

# Traffic Rate Network Tomography via Moment Generating Function Matching

Yariv Ephraim, Joshua Coblenz, Brian L. Mark  
Dept. of Electrical and Computer Engineering  
George Mason University  
Fairfax, VA 22030, U.S.A.  
{ yephraim, jcoblen2, bmark }@gmu.edu

Hanoch Lev-Ari  
Dept. of Electrical and Computer Engineering  
Northeastern University  
Boston, MA 02115, U.S.A.  
levari@ece.neu.edu

**Abstract**—Network tomography aims at estimating source-destination traffic rates from link traffic measurements. This inverse problem was formulated by Vardi in 1996 for independent Poisson traffic over networks operating under deterministic as well as random routing regimes. Vardi used a second-order moment matching approach to estimate the rates where a solution for the resulting linear matrix equation was obtained using an iterative minimum I-divergence procedure. Vardi’s second-order moment matching approach was recently extended to higher order cumulant matching approach with the goal of improving the rank of the system of linear equations. In this paper we go one step further and develop a moment generating function matching approach for rate estimation, and seek a least squares as well as an iterative minimum I-divergence solution of the resulting linear equations. We also specialize this approach to a characteristic function matching approach which exhibits some advantages. These follow from the fact that the characteristic function matching approach results in fewer conflicting equations involving the empirical estimates. We demonstrate that the new approach outperforms the cumulant matching approach while being conceptually simpler.

**Index Terms**—Network traffic, network tomography, inverse problem, moment generating function.

## I. INTRODUCTION

The goal of network tomography is to estimate the rates of traffic flows over source-destination pairs from traffic flows over links of the networks. Traffic flow rates are measured by the number of packets per second. Network tomography was formulated by Y. Vardi in the seminal paper [1]. Suppose that  $X$  denotes a vector of  $L$  source-destination traffic flows, and that  $Y$  denotes a vector of  $M$  link traffic flows. Normally,  $M \ll L$ . All source-destination traffic flows are assumed independent Poisson random variables. Thus,  $X$  comprises a vector of  $L$  independent Poisson random variables. Link traffic flow measurements are assumed passive and require no probes. For networks operating under a deterministic routing regime, traffic flows from each source node is routed to the destination node over a fixed path. Traffic flows over each link may originate from multiple source-destination traffic flows. Thus, we define a binary variable  $a_{ij}$  such that  $a_{ij} = 1$  when traffic over source-destination  $j$  passes through link  $i$  and  $a_{ij} = 0$  otherwise. We also define an  $M \times L$  binary routing matrix

$A = \{a_{ij}, i = 1, \dots, M; j = 1, \dots, L\}$  which is assumed known throughout the paper. It follows that  $Y = AX$ . The rate vector is the expected value of  $X$  which we denote by  $\lambda = E\{X\}$ .

The network tomography problem is that of estimating the rate vector  $\lambda$  from  $N$  independent realizations of  $Y$ . This is a rather challenging inverse problem since  $Y = AX$  is an underdetermined set of equations, and the estimate of  $\lambda$  must be non-negative. Vardi referred to the problem as a “LININPOS” which stands for LINear INverse POSitive problem. Maximum likelihood estimation of  $\lambda$  is not feasible since the components of  $Y$  are dependent Poisson random variables with no explicitly known distribution. The expectation-maximization (EM) algorithm is also not useful for this problem as it requires calculation of the conditional mean  $E\{X | Y\}$  which in turns requires the infinite solutions of  $Y = AX$ . Vanderbei and Iannone [2] developed an EM algorithm in which  $E\{X | Y\}$  is simulated. An alternative popular approach is to rely on the explicit form of  $E\{X | Y\}$  for jointly Gaussian  $X$  and  $Y$ , see, e.g., [1], [3], [4].

Vardi resorted to moment matching in which  $\lambda$  is estimated from a linear matrix equation relating the first two empirical moments of  $Y$  to the corresponding first two theoretical moments of  $AX$ . The approach is applicable to networks operating under deterministic as well as random routing regimes. In the latter case, there are multiple alternative paths for each source-destination pair which can be selected according to some probability law. Vardi invoked an iterative procedure for estimating  $\lambda$  which was previously developed for image deblurring [5] and as an EM iteration for Positron Emission Tomography (PET). This procedure will be specified in Eq. (17). Vardi’s work led to extensive research in the areas of network inference and medical tomography.

