10-1-1992

Time Complexity of a Path-Formulated Optimal Routing Algorithm

John K. Antonio
Purdue University School of Electrical Engineering

Wei K. Tsai
University of California, Irvine, Department of Electrical Engineering

G. M. Huang
Texas A&M University, Department of Electrical Engineering

TIME COMPLEXITY OF A PATH-FORMULATED OPTIMAL ROUTING ALGORITHM

JOHN K. ANTONIO
WEI K. TSAI
G. M. HUANG

Time Complexity of a Path-Formulated Optimal Routing Algorithm

John K. Antonio
School of Electrical Engineering
Purdue University
West Lafayette, Indiana 47907-1285

Wei K. Tsai
Department of Electrical Engineering
University of California, Irvine
Irvine, California 92717

G. M. Huang
Department of Electrical Engineering
Texas A&M University
College Station, Texas 77843

October 1992

1

Abstract

A detailed analysis of convergence rate is presented for an iterative path formulated optimal routing algorithm. The primary objective is to quantify, analytically, how the convergence rate changes as the number of nodes in the underlying graph increases. The analysis is motivated by a particular path formulated gradient projection algorithm that has demonstrated excellent convergence rate properties through extensive numerical studies. In particular, the empirical data suggests that the number of iterations required for convergence to within a small fraction of the optimal cost is relatively independent of the number of nodes. Deriving a correspondingly tight analytical bound for the number of iterations required for convergence, as a function of problem size, proves to be a formidable task, primarily because the dimension of the underlying optimization problem (i.e., the total number of paths connecting all origin-destination pairs) generically grows with a super-polynomial function of the number of nodes. The main result of this paper is that the number of iterations for convergence depends on the number of nodes only through the network diameter.

I. INTRODUCTION AND PROBLEM FORMULATION

A. Basics

The primary objective of virtually all routing algorithms is to select routes for those origin destination (OD) pairs that request data communication. A secondary objective is to insure that messages transmitted along the selected routes are delivered to the correct destinations. This latter objective is accomplished by using standard techniques involving protocols and routing tables. In this paper the focus is on the former objective— the route selection problem. It is known that route selection has a substantial impact on the performance of data networks [1-9,12,13,15,16]. Roughly speaking, an optimal routing is a set of routes that yields the "best" network performance—based on some quantitative measure. The types of performance measures employed by most optimal routing formulations, estimate, in some sense, the average delay associated with sending a packet of data to a typical destination node.

An important issue to consider when implementing a routing algorithm in a large distributed data network is the question of whether the computation should be done in a centralized or distributed manner. Centralized implementations are fairly straight-forward: a designated "central" node is sent data (which characterizes the state of the network) from the other network nodes; then, based on this information, the central node solves the optimal routing problem and broadcasts the solution back out to the network. One of the obvious problems with this type of scheme is the associated communication overhead (i.e., bottleneck). In contrast, certain distributed implementations can reduce this communication overhead, by requiring, for example, only nearest neighbor communication. Due to the potentially high degree of fault tolerance, fast convergence rates, low communication overhead, and other reasons, distributed implementations have received a great deal of attention in the literature over the past decade or so.

One fundamental question associated with distributed routing algorithms is that of convergence. Namely, because the order of events in distributed algorithms occur asynchronously (to one degree or another), the question of whether the algorithm will converge becomes a non-trivial one. In references [5] and [6], convergence of a class of distributed optimal routing algorithms is proven under very mild assumptions. The present authors have proven convergence of a class of distributed iterative aggregation algorithms [15], which have applications in optimal routing.

While the question of distributed asynchronous convergence has been addressed in the above cited works (and others), the goal in the present paper is to determine the amount of time required for a class of iterative path formulated optimal routing algorithms to converge. The time complexity of routing algorithms is an important practical as well as theoretical issue. In practice, it is imperative that the routing algorithm converge within a certain amount of time, otherwise the eventually arrived upon solution may be of little or no value. In the present paper it is shown how network parameters such as maximum link utilization factors, traffic demand values, link capacity values, and the number of network nodes affect the time required for convergence. In order to achieve meaningful bounds for the convergence rate, a certain price was paid in that the assumed model for computation is essentially synchronous (in terms of the order in which iterations are executed). However, it is believed that the ground-work laid out in this paper should serve well as a guide for future work under more relaxed (i.e., asynchronous) assumptions. The main time complexity results are for a class of path-formulated gradient projection-based algorithms.

B. *Formulation of* the Optimal Routing *Problem*

The following formulation uses the same notation and is based on the same approximating assumptions as set forth by Bertsekas and Gallager in reference [2].

Delay Models

Queuing theory is the primary methodological framework for analyzing network performance. Oftentimes its use requires simplifying assumptions for the sake of mathematical tractability. Due to the complexity of realistic networks, it is typically impossible to obtain accurate quantitative delay predictions, however, the models used often provide valuable qualitative results and insights [2].

Perhaps the simplest queuing model is the so-called $M/M/1$ queuing system that consists of a single queuing station and a single server. It is assumed that customers (i.e., packets of data) arrive according to a Poisson process with rate F, and the probability distribution of the service rate is exponential with mean C. By applying Little's Theorem, the average delay for a packet to traverse link (i, j) is given by

$$D_{ij} = \frac{1}{C_{ij} - F_{ij}},$$ (1)

4

where C_{ij} and F_{ij} denote the service rate and arrival rate respectively, associated with link (i,j).

Jackson's Theorem states that in a network of single server queues in which customers arrive from outside the network at each queue according to independent Poisson processes, the average number of outstanding packets in the (steady-state) system can be derived as if each queue in the network is an $M/M/1$ queue. So, for the purpose of measuring network performance, modeling the entire network with simple $M/M/1$ queues is justified.

Based on Jackson's Theorem and Equation (1), the cost function is defined as a weighted sum of all link delays:

$$D(F) = \sum_{(i,j)\in\mathcal{L}} \frac{F_{ij}}{C_{ij} - F_{ij}}, \tag{2}$$

where $\mathcal{L}$ is the set of all links and links having more traffic flow are given higher relative weightings, i.e., Equation (1) is multipled by F_{ij}. Note that each term in the sum represents the average size of the queue associated with link (i,j). Therefore, $D(F)$ is an estimate of the total number of outstanding packets in the network. For the purposes of this paper, determining routes that minimize $D(F)$, for a given set of OD traffic demands, will constitute the notion of an optimal routing.

Preliminary Notation

The following notation is needed in order to formally state the optimal routing problem. Throughout the paper, script fonts such as $\mathcal{W}$ and $\mathcal{P}$ are used exclusively to denote sets.

$\mathcal{W}$: The set of OD pairs requesting communication.

w : A generic OD pair in **W**.

r_w : The arrival rate (traffic demand) measured in packets/sec, for the OD pair w.

$\mathcal{P}_w$: For the OD pair w, this is the set of all logical paths connecting the origin node to the destination node.

p : A generic path in $\mathcal{P}_w$.

x_p : The flow rate on the logical path p.

Constraint Equations

The following constraint equations arise naturally due to conservation of flow.

$$F_{ij} = \sum_{\substack{\text{all paths } p \\ \text{containing } (i,j)}} x_p, \tag{3}$$

$$\sum_{p \in \mathcal{P}_w} x_p = r_w \qquad \text{for all } w \in W \tag{4}$$

and

$$x_p \geq 0 \qquad \text{for all } p \in \mathcal{P}_w, \ w \in \mathcal{W}. \tag{5}$$

Note that the cost function being minimized, see Equation (2), can be expressed in terms of the path vector x, defined as $\mathrm{x} = [x_p]_{\substack{p \in \mathcal{P}_w \\ w \in \mathcal{W}}}$. By combining constraint Equation (3) with the definition of the path vector, the cost function of (2) can be written as

$$D(F) = D(x) \stackrel{\text{def}}{=} \sum_{(i,j) \in \mathcal{L}} \frac{K_{ij} x}{C_{ij} - K_{ij} x}, \tag{6}$$

where K_{ij} is a row vector with components equal to either zero or one. Specifically, the pth component of K_{ij} is one if link (i,j) is on path p, otherwise, the pth component of K_{ij} is zero.

The Path-Formulated Optimal Routing Problem

$$\text{Given } r_w, \qquad \text{for each } w \in W, \tag{7}$$

$$\text{minimize } \{D(x)\}, \tag{8}$$

such that Equations (4) and (5) are satisfied.

II. THE PATH FORMULATED GRADIENT PROJECTION (PFGP) ALGORITHM

It can be shown that the path-formulated optimal routing problem can be transformed into an equivalent box-constrained problem, see [15] for more details. Also, the function $D(x)$ is a differentiable convex function of the path vector x. Therefore, the path-formulated optimal routing problem can be solved numerically by using well established techniques from nonlinear programming; the focus here is on the gradient projection method. The main idea of the gradient projection technique is that after a step is made

in the direction of the negative gradient, the result is orthogonally projected onto the positive orthant.

The iteration equation that results from applying the gradient projection method to the path-formulated optimal routing problem necessitates the definition of the first derivative length (FDL) of a path. The FDL of path p, denoted d_p, is defined by

$$d_p \stackrel{\text{def}}{=} \frac{\partial D(x)}{\partial x_p} = \sum_{\substack{\text{all links } (i,j) \\ \text{on path } p}} \frac{\partial D_{ij}(F_{ij})}{\partial F_{ij}}, \tag{9}$$

where

$$D_{ij}(F_{ij}) = \frac{F_{ij}}{C_{ij} - F_{ij}}. \tag{10}$$

Next, the minimum FDL (MFDL) paths, denoted as $\overline{p}_w$ for each $w \in W$, are defined by

$$d_{\overline{p}_w} = \min_{p \in \mathcal{P}_w} \{d_p\}, \qquad \text{for all } w \in W. \tag{11}$$

Note that for any particular w, there may be more than one MFDL path. In case of such an event, $\overline{p}_w$ is an arbitrarily chosen MDFL path.

