May 1979 LIDS-R-907

CONVERGENCE OF A ROUTING ALGORITHM

by

Eliezer M. Gafni

This research was supported in part in the Laboratory for Information and Decision Systems under grant NSF-ENG-79-06332.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

We model a computer communication network as a graph, where nodes correspond to computers and arcs correspond to communication links connecting them. The computers (nodes) are engaged in sending messages to each other. A message goes from its origin to its destination via intermediate nodes. Every intermediate node must have a rule to decide upon the next node the message will be forwarded to, on its way to the destination. A set of such rules, for every node and destination will be called a routing strategy.

Due to physical limitations such as, finite link transmission capacity, processing time at each node, etc., a message arriving at an intermediate node is queued up with other messages waiting for transmission on a link leading to its next intermediate node. The time a message spends in the network, which is the time elapsed since it entered the network at its origin until it is received at its destination, is the delay the message undergoes. The problem of finding a routing strategy to minimize average delay per message, given the network parameters and the network input statistics, is called the routing problem. When the network input statistics are fixed in time, the problem is called the static routing problem. The static routing problem can be formulated, under various assumptions, as a nonlinear multicommodity flow problem.

In [1], Gallager proposed a distributed algorithm to solve the nonlinear multicommodity flow problem. The algorithm is distributed in the sense that computation is divided equally among all nodes in the network, and every node computes the value of its own routing variables. In view of the fact that the algorithm can be operated "on-line", it is also suitable for solving the quasi-static routing problem, in which the network input statistics change "slowly" in time, thereby causing the minimizing routing strategy to vary. Generally, one can assume that if the algorithm converges fast enough, it will be able to "track" the statistical variation of the process, and keep the routing close to optimal at all times.

Gallager's algorithm uses only first derivative information, and in a distributed operation context cannot utilize a line search of the type common in nonlinear programming algorithms. As a result, it suffers from the usual drawback of most optimization algorithms without a line search; namely, convergence is guaranteed only for a sufficiently small, but otherwise unknown, stepsize. Computational results for Gallager's algorithm reported by Poulos [5] suggest that the range of appropriate stepsizes varies significantly with the traffic input data, and indicate slow convergence. Gallager gives a convergence proof that relies heavily on an anti-zigzaging device which is embedded in the definition of blocking [[1], equ.(15)] and is an integral part of the algorithm.

In [2] Bertsekas, following [1], proposes a class of algorithms which can utilize second derivatives. This class of algorithms is motivated by the Goldstein-Levitin-Polyak method [3],[4] from nonlinear programming. It is shown there that Gallager's algorithm belongs to this class. Bertsekas suggests a member of this class which can be implemented in a distributed manner.

Incorporation of second derivatives may result in automatic stepsize scaling, good direction choice, and superlinear convergence rate. Yet, the particular algorithm, suggested by Bertsekas, neglects cross term second derivatives, and therefore is not guaranteed to have these nice features. Nevertheless computational results reported in Bertsekas, Gafni and Vastola [6] are encouraging. An appropriate stepsize was found, under which rather fast convergence was achieved for most of the networks, with various levels of traffic input data, that were tested. Pathological cases, where Bertsekas' algorithm behaves poorly still exist.

Bertsekas proves convergence of his general class of algorithms, but under the restrictive assumption that all network inputs are positive. The class of algorithms of [2] can be operated either in a "one-at-a-time" mode, whereby the routing strategies for various destinations are updated cyclically one after the other; or in an "all-at-once" mode where the routing strategies for all destinations are updated simultaneously. The convergence proof of [2] addresses only the case of the "all-at-once" mode.

This work can be considered a sequel to [1] and [2]. In the second chapter we familiarize the reader with the content of [1] and [2]. In the third chapter we give a convergence proof for the general class of algorithms of [2] which bypasses the need for the restrictive assumption on the network input mentioned earlier. It also bypasses the need for the antizigzaging device used in [1]. This is significant in that extension of the antizigzaging device of [1] within the context of the class of algorithms of [2] introduces complications and requires additional computation overhead. Simultaneously, convergence is proved for the "one-at-a-time" and the "all-at-once" modes mentioned above.

CHAPTER 2

BACKGROUND

2.1 Problem Formulation:

A common statement of the static routing problem as a nonlinear multicommodity flow problem is the following:

We are given a network consisting of N nodes denoted by $1,2,\ldots, N$ and L directed links. We denote (i,ℓ), the link from node i to node ℓ. We assume that the network is connected in the sense that for any two nodes m,n there is a directed path from m to n. The set of links is also denoted by L. Consider the following problem:

$$\text{(MFP)} \qquad \text{minimize} \sum_{(i,\ell)\in L} D_{i\ell}\left[\sum_{j=1}^{N} f_{i\ell}(j)\right]$$

subject to:

$$\sum_{\ell\in O(i)} f_{i\ell}(j) - \sum_{m\in I(i)} f_{mi}(j) = r_i(j) \quad \forall\, i = 1,\ldots, N \;\; j = 1,\ldots,N \;\; i \neq j$$

$$f_{i}(j) \geq 0 \qquad \forall (i,\ell) \in L \qquad j = 1,\ldots, N$$

$$f_{j\ell}(j) = 0 \qquad \forall\, (j,\ell) \in L \qquad j = 1,\ldots, N$$

where $f_{i\ell}(j)$ is the flow in link (i,ℓ) destined for node j, $O(i)$ and $I(i)$ are sets of nodes ℓ for which $(i,\ell) \in L$ and $(\ell,i) \in L$, respectively, and for $i \neq j$ $r_i(j)$ is a known traffic input at node i destined for j. The standing assumptions throughout are:

a) $r_i(j) \geq 0 \quad \forall\, i = 1,\ldots, N \quad j = 1,\ldots, N \quad i \neq j$

b) The functions $D_{i\ell}$ are defined on some domain $[0, c_{i\ell})$, where $c_{i\ell}$ is either a positive scalar or $+\infty$. Furthermore, $D_{i\ell}$ is convex, everywhere continuous, twice continuously differentiable with positive first and second derivative in the relative interior of its domain, and satisfies

b(1) $D'_{i\ell}(0) > 0 \quad \forall\ (i,\ell) \in L$ where prime denotes first derivative.

b(2) $\lim_{f_{i\ell} \to c_{i\ell}} D_{i\ell}(f_{i\ell}) = +\infty$

The formulation above, which will be called the <u>f-space formulation</u>, is not suitable for distributed algorithmic operation, since the control variables $f_{i\ell}(j)$, $(i\ell) \in L$ $j = 1,\ldots, N$ are not determined by a single node, but rather are the result of cumulative decisions of several nodes. An equivalent formulation where every control variable is determined by a single node, uses the transformation:

$$\phi_{ik}(j) = \frac{f_{ik}(j)}{\sum_{\ell \in O(i)} f_{i\ell}(j)} \qquad \forall\ i,j \text{ such that } \sum_{\ell \in O(i)} f_{i\ell}(j) > 0 \tag{2.1}$$

Denote by ϕ the vector $[\phi_{11}(1), \phi_{12}(1),\ldots, \phi_{1N}(1), \phi_{21}(1),\ldots, \phi_{NN}(N)]^T$ where T denotes transposition. Let Φ be the set of vectors ϕ whose entries $\phi_{ik}(j)$ $1 \le i,k,j \le N$ satisfy the following conditions:

a) $\phi_{ik}(j) \geq 0, \quad \phi_{ik}(j) = 0 \quad \forall\ (i,k) \notin L$ or if $i = j$

b) $\sum_{k=1}^{N} \phi_{ik}(j) = 1 \quad \forall\ j = 1,\ldots, N \quad i = 1,\ldots, N \quad i \neq j$

c) For each j, there is a subset of L, consisting of links (i,ℓ) such that $\phi_{i\ell}(j) > 0$, that forms a directed network in which every node has a directed path to j.

It is shown in [1] that for every $\phi \in \Phi$ the set of equations:

$$t_i(j) = r_i(j) + \sum_{\ell=1}^{N} t_\ell(j)\phi_{\ell i}(j) \quad \forall\ i = 1,\ldots, N \quad j = 1,\ldots, N \quad i \neq j \tag{2.2}$$

has a unique solution for $t_i(j)$ which is nonnegative continuously differentiable as a function of $r_i(j)$, $\phi_{ik}(j)$ $1 \le i, k, j \le N$, and satisfies:

$$t_i(j) = \sum_{\ell=1}^{N} \frac{\partial t_i(j)}{\partial r_\ell(j)} r_\ell(j) \qquad i = 1,\ldots N \quad j = 1,\ldots, N \quad i \neq j \tag{2.3}$$

where $\frac{\partial t_i(j)}{\partial r_\ell(j)}$ depends only on $\phi_{ik}(j)$ $1 \le i, \ \le N$ (i.e., for fixed $\phi \in \Phi$, $t_i(j)$ is a linear function of $r_\ell(j)$ $1 \le \ell \le N$).

Moreover, every $f_{ik}(j)$ $1 \le i, j, k \le N$ which is feasible for the (MFP) satisfies

$$f_{ik}(j) = t_i(j)(\phi, r)\phi_{ik}(j) \qquad 1 \le i, j, k \le N \tag{2.4}$$

for some (possibly many) $\phi \in \Phi$ where $r_i = [r_i(1), r_i(2),\ldots, r_i(N)]^T$ $r = [r_1^T, r_2^T, \ldots, r_N^T]^T$ and $t_i(j)(\phi, r)$ $1 \le i, j \le N$ solve (2.2) for the corresponding ϕ and r. Conversely, for every $\phi \in \Phi$ the link flows determined by (2.4) are feasible for the (MFP).