Extension of Vardi’s second-order moment matching approach to higher-order cumulants matching was considered in [6]. The higher-order cumulants introduce new useful information on the unavailable distribution of  $Y$  which can be leveraged in estimating the source-destination flow rates. By using a sufficient number of empirical cumulants, the linear mapping involved in the cumulant matching approach can achieve full column rank. In an ideal scenario where the cumulant matching equations are consistent, and the cumulants

of the link traffic flows are accurately known, the rate vector can be recovered without error. It was demonstrated in [6] that this ideal situation is approachable when a sufficiently large number of realizations of  $Y$  is available.

In this paper we explore estimation of the rate vector  $\lambda$  using moment generating function (MGF) matching. The rate is estimated from equations relating the theoretical and empirical MGF of  $Y$  and its derivatives. This is an extension of the moment matching approach to a function that represents moments of all orders. The approach was proposed by Quandt and Ramsey in 1978 [7]–[9] for estimating a mixture of distributions. Under independent Poisson path traffic flows, the MGF matching equations are linear just like in the cumulant matching approach.

A vast literature exists on network tomography. Key references include the work of Vanderbei and Iannone [2] mentioned earlier, and the work of Tebaldi and West [10] on Bayesian rate estimation using Markov Chain Monte Carlo simulation. See also [3], [4], [11]–[13] on related approaches. Other work discuss network topology discovery from traffic measurements [14]–[17]. A somewhat outdated survey can be found in [18]. A more recent survey [19] discusses network tomography in conjunction with network coding. The work in [3], [4], [18] attempt to implement maximum likelihood estimation of the source-destination rates from link data under a Gaussian rather than a Poisson traffic model. In [12], conditions for identifiability of higher order cumulants in estimation of source-destination traffic from link measurements were established.

The plan for this paper is as follows. In Section II we address rate estimation in networks with deterministic routing. We develop the rate estimation equations which include matching of the MGF and its first three derivatives. Rate estimation from these equations is performed using least squares and Vardi's iterative approach. That approach was shown to a minimum  $I$ -divergence recursion (17). Rate estimation in networks operating under a random routing regime follows from application of the rate estimation approach for deterministic routing networks to an appropriately constructed super-network [6], [10]. Numerical results for the NSFnet [20] are presented in Section III. Concluding remarks are given in Section IV.

## II. MGF MATCHING IN DETERMINISTIC ROUTING

In this section we present our MGF matching approach. The equations include the MGF and its first three derivatives.

### A. MGF Matching

Let  $\theta = \text{col}\{\theta_1, \dots, \theta_M\}$  denote the parameter of the MGF. Expressing  $A = [a_1, \dots, a_L]$  where  $a_j$  is the  $j$ th column of  $A$ , the MGF of  $Y$  is given by

$$M_\theta(Y) = E\{e^{\theta'Y}\} = E\{e^{\theta'AX}\} = M_{\theta'A}\{X\} \quad (1)$$

where  $\cdot'$  denotes matrix transpose. By independence of the components  $\{X_j\}$  of  $X$ , and the Poisson distribution of each  $X_j$  with mean  $\lambda_j$ , we have

$$\begin{aligned} M_\theta(Y) &= \prod_{j=1}^L M_{\theta'a_j}(X_j) \\ &= \prod_{j=1}^L \exp[\lambda_j(e^{\theta'a_j} - 1)]. \end{aligned} \quad (2)$$

Hence,

$$\log M_\theta(Y) = (e^{\theta'A} - \mathbf{1}) \lambda \quad (3)$$

where

$$e^{\theta'A} - \mathbf{1} := (e^{\theta'a_1} - 1, \dots, e^{\theta'a_L} - 1). \quad (4)$$

The empirical MGF of  $Y$  given  $N$  realizations  $\{y_1, \dots, y_N\}$  is given by

$$\begin{aligned} \log M_\theta(Y) &= \log E\{e^{\theta'Y}\} \\ &\approx \log \frac{1}{N} \sum_{n=1}^N e^{\theta'y_n}. \end{aligned} \quad (5)$$

Hence, the zeroth-order MGF matching equation is

$$\log \frac{1}{N} \sum_{n=1}^N e^{\theta'y_n} = (e^{\theta'A} - \mathbf{1}) \lambda. \quad (6)$$

We next supplement this equation with some derivatives of the MGF evaluated at the same  $\theta$ . This will enable increasing the column rank of the linear mapping from which  $\lambda$  is estimated. Let  $g : R^K \rightarrow R^K$  be any function of  $y_n$  where  $K \in \{M, M^2, M^3\}$ . Define

$$\varphi(g(y_n)) := \sum_{n=1}^N \frac{e^{\theta'y_n}}{\sum_{m=1}^N e^{\theta'y_m}} g(y_n) \quad (7)$$

which can be interpreted as the expected value of  $g(Y)$  under the probability mass function

$$p_Y(y_n) = \frac{e^{\theta'y_n}}{\sum_{m=1}^N e^{\theta'y_m}} \quad (8)$$

for  $Y \in \{y_1, \dots, y_N\}$ . Let  $D(e^{\theta'A})$  denote a diagonal matrix with  $(e^{\theta'a_1}, \dots, e^{\theta'a_M})$  on the main diagonal. Higher-order derivatives of (6) are concisely expressed in terms of the Khatri-Rao product defined by

$$A \odot A := [a_1 \otimes a_1, a_2 \otimes a_2, \dots, a_L \otimes a_L] \quad (9)$$

where  $\otimes$  denotes the Kronecker product.