The iteration equation associated with the PFGP algorithm [5,20] can now be stated:

$$x_p^{(k+1)} = \max\{0, \ x_p^{(k)} - \alpha^{(k)}(H_{pp}^{(k)})^{-1}(d_p^{(k)} - d_{\overline{p}_w}^{(k)})\}, \qquad \text{for all } w \in \mathcal{W}, \ p \in \mathcal{P}_w, \ p \neq \overline{p}_w, \tag{12}$$

where k is the iteration count. The term $\alpha^{(k)}$ denotes the step size and the term $H_{pp}^{(k)}$ is a scaling factor that is related to the second derivative length of path p. It is easy to verify that the term $\alpha^{(k)}(H_{pp}^{(k)})^{-1}(d_p^{(k)} - d_{\overline{p}_w}^{(k)}) \geq 0$, for all $p \in \mathcal{P}_w$, and therefore, the above iteration equation need not be applied to those paths for which $x_p^{(k)} = 0$. Thus, the set of active paths at iteration k, denoted by $\tilde{\mathcal{P}}_w^{(k)}$, is defined as as

$$\tilde{\mathcal{P}}_w^{(k)} = \{p \in \mathcal{P}_w | x_p^{(k)} > 0\}.$$

So, a more efficient version of the PFGP algorithm (as described originally in [20]) is the following:

$$x_p^{(k+1)} = \max\{0, \ x_p^{(k)} - \alpha^{(k)}(H_{pp}^{(k)})^{-1}(d_p^{(k)} - d_{\overline{p}_w}^{(k)})\}, \qquad \text{for all } w \in W, \ p \in \tilde{\mathcal{P}}_w^{(k)}, \ p \neq \overline{p}_w, \tag{13}$$

$$x_{\bar{p}_w}^{(k+1)} = r_w - \sum_{p \in \tilde{\mathcal{P}}_w, p \neq \bar{p}_w} x_p^{(k)}, \qquad \text{for all } w \in W, \tag{14}$$

$$\tilde{\mathcal{P}}_w^{(k+1)} = \left\{ p \in \tilde{\mathcal{P}}_w^k \mid x_p^{(k)} - \alpha^{(k)} (H_{bb}^{(k)})^{-1} (d_{\bar{p}}^{(k)} - d_{\bar{p}_w}^{(k)}) > 0 \right\} \cup \{\bar{p}_w\}, \qquad \text{for all } w \in W. \tag{15}$$

The PFGP algorithm of Equations (13) through (15) has been efficiently implemented as a serial FORTRAN code, see reference [1]. This code uses a constant step size and the value of $H_{pp}^{(k)}$ is an approximation of the pth diagonal element of the Hessian matrix. The set of active paths for each $w \in W$ are initialized with a single (randomly selected) shortest hop path.

III. General Time Complexity Issues

A. Basics

The overall time complexity of the PFGP algorithm is given by the product of the complexity of each iteration and the complexity of the number of iterations required for convergence to an acceptably small neighborhood of the optimal solution.

From Equations (13) through (15), the complexity of each iteration is clearly dependent on the following three quantities: (i) the number of active paths: $|\tilde{\mathcal{P}}_w^{(k)}|$; (ii) the number of OD pairs: $|\mathcal{W}|$; and (iii) the number of nodes in the network: n. The dependence on $|\tilde{\mathcal{P}}_w^{(k)}|$ is due to the fact that the flow on each (non-MFDL) active path must be updated according to Equation (13). The dependence on $|\mathcal{W}|$ comes from the fact that a MFDL path must be determined for each $w \in W$. Finally, the dependence on the number of nodes, denoted by n, is due to the fact that solving shortest path problems (i.e., finding the MFDL paths) generically depends on the size of the graph. The complexity of each iteration (E1) is therefore denoted as $T_{EI}(k, |W|, n)$, where k actually denotes the dependence on $|\tilde{\mathcal{P}}_w^{(k)}|$. $T_{EI}(k, |W|, n)$ is fairly straightforward to estimate—the only difficulty comes in estimating the maximum number of active paths used in any single iteration. The following is an obvious upper bound for $|\tilde{\mathcal{P}}_w^{(k)}|$:

$$|\tilde{\mathcal{P}}_w^{(k)}| \leq k + |\tilde{\mathcal{P}}_w^{(0)}|, \qquad \text{for all } w \in \mathcal{W}, k \geq 0, \tag{16}$$

because at each iteration at most one new active path is added to each set $\tilde{\mathcal{P}}_w^{(k)}$.

In contrast to the fairly straightforward task associated with estimating the complexity of each iteration (described above), the main concern in this paper is to estimate the

8

complexity associated with the number of iterations, say N_I, required for the PFGP algorithm to converge. Most of the classical results related to convergence rates of numerical optimization algorithms depend on the values of the largest and smallest eigenvalues of the Hessian. For example, it is shown in [10, p. 338–340] that by using a special step size rule, the convergence rate for gradient projection algorithms is bounded by

$$D^{(k+1)} \le \left[\frac{(B-b)}{(B+b)} \right]^2 D^{(k)} \tag{17}$$

where $D^{(k)}$ denotes the value of the cost function at iteration k, and B and b are, respectively, the largest and smallest eigenvalues (in magnitude) of the Hessian. From Equation (17) it is easy to see that if the difference $B - b$ is large (or $b \to 0$), then $[(B-b)/(B+b)]^2 \to 1$. On the other hand, if $b \to B$, then $[(B-b)/(B+b)]^2 \to 0$. Clearly, the smaller the value of $[(B-b)/(B+b)]^2$, the faster the convergence rate, which implies fewer iterations are required for convergence to within a fixed neighborhood of the optimal solution. Unfortunately, the convergence rate of Equation (17) has some practical problems when considering the application of the gradient projection technique to the optimal routing problem. First, the assumed step size rule used to derive Equation (17) is based on a type of line minimization technique which would be difficult to implement in a large distributed network—in practice a constant (or simple) step size rule is used. Second, it is difficult to determine a meaningful lower bound for b, primarily because the number of active paths can (potentially) grow according to a super-polynomial function of n. In particular, $|\tilde{\mathcal{P}}_w^{(k)}|$ is bounded above by $|\mathcal{P}_w^{(n)}|$, where $|\mathcal{P}_w^{(n)}|$ denotes the total number of (potential) paths that interconnect the OD pair w. For all but the sparsest of graphs, $|\mathcal{P}_w^{(n)}|$ grows as a super-polynomial function of n. (Consider, for example, the fact that there exists $O(2^n)$ distinct paths that interconnect various OD pairs in a simple n-node planar mesh.)

In estimating the number of iterations for convergence for the PFGP algorithm, one of the most crucial issues is the assumed bound for $|\tilde{\mathcal{P}}_w^{(k)}|$. If one uses the fact that $|\tilde{\mathcal{P}}_w^{(k)}| \le |\mathcal{P}_w^{(n)}|$, then the resulting analysis indicates that the number of iterations for convergence is bounded by a function that depends on $|\mathcal{P}_w^{(n)}|$, which can result in an overall bound that grows with a super-polynomial function of n. This assumption and the resulting convergence rate result are apparently too loose because the empirical data suggests that the number of iterations for convergence actually grows (at most) slowly as the number of nodes in the graph is increased.

In the paper [19], it is shown that if one *assumes* that $|\tilde{\mathcal{P}}_w^{(k)}| \leq |\mathcal{P}_{\mathbf{max}}|$, where $|\mathcal{P}_{\mathbf{max}}|$ is a constant (independent of both k and n), then the number of iterations for convergence is indeed bounded by a slowly increasing function of n. In [19], the assumption that $|\tilde{\mathcal{P}}_w^{(k)}| \leq |\mathcal{P}_{\mathbf{max}}|$ is argued to be reasonable because numerous simulation studies indicated that $|\tilde{\mathcal{P}}_w^{(k)}|$ rarely exceeded ten, regardless of the size of the network. Also, the main thrust in [19] is not in getting a necessarily tight bound for the convergence rate but rather in showing that the convergence rate assuming a Jacobi-type updating rule is the roughly the same as the convergence rate associated with a Gauss-Seidel updating rule.

In the present paper, we allow $|\tilde{\mathcal{P}}_w^{(k)}|$ to grow according to $k + |\tilde{\mathcal{P}}_w^{(0)}|$ (i.e., no uniform bound is assumed) and show that the number of iterations for convergence is still bounded by (at most) a slowly increasing function of n (and thus, from this result, we confirm indeed that only a small number of active paths are required to achieve convergence). Because no uniform bound is assumed for $|\tilde{\mathcal{P}}_w^{(k)}|$ in the present paper, the analysis techniques are significantly different from those used in reference [19].

B. *Serial Versus Distributed Time Complexities*

Thus far a distinction has not been made between the time complexity of the PFGP algorithm relative to serial and distributed implementations. In a serial (single processor) implementation, the value for $T_{\mathcal{E}I}(k, |W|, n)$ is the sum of the time required to solve the shortest path problem for each $w \in W$ and the time required to update all active path flows in $\tilde{\mathcal{P}}_w^{(k)}$. In a distributed implementation, a distributed shortest path algorithm could be employed (such as the distributed Bellman-Ford algorithm [2,14]) and each node i could assume the responsibility of updating all active path flows originating at node i. Of course one of the main difficulties associated with distributed algorithms (in general) is the asynchronous nature of the communication overhead.

For the purposes of this paper it shall be assumed that iteration $k + 1$ is executed only after iteration k is completed. Under this simplifying assumption, the complexity for the number of iterations is the same for both the serial and distributed implementations of the PFGP algorithm. In terms of the distributed implementation, this assumption implicitly assumes the existence of a uniform upper bound for the communication time complexity of each iteration. This type of assumption results in what is typically called a **partially** asynchronous distributed algorithm. It has been proven that the distributed PFGP algorithm will actually converge (eventually) in a virtually totally asynchronous

computing environment, see [5,6]. However, with such mild restrictions on the ordering of events it becomes very difficult to tightly bound the rate of convergence.

The analysis technique introduced in the present paper for estimating the convergence rate (for the partially asynchronous case) serves a twofold purpose. First, the resulting bound may serve as a nominal estimate of the convergence rates associated with a more asynchronous model. Second, the analysis techniques developed in deriving the bound for the partially asynchronous case may serve as a guide for deriving similar results under more relaxed assumptions.

IV. THE COMPLEXITY OF THE CONVERGENCE RATE

In this section upper bound results for the convergence rate of the PFGP algorithm are derived. First, all necessary notation for stating the main results is introduced.

A. *Notation*

From this point on, the superscript "(n)" is placed on those variables or sets that are explicitly dependent on the number of nodes in the network. Likewise, the superscript "(k)" is used to indicate dependence on the iteration count, k. Variables with neither a "(n)" nor a "(k)" superscript are assumed to be constants, independent of both n and k. One particularly important yet subtle point is that the set of all logical paths associated with the OD pair w is denoted by $\mathcal{P}_w^{(n)}$, while the set of active paths at iteration k associated with OD pair w is denoted by $\tilde{\mathcal{P}}_w^{(k)}$.

