SECOND DERIVATIVE ALGORITHMS FOR MINIMUM
DELAY DISTRIBUTED ROUTING $1^{\prime \prime}$ NETKORKS ${ }^{\dagger}$
by

Dimitri P. Bertsekas*<br>Eli M. Gafni**<br>Robert G. Gailager***


#### Abstract

We pronose a class of algorithms for finding an optimal quasistatic routing in a communication network. The algorithms are based on Ganlager's method [1]. Their main feature is that they utilize second derivatives of the objective function and may be viewed as approximations to a constrained version of Newton's method. The use of second derivatives results in improved speed of convergence and automatic stepsize scaling with respect to level of traffic input. These advantages are of crucial importance for the practical implementation of the algorithm using distributed computation.


[^0]
## 1. Introduction

We consider the problem of optimal routing of messages in a communication network so as to minimize average delay per message. We primarily have in mind a situation where the statistics of external traffic inputs change slowly with time as described in the paper by Gallager [1]. While algorithms of the type to be described can also be used for centralized computation, we place primary emphasis on algorithms that are well suited for distributed computation

Two critical requirements for the success of a distributed routing algorithm are speed of convergence and relative insensitivity of performance to variations in the statistics of external traffic inputs. Unfortunately the algorithm of [1] is not entirely satisfactory in these respects. In particular it is impossible to select in this algorithm a stepsize that will guarantee convergence and good rate of convergence for a broad range of external traffic inputs. The work described in this paper was motivated primarily by this consideration.
A. standard approach for improving the rate of convergence and facilitating stepsize selection in optimization algorithms is to scale the descent direction using second derivatives of the objective function as for example in Newton's method. This is also the approach taken here. On the other hand the straightforward use of Newton's method is inappropriate for our problem primarily because of large dimensionality. We have thus introduced various approximations to Newton's method which exploit the network structure of the problem and facilitate distributed computation.

In Section 2 we describe a broad class of algorithms for minimum delay routing. This class is patterned after a gradient projection method
for nonlinear programing [2],[3] as explained in [4], and contains as a special case Gallager's original algorithn except for a variation in the definition of a blocked node [compare with equation (15) of [1]]. This variation is essential in order to avoid unnecessary complications in the statement and operation of our algorithms and despite, its seemingly minor significance, it has necessitated a major divergence in the proof of convergence from the corresponding proof of [1].

Section 3 describes in more detail a particular algorithm from the class of Section 2. This algorithm employs second derivatives in a manner which approximates a constrained version of Newton's method [3] and is well suited for distributed computation.

The algorithm of Section 3 seems to work well for most quasistatic routing problems likely to appear in practice as extensive computational experience has shown [5]. However there are situations where the unity stepsize employed by this algorithm may be inappropriate. In Section 4 we present another distributed algorithm which automatically corrects this potential difficulty whenever it arises at the expense of additional computation per iteration. This algorithm also employs second derivatives, and is based on minimizing at each iteration a suitable upper bound to a quadratic approximation of the objective function.

Proofs of convergence have been relegated to Appendices. Both algorithms of Sections 3 and 4 have been rested extensively and computational results have been documented in [5] and [6]. These results substantiate the as- ions made here regarding the practical properties of the algorithms. There are also other related second
-4-
derivative algorithns [7], [8] that operate in the space of path flows and exhibit similar behavior as the ones of this paper. These algorithms are well suited for centralized computation and virtual circuit networks but, in contrast with the ones of the present paper, recuire global information at each node regarding the network topology and the total flow on each link.

We finally mention that while we have restricted attention to the problem of routing, the algorithms of this paper can be applied to other problems of interest in commuication networks. For example problems of optimal adaptive flow control or combined routing and flow control have been formulated in [9],[10] as nonlinear multicoinodity flow problems of the type considered here, and the algorithms of this paper are suitable for their solution.

2. A Class of Routing Algorithms

Consijer a network consisting of $N$ nodes denoted by $1,2, \ldots, N$ and $L$ directed links. The set of links is denoted by $L$. We denote by ( $i, \ell$ ) the link from node $i$ to node $\ell$ and 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 flow on each link $(i, \ell)$ for any destination $j$ is denoted by $f_{i \ell}(j)$. The total flow on each link ( $i, \ell$ ) is denoted by $F_{i \ell}$, i.e.

$$
F_{i \ell}=\sum_{j=1}^{N} f_{i \ell}(j)
$$

The vector of all flows $f_{i \ell}(j),(i, \ell) \varepsilon L, j=1, \ldots, N$ is denoted by $f$.
We are interested in numerical solution of the following multiconmodity network flow problem:

$$
\begin{align*}
& \text { minimize } \sum_{(i, \ell) \varepsilon L} D_{i \ell}\left(F_{i \ell}\right)  \tag{MFP}\\
& \text { subject to } \sum_{\ell \in 0(i)} f_{i \ell}(j)-\sum_{m \in I(j)} f_{m i}(j)=r_{i}(j), \\
& \forall i=1, \ldots, N, i \neq j \\
& f_{i \ell}(j)=0, \quad \forall(i, \ell) \varepsilon L, i=1, \ldots, N, j=1, \ldots, N \\
& f_{j \ell}(j)=0, \quad V(j, \ell) \varepsilon L, j=1, \ldots, N,
\end{align*}
$$

where, for $i \neq j, r_{i}(j)$ is a known traffic input at node $i$ destined for $j$, and $O(i)$ and $I(i)$ are the sets of nodes $\ell$ for which $(i, \ell) \varepsilon L$ and $(\ell, i) \varepsilon L$ respectively.

The standing assumptions throughoat the paper are:
a) $r_{i}(j) \geq 0, \quad i, j=1, \ldots, N, i \neq j$
b) Each function $D_{i \ell}$ is defined on an interval $\left[0, C_{i \ell}\right.$ ) where $C_{i \ell}$ is either a positive number (the link capacity) or $+\infty$; $D_{i \ell}$ is twice continuously differentiable on ( $0, \mathrm{C}_{\mathrm{i} \ell}$ ). The first and second derivatives of $\mathrm{D}_{\mathrm{i} \ell}$ at zero are defined by taking the linit from the right. Furthermore $D_{i \ell}$ is convex, continuous, and ras strictly positive first and second derivatives on $\left[0, C_{i 2}\right]$.
c) (MFP) has at least one feasible solution, $f$ satisfying $F_{i \ell}<C_{i \ell}$ For all (i, 2) $\varepsilon$.

For notational convenience in describing various algorithas we will suppress in what follows the destination index and concentrate on a single destination chosen for concreteness to be node N. Our definitions, optimality conditions, and algorithms are essentially identical for each destination, so this notational simplification should not become a source of confusion. In the case where there are multiple destinations it is possible to implement our algorithns in at least two different ways. Either iterate simultaneously for all destinations (the "all-at-once" version), or iterate sequentially one destination at a tife in a cyclic manner with intermediate readjustnent of link flows (the "one-at-a-time" version). The remainder of our notation follows in large measure the one employed in [1]. In addition all vectors will be considered to be column vectors, transposition wiil be denothed by a superscript $T$, and the stanciard Euclidean norm of a vector will be denoted by $|\cdot|$, i.e. $x^{T} x=|x|^{2}$ for any vector $x$. Vector inequalities are meant to be componentwise, i.e. for $x=\left(x_{1}, \ldots, x_{n}\right)$ we write $x \geq 0$ if $x_{i .} \geq 0$ for all $i=1, \ldots, n$. Let $t_{i}$ be the total incoming traffic at node $i$

$$
\begin{equation*}
t_{i}=r_{i}+\sum_{\substack{m \in I \\ m \neq N}} f_{\operatorname{mi}}, \quad i=1, \ldots, N-1, \tag{l}
\end{equation*}
$$

-7-
and for $t_{i} \neq 0$ let $\phi_{i \ell}$ be the fraction of $t_{i}$ that -ravels on link ( $i, 2$ )

$$
\phi_{i \ell}=\frac{f_{i \ell}}{t_{i}}, \quad i=1, \ldots, N-1(i, \ell) \varepsilon i
$$

Then it is possible to reformulate the problem in terms of the variables $\phi_{i \ell}$ as follows [1].

For each node $i \neq N$ we fix an order of the outgoing links ( $i, \ell$ ), $\ell \in O(i)$. We identify with each collection $\left\{\phi_{i g} \mid(i, \ell) \varepsilon L, i=i, \ldots, N-1\right\}$ a column vector $\phi=\left(\phi_{1}^{T}, \phi_{2}^{T} \ldots=\phi_{\mathrm{N}-1}^{\mathrm{T}}\right)^{-}$, where $\phi_{i}$ is the colum vector with coordinates $\varphi_{i \ell}, \ell \varepsilon 0(i)$. Let

$$
\begin{equation*}
\bar{\phi}=\left\{\phi_{1}^{\mid} \phi_{i \ell} \geq 0, \sum_{\ell \varepsilon 0(i)} \phi_{i \ell}=1, \quad(i, \ell) \varepsilon L, i=1, \ldots, N-1\right\} \tag{2}
\end{equation*}
$$

and let $\phi$ be the subset of $\bar{\phi}$ consisting of all $\phi$ for which there exists a directed path $(i, 2), \ldots,(m, N)$ from every node $i=1, \ldots, N-1$ to the destination $N$ along which $\phi_{i 2}>0, \ldots, \phi_{m N}>0$. Clearly $\Phi$ and $\bar{\phi}$ are convex sets, and the closure of $\Phi$ is $\bar{\Phi}$. It is shown in [1] that for every $\phi \varepsilon \Phi$ and $r=\left(r_{1}, r_{2}, \ldots, r_{N-1}\right)$ with $r_{i} \geq 0, i=1, \ldots, N-1$ there exist unique vectors $t(\phi, r)=\left(t_{1}(\phi, r), \ldots, t_{N-1}(\phi, r)\right)$ and $f(\phi, r)$ with coordinates $f_{i \ell}(\phi, r),(i, \ell) \varepsilon L, i \neq N$ satisfying

$$
\begin{aligned}
& \tau(\phi, r) \geq 0, f(\phi, r) \geq 0 \\
& t_{i}(\phi,=)=r_{i}+\sum_{\substack{\operatorname{meI}(i) \\
m \neq N}} f_{m i}(\phi, r), \quad i=1,2, \ldots, N-1 \\
& \sum_{\ell \in 0(i)} f_{i \ell}(\phi, r)-\sum_{\substack{\operatorname{mel}(i) \\
m \neq N}} f_{m i}(\phi, r)=r_{i}, \quad i=1, \ldots, N-1 \\
& f_{i \ell}(\phi, r)=t_{i}(\phi, r) \phi_{i \ell}, \quad i=1, \ldots, N-1 ;(\bar{i} ; \ell) \varepsilon L .
\end{aligned}
$$

Furthermore the functions $t(\phi, r), f(\phi, r)$ are twice continuously differentiable in the relative interior of their domain of definition $\Phi x\{r \mid r \geq 0\}$. The derivatives at the relative boundary can also be defined by taking the limit through the relative interior. Furthermore for every $r \geq 0$ and every $f$ which is feasible for (MFP) there exists a $\phi \varepsilon \Phi$ such that $f=f(\phi, r)$.

It follows from the above discussion that the problem can be written in terms of the variables $\phi_{i \ell}$ as

$$
\begin{equation*}
\operatorname{minimize} \quad D(\phi, r) \triangleq \sum_{(i, \ell) \varepsilon L} D_{i \ell}\left[f_{i \ell}(\phi, r)\right] \tag{3}
\end{equation*}
$$

subject to фєФ,
where we write $D(\phi, r)=\infty$ if $\varepsilon_{i \ell}(\psi, r) \geq C_{i \ell}$ for some (i, $\left.\ell\right) \varepsilon L$.
Similarly as in [1], our algorithms generate sequences of loopfree routing variables $\phi$ and this allows efficient computation of various derivatives of $D$. Thus for a given $\phi \varepsilon \Phi$ w: say that node $k$ is downstream from node if if there is a directed path from $i$ to $k$, and for every link ( $\ell$, m) on the path we have $\phi_{\ell m}>0$. We say that node $i$ is upstream from node $k$ if $k$ is downstream from $i$. We say that $\phi$ is loopfree if there is no pair of nodes $i, k$ such that $i$ is both upstream and dowstream from $k$. For any $\phi \varepsilon \Psi$ and $r \geq 0$ for which
$D(\phi, r)<\infty$ the partial derivatives $\frac{2 D(\phi, r)}{\partial \phi_{i \ell}}$ can be computed using the following equations [1]

$$
\begin{align*}
& \frac{\partial D}{\partial \phi_{i \ell}}=t_{i}\left(D_{i \ell}^{\prime}+\frac{\partial D}{\partial r_{\ell}}\right), \quad(i, \ell) \varepsilon L, i=1, \ldots, N-1  \tag{4}\\
& \frac{\partial D}{\partial r_{i}}=\sum_{\ell \varepsilon O(i)} \phi_{i \ell}\left(D_{i \ell}^{\prime}+\frac{\partial D}{\partial r_{\ell}}\right), i=1, \ldots, N-1  \tag{5}\\
& \frac{\partial D}{\partial r_{N}}=0
\end{align*}
$$

where $D_{i \ell}^{\prime}$ denotes the first derivative of $D_{i \ell}$ with respect to $f_{i \ell}$. The equations above uniquely determine $\frac{\partial D}{\partial \phi_{i \ell}}$ and $\frac{\partial D}{\partial r_{i}}$ and their computation is particularly simple if $\phi$ is loopfree. In a distributed setting each node $i$ computes $\frac{\partial D}{\partial \phi_{i l}}$ and $\frac{\partial D}{\partial r_{i}}$ via (4),(5) after receiving the value of $\frac{\partial D}{\partial r_{\ell}}$ from all its immediate downstream neighbors. Because $\phi$ is loopfree the computation can be organized in a deadlock-free manner starting from the destination node N and proceeding upstream [1].

