Received: 28 September 2021 | Revised: 20 June 2022 | Accepted: 17 September 2022

DOI: 10.1002/net.22127

RESEARCH ARTICLE

WILEY

Traffic rate network tomography with higher-order cumulants

Hanoch Lev-Ari[1] | Yariv Ephraim[2] | Brian L. Mark[2]

[1]Department of Electrical and Computer Engineering, Northeastern University, Boston, Massachusetts, USA

[2]Department of Electrical and Computer Engineering, George Mason University, Fairfax, Virginia, USA

Correspondence

Brian L. Mark, Department of Electrical and Computer Engineering, George Mason University, Fairfax, VA 22030, USA.

Email: bmark@gmu.edu

Funding information

National Science Foundation, Grant/Award Number: 1717033

Abstract

Network tomography aims at estimating source–destination traffic rates from link traffic measurements. This inverse problem was formulated by Vardi in 1996 for Poisson traffic over networks operating under deterministic as well as random routing regimes. In this article, we expand Vardi's second-order moment matching rate estimation approach to higher-order cumulant matching with the goal of increasing the column rank of the mapping and consequently improving the rate estimation accuracy. We develop a systematic set of linear cumulant matching equations and express them compactly in terms of the Khatri–Rao product. Both least squares estimation and iterative minimum I-divergence estimation are considered. We develop an upper bound on the mean squared error (MSE) in least squares rate estimation from empirical cumulants. We demonstrate that supplementing Vardi's approach with the third-order empirical cumulant reduces its minimum averaged normalized MSE in rate estimation by almost 20% when iterative minimum I-divergence estimation was used.

KEYWORDS

higher-order cumulants, inverse problem, mean squared error, network tomography, network traffic, Poisson model

1 | INTRODUCTION

Network tomography was formulated by Vardi in the seminal paper [34]. The goal is to estimate the rates of traffic flows over source–destination pairs from traffic flows over links of a network. For concreteness, we shall discuss the problem in terms of "packet" arrivals in a communication network as in [34]. A packet comprises several hundreds of bytes. The model has also attracted considerable interest in the context of transportation networks, see, for example, [11, 30, 32]. Mathematically, network tomography is a challenging inverse problem defined on a network with random independent source–destination arrival counts, which induce dependent link arrival counts. Traffic flow rates are measured, for example, by the number of packets per second. We use X to denote a vector of L source–destination traffic counts, and Y to denote a vector of M link traffic counts. Normally, $M \ll L$. The packet counts for all source–destination traffic flows are assumed to be 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. A probe is a special packet transmitted to a certain destination or destinations in the network for the purpose of obtaining or measuring a particular performance metric, such as end-to-end delay. For networks operating under a deterministic routing regime, a traffic flow is routed from a source node 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 article. It follows that $Y = AX$. The expected value of the jth Poisson source–destination traffic flow constitutes its traffic flow rate. Thus, the vector of source–destination traffic flow rates is the expected value of X which we denote by $\lambda = E\{X\}$.

The network tomography problem is that of estimating the rate vector λ from given realizations of Y. In this inverse problem, $Y = AX$ is an underdetermined set of equations and the estimate of λ must be non-negative. Vardi invoked the terminology "LININPOS" for linear inverse positive problem. Maximum likelihood estimation of λ 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 [32] developed an EM algorithm in which $E\{X|Y\}$ is simulated. An alternative common approach is to rely on the explicit form of $E\{X|Y\}$ for jointly Gaussian X and Y, see, for example, [4, 18, 34].

Vardi resorted to moment matching in which λ 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 λ which was previously developed for image deblurring [25]. This useful procedure will be discussed in Section 2. Vardi's work ignited extensive research in the areas of network inference and medical tomography.

In this article, we explore the benefits of supplementing Vardi's second-order moment matching approach with higher-order empirical cumulants. Here too the cumulant matching equation is linear in λ, and the approach is applicable to networks operating under deterministic as well as random routing regimes. 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 may 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 as the unique solution of the cumulant matching equations. As we demonstrate in Section 5, this ideal situation is approachable when a sufficiently large number of realizations of Y is available.

A vast literature exists on network tomography. Several forms of network tomography have been studied in the literature depending on the type of measurement data and the network performance parameters (e.g., rate, delay) of interest: (i) source–destination path-level traffic rate estimation based on link-level traffic measurements [4, 11, 18, 21, 27, 30, 34]; (ii) link-level parameter estimation based on end-to-end path-level traffic measurements [3, 10, 19, 20, 31, 32]; (iii) network topology discovery from traffic measurements [1, 9, 12, 22]. Approaches to network tomography can also be characterized as active (e.g., [3, 15, 19, 24]), whereby explicit control traffic is injected into the network to extract measurements, or passive (e.g., [4, 10, 34, 35]), whereby the observation data are obtained from existing network traffic. A somewhat outdated survey circa 2004 can be found in [5]. A more recent survey [23] discusses network tomography in conjunction with network coding.

