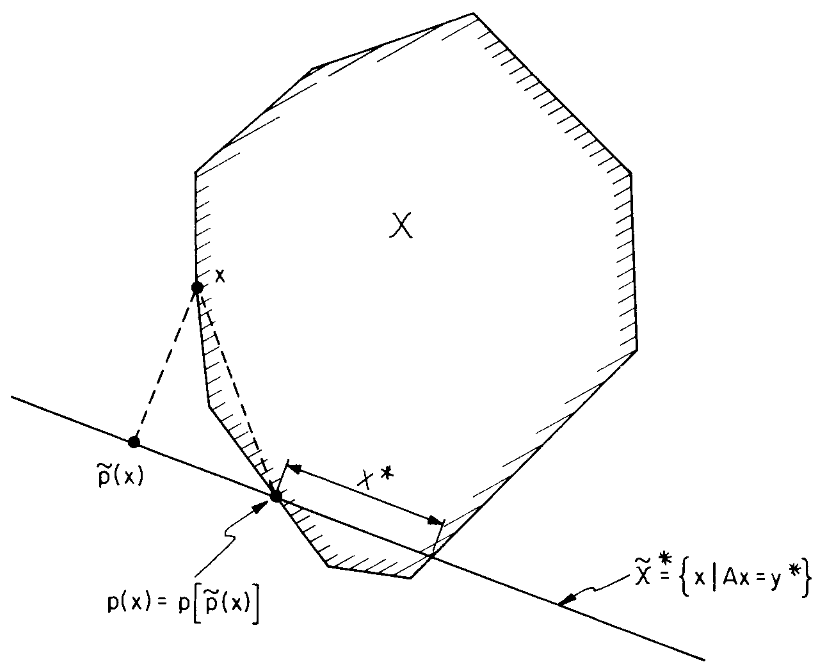

Fig. 1. Geometric interpretation of Lemma 1.

Then $p(x) = p[\tilde{p}(x)]$ *and then there exists a positive scalar* $\eta(S)$ *depending continuously on* S *such that for all* $x \in X$

$$\|x - \tilde{p}(x)\|^2 \geq \eta(s)\|p(x) - \tilde{p}(x)\|^2. \tag{21}$$

Proof of Proposition 1. (a) From (10) we obtain

$$\|\hat{x}(x, \alpha) - p[\hat{x}(x, \alpha)]\|^2 \leq \|\hat{x}(x, \alpha) - p(x)\|^2. \tag{22}$$

Also, using (17) and (11) we have

$$\begin{aligned} \|\hat{x}(x, \alpha) - p(x)\|^2 &= \|\hat{x}(x, \alpha) - \hat{x}[p(x), \alpha]\|^2 \\ &\leq \|[x - p(x)] - \alpha S^{-1}[\bar{T}(x) - \bar{T}[p(x)]]\|^2 \\ &= \|x - p(x)\|^2 - 2\alpha[\bar{T}(x) - \bar{T}[p(x)]]'[x - p(x)] \\ &\quad + \alpha^2\|S^{-1}[\bar{T}(x) - \bar{T}[p(x)]]\|^2. \end{aligned} \tag{23}$$

In view of (22) and (23) it will be sufficient to show that there exist positive scalars $q(S)$ and $r(S)$ depending continuously on S such that

$$[\bar{T}(x) - \bar{T}[p(x)]]'[x - p(x)] \geq q(S)\|x - p(x)\|^2, \quad \forall x \in X, \tag{24a}$$

$$\|S^{-1}[\bar{T}(x) - \bar{T}[p(x)]]\|^2 \leq r(S)\|x - p(x)\|^2, \quad \forall x \in X. \tag{24b}$$

To prove (24a) we first use (3) and (12) to obtain for all $x \in X$ and $y = Ax$

$$\begin{aligned}[\bar{T}(x) - \bar{T}[p(x)]]'[x - p(x)] &= [T(Ax) - T[Ap(x)]]'[Ax - Ap(x)] \\ &= [T(y) - T(y^*)]'(y - y^*) \geq \lambda |y - y^*|^2. \quad (25)\end{aligned}$$

We now consider next the range $R(A)$ and the nullspace $N(A)$ of A, and the orthogonal complement

$$N(A)^+ = \{z \in \mathbf{R}^n \mid z'x = 0, \quad \forall\, x \in N(A)\}.$$

When viewed as a mapping from $N(A)^+$ to $R(A)$, A is one-to-one and onto, and since we are dealing with finite dimensional spaces we have that there exists a positive scalar $\beta(S)$ depending continuously on S such that

$$|w|^2 \geq \beta(S)\|z\|^2, \quad \forall\, w \in R(A),\, z \in N(A)^+,\, w = Az. \quad (26)$$

Since for all $x \in X$, $y = Ax$, and $\tilde{p}(x)$ defined by (20), we have

$$y - y^* = A[x - \tilde{p}(x)] \quad \text{and} \quad x - \tilde{p}(x) \in N(A)^+$$

we obtain from (26)

$$|y - y^*|^2 \geq \beta(S)\|x - \tilde{p}(x)\|^2. \quad (27)$$

Using the Pythagorean theorem and (21) we have

$$\begin{aligned}\|x - p(x)\|^2 &= \|x - \tilde{p}(x)\|^2 + \|p(x) - \tilde{p}(x)\|^2 \\ &\leq \left[1 + \frac{1}{\eta(S)}\right]\|x - \tilde{p}(x)\|^2. \quad (28)\end{aligned}$$

By combining (25), (27) and (28) we obtain

$$[\bar{T}(x) - \bar{T}[p(x)]]'[x - p(x)] \geq \lambda\beta(S)\left[1 + \frac{1}{\eta(S)}\right]^{-1}\|x - p(x)\|^2$$

so (24a) holds with

$$q(S) = \lambda\beta(S)\left[1 + \frac{1}{\eta(S)}\right]^{-1}.$$

To show (24b) we first use (2) and (12) to write

$$\begin{aligned}\|S^{-1}[\bar{T}(x) - \bar{T}[p(x)]]\|^2 &= [T(y) - T(y^*)]'AS^{-1}A'[T(y) - T(y^*)] \\ &\leq \Lambda(S)|T(y) - T(y^*)|^2 \leq \Lambda(S)L^2|y - y^*|^2 \quad (29)\end{aligned}$$

where $\Lambda(S)$ is the largest eigenvalue of $AS^{-1}A'$.

By the same argument used to derive (26) we can assert that there exists a positive scalar $\gamma(S)$ depending continuously on S such that

$$|w|^2 \leq \gamma(S)\|z\|^2, \quad \forall\, w \in R(A),\, z \in N(A)^+, \quad w = Az. \quad (30)$$

Therefore

$$|y - y^*|^2 \leq \gamma(S)\|x - \tilde{p}(x)\|^2 \leq \gamma(S)\|x - p(x)\|^2. \quad (31)$$

Combination of (29) and (31) yields (24b) for $r(S) = \Lambda(S)L^2\gamma(S)$. This completes the proof of part (a).