By virtue of the discussion above, (MFP) can be reformulated as

$$\text{(MFP)*} \quad \text{minimize } D_r(\phi, r) = \sum_{(i,\ell) \in L} D_{i\ell}\left[\sum_{i=j}^{N} f_{i\ell}(j)(\phi, r)\right]$$

subject to $\phi \in \Phi$

and this version of the problem will be called the <u>ϕ - space formulation</u>.

2.2 Optimality Conditions

Let $f_{ik}(\phi,r) = \sum_{j=1}^{N} f_{ik}(j)(\phi,r)$. We have

$$(2.4)\qquad \frac{\partial D_T(\phi,r)}{\partial r_i(j)} = \sum_{(i,k)\in L} D'_{ik}(f_{ik}(\phi,r)) \cdot \frac{\partial f_{ik}(\phi,r)}{\partial r_i(j)} \quad 1 \le i,\ j \le N,\ i \ne j$$

$$(2.5)\qquad \frac{\partial D_T(\phi,r)}{\partial \phi_{ik}(j)} = \sum_{(i,k)\in L} D'_{ik}(f_{ik}(\phi,r)) \frac{\partial f_{ik}(\phi,r)}{\partial \phi_{ik}(j)} \quad 1 \le i,k,\ j \le N\ .\ i \ne j$$

It it shown in [1] that

$$(2.6)\qquad \frac{\partial D_T(\phi,r)}{\partial r_i(j)} = \sum_k \phi_{ik}(j) \left[D'_{ik}(f_{ik}(\phi,r)) + \frac{\partial D_T(\phi,r)}{\partial r_k(j)}\right] \quad 1 \le i,\ j \le N,\ i \ne j$$

$$(2.7)\qquad \frac{\partial D_T(\phi,r)}{\partial \phi_{ik}(j)} = t_i(j)(\phi,r)\left[D'_{ik}(f_{ik}(\phi,r)) + \frac{\partial D_T(\phi,r)}{\partial r_k(j)}\right] \quad 1 \le i,k,\ j \le N,\ i \ne j$$

Moreover, (2.4) is the unique solution of (2.6) and both (2.4) and (2.5) are continuously differentiable in ϕ and r. We will use the notation $\delta_{ik}(\phi,r) = D'_{ik}(f_{ik}(\phi,r)) + \frac{\partial D_T(\phi,r)}{\partial r_k(j)}$. The quantity $\frac{\partial D_T(\phi,r)}{\partial r_i(j)}$ will be called the "node i marginal delay with respect to destination j"; and the quantity $\delta_{ik}(j)(\phi,r)$ will be called the "link (i,k) marginal delay with respect to destination j". Equation (2.6) asserts that a node marginal delay is a convex sum of the marginal delays of its outgoing links with respect to the same destination.

It is shown in [1] that a sufficient condition for optimality of a routing $\hat{\phi}$ is:

$$
(2.8)\qquad
\begin{aligned}
&\frac{\partial D_T(\hat{\phi},r)}{\partial r_i(j)} = \delta_{ik}(j)(\hat{\phi},r) \text{ if } \phi_{ik}(j) > 0 \\
&\frac{\partial D_T(\hat{\phi},r)}{\partial r_i(j)} \leq \delta_{ik}(j)(\hat{\phi},r) \text{ if } \phi_{ik}(j) = 0
\end{aligned}
\qquad
\begin{aligned}
i &= 1,\ldots, N \\
j &= 1,\ldots, N
\end{aligned}
$$

namely, that the marginal delay of every node for any given destination attains the minimum over all possible convex sums of the marginal delays, of its outgoing links, with respect to the same destination.

2.3 A Class of Algorithms:

The class of algorithms suggested in [2] is defined by:

$$\phi_i^{k+1}(j) = \overline{\phi}_i^k(j) \qquad i \neq j$$

where $\overline{\phi}_i^k(j)$ is the solution to the problem

$$
(2.9)\qquad
\begin{aligned}
\text{minimize}\quad & \delta_i(j)(\phi^k,r)^T(\phi_i(j) - \phi_i^k(j)) \\
& + \frac{1}{2} \cdot \frac{t_i(j)(\phi^k,r)}{\alpha} (\phi_i(j) - \phi_i^k(j))^T M_i^k(j)(\phi_i(j)-\phi_i^k(j))
\end{aligned}
$$

$$\text{subject to } \phi_i(j) \geq 0 \sum_{\ell} \phi_{i\ell}(j) = 1,\ \phi_{i\ell}(j) = 0 \ \forall\ \ell \in B(i, \phi^k)(j)$$

where

$$\delta_i(j)(\phi,r)^T = [\delta_{i1}(j)\ (\phi,r), \delta_{i2}(j)(\phi,r),\ldots,\delta_{iN}(j)(\phi,r)]$$

$$\phi_i(j) = [\phi_{i1}(j),\ \phi_{i2}(j),\ldots,\ \phi_{iN}(j)]^T$$

and $\alpha > 0$ is some scalar. The matrix $M_i^k(j)$ is symmetric and satisfies

(2.10) $$M_i^k(j) = \{m_{p\ell}^k\}, \quad |m_{p\ell}^k| \le \Lambda \quad 1 \le p,\ell \le N$$

$$\lambda \|v_i(j)\|^2 \le v_i^T(j) M_i(j) v_i(j)$$

for all $v_i(j)$ in the subspace $\{v_i(j) \mid \sum_{\ell \in B(i,\phi^k)(j)} v_{i\ell}(j) = 0, \quad v_{i\ell}(j) = 0$ $\forall\ \ell \in B(i,\phi^k)(j)\}$, where Λ, λ are some positive scalars and $\| \ \|$ represents the Eucledian norm. The set of indices $B(i,\phi^k)(j)$ is specified by the following definitions:

Definition 2.1: A node n is said to be downstream to m with respect to j, if there exists a path $(m,k)\ (k,\ell) \ldots (s,n)$ such that $\phi_{mk}(j) > 0$, $\phi_{k\ell}(j) > 0, \ldots, \phi_{sn}(j) > 0$.

Definition 2.2: A node m is said to be upstream to n with respect to j, if n is downstream to m with respect to j.

Definition 2.3: An element $\phi \in \Phi$ is said to be loop free with respect to j if there do not exist two nodes m,n which are downstream to each other with respect to j.

Definition 2.4: For any $\phi \in \Phi$, $i = 1,\ldots, N$ and $j = 1,\ldots, N$, the set $B(i,\phi)(j)$ referred to as the set of blocked nodes for ϕ at i with respect to j, is the set of all ℓ such that $(i,\ell) \notin L$ or $(i,\ell) \in L$ $\phi_{i\ell}(j) = 0$ and either $\frac{\partial D_T(\phi,r)}{\partial r_i(j)} \le \frac{\partial D_T(\phi,r)}{\partial r_\ell(j)}$, or there exists a link (m,n) referred to as an improper link such that $m = \ell$ or m is downstream to ℓ with respect to j and we have $\phi_{mn}(j) > 0$, and $\frac{\partial D_T(\phi,r)}{\partial r_m(j)} \leq \frac{\partial D_T(\phi,r)}{\partial r_n(j)}$.

Consider the mappings:

$$A_i(j): \phi \in R^{NxNxN} \rightarrow \phi_i(j) \in R^N \qquad i = 1,\dots,N \; j = 1,\dots,N \qquad i \neq j$$

$$A_i(i): \phi \in R^{NxNxN} \rightarrow 0 \in R^N \qquad i = 1,\dots,N$$

defined by $A_i(j)(\phi^k) = \overline{\phi}_i^k(j)$ $A_i(i)(\phi^k) = 0$. The algorithm which generates ϕ^{k+1} given ϕ^k can be implemented in two modes. The "all-at-once" mode of implementation of the algorithm corresponds to:

$$\phi^{k+1} = [\phi^{k+1^T}(1), \; \phi^{k+1^T}(2),\dots,\phi^{k+1^T}(N)]^T,$$

$$\phi^{k+1}(j) = [A_1(j)^T(\phi_1^k(j)), \; A_2^T(j)(\phi_2^k(j)),\dots,A_N^T(j)(\phi_N^k(j))]^T$$

and will be denoted as $\phi^{k+1} = A(\phi^k)$; while the "one-at-a-time" mode corresponds to

$$\phi^{k+1} = [\phi^{k^T}(1),\dots, \phi^{k^T}(m-1), \; \phi^{k+1^T}(m), \; \phi^{k^T}(m+1),\dots, \phi^{k^T}(N)]^T$$

$$\phi^k(j) = [\phi_1^T(j), \; \phi_2^T(j),\dots, \phi_N^T(j)]^T,$$

$\phi^{k+1}(m)$ is defined above and $m = (k \bmod N) + 1$.