The first-order derivative of (6) is given by

$$\varphi(y_n) = A \cdot D(e^{\theta'A}) \cdot \lambda. \quad (10)$$

Let  $\bar{y}_n = y_n - \varphi(y_n)$ . The second-order derivative of (6) is given by

$$\varphi(\bar{y}_n \otimes \bar{y}_n) = (A \odot A) \cdot D(e^{\theta'A}) \cdot \lambda. \quad (11)$$

The third-order derivative of (6) is given by

$$\varphi(\bar{y}_n \otimes \bar{y}_n \otimes \bar{y}_n) = (A \odot A \odot A) \cdot D(e^{\theta' A}) \cdot \lambda. \quad (12)$$

The high-order moment matching approach of [6] is obtained from (10)-(12) when  $\theta \rightarrow 0$ . We have

$$\begin{aligned} \frac{1}{N} \sum_{n=1}^N y_n &= A\lambda \\ \frac{1}{N} \sum_{n=1}^N \bar{y}_n \otimes \bar{y}_n &= (A \odot A)\lambda \\ \frac{1}{N} \sum_{n=1}^N \bar{y}_n \otimes \bar{y}_n \otimes \bar{y}_n &= (A \odot A \odot A)\lambda \end{aligned} \quad (13)$$

where  $\bar{y}_n = y_n - \sum_{k=1}^N y_k / N$ . The LHS of (13) constitutes the empirical moments of the observed link traffic measurement vector. In [6] the  $K$ -statistics were used instead. For any practical value of  $N$ , the  $K$ -statistics are essentially the empirical moments shown in (13).

The rate vector  $\lambda$  can be estimated as the solution of any subset of the four equations in (6), (10), (11) and (12). Stacking the first  $r$  equations for  $r = 1, 2, 3, 4$  yields a linear matrix equation in  $\lambda$  which can be expressed as

$$\hat{\eta}_r(Y) = \mathcal{A}_r \lambda. \quad (14)$$

In this equation  $\hat{\eta}_r(Y)$ ,  $r = 2, 3, 4$ , contains values of the empirical MGF of  $Y$  for the mentioned orders of  $r = 2, 3, 4$ , and  $\hat{\eta}_1(Y)$  is the LHS of (6). The matrix  $\mathcal{A}_r$ ,  $r = 2, 3, 4$ , represents a matrix with stacked Khatri-Rao products of  $A$ , and from (6),  $\mathcal{A}_1 = (e^{\theta' A} - \mathbf{1})$ .

Estimation of  $\lambda$  can be performed using the least squares approach. The unique Tikhonov regularized least squares solution for the possibly inconsistent set of equations (14), when  $\mathcal{A}_r$  is not necessarily full column rank, is given by [21, p. 51]

$$\hat{\lambda}_r = (\mathcal{A}_r^* \mathcal{A}_r + \gamma I)^{-1} \mathcal{A}_r^* \hat{\eta}_r(Y) \quad (15)$$

for some  $\gamma > 0$ . Note that the regularized estimator applies to a skinny as well as a fat matrix  $\mathcal{A}_r$ .

We could change the relative weight of the equations in (14) by multiplying the equations corresponding to a specific derivative of the MGF by some factor smaller than one. We found it useful to multiply the equations corresponding to the third-order derivative by  $\epsilon = .01$ . The rationale here is similar to the regularization of the moments used by Vardi in [1]. Following this approach, we denote the  $\epsilon$ -weighted matrix  $\mathcal{A}_r$  by  $\mathcal{A}_{r,\epsilon}$ , and let the vector  $\hat{\eta}_r(Y)$  be denoted by  $\hat{\eta}_{r,\epsilon}(Y)$ . Then, from (15), the least squares rate vector estimate is given by

$$\hat{\lambda}_{r,\epsilon} = (\mathcal{A}_{r,\epsilon}^* \mathcal{A}_{r,\epsilon} + \gamma I)^{-1} \mathcal{A}_{r,\epsilon}^* \hat{\eta}_{r,\epsilon}(Y). \quad (16)$$

Note that the estimator (16) is not guaranteed to be non-negative. A non-negative estimate of  $\lambda$  can be obtained by using non-negative least squares optimization [22]. This approach did not lead to good results in our earlier work [6]. Instead, we have arbitrarily substituted negative estimates in our numeric examples with the value of .001. This approach

resulted in substantially lower MSE compared to using the constrained optimization algorithm of [22, p. 161]. The performance of the algorithm should not be affected by this arbitrary substitution since negative estimates are rare at our working point.