A necessary condition for optimality is given by (see [1])

$$
\begin{array}{ll}
\frac{\partial D}{\partial \phi_{i \ell}}=\min _{\min (i)} \frac{\partial D}{\partial \phi_{i m}} & \text { if } \phi_{i \ell}>0 \\
\frac{\partial D}{\partial \phi_{i \ell}} \geq \min _{\min 0(i)} \frac{\partial D}{\partial \phi_{i m}} & \text { if } \phi_{i \ell}=0,
\end{array}
$$

where all derivatives are evaluated at the optimum. In view of $44^{-}$, this condition can be written for $\tau_{i}>0$

$$
\begin{array}{ll}
D_{i \ell}^{\prime}+\frac{\partial D}{\partial r_{\ell}}=\min _{m \in 0(i)}\left(D_{i m}^{\prime}+\frac{\partial D}{\partial r_{m}}\right) & \text { if } \phi_{i \ell}>0 \\
D_{i \ell}^{\prime}+\frac{\partial D}{\partial r_{\ell}} \geq \min _{m \varepsilon 0(i)}\left(D_{i m}^{\prime}+\frac{\partial D}{\partial r_{m}}\right) & \text { if } \phi_{i \ell}=0 .
\end{array}
$$

Combining these relations with (5) we have that if $t_{i} \neq 0$ then

$$
\begin{equation*}
\frac{\partial D}{\partial r_{i}}=\min _{\operatorname{me0}(\mathrm{i})} \delta_{i m} \tag{6}
\end{equation*}
$$

where

$$
\begin{equation*}
\delta_{i m}=D_{i m}^{\prime}+\frac{\partial D}{\partial r_{n}}, \quad \forall m \varepsilon 0(i) \tag{7}
\end{equation*}
$$

In fact if (6) holds for all $i$ (whether $t_{i}=0$ or $t_{i}>0$ ) then it is surficient to guarantee optimality (see [1], Theorem 3).

We consider the class of algorithms

$$
\begin{equation*}
\phi_{i}^{k+1}=\phi_{i}^{k}+\Delta \phi_{i}^{k} \quad, \quad i=1, \ldots, k-1 \tag{8}
\end{equation*}
$$

where, for each $i$, the vector $\Delta \phi_{i}^{k}$ with components $\Delta \phi_{i \ell}^{k}, \ell \varepsilon 0$ (i) is any solution of the problem

$$
\begin{gather*}
\text { minimize } \hat{o}_{i}^{T} \Delta \phi_{i}+\frac{t_{i}}{2 \alpha} \Delta \phi_{i}^{\mathrm{M}_{i}^{k}} \Delta \phi_{i}  \tag{9}\\
\text { subject to } \phi_{i}^{k}+\Delta \phi_{i} \geq 0, \quad \sum_{i} \Delta \phi_{i \ell}=0, \\
\Delta \phi_{i \ell}=0, \quad \forall \ell \varepsilon B\left(i ; \phi^{k}\right) .
\end{gather*}
$$

The scalar $\alpha$ is a positive parameter. The vector $\delta_{i}$ has components [cf. (7)]

$$
\delta_{i \ell}=D!_{i \ell}+\frac{\partial D}{\partial r_{\ell}}, \quad \forall \ell \varepsilon 0(i)
$$

where all derivatives are evaluated at $\dot{\phi}^{k}$ and $f\left(\phi^{k}, r\right)$, and $\dot{\delta}_{i}^{T}$ (or $\Delta \phi_{i}^{T}$ ) denotes transpose of $\hat{o}_{i}\left(\right.$ or $\left.\Delta \phi_{i}\right)$. For each $i$ for which $t_{i}\left(\phi^{k}, r\right)>0$, the matrix $N_{i}^{k}$ is some symmetric matrix which is positive definite on the subspace $\left\{v_{i} \mid \sum_{\ell \in 0(i)} v_{i}=0\right\}$, i.e.

$$
v_{i}^{T} v_{i}^{k} v_{i}>0, \quad \forall v_{i} \neq 0, \sum_{i \in 0(i)} \quad v_{i \ell}=0 .
$$

This condition guarantees that the solution to problem (3) exists and is unique. For nodes $i$ for which $t_{i}\left(\oint^{k}, r\right)=0$ the definition of $H_{i}^{k}$ is inmaterial. The set of indices $B\left(i ; \dot{o}^{k}\right)$ is specified in the following definition:

Definition: For any $\phi E \phi$ and $i=1 ; \ldots, \ldots=1$ the set $B(i ; \phi)$, referred to as the set of blocked nodes for $\phi$ at $i$, is the set of all $2 E 0$ (i) such that $\phi_{i 2}=0$, and either $\frac{\partial D(\phi, r)}{\partial r_{i}} \leq \frac{\partial D(\bar{\phi}, \bar{r})}{\partial r_{2}}$, or there exists a link ( $m, n$ ) referred to as an improper link such that $m=\ell$ or $n$ is downstream of $\ell$ and we have $\phi_{m \mathrm{~m}}>0, \frac{\partial \bar{D}(\phi, \dot{r})}{\partial r_{m}} \leq \frac{\partial \bar{D}(\phi, r)}{\partial r_{i}}$.

It is shown below thas if $\phi^{k}$ is loopfree, then $\phi^{k+1}$ generated by the algorithm is also looperee. Thus the algorithm generates a sequence of loopfree routinge if the starting $\dot{\oint}^{\circ}$ is loopfree. He refer to [1] for a rescription of the method for generating the sets $B\left(i ; \phi^{k}\right)$ in a manner suitable for distributed computation. Our definition of $B\left(i ; \phi^{k}\right)$ differs from the one of [1] primarily in that a special device that facilitated the proof of convergence given in [1] is not employed (compare with equ. (is) of [1]).

He now demonstrate some of the properties of the algorithm in the following proposition.

Proposition 1: a) If $\phi^{k}$ is loopfree then $\phi^{i+1}$ is loopfree.
b) If $\phi^{k}$ is loopfree and $\Delta \phi^{k}=0$ solves problem (9) then $\phi^{k}$ is optimal.
c) If $\phi^{k}$ is optimal then $\phi^{k+1}$ is also optimal.
d) If $\Delta \phi_{i}^{k} \neq 0$ for some $i$ for which $t_{i}\left(\phi^{k}, r\right)>0$ then there exists a positive scalar $\eta_{k}$ such tilt

$$
D\left(\phi^{k}+\Pi \Delta \phi^{k}, r\right)<D\left(\phi^{k}, r\right), \quad \forall \eta \varepsilon\left(0, \eta_{k}\right]
$$

Proof: a) Assume that $\phi^{k+1}$ is not loopfree and there exists a sequence of inks forming a directed loop al.. rich $^{\mathrm{k}+1}$ is positive. Then there must exist a link $(m, n)$ on the loop for which $\frac{\partial D\left(\phi^{k}, r\right)}{\partial r_{m}} \leq \frac{\partial D\left(\phi^{k}, r\right)}{\partial r_{n}}$. From the definition of $B\left(m ; \phi^{k}\right)$ we must have $\phi_{\operatorname{mn}}^{k}>0$ and hence $(m, n)$ is an improper link. Now move backwards around the loop to the first link ( $i, \ell$ ) for which $\phi_{i \ell}^{k}=0$. Such a link must exist since $\phi^{k}$ is loopfree. Since $\ell$, is upstream of $m$ and $(m, n)$ is improper, we have $2 \varepsilon B\left(i ; \phi^{k}\right)$ which contradicts the hypothesis $\phi_{i l}^{k+1}>0$.
b) If $\Delta \phi^{k}=0$ solves problem (9) then we must have $\delta_{i}^{T} \Delta \phi_{i} \geq 0$ for each $i$ and $\Delta \phi_{i}$ satisfying the constraints of (9)

$$
\begin{equation*}
\Delta \phi_{i} \geq-\phi_{i}^{k}, \sum_{\ell} \Delta \phi_{i \ell}=0, \quad \Delta \phi_{i \ell}=0, \quad \forall \ell \varepsilon B\left(i ; \phi^{k}\right) \tag{10}
\end{equation*}
$$

By writing $\Delta \phi_{i}=\phi_{i}-\phi_{i}^{k}$ and using (5), (7) we have

$$
\begin{aligned}
\delta_{i}^{T}\left(\phi_{i}-\phi_{i}^{k}\right) & =\sum_{\ell} \delta_{i \ell} \phi_{i \ell}-\sum_{\ell} \delta_{i \ell} \phi_{i \ell}^{k} \\
& =\sum_{\ell} \delta_{i \ell} \phi_{i \ell}-\frac{\partial D}{\partial r_{i}} \geq 0 .
\end{aligned}
$$

By considering $\phi_{i \ell}=1$ individually for each $\ell \notin F\left(i ; \phi^{k}\right)$, we obtain

$$
\frac{\partial D}{\partial r_{i}} \leq \delta_{i \ell}, \quad \forall \ell \& B\left(i ; \phi^{k}\right)
$$

From (5) and (7) then

$$
\frac{\partial D}{\partial r_{i}}=\delta_{i \ell}, \quad \forall \ell \varepsilon B\left(i ; \phi^{k}\right), w i \not h \phi_{i \ell}^{k}>0 .
$$

Since $D_{i \ell}^{\prime}>0$ for all ( $i, \ell$ ) $\varepsilon L$ it follows from (5), (7) and the relation abuve that there are no improper links, and using the definition of $B\left(i ; \phi^{k}\right)$ we obtain

$$
\frac{\partial \nu}{\partial r_{i}}=\min _{\ell \varepsilon 0(i)} \delta_{i \ell}
$$

which is a sufficient condition for optimality of $\phi^{k}$ [cf. (6)]. c) If $\phi^{k}$ is optimal then from the necessary condition for optjmality (6) we have that for all $i$ with $t_{i}>0$

$$
\frac{\partial D}{\partial r_{i}}=\min _{m \in 0(i)} \delta_{i m}
$$

It follows using a reverse argument to the one in b) above that

$$
\Delta \phi_{i}^{k}=0 \quad \text { if } t_{i}>0 .
$$

Since changing only routing variables of nodes $i$ for which $t_{i}=0$ does not affect the flow through each link we have $D\left(\phi^{k}, r\right)=D\left(\phi^{k+1}, r\right)$ and $\phi^{k+1}$ is optimal.
d) Since $M_{i}^{k}$ is positive semidefinite for all $i$ with $t_{i}>0$ and $\Delta \phi_{i}^{k}$ is a solution af problem (9) we have

$$
\overbrace{\mathrm{i}}{ }_{\mathrm{i}}^{\mathrm{r}} \phi_{\mathrm{i}}^{\mathrm{k}} \leq 0
$$

If $t_{i}>0$ then $M_{i}^{k}$ is positive definite on the appropriate subspace and
the solution of problem ( 9 ) ;- unique, so if in addition $\Delta d^{k} \neq 0$ then we have

$$
\delta_{i}^{T} \Delta \phi_{i}^{k}<0
$$

Using the fact [cf. (4), (T)]

$$
\frac{\partial D}{\partial \phi_{i}}=t_{i} \delta_{i}
$$

we obtain that

$$
\frac{\partial D^{T}}{\partial \phi_{\mathbf{i}}^{*}} \Delta \phi_{i}^{k}<0
$$

Hence $\Delta \phi^{k}$ is a direction of descent at $\phi^{k}$ and the result follows. Q.E.D.
The following proposition is the main convergence result regarding algorithm (8), (9). Its proof is quite complex and is given in Appendix
A. The proposition applies to the multiple destination case in the "all-at-once" and the "one-at-a-time" version.

Proposition 2: Let the initial routing $\phi^{0}$ be loopfree and satisfy $D\left(\phi^{0}, r\right) \leq D_{0}$ where $D_{u}$ is some scalar. Assume also that there exist two positive scalars $\lambda, \Lambda$ such that the sequences of matri.ces $\left\{M_{\dot{i}}^{k}\right\}$ satisfy the following two conditions:
a) The absolute value of each element of $M_{i}^{k}$ is bounded above by $\Lambda$.
b) There holds

$$
\lambda\left|v_{i}\right|^{2} \leq v_{i}^{T} M_{i}^{k} v_{i}
$$

for all $v_{i}$ in the suhspace $\left.\left\{v_{i} \mid \sum_{\ell \notin B(i ; \phi}{ }^{k}\right) v_{i \ell}=0\right\}$.
Then there exists a positive scalar $\bar{\alpha}$ (depending on $D_{0}, \lambda$, and $\Lambda$ ) such that for all
$\alpha \varepsilon(0, \bar{\alpha}]$ and $k=0,1, \ldots$ the sequence $\left\{\phi^{k}\right\}$ generated by algorithm (8), (9) satisfies

$$
D\left(\phi^{k+1}, r\right) \leq D\left(\phi^{k}, r\right), \lim _{k \rightarrow \infty} D\left(\phi^{k}, r\right)=\min _{\phi \varepsilon \Phi} D(\phi, r)
$$

Furthermore every limit point of $\left.i \phi^{\mathrm{k}}\right\}$ is an optimal solution of problem (3).

Another interesting result which will not be given here but can be found in [11] states that, after a finite rumber of iterations, improper links do not appear further in the algorithm so that for rate of convergence analysis purfoses the potential presence of improper links can be ignured. Based on this fact it can be shown under a mild assumption that for the single destination case the rate of convergence of the algorithm is linear [11].