$\mathcal{W}^{(n)}$: The set of OD pairs requesting communication.

w : A generic OD pair in $\mathcal{W}^{(n)}$.

r_w : The arrival rate for the OD pair w $\in \mathcal{W}^{(n)}$.

$r_{\min}$: The minimum arrival rate, for all w $\in \mathcal{W}^{(n)}$:

$$r_{\min} = \min_{w \in \mathcal{W}^{(n)}} \{r_w\}.$$

$\mathcal{P}_w^{(n)}$: For the OD pair w, this is the set of all logical paths that connect the origin node to the destination node.

p : A generic path in $\mathcal{P}_w^{(n)}$.

11

x : The flow rate on path p at iteration k.

$\tilde{\mathcal{P}}_w^{(k)}$: This is the set of active paths in $\mathcal{P}_w^{(n)}$ at iteration k:

$$\tilde{\mathcal{P}}_w^{(k)} = \{p \in \mathcal{P}_w^{(n)} | x_p^{(k)} > 0\}.$$

$|\tilde{\mathcal{P}}_{\max}^{(0)}|$: The maximum number of initially active paths associated with any single OD pair:

$$|\tilde{\mathcal{P}}_{\max}^{(0)}| = \max_{w \in \mathcal{W}} \left\{ |\tilde{\mathcal{P}}_w^{(0)}| \right\}.$$

$F_{ij}^{(k)}$: The flow rate on link (i, j) at iteration k.

C_{ij} : The capacity of link (i, j).

$C_{\min}$: The minimum value of C_{ij}, for all $(i, j) \in \mathcal{L}$.

$C_{\max}$: The maximum value of C_{ij}, for all $(i, j) \in \mathcal{L}$.

$\rho_{ij}^{(k)}$: The utilization factor of link (i, j) at iteration k:

$$\rho_{ij}^{(k)} = \frac{F_{ij}^{(k)}}{C_{ij}}.$$

$\rho_{\max}$: The maximum link utilization factor for all $(i, j) \in \mathcal{L}$, and all k:

$$\rho_{\max} = \max_{\substack{(i,j) \in \mathcal{L} \\ k \geq 0}} \left\{ \rho_{ij}^{(k)} \right\}.$$

$D^{(k)}$: The value of the cost function at iteration k:

$$D^{(k)} = \sum_{(i,j) \in \mathcal{L}} \frac{F_{ij}^{(k)}}{C_{ij} - F_{ij}^{(k)}} = \sum_{(i,j) \in \mathcal{L}} \frac{K_{ij} x^{(k)}}{C_{ij} - K_{ij} x^{(k)}}.$$

D^* : The optimal value of the cost function.

$E^{(k)}$: The relative error:

$$E^{(k)} = \frac{D^{(k)} - D^*}{D^*}$$

$h_w^{(n)}$: The minimum hop distance between the origin and destination of each OD pair $w \in \mathcal{W}^{(n)}$.

$h_{\max}^{(n)}$: The maximum value of $h_w^{(n)}$, for all $w \in \mathcal{W}^{(n)}$:

$$h_{\max}^{(n)} = \max_{w \in \mathcal{W}^{(n)}} \{h_w^{(n)}\}.$$

$h_{\min}^{(n)}$: The minimum value of $h_w^{(n)}$, for all $w \in \mathcal{W}^{(n)}$:

$$h_{\min}^{(n)} = \min_{w \in \mathcal{W}^{(n)}} \{h_w^{(n)}\}.$$

h   : The average value of $h_w^{(n)}$, for all $w \in \overline{\mathcal{W}}^{(n)}$:

$$h_{avg}^{(n)} = \frac{1}{|\mathcal{W}^{(n)}|} \sum_{w \in \mathcal{W}^{(n)}} h^{\bullet \bullet}$$

B.   Assumptions:

(A1) Iteration k is completed before iteration k $+$ 1 begins.

(A2) $0 \le \rho_{\max} < 1$.

(A3) There exists constants $\zeta_{\min}$ and $\zeta_{\max}$, such that $\alpha^{(k)}$ (the **stepsize** at iteration k) satisfies

$$(\zeta_{\min})\overline{\alpha}^{(k)} \le \alpha^{(k)} \le (\zeta_{\max})\overline{\alpha}^{(k)},$$

where

$$0 < \zeta_{\min} \le \zeta_{\max} < 1$$

and

$$\overline{\alpha}^{(k)} \stackrel{\text{def}}{=} \left(\frac{C_{\min}}{C_{\max}}\right)^3 \left(\frac{(1-\rho_{\max})^5}{3}\right) \left(\frac{\min_{w \in \mathcal{W}^{(n)}} \{|\tilde{\mathcal{P}}_w^{(k)}|\}}{\max_{w \in \mathcal{W}^{(n)}} \{|\tilde{\mathcal{P}}_w^{(k)}|\}}\right) \left(\frac{(h_{\min}^{(n)})}{(|\mathcal{W}^{(n)}|)(h_{avg}^{(n)})}\right).$$

The need for assumption (A1) was mentioned previously. Namely, without assuming this type of synchronization (with respect to the execution of the iterations) the analysis of convergence rates becomes extremely difficult. (Note: In a more relaxed asynchronous computing environment the question of interest is not typically one of "How fast does the algorithm converge?" but more fundamentally "Does the algorithm converge?")

Assumption (A2) (i.e., $0 \le \rho_{\max} < 1$) implies that the routings produced by each iteration of the PFGP algorithm be "valid" routings, i.e., $F_{ij}^{(k)} < C_{ij}$, for all $(i,j) \in \mathcal{L}$ and all $k \ge 0$. However, by the descent property of the gradient projection method

13

and the convexity of the optimization problem at hand, it can be established that if the initial routing is valid, i.e., if $F_{ij}^{(0)} < C_{ij}$, for all $(i,j) \in \mathcal{L}$, then all routings produced by subsequent iterations will also be valid. Therefore, a question is raised as to how to handle initial routings that are not valid, i.e., those for which $F_{ij}^{(0)} > C_{ij}$, for some $(i,j) \in L$. In the implementation of the algorithm in [1], this potential problem is handled by cleverly redefining the penalty function for each link as follows:

$$D_{ij}(F_{ij}) = \begin{cases} \frac{F_{ij}}{C_{ij} - F_{ij}}, & \text{if } F_{ij} < 0.95 C_{ij} \\ a(F_{ij})^2 + bF_{ij} + c, & \text{if } F_{ij} \geq 0.95 C_{ij} \end{cases} \tag{18}$$

where the coefficients a, b, and c are chosen so that $D_{ij}(F_{ij})$ and its first two derivatives (with respect to F_{ij}) are all smooth at the point $F_{ij} = 0.95C$. The advantage in using Equation (18) to define the link penalty function (instead of the strict $M/M/1$-based penalty function) is the fact that the functions $D_{ij}(F_{ij})$, $D'_{ij}(F_{ij})$, and $D''_{ij}(F_{ij})$ are all well-defined for all $F_{ij} \geq 0$ (instead of just $0 \leq F_{ij} < C_{ij}$). Also note that $D''_{ij}(F_{ij}) \leq a$, for all $F_{ij} \geq 0$. Therefore, the size of each element in the associated Hessian is bounded above by the constant a, regardless of the values of link flows F_{ij}. Thus, the convergence rate associated with the penalty function of Equation (18) is the same as the convergence rate associated with the strict $M/M/1$-based penalty function under the assumption that $\rho_{\max} \leq 0.95$. Therefore, under the practical assumption that the utilization of all links will eventually be below 0.95 (e.g., at the optimal solution), using Equation (18) enables us to (effectively) bound $\rho_{\max}$ by 0.95. Clearly, the factor of 0.95 can be further lowered in Equation (18) if lower optimal (i.e., eventual) utilization factors are expected. However, lowering this factor below what turns out to be the actual optimal utilization factor can result in a suboptimal routing, with respect to the $M/M/1$-based cost function.

Assumption (A3) requires the **stepsize** to lie within a specified interval. (Note: It is well-known — for general gradient-type algorithms — that if the **stepsize** is too large, then the algorithm may not converge. On the other hand, if the **stepsize** is too small, then the convergence rate may be arbitrarily slow.)

C. *The* Dynamics of $E^{(k)}$

The main convergence results hinge around the derivation of a closed form bound for a nonlinear recurrence involving $E^{(k)}$. This recurrence equation, which captures the dynamic response of $E^{(k)}$, is stated in Lemma 1. The derivation of Lemma 1 involves

14

a string of preliminary results which are included in detail in the Appendix. A general closed form bound for the recurrence is proven in Lemma **2.**

__Lemma 1:__ Given that assumptions (A1) through (A3) are satisfied, the following holds for all k $\geq$ 0:

$$E^{(k+1)} \leq E^{(k)} \left(1 - E^{(k)} K_1 f^{(n)} g^{(k)} \right), \qquad (L1)$$

where

$$K_1 = \left(\frac{(\zeta_{\min})^2 (1 - \zeta_{\max})(r_{\min})(C_{\min})^9 (1 - \rho_{\max})^{15}}{6(\zeta_{\max})(C_{\max})^8 (1 + C_{\max})^2} \right), \qquad (L1.1)$$

$$f^{(n)} = \frac{(h_{\min}^{(n)})^2}{(h_{\max}^{(n)})^2}, \qquad (L1.2)$$

and

$$g^{(k)} = \frac{\left(\min_{w \in \mathcal{W}^{(n)}} \left\{ \left| \tilde{\mathcal{P}}_w^{(k)} \right| \right\} \right)^2}{\left(\max_{w \in \mathcal{W}^{(n)}} \left\{ \left| \tilde{\mathcal{P}}_w^{(k)} \right| \right\} \right)}. \qquad (L1.3)$$

__Lemma 2:__ Given that assumptions (A1) through (A3) are satisfied, the following holds for all k $\geq$ 0:

$$E^{(k)} \leq \frac{1}{K_2 + K_1 f^{(n)} \sum_{i=0}^{k-1} g^{(i)}}, \qquad (L2)$$

where K_1, $f^{(n)}$ and $g^{(k)}$ are defined by Equations $(L1.1)$ through (L1.3) and

$$K_2 = \frac{(1 - \rho_{\max}) r_{\min}}{|\tilde{\mathcal{P}}_{\max}^{(0)}| \rho_{\max} C_{\max}}. \qquad (L2.1)$$

__Proof:__ Obviously, $E^{(k)} \geq 0$ for all k $\geq$ 0. Also, from Equation $(L1)$ note that $E^{(k+1)} \leq E^{(k)}$. Therefore, an alternate expression for $E^{(k+1)}$ is obtained by bounding the right hand side of Equation $(L1)$, yielding

$$E^{(k+1)} \leq E^{(k)} \left(1 - E^{(k+1)} K_1 f^{(n)} g^{(k)} \right).$$

Now the following bound for $E^{(k+1)}$ is obtained:

$$E^{(k+1)} \leq \frac{E^{(k)}}{1 + E^{(k)} K_1 f^{(n)} g^{(k)}}.$$

15

A closed form bound is then found for all $k \geq 0$:

$$E^{(k)} \leq \frac{E^{(0)}}{1 + E^{(0)} K_1 f^{(n)} \sum_{i=0}^{k-1} g^{(i)}}.$$

Finally, from Proposition 4 (see Appendix), $E^{(0)}$ is bounded as follows:

$$E^{(0)} \leq \frac{|\tilde{\mathcal{P}}_{\max}^{(0)}| \rho_{\max} C_{\max}}{(1 - \rho_{\max}) r_{\min}}.$$

□

D.   The Main Results

In order to determine a bound for the number of iterations required to reduce the relative error to a small value, the right hand side of Equation (L2) is set equal to the desired small value and solved. The following theorem states this fundamental result.

