

Technical Notes and Correspondence

Time Complexity of a Path Formulated Optimal Routing Algorithm (Second Printing)¹

John K. Antonio, Wei K. Tsai, and G. M. Huang

Abstract—A detailed analysis of convergence rate is presented for an iterative path formulated optimal routing algorithm. In particular, it is quantified, 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. The analytical result proven in this note 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

It is known that route selection has a substantial impact on the performance of data networks [1]–[4], [6]–[8]. 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.

The goal in the present note is to determine the amount of time required for a class of iterative path formulated optimal routing algorithms to converge. The time complexity of a routing algorithm 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 note, 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 is paid in that the assumed model for computation is essentially synchronous (in terms of the order in which iterations are executed). 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 [2].

Manuscript received July 16, 1990; revised December 9, 1991, May 13, 1992, and November 2, 1992.

J. K. Antonio is with the School of Electrical Engineering, Purdue University, West Lafayette, IN 47907-1285 USA.

W. K. Tsai is with the Department of Electrical Engineering, University of California, Irvine, Irvine, CA 92717 USA.

G. M. Huang is with the Department of Electrical Engineering, Texas A&M University, College Station, TX 77843 USA.

IEEE Log Number 9213962.

¹This article was previously printed in the February 1994 issue, but is being reprinted due to errors made because of font incompatibilities with the typesetter.

Delay Models

Perhaps the simplest queuing model is the so-called $M/M/1$ queueing system that consists of a single queuing station and a single server. In such a system, the average delay for a packet to traverse link (i, j) is given by

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

where C_{ij} and F_{ij} denote the service rate and arrival rate respectively, associated with link (i, j) . Based on Jackson's Theorem and (1), the cost function assumed here is defined as a weighted sum of all link delays

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

where \mathcal{L} denotes the set of all network links and $D(F)$ is an estimate of the total number of outstanding packets in the network. For the purposes of this note, 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 to formally state the optimal routing problem. Throughout the note, 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 \mathcal{W} .
- r_w : The arrival rate (traffic demand) measured in packets/s, for the OD pair w .
- \mathcal{P}_w : For the OD pair w , this is the set of all paths connecting the origin node to the destination node.
- p : A generic path in \mathcal{P}_w .
- x_p : The flow rate on the p .

Constraint Equations

The following constraint equations arise naturally due to conservation of flow

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

$$\sum_{p \in \mathcal{P}_w} x_p = r_w \quad \text{for all } w \in \mathcal{W} \quad (4)$$

and

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

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

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

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

The Path Formulated Optimal Routing Problem

$$\text{Given } r_w, \quad \text{for each } w \in \mathcal{W}, \quad (7)$$

$$\text{minimize } \{D(x)\} \quad (8)$$

such that (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 [8] 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 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

$$d_p \stackrel{\text{def}}{=} \frac{\partial D(x)}{\partial x_p} = \sum_{\substack{\text{all links } (i, j) \\ \text{on path } p}} \frac{\partial}{\partial F_{ij}} \left\{ \frac{F_{ij}}{C_{ij} - F_{ij}} \right\}. \quad (9)$$

The minimum FDL (MFDL) paths, denoted as \bar{p}_w for each $w \in \mathcal{W}$, are defined by

$$d_{\bar{p}_w} = \min_{p \in \mathcal{P}_w} \{d_p\}, \quad \text{for each } w \in \mathcal{W}. \quad (10)$$

The iteration equation associated with the PFGP algorithm [3], [12] can now be stated

$$x_p^{(k+1)} = \max \{0, x_p^{(k)} - \alpha^{(k)} (H_{pp}^{(k)})^{-1} (d_p^{(k)} - d_{\bar{p}_w}^{(k)})\}, \quad \text{for all } w \in \mathcal{W}, p \in \mathcal{P}_w, p \neq \bar{p}_w \quad (11)$$

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 $\alpha^{(k)} (H_{pp}^{(k)})^{-1} (d_p^{(k)} - d_{\bar{p}_w}^{(k)}) \geq 0$, 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

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

A more efficient version of the PFGP algorithm (as described originally in [12]) is