Estimation of  $\lambda$  could also be performed using the iterative approach of Vardi [1]. This approach guarantees non-negative rate estimates. In this paper we study the least squares estimator (16) and the iterative approach and compare their performance. The iterative solution of (14) is stated as follows:

$$\lambda_j^{\text{new}} = \lambda_j^{\text{old}} \sum_{i=1}^{M_a} \bar{\mathcal{A}}_r(i,j) \frac{\hat{\eta}_{r,i}(Y)}{(\mathcal{A}_r \lambda^{\text{old}})_i}, \quad j = 1, \dots, L, \quad (17)$$

where  $\lambda_j^{\text{old}}$  denotes a current estimate of the  $j$ th component of  $\lambda$ , and  $\lambda_j^{\text{new}}$  denotes the new estimate of that component at the conclusion of the iteration;  $\hat{\eta}_{r,i}(Y)$  is the  $i$ th component of  $\hat{\eta}_r(Y)$  in (14);

$$\bar{\mathcal{A}}_r(i,j) := \frac{\mathcal{A}_r(i,j)}{\sum_{t=1}^{M_a} \mathcal{A}_r(t,j)} \quad (18)$$

with  $\mathcal{A}_r(i,j)$  being the  $(i,j)$ th component of  $\mathcal{A}_r$ ; and  $M_a$  is the number of equations in (14). This procedure is in fact an EM iteration in the positron emission tomography problem, which follows a similar formulation as network tomography, but with the crucially facilitating difference that the Poisson components of  $Y$  are now *independent* random variables [23], [24]. Clearly this assumption does not hold for the network tomography problem. The iteration was independently but heuristically developed for image deblurring [5], [25]. In that context, it was shown by Snyder, Schulz and O'Sullivan [25] to monotonically decrease Csiszár's I-divergence [26] between the original image convolved with the kernel, and the observed blurred image.

When the characteristic function  $M_{i\theta}(Y) = E\{e^{i\theta' Y}\}$  is used instead of the MGF, the above equations (7), (6), (10), (11) and (12) become, respectively,

$$\varphi(g(y_n)) := \sum_{n=1}^N \frac{e^{i\theta' y_n}}{\sum_{m=1}^N e^{i\theta' y_m}} g(y_n) \quad (19)$$

$$\log \frac{1}{N} \sum_{n=1}^N e^{i\theta' y_n} = (e^{i\theta' A} - \mathbf{1}) \lambda. \quad (20)$$

$$\varphi(iy_n) = iA \cdot D(e^{i\theta' A}) \cdot \lambda. \quad (21)$$

$$\varphi(i\bar{y}_n \otimes i\bar{y}_n) = i^2 (A \odot A) \cdot D(e^{i\theta' A}) \cdot \lambda. \quad (22)$$

$$\varphi(i\bar{y}_n \otimes i\bar{y}_n \otimes i\bar{y}_n) = i^3 (A \odot A \odot A) \cdot D(e^{i\theta' A}) \cdot \lambda. \quad (23)$$

Arranging Eqs. (20)-(23) as in (14) results in a linear equation for  $\lambda$ . We shall not repeat this equation here and refer to (14) as either the MGF matching equation or as the characteristic function matching equation, respectively. Estimating  $\lambda$  by matching the characteristic function rather than the MGF is

advantageous since some equations in the MGF approach may be inconsistent and hence are ignored in the rate estimation process [6]. Inconsistency occurs when the centralized moments in (10)-(12) are negative while the RHS of each of these equations is always non-negative. This problem does not arise when matching the characteristic function as all equations involved are complex. Thus, rows in the characteristic function matching equation are not removed.

### B. Complexity

The computational effort in the MGF matching approach is similar to that seen in the cumulant matching approach of [6]. The set of equations (14) contains  $n_r(M) = 1 + M + M^2 + M^3$  when the first three derivatives of the MGF are used. Construction of the right hand side of (14) requires  $(M^2 + M^3)L$  operations. Construction of the left hand side of (14) requires  $(M^2 + M^3)N$  operations where  $N$  is the number vectors used to estimate the MGF and its derivatives. Solving the equations requires effort that depends only on  $n_r(M)$ ,  $M$  and the number of iterations in (17). The combined effort is dominated by  $n_r(M)N$  since  $N$  must be large to produce meaningful MGF estimates. Thus the computational effort of the MGF matching approach is approximately linear in  $N$  when  $N$  is large which is always the case.