(b) Let $\bar{\alpha}$ be any positive scalar such that $\bar{\alpha} < (2q(S))/r(S)$. Then using (18) we have for all $\alpha \in (0, \bar{\alpha}]$.

$$\|x_{k+1} - p(x_{k+1})\|^2 \leq t(\alpha)\|x_k - p(x_k)\|^2 \tag{32}$$

where $t(\alpha) = 1 - 2\alpha q(S) + \alpha^2 r(S)$. In view of the fact $\alpha \leq \bar{\alpha} < (2q(S))/r(S)$ it is easily seen that $0 \leq t(\alpha) < 1$. We have from (32)

$$\|x_k - p(x_k)\|^2 \leq t(\alpha)^k \|x_0 - p(x_0)\|^2 \tag{33}$$

and using (23) and (24) we obtain

$$\|x_{k+1} - p(x_k)\| \leq t(\alpha)^{1/2}\|x_k - p(x_k)\|. \tag{34}$$

By the triangle inequality

$$\|x_{k+1} - x_k\| \leq \|x_{k+1} - p(x_k)\| + \|x_k - p(x_k)\|. \tag{35}$$

Combining (35) with (33) and (34) we obtain

$$\begin{aligned} \|x_{k+1} - x_k\| &\leq [t(\alpha)^{1/2} + 1]\|x_k - p(x_k)\| \\ &\leq [t(\alpha)^{1/2} + 1] t(\alpha)^{k/2}\|x_0 - p(x_0)\|. \end{aligned} \tag{36}$$

Let $\beta = t(\alpha)^{1/2}$ and $\bar{q} = [t(\alpha)^{1/2} + 1]\|x_0 - p(x_0)\|$. Then (36) can be written as

$$\|x_{k+1} - x_k\| \leq \bar{q}\beta^k.$$

For all $k \geq 0$, $m \geq 1$ we have

$$\begin{aligned} \|x_{k+m} - x_k\| &\leq \|x_{k+m} - x_{k+m-1}\| + \cdots + \|x_{k+1} - x_k\| \\ &\leq \bar{q}(\beta^{k+m-1} + \cdots + \beta^k) = \frac{\bar{q}\beta^k(1-\beta^m)}{1-\beta}. \end{aligned} \tag{37}$$

Since in view of $t(\alpha) < 1$ we have $\beta < 1$, it follows that $\{x_k\}$ is a Cauchy sequence and hence converges to a vector x^*. Since by (33) $\{x_k - p(x_k)\}$ converges to zero we must have $x^* = p(x^*)$ which implies $x^* \in X^*$. By taking limit in (37) as $m \to \infty$ we obtain for all $k = 0, 1, \ldots$

$$\|x_k - x^*\| \leq q\beta^k$$

where $q = \bar{q}(1-\beta)^{-1}$. This completes the proof.

It is quite remarkable that as shown in Proposition 1(b), the sequence $\{x_k\}$ converges to a single limit point even though the set of solutions X^* may contain an infinite number of points.

3. An algorithm for the traffic assignment problem

In this section we consider the following network flow problem. A network consisting of a set of nodes $\mathcal{N}$ and a set of directed links $\mathcal{L}$ is given, together with a set W of node pairs referred to as *origin–destination* (*OD*) *pairs*. For OD pair $w \in W$ there is a known demand $d_w > 0$ representing traffic entering the network at the origin and exiting at the destination. For each OD pair w, the demand d_w is to be distributed among a given collection P_w of simple directed paths joining w. We denote by x_p the flow carried by path p. Thus the set of feasible path flow vectors $x = \{x_p \mid p \in P_w, w \in W\}$ is given by

$$X = \left\{x \,\middle|\, \sum_{p \in P_w} x_p = d_w, x_p \geq 0, \forall\, p \in P_w, w \in W\right\}. \tag{38}$$

Each collection of path flows $x \in X$ defines a collection of link flows y_{ij}, $(i, j) \in \mathcal{L}$ by means of the equation

$$y_{ij} = \sum_{w \in W} \sum_{p \in P_w} \delta_p(i, j) x_p, \quad \forall (i, j) \in \mathcal{L} \tag{39}$$

where $\delta_p(i, j) = 1$ if path p contains link (i, j) and $\delta_p(i, j) = 0$ otherwise. The vector of link flows $y = \{y_{ij} \mid (i, j) \in \mathcal{L}\}$ corresponding to $x \in X$ can be written as $y = Ax$ where A is the arc-chain matrix defined by (39). The set of feasible link flows is thus

$$Y = AX = \{y \mid y = Ax, x \in X\}. \tag{40}$$

We assume that for each link $(i, j) \in \mathcal{L}$ there is given a function $T_{ij} : Y \to \mathbf{R}$ such that $T_{ij}(y) > 0$ for all $y \in Y$. The value of $T_{ij}(y)$ represents a measure of delay in traversing link (i, j) when the set of link flows is y (travel time in transportation networks [1, 8], marginal delay in communication networks [4, 13]. The vector with components $T_{ij}(y)$ is denoted $T(y)$. We assume that $T(y)$ is Lipschitz continuous and strongly monotone. This is a reasonable assumption for transportation networks, as well as for communication networks.

For each $x \in X$ and corresponding $y = Ax$ the vector $T(y)$ defines for each $w \in W$ and $p \in P_w$ a length

$$\bar{T}_p(x) = \sum_{(i, j) \in \mathcal{L}} \delta_p(i, j) T_{ij}(y) \tag{41}$$

which may be viewed as the total travel time of path p. The problem is to find $x^* \in X$ such that for all $\bar{p} \in P_w$ and $w \in W$

$$\bar{T}_{\bar{p}}(x^*) = \min_{p \in P_w} \bar{T}_p(x^*), \quad \text{if } x^*_{\bar{p}} > 0. \tag{42}$$

This problem is based on the user-optimization principle which asserts that a traffic network equilibrium is established when no user may decrease his travel time by making a unilateral decision to change his route. If we denote by $\bar{T}(x)$ the vector of lengths $\bar{T}_p(x)$, $p \in P_w$, $w \in W$ then it is easy to see [1, 8] using (39) and (41) that

$$\bar{T}(x) = A'T(Ax) \tag{43}$$

and that the problem defined earlier via (42) is equivalent to finding a solution $x^* \in X$ of the variational inequality

$$(x - x^*)'\bar{T}(x^*) \geq 0, \quad \forall\, x \in X, \tag{44}$$

which is of the form (1).