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

$$x_{\bar{p}_w}^{(k+1)} = r_w - \sum_{p \in \tilde{\mathcal{P}}_w, p \neq \bar{p}_w} x_p^{(k)}, \quad \text{for all } w \in \mathcal{W}, \quad (13)$$

$$\tilde{\mathcal{P}}_w^{(k+1)} = \{p \in \tilde{\mathcal{P}}_w^{(k)} | x_p^{(k)} - \alpha^{(k)} (H_{pp}^{(k)})^{-1} (d_p^{(k)} - d_{\bar{p}_w}^{(k)}) > 0\} \cup \{\bar{p}_w\}, \quad \text{for all } w \in \mathcal{W}. \quad (14)$$

The PFGP algorithm of (12)–(14) has been efficiently implemented as a serial FORTRAN code, see [1].

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 (12)–(14), 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 complexity of each iteration 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 $|\tilde{\mathcal{P}}_w^{(k)}|$

$$|\tilde{\mathcal{P}}_w^{(k)}| \leq k + |\tilde{\mathcal{P}}_w^{(0)}|, \quad \text{for all } w \in \mathcal{W}, k \geq 0 \quad (15)$$

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, the main concern in this note is to estimate the 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, see for example, [5, pp. 338–340]. Unfortunately, it is difficult to determine meaningful bounds for the smallest and largest eigenvalues, primarily because the number of active paths can 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 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.)

B. Serial Versus Distributed Time Complexities

For the purposes of this note 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. It is known that the distributed PFGP algorithm will actually converge (eventually) in a virtually totally asynchronous computing environment, see [3].

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 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 $\tilde{\mathcal{P}}_w^{(k)}$.

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 paths that connect the origin node to the destination node.

$\tilde{\mathcal{P}}_w^{(k)}$: 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

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

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

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

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

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

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

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

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 $w \in \mathcal{W}^{(n)}$:

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

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

$$h_{\text{avg}}^{(n)} = \frac{1}{|\mathcal{W}^{(n)}|} \sum_{w \in \mathcal{W}^{(n)}} h_w^{(n)}.$$

B. Assumptions

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

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

A3) There exists constants ζ_{\min} and ζ_{\max} , such that $\alpha^{(k)}$ (the step size at iteration k) satisfies

$$(\zeta_{\min})\bar{\alpha}^{(k)} \leq \alpha^{(k)} \leq (\zeta_{\max})\bar{\alpha}^{(k)},$$

where

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

and

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

Assumption A1) ensures synchronous execution of the iterations. Assumption A2) requires that 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 $k \geq 0$. In practical networks, this assumption is realistic provided that a flow control mechanism is used to regulate the total amount of traffic allowed to enter the network. Finally, assumption A3) requires the step size to lie within a specified interval.

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

The main convergence rate results hinge around the derivation of a bound for a nonlinear recurrence involving $E^{(k)}$. The recurrence, which captures the dynamic response of $E^{(k)}$, is stated in Lemma 1. A general closed form bound for this recurrence is proven in Lemma 2.

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

$$E^{(k+1)} \leq E^{(k)} [1 - E^{(k)} K f^{(n)} g^{(k)}] \quad (\text{L1})$$

where

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

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

and

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

Proof: A series of eight initial propositions are needed before beginning the proof of the lemma. These initial propositions are stated in the appendix.

To prove the lemma, start with the Taylor's series expansion

$$\begin{aligned} D^{(k+1)} - D^* &\leq D^{(k)} - D^* + \nabla_x D^{(k)} (x^{(k+1)} - x^{(k)}) \\ &\quad + (x^{(k+1)} - x^{(k)})^T \nabla_x^2 D^{(k)} (x^{(k+1)} - x^{(k)}) \end{aligned}$$

which implies

$$\begin{aligned} D^{(k+1)} - D^* &\leq D^{(k)} - D^* + \nabla_x D^{(k)} (x^{(k+1)} - x^{(k)}) \\ &\quad + (x^{(k+1)} - x^{(k)})^T \nabla_x^2 D^{(k, k+1)} (x^{(k+1)} - x^{(k)}). \quad (\text{L1.4}) \end{aligned}$$