## III. NUMERICAL EXAMPLE

In this section we demonstrate the performance of the moment generating function matching approach and of the characteristic function matching approach in estimating the rate vector of the source-destination traffic in the NSFnet network [20]. We compare these two new approaches with the recently developed high-order cumulant matching approach [6]. In the moment generating function matching approach the rate vector  $\lambda$  is estimated as the solution of (6), (10)-(12) with  $r = 3$ . In the characteristic function matching approach the rate vector  $\lambda$  is estimated from (20)-(23) with  $r = 3$ .

We study the NSFnet [20] whose topology is shown in Fig. 1. The network consists of 14 nodes and 21 bidirectional links. Hence, it contains  $L = 14 \cdot 13/2 = 91$  source-destination pairs. This size network may represent a private network, a transportation network or a subset of interest of a large-scale network. The link weights in Fig. 1 are exclusively used to determine  $k \geq 1$  shortest paths for each source-destination pair. Otherwise, they play no role in the traffic rate estimation problem. To determine the  $k$  shortest paths between a given source-destination pair, we used the shortest simple paths function from the NetworkX Python library, which is based on the algorithm of Yen [27]. When  $k = 1$ , the number of source-destination paths equals the number of source-destination pairs and the routing matrix  $A$  is a  $21 \times 91$  matrix. The augmented routing matrix  $\mathcal{A}_r$  achieves full column rank when  $r = 2$ .

With  $k \geq 2$ , we can assign multiple paths to each source-destination pair and treat them as distinguishable new source-destination pairs. The routing matrix  $A$  thus becomes fatter and using higher-order empirical cumulants becomes beneficial.

For example, when  $k = 2$ , we have  $L = 182$  source-destination paths, the column rank of  $\mathcal{A}_2$  is 162 and  $\mathcal{A}_3$  has full column rank. Thus, using this example we focus on third-order MGF matching.

The network with  $k = 2$  and a  $21 \times 182$  routing matrix  $A$  could also be seen as a super-network in the Tebaldi-West sense [10] for a network with  $L = 91$  source-destination pairs operating under a random routing regime with two possible paths per each source-destination pair. The accuracy of the rate estimation for the random routing network is determined by the accuracy of the rate estimation in the deterministic routing super-network. Thus, it suffices to focus on rate estimation in the deterministic routing network with the  $21 \times 182$  routing matrix  $A$ .

In our numerical examples the source-destination rates comprising the arrival rate vector  $\lambda$  are drawn independently from a uniform distribution on the interval  $[0, 4]$ , i.e.,

$$\lambda_j \sim U[0, 4], \quad j = 1, \dots, L. \quad (24)$$

In this case, the mean and variance of each component of  $\lambda$  are given by 2 and  $16/12 \approx 1.333$ , respectively.

In each of  $T = 500$  simulation runs, a rate vector  $\lambda$  was generated according to (24). The rate vector was subsequently used in generating  $N$  statistically independent identically distributed Poisson vectors  $\{X_n\}$  which were transformed into the vectors  $\{Y_n = AX_n\}$  using the assumed known routing matrix  $A$ . The  $N$  statistically independent identically distributed Poisson vectors  $\{Y_n\}$  were used in estimating the rate vector in the current run. Equations (6) and (10)-(12) were used for MGF matching, and Equations (19)-(23) were used for characteristic function matching. For the third-order derivative we have used the regularization value of  $\epsilon_3 = .01$ . The least squares regularization factor was set to  $\gamma = .0005$ . The iteration (17) was initialized uniformly with each  $\lambda_j = .01$ . It was terminated at the conclusion of 300 iterations. We experimented with  $N = 5000$ ,  $N = 10000$  and  $N = 20000$  samples. Since the matrix  $A$  is binary  $\{0, 1\}$ , the matrix  $\mathcal{A}_r$  in (14) may contain duplicate rows. All duplicates were removed before the rate vector was estimated. The parameter  $\theta$  of the MGF is generated randomly in each run. It is drawn uniformly from  $[0, 0.001]$ .