The class of algorithms (8), (9) is quite broad since different choices of matrices $M_{i}^{k}$ yield different algorithms. A specific choice of $M_{i}^{k}$ yields Gallager's algorithm [1] [except for the difference in the definition of $B\left(i ; \phi^{k}\right)$ mentioned earlier]. This choice is the one for which $M_{i}^{k}$ is diagonal with all elements along the diagonal being unity except the $(\bar{l}, \bar{l})$ th element which is zero where $\bar{l}$ is a node for which

$$
\delta_{i \bar{i}}=\min _{\ell ट 0(i)} \delta_{i \ell} .
$$

We leave the verification of this fact to the reader. In the next section we describe a specific algorithm involving a choice of $M_{i}^{k}$ based on second derivatives of $D_{i l}$. The convergence result of Proposition 2 is applicable to this algorithri.

## 3. An Algorithm Based on Second Derivatives

A drawback of the algorithn of [1] is that a proper range of the stepsize parameter $\alpha$ is hard to determine. In order for the algorithm to have guaranteed convergence for a broad range of inputs $r$, one must take $\alpha$ quite small but thiswill lead to a poor speed of convergence for most of these inputs. It appears that in this respect a better choice of the matrices $M_{i}^{k}$ can be based on second derivatives. This tends to make the algorithm to a large extent scale free, and for most problems likely to aprear in practice, a choice of the stepsize $\alpha$ near unity results in both convergence and reasonably good speed of convergence for a broad range of inputs $r$. This is supported by extensive computational experience some of which is reported in [5] and [6].

We use the notation

$$
D_{i \ell}^{\prime \prime}=\frac{\partial^{2} D_{i \ell}}{\left[\partial f_{i \ell}\right]^{2}}
$$

We have already assumed that $D_{i \ell}^{\prime \prime}$ is positive in the set $\left[0, C_{i \ell}\right)$. We vould like to choose the matrices $M_{i}^{k}$ to be diagonal with $\tau_{i}^{-2} \frac{\partial^{2} D\left(\phi^{k}, r\right)}{\left[\partial \phi_{i \ell}\right]^{2}}$ along the diagonal. This corresponds to an approximation of a constrained version of Newton's method (see [3]), where the off-diagonal terms of the Hessian matrix of $D$ are set to zero. This type of approximated version of Newton's method is often employed in solving large scale unconstrained optimization pioblems. Unfortunately the second derivatives $\frac{\partial^{2} D}{\left[\partial \phi_{i \ell}\right]^{2}}$ are difficult to compute. However, it is possible to compute easily upper and lower bounds to them which, as shown by computational ex-
periments, are sufficiently accurate for practical purposes.

Calculation of Upper and Lower Bounds to Second Derivatives
We compuie $\frac{\partial^{2} D}{\left[\partial \phi_{i \ell}\right]^{2}}$ evaluated at a loopfree $\phi Е \Phi$, for all links $(i, \ell) \varepsilon L$ fo ..sch $\ell \in B(i ; \phi)$. We have using (4)

$$
\frac{\partial^{2} D}{\left[\partial \phi_{i \ell}\right]^{2}}=\frac{\partial}{\partial \phi_{i \ell}}\left\{t_{i}\left(D_{i \ell}^{\prime}+\frac{\partial D}{\partial r_{\ell}}\right)\right\}
$$

Since $\ell \neq B(i ; \phi)$ and $\phi$ is loopfree, the node $\ell$ is not upstream of i. It follows that $\frac{\partial t_{i}}{\partial \phi_{i \ell}}=0$ and $\frac{\partial D_{i \ell}^{\prime}}{\partial \phi_{i \ell}}=D_{i \ell}^{\prime \prime \prime} t_{i}$. Using again the fact that $\ell$ is not upstream of $i$ we have $\frac{\partial t_{i}}{\partial r_{\ell}}=0, \frac{\partial D_{i \ell}}{\partial r_{\ell}}=0$ and it follows that

$$
\frac{\partial^{2} D}{\partial \phi_{i \ell} \partial r_{\ell}}=\frac{\partial}{\partial r_{\ell}} \frac{\partial D}{\partial \phi_{i \ell}}=\frac{\partial}{\partial r_{\ell}}\left\{t_{i}\left(D_{i \ell}^{\prime}+\frac{\partial D}{\partial r_{\ell}}\right)\right\}=t_{i} \frac{\partial^{2} D}{\left[\partial r_{\ell}\right]^{2}} .
$$

Thus we finally obtain

$$
\begin{equation*}
\frac{\partial^{2} D}{\left[\partial \phi_{i \ell}\right]^{2}}=t_{i}^{2}\left(D_{i \ell}^{\prime \prime}+\frac{\partial^{2} D}{\left[\partial r_{\ell}\right]^{2}}\right) \tag{11}
\end{equation*}
$$

A little thought shows that the second derivative $\frac{\partial^{2} D}{\left[\partial r_{\ell}\right]^{2}}$ is given by
the more general formula the more general formula

$$
\begin{equation*}
\frac{\partial^{2} D}{\partial r_{\ell} \partial r_{m}}=\sum_{(j, k) \in L} q_{j k}(\ell) q_{j k}(m) D_{i k}^{\prime \prime}, \quad \forall \ell, m=1, \ldots, N-1 \tag{i2}
\end{equation*}
$$

where $q_{j k}(\ell)$ is the portion of a unit of flow originating at $\ell$ which goes through link $(j, k)$. However calculation of $\frac{\partial^{2} D}{\left[\partial r_{\ell}\right]^{2}}$ using this
formula is complicated, and in fact tibere seems to be no easy way to compute this second derivative. However upper and lower bounds to it can be easily computed as we now show. By using (5) we obtain

$$
\frac{\partial^{2} D}{\left[\partial r_{\ell}\right]^{2}}=\frac{\partial}{\partial r_{\ell}}\left\{\sum_{\text {m }} \phi_{\ell m}\left(D_{\ell m}^{\prime}+\frac{\partial D}{\partial r_{m}}\right)\right\}
$$

Since $\phi$ is loopfree we have that if $\phi_{\ell m}>0$ then $n$ is not upstream of $\ell$ and therefore $\frac{\partial t_{\ell}}{\partial r_{\ell}}=1$ and $\frac{\partial D_{\ell m}^{\prime}}{\partial r_{\ell}}=D_{\ell}^{\prime \prime} \phi_{\ell \text { III }}$. A similar reasoning shows that

$$
\frac{\partial^{2} D}{\partial r_{\ell} \partial r_{m}}=\frac{\partial}{\partial r_{m}}\left\{\sum_{n} \phi_{\ell n}\left(D_{2 n}^{\prime}+\frac{\partial D}{\partial r_{n}}\right)\right\}=\sum_{n} \phi_{\ell n} \frac{\partial^{2} D}{\partial r_{m} \partial r_{n}}
$$

Combining the above relations we obtain

$$
\begin{equation*}
\frac{\partial^{2} \mathrm{D}}{\left[\partial r_{\ell}\right]^{2}}=\sum_{\mathrm{m}} \phi_{2 m}^{2} D_{\ell m}^{\prime \prime}+\sum_{\mathrm{m}} \sum_{\mathrm{n}} \phi_{\ell m} \phi_{\ell n} \frac{\partial^{2} \mathrm{D}}{\left\langle\mathrm{rr}_{\mathrm{m}} \partial r_{\mathrm{n}}\right.} \tag{13}
\end{equation*}
$$

Since $\frac{\partial^{2} D}{\partial r_{m} \partial r_{n}} \geq 0$, by setting $\frac{\partial^{2} D}{\partial r_{m} \partial r_{n}}$ to zero for mғn we obtain the lower bound

$$
\sum_{m} \phi_{2 m}^{2}\left(D_{l m}^{\prime \prime}+\frac{\partial^{2} D}{\left[\partial r_{m}\right]^{2}}\right) .
$$

By applying the Cauchy-Schwartz inequality in conjunction with (12) we also obtain

$$
\frac{\partial^{2} D}{\partial r_{m} \partial r_{n}} \leq \sqrt{\frac{\partial^{2} D}{\left[\partial r_{m}\right]^{2}}} \frac{\partial^{2} D}{\left[\partial r_{n}\right]^{2}} .
$$

Using this fact in (13) we obtain the upper bound

$$
\sum_{i n} \phi_{\ell m}^{2} D_{\ell m}^{\prime \prime}+\left(\sum_{m} \phi_{\ell m} \sqrt{\frac{\partial^{2} D}{\left[\partial r_{m}\right]^{2}}}\right)^{2}
$$

It is now easy to see that we have for all $k$

$$
\underline{R}_{\ell} \leq \frac{\partial^{2} D}{\left[\partial r_{\ell}\right]^{2}} \leq \bar{R}_{\ell}
$$

where $\underline{R}_{\ell}$ and $\overline{\mathrm{R}}_{\ell}$ are generated by

$$
\begin{align*}
& \underline{R}_{\ell}=\sum_{m} \phi_{\ell m}^{2}\left(D_{\ell m}^{\prime \prime}+R_{m}\right)  \tag{14}\\
& \bar{R}_{\ell}=\sum_{m} \phi_{\ell m}^{2} D_{\ell m}^{\prime \prime \prime}+\left(\sum_{m} \phi_{\ell m} \sqrt{\bar{R}_{m}}\right)^{2}  \tag{15}\\
& {\underset{A}{N}}^{R_{N}}=\bar{R}_{N}=0 \tag{16}
\end{align*}
$$

The computation is carried out by passing ${\underset{2}{2}}^{2}$ and $\overline{\mathrm{R}}_{2}$ upstream together with $\frac{\partial D}{\partial r_{\ell}}$ and this is well suited for a distributed algorithm. Upper and lower ${ }_{\text {bounds }} \Phi_{i \ell} \ell \bar{\Phi}_{i \ell}$ for $\frac{\partial^{2} D}{\left[\partial_{i \ell}\right]^{2}}, l \neq B(i ; \phi)$ are obtained simultaneoulsy by means of the equation [cf. (11)]

$$
\begin{align*}
& \Phi_{-i l}=t_{i}^{2}\left(D_{i \ell}^{\prime \prime}+R_{i}\right)  \tag{17}\\
& \Phi_{i \ell}=t_{i}^{2}\left(D_{i \ell}^{\prime \prime}+\bar{R}_{\ell}\right) . \tag{18}
\end{align*}
$$

It is to be noted that in some situations occuring frequently in practice the upper and lower bounds $\Phi_{-i \ell}$ and $\bar{\Phi}_{\text {il }}$ coincide and are equal to the true
second derivative. This will occur if $\phi_{\ell_{m a}} \phi_{n_{n}} \frac{\partial^{2} D}{\partial r_{\mathbb{I}} \partial_{n}}=0$ for $m \neq n$. For example if the routing pattern is as shown in Figure 1 (only links that carry flow are shown) then $\Phi_{i \ell}=\Phi_{i \ell}=\frac{\partial^{2} D}{\left[\partial \phi_{i \ell}\right]^{2}}$ for all (i, $\left.\ell\right) \varepsilon L$, $\ell \notin B(i ; \phi)$.


Figure 1

A typical case where $\bar{\Phi}_{i \ell} \neq \Phi_{i \ell}$ and the discrepancy affects materially the algorithm to be presented is when flow originating at $i$ splits and joins again twice on its way to $N$ as shown in Figure 2.


Figure 2

The Algorithm
The following algorithm seems to be a reasonable choice. If $\mathrm{t}_{\mathrm{i}} \neq 0$ we take $M_{i}^{k}$ in (9) to be the diagonal matrix with $\frac{1}{t_{i}^{2}} \bar{\phi}_{i \ell}, \ell E 0(i)$ along the diagonal whe::e $\bar{\Phi}_{i l}$ is the upper bound computed from ${ }_{(18)}$ and (14)-(16) and $\alpha$ is a positive scalar chosen experimentally. (In most cases $\alpha=1$ is satisfactory.)

Convergence of this algorithm can be easily established by verifying that the assumption of Proposition 2 is satisfied. A variation of the method results if we use in place of the upper bound $\bar{\phi}_{i \ell}$ the average of the upper and lower bounds $\frac{\bar{\Phi}_{i \ell}+\Phi_{i \ell}}{2}$. This however requires additional computation and commonication between modes.

$$
\begin{align*}
& \text { Problem (9) can be written for } t_{i} \neq 0 \text { as } \\
& \text { minimize } \\
& \sum_{\ell}\left\{\delta_{i \ell} \Delta \phi_{i \ell}+\frac{\bar{\phi}_{i \ell}}{2 \alpha t_{i}}\left(\Delta \phi_{i \ell}\right)^{2}\right\}  \tag{19}\\
& \text { subject to } \quad \Delta \phi_{i \ell} \geq-\phi_{i \ell}^{k}, \sum_{\ell} \Delta \phi_{i \ell}=0, \Delta \phi_{i \ell}=0 \quad \forall \ell \in B\left(i ; \phi^{k}\right)
\end{align*}
$$

and can be solved using a Lagrange multiplier technique. By introducing the expression (18) for $\bar{\Phi}_{i \ell}$ and carrying out the straightforward calculation we can write the corresponding iteration (8) as

$$
\begin{equation*}
\phi_{i \ell}^{k+1}=\max \left\{0, \phi_{i \ell}^{k}-\frac{\alpha\left(\delta_{i \ell}-\mu\right)}{t_{i}\left(D_{i \ell}^{\prime \prime}+\bar{R}_{\ell}\right)}\right\} \tag{20}
\end{equation*}
$$

where $\mu$ is a Lagrange multiplier determined from the condition

$$
\begin{equation*}
\sum_{\ell \neq B\left(i ; \phi^{k}\right)} \max \left\{0, \phi_{i \ell}^{k}-\frac{\alpha\left(\delta_{i \ell}-\mu\right)}{t_{i}\left(D_{i \ell}^{\prime \prime}+\bar{R}_{\ell}\right)}\right\}=1 \tag{21}
\end{equation*}
$$

The equation above is piecewise linear in the single variable $\mu$ and is nearly trivial computationally. Note from (20) that $\alpha$ plays the role of a stepsize parameter.