The sets $B(i,\phi^k)(j)$ $1 \leq i, j \leq N$ are introduced in order to ensure that the sequence $\{\phi^k\}$ generated by the algorithm consists of loopfree elements. This in turn is essential in order to make (2.9) suitable for distributed implementation, and to facilitate the computation of the quantities needed in it. The distributed implementation of the algorithm is essentially the same as the one of Gallager and the reader is referred to [1] for details. The presence of the sets $B(i,\phi^k)(j)$, however, introduces difficulties in proving convergence. Gallager defines $B(i,\phi^k)(j)$ in a somewhat different manner, introducing what can be considered as an antizigzaging device. Generalization of this device to (2.9) is possible but at the expense of making the distributed implementation of (2.9) cumbersome and computationally more expensive.

It is shown in [2, proposition 1] that if $\phi^k \in \Phi$ and ϕ^k is loop free, then ϕ^{k+1} is loopfree and ϕ^k is optimal if and only if $\phi^k = \phi^{k+1} = \ldots = \phi^{k+N}$.

2.4 Gallager's Algorithm (without an antizigzaging device):

Gallager's algorithm corresponds to a diagonal $M_i^k(j)$, where the elements on the diagonal are 1 except for one element $m_{n,n}$ which is zero, and n satisfies

$$\delta_{in}(j)(\phi^k,r) \leq \delta_{ik}(j)(\phi^k,r) \ \forall\ k, \quad k \notin B(i,\phi^k)(j).$$

The last equation makes $M_i^k(j)$ discontinuous in ϕ, which inturn makes the algorithm discontinuous in ϕ. The discontinuity of the algorithm is a major difficulty in proving convergence. It is trivial to check that (2.9) with the above substitution for $M_i^k(j)$ satisfies equation (12) of [1]. An "Almost" fixed $M_i^k(j)$ $1 \leq i,\ j \leq N$, $k = 1,2,\ldots$ is very attractive since it reduces the information-transfer and computation needed for distributed implementation of (2.9), yet is makes the algorithm somewhat inefficient.

2.5 Bertsekas' Algorithm:

The particular algorithm proposed by Bertsekas corresponds to a diagonal $M_i^k(j)$ where the element $m_{\ell\ell}$ is an upper bound to

$D''_{i\ell}(f_{i\ell}(\phi^k,r)) + \dfrac{\partial D^2_T(\phi^k,r)}{\partial r_\ell(j)}$, which can be computed in a manner which is consistent with the distributed implementation suggested by Gallager. Bertsekas' algorithm is continuous in ϕ.

CHAPTER 3

CONVERGENCE PROOF

3.1 Introduction

The proof consists of three parts. The first establishes the descent properties of the algorithm. The second examines some aspects of the limiting behaviour of the algorithm as the number of iterations tends to infinity. In the third part all the properties of the algorithm, shown in the second part are combined together to form the convergence proof. The proof is given for the "all-at-once" mode. The proof of the "one-at-a-time" mode involves minor modifications and these are sketched subsequently.

Before we proceed, we change slightly the form of the sufficient conditions for optimality:

Lemma 1: For $\phi \in \Phi$, let $S_i(\phi,r)(j)$ be the length of a shortest path from i to destination j, where the length of each link $(i,\ell) \in L$ is taken to be $D'_{i\ell}(f_{i\ell}(\phi,r))$. A sufficient condition for optimality of ϕ is

$$(3.1) \qquad \frac{\partial D_T(\phi,r)}{\partial r_i(j)} = S_i(\phi,r)(j) \quad \forall\ i = 1,\ldots,N \quad j = 1,\ldots,N$$

Proof: We just have to show that (3.1) implies (2.8). Substituting (3.1) in (2.6) we get

$$(3.2) \qquad S_i(\phi,r)(j) = \sum_k \phi_{ik}(j)[D'_{ik}(f_{ik}(\phi,r)) + S_k(\phi,r)(j)] \ .$$

On the other hand by definition of shortest path and Bellman's equation:

$$(3.3) \qquad S_i(\phi,r)(j) = \min_k \{D'_{ik}(f_{ik}(\phi,r)) + S_k(\phi,r)(j)\} .$$

Equations (3.2), (3.3) and the conditions of ϕ imply that

$$(3.4) \qquad \begin{aligned} S_i(\phi,r)(j) &= D'_{ik}(f_{ik}(\phi,r)) + S_k(\phi,r)(j) \quad \forall\, k, \quad \phi_{ik}(j) > 0 \\ S_i(\phi,r)(j) &\leq D'_{ik}(f_{ik}(\phi,r)) + S_k(\phi,r)(j) \quad \forall\, k, \quad \phi_{ik}(j) = 0. \end{aligned}$$

Substituting (3.1), again, in (3.4) we get (2.8). Q.E.D.

<u>Lemma 2</u>: Let $S_i(\phi,r)(j)$ be defined as in lemma 1. There holds

$$(3.5) \qquad \frac{\partial D_T(\phi,r)}{\partial r_i(j)} \geq S_i(\phi,r)(j) \quad \forall\, \phi \in \Phi,\ r,\ i = 1,\ldots,N \quad j = 1,\ldots, N .$$

Although, lemma 2 is true as stated, we will prove it only for $\phi \in \hat{\Phi}$, where $\hat{\Phi} = \{\phi | \phi \in \Phi,\ \phi \text{ is loopfree}\}$.

<u>Proof</u>: Fix the destination j. For every $\phi \in \hat{\Phi}$ the directed network induced by $\phi_{ik}(j) > 0$ $1 \leq i, k \leq N$ (i.e., the network obtained from the original by eliminating all links (i,k) with $\phi_{ik}(j) = 0$), is an acyclic directed network and every node has a directed path to the destination. In reference to this network define

$$(3.6) \qquad S_k = \{i \in N | i \text{ is at most } k \text{ hops from the destination}\} .$$

Since $\phi \in \hat{\Phi} \subset \Phi$ we have that

(3.7) $\quad S_0 = \{j\}$

and (3.5) holds for $i = j$ with equality.

We now prove that if (3.5) holds for all $i \in S_\ell$ then it holds for all $i \in S_{\ell+1}$.

Let $i \in S_{\ell+1}|S_\ell$ where $|$ denotes set subtraction. We have

$$
\begin{aligned}
(3.8) \qquad \frac{\partial D_T(\phi,r)}{\partial r_i(j)} &= \sum_{k=1}^{N} \phi_{ik}(j)\left[D'_{ik}(f_{ik}(\phi,r)) + \frac{\partial D_T(\phi,r)}{\partial r_k(j)}\right] = \\
&= \sum_{k\in\{1,\dots,N\}\cap S_\ell} \phi_{ik}(j)\left[D'_{ik}(f_{ik}(\phi,r)) + \frac{\partial D_T(\phi,r)}{\partial r_k(j)}\right] \\
&\geq \sum_{k\in\{1,\dots,N\}\cap S_\ell} \phi_{ik}(j)\,[D'_{ik}(f_{ik}(\phi,r)) + s_k(\phi,r)(j)] \\
&\geq \min_{\substack{k \\ \phi_{ik}(j)>0}} \{D'_{ik}(f_{ik}(\phi,r))) + s_k(\phi,r)(j)\} = s_i(\phi,r)(j) .
\end{aligned}
$$

The second equality follows from the fact that $\phi \in \hat{\Phi}$, and the third relation follows from our hypothesis.

Since $\phi \in \hat{\hat{\Phi}}$, $S_{N-1} = \{1,\dots,N\}$ and $S_k|S_{k-1} = \phi$ for all $k \geq N$, the induction will be complete after N steps. Relations (3.7) and (3.8) establish (3.5).

Q.E.D.

3.2 Descent Properties of the Algorithm:

We establish the descent property of the algorithm, by upper bounding the objective function along a straight line connecting the current point with the next. In our case one can either consider a straight line in the f-space, which is natural to the f-space formulation or a straight line in the ϕ-space, which is natural to the ϕ-space formulation. A straight line in one formulation maps into a curve in the other. The nature of the algorithm [cf. (2.9)] calls for the ϕ-space formulation. Unfortunately we were not able to obtain the needed result using the ϕ-formulation. Gallager uses the f-space formulation to appropriately upper bound the objective function ([1] appendix C, lemmas 1-4). We use a similar argument to prove the result we need for our case.

We state the desired proposition and present the proof in Appendix A.

Proposition 1: Let $\phi^{k+1} = A(\phi^k)$ and let $D_0 \in (0,\infty)$ be some real number, and assume that

$$\phi^0 \in \hat{\Phi}, \quad D(\phi^0,r) \leq D_0 .$$

Then there exists a scalar $\bar{\alpha} \in (0,1]$ (depending on N, D_0, λ and Λ) such that for every $\alpha \in (0,\alpha]$ used in (2.9) we have

$$D_T(\phi^{k+1},r) - D_r(\phi^k,r) \leq -p \sum_{i,j} t_i(j)(\phi^k,r)^2 \cdot \|\phi_i^{k+1}(j) - \phi_i^k(j)\|^2 \tag{3.9}$$

where p is some positive constant which depends on α, D_0, λ, Λ and N.

3.3 The Limiting Behaviour of the Algorithm:

If we use a value of α such that (3.9) holds, the sequence $\{D_T(A^k(\phi^0),r)\}_{k=1}^{\infty}$, where $A^k(\phi) = A^{k-1}(A(\phi))$, is a nonincreasing sequence.