Theorem 1: Given that assumptions (A1) through (A3) are satisfied, then for any constant $\epsilon > 0$, the number of iterations, N_I, required to achieve $E^{(N_I)} \leq \epsilon$ is bounded by

$$N_I \begin{cases} = 0, & \text{if } \frac{1}{\epsilon} - K_2 \leq 0 \\ \leq G^{-1}\left(\frac{1}{K_1 f^{(n)}}\left(\frac{1}{\epsilon} - K_2\right)\right), & \text{if } \frac{1}{\epsilon} - K_2 > 0 \end{cases} \qquad (T1)$$

where $G^{-1}(\cdot)$ is the inverse mapping of $G^{(k)}$:

$$G^{(k)} \stackrel{\text{def}}{=} \sum_{i=0}^{k-1} g^{(i)} \qquad (T1.1)$$

and

$$G^{-1}(z) = \min\left\{ k \mid \sum_{i=0}^{k-1} g^{(i)} \geq z \right\}. \qquad (T1.2)$$

Proof: Set the right hand side of Equation (L2) equal to ϵ and solve.
□

Bounding $G^{(k)}$ and $f^{(n)}$

Because $1 \leq \left|\tilde{\mathcal{P}}_w^{(k)}\right| \leq k + \left|\tilde{\mathcal{P}}_{\max}^{(0)}\right|$, for all $w \in \mathcal{W}^{(n)}$, $g^{(k)}$ (refer to Equation (L1.3)) can be bounded as follows:

$$\frac{1}{k + \left|\tilde{\mathcal{P}}_{\max}^{(0)}\right|} \leq g^{(k)} \leq k + \left|\tilde{\mathcal{P}}_{\max}^{(0)}\right|.$$

16

Therefore, based on Equation (T1.1) the following bounds for $G^{(k)}$ are obtained:

$$\left(\frac{1}{\left|\tilde{\mathcal{P}}_{\max}^{(0)}\right|}\right) \log_2 k \leq G^{(k)} \leq \left(\frac{\left|\tilde{\mathcal{P}}_{\max}^{(0)}\right|}{2}\right) k(k+1).$$

So, without making any assumptions on the rate at which the important quantity $g^{(k)}$ grows, it must generically be assumed that the mapping G^{-1} is exponential, in spite of the fact that it could be as small as the square-root function. In the following lemma, however, it is shown that if there exists a constant $0 < \gamma \leq 1$ such that $\left|\tilde{\mathcal{P}}_w^{(k)}\right| \leq (k+1)^{1-\gamma}$, for all w, then the mapping G^{-1} is polynomial. Thus, there is reason to believe that in practice, G^{-1} may rarely be exponential.

<u>Lemma 3:</u> If there exists a constant $0 < y \leq 1$ such that $\left|\tilde{\mathcal{P}}_w^{(k)}\right| \leq (k+1)^{1-\gamma}$, for all $w \in \mathcal{W}^{(n)}$, then

$$G^{(k)} \geq \frac{k^\gamma}{\gamma}.$$

<u>Proof:</u> Applying the assumption to Equation $(L1.3)$, note that

$$g^{(k)} \geq \frac{1}{(k+1)^{1-\gamma}}.$$

Therefore, applying the definition of $G^{(k)}$ (Equation (T1.1)) yields

$$G^{(k)} \geq \sum_{i=1}^{k} \frac{1}{i^{1-\gamma}}.$$

Now, by bounding the above sum using a standard integral approximation technique,

$$G^{(k)} \geq \frac{k^\gamma}{\gamma}.$$

□

The next task involves upper bounding the quantity $\frac{1}{f^{(n)}}$.

Because $1 \leq h_w^{(n)} \leq d^{(n)}$, for all w (where $d^{(n)}$ denotes the hop diameter of the network), an upper bound $\frac{1}{f^{(n)}}$ (refer to Equation (L1.2)) is as follows:

$$\frac{1}{f^{(n)}} \leq \left(d^{(n)}\right)^2.$$

Fortunately, it turns out that under very mild assumptions, the diameters of large random graphs increase (in the worst case) proportional to $\log n$, with probability one. In particular, in reference [18, pp. 233-236], it is proven that if the probability of any link

being in the graph is p, then with probability one (as $n \to \infty$) the diameter of the graph will equal either d or $d + 1$, where d satisfies the following equation

$$p^d n^{d-1} = \log(n^2). \tag{19}$$

The following lemmas result from Equation (19).

Lemma 4: Provided that $P[(i,j) \in \mathcal{L}] \geq \frac{2}{n}$, for all $i \neq j$, then with probability one, the diameter of the graph is bounded by a logarithmic function of n. In particular

$$d^{(n)} \leq 2(\log n + 1).$$

Proof: Set $p = \frac{2}{n}$ in Equation (19) and solve for the desired bound on d.
□

Lemma 5: For any constant $0 < 6 \leq 1$, provided that $P[(i,j) \in \mathcal{L}] \geq \frac{2}{n^{1-\delta}}$, for all $i \neq j$, then with probability one, the diameter of the graph is bounded by a constant. In particular,

$$d^{(n)} \leq \frac{1+\delta}{\delta}.$$

Proof: By setting $p = \frac{2}{n^{1-\delta}}$, note that

$$2^{d-1} n^{\delta(d-1)} = n^{1-\delta} \log n.$$

Now, apply the logarithm to both sides of above Equation and solve for $d - 1$

$$d - 1 = \frac{\log\left(n^{1-\delta}\right) + \log(\log n)}{1 + \delta \log n}$$

Now by noting that $\log\left(n^{1-\delta}\right) = (1-\delta) \log n$ and by replacing the denominator by $\delta \log n$, (after letting $n \to \infty$)

$$d \mid \frac{1}{6}.$$

Finally, recalling that the diameter is bounded by $d + 1$, the result is proven.
□

E. *Summary of the Results*

The main results are summarized below as corollaries. The corollaries come by applying the various conditions and results of Lemmas 3 through 5 to Theorem 1. Assumptions

18

(A1) through (A3) are assumed for all three corollaries. Also, it is assumed that $\frac{1}{\epsilon} > K_2$ (otherwise, $N_I = 0$).

Corollary 1: For any three constants $0 < \epsilon$, $0 < \delta \leq 1$ and $0 < \gamma \leq 1$, provided that $P[(i,j) \in \mathcal{L}] \geq \frac{2}{n^{1-\delta}}$, for all $i \neq j$, and provided that $\left| \mathcal{P}_w^{(k)} \right| \leq (k+1)^{1-\gamma}$, for all k, then with probability one, the number of iterations for convergence to $E^{(N_I)} \leq \epsilon$ is bounded by the following constant

$$N_I \leq \left(\frac{4\gamma}{K_1 \delta^2} \left(\frac{1}{\epsilon} - K_2 \right) \right)^{\frac{1}{\gamma}}.$$

Corollary 2: For any two constants $0 < \epsilon$ and $0 < \delta \leq 1$, provided that $P[(i,j) \in \mathcal{L}] \geq \frac{2}{n^{1-\delta}}$, for all $i \neq j$, then with probability one, the number of iterations for convergence to $E^{(N_I)} \leq \epsilon$ is bounded by the following constant

$$N_I \leq 2^{\frac{(1+\delta)^2}{K_1 \delta^2} \left(\frac{1}{\epsilon} - K_2 \right)}.$$

Corollary 3: For any two constants $0 < \epsilon$ and $0 < \gamma \leq 1$, provided that $P[(i,j) \in \mathcal{L}] \geq \frac{2}{n}$, for all $i \neq j$, and provided that $\left| \mathcal{P}_w^{(k)} \right| \leq (k+1)^{1-\gamma}$, for all k, then with probability one, the number of iterations for convergence to $E^{(N_I)} \leq \epsilon$ is bounded by the following poly-logarithmic function of n

$$N_I \leq \left(\frac{4 (\log n + 1)^2}{K_1} \left(\frac{1}{\epsilon} - K_2 \right) \right)^{\frac{1}{\gamma}}.$$

V.   COMPARISON WITH PREVIOUS RESULTS AND CONCLUSIONS

A.   *Comparison with Previous Results*

The main results are *not incowistent* with the only other known convergence rate results for the algorithm under consideration. In reference [13] it is proven that for any fixed number of nodes there exists a parameter $\beta < 1$ such that $\| x^{(k)} - x^* \| \leq K \beta^k$, where K and β are constants, $x^{(k)}$ is the vector of path flow variables at iteration k and x^* is an optimal vector of path flows. One fundamental issue not addressed in [13], however, is the rate with which the parameter β approaches unity as the number of nodes n is increased. Furthermore, because the result in [13] characterizes the convergence rate of the quantity $\| x^{(k)} - x^* \|$ while the results of the present paper bound the convergence rate of the quantity $E^{(k)} = \frac{D^{(k)} - D^*}{D^*}$, it is difficult to make a fair comparison as to which convergence rate estimate is tighter. This difficulty is highlighted by the construction of a simple example network, see Fig. 1. In the example network, for any given $0 < \epsilon \leq \frac{1}{2}$,
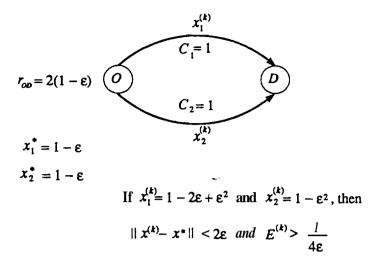
Fig. 1. A simple example network that shows closeness in the sense of the measure $\|x^{(k)} - x^*\|$ does not necessarily imply closeness according to the measure $E^{(k)}$.

define a set of path flows, call them $x^{(k)}$, for which $\|x^{(k)} - x^*\| < 2\epsilon$, while, on the other hand, the relative error of the cost function for the same flows satisfies $E^{(k)} > \frac{1}{4\epsilon}$. Thus, the example shows that being "close" in the sense of the measure $\|x^{(k)} - x^*\|$ does not **necessarily** imply closeness in the sense of the relative error of the cost function. Admittedly, the example in Fig. 1 is somewhat pathological because the path flow values depend on the convergence parameter ϵ. Nevertheless, it demonstrates the **existence** of sets of path flow values (perhaps initial path flows) for which the error measure $\|x^{(k)} - x^*\|$ does not provide a reasonable estimate of the relative error of the cost function.