Let $W = \{w_1, w_2, \ldots, w_M\}$ and consider the algorithm of the previous section with a matrix S which is block diagonal of the form

$$S = \begin{bmatrix} S_1 & & 0 \\ & S_2 & \\ 0 & & S_M \end{bmatrix}.$$

Here each matrix S_i corresponds to the OD pair w_i. We assume that each matrix S_i is positive definite symmetric. The projection iteration (4) can be implemented by finding x_{k+1} solving the quadratic program (16). In view of the block diagonal form of S and the decomposable nature of the constraint set X (cf. (38)) the quadratic program can be decomposed into a collection of smaller quadratic programs—one per OD pair $w \in W$. The form of these programs for the case where each matrix S_i is diagonal with elements s_p, $p \in P_{w_i}$ along the diagonal is

$$\begin{aligned} &\text{minimize} && \sum_{p \in P_{w_i}} \left\{ (x_p - x_p^k)\bar{T}_p(x_k) + \frac{s_p}{2\alpha}(x_p - x_p^k)^2 \right\}, \\ &\text{subject to} && \sum_{p \in P_{w_i}} x_p = d_{w_i}, \\ & && x_p \geq 0, \quad \forall\, p \in P_{w_i} \end{aligned} \tag{45}$$

where x_p^k, $p \in P_{w_i}$, $i = 1, \ldots, M$ are the components of the vector x_k. Problem (45) involves a single equality constraint and can be solved very easily—essentially in closed form [4, 12]. The convergence and linear rate of convergence results of Proposition 1 apply to this algorithm.

The preceding algorithm is satisfactory if each set P_{w_i} contains relatively few paths. In some problems however the number of paths in P_{w_i} can be very large (for example P_{w_i} may contain all simple paths joining w_i). In this case it is preferable to start with a subset of each P_{w_i} and augment this subset as necessary as suggested by Aashtiani [1]. The corresponding algorithm is as follows:

We begin with a subset $P_{w_i}^0 \subset P_{w_i}$ for each $w_i \in W$ and a vector of initial path flows x_0 such that for all $w_i \in W$

$$x_p^0 = 0, \quad \text{if } p \notin P_{w_i}^0.$$

At the kth iteration we have for each OD pair $w_i \in W$ a subset of paths $P^k_{w_i} \subset P_{w_i}$ and a corresponding vector of path flows x_k satisfying

$$x^k_p = 0, \quad \text{if } p \notin P^k_{w_i}.$$

We compute, for each $w_i \in W$, a shortest path $p^k_{w_i} \in P_{w_i}$ using $T_{ij}(Ax_k)$ as length for each link (i, j). We set

$$P^{k+1}_{w_i} = P^k_{w_i} \cup \{p^k_{w_i}\}.$$

(Note here that $p^k_{w_i}$ may already belong to $P^k_{w_i}$). We then solve the quadratic programming problem

$$\begin{aligned} &\text{minimize} \quad \sum_{p \in P^{k+1}_{w_i}} \left\{(x_p - x^k_p)\bar{T}_p(x_k) + \frac{s_p}{2\alpha}(x_p - x^k_p)^2\right\} \\ &\text{subject to} \quad \sum_{p \in P^{k+1}_{w_i}} x_p = d_{w_i}, \\ &\qquad\qquad\quad x_p \geq 0, \quad \forall\, p \in P^{k+1}_{w_i}. \end{aligned} \tag{46}$$

If $\bar{x}^k_p$, $p \in P^{k+1}_{w_i}$ is the solution of this problem we set

$$x^{k+1}_p = \begin{cases} \bar{x}^k_p, & \text{if } p \in P^{k+1}_{w_i}, \\ 0, & \text{if } p \notin P^{k+1}_{w_i}. \end{cases} \tag{47}$$

Note that the quadratic programming problem (46) is the same as (45) except for the fact that it involves only paths in the subset $P^{k+1}_{w_i}$. The subset is possibly augmented at each iteration k to include the current shortest path $p^k_{w_i}$. The expectation here is that, while P_{w_i} may contain a very large number of paths, the actual number of paths generated and included in the set $P^k_{w_i}$ remains small as k increases. This expectation has been supported by computational experiments [1]. The convergence and rate of convergence properties of the algorithm (46), (47) are identical with those of the earlier algorithm based on (45) as the reader can easily verify. The key idea is based on the fact that for each i, P_{w_i} contains a finite number of paths and $P^k_{w_i}$ grows monotonically so that the sequence $\{P^k_{w_i}\}$ converges to some subset of paths $\bar{P}_{w_i} \subset P_{w_i}$. By applying Proposition 1 it follows that the algorithm converges to a solution of the problem obtained when P_{w_i} is replaced by $\bar{P}_{w_i}$. Because $P^k_{w_i}$ is augmented at each iteration with the current shortest path it is a simple matter to conclude that the solution x^* obtained satisfies the required minimum travel time condition (42).

The results and algorithms of this section can be strengthened considerably in the case where T is in addition the gradient of convex function F in which case the problem is equivalent to the convex programming problem of minimizing $F(y)$ subject to $y \in Y$ or minimizing $F(Ax)$ subject to $x \in X$. Under these circumstances there are convergence results [4, 5, 12] relating to projection type algorithms which allow for a variable matrix S and for coordinate descent type iterations whereby each iteration is performed with respect to a single OD pair (or a small group of OD pairs) and all OD pairs are taken up in sequence. In fact for such algorithms it seems that it is easier to select an appropriate value for the stepsize α. Aashtiani's computational experience [1] suggests that such algorithms also work well in many cases where T is not a gradient mapping. We have been unable however to obtain a general convergence result for coordinate descent versions of the projection algorithm. The possibility of changing the matrix S from one iteration to the next is considered in the next section.