It can be seen that (20) is such that all routing variables $\phi_{i \ell}$ such that $\delta_{i \ell}<\mu$ will be increased or stay fixed at unity, while all routing variables $\phi_{i \ell}$ such that $\delta_{i \ell}>\mu$ will be decreased or stay fixed at zero. In particular the routing variable with smallest $\delta_{i \ell}$ will either be increased or stay fixed at unity, similarly as in Gallager's algorithm.
4. An Algorithm Based on an Upper Bound to Newton's Method

While the introduction of a diagonal scaling based on second derivatives alleviates substantially the problem of stepsize selection, it is still possible that in some iterations a unity stepsize will not lead to a reduction of the objective function and may even cause divergence of the algorithm of the previous section. This can be corrected by using a smaller stepsize as shown in Proposition 2 but the proper range of step!ize magritude depends on the network topology and may not be easy to detemmine. This dependence stems from the replacement of the Hessian matrix of $D$ by a diagonal approximation which in turn facilitates the computation of upper bounds to second derivatives in a distributed manner. Neglecting the off-diagonal terms of the Hessian means that while operating the algorithm for one destination we ignore changes which are caused by other destinations. The potential difficulties resulting from this can be alleviated (and for most practical problems eliminated) by operating the algorithm in a "one-at-a-time" version as discussed in Section 2. However the effect of neglecting the off-diagonal terms can still be detrimental in some situations such as the one depicted by Figure 3. Here $r_{1}=r_{2}=r_{3}=r_{4}>0, r_{5}=r_{6}=0$ and node 7


## -24-

is the only destination. If the algorithm of the previous section is applied to this example with $\alpha=1$, then it can be verified that each of the nodes $1,2,3$ and 4 will adjust its routing variables according to what would be Newton's method if all other variables remained unchanged. If we assume symmetric initial conditions and that the first and second derivatives $D_{57}^{\prime}$ s $D_{57}^{\prime \prime \prime}$ and $D_{67}^{\prime}, D_{67}^{\prime \prime}$ are much larger than the correspionding derivatives of all other links, then the algorithm would lead to a change of flow about four times larger than appropriate. Thus for example a value of $\alpha=1 / 4$ is appropriate, while $\alpha=1$ can lead tu divergence.

The algorithm proposed in this section bypasses these difficulties at the expense of additional computation per iteration. We show that if the initial flow vector is near optimal then the algorithm is gaaranteed to reduce the value of the objective function at each iteration and to converge to the optimum with a unity stepsize. The algorithm "upper bounds" a quadratic approximation to the objective function $D$. This is done by first making a trial change in the routing variabies using algorithm (8), (9). The link flows that would result from this change are then calculated going from the "most upstream" nodes downstream towards the destination. Based on the calculated flows the algorithm "senses" situations like the one in Figure 3 and automatically "reduces" the stepsize. He describe the algorithm for the case of a single destination (node N). The algorithm for the case of more than one destination consists of sequences of single destination iterations whereby all destinations are taken up cyclically (i.e. the one-at-a-time mode of operation).

## Tre Algorithm

At the typical iteration of the algorithm we have a vector of loopfree routing variables $\phi$ and a correspoliding flow vector $f$. We first carry out ateration (8), (9) with the choice of $\mathrm{M}_{\mathrm{i}}^{\mathrm{k}}$ described in Section 3 and a unity stepsize, and obtain a trial increment of routing variables denoted by $\Delta \phi^{*}$. Based on $\Delta \varphi^{*}$ we calculate the new (and final) increment of routing variables $\Delta \tilde{\phi}$ and the new routing vector

$$
\begin{equation*}
\bar{\phi}=\phi+\Delta \zeta \tag{22}
\end{equation*}
$$

by weans of a procedure of the following type. Esch node $i$ computes the corresponding vector of routing variable increeents $\Delta \phi_{i}$ by solving a problen of the form

$$
\begin{equation*}
\text { ninimize } \quad Q_{i}\left(\Delta \phi_{i}\right) \tag{23}
\end{equation*}
$$

subject to the constraints

$$
\begin{align*}
& \Delta \phi_{i \ell} \geq 0 \text { if } \Delta \phi_{i \ell}^{*}>0  \tag{24a}\\
& \Delta \phi_{i \ell} \leq 0 \text { if } \Delta \phi_{i \ell}^{*}<0  \tag{24b}\\
& \Delta \phi_{i \ell}=0 \text { if } \Delta \phi_{i \ell}^{*}=0  \tag{24c}\\
& \sum_{\phi} \Delta \phi_{i \ell}=0  \tag{24d}\\
& \phi_{i \ell}+\Delta \phi_{i \ell} \geq 0 \tag{24e}
\end{align*}
$$

where $Q_{i}\left(\Delta \phi_{i}\right)$ is a quadratic function of $\Delta \phi_{i}$ wich depends on $\phi$ and $A \phi^{*}$, and will be defined shortly. Notice that the constraint (24) guarantees that the new vector of routing variables $\tilde{\phi}$ is loopfree. In what follows He describe the procedure and rationale for obtaining the form of the quadratic function $Q_{i}$ of (23), and show that all computations can be
carried out in a distributed manner.
Let $\Delta f$ denote an increment of flow such that $f+\Delta f$ is feasible. A constrained version of Newton's method [3] is obtained if $\Delta f$ is chosen to minimize the quadratic objective function

$$
\begin{equation*}
N(\Delta f)=\sum_{i, \ell} D_{i \ell}^{\prime} \Delta^{f} i_{i \ell}+\frac{1}{2} \sum_{i, \ell} D_{i \ell}^{\prime \prime}\left(\Delta f_{i \ell}\right)^{2} \tag{25}
\end{equation*}
$$

subject to $f+\Delta f \varepsilon F$ where $F$ is the set of all feasible flow vectors. Let $\Delta \phi$ be the change in $\phi$ that corresponds to $\Delta f$. We write

$$
\bar{\phi}=\phi+\Delta \phi
$$

Finally let $t$ and $\Delta t$ be the vectors of tctal incoming traffic at the networl. nodes and corresponding changes [cf. (1)]. Then we have

$$
\begin{align*}
& \Delta t_{i}=\sum_{\ell} \Delta f_{\ell i}  \tag{26}\\
& \Delta f_{i \ell}=\Delta t_{i} \bar{\phi}_{i \ell}+t_{i} \Delta \phi_{i \ell} . \tag{27}
\end{align*}
$$

Substicting (27) in (25) we obtain

$$
\begin{align*}
N(\Delta f) & =\sum_{i, \ell} D_{i \ell}^{\prime} \Delta t_{i} \bar{\phi}_{i \ell}+\sum_{i, \ell} D_{i \ell}^{\prime} t_{i} \Delta \varphi_{i \ell}  \tag{28}\\
& +\frac{1}{2} \sum_{i, \ell} D_{i \ell}^{\prime \prime}\left[\left(\Delta t_{i} \bar{\phi}_{i \ell}\right)^{2}+2 \Delta t_{i} \bar{\phi}_{i \ell} t_{i} \Delta \phi_{i \ell}+\left(t_{i} \Delta \phi_{i \ell}\right)^{2}\right]
\end{align*}
$$

$$
\begin{align*}
& \bar{D}_{i \ell}^{\prime}=D_{i \ell}^{\prime}+D_{i \ell, i}^{\prime \prime} t_{i} \Delta \phi_{i \ell}  \tag{29}\\
& \bar{D}_{i}^{\prime}=\sum_{\ell} \bar{\phi}_{i \ell}\left(\bar{D}_{i \ell}^{\prime}+\bar{D}_{\ell \ell}^{\prime}\right), \bar{D}_{N}^{\prime}=0 \tag{30}
\end{align*}
$$

By multiplying (30) by $\Delta t_{1}$, summing over $i$ and using (27) we obtain

$$
\begin{aligned}
\sum_{i, \ell} \bar{D}_{i \ell}^{\prime} \bar{\phi}_{i \ell} \Delta t_{i} & =\sum_{i} \bar{D}_{i}^{\prime} \Delta t_{i}-\sum_{i, \ell} \bar{\phi}_{i \ell l} \bar{D}_{\ell}^{\prime} \Delta t_{i} \\
& =\sum_{i} \bar{D}_{i}^{\prime} \Delta t_{i}-\sum_{i, \ell} \Delta f_{i \ell} \bar{D}_{\ell}^{\prime}+\sum_{i, \ell} t_{i} \Delta \phi_{i \ell} \bar{D}_{\ell}^{\prime} \\
& =\sum_{i} \bar{D}_{i}^{\prime} \Delta t_{i}-\sum_{\ell} \Delta t_{\ell} \bar{D}_{\ell}^{;}+\sum_{i \ell \ell} i_{i} \Delta \phi_{i \ell} \bar{D}_{\ell}^{\prime} \\
& =\sum_{i, \ell}^{t_{i} \Delta \phi_{i \ell} \bar{D}_{\ell}^{\prime}}
\end{aligned}
$$

By using this relation together with (29) we can write (28) as

$$
\begin{align*}
N(\Delta f)= & \sum_{i}\left\{\sum_{\ell} t_{i} \Delta \phi_{i \ell}\left(D_{i \ell}^{\prime}+\bar{D}_{\ell}^{\prime}\right)\right. \\
& \left.+\frac{1}{2} \sum_{\ell} D_{i \ell}^{\prime \prime}\left[\left(\Delta t_{i} \bar{\phi}_{i \ell}\right)^{2}+\left(t_{i} \Delta \phi_{i \ell}\right)^{2}\right]\right\} \tag{31}
\end{align*}
$$

Now if $\left(\Delta t_{i}\right)^{2}$ were available then we could conceive of a recursive schene whereby node $i$ would obtain the vector $\Delta \phi_{i}$ which minimizes the corresponding term in the right hand side of (31) after receiving the value of $\overline{\mathrm{D}}_{\ell}$ from its downstream neighbors $\ell$, and in fact it can be seen that such a computation can be carried out in distributed fashion starting from the destination and proceeding upstream similarly as for algorithm (8), (9). Unfortunately $\left(\Delta t_{i}\right)^{2}$ depends on the values of $\Delta \phi_{m}$ for nodes $m$ that lie upstream of $i$. To bypass this difficulty we develop in what follows an upper bound for the troublesome term $\sum_{i, \ell} D_{i \ell}^{\prime \prime}\left(\Delta t_{i} \bar{\phi}_{i \ell}\right)^{2}$ by making use of the increment $\Delta \phi^{*}$ obtained through an iteration of algorithm ( 8 ), ( 9 ). When this upperbound is substituted in (31) we will obtain an upper bound to
$N(\Delta f)$ of the form

$$
N(\Delta f) \leq \sum_{i} t_{i} Q_{i}\left(\Delta \phi_{i}\right)
$$

where $Q_{i}\left(\Delta \phi_{i}\right)$ is precisely the expression to be used in the algorithn [cf. (23)].

Let us define for all $i=1, \ldots, N-1,(i, \ell) \varepsilon L$ and $\Delta \phi$ satisfying the constraint (24)

$$
\begin{align*}
& \Delta \phi_{\dot{i} \ell}^{+}=\max \left(0, \Delta \phi_{i \ell}^{*}\right), \quad \Delta \phi_{\dot{i} \ell}^{*}=\left|\min \left(0, \Delta \phi_{i \ell}^{*}\right)\right|  \tag{32}\\
& \Delta \phi_{i \ell}^{+}=\max \left(0, \Delta \phi_{i \ell}\right), \quad \Delta \phi_{i \ell}^{-}=\left|\min \left(0, \Delta \phi_{i \ell}\right)\right| \tag{33}
\end{align*}
$$

$$
\begin{align*}
& \Delta t_{i}^{\star_{i}^{-}}=\sum_{\ell}\left[t_{\ell} \Delta \phi_{\ell i}^{\star-}+\Delta t_{\ell}^{\star_{L}^{-}}\left(\varphi_{\ell i}+\Delta \phi_{l i \mathrm{i}}^{{ }^{-}}\right)\right]  \tag{35}\\
& \Delta \mathrm{t}_{\mathrm{i}}^{+}=\sum_{\ell}\left[\mathrm{t}_{\ell} \Delta \phi_{\ell i}^{+}+\Delta t_{\ell}^{+}\left(\phi_{\ell i}+\Delta \phi_{\ell i}^{+}\right)\right]  \tag{36}\\
& \Delta t_{i}^{-}=\sum_{\ell}\left[t_{\ell} \Delta \phi_{\ell i}^{-}+\Delta t_{\ell}^{-}\left(\phi_{\ell i}+\Delta \phi_{\ell i}^{-}\right)\right] . \tag{37}
\end{align*}
$$

The quantities $\Delta t_{i^{*}}^{+}, \Delta t_{i^{*}}^{-}$are well defined by virtue of the fact that the set of links

$$
L *=\left\{(i, \ell) \varepsilon L \mid \phi_{i \ell}>0, \text { or } \phi_{i \ell}+\Delta \phi_{i \ell}^{*}>0\right\}
$$

forms an acyclic network [in view of the manner that the sets of blocked nodes $B(\phi ; i)$ are defined in algorithm (8),(9)]. As a result $\Delta t_{i}^{+}{ }^{+}$and $\Delta t_{i}^{*^{-}}$are zero for all nodes $i$ which are the "most upstream" in this acyclic network. Starting from these nodes and proceeding downstream the computation of $\Delta t_{i}^{*^{+}}$and $\Delta t_{i}^{\star^{-}}$can be carried out in a distributed
manner for all nodes i using (34) and (35). Similarly [in view of the constraint (24)] the quantities $\Delta t_{i}^{+}, \Delta t_{i}^{-}$are well defined. It can be easily seen that for all i we have

$$
-\Delta t_{i}^{-} \leq \Delta t_{i} \leq \Delta t_{i}^{+}
$$

As a result it follows that

$$
\begin{equation*}
\left(\Delta t_{i}\right)^{2} \leq\left(\Delta t_{i}^{+}\right)^{2}+\left(\Delta t_{i}^{-}\right)^{2} \tag{38}
\end{equation*}
$$

We will develop upper bounds to the terms $\left(\Delta t_{i}^{+}\right)^{2}$ and $\left(\Delta t_{i}^{-}\right)^{2}$. To this end we need the following lemma the straightforward proof of which is left to the reader.