Our assumption on the functions $D_{i\ell}$ implies that D_T is bounded from below. As a result we have

$$(3.10) \qquad \lim_{k\to\infty} t_i(j)(\phi^k,r)^2 \cdot \|\phi_i^{k+1}(j) - \phi_i^k(j)\|^2 = 0 \qquad \forall\ i = 1,\dots, N \quad j = 1,\dots, N\ i \neq j$$

<u>Lemma 3</u>: For all α such that (3.9) holds we have

$$(3.11) \qquad \lim_{K\to\infty} |f_{i\ell}(j)(A(\phi^k),r)) - f_{i\ell}(j)(\phi^k,r)| = 0 \quad \forall\ (i,\ell) \in L \quad j = 1,\dots,N$$

<u>Proof</u>: Subtracting equation (2.2) that corresponds to $r_i(j)$ and ϕ^k from the one which corresponds to $r_i(j)$ and ϕ^{k+1} one gets (see [1] (c23)):

$$t_i(j)(\phi^{k+1},r) - t_i(j)(\phi^k,r) = \sum_\ell [t_\ell(j)(\phi^{k+1},r) - t_\ell(j)(\phi^k,r)] \cdot \phi_{\ell i}^{k+1}$$

$$+ \sum_\ell t_\ell(j)(\phi^k,r)(\phi_{\ell i}^{k+1}(j) - \phi_{\ell i}^k(j)) \ .$$

The equation above is in the standard form of (2.2), and using (2.3) one gets

$$(3.12) \qquad t_i(j)(\phi^{k+1},r) - t_i(j)(\phi^k,r) = \sum_\ell \frac{\partial t_i(j)(\phi^{k+1},r)}{\partial r_\ell(j)} \sum_n t_n(j)(\phi_{n\ell}^{k+1}(j) - \phi_{n\ell}^k(j)) \ .$$

As mentioned earlier, since $\phi^0 \in \hat{\Phi}$ we have that $\phi^k \in \hat{\Phi}$ $k = 0,1,\dots$ and therefore $0 \le \dfrac{\partial t_i(j)(\phi^{k+1},r)}{\partial r_\ell(j)} \le 1$. Moreover, according to the definition

Taking the limit as $k \to \infty$ on both sides of (3.15) and using (3.10) we get (3.11).

Q.E.D.

Lemma 4: For all α such that (3.9) holds we have

$$\lim_{K\to\infty} [\bar{\Delta}_i^k(j) - \underline{\Delta}_i^k(j)] = 0 \qquad i = 1,\ldots, N \quad j = 1,\ldots, N \quad i \neq j \tag{3.16}$$

where

$$\bar{\Delta}_i^k(j) = \max_{\ell} \{\delta_{i\ell}(j)(\phi^k,r) \mid \ell \in O(i), \ \ \phi_{i\ell}^{k+1} > 0\}$$

$$\underline{\Delta}_i^k(j) = \min_{\ell} \{\delta_{i\ell}(j)(\phi^k,r) \mid \ell \in O(i), \ \ \ell \notin B(i,\phi^k)(j)\}$$

Proof: We have that $\phi_i^{k+1}(j)$ solves problem (2.9). By the necessary condition for optimality for (2.9) we obtain

$$[\delta_i(j)(\phi^k,r) + \frac{t_i(j)(\phi^k,r)}{\alpha} \cdot M_i^k(j)(\phi_i^{k+1}(j) - \phi_i^k(j))]^T (\phi_i(j) - \phi_i^{k+1}(j)) \geq 0 \tag{3.17}$$

for all $\phi_i(j)$ which are feasible for problem (2.9).

Let $\bar{\ell}$ and $\underline{\ell}$ be such that

$$\delta_{i\bar{\ell}}(\phi^k,r)(j) = \bar{\Delta}_i^k(j)$$

$$\delta_{i\underline{\ell}}(\phi^k,r)(j) = \underline{\Delta}_i^k(j) \ .$$

If $\bar{\ell} = \underline{\ell}$ we have $\bar{\Delta}_i^k(j) = \underline{\Delta}_i^k(j)$. Assume $\bar{\ell} \neq \underline{\ell}$, we define $\phi_i^*(j)$ to be

$$\phi^*_{i\ell}(j) = \begin{cases} \phi^{k+1}_{i\ell}(j) - \varepsilon & \text{if } \ell = \overline{\ell} \\ \phi^{k+1}_{i\ell}(j) + \varepsilon & \text{if } \ell = \underline{\ell} \\ \phi^{k+1}_{i\ell}(j) & \text{otherwise} \end{cases} .$$

where $\varepsilon > 0$ is small enough so that $\phi^{k+1}_{i\overline{\ell}}(j) - \varepsilon > 0$. By definition of $\overline{\Delta}^k_i$ (j) such an ε exists and by feasibility of $\phi^{k+1}_i(j)$ we have that $\phi^*_i(j)$ is also feasible.

Plugging $\phi^*_i(j)$ in (3.17) in place of $\phi_i(j)$ we get

$$(3.18) \qquad (\overline{\Delta}^k_k(j) - \underline{\Delta}^k_i(j))\varepsilon \leq \frac{1}{\alpha}(M^k_{i\underline{\ell}}(j) - M^k_{i\overline{\ell}}(j))\varepsilon$$

where $M^k_{i\underline{\ell}}(j)$ and $M^k_{i\overline{\ell}}(j)$ are the $\underline{\ell}$ and $\overline{\ell}$ elements of the vector $M^k_i(j)t_i(j)(\phi^k,r) \cdot (\phi^{k+1}_i(j) - \phi^k_i(j))$. All elements of $M^k_i(j)$ are uniformly bounded from above by Λ and by (3.10) every element of $t_i(j)(\phi^k,r)(\phi^{k+1}_i(j) - \phi^k_i(j))$ goes to zero as k tends to infinity. Therefore,

$$0 \leq \overline{\Delta}^k_i(j) - \underline{\Delta}^k_i(j) \leq \frac{2\Lambda}{\alpha} t_i(j)(\phi^k,r)\sum_{\ell}|\phi^{k+1}_{i\ell}(j) - \phi^k_{i\ell}(j)|,$$

and this equation also holds if $\underline{\ell} = \overline{\ell}$. Equation (3.10) implies that the R.H.S. tends to zero as k tends to infinity and therefore

$$(3.19) \qquad \lim_{k\to\infty} [\overline{\Delta}^k_i(j) - \underline{\Delta}^k_i(j)] = 0$$

Q.E.D.

Lemma 5: Let α be such that (3.9) holds. Then there exists a set of indices $\tilde{K}$ such that the subsequences $\{\phi^\ell\}_{\ell\in\tilde{K}} = \{A^\ell(\phi^0)\}_{\ell\in\tilde{K}}$ and of $B(i,\phi^k)(j)$ the directed network induced by $\phi^k_{i\ell}(j) > 0$ and $\phi^{k+1}_{i\ell}(j) > 0$ $\forall (i,\ell) \in L$ is loopfree and therefore

$$\frac{\partial t_i(j)(\phi^{k+1},r)}{\partial r_\ell(j)} \cdot (\phi^{k+1}_{i\ell}(j) - \phi^k_{i\ell}(j)) = 0$$

Upper bounding (3.12) we get

$$(3.13) \qquad |t_i(j)(\phi^{k+1},r) - t_i(j)(\phi^k,r)| \le \sum_{n\neq i} t_n(j)(\phi^k,r)[\sum_\ell |\phi^{k+1}_{n\ell}(j) - \phi^k_{n\ell}(j)|] .$$

Writing down the expression for $|f_{i\ell}(j)(\phi^{k+1},r) - f_{i\ell}(j)(\phi^k,r)|$ one gets

$$(3.14) \qquad |f_{i\ell}(j)(\phi^{k+1},r) - f_{i\ell}(j)(\phi^k,r)| = |t_i(j)(\phi^{k+1},r)\phi^{k+1}_{i\ell}(j) - t_i(j)(\phi^k,r)\phi^k_{i\ell}(j)|$$

$$= |(t_i(j)(\phi^{k+1},r) - t_i(j)(\phi^k,r))\phi^{k+1}_{i\ell}(j) + t_i(j)(\phi^k,r)(\phi^{k+1}_{i\ell}(j) - \phi^k_{i\ell}(j))|$$

$$\le |t_i(j)(\phi^{k+1},r) - t_i(j)(\phi^k,r)| + |t_i(j)(\phi^k,r)(\phi^{k+1}_{i\ell}(j) - \phi^k_{i\ell}(j)|$$

$$\le \sum_n t_n(j)(\phi^k,r) \cdot \sum_\ell |\phi^{k+1}_{n\ell}(j) - \phi^k_{n\ell}(j)|$$

where in the first inequality we used the triangle inequality and the fact $0 \le \phi_{i\ell}(j) \le 1$, and in the second we used (3.13) and the fact that $t_n(j) \ge 0$. Squaring (3.14) and using successively the Cauchy-Swartz inequality we get

$$(3.15) \qquad (f_{i\ell}(j)(\phi^{k+1},r) - f_{i\ell}(j)(\phi^k,r))^2 \le (\sum_n t_n(j)(\phi^k,r) \cdot \sum_\ell |\phi^{k+1}_{n\ell}(j) - \phi^k_{n\ell}(j)|)^2$$

$$\le (N-1) \cdot \sum_n t_n(j)(\phi^k,r)^2 (\sum_\ell |\phi^{k+1}_{n\ell}(j) - \phi^k_{n\ell}(j)|)^2$$

$$\le (N-1)^2 \sum_n t_n(j)(\phi^k,r)^2 \| \phi^{k+1}_n(j) - \phi^k_n(j)\|^2 .$$

and $\{\phi^{\ell-1}\}_{\ell \in \tilde{K}}$ coverge. Moreover we have that

$$f_{i\ell}(j)(\hat{\phi},r) = f_{i\ell}(j)(\tilde{\phi},r) \quad \forall\ (i,\ell) \in L \quad j = 1,\ldots, N \tag{3.20}$$

$$S_i(\hat{\phi},r)(j) = S_i(\tilde{\phi},r)(j) \quad i = 1,\ldots, N \quad j = 1,\ldots, N \tag{3.21}$$

where

$$\hat{\phi} = \lim_{\substack{\ell\to\infty \\ \ell\in\tilde{K}}} \phi^{\ell}, \qquad \tilde{\phi} = \lim_{\substack{\ell\to\infty \\ \ell\in\tilde{K}}} \phi^{\ell-1}$$