4. An algorithm with variable projection metric

A drawback of the algorithms of Section 2 and 3 is that the matrix S is restricted to be the same at each iteration. Computational experience with optimization problems as well as network flow problems [1, 6] suggests that, if T is differentiable, better results can be obtained if the matrix S is varied from one iteration to the next and is made suitably dependent on first derivatives of the mapping T in a manner which approximates Newton's method. We have not been able to show the result of Proposition 1(b) for algorithms in which the matrix S may change arbitrarily. On the other hand it is possible to construct an algorithmic scheme that allows for a variable matrix S but at the same time incorporates a mechanism that safeguards against divergence. The main idea in this scheme is *to allow a change in the matrix S only when the algorithm makes satisfactory progress towards convergence.* The algorithm is as follows:

A set $\mathcal{S}$ of positive definite symmetric matrices is given. It is assumed that all eigenvalues of all matrices $S \in \mathcal{S}$ lie in some compact interval of positive real numbers, i.e., there exist $m_1, m_2 > 0$ such that $m_1|z|^2 \leq z'Sz \leq m_2|z|^2$, for all $z \in \mathbf{R}^n$, $S \in \mathcal{S}$. We consider the algorithm

$$x_{k+1} = P_X^{S_k}[x_k - \alpha S_k^{-1}\bar{T}(x_k)], \quad x_0 \in X \tag{48}$$

where $S_k \in \mathcal{S}$ for all $k = 0, 1, \ldots$. The stepsize α is such that

$$h \overset{\Delta}{=} \max_{S \in \mathcal{S}}[1 - 2\alpha q(S) + \alpha^2 r(S)] < 1, \tag{49}$$

where $q(S)$ and $r(S)$ are as in Proposition 1(a) [c.f. (18)]. The maximum in (49) is attained by continuity of $q(\cdot)$ and $r(\cdot)$. It is clear that there exists $\bar{\alpha} > 0$ such that (49) is satisfied for all $\alpha \in (0, \bar{\alpha}]$. Given x_{k+1} the matrix S_{k+1} is either chosen arbitrarily from $\mathcal{S}$ or else $S_{k+1} = S_k$ depending on whether the quantity

$$w_k = (x_{k+1} - x_k)' S_k (x_{k+1} - x_k) \tag{50}$$

has decreased or not by a certain factor over the last time the matrix S was changed. More specifically a scalar $\bar{\beta} \in (0, 1)$ (typically close to unity) is chosen, and at each iteration k a scalar $\bar{w}_{k+1}$ is computed according to

$$\bar{w}_{k+1} = \begin{cases} \bar{\beta} w_k, & \text{if } w_k \leq \bar{w}_k, \\ \bar{w}_k, & \text{if } w_k > \bar{w}_k \end{cases} \tag{51}$$

where w_k is given by (50) and initially $\bar{w}_0 = \infty$. We select

$$S_{k+1} = S_k, \quad \text{if } \bar{w}_{k+1} = \bar{w}_k, \tag{52a}$$

$$S_{k+1} \in \mathcal{S}, \quad \text{if } \bar{w}_{k+1} < \bar{w}_k. \tag{52b}$$

Thus for each k, the scalar $\bar{w}_k$ represents a target value below which w_k must drop in order for a change in S to be allowed in the next iteration.

We first show that if $\{x_k\}$ is a sequence generated by the algorithm just described, then

$$\liminf_{k\to\infty} w_k = 0. \tag{53}$$

Indeed if $\liminf_{k\to\infty} w_k > 0$, then S_k must have been allowed to change only a finite number of times in which case it follows from Propostion 1 that $\{x_k\}$ converges to a solution x^*. As a result we have $w_k \to 0$ contradicting the earlier assertion.

Let us denote by $\|\cdot\|_k$ the norm corresponding to S_k and by $p_k(z)$ the projection of a vector $z \in \mathbf{R}^n$ on X^* with respect to $\|\cdot\|_k$. We have by using the triangle inequality, (49), (50) and Proposition 1(a)

$$\begin{aligned} w_k^{1/2} &= \|x_{k+1} - x_k\|_k \\ &\geq \|x_k - p_k(x_{k+1})\|_k - \|x_{k+1} - p_k(x_{k+1})\|_k \\ &\geq \|x_k - p_k(x_k)\|_k - \sqrt{h}\|x_k - p_k(x_k)\|_k \\ &= (1-\sqrt{h})\|x_k - p_k(x_k)\|_k. \end{aligned} \tag{54}$$

Hence (53) implies $\liminf_{k\to\infty}\|x_k - p_k(x_k)\|_k = 0$ or equivalently (in view of the fact $\|z\|_k^2 \geq m_1|z|^2$, $m_1 > 0$)

$$\liminf_{k\to\infty} |x_k - p_k(x_k)| = 0. \tag{55}$$

This means that at least a subsequence of $\{x_k\}$ converges to the solution set X^*. We will show that in fact for some vector $x^* \in X^*$, and some scalars $q > 0$ and $\beta \in (0, 1)$ we have

$$|x_k - x^*| \leq q\beta^k, \quad \forall\, k = 0, 1, \ldots$$

i.e., $\{x_k\}$ converges to a solution x^* at a rate which is at least linear.

We have

$$\begin{aligned} w_k &= \|x_k - p_k(x_k) + p_k(x_k) - x_{k+1}\|_k^2 \\ &\leq \|x_k - p_k(x_k)\|_k^2 + \|p_k(x_k) - x_{k+1}\|_k^2 \\ &\qquad + 2\|x_k - p_k(x_k)\|_k \|p_k(x_k) - x_{k+1}\|_k. \end{aligned} \tag{56}$$

Using (23), (24) and (49) we obtain

$$\|p_k(x_k) - x_{k+1}\|_k^2 \le h\|x_k - p_k(x_k)\|_k^2. \tag{57}$$

Combination of (56) and (57) yields

$$w_k \le (1+\sqrt{h})^2\|x_k - p_k(x_k)\|_k^2. \tag{58}$$

Also in view of (49) and Proposition 1 we have that there exists a scalar $d \ge 1$ such that for all k

$$\begin{aligned}\|x_{k+1} - p_{k+1}(x_{k+1})\|_{k+1}^2 &\le \|x_{k+1} - p_k(x_{k+1})\|_{k+1}^2\\ &\le d\|x_{k+1} - p_k(x_{k+1})\|_k^2\\ &\le dh\|x_k - p_k(x_k)\|_k^2\end{aligned}$$

and finally using (54)

$$\|x_{k+1} - p_{k+1}(x_{k+1})\|_{k+1}^2 \le \frac{dh}{(1-\sqrt{h})^2}\, w_k. \tag{59}$$

Let

$$\mathscr{K} = \{k \mid w_k \le \bar{w}_k\}. \tag{60}$$

By (51) and (52) we have

$$S_{k+1} = S_k, \quad \forall\, k \notin \mathscr{K}. \tag{61}$$

Also from (51) and (53) and the fact that if $w_{\bar{k}} = 0$, then $w_k = 0$ for all $k \ge \bar{k}$, it follows that $\mathscr{K}$ contains an infinite number of indices. Let k_1 and k_2 be two successive indices in $\mathscr{K}$ with $k_1 < k_2$. Then

$$w_{k_2} \le \bar{w}_{k_2} = \bar{\beta} w_{k_1} \tag{62}$$

while, if $k_2 - k_1 > 1$, we have

$$w_{k_1+m} > \bar{w}_{k_2} = \bar{\beta} w_{k_1}, \quad \forall\, m = 1, \dots, (k_2 - k_1 - 1). \tag{63}$$