Lemma 1: Under the constraint (24)

$$
\begin{aligned}
& \Delta t_{i}^{\star}=0 \Rightarrow \Delta t_{i}^{+}=0, \forall i=1, \ldots, N-1 \\
& \Delta t_{i}^{\star}=0 \Rightarrow \Delta t_{i}^{-}=0, \quad \forall i=1, \ldots, N-1
\end{aligned}
$$

By using (36), (34) and the Cauchy-Schwartz inequality we obtain for all $i=7, \ldots, N-1$ with $\Delta t_{i}^{+}>0$

$$
\begin{align*}
& \left(\Delta t_{i}^{+}\right)^{2}=\left[\sum_{\ell}^{[ }\left(t_{\ell} \Delta \phi_{\ell i}^{+} * \Delta t_{\ell}^{+} \phi_{\ell i}+\Delta t_{\ell}^{+} \Delta \phi_{\ell i}^{+}\right]^{2}\right. \tag{39}
\end{align*}
$$

$$
\begin{aligned}
& \leq\left[\sum_{\ell} \frac{t_{l}\left(\Delta \phi_{\ell, i}^{+}\right)^{2}}{\Delta \phi_{\ell, i}^{*}}+\sum_{\ell} \frac{\left(\Delta t_{\ell}^{+}\right)^{2} \phi_{\ell i}}{\Delta t_{l}^{+}}+\sum_{\ell} \frac{\left(\Delta t_{l}^{+}\right)^{2}\left(\Delta \phi_{\ell i}^{+}\right)^{2}}{\Delta t_{l}^{*+} \Delta \phi_{\ell i}^{+}}\right] \Delta t_{i}^{*}
\end{aligned}
$$

where from sach summation above we exclude all nodes $\ell$ for which the corresponding denominator [and hence also the numerator by (24) and Lemma 1] is zero. Similarly we obtain

$$
\begin{equation*}
\left(\Delta t_{i}^{-}\right)^{2} \leq\left[\sum_{\ell} \frac{t_{\ell}\left(\Delta \phi_{\ell i}^{-}\right)^{2}}{\Delta \phi_{\ell i}^{*}}+\sum_{\ell} \frac{\left(\Delta t_{\ell}^{-}\right)^{2} \phi_{2 i}}{\Delta t_{\ell}^{*}}+\sum_{\ell} \frac{\left(\Delta t_{\ell}^{-}\right)^{2}\left(\Delta \phi_{\ell i}^{-}\right)^{2}}{\Delta t_{l}^{*}-\Delta \phi_{l i}^{*}}\right] \Delta t_{i}^{*} \tag{40}
\end{equation*}
$$

Define now for all $i=1, \ldots N-1$

$$
\begin{align*}
& {D_{i}^{\prime \prime}}^{\prime+}=\sum_{\ell}\left\{D_{i \ell}^{\prime \prime}\left(\phi_{i \ell}\right)^{2} \Delta t_{i}^{*}+D_{\ell}^{\prime \prime+}\left[\phi_{i \ell}+\frac{\left(\Delta \phi_{i \ell}^{+}\right)^{2}}{\Delta \phi_{i \ell}^{+}}\right]\right\}  \tag{41}\\
& D_{i}^{\prime \prime-}=\sum_{\ell}\left\{D_{i \ell}^{\prime \prime}\left(\phi_{i \ell}\right)^{2} \Delta t_{i}^{*}+D_{\ell}^{\prime \prime}\left[\phi_{i \ell}+\frac{\left(\Delta \phi_{i \ell}^{-}\right)^{2}}{\Delta \phi_{i \ell}^{*}}\right]\right\} \tag{42}
\end{align*}
$$

where the summation in (41) [(42)] is over all nodes 2 such that $\Delta \phi_{i \ell}^{+} \neq 0$
$\left(\begin{array}{c}\phi_{i \ell}^{\star} \\ \star^{-}\end{array} \neq 0\right)$. Define also

$$
\begin{equation*}
\mathrm{D}_{\mathrm{N}}^{1^{+}}=0, \quad \mathrm{D}_{\mathrm{N}}^{\mathbf{N}^{-}}=0 . \tag{43}
\end{equation*}
$$

Notice that given $\Delta \phi_{i \ell}$ and $D_{2}^{\prime \prime+}$ for all downstream neighbors $\ell$ ic is possible
 quantities $D_{i}^{0^{+}}, D_{i}^{10^{-}}$are well defined and can be computed recursively starting from the destination N and proceeding upstream in a distributed manner.

The following proposition yields the desired upper bound.
Proposition 3: Under the constraint (24) we have

$$
\begin{equation*}
N(\Delta f) \leq \sum_{i} t_{i} Q_{i}\left(\Delta \phi_{i}\right) \tag{44}
\end{equation*}
$$

where

$$
\begin{equation*}
Q_{i}\left(\Delta \phi_{i}\right)=\sum_{\ell}\left[\left(D_{i \ell}^{\prime}+\bar{D}_{\ell}^{\prime}\right) \Delta \phi_{i \ell}+\frac{1}{2}\left(t_{i} D_{i \ell}^{\prime \prime}+\beta_{i \ell}\right)\left(\Delta \phi_{i \ell}\right)^{2}\right] \tag{45}
\end{equation*}
$$

and

$$
\beta_{i \ell}= \begin{cases}\frac{D_{l}^{\prime+}}{\Delta \phi_{i \ell}^{+}} & \text {if } \Delta \phi_{i \ell}^{*}>0  \tag{46}\\ \frac{D_{l}^{\prime-}}{\Delta \phi_{i \ell}^{+}} & \text {if } \Delta \phi_{i \ell}^{*}<0 \\ 0 & \text { if } \Delta \phi_{i \ell}^{*}=0\end{cases}
$$

Proof: In view of (31) it will suffice to show that

$$
\begin{equation*}
\sum_{i, \ell} D_{i \ell}^{\prime \prime}\left(\Delta t_{i} \bar{\phi}_{i \ell}\right)^{2} \leq \sum_{i, \ell} t_{i} \beta_{i \ell}\left(\Delta \phi_{i \ell}\right)^{2} \tag{47}
\end{equation*}
$$

From (38) we have

$$
\begin{align*}
& \sum_{i, \ell} D_{i \ell}^{\prime \prime}\left(\Delta t_{i} \bar{\phi}_{i \ell}\right)^{2}=\sum_{i}^{\Gamma}\left(\Delta t_{i}\right)^{2} \sum_{\ell} D_{i \ell}^{\prime \prime}\left(\bar{\phi}_{i \ell}\right)^{2}  \tag{48}\\
& \leq \sum_{i}\left(\Delta t_{i}^{+}\right)^{2} \sum_{\ell} D_{i \ell}^{\prime \prime}\left(\bar{\phi}_{i \ell}\right)^{2}+\sum_{i}\left(\Delta t_{i}^{-}\right)^{2} \sum_{\ell} D_{i \ell}^{\prime \prime}\left(\bar{\phi}_{i \ell}\right)^{2} . \tag{41}
\end{align*}
$$

For all $i$ with $\Delta t_{i}^{+}>0$ we have, using Lemma $1, \Delta t_{i}^{*}>0$ so by dividing by $\Delta t_{i}^{*}{ }^{*}$ we obtain
$\sum_{i}\left(\Delta t_{i}^{+}\right)^{2} \sum_{\ell} D_{i \ell}^{\prime \prime}\left(\bar{\phi}_{i \ell}\right)^{2}=\sum_{i}\left(\Delta t_{i}^{+}\right)^{2}\left[\frac{D_{i}^{\prime 1^{+}}}{\Delta t_{i}^{*}}-\sum_{\ell} \frac{D_{\ell}^{\prime{ }^{+}}}{\Delta t_{i}^{+}}\left[\phi_{i \ell}+\frac{\left(\Delta \phi_{i \ell}^{+}\right)^{2}}{\Delta \phi_{i \ell}^{+}}\right]\right]$
$=\sum_{i} \frac{\left(\Delta t_{i}^{+}\right)^{2} D_{i}^{\prime 0^{+}}}{\Delta t_{i}^{*^{+}}}-\sum_{i, \ell} D_{\ell}^{\prime \prime}+\left[\frac{\left(\Delta t_{i}^{+}\right)^{2} \phi_{i \ell}}{\Delta t_{i}^{*}}+\frac{\left(\Delta t_{i}^{+}\right]^{2}\left(\Delta \phi_{i \ell}^{+}\right)^{2}}{\Delta t_{i}^{+}} \Delta \Delta \phi_{i \ell}^{+}{ }^{+}\right]$.

By using (39) we obtain

$$
\begin{gather*}
\sum_{i}\left(\Delta t_{i}^{+}\right)^{2} \sum_{\ell} D_{i \ell}^{\prime \prime}\left(\bar{\phi}_{i \ell}\right)^{2} \leq \sum_{i} \frac{\left(\Delta t_{i}^{+}\right)^{2} D_{i}^{\prime 1^{+}}}{\Delta t_{i}^{*}}-\sum_{\ell} \frac{\left(\Delta t_{l}^{+}\right)^{2} D_{l}^{1^{+}}}{\Delta t_{l}^{+}}+\sum_{i, \ell} \frac{D_{l}^{\prime \prime+} t_{i}\left(\Delta \phi_{i \ell}^{+}\right)^{2}}{\Delta \phi_{i \ell}^{+}} \\
=\sum_{i, \ell} \frac{D_{l}^{\prime \prime+} t_{i}\left(\Delta \phi_{i \ell}^{+}\right)^{2}}{\Delta \phi_{i \ell}^{+}}
\end{gather*}
$$

Similarly we obtain

$$
\begin{equation*}
\sum_{i}\left(\Delta t_{i}^{-}\right)^{2} \sum_{\ell} D_{i \ell}^{\prime \prime}\left(\bar{\phi}_{i \ell}\right)^{2} \leq \sum_{i, \ell} \frac{D_{\ell}^{\prime \prime-} t_{i}\left(\Delta \phi_{i \ell}^{-}\right)^{2}}{\Delta \phi_{i \ell}^{*}} \tag{50}
\end{equation*}
$$

By combining (48)-(50) and using the constraint (24) we obtain the desired relation (47).
Q.E.D.

The algorithm can now be completely defined. After the routing increment $\Delta \phi^{*}$ is calculated in a distributed manner by means of algorithm (8), (9), each node $i$ computes the quantities $\Delta t_{i}^{+}$and $\Delta t_{i}^{*}$. This is done recursively and in a distributed manner by means of equations (34), (35) starting from the "most upstream" nodes and proceeding downstream towards the destination. When this downstream propagation of information reaches the destination indicating that all nodes have completed the computation of $\Delta t_{i}^{*^{+}}$and $\Delta t_{i}^{*^{-}}$, the destination gives the signal for initiation of the second phase of the iteration which consists of computation of the actual routing increments $\Delta \tilde{\phi}_{i}$. To do this each node $i$ must receive the values of $\overline{\mathrm{D}}_{\ell}^{\prime}, \mathrm{D}_{\ell}^{\prime \prime+}$, and $\mathrm{D}_{\ell}^{\prime \prime-}$ fros its downstream neighbors $\ell$ and then detersine the increments $\Delta \tilde{\phi}_{i \ell}$ which minimize $Q_{i}\left(\Delta \phi_{i}\right)$ subject to the constraint (24) and the new routing variables

$$
\bar{\phi}_{i \ell}=\phi_{i \ell}+\Delta \tilde{\phi}_{i \ell}
$$

Then node $i$ proceeds to compute $\overline{D_{i}^{\prime}}, D_{i}^{\prime \prime+}$, and $D_{i}^{\prime \prime}{ }^{\prime \prime}$ via (30), (41), and (42)
and broadcasts these values to all upstream neighbors. Thus proceeding recursively upstream from the destination each node computes the actual routing increments $\Delta \tilde{\phi}_{i}$ in much the same way as the trial routing increments $\Delta \phi_{i}^{*}$ were computed earlier.

It is shown in Appendix $B$ that if the starting flow vector $f^{0}$ is sufficiently close to being optimal then the algorithr just described reduces the value of the objective function at each iteration and converges to the optimal value. He cannot expect to be able to guarantee theoretical convergence when the starting routing variables are far from optimal since this is not a generic property of Newton's method which the algorithm attampts to approximate. However in a large number of computational experiments with objective functions typically arising in communication networks and starting flow vectors which were far from optimal [5] we have never observed divergence or an increase of the value of the objective function in a single iteration. In any case it is possible to prove a global. convergence result for the version of the algorithm whereby the expression $Q_{i}\left(\Delta \phi_{i}\right)$ is replaced by
$Q_{i}^{\alpha}\left(\Delta \phi_{i}\right)=\sum_{\ell}\left[\left(D_{i \ell}^{\prime}+\bar{D}_{\ell}^{\prime}\right) \Delta \phi_{i \ell}+\frac{1}{2 \alpha}\left(\tau_{i} D_{i \ell}^{\prime \prime}+\beta_{i \ell}\right)\left(\Delta \phi_{i \ell}\right)^{2}\right]$
where $\alpha$ is a sufficiently small positive scalar stepsize. In Appendix B we show that by choosing $\alpha$ sufficiently swall it is possible to quarantee a reduction of the objective function at each iteration for any starting point $\phi^{0} \varepsilon \Phi$. This fact can be used to prove a convergence result similar to the one of Proposition 2.