Let  $\lambda_t(j)$  and  $\hat{\lambda}_t(j)$  denote, respectively, the  $j$ th component of  $\lambda$  and its estimate at the  $t$ th run where  $j = 1, \dots, L$  and  $t = 1, \dots, T$ . For each estimate we evaluated the *normalized MSE* defined by

$$\xi_j^2 = \frac{\frac{1}{T} \sum_{t=1}^T (\lambda_t(j) - \hat{\lambda}_t(j))^2}{\frac{1}{T} \sum_{t=1}^T (\lambda_t(j))^2} \quad (25)$$

and the *averaged normalized MSE* defined by

$$\bar{\xi}^2 = \frac{1}{L} \sum_{j=1}^L \xi_j^2. \quad (26)$$

The MSE in estimating  $\lambda_j$  is approximately  $\xi_j^2 \cdot E\{\lambda^2(j)\}$  when  $T$  is sufficiently large.

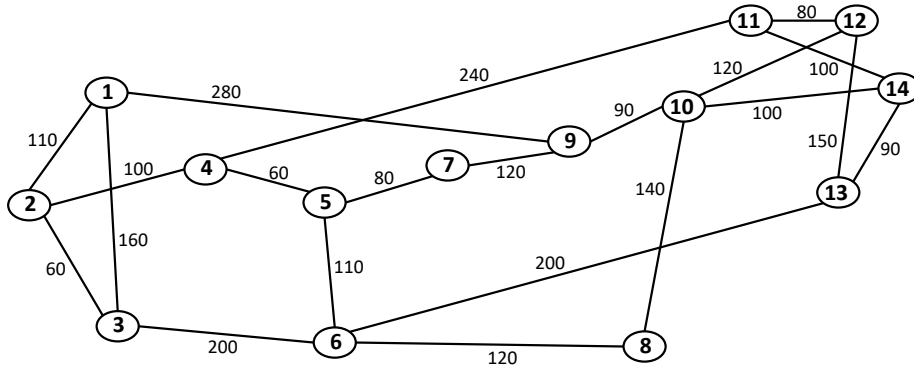


Fig. 1. NSFnet topology with link weights as in [20, Fig. 4].

$N$	MGF			Characteristic function		
	5K	10K	20K	5K	10K	20K
$\xi^2$	0.1585	0.0960	0.0629	0.1517	0.0898	0.0579
Percent Neg.	6.4055	4.2582	3.0560	7.4440	4.9956	3.4363

TABLE I

COMPARISON OF MGF AND CHARACTERISTIC FUNCTION MATCHING IN ESTIMATING THE RATES OF SOURCE-DESTINATION PATHS IN THE NSFNET.  $r = 3$  IN THIS EXPERIMENT.  $\xi^2$  IS THE AVERAGED NORMALIZED MSE IN LEAST SQUARES ESTIMATION. PERCENT NEG. IS THE PERCENT OF RATES THAT WERE ESTIMATED AS NEGATIVE RATES BY THE LEAST SQUARES ESTIMATOR

$N$	Cumulants			Characteristic function		
	5K	10K	20K	5K	10K	20K
$\xi^2$	0.1934	0.0950	0.0625	0.1517	0.0898	0.0579
Percent Neg.	6.2418	4.2275	3.0626	7.4440	4.9956	3.4363

TABLE II

COMPARISON OF CUMULANTS MATCHING AND CHARACTERISTIC FUNCTION MATCHING IN ESTIMATING THE RATES OF SOURCE-DESTINATION PATHS IN THE NSFNET.  $r = 3$  IN THIS EXPERIMENT.  $\xi^2$  IS THE AVERAGED NORMALIZED MSE IN LEAST SQUARES ESTIMATION. PERCENT NEG. IS THE PERCENT OF RATES THAT WERE ESTIMATED AS NEGATIVE RATES BY THE LEAST SQUARES ESTIMATOR

Our numerical results are shown in Tables I-V. Table I provides a comparison between the rate accuracy in least squares MGF matching and least squares characteristic function matching where in both cases the function and its three derivatives are matched. Clearly, rates estimated by the characteristic function matching approach are more accurate for the three values of  $N$ . Table II provides a comparison between rate accuracy in the cumulant matching approach of [6] and the least squares characteristic function matching discussed here based on three derivatives, i.e.,  $r = 3$ . The table shows lower averaged normalized MSE for the least squares characteristic function matching approach. We emphasize that similar results are obtained when the characteristic function matching equations (19)-(23) are replaced by the moment matching equations (13) evaluated at  $\theta = 0$ . This suggests that the better results obtained by the characteristic function matching approach are due to its use of the complete available data without removal of inconsistent equations as was done in [6]. Table III provides the averaged normalized MSE and percent of negative estimates for the least squares characteristic function matching approach using  $r = 1, 2, 3$  derivatives and  $N = 5000, 10000$  and  $20000$  link traffic vectors. Generally, the accuracy improves with increasing  $r$  and  $N$ .