It is noted that if the number of nodes were assumed to be fixed (i.e., a constant) in the present paper, then $f^{(n)}$ and $g^{(k)}$ could be bounded by constants. Therefore, for $E^{(k)} \geq 1$ (refer to Equation $(L1)$) $E^{(k)} \leq E^{(0)}\bar{\beta}^k$, where $\bar{\beta} = (1 - K_1\bar{f}\bar{g})$, and $\bar{f}$ and $\bar{g}$ are constant upper bounds for $f^{(n)}$ and $g^{(k)}$, respectively. Likewise, for $E^{(k)} < 1$, $E^{(k)} \leq \frac{1}{1+kK_1\bar{f}\bar{g}}$.

In the paper [19], the results indicate that if one **assumes** the number of active paths for each OD pair is bounded by a constant, then the number of iterations for convergence increases slowly as the number of nodes in the graph is increased. In the present paper, no bound is placed on the the number of active paths for each OD pair (i.e., each OD pair is allowed to increase its set of active paths by one at each iteration). Our analysis shows that even under this relaxed condition, the number of iterations for convergence

still increases (at most) slowly as the number of nodes is increased.

B. Conclusions

Bounds have been derived for the number of iterations required for a class of path-formulated gradient projection-based algorithms to converge. The bounds confirm observations made through experimentation and experience, and (more importantly) also offer new insights. First, under relatively mild assumptions on the denseness of the network graph, it was proven that the number of iterations for convergence is independent of the size of the network (with probability one). Second, with essentially no restrictions on the graph density, it is proven that the number of iterations for convergence is bounded by a poly-logarithmic (i.e., sub-linear) function of the number of nodes, provided that the number of active paths for each OD pair (at iteration k) is bounded by a strictly sub-linear function of k. Also, the results show that the number of iterations required for convergence increases as the maximum link utilization factor increases, a fact that has been well established through experimentation. The goal of this paper has been to study how the convergence rate changes as the number of nodes in the graph is increased. Therefore, in the "bound-finding" analysis, the size of the constants were sacrificed in order to get tight asymptotic bounds in n. The constants associated with the bounds are probably overly conservative, and it is not claimed that they are the tightest by any means. Nevertheless, the results appear to be new and insightful.

Two directions of future work are planned. First, a similar derivation of bounds for a competing link-formulated optimal routing algorithm [17] shall be attempted. Second, an extension of the results of the present paper to include the class of iterative aggregation/disaggregation algorithms described in [15]. It seems possible to show that by aggregating paths associated with appropriately chosen groups of OD pairs, the number of iterations for convergence may decrease.

APPENDIX

The purpose of this appendix is to prove Lemma 1 of Section IV. A string of eight initial propositions is given followed by a restatement and proof of Lemma 1. In addition to the notation introduced in Section IV, the following notation is used.

Additional Notation:

$\tilde{\ell}_p$: The number of links along the active path p.

$\nabla_x D^{(k)}$: The gradient of the cost function with respect to the path variables, at iteration k:

$$\nabla_x D^{(k)} = \frac{\partial D^{(k)}}{\partial x}.$$

$\nabla_x^2 D^{(k)}$: The Hessian of the cost function with respect to the path variables, at iteration k:

$$\nabla_x^2 D^{(k)} = \frac{\partial^2 D^{(k)}}{\partial x^2}.$$

$\tilde{\mathcal{P}}^{(k)}$: The set of all active paths at iteration k:

$$\tilde{\mathcal{P}}^{(k)} = \cup_{w \in \mathcal{W}(n)} \tilde{\mathcal{P}}_w^{(k)}.$$

$\tilde{\mathcal{P}}^{(k,k+1)}$: The set of all paths that are active at iteration k or $k+1$:

$$\tilde{\mathcal{P}}^{(k,k+1)} = \tilde{\mathcal{P}}^{(k)} \cup \tilde{\mathcal{P}}^{(k+1)}.$$

$\nabla_x^2 D^{(k,k+1)}$: The Hessian of the cost function restricted to the subspace of active paths $p \in \tilde{\mathcal{P}}^{(k,k+1)}$.

Next, the eight preliminary propositions and Lemma 1 are stated and proven. Assumptions (A1) through (A3) are implicitly assumed.

Proposition 1: For all $(i, j) \in \mathcal{L}$ and for all k; the following inequalities hold.

$$\text{(i)} \quad 0 \quad \leq D_{ij}^{(k)} \leq \frac{\rho_{\max}}{(1 - \rho_{\max})},$$

$$\text{(ii)} \quad \frac{1}{C_{\max}} \quad \leq \frac{\partial D_{ij}^{(k)}}{\partial F_{ij}} \leq \frac{1}{(C_{\min})(1 - \rho_{\max})^2},$$

$$\text{(iii)} \quad \frac{2}{(C_{\max})^2} \quad \leq \frac{\partial^2 D_{ij}^{(k)}}{(\partial F_{ij})^2} \leq \frac{2}{(C_{\min})^2(1 - \rho_{\max})^3},$$

where

$$D_{ij}^{(k)} \stackrel{\text{def}}{=} \frac{F_{ij}^{(k)}}{C_{ij} - F_{ij}^{(k)}}.$$

Proof:

(i) Due to the convexity of $D_{ij}^{(k)}$, the lower and upper bounds are obtained by simply evaluating $D_{ij}^{(k)}$ at $F_{ij}^{(k)} = 0$ and $F_{ij}^{(k)} = (\rho_{\max})(C_{ij})$, respectively, which gives the desired result.

(ii) Note that the first derivative length of link (i,j) is given by

$$\frac{\partial D_{ij}^{(k)}}{\partial F_{ij}} = \frac{C_{ij}}{(C_{ij} - F_{ij}^{(k)})^2}.$$

As in the previous part, the convexity of $\frac{\partial D_{ij}^{(k)}}{\partial F_{ij}}$ enables us to find lower and upper bounds by evaluating at $F_{ij}^{(k)} = 0$ and $F_{ij}^{(k)} = (\rho_{\max})(C_{ij})$, which yields

$$\frac{1}{C_{ij}} \leq \frac{\partial D_{ij}^{(k)}}{\partial F_{ij}} \leq \frac{1}{(C_{ij})(1 - \rho_{\max})^2}, \qquad \text{for all } (i,j) \in \mathcal{L}, k.$$

Now, because $\frac{1}{C_{\max}} \leq \frac{1}{C_{ij}} \leq \frac{1}{C_{\min}}$, for all $(i,j) \in \mathcal{L}$, the result is proven.

(iii) Again, due to the convexity of $\frac{\partial^2 D_{ij}^{(k)}}{(\partial F_{ij})^2}$, the bounds are determined in the same manner as the previous part.

□

Proposition 2: The number of links along the active path p, defined as $\tilde{\ell}_p$, is bounded by:

$$h_w^{(n)} \leq \tilde{\ell}_p \leq \left(\frac{C_{\max}}{(C_{\min})(1 - \rho_{\max})}\right) h_w^{(n)}, \qquad \text{for all } p \in \tilde{\mathcal{P}}_w^{(k)}, w \in \mathcal{W}^{(n)}, k.$$

Proof: The lower bound is obvious. (By definition of $h_w^{(n)}$, every path associated with the OD pair w has at least $h_w^{(n)}$ hops.) To prove the upper bound, note first that for every active path $p \in \tilde{\mathcal{P}}_w^{(k)}$ there exists at least one iteration count, say $k_p \leq k$, for which the flows on the network links (i.e., $F_{ij}^{(k_p)}$) were such that path p was a minimum first derivative length (MFDL) path at iteration k_p. In other words, in order for a path to be active at iteration k, it must have been a MFDL at a previous (or the current) iteration. Now, because the first derivative of each link is strictly greater than zero, we have that

$$\tilde{\ell}_p \leq \left(\frac{D_{\max}'^{(k_p)}}{D_{\min}'^{(k_p)}}\right) h_w^{(n)}, \qquad \text{for all } w \in \mathcal{W}^{(n)}, p \in \tilde{\mathcal{P}}_w^{(k_p)}, \tag{A.1}$$

where

$$D_{\max}'^{(k)} \overset{\text{def}}{=} \max_{(i,j) \in \mathcal{L}} \left\{\frac{\partial D_{ij}^{(k)}}{\partial F_{ij}}\right\}$$

23

and

$$D_{\min}^{'(k)} \overset{\text{def}}{=} \min_{(i,j)\in\mathcal{L}} \left\{ \frac{\partial D_{ij}^{(k)}}{\partial F_{ij}} \right\}.$$

Substituting the result of Proposition 1-part (ii) into Equation (A.1), the proof is complete.