Proof: It is established ([1], [2]) that $\{\phi^k\}$ belongs to a compact set and therefore has a convergent subsequence $\{\phi^i\}_{i\in\hat{K}}$, where $\hat{K}$ is some set of indices. The subsequence $\{\phi^{i-1}\}_{i\in\hat{K}}$ also belongs to a compact set and has a convergent subsequence $\{\phi^{\ell-1}\}_{\ell\in\tilde{K}}$, where $\tilde{K}$ is some set of indices and $\tilde{K} \subset \hat{K}$. The sequence $\{\phi^{\ell}\}_{\ell\in\tilde{K}}$ converges since it is a subsequence of a convergent sequence. Equation (3.11) and the last observation implies (3.20). Equation (3.21) follows now since the shortest distance is flow dependent only.

Q.E.D.

Corollary: Let $\{\phi^{k-i}\}_{k\in\tilde{K}}$ and $\{\phi^k\}_{k\in\tilde{K}}$ be two convergent subsequences of $\{A^k(\phi^0)\}_{k=0}^{\infty}$ which converge to $\tilde{\phi}$ and $\hat{\phi}$ respectively, and i is some positive integer. Then equation (3.20) and (3.21) hold for the corresponding $\hat{\phi}$ and $\tilde{\phi}$.

Proof:

$$|f_{i\ell}(j)(\phi^k_{1r}) - f_{i\ell}(j)(\phi^{k-1},r)| \leq |f_{i\ell}(j)(\phi^k,r) - f_{i\ell}(j)(\phi^{k-1},r)| + \ldots$$
$$+ |f_{i\ell}(j)(\phi^{k-i+1},r) - f_{i\ell}(j)(\phi^{k-i},r)|$$

Take the limit on both sides of the equation above and use (3.11).

Q.E.D.

3.4 Convergence Proof:

We use "two dimensional induction" to show that the limit point of any convergent subsequence of $\{\phi^k\}_{k=0}^{\infty}$, satisfies the sufficient conditions for optimality (3.1). Lemma 6 that follows represents a step of the induction proof and is the cornerstone of the proof. We use repeatedly the fact that if some property 1 holds for all k, $k > k_1$ and some property 2 holds for all k, $k > k_2$ then both hold for all k, $k > \max(k_1, k_2)$. In what follows we will express this by writing "If 1 holds for all k large enough and 2 holds for all k large enough, then both hold for all k large enough".

Lemma 6: Let α be such that (3.9) holds and let $\{\phi^{k-1}\}_{k\in\tilde{K}} \to \tilde{\phi}$ and $\{\phi^k\}_{k\in\tilde{K}} \to \hat{\phi}$ be two convergent subsequences of $\{A^k(\phi^0)\}_{k=0}^{\infty}$.

Let $S(\tilde{\phi},r)(j) = \{S_i(\tilde{\phi},r)(j) \mid \forall i \in N\}$. Let $S_1(j),\ldots, S_p(j)$, $p \leq N$ be the distinct elements of the set $S(\tilde{\phi},r)(j)$ and assume without loss of generality that $0 = S_1(j) < S_2(j) <\ldots< S_p(j)$. Denote

$$I_m(j) = \{i \mid S_i(\tilde{\phi},r)(j) \leq S_m(j)\}, \quad m = 1,\ldots, p.$$

Assume that for some integer q we have:

a)

$$\frac{\partial D_T(\hat{\phi},r)}{\partial r_i(j)} = \frac{\partial D_T(\tilde{\phi},r)}{\partial r_i(j)} = S_i(\tilde{\phi},r)(j) \quad \forall\, i \in I_q(j), \quad j = 1,\ldots, N \tag{3.22}$$

b) For all k large enough, $k \in \tilde{K}$, and for any j, if $\phi_{mn}^{k-1}(j) > 0$ and $m \in I_q(j)$ then $\dfrac{\partial D_T(\phi^{k-1},r)}{\partial r_m(j)} > \dfrac{\partial D_T(\phi^{k-1},r)}{\partial r_n(j)}$. Then a')

$$\frac{\partial D_T(\hat{\phi},r)}{\partial r_i(j)} = S_i(\hat{\phi},r)(j) \quad \forall\, i \in I_{q+1}(j), \quad j = 1,\ldots,N \tag{3.23}$$

b') For all k large enough, $k \in \tilde{K}$, and for any j, if $\phi^k_{mn}(j) > 0$

and $m \in I_{q+1}(j)$ then $\dfrac{\partial D_T(\phi^k,r)}{\partial r_m(j)} > \dfrac{\partial D_T(\phi^k,r)}{\partial r_n(j)}$

Proof: Let $i \in I_{q+1}(j)|I_q(j)$, $O_i \triangleq \{\ell|(i,\ell) \in L\}$ and

$$\ell_i(j) = \{\ell|D'_{i\ell}(f_{i\ell}(\tilde{\phi},r)) + S_\ell(\tilde{\phi},r)(j) = S_i(\tilde{\phi},r)(j), \quad \ell \in O_i\} . \tag{3.24}$$

By the definition of shortest distance we have

$$S_i(\tilde{\phi},r)(j) < D'_{i\ell}(f_{i\ell}(\tilde{\phi},r)) + S_\ell(\tilde{\phi},r) \quad \forall\, \ell \in O_i|\ell_i(j) \quad j = 1,\ldots,N. \tag{3.25}$$

Using lemma 2

$$S_i(\tilde{\phi},r)(j) < D_{i\ell}(f_{i\ell}(\tilde{\phi},r)) + S_\ell(\tilde{\phi},r) \leq D'_{i\ell}(f_{i\ell}(\tilde{\phi},r)) + \frac{\partial D_T(\tilde{\phi},r)}{\partial r_\ell(j)} \tag{3.26}$$

$$= \delta_{i\ell}(j)(\tilde{\phi},r) \quad \forall\, \ell \in O_i|\ell_i(j) \quad m = 1,\ldots, N .$$

By the assumption $D'_{i\ell}(f) > 0\ \forall f$ and the fact $i \in I_{q+1}(j)|I_q(j)$, we have

$$\ell_i(j) \subset I_q(j) \qquad j = 1,\ldots,N . \tag{3.27}$$

Therefore by using hypothesis a) we have

$$\delta_{i\ell}(j)(\tilde{\phi},r) = D'_{i\ell}(f_{i\ell}(\tilde{\phi},r)) + \frac{\partial D_T(\tilde{\phi},r)}{\partial r_\ell(j)} = D'_{i\ell}(f_{i\ell}(\tilde{\phi},r)) + S_\ell(\tilde{\phi},r)(j) \tag{3.28}$$

$$= S_i(\tilde{\phi},r)(j) \quad \forall\, \ell \in \ell_i(j) \quad j = 1,\ldots, N .$$

The cardinality of O_i is less then N, i.e. is finite, and therefore there exists $\varepsilon > 0$ such that

$$\delta_{i\omega}(j)(\tilde{\phi},r) - \varepsilon > \delta_{i\ell}(j)(\tilde{\phi},r) \quad \forall\, \omega \in O_i|\ell_i(j),\ \ell \in \ell_i(j) \quad j = 1,\ldots,N \tag{3.29}$$

Since $S_{i\ell}(j)(\phi,r)$ is continuous in ϕ and $\{\phi^{k-1}\}_{k\in\tilde{K}}$ is a convergent sequence, we get that for all k large enough, $k \in \tilde{K}$

$$(3.30) \qquad \delta_{i\omega}(j)(\phi^{k-1},r) > \delta_{i\ell}(\phi^{k-1},r) + \frac{\varepsilon}{2} \quad \forall\, \omega \in O_i \,|\, \ell_i(j),\ \ell \in \ell_i(j),\ j=1,\ldots,N.$$

Also $\dfrac{\partial D_T(\phi,r)}{\partial r_i(j)}$, $1 \le i,j \le N$, is continuous in ϕ and therefore by lemma 2, (3.28) and hypothesis a), for all k large enough, $k \in \tilde{K}$

$$\frac{\partial D_T(\phi^{k-1},r)}{\partial r_i(j)} > \frac{\partial D_T(\phi^{k-1},r)}{\partial r_\ell(j)} \quad \forall\, \ell \in \ell_i(j), \quad j = 1,\ldots, N$$

which together with hypothesis b) and the definition of $B(\phi,i)(j)$ implies that for all k large enough, $k \in \tilde{K}$

$$(3.31) \qquad \ell_i(j) \cap B(\phi^{k-1},i)(j) = \emptyset \qquad j = 1,\ldots, N \quad .$$