In Table IV we have constructed the matrix  $\mathcal{A}_r$  in (14) by stacking  $N_\theta = 10$  sets of equations (19)-(23) where each set is evaluated at a different  $\theta$ . This approach represents an attempt to complete the column rank of the matrix  $\mathcal{A}_r$  and thus aim at a unique least squares estimate. The results show that this

$N$	$\xi^2$			Percent Neg.		
	$r = 1$	$r = 2$	$r = 3$	$r = 1$	$r = 2$	$r = 3$
5K	0.3037	0.1476	0.1517	0.1934	7.1571	7.444
10K	0.3037	0.0897	0.0898	0.1890	4.8110	4.9956
20K	0.3037	0.0595	0.0579	0.1868	3.3945	3.4363

TABLE III

AVERAGED NORMALIZED MSE ( $\xi^2$ ) AND PERCENT NEGATIVE ESTIMATED RATES IN LEAST SQUARES CHARACTERISTIC FUNCTION MATCHING ESTIMATION AS A FUNCTION OF THE NUMBER OF DERIVATIVES ( $r$ ) AND NUMBER OF LINK TRAFFIC MEASUREMENTS

approach is not warranted. Using  $N_\theta = 1$  outperformed the case with  $N_\theta = 10$  in all working points of interest. Table V provides the averaged normalized MSE in estimating the rate vector  $\lambda$  by applying the minimum  $I$ -divergence iteration (17) to the characteristic function and its  $r$  derivatives for  $r = 1, 2, 3$  and  $N = 5000, 10000$  and  $20000$ . Comparing with

$N_\theta$	$\xi^2$			Percent Neg.		
	$r = 1$	$r = 2$	$r = 3$	$r = 1$	$r = 2$	$r = 3$
1	0.3037	0.0897	0.0897	0.1890	4.8110	4.9956
10	0.3003	0.0900	0.1928	0.1791	4.8484	9.1407

TABLE IV

AVERAGED NORMALIZED MSE ( $\xi^2$ ) AND PERCENT NEGATIVE ESTIMATED RATES IN LEAST SQUARES ESTIMATION AS A FUNCTION OF THE NUMBER OF DERIVATIVES ( $r$ ) OF THE MGF USING  $N_\theta = 10$  DIFFERENT  $\theta$  VECTORS AND  $N = 10000$  SAMPLES

$N$	$\xi^2$		
	$r = 1$	$r = 2$	$r = 3$
5K	0.2293	0.1534	0.1638
10K	0.2293	0.0996	0.1035
20K	0.2293	0.0703	0.0704

TABLE V

AVERAGED NORMALIZED MSE ( $\xi^2$ ) IN MINIMUM I-DIVERGENCE RATE ESTIMATION (17) AS A FUNCTION OF THE NUMBER OF DERIVATIVES ( $r$ ) OF THE MGF

Table III shows that the results are generally comparable but the least squares estimator is better at the important working point of  $r = 3, N = 20000$ . The advantage of the iteration (17) is that the rate estimates are always positive.

#### IV. CONCLUDING REMARKS

We have developed a framework for MGF matching approach for estimating the rates of source-destination Poisson traffic flows from link traffic flows. The approach is equally applicable to networks operating under deterministic as well as random routing strategies. Under independent Poisson source-destination traffic flows, the approach boils down to a set of linear equations relating values of the MGF of the link measurements  $Y$ , and its derivatives, to the rate vector  $\lambda$ , via a matrix involving Khatri-Rao products of the routing matrix. We studied least squares rate estimation as well as iterative minimum I-divergence estimation. We demonstrated that the related characteristic function matching approach outperforms the MGF matching approach. It also outperforms the cumulant matching approach of [28] while being conceptually simpler. The advantage of the characteristic function matching approach follows from its tolerance to equations that are considered inconsistent in the MGF matching approach.