□

Proposition 3: The maximum eigenvalue of $\widetilde{\nabla}_x^2 D^{(k,k+1)}$, denoted as $\tilde{\lambda}_{\max}^{(k)}$, satisfies the following inequality:

$$\tilde{\lambda}_{\max}^{(k)} < \left(\frac{6(C_{\max})\left(\max_{w\in\mathcal{W}^{(n)}}\{|\tilde{\mathcal{P}}_w^{(k)}|\}\right)}{(C_{\min})^3(1-\rho_{\max})^5} \right) \left(|\mathcal{W}^{(n)}|\right)\left(h_{\text{avg}}^{(n)}\right), \quad \text{for all } k.$$

Proof: Let $\nabla_x^2 D^{(k)} = [H_{pq}^{(k)}]$. So for any two paths p and q we have

$$H_{pq}^{(k)} = \sum_{\substack{\text{all common links } (i,j) \\ \text{on paths p and q}}} \frac{\partial^2 D_{ij}(F_{ij}^{(k)})}{(\partial F_{ij})^2}.$$

In order to bound $\tilde{\lambda}_{\max}^{(k)}$, we apply the Gerschgorin Theorem [11]. In particular, we shall determine an upper bound for the sum of elements in each row of the restricted Hessian (i.e., restricted to the set of active paths in $\tilde{\mathcal{P}}^{(k,k+1)}$). First, note that if we consider the sum of elements for each row of the Hessian which is restricted to the set of paths in $\tilde{\mathcal{P}}^{(k)}$, we have that

$$\sum_{q\in\tilde{\mathcal{P}}^{(k)}} H_{pq}^{(k)} \le \sum_{q\in\tilde{\mathcal{P}}^{(k)}} H_{qq}^{(k)}, \quad \text{for each } p \in \tilde{\mathcal{P}}^{(k)},$$

because $H_{pq}^{(k)} \le H_{qq}^{(k)}$, for all p and q. Similarly, if we consider the sum of elements for each row of the Hessian which is restricted to the set of paths in $\tilde{\mathcal{P}}^{(k+1)}$, we have

$$\sum_{q\in\tilde{\mathcal{P}}^{(k+1)}} H_{pq}^{(k)} \le \sum_{q\in\tilde{\mathcal{P}}^{(k+1)}} H_{qq}^{(k)}, \quad \text{for each } p \in \tilde{\mathcal{P}}^{(k+1)}.$$

Therefore, we have that

$$\sum_{q\in\tilde{\mathcal{P}}^{(k,k+1)}} H_{pq}^{(k)} \le \sum_{q\in\tilde{\mathcal{P}}^{(k)}} H_{qq}^{(k)} + \sum_{q\in\tilde{\mathcal{P}}^{(k+1)}} H_{qq}^{(k)}, \quad \text{for each } p \in \tilde{\mathcal{P}}^{(k,k+1)}.$$

Since $H_{qq}^{(k)}$ is the summation of all second derivative link lengths along path q, we have by Proposition 1-part (iii) that

$$H_{qq}^{(k)} \le A_1(\tilde{\ell}_q), \quad \text{for all } q,$$

24

where

$$A_1 \overset{\text{def}}{=} \frac{2}{(C_{\min})^2(1 - \rho_{\max})^3}.$$

So we can write

$$\sum_{q \in \tilde{\mathcal{P}}(k,k+1)} H_{pq}^{(k)} \leq A_1 \left(\sum_{q \in \tilde{\mathcal{P}}(k)} \tilde{\ell}_q + \sum_{q \in \tilde{\mathcal{P}}(k+1)} \tilde{\ell}_q \right)$$

$$= A_1 \left(\sum_{w \in \mathcal{W}^{(n)}} \sum_{q \in \tilde{\mathcal{P}}_w^{(k)}} \tilde{\ell}_q + \sum_{w \in \mathcal{W}^{(n)}} \sum_{q \in \tilde{\mathcal{P}}_w^{(k+1)}} \tilde{\ell}_q \right)$$

$$\leq A_1 \left(\frac{C_{\max}}{(C_{\min})(1 - \rho_{\max})^2} \right) \left(\sum_{w \in \mathcal{W}^{(n)}} \sum_{q \in \tilde{\mathcal{P}}_w^{(k)}} h_w^{(n)} + \sum_{w \in \mathcal{W}^{(n)}} \sum_{q \in \tilde{\mathcal{P}}_w^{(k+1)}} h_w^{(n)} \right)$$

(By Proposition 2)

$$= A_1 \left(\frac{C_{\max}}{(C_{\min})(1 - \rho_{\max})^2} \right) \left(\sum_{w \in \mathcal{W}^{(n)}} h_w^{(n)} \sum_{q \in \tilde{\mathcal{P}}_w^{(k)}} 1 + \sum_{w \in \mathcal{W}^{(n)}} h_w^{(n)} \sum_{q \in \tilde{\mathcal{P}}_w^{(k+1)}} 1 \right)$$

$$= A_1 \left(\frac{C_{\max}}{(C_{\min})(1 - \rho_{\max})^2} \right) \left(\sum_{w \in \mathcal{W}^{(n)}} h_w^{(n)} |\tilde{\mathcal{P}}_w^{(k)}| + \sum_{w \in \mathcal{W}^{(n)}} h_w^{(n)} |\tilde{\mathcal{P}}_w^{(k+1)}| \right)$$

$$\leq A_1 \left(\frac{C_{\max}}{(C_{\min})(1 - \rho_{\max})^2} \right)$$

$$\left(\max_{w \in \mathcal{W}^{(n)}} \left\{ \left| \tilde{\mathcal{P}}_w^{(k)} \right| \right\} \sum_{w \in \mathcal{W}^{(n)}} h_w^{(n)} + \max_{w \in \mathcal{W}^{(n)}} \left\{ \left| \tilde{\mathcal{P}}_w^{(k+1)} \right| \right\} \sum_{w \in \mathcal{W}^{(n)}} h_w^{(n)} \right)$$

$$= A_1 \left(\frac{C_{\max}}{(C_{\min})(1 - \rho_{\max})^2} \right)$$

$$\left(\max_{w \in \mathcal{W}^{(n)}} \left\{ \left| \tilde{\mathcal{P}}_w^{(k)} \right| \right\} + \max_{w \in \mathcal{W}^{(n)}} \left\{ \left| \tilde{\mathcal{P}}_w^{(k)} \right| \right\} + 1 \right) (|\mathcal{W}^{(n)}|)(h_{\text{avg}}^{(n)})$$

$$\leq A_1 \left(\frac{C_{\max}}{(C_{\min})(1 - \rho_{\max})^2} \right) \left(3 \max_{w \in \mathcal{W}^{(n)}} \left\{ \left| \tilde{\mathcal{P}}_w^{(k)} \right| \right\} \right) (|\mathcal{W}^{(n)}|)(h_{\text{avg}}^{(n)})$$

$$\left(\text{Since } 1 \leq \max_{w \in \mathcal{W}^{(n)}} \left\{ \left| \tilde{\mathcal{P}}_w^{(k)} \right| \right\} \right)$$

$$= \left(\frac{6(C_{\max}) \left(\max_{w \in \mathcal{W}^{(n)}} \left\{ \left| \tilde{\mathcal{P}}_w^{(k)} \right| \right\} \right)}{(C_{\min})^3(1 - \rho_{\max})^5} \right) (|\mathcal{W}^{(n)}|)(h_{\text{avg}}^{(n)}).$$

$\square$

Proposition 4:

$$\text{(i)} \quad D^* \geq \left(\frac{r_{\min}}{C_{\max}^{\max}} \right) (|\mathcal{W}^{(n)}|)(h_{\text{avg}}^{(n)}),$$

$$\text{(ii)} \quad D^{(0)} \leq \left(\frac{|\tilde{\mathcal{P}}_{\max}^{(0)}| \rho_{\max}}{1 - \rho_{\max}} \right) (|\mathcal{W}^{(n)}|)(h_{\text{avg}}^{(n)}).$$

Proof— Part (i): For any k, we have that

$$
\begin{aligned}
D^{(k)} &= \sum_{(i,j) \in \mathcal{L}} \frac{F_{ij}^{(k)}}{C_{ij} - F_{ij}^{(k)}} \\
&\geq \sum_{(i,j) \in \mathcal{L}} \frac{F_{ij}^{(k)}}{C_{ij}} \\
&\geq \frac{1}{C_{\max}} \sum_{(i,j) \in \mathcal{L}} F_{ij}^{(k)} \\
&= \frac{1}{C_{\max}} \sum_{w \in \mathcal{W}^{(n)}} \sum_{p \in \tilde{\mathcal{P}}_w^{(k)}} x_p^{(k)} \tilde{\ell}_p
\end{aligned}
$$

(By linearity and constraint Equation (**3**))

$$
\geq \frac{1}{C_{\max}} \sum_{w \in \mathcal{W}^{(n)}} \sum_{p \in \tilde{\mathcal{P}}_w^{(k)}} x_p^{(k)} h_w^{(n)}
$$

(By Proposition 2)

$$
= \frac{1}{C_{\max}} \sum_{w \in \mathcal{W}^{(n)}} h_w^{(n)} \sum_{p \in \tilde{\mathcal{P}}_w^{(k)}} x_p^{(k)}
$$

$$
= \frac{1}{C_{\max}} \sum_{w \in \mathcal{W}^{(n)}} h_w^{(n)} r_w
$$

(By constraint Equation (4) and the definition of $\tilde{\mathcal{P}}_w^{(k)}$)

$$
2 \left(\frac{r_{\min}}{C_{\max}} \right) \sum_{w \in \mathcal{W}^{(n)}} h_w^{(n)}
$$

$$
= \left(\frac{r_{\min}}{C_{\max}} \right) (|\mathcal{W}^{(n)}|)(h_{\text{avg}}^{(n)}).
$$

Now, because $D^* \leq D^{(k)}$ (by definition of D*) the result of part (i) is proven.

    Proof— Part (ii): Let $|\tilde{\mathcal{L}}^{(0)}|$ denote the number of active links at iteration 0, i.e., $|\tilde{\mathcal{L}}^{(0)}| \stackrel{\text{def}}{=} |\{(i,j) : F_{ij}^{(0)} > 0, (i,j) \in \mathcal{L}\}|.$

$$
\begin{aligned}
|\tilde{\mathcal{L}}^{(0)}| &\leq \sum_{w \in \mathcal{W}^{(n)}} \sum_{p \in \tilde{\mathcal{P}}_w^{(0)}} \tilde{\ell}_p \\
&\leq \sum_{w \in \mathcal{W}^{(n)}} \sum_{p \in \tilde{\mathcal{P}}_w^{(0)}} h_w^{(n)}
\end{aligned}
$$

(Since the initial paths are chosen to be shortest hop paths)

$$= \sum_{w \in \mathcal{W}^{(n)}} h_w^{(n)} \sum_{p \in \tilde{\mathcal{P}}_w^{(0)}} 1$$

$$\leq |\tilde{\mathcal{P}}_{\max}^{(0)}| \sum_{w \in \mathcal{W}^{(n)}} h_w^{(n)}$$

$$= (|\tilde{\mathcal{P}}_{\max}^{(0)}|)(|\mathcal{W}^{(n)}|)(h_{\text{avg}}^{(n)})$$

Next, by applying the above bound to the fact that

$$D^{(0)} \leq |\tilde{\mathcal{L}}^{(0)}| \left(\frac{\rho_{\max}}{1 - \rho_{\max}} \right),$$

the result of part (ii) is proven.