## Appendix A: Proof of Proposition 2

The proof of proposition 2 to be given in this appendix applies to the "all-at-once" vetsion of algorithm (8), (9), i.e. the one where at each iteration $k$ every node $i$ solves problen (9) for all destinations $j$ and adjusts the corresponding routing variables according to (8). A nearly identical proof applies to the "one-at-a-time" version (see Gafni [11]). The destination of flows, routing variables, etc. will be denoted within parentheses. Thus for example $\phi_{i \ell}(j)$ denotes the routing variable of lini: ( $i, \ell$ ) for destination $j$.

The following llema bears close similarity in woth statement and proof as Lema 5 of Gallager [1]. The proof sili be critari, but may be found in [11].

Lemma A.1: Let the assumptions of Proposition 2 hid. There exists a scalar $\bar{\alpha} \in(0,1]$ (deperding on $D_{0}, \lambda$, and $\Lambda$ ) such that, for every $\alpha \in(0, \bar{\alpha}]$, the corresponding sequence $\left\{\oint^{k}\right\}$ generated by algorithm (8), (9) satisfies

$$
\begin{align*}
& D\left(\phi^{k+1}, r\right)-D\left(\phi^{k}, r\right) \leq-\rho \sum_{i, j}\left[t_{i}^{k}(j)\right]^{2}\left|\Delta \phi_{i}^{k}(j)\right|^{2}, \quad k=0,1, \ldots  \tag{A.1}\\
& \lim _{k \rightarrow \infty} t_{i}^{k}\left(j ;\left|\Delta \phi_{i}^{k}(j)\right|=0, \quad \forall i, j=1,2, \ldots, N, i \neq j\right.  \tag{A.2}\\
& \lim _{k \rightarrow \infty}\left|f_{i \ell}^{k+1}(j)-f_{i \ell}^{k}(j)\right|=0, \quad v(i, \ell) \varepsilon L, i, j=1,2, \ldots, N, i \neq j \tag{A.3}
\end{align*}
$$

where $\rho$ is some positive scalar (depending on $\left.\alpha, D_{0}, \lambda, \Lambda\right), t_{i}^{k}(j)$ denotes the total traffic arriving at node $i$ which is destined for $j$ when the routing is $\phi^{k}, \Delta \phi_{i}^{k}(j)=\phi_{i}^{k+1}(j)-\phi_{i}^{k}(j)$, and $f_{i \ell}^{k}(j), f_{i \ell}^{k+1}(j)$ are the flows
on link ( $i, \ell$ ) destined for $j$ and corresponding to $\phi^{k}, \phi^{k+1}$ respectively.
The following leman provides a key fact.
Lemma A.2: If $\alpha \in(0, \bar{\alpha}]$ where $\bar{\alpha}$ is as in Lemma A. 1 and $\left\{\phi^{k}\right\}$ is a corresponding sequence generated by algorithm (8), (9) there holds

$$
\begin{equation*}
\lim _{k \rightarrow \infty}\left[\Delta_{i}^{k}(j)-\Delta_{i}^{k}(j)\right]=0, \quad \forall i, j=1, \ldots, N, i \neq j \tag{A.4}
\end{equation*}
$$

where for all $i, j, k$

$$
\begin{align*}
& \Delta_{i}^{k}(j)=\max _{\ell}\left\{\delta_{i \ell}^{k}(j) \mid \ell \varepsilon 0(i), \phi_{i \ell}^{k+1}(j)>0\right\}  \tag{A.5}\\
& \Delta_{i}^{k}(j)=\min _{\ell}\left\{\delta_{i \ell}^{k}(j) \mid \ell \varepsilon 0(i), \ell \in B\left(i, \phi^{k}\right)(j)\right\}  \tag{A.6}\\
& \delta_{i \ell}^{k}(j)=D_{i \ell}^{\prime}\left(f_{i \ell}^{k}\right)+\frac{\hat{2}\left(\phi^{k}, r\right)}{\partial r_{i}(j)} . \tag{A.7}
\end{align*}
$$

Proof: From a necessary condition for optimality for problem (9) we obtain

$$
\begin{equation*}
\left[\delta_{i}^{k}(j)+\frac{t_{i}^{k}(j)}{\alpha} M_{i}^{k}(j) \Delta \phi_{i}^{k}(j)\right]^{T}\left[\phi_{i}(j)-\phi_{i}^{k+1}(j)\right] \geq 0 \tag{A.8}
\end{equation*}
$$

for all $\phi_{i}(j)$ which are feasible in problem (9). Let $\bar{\ell}$ and $\underline{\ell}$ be such that

$$
\delta_{i \dot{l}}^{k}(j)=\bar{\Delta}_{i}(j), \quad \delta_{i \underline{l}}^{k}(j)=\Delta_{i}^{k}(j)
$$

If $\bar{\ell} \neq \underline{\ell} w=$ define $\phi_{i}^{*}(j)$ to be the vector with components

$$
\phi_{i \ell}^{*}(j)= \begin{cases}\phi_{i \ell}^{k+1}(j)-\varepsilon & \text { if } \ell=\bar{\ell} \\ \phi_{i \ell}^{k+1}(j)+\varepsilon & \text { if } \ell=\ell \\ \phi_{i \ell}^{k+1}(j) & \text { otherwise }\end{cases}
$$

where $\varepsilon>0$ is small enough so that $\phi_{i}^{k+1}(j)-\varepsilon>0$. By definition of $\vec{\Delta}_{i}^{k}(j)$ such an $\varepsilon$ exists and by feasibility of $\phi_{i}^{k+1}(j)$ we have that $\phi_{i}^{*}(j)$ is also feasible. Substituting $\phi_{i}^{\star}(j)$ irs (A.8) in place of $\phi_{i}(j)$ we obtain

$$
\varepsilon\left[\Delta_{i}^{k}(j)-{\underset{\Delta}{i}}_{k}^{\Delta_{i}}(j)\right] \leq \frac{\varepsilon}{\alpha}\left[\mu_{i \ell}^{k}(j)-\mu_{i \ell}^{k}-(j)\right]
$$

where $\mu_{i \ell}^{k}(j)$ and $\mu_{i \ell}^{k}(j)$ are the $\underline{\ell}$ and $\bar{\ell}$ elements of the vector $t_{i}^{k}(j) M_{i}^{k}(j) \Delta \phi_{i}^{k}(j)$. Using the assumption that all elements of $M_{i}^{k}(j)$ are bounded above by $\Lambda$ we obtain

$$
0 \leq \frac{\Delta_{i}^{k}}{\Delta_{i}}(j)-\Delta_{i}^{k}(j) \leq \frac{2 \Lambda}{\alpha} t_{i}^{k}(j) \sum_{\ell}\left|\Delta \phi_{i \ell}^{k}(j)\right|
$$

This relation holds also it $\underline{\ell}=\bar{Z}$ since then $\vec{l}_{i} \dot{j}(j)=\Delta_{i}^{k}(j)$. From (A.2) we see that the right hand side tends to zero. Equation (A.4) follows. Q.E.D.

Given any set of routing variables $\phi \varepsilon \Phi$ there is a unique corresponding set of flows $f_{i \ell}(j)$. If we view the first derivative $D_{i \ell}^{\prime}\left(f_{i \ell}\right)$ as the length of link ( $i, \ell$ ) then the corresponding shortest distance from any node $i$ to any other node $j$ is well defined and will be denoted by $S_{i j}(\phi)$. It is easily seen using equation (6) that a sufficicnt. condition for optimaiity of a set of routing variables $\hat{\phi} \varepsilon \Phi$ is

$$
\begin{equation*}
S_{i j}(\hat{\phi})=\frac{\partial \hat{\eta}(\hat{\phi}, r)}{\partial r_{i}(j)} \quad, \quad V \ldots, j=1, \ldots, N_{z} i \neq j \tag{A.9}
\end{equation*}
$$

Furthermore there holds

$$
\begin{equation*}
S_{i j}(\phi) \leq \frac{\partial D(\phi, r)}{\partial r_{i}(j)}, \forall \phi \varepsilon \Phi \text { and } i, j=1, \ldots, N, i \neq j \tag{A.10}
\end{equation*}
$$

We have the following lemma:
Lema A.3: If $\alpha \in(0, \bar{\alpha}]$ where $\bar{\alpha}$ is as in Lemma A.1, $\left\{\phi^{k}\right\}$ is a corresponding sequence of algorithm (8), (9), $m \geq 1$ is an integer, and $\tilde{K}$ is an infinite
index set such that the subsequences $\left\{\phi^{k}\right\}_{k \varepsilon} \tilde{K}^{\text {and }\left\{\phi^{k-m}\right\}_{k \varepsilon}^{K}} \tilde{K}^{\text {converge to }} \hat{\phi}$ and $\tilde{\phi}$ respectively then

$$
\begin{align*}
& f_{i \ell}(j)(\hat{\phi}, r)=f_{i \ell}(j)(\tilde{\phi}, r), \quad \forall(i, \ell) \varepsilon L, j=1, \ldots, N  \tag{A.11}\\
& S_{i j}(\hat{\phi})=S_{i j} \tilde{(\phi)} \quad, \quad \forall i, j=1,2, \ldots, N \tag{A.12}
\end{align*}
$$

Proof: Equation (A.11) follows from (A.3), and equation (A.12) follows from the fact that $S_{i j}(\phi)$ depends on $\phi$ only through the flows $f_{i \ell}(j)(\phi, r)$. Q.E.D.

We will use "two dimensional induction" to show that the limit of any convergent subsequence of $\left\{\phi^{k}\right\}$ satisfies the sufficient condition for optimailty (A.9). Leman A. 4 that follows represents the basic step of the induction proof. He use repeatedly the fact that if some property 1 holds for all $k$ with $k>k_{1}$ and some property 2 holds for all $k$ with $k>k_{2}$ then both hold for all $k$ with $k>\max \left(k_{1}, k_{2}\right)$. 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 A.4: Let $\alpha \in(0, \bar{\alpha}]$ where $\bar{\alpha}$ is as in Lemma A.1, let $\left\{\phi^{k}\right\}$ be a corresponding sequence generated by algorithm (3), (9) and let $\left\{\phi^{k-1}\right\}_{k \in \tilde{K}} \cdot \tilde{\phi}$ and $\left\{\phi^{k}\right\}_{k \varepsilon K} \tilde{K}+\hat{\phi}$ be two convergent subsequences of $\left\{\phi^{k}\right\}$. For each $j$ let $\left.S_{j} \tilde{(\tilde{\phi}}\right)$ be the set of distances $\left\{S_{i j}(\tilde{\phi}) \mid \quad i \in N\right\}$. Let $S_{1}(j), \ldots, S_{p}(j), p \leq N$ be the distinct elements of the set $S_{j}(\tilde{\phi})$ and assume without loss of generality that $0=S_{1}(j)<S_{2}(j)<\ldots<S_{p}(j)$. Denote

$$
\begin{equation*}
I_{q}(j)=\left\{i \mid S_{i j}(\tilde{\phi}) \leq S_{q}(j)\right\}, \quad q=1, \ldots, p \tag{A.13}
\end{equation*}
$$

Assume that for some integer $q$ we have:
a) $\frac{\partial D(\hat{\phi}, x)}{\partial r_{i}(j)}=\frac{\partial D(\tilde{\phi}, r)}{\partial r_{i}(j)}=S_{i j}(\tilde{\phi})$ ai $\varepsilon I_{q}(j), j=1, \ldots, N$
b) For all $k$ large enough, $k \in \tilde{k}$, and for any $j$, if $\phi_{\min }^{k-1}(j)>0$ and m $\varepsilon I_{q}(j)$ then $\frac{\partial D\left(\phi^{k-1}, r\right)}{\partial r_{m}(j)}>\frac{\partial D\left(\phi^{k-1}, r\right)}{\partial r_{n}(j)}$.

Then:
a')

$$
\begin{equation*}
\frac{\partial D(\hat{\phi}, r)}{\partial r_{i}(j)}=S_{i j}(\hat{\phi}) \quad \forall i \varepsilon I_{q+1}(j), j=1, \ldots, N \tag{A.15}
\end{equation*}
$$

$b^{\prime}$ ) For all $k$ large enough, $k \in \tilde{K}$, and for any $j$, if $\phi_{\min }^{k}(j)>0$ and $m \in I_{q+1}(j)$ then $\frac{\partial D\left(\phi^{k}, r\right)}{\partial r_{m}(j)}>\frac{\partial D\left(\phi^{k}, r\right)}{\partial r_{n}(j)}$.