Lemma 4 combined with (3.30) and (3.31) implies that for all k large enough $k \in \tilde{K}$

$$(3.32) \qquad \phi^k_{i\omega}(j) = 0 \qquad \forall\, \omega \in O_i \,|\, \ell_i(j), \qquad j = 1,\ldots,N$$

and taking the limit

$$(3.33) \qquad \hat{\phi}_{i\omega}(j) = 0 \qquad \forall\, \omega \in O_i \,|\, \ell_i(j), \qquad j = 1,\ldots,N \quad .$$

Using (3.33), lemma 5 and hypothesis a) we have

$$\frac{\partial D_T(\hat{\phi},r)}{\partial r_i(j)} = \sum_{\ell} \hat{\phi}_{i\ell}(j) \left[D'_{ik}(f_{ik}(\hat{\phi},r)) + \frac{\partial D_T(\hat{\phi},r)}{\partial r_\ell(j)} \right] =$$

$$= \sum_{\ell \in \ell_i(j)} \hat{\phi}_{i\ell}(j) \left[D'_{ik}(f_{ik}(\hat{\phi},r)) + \frac{\partial D_T(\hat{\phi},r)}{\partial r_\ell(j)} \right] =$$

$$= \sum_{\ell \in \ell_i(j)} \hat{\phi}_{i\ell}(j) \left[D'_{ik}(f_{ik}(\hat{\phi},r)) + \frac{\partial D_T(\tilde{\phi},r)}{\partial r_\ell(j)} \right] =$$

$$= \sum_{\ell \in \ell_i(j)} \hat{\phi}_{i\ell}(j) \left[D'_{ik}(f_{ik}(\tilde{\phi},r)) + \frac{\partial D_T(\tilde{\phi},r)}{\partial r_\ell(j)} \right] =$$

$$= S_i(\tilde{\phi},r)(j) = S_i(\hat{\phi},r)(j)$$

This together with part a) of the hypothesis establishes a').

To see b'), notice that by continuity of $\frac{\partial D_T(\phi,r)}{\partial r_i(j)}$ in ϕ and the preceding equation we have that for all k large enough, $k \in \tilde{K}$

$$\frac{\partial D_T(\phi^k,r)}{\partial r_i(j)} > \frac{\partial D_T(\phi^k,r)}{\partial r_\ell(j)} \quad \forall\ \ell \in \ell_i(j), \quad j = 1,\ldots,N\ . \tag{3.34}$$

Equations (3.34) and (3.32) hold for all $i \in I_{q+1}(j)/I_q(j)$. They also hold for all $i \in I_q(j)$, since the preceding analysis can be carried out for any $i \in I_m(j)/I_{m-1}(j)$, $m \leq q+1$, by using only part of the hypothesis. Therefore equations (3.34) and (3.32) hold for every $i \in I_{q+1}(j)$ $j = 1,..,N$ and b') follows.

Q.E.D.

By now we have developed all the machinery for the convergence proof. We will simply make repeated application of lemma 6 for the proper sequences.

Theorem 1: Let α be such that (3.9) holds. Then every limit point of $\{A^k(\phi^0)\}_{k=0}^{\infty}$ is optimal for the (MFP).

Proof: The sequence $\{A^k(\phi^0)\}_{k=0}^{\infty}$ belongs to a compact set and therefore there exists a convergent subsequence $\{\phi^k\}_{k\in\hat{K}} \to \hat{\phi}$. The sequence $\{\phi^{k-1}\}_{k\in\hat{K}}$ has a convergent subsequence $\{\phi^{k-1}\}_{k\in\hat{K}_1} \to \hat{\phi}_1$, $\hat{K}_1 \subset \hat{K}$. The sequence $\{\phi^{k-2}\}_{k\in\hat{K}_1}$ has a convergence subsequence $\{\phi^{k-2}\}_{k\in\hat{K}_2} \to \hat{\phi}_2$, $\hat{K}_2 \subset \hat{K}_1$.

Proceeding this way we get a convergent subsequence

$$\{\phi^{k-N+1}\}_{k\in\hat{K}_{N-1}} \to \hat{\phi}_{N-1}, \; K_{N-1} \subset K_{N-2} \quad .$$

We have $\hat{K}_{N-1} \subset \hat{K}_{N-2} \subset \cdots \subset \hat{K}$ and

$$\{\phi^{k-N+1}\}_{k\in\hat{K}_{N-1}} \to \hat{\phi}_{N-1}, \ldots, \{\phi^{k-1}\}_{k\in\hat{K}_{N-1}} \to \hat{\phi}_1, \{\phi^k\}_{k\in\hat{K}_{N-1}} \to \hat{\phi}$$

By the corollary to lemma 5 the shortest route distances which correspond to $\hat{\phi}_{N-1}$, $\hat{\phi}_{N-2}, \ldots, \hat{\phi}$ are the same. As a result, in what follows, when we mention the set $I_q(j)$ we need not specify the limit point $\hat{\phi}_i$ to which it corresponds.

Let $\tilde{K}$ in lemma 6 be $\hat{K}_{n-1}$. For each destination j, j is the only element in $I_1(j)$ and therefore the assumptions of lemma 6 hold for $I_1(j)$ and the pairs of sequences $\left(\{\phi^k\}_{k\in\tilde{K}}, \{\phi^{k-1}\}_{k\in\tilde{K}}\right)$,

$$\left(\{\phi^{k-1}\}_{k\in\tilde{K}}, \{\phi^{k-2}\}_{k\in\tilde{K}}\right), \ldots, \left(\{\phi^{k-N+2}\}_{k\in\tilde{K}}, \{\phi^{k-N+1}\}_{k\subset\tilde{K}}\right)$$

Applying lemma 6 for $q=1$, we obtain that the hypothesis of lemma 6 holds for $q = 2$ and the pairs of sequences $\left(\{\phi^k\}_{k\subset\hat{K}}, \{\phi^{k-1}\}_{k\subset\hat{K}}\right), \ldots,$ $\left(\{\phi^{k-N+3}\}_{k\in\tilde{K}}, \{\phi^{k-N+2}\}_{k\in\tilde{K}}\right)$. Proceeding this way we obtain that the hypothesis of lemma 6 holds for $q = N-1$ and the pair $\left(\{\phi^k\}_{k\in\tilde{K}}, \{\phi^{k-1}\}_{k\in\tilde{K}}\right)$. Applying lemma 6 again we obtain that part a') of lemma 6 holds for $q = N-1$, i,e equation (3.23) holds for $I_N(j)$ $j = 1,\ldots,N$. Since every node in the network belongs to $I_N(j)$ $j = 1,\dots,N$ (3.1) is satisfied for $\phi = \hat{\phi}$, and by lemma 1 $\hat{\phi}$ is optimal for the (MFP).

Q.E.D.

Theorem 1 combined with Lemma 6 implies the following corollary:

Corollary 1: Let α be such that (3.9) holds and let $\{A^k(\phi^0)\}_{k\in\tilde{K}}$ be a convergent sequence. Then there exists a subsequence $\{\phi^\ell\}_{\ell\in\tilde{L}}$, $\tilde{L}\subset\tilde{K}$ such that for every link (m,n) and destination there exists some $\bar{k}$ such that for all $k > \bar{k}$, $k \in \tilde{L}$, $j = 1,\dots,N$, $\phi^k_{mm}(j) > 0$ implies that

$$\frac{\partial D_T(\phi^k,r)}{\partial r_m(j)} > \frac{\partial D_T(\phi^k,r)}{\partial r_n(j)}.$$

Corollary 2: For every link (m,n) and destination j there exists some $\bar{k}$ such that for all $k > \bar{k}$, $j = 1,\dots, N$, $\phi^k_{mn}(j) > 0$ implies

$\dfrac{\partial D_T(\phi^k,r)}{\partial r_m(j)} > \dfrac{\partial D_T(\phi^k,r)}{\partial r_n(j)}$ and hence no improper links appear in the algorithm for all k sufficiently large.

Proof: The counter assertion is that for some j there exists an infinite subsequence $\{\phi^\ell\}_{\ell \in L}$ such that, for all $\ell \in L$, ϕ^ℓ has an improper link with respect to j. Extract a convergent subsequence $\{\phi^m\}_{m\in\bar{L}}$, $\bar{L}\subset L$. According to the preceding corollary, the sequence $\{\phi^m\}_{m\in\bar{L}}$ has a subsequence $\{\phi^n\}_{n\in\tilde{L}}$, $\tilde{L} \subset \bar{L}$ and for n large enough, $n \in \tilde{L}$, ϕ^n does not have an improper link. Since $n \in \tilde{L}\subset \bar{L}\subset L$ we have got a contradiction to the assertion.

Q.E.D.

Corollary 2 implies that for k large enough no improper links exist in the network. The set $B(\phi^k,r)(j)$ includes also nodes n such that

$\phi_{in}(j) = 0$, $\frac{\partial D_T(f^k,r)}{\partial r_i(j)} \leq \frac{\partial D_T(\phi^k,r)}{\partial r_n(j)}$. Using a similar argument, to the one which was used to prove lemma 4, one can show that for all k large enough these constraints are not binding. Together, corollary 2 and the observation above, implies that for all k large enough the constraints corresponding to blocked nodes are not active and can be dropped all together. Without these constraints (2.9) is exactly the Goldstein-Levitin-Polyak method, and all the rate of convergence results for this method apply.