We also have $S_{k_1+1} = \cdots = S_{k_2}$. In the case where $k_2 - k_1 > 1$, Proposition 1(a) together with (49) and (59) yields for all $m = 1, \dots, (k_2 - k_1 - 1)$

$$\begin{aligned}\|x_{k_1+m} - p_{k_1+m}(x_{k_1+m})\|_{k_1+1}^2 &\le h^{m-1}\|x_{k_1+1} - p_{k_1+1}(x_{k_1})\|_{k_1+1}^2\\ &\le \frac{dh^{m-1}}{(1-\sqrt{h})^2}\, w_{k_1}.\end{aligned} \tag{64}$$

Using (58) and (64) we have

$$w_{k_1+m} \le \frac{dh^{m-1}(1+\sqrt{h})^2}{(1-\sqrt{h})^2}\, w_{k_1}, \quad \forall\, m = 1, \dots, (k_2 - k_1 - 1). \tag{65}$$

Inequalities (63) and (65) yield

$$\bar{\beta} w_{k_1} < w_{k_2-1} \le \frac{dh^{k_2-k_1-2}(1+\sqrt{h})^2}{(1-\sqrt{h})^2}\, w_{k_1}.$$

It follows that if $k_2 - k_1 \geq 1$, then

$$k_2 - k_1 \leq 2 + \frac{\ln \dfrac{\bar{\beta}(1-\sqrt{\bar{h}})^2}{d(1+\sqrt{\bar{h}})^2}}{\ln h}, \tag{66}$$

so if $\bar{m}$ is any positive integer such that

$$2 + \frac{\ln \dfrac{\bar{\beta}(1-\sqrt{\bar{h}})^2}{d(1+\sqrt{\bar{h}})^2}}{\ln h} \leq \bar{m} \tag{67}$$

we have

$$k_2 - k_1 \leq \bar{m} \tag{68}$$

for any two successive indices k_1, k_2 in $\mathcal{K}$. It follows using (62) that

$$w_k \leq w_0(\bar{\beta}^{1/\bar{m}})^k, \quad \forall\, k \in \mathcal{K}. \tag{69}$$

Using (65), (68) and (69) we also obtain

$$w_k \leq \frac{d(1+\sqrt{\bar{h}})^2 w_0}{(1-\sqrt{\bar{h}})^2}(\bar{\beta}^{1/\bar{m}})^{k-\bar{m}}, \quad \forall\, k \notin \mathcal{K}. \tag{70}$$

Combining (69) and (70) we have for some scalar $\bar{q} > 0$

$$w_k \leq \bar{q}(\bar{\beta}^{1/\bar{m}})^k, \quad \forall\, k = 0, 1, \ldots \tag{71}$$

Since

$$m_1|x_{k+1} - x_k|^2 \leq \|x_{k+1} - x_k\|_k^2 = w_k$$

relation (71) yields

$$|x_{k+1} - x_k|^2 \leq \frac{\bar{q}}{m_1}(\bar{\beta}^{1/\bar{m}})^k, \quad \forall\, k = 0, 1, \ldots$$

Since $\bar{\beta} < 1$, it follows in exactly the same manner as in the proof of Proposition 1(b) that $\{x_k\}$ is a Cauchy sequence which converges to a vector $x^* \in X^*$. Furthermore for some $q > 0$ and $\beta \in (0, 1)$ we have

$$|x_k - x^*| \leq q\beta^k, \quad k = 0, 1, \ldots \tag{72}$$

We have thus proved the following proposition.

Proposition 2. *There exists $\bar{\alpha} > 0$ such that if $\alpha \in (0, \bar{\alpha}]$, a sequence $\{x_k\}$ generated by iteration* (48) *with $\{S_k\}$ selected according to* (50)–(52) *converges to a solution x^* of the variational inequality* (1). *Furthermore there exist scalars $q > 0$ and $\beta \in (0, 1)$ such that*

$$|x_k - x^*| \leq q\beta^k, \quad \forall\, k = 0, 1, \ldots$$

5. A computational example

In this section we report computational results for a traffic assignment problem. The corresponding network is shown in Fig. 2, and may be viewed as a model of a circular highway. There are five origins and destinations numbered 1, 2, 3, 4, 5 and connected through the highway via entrance and exit ramps. We consider the five OD pairs (1, 4), (2, 5), (3, 1), (4, 2), (5, 3). Each OD pair has two paths associated with it—the clockwise and counterclockwise paths on the corresponding circle. The expressions for the travel time on each link are shown in Fig. 2 where the function g is given by $g(x) = 1 + x + x^2$. Different values of the nonnegative scalar γ represent different degrees of dependence of the travel times of some links on the flows of other links. The problem is equivalent to an optimization problem if and only if there is no such dependence ($\gamma = 0$).

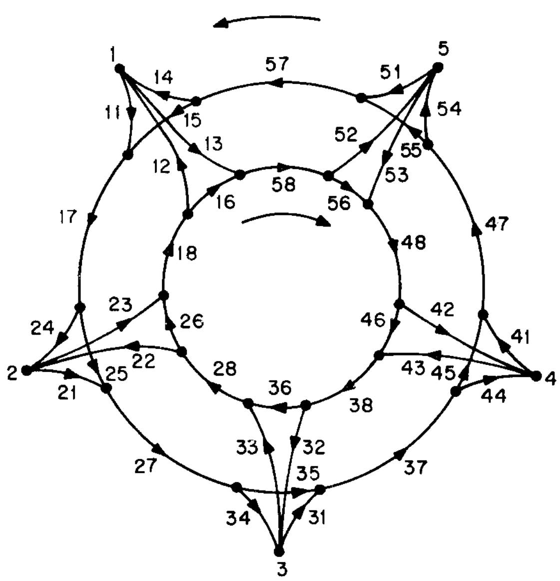

Fig. 2.

Types of links:

(1) Highway links: 17, 27, 37, 47, 57, 18, 28, 38, 48, 58.

(2) Exit ramps: 14, 24, 34, 44, 54, 12, 22, 32, 42, 52.

(3) Entrance ramps: 11, 21, 31, 41, 51, 13, 23, 33, 43, 53.

(4) Bypass links: 15, 25, 35, 45, 55, 16, 26, 36, 46, 56.

Delay on links [where g is defined by $g(x) = 1 + x + x^2$]:

(1) Delay on highway link k: $10 \cdot g$[Flow on k]$+ 2 \cdot \gamma \cdot g$[Flow on exit ramp from k].

(2) Delay on exit ramp k: g[Flow on k].

(3) Delay on entrance ramp k: g[Flow on k]$+ \gamma \cdot g$[Flow on bypass link merging with k].

(4) Delay on bypass link k: g[Flow on k].

Remark: Flow is not allowed to use exit ramp not leading to its destination.