Proof: Let $i$ be such that $i \varepsilon I_{q+1}(j)$
and denote

$$
\ell_{i}(j)=\left\{\ell \mid s_{i j}(\tilde{\phi})=D_{i \ell}^{1}\left[f_{i \ell}(\tilde{\phi}, r)\right]+s_{\ell j}(\tilde{\phi}), \ell \in 0(i)\right\} .
$$

By the definition of shortest distance we have

$$
S_{i j}(\tilde{\phi})<D_{i \ell}^{\prime}\left[f_{i \ell}(\tilde{\phi}, r)\right]+S_{\ell j} \tilde{(\underline{\phi})} \quad \forall \ell \notin \ell_{i}(j), \ell \varepsilon 0(\mathrm{i}) .
$$

Using (A.10) and the above equation
$s_{i j} \tilde{(\phi)}<D_{i \ell}^{\prime}\left[f_{i \ell}(\tilde{\phi}, r)+\frac{\partial D(\tilde{\phi}, r)}{\partial r_{\rho}(j)}\right.$
or equivalently
$\mathrm{S}_{\mathrm{ij}}(\tilde{\phi})<\delta_{i \ell}(\mathrm{j})(\tilde{\phi}, \mathrm{r}) \quad \forall \ell \notin \ell_{\mathrm{i}}(\mathrm{j}), \ell \varepsilon \sigma(\mathrm{i})$.
By the assumption $D_{i \ell}>0$ and the fact $i \varepsilon I_{q+1}(j)$, we have

$$
\ell_{i}(j) \subset I_{q}(j) \quad j=1, \ldots, N
$$

Therefore by using hypothesis a) we have

$$
\begin{align*}
\delta_{i \ell}(j) \tilde{(\phi, r)} & =D_{i \ell \ell}^{\prime}\left[f_{i \ell}(\tilde{\phi}, r)\right]+\frac{\partial D(\tilde{\phi}, r)}{\partial r_{\ell}(j)}=D_{i \ell}^{\prime}\left[f_{i \ell}(\tilde{\phi}, r)\right]+S_{\ell j}(\tilde{\phi})  \tag{A.16}\\
& =S_{i j}(\tilde{\phi}) \quad \forall \ell \varepsilon \ell_{i}(j), \quad j=1, \ldots, N .
\end{align*}
$$

Since $O$ (i) is a finite set there exists $\varepsilon>0$ s:ch that $\delta_{i w}(j)(\tilde{\phi}, r)-\varepsilon>\delta_{i \ell}(j)(\tilde{\phi}, r) \quad \forall w \in \ell_{i}(j), w \in 0(i), \ell \varepsilon \ell_{i}(j), j=1, \ldots, N$

Since $\delta_{i 2}(j)(\phi, T)$ is continuous in $\phi$ and $\left\{\phi^{k-1}\right\}_{k \varepsilon} \tilde{\gamma}$. is a convergent sequence, we get that for all $k$ large enough, $k \varepsilon \tilde{K}$
$\delta_{j_{w}}(j)\left(\phi^{k-1}, r\right)>\delta_{i \ell}\left(\phi^{k-1}, r\right)+\frac{\varepsilon}{2} \quad \forall \geqslant \neq \ell_{i}(i), \operatorname{weO}(i)=\ell \varepsilon \ell_{i}(j), j=1, \ldots, N$.

Also $\frac{\partial D(\phi, r)}{\partial r_{i}(j)}, 1 \leq i, j \leq N$, is continuous in $\phi$ and therefore by Leman
2, (1.16) and hypothesis a), for all k large enough, $k \varepsilon \vec{K}$

$$
\frac{\partial D\left(\phi^{k-1}, r\right)}{\partial r_{i}(j)}>\frac{\partial D\left(0^{k-1}, r\right)}{\partial r_{i}(j)} \quad \forall \ell \varepsilon \ell_{i}(j), \quad j=1, \ldots, N
$$

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

$$
\begin{equation*}
2_{i}\left(j ; \cap B\left(\phi^{k-1} ; i\right)(j)=f, \quad j=1, \ldots, N .\right. \tag{A.18}
\end{equation*}
$$

Lemma A. 2 combined with (A.17) and (A.18) implies that for all $k$ large e:.ough $k \in \tilde{K}$

$$
\begin{equation*}
\phi_{i k}^{k}(j)=0 \quad \forall w \in \hat{t}_{i}(j) ; \quad j=1, \ldots, N \tag{A.19}
\end{equation*}
$$

and taking the limit

$$
\begin{equation*}
\hat{\phi}_{i w}(j)=0 \quad w \in l_{i}(j), \quad j=1, \ldots, N \tag{A.20}
\end{equation*}
$$

Using (A.20), Lemon A. 3 and hypothesis a) we have

$$
\begin{aligned}
\frac{\partial L(\hat{\phi}, r)}{\partial r_{i}(j)} & =\sum_{\ell} \hat{\phi}_{i \ell}(j)\left[D_{i k}^{\prime}\left[f_{i k}(\hat{\phi}, r)\right]+\frac{\partial D(\hat{\phi}, r)}{\partial r_{\ell}(j)}\right] \\
& =\sum_{\ell \varepsilon \ell_{i}(j)} \hat{\phi}_{i \ell}(j)\left[D_{i k}^{\prime}\left[f_{i k}(\hat{\phi}, r)\right]+\frac{\partial D(\hat{\phi}, r)}{\partial r_{\ell}(j)}\right] \\
& =\sum_{\ell \hat{\varepsilon}_{i}(j)} \hat{\phi}_{i \ell}(j)\left[D_{i k}^{\prime}\left[f_{i k}(\tilde{\phi}, r)\right]+\frac{\partial D(\tilde{\phi}, r)}{\partial r_{\ell}(j)}\right] \\
& =S_{i j}(\tilde{\phi})=S_{i j}(\hat{\phi})
\end{aligned}
$$

This together with part a) of the hypothesis establishes a').
To see $b^{\prime}$ ), notice that by continuity of $\frac{G D(\phi, r)}{\partial x_{i}(j)}$ in $\phi$ and the preceding equation we have that for all $k$ las: nough, $k \in \tilde{K}$

$$
\begin{equation*}
\frac{\partial D\left(\phi^{k}, r\right)}{\partial r_{i}(j)}>\frac{\partial D\left(\phi^{k}, r\right)}{\partial r_{\ell}^{(j)}} \quad \forall \ell \varepsilon \ell_{i}(j), \quad j=1, \ldots, N . \tag{A.21}
\end{equation*}
$$

Equations (A.21) and (A.19) hold for all $i \varepsilon I_{q+1}(j)$ and $b^{\prime}$ ) follows.
Q.E.D.

By now we have developed all the machinery for the convergence proof of Proposition 2. He will simply make repeated application of Lerinna A. 4 for the proper sequences.

Proof of Proposition 2: Take $\bar{\alpha}$ to be as in Lemma A.1, let $\alpha \in(0, \bar{\alpha}]$ and let $\left\{\phi^{k}\right\}$ be a corresponding sequence generated by algorithm (8), (9). The sequence $\left\{\phi^{k}\right\}$ belongs to a compact set and therefore there exists a convergent subsequence $\left\{\phi^{\mathrm{k}}\right\}_{k \varepsilon K} \rightarrow \phi$. The sequence $\left\{\phi^{\mathrm{k}-1}\right\}_{\mathrm{k} \varepsilon K}$ has a convergent subsequence $\left\{\phi^{k-1}\right\}_{k \varepsilon K_{1}} \rightarrow \phi_{1}, K_{1} \subset K$. The sequence $\left\{\phi^{k-2}\right\}_{k \in K_{1}}$ has a convergent subsequence $\left\{\phi^{k-2}\right\}_{k \in K_{2}} \rightarrow \phi_{2}, K_{2} \subset K_{1}$. Proceeding this way k 2 get a convergent subsequence

$$
\left\{\phi^{\mathrm{k}-\mathrm{N}+1}\right\}_{\mathrm{k} \in K_{\mathrm{N}-1}}+\phi_{\mathrm{N}-1}, K_{\mathrm{N}-1} \subset K_{\mathrm{N}-2}
$$

We have $K_{N-1} \subset \kappa_{N-2}^{\prime} \subset \ldots \subset K$ and

$$
\left\{\phi^{i \cdot-\mathrm{N}+1}\right\}_{\mathrm{k} \varepsilon K_{\mathrm{N}-1}} \rightarrow \phi_{\mathrm{N}-1}, \ldots,\left\{\phi^{\mathrm{k}-1}\right\}_{\mathrm{k} \varepsilon K_{\mathrm{N}-1}} \rightarrow \phi_{1},\left\{\phi^{\mathrm{k}}\right\}_{\mathrm{k} \varepsilon K_{\mathrm{N}-1}} \rightarrow \phi
$$

By Lemma A. 3 the shortest distances which correspond to $\phi_{\mathrm{N}-1}, \phi_{\mathrm{N}-2}, \ldots, \phi$ are the same. As a result, in what follows, when we mention the set $I_{q}(j)$ we need not specify the limit point $\dot{\varphi}_{i}$ to which it corresponds.

Let $\tilde{K}$ in Lemma $A .4$ be $K_{N-1}$. For each destination $j$, the only element in $I_{1}(j)$ is $j$ and therefore the assumptions of Lenma A. 4 hold for $I_{1}(j)$ and the rairs of sequences


Applying Lemma A. 4 for $q=1$, we obtain that its hypothesis holds for $q=2$ and the pairs of sequences $\left(\left\{\left\{\phi^{k}\right\}_{k \varepsilon \tilde{K}},\left\{\phi^{k-1}\right\}_{k \varepsilon} \tilde{K}\right], \ldots\right.$,
$\left.\left[\left\{\phi^{k-N+3}\right\}_{k \in K} \tilde{K},\left\{\phi^{k-N+2}\right\}_{k \in \tilde{K}}\right]\right)$. Proceeding this way we note that the
hypothesis of Lemma. 4.4 holds for $q=N-1$ and the pair $\left(\left\{\phi^{k}\right\}_{k \varepsilon K^{\prime}}\left\{\phi^{k-1}\right\}_{k \varepsilon K}\right)$. Applying Lemma A. 4 again we see that the conclusion of its part $a^{\prime}$ ) holds for $q=N$-1, i.e., equation (A.15) holds for $I_{N}(j), j=1, \ldots, N$. Since every node in the network belongs to $I_{N}(j), j=1, \ldots, N$, it follows that (A.9) is satisfied, and hence $\phi$ is optimal.
Q.E.D.

## Appendix B

In this appendix we analyze the descent properties of the algorithm of Section 4. We assume a single destination but the proof extends trivially to the case where we have multiple destinations and the algorithm is operated in the one destination at a time mode. In view of the fact that each function $D_{i \ell}$ is strictly convex it follows that there is a unique optimal set of total link flows $\left\{f_{i \ell}^{*} \mid(i, \ell) \varepsilon L\right\}$. It is clear that given any $\varepsilon>0$ there exists a scalar $\gamma_{\varepsilon}$ such that for all feasible total link flow vectors $f$ satisfying

$$
\begin{equation*}
\left|f_{i \ell}-f_{i \ell}^{\star}\right| \leq Y_{\varepsilon} \quad, \quad \forall(i, \ell) \varepsilon L \tag{B.1}
\end{equation*}
$$

we have

$$
\begin{equation*}
\frac{1}{1+\varepsilon} D_{i . \ell}^{\prime \prime}\left(f_{: \ell}^{*}: \leq D_{i \ell}^{\prime \prime}\left(f_{i \lambda}\right) \quad \dot{s} \cdot \bar{c}\right) D_{i \ell}^{\prime \prime}\left(f_{i \ell}^{*}\right), \quad \forall(i, \ell) \varepsilon L . \tag{B.C}
\end{equation*}
$$

The strict positiviey assumpeion $r$ " $\gamma_{\varepsilon}>0$ there $\in x_{i} s t s$ q scalar $j\left(Y_{\varepsilon}\right)$ ach thai every feasible fatisfying $\sum_{i, \ell} D_{i \ell}\left(f_{i \ell}\right) \leq \delta\left(\gamma_{\varepsilon}\right)$ aise secisties (B.1) and hence also (B.2). Furthermore $\delta\left(\gamma_{\varepsilon}\right)$ can be taken arbitrarily large provided $\gamma_{\varepsilon}$ is sufficiently large. We will make use of this fact in the proof of the subsequent result.

Proposition B.1: Let $\phi$ and $\bar{\phi}$ be two successive vectors of routing variables generated by the algorithm of Section 4 (with stepsize $\alpha=1$ ) and let $f$ and $\bar{f}$ be the corresponding vectors of link flows. Assume that for some $\varepsilon$ with $0<\varepsilon<\frac{2}{\sqrt{3}}-1$ we have

$$
\begin{equation*}
\sum_{i, 0} D_{i \ell}\left(f_{i \ell}\right) \leq \delta\left(\gamma_{\varepsilon}\right) \tag{B.3}
\end{equation*}
$$

where $\gamma_{\varepsilon}$ is the scalax corresponding to $\varepsilon$ as in (B.1), (B.2), and $\delta\left(\gamma_{\varepsilon}\right)$ is such tha: (B.1) [and hence also (B.2)] holds for all feasible $f$ satisfying (B.3). Then
$D(\bar{\phi}, r)-D(\phi, r) \leq-\rho(\varepsilon) \sum_{(\bar{i}, \ell) \varepsilon L} t_{i}\left(D_{i \ell}^{\prime \prime}+\beta_{i \ell}\right)\left(\bar{\phi}_{i \ell}-\phi_{i \ell}\right)^{2}$
where $\rho(\varepsilon)=\frac{1-4 \varepsilon-2 \varepsilon^{2}}{2}>0$ for all $\varepsilon$ with $0<\varepsilon<\sqrt{\frac{3}{2}}-1$.
Proof: Let $\Delta \tilde{f}$ be the increment of flow corresponding to the increment $\Delta \tilde{\phi}=\bar{\phi}-\phi$. We have
$D(\bar{W}, I)-D(\phi, r)=\left.\sum_{i, \ell} \Delta f_{i \ell} D_{i \ell}^{\prime}\left(f_{i \ell}+\eta \Delta f_{i \ell}\right)\right|_{\eta=0}+\left.\frac{1}{2} \sum_{i, \ell}\left(\Delta f_{j \ell}\right)^{2} D_{i \ell}^{\prime \prime}\left(f_{i \ell}+\eta \Delta f_{i \ell}\right)\right|_{\eta=n}{ }^{*}$
for some $n^{\star} \varepsilon[0,1]$. Denoting $D_{i \lambda}^{\prime \prime}\left(f_{i \ell}+\eta^{\star} \Delta f_{i \ell}\right)=\hat{\mathrm{p}}_{i \ell}^{\prime \prime}$ and using an argument similar to the one employed in Section 4 [cf. (28)-(31)] we obtain

$$
\begin{align*}
D(\bar{\phi}, r)-D(\phi, r)= & \sum_{i, \ell} t_{i} \tilde{\hat{\phi}}_{i \ell}\left(D_{i \ell}^{\prime}+\bar{D}_{\ell}^{\prime}\right)+\sum_{i, \ell}\left(\hat{D}_{i \ell}^{\prime \prime}-D_{i \ell}^{\prime \prime}\right) \Delta t_{i} \bar{\phi}_{i \ell} t_{i} \Delta \tilde{\phi}_{i \ell} \\
& +\frac{1}{2} \sum_{i, 2} \hat{D}_{i \ell}^{\prime \prime}\left[\left(\Delta t_{i} \bar{\phi}_{i \ell}\right)^{2}+\left(t_{i} \Delta \tilde{\phi}_{i \ell}\right)^{2}\right] \tag{B.5}
\end{align*}
$$

We will derive upper bounds fo: each of the three terms in the right side of (B.5).