3.5 <u>The "one-at-a-time" Mode of Implementation:</u>

The proof of the "one-at-a-time" mode of implementation is almost the same as the preceeding proof. Let's just sketch the general lines along which the proof evolves:

1. Notice that all lemmas hold for the "one-at-a-time" mode, where only some proportionality constants change. (Actually they become smaller, which allows α to be larger).

2. Extract a convergent subsequence $\{\phi^k\}_{k\in\hat{K}} \to \hat{\phi}$. Fix destination j and consider the sequence $\left\{\phi^{N\cdot\left\lfloor\frac{k-j}{N}\right\rfloor + j}\right\}_{k\in\hat{K}}$. By definition of the "one-at-a-time" node we have that $\phi^{N\cdot\left\lfloor\frac{k-j}{N}\right\rfloor + j}(j) = \phi^k(j)$ where $\phi(j) = [\phi_1(j)^T, \phi_2(j)^T, \ldots, \phi_N(j)^T]^T$ and $\lfloor q \rfloor$ denotes the maximal integer which is less then q. Take a convergent subsequence of $\left\{\phi^{N\cdot\left\lfloor\frac{k-j}{N}\right\rfloor + j}\right\}_{k\in K}$ to be

$\{\phi^{\ell_j}\}_{\ell_j \in K_j} \to \phi^*$. We have $\phi^*(j) = \hat{\phi}(j)$ and

$f_{ik}(j)(\phi^*,r) = f_{ik}(j)(\hat{\phi},r)$, $(i,k) \in L \quad j = 1,\ldots,N$

by slight modification of (3.20). Therefore, if ϕ^* satisfy the optimality conditions with respect to destination j so is $\hat{\phi}$.

3. Prove optimality of ϕ^* with respect to destination j by considering the pairs of sequences $\left\{\{\phi^{\ell_k}\}_{\ell_k \in K_{j_{N(N-1)}}}, \{\phi^{\ell_k - 1}\}\ldots\right\}$, $\left\{\{\phi^{\ell_k - N}\}_{\ldots}, \{\phi^{\ell_k - N - 1}\}_{\ldots}\right\}, \ldots, \left\{\{\phi^{\ell_k - N(N-1)}\}_{\ldots}, \{\phi^{\ell_k - N(N-1)-1}\}_{\ldots}\right\}$. Apply lemma 6, and use (3.20) and the fact that $\phi^{\ell_k - N\cdot P}(j) = \phi^{\ell_k - N(P-1)-1}(j)$ for all integers P, to create new pairs and so forth.

4. Repeat steps 2 and 3 for every destination to show that $\hat{\phi}$ is optimal with respect to all destinations.

Alternatively, one can modify lemma 6 and Theorem 1 and use (N-1)N subsequences.

APPENDIX A

Proof of Proposition 1:

As mentioned, this appendix follows similar arguments as appendix C of [1]. We establish the proof through a sequence of four lemmas.

Lemma A1:

$$\text{(A1)} \qquad \sum_{i,k} D'_{ik}(f_{ik}(\phi,r)) \cdot f_{ik}(\phi,r) = \sum_{j,k} r_k(j) \frac{\partial D_T(\phi,r)}{\partial r_k(j)}$$

Proof: By (2.2) and (2.4) we have

$$f_{ik}(\phi,r \cdot \beta) = f_{ik}(\phi,r) \cdot \beta \qquad \forall\, \beta > 0$$

and therefore

$$D_T(\phi,r \cdot \beta) = \sum_{i,k} D_{ik}[f_{ik}(\phi,r) \cdot \beta]$$

Differentiate (A2) with respect to β at $\beta = 1$ to get

$$\sum_{k,j} \frac{\partial D_T(\phi,r)}{\partial r_k(j)} \cdot \frac{\partial (r_k(j)\ \beta)}{\partial \beta} = \sum_{i,k} \frac{\partial D_{ik}(f_{ik}(\phi,r))}{\partial f_{ik}(\phi,r)} \frac{\partial (f_{ik}(\phi,r)\cdot\beta)}{\partial \beta}$$

or

$$\sum_{k,j} \frac{\partial D_T(\phi,r)}{\partial r_k(j)} r_k(j) = \sum_{i,k} D'_{ik}(f_{ik}(\phi,r)) \cdot f_{ik}(\phi,r)$$

Q.E.D.

Lemma A2:

$$\text{(A2)}\qquad \sum_{i,j} t_i(j)(\phi^k,r)\cdot t_i(j)(\phi^{k+1},r)\cdot \|\phi_i^{k+1}(j)-\phi_i^k(j)\|^2 \geq$$

$$p\sum_{ij} t_i^2(j)(\phi^k,r)\|\phi_i^{k+1}(j)-\phi_i^k(j)\|^2 \qquad i=1,\dots,N,\ j=1,\dots,N\ \ i\neq j$$

where p is some positive scalar.

Proof: Define ϕ^* as follows:

$$\text{(A4)}\qquad \phi^*_{in}(j) = \min(\phi^k_{in}(j),\ \phi^{k+1}_{in}(j))$$

Denote $t_i(j) = t_i(j)(\phi^k,r)$, $\overline{t_i(j)} = t_i(j)(\phi^{k+1},r)$ and $t_i^*(j) = t_i(j)(\phi^*,r)$ where $t_i^*(j)$ is mathematically well defined since ϕ^* is loopfree by definition of $B(i,\phi^k)(j)$. We have that

$$\text{(A5)}\qquad t_i^*(j) = \sum_\ell t_\ell^*(j)\phi^*_{\ell i}(j) + r_i(j) = \sum_\ell t_\ell^*(j)(\phi^k_{\ell i}(j) + \phi^*_{\ell i}(j) - \phi^k_{\ell i}(j)) + r_i(j),$$

Subtracting (A5) from (2.2) one gets

$$\text{(A6)}\qquad t_i(j) - t_i^*(j) = \sum_\ell (t_\ell(j) - t_\ell^*(j)\phi^k_{\ell i}(j) + \sum_\ell t_\ell^*(j)(\phi^k_{\ell i}(j) - \phi^*_{\ell i}(j)).$$

Equation (A6) has the general structure of (2.2) and therefore by (2.3) we get

$$\text{(A7)}\qquad t_i(j) - t_i^*(j) = \sum_\ell \frac{\partial t_i(j)}{\partial r_\ell(j)} \sum_k t_k^*(\phi^k_{k\ell}(j) - \phi^*_{k\ell}(j)) .$$

Since ϕ^k is loopfree we have $0 \leq \frac{\partial t_i(j)}{\partial r_\ell(j)} \leq 1$, and by definition of ϕ^* and loopfree of ϕ^k, one gets $\frac{\partial t_i(j)}{\partial r_\ell(j)} (\phi^k_{i\ell}(j) - \phi^*_{i\ell}(j)) = 0$. Therefore

(A8) $$t_i(j) - t_i(j) \leq \sum_{k\neq i} t^*_k(j) \sum_\ell (\phi^k_{k\ell}(j) - \phi^*_{k\ell}(j)) \; .$$

Since $1 \geq \sum_\ell (\phi^k_{i\ell}(j) - \phi^*_{i\ell}(j)) \geq 0$ we can write

(A9) $$t_i(j) \cdot \sum_\ell (\phi^k_{i\ell}(j) - \phi^*_{i\ell}(j) \leq \sum_{k\neq i} t^*_k(j) \sum_\ell (\phi^k_{k\ell}(j) - \phi^*_{k\ell}(j))$$

$$+ t^*_i(j) \sum_\ell (\phi^k_{i\ell} - \phi^*_{i\ell}(j))$$

$$\leq \sum_n t^*_n(j) \cdot \sum_\ell (\phi^k_{n\ell}(j) - \phi^*_{n\ell}(j)) \qquad i = 1,\ldots,N \;\; j=1,\ldots,N$$

Since $\phi^{k+1}(j) \geq \phi^*(j)$, $\phi^k(j) \geq \phi^*(j)$ and $t_i(j)(\phi,r) \geq 0$ we have that $t_i(j) \geq t^*_i(j)$ and $\overline{t_i(j)} \geq t^*_i(j)$, and as a result

(A10) $$\sum_i \overline{t_i(j)} \cdot t_i(j) \cdot \sum_\ell (\phi^k_{i\ell}(j) - \phi^*_{i\ell}(j))^2 \geq \sum_i t^*_i(j)^2 \cdot (\sum_\ell \phi^k_{i\ell}(j) \cdot \phi^*_{i\ell}(j))^2$$

$$\geq \frac{1}{N}(\sum_i t^*_i(j) \sum_\ell (\phi^k_{i\ell}(j) - \phi^*_{i\ell}(j)))^2 \geq \frac{1}{N} t^2_n(j) (\sum_\ell (\phi^k_{n\ell}(j) - \phi^*_{n\ell}(j))^2$$

where the second inequality follows from the Cauchy-Shwartz inequality and the last follows from (A9). Since (A10) holds for n=1,...,N we get

(A11) $$\sum_i \overline{t_i(j)}\, t_i(j)\cdot\left(\sum_\ell(\phi^k_{i\ell}(j) - \phi^*_{i\ell}(j))\right)^2 \geq \frac{1}{N^2}\sum_n t_n^2(j)\left(\sum_\ell(\phi^k_{n\ell}(j) - \phi^*_{n\ell}(j))\right)^2 .$$