OD Pairs: (1, 4), (2, 5), (3, 1), (4, 2), (5, 3).

Tables 1 and 2 list representative computational results for two demand patterns, three values of γ, and fifteen iterations of two different algorithms labeled 'all-at-once' and 'one-at-a-time' and described below. The number shown in the tables for each iteration k is the following normalized measure of convergence

$$\sum_{\text{all OD pairs } w} \frac{\Delta x_{w,k}}{d_w} \frac{\Delta \bar{T}_{w,k}}{\bar{T}_{\min, w, k}} \tag{73}$$

where d_w is the demand of the OD pair w, $\Delta x_{w,k}$ is the portion of the demand that does not lie on the shortest path of the OD pair w at the end of iteration k, $\Delta \bar{T}_{w,k}$ is the difference of the travel times of the longest and shortest paths and $\bar{T}_{\min, w, k}$ is the travel time of the shortest path. Clearly the expression (73) is zero if and only if the corresponding traffic assignment is optimal. The starting flow pattern in all runs was the worst possible whereby all the demand of each OD pair is routed on the counter-clockwise path. The results suggest that the algorithms yield near optimal flow patterns after very few iterations and subsequently continue their progress at a fairly satisfactory rate. This type of convergence behavior is consistent with the one observed for related algorithms tested in [6].

The 'all-at-once' algorithm is the one of the previous section [c.f. (48), (50)–(52)] with the projection matrix S_k being diagonal [c.f. (45)]. For each

Table 1
Demands $d(1,4) = 0.1$; $d(2,5) = 0.2$; $d(3,1) = 0.3$; $d(4,2) = 0.4$; $d(5,3) = 0.5$

	All-at-once, $\alpha = 0.8$, $\bar{\beta} = 0.99$			One-at-a-time, $\alpha = 1$, $\bar{\beta} = 0.99$		
K	$\gamma = 0$	$\gamma = 0.5$	$\gamma = 4$	$\gamma = 0$	$\gamma = 0.5$	$\gamma = 4$
0	0.14417×10^{2}	0.14793×10^{2}	0.17426×10^{2}	0.14417×10^{2}	0.14793×10^{2}	0.17426×10^{2}
1	0.14897×10^{1}	0.15079×10^{1}	0.16436×10^{1}	0.43831×10^{0}	0.46175×10^{0}	0.36140×10^{0}
2	0.39463×10^{0}	0.36291×10^{0}	0.23633×10^{0}	0.73026×10^{-1}	0.80765×10^{-1}	0.13319×10^{0}
3	0.35901×10^{0}	0.29642×10^{0}	0.12340×10^{0}	0.28867×10^{-1}	0.37109×10^{-1}	0.39248×10^{-1}
4	0.55230×10^{-1}	0.41860×10^{-1}	0.11996×10^{-1}	0.11671×10^{-1}	0.17501×10^{-1}	0.15697×10^{-1}
5	0.80434×10^{-1}	0.52264×10^{-1}	0.55236×10^{-2}	0.50979×10^{-2}	0.86871×10^{-2}	0.10786×10^{-1}
6	0.11485×10^{-1}	0.70972×10^{-2}	0.14528×10^{-2}	0.23552×10^{-2}	0.45007×10^{-2}	0.76687×10^{-2}
7	0.19034×10^{-1}	0.93712×10^{-2}	0.71098×10^{-3}	0.11224×10^{-2}	0.23972×10^{-2}	0.55389×10^{-2}
8	0.26034×10^{-2}	0.13097×10^{-2}	0.34892×10^{-3}	0.54303×10^{-3}	0.12962×10^{-2}	0.40331×10^{-2}
9	0.43683×10^{-2}	0.21587×10^{-2}	0.18026×10^{-3}	0.26555×10^{-3}	0.70754×10^{-3}	0.29498×10^{-2}
10	0.55167×10^{-3}	0.11845×10^{-2}	0.99838×10^{-4}	0.12989×10^{-3}	0.38809×10^{-3}	0.21641×10^{-2}
11	0.10228×10^{-2}	0.11548×10^{-2}	0.80656×10^{-4}	0.64334×10^{-4}	0.21363×10^{-3}	0.15911×10^{-2}
12	0.44825×10^{-3}	0.80420×10^{-3}	0.68090×10^{-4}	0.31281×10^{-4}	0.11782×10^{-3}	0.11698×10^{-2}
13	0.51164×10^{-3}	0.67664×10^{-3}	0.59269×10^{-4}	0.15289×10^{-4}	0.64249×10^{-4}	0.86171×10^{-3}
14	0.30760×10^{-3}	0.51037×10^{-3}	0.51284×10^{-4}	0.77946×10^{-5}	0.35938×10^{-4}	0.63501×10^{-3}
15	0.27834×10^{-3}	0.41039×10^{-3}	0.44031×10^{-4}	0.41734×10^{-5}	0.19540×10^{-4}	0.46808×10^{-3}

Table 2
Demands $d(1,4)=1$, $d(2,5)=8$, $d(3,1)=1$, $d(4,2)=8$, $d(5,3)=1$

	All-at-once, $\alpha=0.8$, $\bar{\beta}=0.99$			One-at-a-time, $\alpha=1$, $\bar{\beta}=0.99$		
K	$\gamma=0.0$	$\gamma=0.5$	$\gamma=4$	$\gamma=0$	$\gamma=0.5$	$\gamma=4$
0	0.10203×10^{4}	0.10478×10^{4}	0.12404×10^{4}	0.10203×10^{4}	0.10478×10^{4}	0.12405×10^{4}
1	0.19446×10^{1}	0.15853×10^{1}	0.33203×10^{0}	0.28891×10^{1}	0.24262×10^{1}	0.86532×10^{0}
2	0.83731×10^{0}	0.57695×10^{0}	0.27719×10^{-1}	0.65950×10^{0}	0.49153×10^{0}	0.59564×10^{-1}
3	0.12902×10^{1}	0.90467×10^{0}	0.37155×10^{-1}	0.78641×10^{-1}	0.56482×10^{-1}	0.13636×10^{-1}
4	0.45269×10^{0}	0.24701×10^{0}	0.24096×10^{-1}	0.63059×10^{-2}	0.70762×10^{-2}	0.56463×10^{-2}
5	0.83315×10^{0}	0.42439×10^{0}	0.17847×10^{-1}	0.70066×10^{-3}	0.27832×10^{-2}	0.29389×10^{-2}
6	0.23621×10^{0}	0.79849×10^{-1}	0.12985×10^{-1}	0.10199×10^{-3}	0.12097×10^{-2}	0.16215×10^{-2}
7	0.50649×10^{0}	0.11570×10^{0}	0.95892×10^{-2}	0.79205×10^{-4}	0.45572×10^{-3}	0.90603×10^{-3}
8	0.11333×10^{0}	0.10000×10^{-1}	0.71098×10^{-2}	0.14878×10^{-4}	0.16733×10^{-3}	0.50823×10^{-3}
9	0.26874×10^{0}	0.78683×10^{-2}	0.52748×10^{-2}	0.94583×10^{-4}	0.61370×10^{-4}	0.28569×10^{-3}
10	0.43230×10^{-1}	0.10397×10^{-2}	0.39184×10^{-2}	0.65018×10^{-4}	0.22825×10^{-4}	0.16119×10^{-3}
11	0.11434×10^{0}	0.13673×10^{-2}	0.29147×10^{-2}	0.39872×10^{-4}	0.90891×10^{-5}	0.90105×10^{-4}
12	0.12722×10^{-1}	0.68513×10^{-3}	0.21694×10^{-2}	0.24651×10^{-4}	0.33388×10^{-5}	0.51215×10^{-4}
13	0.43756×10^{-1}	0.50364×10^{-3}	0.16170×10^{-2}	0.15648×10^{-4}	0.10026×10^{-5}	0.27883×10^{-4}
14	0.45550×10^{-2}	0.32649×10^{-3}	0.12047×10^{-2}	0.92416×10^{-5}	0.40281×10^{-6}	0.15870×10^{-4}
15	0.20089×10^{-1}	0.22516×10^{-3}	0.89921×10^{-3}	0.68895×10^{-5}	0.47333×10^{-6}	0.89927×10^{-5}