□

Proposition 5:

$$\nabla_x D^{(k)}(x^{(k+1)} - x^{(k)}) \leq - \left(\frac{\min_{w \in \mathcal{W}^{(n)}} \left\{ |\tilde{\mathcal{P}}_w^{(k)}| \right\}}{(C_{\max})^2} \right) \left(\frac{2(h_{\min}^{(n)})}{\alpha^{(k)}} \right) \| x^{(k+1)} - x^{(k)} \|^2, \quad \text{for all } k.$$

where $\alpha^{(k)}$ is the stepsize, $\| \cdot \|$ denotes the standard Euclidean norm:

$$\| x^{(k)} \| \stackrel{\text{def}}{=} \left(\sum_p |x_p^{(k)}|^2 \right)^{\frac{1}{2}}$$

and

$$x^{(k)} \stackrel{\text{def}}{=} [x_p^{(k)}]_{\substack{p \in \mathcal{P}_w^{(n)} \\ w \in \mathcal{W}^{(n)}}} .$$

Proof: This proposition is proven by applying the result of Lemma **4.3** of reference [12] and Proposition 1-part (iii) of the present paper. In [12], it is shown that

$$\nabla_x D^{(k)}(x^{(k+1)} - x^{(k)}) \leq - \left(\frac{\delta \cdot \eta}{\alpha^{(k)}} \right) \| x^{(k+1)} - x^{(k)} \|^2,$$

where

$$0 \leq \delta \leq \min_{p \in \tilde{\mathcal{P}}^{(k)}} \{ H_{pp}^{(k)} \}$$

and

$$0 \leq \eta \leq \min_{w \in \mathcal{W}^{(n)}} \left\{ |\tilde{\mathcal{P}}_w^{(k)}| \right\}.$$

27

For any active path p, $H_{pp}^{(k)}$ is the sum of at least $h_{min}^{(n)}$ second derivative link lengths (Proposition 2); so by Proposition 1-part (iii) we can choose $6 = \frac{2(h_{min}^{(n)})}{(C_{max})^2} \leq \min_{p \in \tilde{\mathcal{P}}(k)} \{H_{pp}^{(k)}\}$.

□

Proposition 6

$$\nabla_x D^{(k)} (x^{(k+1)} - y) \leq \left(\frac{2(C_{max})}{(C_{min})^3 (1 - \rho_{max})^5} \right) \left(\frac{h_{max}^{(n)}}{\alpha^{(k)}} \right) \| x^{(k+1)} - x^{(k)} \|, \quad \text{for all } k.$$

where $\| \cdot \|$ denotes the standard Euclidean norm, $x^{(k)} = [x_p^{(k)}]_{\substack{p \in P(n) \\ w \in \mathcal{W}^{(n)}}}$, and $y = [y_p]_{\substack{p \in \mathcal{P}_w^{(n)} \\ w \in \mathcal{W}^{(n)}}}$ is any nonnegative vector that satisfies

$$\sum_{p \in \mathcal{P}_w^{(n)}} y_p = r_w, \quad \text{for all } w \in \mathcal{W}^{(n)}.$$

Proof: This proposition is proven by applying the result of Lemma 4.4 of reference [12] along with Proposition 1-part (iii) and Proposition 2 of the present paper. In [12], it is shown that

$$\nabla_x D^{(k)} (x^{(k+1)} - y) \leq \left(\frac{A}{\alpha^{(k)}} \right) \| x^{(k+1)} - x^{(k)} \|,$$

where

$$A \geq \max_{p \in \tilde{\mathcal{P}}(k)} \{H_{pp}^{(k)}\}.$$

Since $H_{pp}^{(k)}$ is the sum of at most $\left(\frac{C_{max}}{(C_{min})(1 - \rho_{max})^2} \right) h_{max}^{(n)}$ second derivative link lengths (by Proposition 2), then we can apply Proposition 1-part (iii) in order to choose

$$\Delta = \left(\frac{2(C_{max})}{(C_{min})^3 (1 - \rho_{max})^5} \right) h_{max}^{(n)} \geq \max_{p \in \tilde{\mathcal{P}}(k)} \{H_{pp}^{(k)}\}.$$

□

Proposition 7:

$$D^{(k)} - D^* \leq \left(\frac{2 + \alpha^{(k)} (C_{min})(1 - \rho_{max}) \sqrt{3 \left(\max_{w \in \mathcal{W}^{(n)}} \left\{ |\tilde{\mathcal{P}}_w^{(k)}| \right\} \right) (|\mathcal{W}^{(n)}|)}}{\alpha^{(k)} (C_{min})(1 - \rho_{max})} \right)$$

$$\left(\frac{C_{max}}{(C_{min})^2 (1 - \rho_{max})^4} \right) \left(h_{max}^{(n)} \right) \| x^{(k+1)} - x^{(k)} \|, \quad \text{for all } k.$$

Proof: By convexity we have that

$$D^{(k)} - D^* \leq \nabla_x D^{(k)} (x^{(k)} - x^*)$$

$$= \nabla_x D^{(k)} (x^{(k+1)} - x^*) + \nabla_x D^{(k)} (x^{(k)} - x^{(k+1)})$$

$$\leq \left(\frac{2(C_{\max})}{(C_{\min})^3 (1 - \rho_{\max})^5} \right) \left(\frac{h_{\max}^{(n)}}{\alpha^{(k)}} \right) \parallel x^{(k+1)} - x^{(k)} \parallel + \nabla_x D^{(k)} (x^{(k)} - x^{(k+1)})$$

(By Proposition 6)

$$\leq \left(\frac{2(C_{\max})}{(C_{\min})^3 (1 - \rho_{\max})^5} \right) \left(\frac{h_{\max}^{(n)}}{\alpha^{(k)}} \right) \parallel x^{(k+1)} - x^{(k)} \parallel +$$

$$\left(\frac{C_{\max}}{(C_{\min})^2 (1 - \rho_{\max})^4} \right) \left(h_{\max}^{(n)} \right) \sum_{\substack{p \in \mathcal{P}_w^{(n)} \\ w \in \mathcal{W}^{(n)}}} (x_p^{(k)} - x_p^{(k+1)})$$

(By Proposition 1 − part (ii) and Proposition 2)

$$\leq \left(\frac{2(C_{\max})}{(C_{\min})^3 (1 - \rho_{\max})^5} \right) \left(\frac{h_{\max}^{(n)}}{\alpha^{(k)}} \right) \parallel x^{(k+1)} - x^{(k)} \parallel +$$

$$\left(\frac{C_{\max}}{(C_{\min})^2 (1 - \rho_{\max})^4} \right) \left(h_{\max}^{(n)} \right) \sum_{\substack{p \in \mathcal{P}_w^{(n)} \\ w \in \mathcal{W}^{(n)}}} |x_p^{(k)} - x_p^{(k+1)}|$$

$$= \left(\frac{2(C_{\max})}{(C_{\min})^3 (1 - \rho_{\max})^5} \right) \left(\frac{h_{\max}^{(n)}}{\alpha^{(k)}} \right) \parallel x^{(k+1)} - x^{(k)} \parallel +$$

$$\left(\frac{C_{\max}}{(C_{\min})^2 (1 - \rho_{\max})^4} \right) \left(h_{\max}^{(n)} \right) \parallel x^{(k)} - x^{(k+1)} \parallel_1$$

(Where $\parallel x^{(k)} \parallel_1 \overset{\text{def}}{=} \sum_p |x_p^{(k)}|$)

$$\leq \left(\frac{2(C_{\max})}{(C_{\min})^3 (1 - \rho_{\max})^5} \right) \left(\frac{h_{\max}^{(n)}}{\alpha^{(k)}} \right) \parallel x^{(k+1)} - x^{(k)} \parallel + (h_{\max}^{(n)}) \parallel x^{(k+1)} - x^{(k)} \parallel$$

$$\left(\frac{C_{\max}}{(C_{\min})^2 (1 - \rho_{\max})^4} \right) \sqrt{ \left(\max_{w \in \mathcal{W}^{(n)}} \left\{ |\tilde{\mathcal{P}}_w^{(k)}| \right\} + \max_{w \in \mathcal{W}^{(n)}} \left\{ |\tilde{\mathcal{P}}_w^{(k+1)}| \right\} \right) (|\mathcal{W}^{(n)}|) }$$

(Since the vector $(x^{(k)} - x^{(k+1)})$ has at most

$$\left(\max_{w \in \mathcal{W}^{(n)}} \left\{ |\tilde{\mathcal{P}}_w^{(k)}| \right\} + \max_{w \in \mathcal{W}^{(n)}} \left\{ |\tilde{\mathcal{P}}_w^{(k+1)}| \right\} \right) (|\mathcal{W}^{(n)}|) \text{ nonzero elements)}$$

$$\leq \left(\frac{C_{\max}}{(C_{\min})^2 (1 - \rho_{\max})^4} \right) \left(h_{\max}^{(n)} \right) \parallel x^{(k+1)} - x^{(k)} \parallel$$

$$\left(\left(\frac{2}{\alpha^{(k)} (C_{\min})(1 - \rho_{\max})} \right) + \sqrt{3 \left(\max_{w \in \mathcal{W}^{(n)}} \left\{ |\tilde{\mathcal{P}}_w^{(k)}| \right\} \right) (|\mathcal{W}^{(n)}|)} \right)$$

$$
= \left(\frac{2 + \alpha^{(k)}(C_{\min})(1 - \rho_{\max})\sqrt{3\left(\max_{w \in \mathcal{W}^{(n)}}\left\{\left|\tilde{\mathcal{P}}_w^{(k)}\right|\right\}\right)\left(|\mathcal{W}^{(n)}|\right)}}{\alpha^{(k)}(C_{\min})(1 - \rho_{\max})} \right)
$$

$$
\left(\frac{C_{\max}}{(C_{\min})^2(1 - \rho_{\max})^4} \right) \left(h_{\max}^{(n)} \right) \| x^{(k+1)} - x^{(k)} \| .
$$

$\square$

**Proposition 8:**

$$
\frac{\| x^{(k+1)} - x^{(k)} \|^2}{D^*} \geq \left(E^{(k)} \right)^2 D^* \left(\frac{\epsilon(C_{\min})^2(1 - \rho_{\max})^4}{(C_{\max})(h_{\max}^{(n)})} \right)^2
$$

$$
\left(\frac{\alpha^{(k)}(C_{\min})(1 - \rho_{\max})}{2 + \alpha^{(k)}(C_{\min})(1 - \rho_{\max})\sqrt{3\left(\max_{w \in \mathcal{W}^{(n)}}\left\{\left|\tilde{\mathcal{P}}_w^{(k)}\right|\right\}\right)\left(|\mathcal{W}^{(n)}|\right)}} \right)^2
$$