From the necessary condition for $\Delta \tilde{\phi}_{i}$ to minimize the function $Q_{i}\left(\Delta \phi_{i}\right)$ of (45) subject to the constraint (24) we obtain

$$
\sum_{\ell}\left[D_{i \ell}^{\prime}+\bar{D}_{2}^{\prime}+\left(t_{i} D_{i \ell}^{\prime \prime}+\beta_{i \ell}\right) \tilde{\phi}_{i \ell}\right] \Delta \tilde{\phi}_{i \ell} \leq 0
$$

or

$$
\begin{equation*}
\sum_{\ell}\left(D_{i \ell}^{\prime}+\bar{D}_{\ell}^{\prime}\right) \tilde{\phi}_{i \ell} \leq-\sum_{\ell}\left(t_{i} D_{i \ell}^{\prime \prime}+\beta_{i \ell}\right)\left(\Delta \tilde{\phi}_{i \ell}\right)^{2} \tag{B.6}
\end{equation*}
$$

There is no loss of generality in replacing each function $D_{i \ell}$ by
a function $\overline{\mathrm{D}}_{\mathrm{i}}$ which is continuously differentiable, is identical with $\mathrm{D}_{\mathrm{i} \ell}$ on the set of flows satisfying (B.1) and is quadratic outside this set, provided that, as part of the subsequent proof, we show that
$\sum_{i, \ell} D_{i \ell}\left(f_{i \ell}+n \Delta f_{i \ell}\right) \leq \delta\left(\gamma_{\varepsilon}\right)$ for all $n \varepsilon[0,1]$. By using this device we can assume that $D_{i \ell}^{\prime \prime}$ satisfies (B.2) for all $f_{i \ell}$. Hence from (B.2)

$$
\begin{align*}
& \frac{\hat{D}_{i \ell}^{\prime \prime}-D_{i \ell}^{\prime \prime}}{D_{i \ell}^{\prime \prime}} \leq(1+\varepsilon)^{2}-1  \tag{B.7}\\
& \hat{D}_{i \ell}^{\prime \prime}  \tag{3.8}\\
& \frac{D_{i \ell}^{\prime \prime}}{D_{i \ell}^{\prime \prime}} \leq(1+\varepsilon)^{2} .
\end{align*}
$$

Using (47), (B.7), the Cauchy-Schwartz inequality and the arithsetic-g:metric inequality we have

$$
\begin{align*}
& \sum_{i, \ell}\left(\hat{D}_{i \ell}-D_{i \ell}^{\prime \prime}\right) \Delta t_{i} \bar{\phi}_{i \ell}{ }^{t}{ }_{i} \Delta \phi_{i \ell} \leq\left[(1+\varepsilon)^{2}-1\right] \sum_{i, \ell} D_{i \ell}^{\prime \prime}\left|\Delta t_{i}\right| \bar{\phi}_{i \ell} t_{i}\left|\Delta \tilde{\phi}_{i \ell}\right| \\
& \leq\left[(1+\varepsilon)^{2}-1\right]\left[\sum_{i \ell} D_{i \ell}^{\prime \prime}\left(\Delta t_{i} \bar{\phi}_{i \ell}\right)^{2}\right]^{\frac{1}{2}}\left[\sum_{i, \ell} D_{i \ell}^{\prime \prime}\left(t_{i} \Delta \tilde{\phi}_{i \ell}\right)^{2}\right]^{\frac{1}{2}} \\
& \leq\left[(1+\varepsilon)^{2}-1\right]\left[\sum_{i \ell} \beta_{i \ell} t_{i}\left(\Delta \tilde{\phi}_{i \ell}\right)^{2}\right]^{\frac{1}{2}}\left[\sum_{i, \ell} D_{i \ell}^{\prime \prime}\left(t_{i} \Delta \tilde{\phi}_{i \ell}\right)^{2}\right]^{\frac{1}{2}}  \tag{B.9}\\
& \leq \frac{1}{2}\left[(1+\varepsilon)^{2}-1\right]\left[\sum_{i, \ell} \beta_{i \ell} \tau_{i}\left(\Delta \tilde{\phi}_{i \ell}\right)^{2}+\sum_{i, \ell} D_{i \ell}^{\prime \prime}\left(t_{i} \Delta \tilde{\phi}_{i \ell}\right)^{2}\right] \\
& =\frac{1}{2}\left[(1+\varepsilon)^{2}-1\right] \sum_{i} t_{i} \sum_{\ell}\left(t_{i} D_{i \ell}^{\prime \prime}+\beta_{i \ell}\left(\left(\Delta \dot{\phi}_{1 \ell}\right)^{2} .\right.\right.
\end{align*}
$$

Using again (47) and (B.8) we obtain for each i

$$
\begin{align*}
& \int_{i, \ell} \hat{D}_{i \ell}^{\prime \prime}\left[\left(\Delta t_{i} \bar{\phi}_{i \ell}\right)^{2}+\left(t_{i} \Delta \tilde{\phi}_{i \ell}\right)^{2} j\right. \\
& \leq(1+\varepsilon)^{2} \sum\left[D_{i \ell}^{\prime \prime}\left(\Delta t_{i} \bar{\phi}_{i \ell}\right)^{2}+D_{i \ell}^{\prime \prime}\left(t_{i} \Delta \tilde{\phi}_{i \ell}\right)^{2}\right]  \tag{B.10}\\
& \leq(1+\varepsilon)^{2}\left[\left[t_{i} \beta_{i \ell}\left(\Delta \tilde{\phi}_{i \ell}\right)^{2}+D_{i \ell}^{\prime \prime}\left(t_{i} \Delta \tilde{\phi \phi}_{i \ell}\right)^{2}\right]\right. \\
&=(1+\varepsilon)^{2} \sum_{i} t_{i} \sum_{\ell}\left(t_{i} D_{i}^{\prime \prime}+\beta_{i \ell}\right)\left(\tilde{\phi}_{i \ell}\right)^{2} .
\end{align*}
$$

By combining now (B.5), (B.6), (B.9), and (B.10) we obtain

$$
\begin{aligned}
D(\bar{\phi}, r)-D(\phi, r) & \leq\left[-1+\frac{(1+\varepsilon)^{2}-1}{2}+\frac{(1+\varepsilon)^{2}}{2}\right] \sum_{(i, \ell) \varepsilon L} t_{i}\left(t_{i} D_{i \ell}^{\prime \prime}+\beta_{i \ell}\right)\left(\Delta \tilde{\phi}_{i \ell}\right)^{2} \\
& =-D(\varepsilon) \sum_{(i, \ell) \varepsilon L} t_{i}\left(t_{i} D_{i \ell}^{\prime \prime}+\sigma_{i \ell}\right)\left(\bar{\phi}_{i \ell}-\phi_{i \ell}\right)^{2}
\end{aligned}
$$

and (B.4) is proved. It is also straightforward to verify that $\rho(\varepsilon)>0$ for $\varepsilon$ in the interval $\left(0, \sqrt{\frac{3}{2}}-1\right)$. Q.E.D.

The preceding proposition shows that the algorithm of Section 4 does not increase the value of the objective function once the flow vector $f$ enters a region of the form $\left\{f \mid \sum_{i, \ell} D_{i \ell}\left(\sum_{i \ell}\right) \leq \delta\left(\gamma_{\varepsilon}\right)\right\}$, and that the size of this region increases as the third derivative of $D_{i \ell}$ becomes smalier. Indeed if each function $D_{i \ell}$ is quadracic then (B.2) is satisfied for all $\varepsilon>0$ and the algorith will not increase the value of the objective for all $f$.

The preceding analysis can be easily modified to show that if we introduce a stepsize $\alpha$ as in (51) then the algorithm of Section 4 is a descent algoriths at all flows in the region $\left\{f \mid \sum_{i, \ell:} D_{i \ell}\left(f_{i \ell}\right) \leq \delta\left(\gamma_{\varepsilon}\right)\right\}$ where

$$
0<\varepsilon<\sqrt{\frac{2+\alpha}{2 \alpha}}-1
$$

From this it follows that given any starting point $\phi^{0} \varepsilon \Phi$, there exists a scalar $\bar{\alpha}>0$ such that for all stepsizes $\alpha \in(0, \bar{\alpha}]$ the algorithm of Section 4 does not increase the value of the objective function at each subsequent iteration.

## References

[1] Gallager, R.G., "A Minimum Delay Routing Algorithm Using Distributed Computation", IEEE Trans. on Communication, Vol. COM-25, 1977, pp. 73-85.
[2] Goldstein, À.A., "Convex Programming in Hilbert Space", Bull. Amer. Math. Soc., Vol. 70, 1964, pp. 709-710.
[3] Levitin, E.S. and B.T. Polyak, "Constrained Minimization Problems", USSR Comput. Math. Math. Phys., Vol. 6, 1966, pp. 1-50.
[4] Bertsekas, D.P., "Algorithms for Nonlinear Multicommodity Network Flow Problems", in Proc. of International Symposium on Systems Optimization and Analysis, A. Beñsoussan and J.L. Lions (eds.), Springer-Verlag, 1979, Pp. 210-224.
[5] Rertsekas, D.P., Gafni, E., and Vastola; K.S., "Validation of Algorithms for Optimal Routing of Flow in Networks", Proc. of IEEE Conf. on Decision and Control, San Diego, Calif., Jan. 1979, pp. 220-227.
[6] Vastola, K.S., "A Numerical Study of Two Measures of Delay for Net. work Routing", M.S. Thesis, Dept. of Electrical Engineering, Univ. of I11, Urbana, I11., Sept. 1979.
[7] Bertsekas, D.P., "A Class of Optimal Routing Algorithms for Communication Networks", Proc. of the Fifth International Conference on Computer Communication (ICCC-80), Atlanta, Ga., Oct. 1980, pp. 71-76.
[8] Bertsekas, D.P., and Gafni, E., "Projection Mechods for Variational Inequalities with Application to the Traffic Assignment Problen", Report LIDS-P-1043, Mass. Institute of Technology, Cambridge, Mass., June 1980, Nath. Programing Studies, (to appear).
[9] Golestaani, S.J., "A Unified Theory of Flow Control and Rcuting in Data Communication Networks", Report LIDS-TH-963, Laboratory for Information and Decision Systems, Mass. Institute of Technology, Cambridge, Mass., Jan. 1980.
[10] Gallager, R.G., and Golestaani, S.J., "Flow Control and Routing Algorithms for Data Networks", Proc. of the Fifth International Conference on Computer Communication (ICCC-80), Atlanta, Ga. Cct., 1980, pp. 779-784.
[11] Gafni, E., "Convergence of a Routing Algorithm", Report LIDS-TH-907, Laboratory for Information and Decision Systems, Mass. Institute of Technology, Cambridge, Mass., May 1979.

## Distrioution List

Defense Documentation Centeri2 Copies
Cameron Station
Alexandria, Virginia 22314
Assistant Chief for Technology ..... 1 Copy
Office of Naval Research, Code 200
Arlington, Virginia 22217
office of Mavai Research ..... 2 Copies
Information Systems Program
Code 437
Arlington, Virginia 22217
Office of Naval Research 1 Copy
Branch Office, Boston
495 Sumer Street
Boston, Massachusetts ..... 02210
Office of Naval Research ..... 1 Copy
Branch Office, Chicago 536 South Clark Street Chicago, Illinois 60605
Office of Naval Research ..... 1 CopyBranch Office, Pasadena1030 East Greet Street
Pasadena, California 91106
Naval Research Laboratory

6 Copies
Technical Information Division, Code 2627
Washington, D.C. 20375
Dr. A. L. Slafkosky
1 Copy
Scientific Advisor
Commandant of the Marine Corps (Code RD-1)
Washington, D.C. 20380
Office of Naval Research
1 Copy

## Code 455

:rrlington, Virginia 22217
Office of Naval Research
1 Copy
code 458
Arlington, Virginia 22217
Naval Electronics Laboratory Center 1 Copy Advanced Software Technology Division Code 5200
San Diego, California 92152
Mr. E. H. Gleissner ..... 1 Copy
Naval Ship Rescarch \& Eevelopment Center Computation and Mathematics Deq-urtment Bethesia, Maryland 20085
Captain Grace M. Hopper 1 Copy
NaICOM/MIS Planning Branch (OP-916D) Office of Chief of Raval Operations Washington, D.C. 20350
Advanced Research Projects Agency
Information Processing Techniques
1400 Wilison Boulevard
Arlington, Virginia 22209
Dr. Stuart L. Brodsky ..... 1 Copy
Office of Naval Researcin
Code 432
Arlington, Virginia 22217
Captain Richard L. Martin, USA1 Copy
Commanding Officer
USS Francis Marion (LPA-249)
FPO New York 09501


[^0]:    $\dagger_{\text {This }}$ research was conducted at the M.I.T. Laboratory for Information ani Decision Systems with partial support provided by ARPA Grant No. ONR-N00014-75-C-1183 and National Science Foundation Grant NSF/ECS 79-19880.
    *Room No. 35-210, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, Mass. 02139
    **Room No. 35-419, same as *.
    ***Room No. 35-206, same as *.