iteration k for which S_k was allowed to change [c.f. (50)–(52)] the diagonal element of S_k corresponding to any one path was taken equal to the sum of the first derivatives of the travel times of links on that path evaluated at the kth flow x_k. This corresponds to a diagonal approximation of Newton's method (c.f. [4, 5]). As a result stepsizes near unity typically give satisfactory convergence behavior. In all runs we used a stepsize $\alpha = 0.8$ which is probably a bit on the high side. The scalar $\bar{\beta}$ used in the test for allowing the matrix S_k to change [c.f. (51)] was taken to be 0.99 in all runs. One of the most interesting observations from our experimentation was that this test was passed at every iteration and so *the matrix S_k was changed at every iteration.*

The 'one-at-a-time' algorithm is similar to the 'all-at-once' algorithm described above. The only difference is that the projection is carried out with respect to a single OD pair with flows corresponding to the other OD pairs being kept fixed, and all OD pairs are taken up in sequence. An iteration consists of a cycle of five projection subiterations (one per OD pair). Each subiteration is, of course, followed by reevaluation of the travel time of each link. Algorithms of this type resemble coordinate descent methods for unconstrained optimization and have been suggested in the context of network flows in [1, 4, 5, 12]. The stepsize α was taken to be unity in all runs. Also $\bar{\beta}$ was chosen to be 0.99 and again it turned out that the matrix S_k was allowed to change in every iteration. It is interesting to note that for this stepsize and this particular example the algorithm tested is equivalent to an algorithm in the class proposed by Aashtiani [1]. Tables 1 and 2 indicate a better performance for the 'one-at-a-time' algorithm. However, there is no convergence proof available for this algorithm at present except in the case where the corresponding variational inequality is equivalent to a convex optimization problem.

Appendix. Proof of Lemma 1

We have by the Pythagorean theorem

$$\begin{aligned}\|x-\tilde{p}(x)\|^2+\|\tilde{p}(x)-p(x)\|^2 &= \|x-p(x)\|^2 \le \|x-p[\tilde{p}(x)]\|^2\\ &= \|x-\tilde{p}(x)\|^2+\|\tilde{p}(x)-p[\tilde{p}(x)]\|^2\end{aligned}$$

Hence

$$\|\tilde{p}(x)-p(x)\|^2 \le \|\tilde{p}(x)-p[\tilde{p}(x)]\|^2. \tag{A.1}$$

Since $p[\tilde{p}(x)]$ by definition is the unique solution of the problem of minimizing $\|\tilde{p}(x)-z\|^2$ over $x \in X^*$, and $p(x) \in X^*$, it follows that equality holds in (A.1) and $p(x)=p[\tilde{p}(x)]$.

Let $\tilde{p}_1(x)$ and $p_1(x)$ be the projections of x on $\tilde{X}^*$ and X^* respectively relative to the standard norm $|\cdot|$, i.e., $\tilde{p}_1(x)=\tilde{p}(x)$, $p_1(x)=p(x)$ for S equal to the identity. In order to show the existence of a continuous $\eta(S)$ such that (21) holds it will suffice to show the existence of a scalar $\eta_1>0$ such that for all $x \in X$

$$|x-\tilde{p}_1(x)|^2 \ge \eta_1 |p_1(x)-\tilde{p}_1(x)|^2. \tag{A.2}$$

In order to see this let $\Lambda(S)$ and $\lambda(S)$ be the largest and smallest eigenvalue of S. We have for all x, $z \in \mathbf{R}^n$

$$\lambda(S)|x-z|^2 \le \|x-z\|^2 \le \Lambda(S)|x-z|^2.$$

It follows that

$$\frac{1}{\lambda(S)}\|x-\tilde{p}(x)\|^2 \ge |x-\tilde{p}_1(x)|^2 \tag{A.3}$$

$$\frac{1}{\Lambda(S)}\|x-p(x)\|^2 \le |x-p_1(x)|^2. \tag{A.4}$$

By the Pythagorean theorem we have

$$|p_1(x)-\tilde{p}_1(x)|^2 = |x-p_1(x)|^2-|x-\tilde{p}_1(x)|^2$$

so (A.2) can be rewritten as

$$|x-\tilde{p}_1(x)|^2 \ge \frac{\eta_1}{1+\eta_1}|x-p_1(x)|^2. \tag{A.5}$$

From (A.3), (A.4) and (A.5) we obtain

$$\|x - \tilde{p}(x)\|^2 \geq \gamma(S)\|x - p(x)\|^2 \tag{A.6}$$

where

$$\gamma(S) = \frac{\eta_1}{1+\eta_1}\frac{\lambda(S)}{\Lambda(S)}.$$

By the Pythagorean theorem we have

$$\|x - p(x)\|^2 = \|x - \tilde{p}(x)\|^2 + \|p(x) - \tilde{p}(x)\|^2$$

and by using this equation in (A.6) we obtain the desired relation

$$\|x - \tilde{p}(x)\|^2 \geq \eta(S)\|p(x) - \tilde{p}(x)\|^2$$

with

$$\eta(S) = \frac{\gamma(S)}{1-\gamma(S)}.$$

We now show existence of an $\eta_1 > 0$ such that (A.2) holds for all $x \in X$. By the preceding analysis this is sufficient to prove the lemma.