Dividing (L1.4) by D^* and applying Propositions 3 and 5, we have

$$E^{(k)} - E^{(k+1)} \geq \left(\frac{A_2^{(k)}}{\alpha^{(k)}} - A_3^{(k)} \right) \frac{\|x^{(k+1)} - x^{(k)}\|^2}{D^*} \quad (\text{L1.5})$$

where

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

and

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

By substituting $\alpha^{(k)} = (\zeta_{\max})\bar{\alpha}^{(k)}$ (i.e., the largest assumed value of $\alpha^{(k)}$) into (L1.5), we have

$$E^{(k)} - E^{(k+1)} \geq A_3^{(k)} \left(\frac{1 - \zeta_{\max}}{\zeta_{\max}} \right) \frac{\|x^{(k+1)} - x^{(k)}\|^2}{D^*}. \quad (\text{L1.6})$$

In a similar fashion, by substituting $\alpha^{(k)} = (\zeta_{\min})\bar{\alpha}^{(k)}$ (i.e., the smallest assumed value for $\alpha^{(k)}$) into the bound given in Proposition 8, it is straightforward (and tedious) to show (L1.7), as shown at the bottom of the page. By applying Proposition 4-part i) to (L1.7), we get (L1.8), shown at the bottom of the page. Finally, by substituting the bound given in (L1.8) into (L1.6), we get (L1.9), as displayed at the bottom of the page. ■

Lemma 2: Given that assumptions A1)–A3) are satisfied, the following holds for all $k \geq 0$

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

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

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

Proof: By the definition of $E^{(k)}$, it follows that $E^{(k)} \geq 0$ for all $k \geq 0$. Also, from (L1), $E^{(k+1)} \leq E^{(k)}$. Therefore, the following expression for $E^{(k+1)}$ is obtained by bounding the right-hand side of (L1)

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

From the above expression, a bound for $E^{(k+1)}$ is given by

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

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

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

D. The Main Results

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

Theorem 1: Given that assumptions A1)–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

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

where

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

and

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

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

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

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

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

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

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

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 $|\tilde{\mathcal{P}}_w^{(k)}| \leq (k+1)^{1-\gamma}$, for all w , then the mapping G^{-1} is polynomial.

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

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

$$\frac{\|x^{(k+1)} - x^{(k)}\|^2}{D^*} \geq (E^{(k)})^2 \left(\frac{D^* (\zeta_{\min})^2 (C_{\min})^{12} (1 - \rho_{\max})^{20} (\min_{w \in \mathcal{W}^{(n)}} \{|\tilde{\mathcal{P}}_w^{(k)}|\})^2 (h_{\min}^{(n)})^2}{36(C_{\max})^8 (1 + C_{\max})^2 (\max_{w \in \mathcal{W}^{(n)}} \{|\tilde{\mathcal{P}}_w^{(k)}|\})^2 (|\mathcal{W}^{(n)}|)^2 (h_{\text{avg}}^{(n)})^2 (h_{\max}^{(n)})^2} \right) \quad (\text{L1.7})$$

$$\frac{\|x^{(k+1)} - x^{(k)}\|^2}{D^*} \geq (E^{(k)})^2 \left(\frac{(\zeta_{\min})^2 (r_{\min}) (C_{\min})^{12} (1 - \rho_{\max})^{20} (\min_{w \in \mathcal{W}^{(n)}} \{|\tilde{\mathcal{P}}_w^{(k)}|\})^2 (h_{\min}^{(n)})^2}{36(C_{\max})^9 (1 + C_{\max})^2 (\max_{w \in \mathcal{W}^{(n)}} \{|\tilde{\mathcal{P}}_w^{(k)}|\})^2 (|\mathcal{W}^{(n)}|) (h_{\text{avg}}^{(n)}) (h_{\max}^{(n)})^2} \right) \quad (\text{L1.8})$$

$$E^{(k)} - E^{(k+1)} \geq (E^{(k)})^2 \left(\frac{(\zeta_{\min})^2 (1 - \zeta_{\max}) (r_{\min}) (C_{\min})^9 (1 - \rho_{\max})^{15} (\min_{w \in \mathcal{W}^{(n)}} \{|\tilde{\mathcal{P}}_w^{(k)}|\})^2 (h_{\min}^{(n)})^2}{6(\zeta_{\max}) (C_{\max})^8 (1 + C_{\max})^2 (\max_{w \in \mathcal{W}^{(n)}} \{|\tilde{\mathcal{P}}_w^{(k)}|\}) (h_{\max}^{(n)})^2} \right) \quad (\text{L1.9})$$