#### REFERENCES

- [1] Y. Vardi, "Network tomography: Estimating source-destination traffic intensities from link data," *J. Am. Stat. Assoc.*, vol. 91, no. 433, pp. 365–377, 1996.
- [2] R. J. Vanderbei and J. Iannone, "An EM approach to OD matrix estimation," Princeton University, Tech. Rep. SOR 94-04, 1994.
- [3] J. Cao, D. Davis, S. Vander Wiel, and B. Yu, "Time-varying network tomography: Router link data," *J. Am. Stat. Assoc.*, vol. 95, pp. 1063–1075, 2004.
- [4] G. Liang and B. Yu, "Maximum pseudo likelihood estimation in network tomography," *IEEE Trans. Signal Process.*, vol. 51, no. 8, pp. 2043–2053, Aug. 2003.
- [5] W. H. Richardson, "Bayesian-based iterative method of image restoration," *J. Opt. Soc. Amer.*, vol. 62, pp. 55–59, Jan 1972.
- [6] H. Lev-Ari, Y. Ephraim, and B. L. Mark, "Traffic rate network tomography with higher-order cumulants," arXiv, Tech. Rep. arXiv:2012.07224 [cs.PF], Dec. 2020. [Online]. Available: <https://arxiv.org/abs/2012.07224>
- [7] R. E. Quandt and J. B. Ramsey, "Estimating mixtures of normal distributions and switching regressions," *J. Am. Stat. Assoc.*, vol. 73, no. 364, pp. 730–738, Dec. 1978.
- [8] —, "Estimating mixtures of normal distributions and switching regressions: Rejoinder," *J. Am. Stat. Assoc.*, vol. 73, no. 364, pp. 751–752, Dec. 1978.
- [9] —, "Estimating mixtures of normal distributions and switching regressions: Rejoinder," *J. Am. Stat. Assoc.*, vol. 74, no. 364, p. 56, Mar. 1979.

- [10] C. Tebaldi and M. West, "Bayesian inference on network traffic using link count data (with discussion)," *J. Am. Stat. Assoc.*, vol. 93, no. 442, pp. 557–576, 1998.
- [11] M. Mardani and G. B. Giannakis, "Estimating traffic and anomaly maps via network tomography," *IEEE/ACM Trans. Netw.*, vol. 24, no. 3, pp. 1533–1547, June 2016.
- [12] H. Singhal and G. Michailidis, "Identifiability of flow distributions from link measurements with applications to computer networks," *Inverse Problems*, vol. 23, pp. 1821–1849, 2007.
- [13] M. L. Hazelton, "Network tomography for integer-valued traffic," *Ann. Appl. Stat.*, vol. 9, no. 1, pp. 474–506, 2015.
- [14] X. Jin, W.-P. Yiu, S.-H. Chan, and Y. Wang, "Network topology inference based on end-to-end measurements," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 12, pp. 2182–2195, Dec. 2006.
- [15] B. Eriksson, G. Dasarathy, P. Barford, and R. Nowak, "Toward the practical use of network tomography for Internet topology discovery," in *Proc. IEEE INFOCOM*, San Diego, CA, USA, March 2010, pp. 1–9.
- [16] J. Ni, H. Xie, S. Tatikonda, and Y. R. Yang, "Efficient and dynamic routing topology inference from end-to-end measurements," *IEEE/ACM Trans. Netw.*, vol. 18, no. 1, pp. 123–135, Feb. 2010.
- [17] A. Anandkumar, A. Hassidim, and J. Kelner, "Topology discovery of sparse random graphs with few participants," in *ACM SIGMETRICS'11*, San Jose, CA, USA, June 2011.
- [18] R. Castro, M. Coates, G. Liang, R. Nowak, and B. Yu, "Network tomography: Recent developments," *Statistical Science*, vol. 19, no. 3, pp. 499–517, 2004.
- [19] P. Qin, B. Dai, B. Huang, G. Xu, and K. Wu, "A survey on network tomography with network coding," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1981–1995, 2014.
- [20] L. H. Bonani, "Modeling an optical network operating with hybrid-switching paradigms," *J. Microw. Optoelectron. Electromagn. Appl.*, vol. 15, no. 4, pp. 275–292, 2016.
- [21] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*, 1st ed. New Jersey: Prentice Hall, 2000.
- [22] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*. New Jersey: Prentice-Hall, 1974.
- [23] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. 1, pp. 113–122, 1982.
- [24] Y. Vardi, "Applications of the EM algorithm to linear inverse problems with positivity constraints," in *Image Models (and their Speech Cousins)*, S. E. Levinson and L. A. Shepp, Eds. Springer, 1966, pp. 183–198.
- [25] D. L. Snyder, T. J. Schulz, and J. A. O'Sullivan, "Deblurring subject to nonnegativity constraints," *IEEE Trans. Sig. Proc.*, vol. 40, 1992.
- [26] I. Csiszár, "Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems," *Ann. Stat.*, vol. 19, no. 4, pp. 2032–2066, Dec. 1991.
- [27] J. Y. Yen, "Finding the k shortest loopless paths in a network," *Management Science*, vol. 17, no. 11, pp. 712–716, Jul. 1971.
- [28] H. Lev-Ari, "Efficient solution of linear matrix equations with application to multistatic antenna array processing," *Commun. in Inf. Syst.*, vol. 5, no. 1, pp. 123–130, 2005.