For each $x \in X$ consider the tangent cone C_x of X at $p(x)$, i.e., the set

$$C_x = \{z \mid \text{there exists } \alpha > 0 \text{ such that } [p(x) + \alpha z] \in X\}. \tag{A.7}$$

Let $\mathscr{C}$ be the collection

$$\mathscr{C} = \{C_x \mid x \in X\}. \tag{A.8}$$

It is easily seen that *$\mathscr{C}$ is a finite collection*, i.e., for some finite set $J \subset X$ we have

$$\mathscr{C} = \{C_j \mid j \in J\}. \tag{A.9}$$

Indeed since X is polyhedral it can be represented by definition [15, Section 19], in terms of a finite number of vectors $d_1, \ldots, d_m \in \mathbf{R}^n$ and scalars $b_1, \ldots, b_m$ as

$$X = \{x \mid d_i'x \leq b_i,\ i = 1, \ldots, m\}.$$

It is easily seen that

$$C_x = \{z \mid d_i'z \leq 0, \forall\ i = 1, \ldots, m \text{ such that } d_i'p(x) = b_i\}.$$

Clearly there is only a finite number of sets of the above form. In what follows $P_\Omega(z)$ denotes the projection of a vector $z \in \mathbf{R}^n$ on a closed convex set $\Omega \subset \mathbf{R}^n$ with respect to the standard norm $|\cdot|$. The essence of the proof of Lemma 1 is contained in the following lemma.

Lemma A.1. *For $j \in J$ let*

$$M_j^* = C_j \cap N(A), \tag{A.10}$$

$$Z_j = \{z \mid z \in C_j, P_{M_j^*}(z) = 0\}, \tag{A.11}$$

where $N(A)$ is the nullspace of A. Then for each $j \in J$ there exists a scalar $\eta_j > 0$ such that

$$|z - P_{N(A)}(z)|^2 \geq \eta_j |z|^2, \quad \forall z \in Z_j \tag{A.12}$$

Proof. Assume the contrary, i.e., that there exist $j \in J$ and sequences $\{z_k\} \subset Z_j$, $\{\eta_j^k\} \subset R$ such that

$$|z_k - P_{N(A)}(z_k)|^2 < \eta_j^k |z_k|, \quad \eta_j^k \to 0. \tag{A.13}$$

We then have $z_k \neq 0$, $\forall k$, and since both M_j^* and Z_j are clearly cones with vertex at the origin we can assume that $|z_k| = 1$, $\forall k$. Let $\bar{z}$ be a limit point of $\{z_k\}$. We have $|\bar{z}| = 1$ and by taking limit in (A.13) we obtain $\bar{z} = P_{N(A)}(\bar{z})$ i.e., $\bar{z} \in N(A)$. Since C_j is closed we also have $\bar{z} \in C_j$ and hence $\bar{z} \in M_j^*$. It follows that

$$P_{M_j^*}(\bar{z}) = \bar{z}. \tag{A.14}$$

On the other hand since $z_k \in Z_j$ we have $P_{M_j^*}(z_k) = 0$, $\forall k$ which implies that $P_{M_j^*}(\bar{z}) = 0$. Since $|\bar{z}| = 1$, this contradicts (A.14).

We now show that the desired relation (A.2) holds with

$$\eta_1 = \min\{\eta_j \mid j \in J\} > 0. \tag{A.15}$$

Choose any $x \in X$ and let $j \in J$ be such that $C_x = C_j$. Let

$$z = x - p_1(x). \tag{A.16}$$

By a simple translation argument, the fact that $p_1(x)$ is the projection of x on $X \cap \{x \mid Ax = y^*\}$ implies that the projection of z on M_j^* is the origin so that $z \in Z_j$. A similar argument shows that

$$\tilde{p}_1(x) - p_1(x) = P_{N(A)}(z). \tag{A.17}$$

Using (A.16) and (A.17) in (A.12) we obtain

$$|x - \tilde{p}_1(x)|^2 \geq \eta_j |x - p_1(x)|^2$$

and (A.2) follows from the definition (A.15) of η_1.

References

[1] H.Z. Aashtiani, "The multi-model assignment problem", Ph.D. Thesis, Sloan School of Management, Massachusetts Institute of Technology (May, 1979).

[2] A. Auslender, *Optimization. Méthodes numériques* (Mason, Paris, 1976).

[3] A.B. Bakushinskij and B.T. Poljak, "On the solution of variational inequalities", *Soviet Mathematics Doklady* 219 (1974) 1705–1710.

[4] D.P. Bertsekas, "Algorithms for nonlinear multicommodity network flow problems", in: A. Bensoussan and J.L. Lions, eds., *International Symposium on Systems Optimization and Analysis* (Springer, New York, 1979) pp. 210–224.

[5] D.P. Bertsekas, "A class of optimal routing algorithms for communication networks", Proceedings of 1980 ICCC, Atlanta, GA (1980) pp. 71–75.

[6] D.P. Bertsekas, E. Gafni and K.S. Vastola, "Validation of algorithms for routing of flow in networks", Proceedings of 1978 Conference on Decision and Control, San Diego, CA (1979) pp. 220–227.

[7] D.G. Cantor and M. Gerla, "Optimal routing in a packet switched computer network", *IEEE Transactions on Computers* C-23 (1974) 1062–1069.

[8] S. Dafermos, "Traffic equilibrium and variational inequalities", *Transportation Science* 14 (1980) 42–54.

[9] J.C. Dunn, "Global and asymptotic convergence rate estimates for a class of projected gradient processes", *SIAM Journal on Control and Optimization* 19 (1981) 368–400.

[10] J.C. Dunn, "Newton's method and the Goldstein step length rule for constrained minimization problems", *SIAM Journal on Control and Optimization* 18 (1980) 659–674.

[11] L. Fratta, M. Gerla, and L. Kleinrock, "The flow deviation method: An approach to store- and forward communication network design", *Networks* 3 (1973) 97–133.

[12] E.M. Gafni, "Convergence of a routing algorithm", Laboratory for Information and Decision Systems Report 907, Massachussetts Institute of Technology, Cambridge, MA (May 1979).

[13] R.G. Gallager, "A minimum delay routing algorithm using distributed computation", *IEEE Transactions on Communication* COM-25 (1977) 73–85.

[14] E.S. Levitin and B.T. Poljak, "Constrained minimization methods", *U.S.S.R. Computational Mathematics and Mathematical Physics* 6 (1966) 1–50.

[15] R.T. Rockafellar, *Convex analysis* (Princeton University Press, Princeton, 1970).

[16] M. Sibony, "Méthodes itératives pour les équations et inéquations aux dérivées partielles nonlinéares de type monotone", *Calcolo* 7 (1970) 65–183.