Proof: Apply the assumption of the lemma to (L1.3) and apply the definition of $G^{(k)}$. Refer to [13] for a detailed proof. ■

The next task involves upper bounding the quantity $(1/f^{(n)})$. 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 for $(1/f^{(n)})$ [refer to (L1.2)] is as follows

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

Fortunately, 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 [10, pp. 233–236], it is proven that if the probability of any link being in the graph is p , denoted by $P[(i, j) \in \mathcal{L}] = p$, then with probability one, the diameter of the graph will equal either d or $d + 1$, where d satisfies the equation

$$p^d n^{d-1} = \log(n^2). \quad (17)$$

The following lemmas result from (17).

Lemma 4: Provided that $P[(i, j) \in \mathcal{L}] \geq (2/n)$, 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 $p = (2/n)$ in (17) and solve for the above bound for $d^{(n)}$. ■

Lemma 5: For any constant $0 < \delta \leq 1$, provided that $P[(i, j) \in \mathcal{L}] \geq (2/n^{1-\delta})$ 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: Set $p = (2/n^{1-\delta})$ in (17) and apply straightforward algebraic bounding techniques, see [13] for a detailed proof. ■

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 3–5 to Theorem 1. Assumptions A1)–A3) are assumed for all three corollaries. Also, it is assumed that $(1/\epsilon) > K_2$ (otherwise, $N_I = 0$).

Corollary 1: For any three constants $0 < \epsilon$, $0 < \delta \leq 1$, and $0 < \gamma \leq 1$, provided that $P[(i, j) \in \mathcal{L}] \geq (2/n^{1-\delta})$, for all $i \neq j$, and provided that $|\mathcal{P}_w^{(k)}| \leq (k+1)^{1-\gamma}$, 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 $P[(i, j) \in \mathcal{L}] \geq (2/n^{1-\delta})$, 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

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

Corollary 3: For any two constants $0 < \epsilon$ and $0 < \gamma \leq 1$, provided that $P[(i, j) \in \mathcal{L}] \geq (2/n)$, for all $i \neq j$, and provided that $|\mathcal{P}_w^{(k)}| \leq (k+1)^{1-\gamma}$, 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 inconsistent with the only other known convergence rate results for the algorithm under consideration. In [7] 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 [7], however, is the rate with which the parameter β approaches unity as the number of nodes n is increased.) It is noted that if the number of nodes were assumed to be fixed (i.e., independent of both n and k) in the present note, then both $f^{(n)}$ and $g^{(k)}$ are bounded by constants, say \bar{f} and \bar{g} , respectively. Therefore, for $E^{(k)} \geq 1$ [refer to (L1)] $E^{(k)} \leq E^{(0)}\bar{\beta}^k$, where $\bar{\beta} = (1 - K_1\bar{f}\bar{g})$. Likewise, for $E^{(k)} < 1$, $E^{(k)} \leq 1/(1 + kK_1\bar{f}\bar{g})$.

In [11], 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 note, however, no bound is placed on 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 were 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. 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., sublinear) 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 sublinear function of k .

APPENDIX

This Appendix contains the eight initial propositions (and the associated notation) needed for the proof of Lemma 1. The proofs of these eight propositions can be found in [13].

Additional Notation:

\tilde{l}_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

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

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

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

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

Next, the eight propositions are stated. Assumptions A1)–A3) are implicitly assumed.

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

- i) $0 \leq D_{ij}^{(k)} \leq \frac{\rho_{\max}}{(1-\rho_{\max})}$,
- ii) $\frac{1}{C_{\max}} \leq \frac{\partial D_{ij}^{(k)}}{\partial F_{ij}^{(k)}} \leq \frac{1}{(C_{\min})^2(1-\rho_{\max})^2}$,
- iii) $\frac{2}{(C_{\max})^2} \leq \frac{\partial^2 D_{ij}^{(k)}}{(\partial F_{ij}^{(k)})^2} \leq \frac{2}{(C_{\min})^2(1-\rho_{\max})^3}$,

where

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

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

$$h_w^{(n)} \leq \tilde{l}_p \leq \left(\frac{C_{\max}}{(C_{\min})(1-\rho_{\max})^2} \right) h_w^{(n)},$$

for all $p \in \tilde{p}_w^{(k)}$, $w \in \mathcal{W}^{(n)}$, k .