**Proof:** Apply the definition of $E^{(k)}$ to the inequality of Proposition 7 and solve for the desired bound.

$\square$

**Lemma 1:** Given that assumptions (A1) through (A3) are satisfied, the following holds for all $k \geq 0$:

$$
E^{(k+1)} \leq E^{(k)} \left[1 - E^{(k)} K f^{(n)} g^{(k)} \right] , \tag{$L1$}
$$

where

$$
K = \left(\frac{(\zeta_{\min})^2(1 - \zeta_{\max})(r_{\min})(C_{\min})^9(1 - \rho_{\max})^{15}}{6(\zeta_{\max})(C_{\max})^8(1 + C_{\max})^2} \right) , \tag{$L1.1$}
$$

$$
f^{(n)} = \frac{(h_{\min}^{(n)})^2}{(h_{\max}^{(n)})^2} , \tag{$L1.2$}
$$

and

$$
g^{(k)} = \frac{\left(\min_{w \in \mathcal{W}^{(n)}}\left\{\left|\tilde{\mathcal{P}}_w^{(k)}\right|\right\} \right)^2}{\left(\max_{w \in \mathcal{W}^{(n)}}\left\{\left|\tilde{\mathcal{P}}_w^{(k)}\right|\right\} \right)} . \tag{$L1.3$}
$$

**Proof:** Start with the Taylor's series expansion,

$$
D^{(k+1)} - D^* \leq D^{(k)} - D^* + \nabla_x D^{(k)}(x^{(k+1)} - x^{(k)}) + (x^{(k+1)} - x^{(k)})^T \nabla_x^2 D^{(k)}(x^{(k+1)} - x^{(k)}),
$$

30

which implies

$$D^{(k+1)} - D^* \leqslant D^{(k)} - D^* + \nabla_x D^{(k)} (x^{(k+1)} - x^{(k)}) + (x^{(k+1)} - x^{(k)})^T \widetilde{\nabla}_x^2 D^{(k,k+1)} (x^{(k+1)} - x^{(k)}).$$

$$(A.2)$$

Dividing Equation (A.2) by D* and applying Propositions 3 and 5, we have

$$E^{(k)} - E^{(k+1)} \geq \left(\frac{A_2^{(k)}}{\alpha^{(k)}} - A_3^{(k)} \right) \frac{\| x^{(k+1)} - x^{(k)} \|^2}{D^*}, \qquad (A.3)$$

where

$$A_2^{(k)} \stackrel{\text{def}}{=} \left(\frac{2(h_{\min}^{(n)}) \left(\min_{w \in \mathcal{W}^{(n)}} \left\{ \left| \tilde{\mathcal{P}}_w^{(k)} \right| \right\} \right)}{(C_{\max})^2} \right)$$

and

$$A_3^{(k)} \stackrel{\text{def}}{=} \left(\frac{6(C_{\max}) \left(\max_{w \in \mathcal{W}^{(n)}} \left\{ \left| \tilde{\mathcal{P}}_w^{(k)} \right| \right\} \right)}{(C_{\min})^3 (1 - \rho_{\max})^5} \right) (|\mathcal{W}^{(n)}|)(h_{\text{avg}}^{(n)}).$$

By substituting $\alpha^{(k)} = (\zeta_{\max}) \overline{\alpha}^{(k)}$ (i.e., the largest assumed value of $\alpha^{(k)}$) into Equation (A.3), we have

$$E^{(k)} - E^{(k+1)} \geq A_3^{(k)} \left(\frac{1 - \zeta_{\max}}{\zeta_{\max}} \right) \frac{\| x^{(k+1)} - x^{(k)} \|^2}{D^*} \qquad (A.4)$$

In a similar fashion, by substituting $\alpha^{(k)} = (\zeta_{\min}) \overline{\alpha}^{(k)}$ (i.e., the smallest assumed value for $\alpha^{(k)}$) into the bound given in Proposition 8, it is straightforward (and tedious) to show that

$$\frac{\| x^{(k+1)} - x^{(k)} \|^2}{D^*} \geq \left(E^{(k)} \right)^2 \left(\frac{D^* (\zeta_{\min})^2 (C_{\min})^{12} (1 - \rho_{\max})^{20} \left(\min_{w \in \mathcal{W}^{(n)}} \left\{ \left| \tilde{\mathcal{P}}_w^{(k)} \right| \right\} \right)^2 (h_{\min}^{(n)})^2}{36 (C_{\max})^8 (1 + C_{\max})^2 \left(\max_{w \in \mathcal{W}^{(n)}} \left\{ \left| \tilde{\mathcal{P}}_w^{(k)} \right| \right\} \right)^2 (|\mathcal{W}^{(n)}|)^2 (h_{\text{avg}}^{(n)})^2 (h_{\max}^{(n)})^2} \right).$$

$$(A.5)$$

By applying Proposition 4-part (i) to Equation (A.5), we get

$$\frac{\| x^{(k+1)} - x^{(k)} \|^2}{D^*} \geq \left(E^{(k)} \right)^2 \left(\frac{(\zeta_{\min})^2 (r_{\min}) (C_{\min})^{12} (1 - \rho_{\max})^{20} \left(\min_{w \in \mathcal{W}^{(n)}} \left\{ \left| \tilde{\mathcal{P}}_w^{(k)} \right| \right\} \right)^2 (h_{\min}^{(n)})^2}{36 (C_{\max})^9 (1 + C_{\max})^2 \left(\max_{w \in \mathcal{W}^{(n)}} \left\{ \left| \tilde{\mathcal{P}}_w^{(k)} \right| \right\} \right)^2 (|\mathcal{W}^{(n)}|)(h_{\text{avg}}^{(n)})(h_{\max}^{(n)})^2} \right).$$

$$(A.6)$$

Finally, by substituting the bound given in Equation (A.6) into Equation (A.4), we get

$$E^{(k)} - E^{(k+1)} \geq \left(E^{(k)} \right)^2 \left(\frac{(\zeta_{\min})^2 (1 - \zeta_{\max}) (r_{\min}) (C_{\min})^9 (1 - \rho_{\max})^{15} \left(\min_{w \in \mathcal{W}^{(n)}} \left\{ \left| \tilde{\mathcal{P}}_w^{(k)} \right| \right\} \right)^2 (h_{\min}^{(n)})^2}{6 (\zeta_{\max}) (C_{\max})^8 (1 + C_{\max})^2 \left(\max_{w \in \mathcal{W}^{(n)}} \left\{ \left| \tilde{\mathcal{P}}_w^{(k)} \right| \right\} \right) (h_{\max}^{(n)})^2} \right)$$

$$(A.7)$$

□

REFERENCES

[1] D. P. Bertsekas, B. Gendron and and W. K. Tsai, "Implementation of an optimal multicornrnodity network flow algorithm based on gradient projection and a path flow formulation," Mass. Inst. Technol., Cambridge, MA, Tech. Rep. LIDS-P-1364, Feb., 1984.

[2] D. P. Bertsekas and R. G. Gallager, Data Networks. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[3] L. R. Ford, Jr. and D. R. Fulkerson, Flows in Networks. Princeton, NJ: Princeton University Press, 1962.

[4] M. Schwartz, 'Analysis of congestion control techniques in computer routing and flow control in data networks," NATO Advanced Study *Inst.:* New Concepts in Multi-User Communications, Norwich, U.K., Aug. 4–16, 1980; Sijthoof and Nordhoof, Neth.

[5] J. N. Tsitsiklis and D. P. Bertsekas, 'Distributed asynchronous optimal routing for data networks," IEEE Trans. *Automat.* Contr. vol. AC-31, pp. 325–332, 1986.

[6] D. P. Bertsekas and J. N. Tsitsiklis, Parallel and Distributed *Computation:* Numerical Methods, Englewood Cliffs, NJ: Prentice-Hall, 1989.

[7] F. Kamoun, and L. Kleinrock, "Stochastic performance evaluation of hierarchical routing for large networks," Comput. Networks, vol. 3, pp. 337–353, 1979.

[8] W. K. Tsai, "Convergence of gradient projection routing methods in an asynchronous stochastic quasi-static virtual circuit network," IEEE Transactions on Automat. Contr., vol. 34, pp. 20–33, Jan. 1989.

[9] M. Schwartz and T. E. Stern, "Routing techniques used in computer communication networks," *IEEE* Trans. Commun., vol. COM-28, pp. 539–552, Apr. 1980.

[10] D. G. Luenberger, Introduction to Linear and Nonlinear Programming. Reading, MA: Addison-Wesley, 1973.

[11] R. S. Varga, *Matrix* Iterative Analysis. Englewood Cliffs, NJ: Prentice-Hall, 1962.

[12] W. K. Tsai, "Optimal quasi-static routing for virtual circuit networks subjected to stochastic inputs," Ph.D. Dissertation, Dep. EECS, Mass. Inst. Technol., Cambridge, MA, 1986.

[13] D. P. Bertsekas and E. M. Gafni, "Projection methods for variational inequalities with application to the traffic assignment problem," *Math. Programming Study 17*, North-Holland Publishing Company, pp. 139–159, 1982.

[14] A. V. Aho, J. E. Hoptcroft, and J. D. Ullman, *The Design and Analysis of Computer Algorithms.* Reading, MA: Addison-Wesley, 1974.

[15] W. K. Tsai, G. M. Huang, J. K. Antonio, and W. T. Tsai, "Distributed iterative aggregation algorithms for box-constrained minimization problems and optimal routing in data networks," *IEEE Trans. on Auto. Control,* Vol. 34, pp. 34–46, Jan. 1989.

[16] J. K. Antonio, "Fast distributed and parallel algorithms for data network control problems," Ph.D. Dissertation, Dep. of Electrical Engineering, Texas A&M University, College Station, TX, 1989.

[17] R. G. Gallager, "A minimum delay routing algorithm using distributed computation," *IEEE Trans. on Communications,* Vol. 25, pp. 73–85, Jan. 1977.

[18] B. Bollobás, *Random Graphs,* New York, NY: Academic Press, 1985.

[19] W. K. Tsai, "Complexity of gradient projection method for optimal routing in data networks," to appear, *IEEE Transaciions on Information Theory.*

[20] D. P. Bertsekas, "A class of optimal routing algorithms for communication networks," *Proceedings of the 5th International Conference on Computer Communications,* Atlanta, GA, 1980.