By the Cauchy-Shwartz inequality

(A12) $$\|\phi_i^{k+1}(j) - \phi_i^*(j)\|^2 \geq \frac{1}{N}\left(\sum_\ell |\phi^{k+1}_{i\ell}(j) - \phi^k_{i\ell}(j)|\right)^2 = \frac{4}{N}\left(\sum_\ell(\phi^k_{i\ell}(j) - \phi^*_{i\ell}(j))\right)^2$$

(A13) $$\left(\sum_\ell(\phi^k_{i\ell}(j) - \phi^*_{i\ell}(j))\right)^2 = \frac{1}{4}\left(\sum_\ell |\phi^{k+1}_{i\ell}(j) - \phi^k_{i\ell}(j)|\right)^2 \geq \frac{1}{4} \|\phi_i^{k+1}(j) - \phi_i^k(j)\|^2 .$$

It follows that

(A14) $$t_i(j)\overline{t_i(j)}\,\|\phi_i^{k+1}(j) - \phi_i^k(j)\|^2 \geq \frac{4}{N}\sum_i \overline{t_i(j)}\cdot t_i(j)\left(\sum_\ell(\phi^k_{i\ell}(j) - \phi^*_{i\ell}(j))\right)^2$$

$$\geq \frac{4}{N^3}\sum_n t_n^2(j)\left(\sum_\ell(\phi^k_{n\ell}(j) - \phi^*_{n\ell}(j))\right)^2 \geq \frac{1}{N^3}\sum_i t_i^2(j)\,\|\phi_i^{k+1}(j) - \phi_i^k(j)\|^2$$

where the first inequality follows by (A12), the second by (A11) and the third by (A13). Summing (A14) over all j we get (A2).

Q.E.D.

<u>Lemma A3</u>:

(A15) $$\sum_{i,k} D'_{ik}(f_{ik}(\phi^k,r))(f_{ik}(\phi^{k+1},r) - f_{ik}(\phi^k,r)) =$$

$$\sum_{i,j} t_i(j)(\phi^{k+1},r)\cdot\delta_i(j)(\phi^k,r)^T(\phi_i^{k+1}(j) - \phi_i^k(j))$$

Proof:

$$\text{(A16)} \quad \sum_{ij} t_i(j)(\phi^{k+1},r) \cdot \delta_i(j)(\phi^k,r)^T(\phi_i^{k+1}(j) - \phi_i^k(j)) =$$

$$= \sum_{i,j} t_i(j)(\phi^{k+1},r) \; [\delta_i(j)(\phi^k,r)^T \cdot \phi_i^{k+1}(j) - \frac{\partial D_T(\phi^k,r)}{\partial r_i(j)}] =$$

$$= \sum_{i,k,j} f_{ik}(j)(\phi^{k+1},r)[D'_{ik}(f_{ik}(\phi^k,r)) + \frac{\partial D_T(\phi^k,r)}{\partial r_k(j)}] - \sum_{i,j} t_i(j)(\phi^k,r) \; \frac{\partial D_T(\phi^k,r)}{\partial r_i(j)}$$

$$= \sum_{i,k} f_{ik}(\phi^{k+1},r) \cdot D'_{ik}(f_{ik}(\phi^k,r)) + \sum_{k,j} (t_k(j)(\phi^{k+1},r) - r_k(j)) \cdot \frac{\partial D_T(\phi^k,r)}{\partial r_i(j)}$$

$$- \sum_{i,j} t_i(j)(\phi^{k+1},r) \cdot \frac{\partial D_T(\phi^k,r)}{\partial r_i(j)} = \sum_{i,k} f_{ik}(\phi^{k+1},r) \cdot D'_{ik}(f_{ik}(\phi^k,r)) -$$

$$\sum_{k,j} r_k(j) \; \frac{\partial D_T(\phi^k,r)}{\partial r_i(j)} = \sum_{i,k} (f_{ik}(\phi^{k+1},r) - f_{ik}(\phi^k,r)) \cdot D'_{ik}(\phi^k,r)$$

where the last equality follows from (A1).

Q.E.D.

We have by now all the machinery we need to prove proposition 1. Define $f_{ik}^{\lambda} = \lambda f_{ik}(\phi^{k+1},r) + (1 - \lambda) f_{ik}(\phi^k,r)$, $D_T(\lambda) = \sum_{i,k} D_{ik}(f_{ik}^{\lambda})$ and $U = \max_{\substack{(i,k)\in L \\ \phi \in \hat{\Phi}}} \{D''_{ik}(f_{ik}(\phi,r)) \mid D_T(\phi) \leq D_T(\phi^0)\}$. Then using (A.15) we obtain for some $\lambda^* \in [0,1]$

(A.17) $$D_T(\phi^{k+1},r) - D_T(\phi^k,r) = D_T(1) = D_T(0) =$$

$$= \left.\frac{dD_T(\lambda)}{d\lambda}\right|_{\lambda=0} + \frac{1}{2}\left.\frac{d^2D_T(\lambda)}{d\lambda^2}\right|_{\lambda=\lambda^*} =$$

$$= \sum_{i,k} D'_{ik}(f_{ik}(\phi^k,r)) \cdot \left.\frac{\partial f^\lambda}{\partial\lambda}\right|_{\lambda=0} + \sum_{i,k} D''_{ik}(f_{ik}^{\lambda*})\left[\left(\left.\frac{df_{ik}^\lambda}{d\lambda}\right|_{\lambda=\lambda*}\right)^2 + \left.\frac{d^2f_{ik}^\lambda}{d\lambda^2}\right|_{\lambda=\lambda*}\right.$$

$$= \sum_{i,k} D'_{ik}(f_{ik}(\phi^k,r)(f_{ik}(\phi^{k+1},r) - f_{ik}(\phi^k,r))$$

$$+ \sum_{i,k} D''_{ik}(f_{ik}^{\lambda*})(f_{ik}(\phi^{k+1},r) - f_{ik}(\phi^k,r))^2$$

$$\leq \sum_{i,j} t_i(j)(\phi^{k+1},r)\delta_i(j)(\phi^k,r)^T(\phi_i^{k+1}(j) - \phi_i^k(j))$$

$$+ \sum_{i,k} U \cdot (f_{ik}(\phi^{k+1},r) - f_{ik}(\phi^k,r))^2 .$$

Using the necessary condition for optimality in (2.9) we obtain

$$[\delta_i(j)(\phi^k,r) + \frac{t_i(j)(\phi^k,r)}{\alpha} . M_i^k(j)(\phi_i^{k+1}(j) - \phi_i^k(j)]^T(\phi_i^k(j) - \phi_i^{k+1}(j)) \geq 0$$

and therefore

$$\delta_i(j)(\phi^k,r)^T(\phi_i^{k+1}(j) - \phi_i^k(j)) \leq$$

$$\leq -\frac{t_i(j)(\phi^k,r)}{\alpha}(\phi_i^{k+1}(j) - \phi_i^k(j))^T M_i(j)(\phi_i^{k+1}(j) - \phi_i^k(j) \leq$$

$$\leq -\frac{\lambda}{\alpha} t_i(j)\|\phi_i^{k+1}(j) - \phi_i^k(j)\|^2.$$

We use last equation and (A2) to upper bound the next to last term in (A17). We upper bound the last term of (A17) using a slight modification of (3.15). As a result we obtain

$$\text{(A18)}\qquad D_T(\phi^{k+1},r) - D_T(\phi^k,r) \le \frac{1}{N^3}\,\frac{\lambda}{\alpha} \sum_{i,j} t_i(j)^2(\phi^k,r)\,\|\,\phi_i^{k+1}(j) - \phi_i^k(j)\,\|^2$$

$$+ UN^4 \sum_{i,j} t_i^2(j)(\phi^k,r)\,\|\,\phi_i^{k+1}(j) - \phi_i^k(j)\,\|^2 =$$

$$= \left(-\frac{1}{N^3}\,\frac{\lambda}{\alpha} + UN^4\right) \sum_{i,j} t_i^{\ell}(j)(\phi^k,r)\,\|\phi_i^{k+1}(j) - \phi_i^k(j)\,\|^2$$

Take $\alpha \in (0, \hat{\alpha}]$, $\hat{\alpha} < \frac{\lambda}{N^7 U}$ in (A18) and proposition 1 follows.

Q.E.D.

REFERENCES

[1] Gallager, R.G., "A Minimum Delay Routing Algorithm Using Distributed Computation", IEEE Trans. on Communication, Vol. COM-25, No. 1, pp. 73-85, Jan., 1977.

[2] Bertsekas, D.P., "Algorithms for Optimal Routing of Flow in Networks," Coordinated Science Laboratory Working Paper University of Illinois at Champaign-Urbana, June, 1978.

[3] Goldstein, A.A., "Convex Programming in Hilbert Space", Bull. Amr. Math. Soc., Vol. 6.70, 1964, pp. 709-710.

[4] Levitin, E.S., and B.T. Polyak, "Constrained Minimization Problems", USSR. Comp. Math. Math. Phys., Vol. 6, 1966, pp. 1-50.

[5] Poulos, J.C., "Simulation of a Distributed Routing Algorithm", M.S. Thesis Dept. of Electrical Engineering and Computer Science, M.I.T., August, 1977.

[6] Bertsekas, D., E. Gafni and Vastola, K.S., "Validation of Algorithms for Routing of Flow in Networks", Proceeding of 1978 IEEE Conference on Decision and Control, San Diego, Calif., Jan. 1979.