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

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

for all k .

Proposition 4:

- i) $D^* \geq \left(\frac{r_{\min}}{C_{\max}} \right) (|\mathcal{W}^{(n)}|)(h_{\text{avg}}^{(n)})$,
- ii) $D^{(0)} \leq \left(\frac{|\tilde{p}_{\max}^{(0)}| \rho_{\max}}{1-\rho_{\max}} \right) (|\mathcal{W}^{(n)}|)(h_{\text{avg}}^{(n)})$.

Proposition 5:

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

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

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

and

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

Proposition 6:

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

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

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

Proposition 7:

$$D^{(k)} - D^* \leq \left(\frac{2 + \alpha^{(k)}(C_{\min})(1-\rho_{\max}) \sqrt{3(\max_{w \in \mathcal{W}^{(n)}} \{|\tilde{p}_w^{(k)}|\}) (|\mathcal{W}^{(n)}|)}}{\alpha^{(k)}(C_{\min})(1-\rho_{\max})} \right) \cdot \left(\frac{C_{\max}}{(C_{\min})^2(1-\rho_{\max})^4} \right) (h_{\max}^{(n)}) \|x^{(k+1)} - x^{(k)}\|, \quad \text{for all } k.$$

Proposition 8:

$$\frac{\|x^{(k+1)} - x^{(k)}\|^2}{D^*} \geq (E^{(k)})^2 D^* \left(\frac{\epsilon(C_{\min})^2(1-\rho_{\max})^4}{(C_{\max})(h_{\max}^{(n)})} \right)^2 \cdot \left(\frac{\alpha^{(k)}(C_{\min})(1-\rho_{\max})}{2 + \alpha^{(k)}(C_{\min})(1-\rho_{\max}) \sqrt{3(\max_{w \in \mathcal{W}^{(n)}} \{|\tilde{p}_w^{(k)}|\}) (|\mathcal{W}^{(n)}|)}} \right)^2.$$

REFERENCES

- [1] D. P. Bertsekas, B. Gendron, and W. K. Tsai, "Implementation of an optimal multicommodity network flow algorithm based on gradient projection and a path flow formulation," Tech. Rep. LIDS-P-1364, Mass. Inst. Technol., Cambridge, MA, Feb. 1984.
- [2] D. P. Bertsekas and R. G. Gallager, *Data Networks*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
- [3] 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.
- [4] M. Schwartz and T. E. Stern, "Routing techniques used in computer communication networks," *IEEE Trans. Commun.*, vol. COM-28, pp. 539–552, Apr. 1980.
- [5] D. G. Luenberger, *Introduction to Linear and Nonlinear Programming*. Reading, MA: Addison-Wesley, 1973.
- [6] 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.
- [7] D. P. Bertsekas and E. M. Gafni, "Projection methods for variational inequalities with application to the traffic assignment problem," *Math. Programming Study 17*. Amsterdam, The Netherlands: North-Holland Publishing Company, pp. 139–159, 1982.
- [8] 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. Automat. Contr.*, vol. 4, pp. 34–46, Jan. 1989.
- [9] R. G. Gallager, "A minimum delay routing algorithm using distributed computation," *IEEE Trans. Communications*, vol. 25, pp. 73–85, Jan. 1977.
- [10] B. Bollobás, *Random Graphs*. New York: Academic Press, 1985.
- [11] W. K. Tsai, J. K. Antonio, and G. M. Huang, "Complexity of gradient projection method for optimal routing in data networks," *IEEE Trans. Inf. Theory*, submitted.
- [12] D. P. Bertsekas, "A class of optimal routing algorithms for communication networks," in *Proc. 5th Int. Conf. Computer Communications*, Atlanta, GA, 1980.
- [13] J. K. Antonio, W. K. Tsai, and G. M. Huang, "Time complexity of a path formulated optimal routing algorithm," Tech. Rep. TR-EE-92-42, Purdue Univ., School of Electrical Engineering, West Lafayette, IN, Oct. 1992.