Closest to Vardi's work is the contemporary work of Vanderbei and Iannone [32] which relies on a Poisson model for incoming traffic. The goal is to estimate the rate of traffic on each link connecting input and output nodes from traffic counts at the input and output nodes. Vanderbei and Iannone did not resort to moment matching but rather developed a simulation-based EM algorithm for maximum likelihood estimation of the rates. A thorough Bayesian approach to Vardi's problem was developed by Tebaldi and West [30] using Markov chain Monte Carlo simulation. See also [11]. Another closely related work to Vardi's problem appeared in [4, 5, 18], where maximum likelihood estimation of the source–destination rates from link data was implemented under a Gaussian rather than a Poisson traffic model. In [21], source–destination rates were estimated by utilizing spatiotemporal correlation of nominal traffic, and the fact that traffic anomalies are sparse. In [27], conditions for identifiability of higher-order cumulants in estimation of source–destination traffic from link measurements were established. In [20], an algorithm was developed for choosing a set of linearly independent source–destination measurement paths from which additive link metrics are individually estimated.

The main contribution of this article is to systematically advance the theory of traffic rate network tomography by increasing the rank of the matrix in the associated inverse problem through the use of higher-order cumulant statistics. As in most of the literature on traffic rate network tomography, we assume that traffic flows follow a Poisson model. The Poisson assumption is common in applications to transportation networks [11]. In modern communication networks, however, traffic at the granularity of packet arrivals tends to be much more "bursty" than Poisson. However, our cumulant-based methodology could be extended to more general traffic models, for example, the mixed Poisson model studied in [8], which can be used to characterize the more bursty traffic processes found in communication networks.

The plan for this article is as follows. In Section 2, we present the minimum I-divergence iterative procedure which plays an important role in this article. In Section 3, we address rate estimation in networks with deterministic routing. We develop a new systematic set of cumulant matching equations for estimating the rate vector from up to the fourth-order empirical cumulant. We also develop an upper bound on the mean squared error (MSE) in least squares estimation of the rate vector from empirical cumulants. We conclude this section with a discussion on some implementation issues of the proposed cumulant matching approach. Complexity of the approach is discussed in Section 3.4. In Section 4, we discuss rate estimation in networks operating under a random routing regime. Numerical results for the NSFnet [2] are presented in Section 5. Concluding remarks are given in Section 6. Technical details of the derivations are deferred to the Appendix.

2 | MINIMUM I-DIVERGENCE ITERATION

Under the Poisson model for X, the moment (and cumulant) matching approach results in a set of linear equations in λ. Generally speaking, the set of equations has the form of $\hat{\eta}(Y) = B\lambda$ where $\hat{\eta}(Y)$ is a vector of M_b empirical moments (first and second-order in [34]) of Y, $B = \{b_{ij}, i = 1, \dots, M_b; j = 1, \dots, L\}$ is a constant zero-one matrix that depends on A but not on λ, and $B\lambda$ represents the corresponding theoretical moments. The exact value of M_b as a function of M is discussed in Section 3.3.

To estimate λ that satisfies $\hat{\eta}(Y) = B\lambda$, Vardi proposed to use an iterative estimation approach which originated in the study of another inverse problem concerning image deblurring [25, 28]. Let λ_j^{old} denote a current estimate of the jth component of λ, and let λ_j^{new} denote the new estimate of that component at the conclusion of the iteration. Let $(B\lambda^{\text{old}})_i$ denote the ith component of $B\lambda^{\text{old}}$. Similarly, let $\hat{\eta}_i(Y)$ denote the ith component of $\hat{\eta}(Y)$. The iteration is given by

$$\lambda_j^{\text{new}} = \lambda_j^{\text{old}} \sum_{i=1}^{M_b} \overline{b}_{ij} \frac{\hat{\eta}_i(Y)}{(B\lambda^{\text{old}})_i}, \quad \text{where } \overline{b}_{ij} := \frac{b_{ij}}{\sum_{t=1}^{M_b} b_{tj}}, \tag{1}$$

for $j = 1, \dots, L$. This iteration is particularly suitable for solving positive inverse problems. It reaches a fixed point when the moment matching equation $\hat{\eta}(Y) = B\lambda^{\text{old}}$ is satisfied. The iteration was studied by Snyder, Schulz, and O'Sullivan [28] in a similarly formulated application of image deblurring. It was shown to monotonically decrease Csiszár's I-divergence [7] between the original image convolved with the kernel, and the observed blurred image. The procedure turned out to be 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 [26, 33]. This iteration will play a central role in our cumulant matching approach.

3 | CUMULANT MATCHING IN DETERMINISTIC ROUTING

In this section, we present our cumulant matching approach and its theoretical performance bound for networks operating under a deterministic routing regime. The goal is to estimate the source–destination rate vector λ from N realizations of the link traffic flow Y where $Y = AX$ and the routing matrix A is known. Conditions for identifiability of λ were given in [34]. Specifically, the parameter λ is identifiable if all columns $\{a_1, \dots, a_L\}$ of A are distinct and none is null.

3.1 | Cumulant matching

Let $\overline{X} := X - E\{X\}$ where X is any real random vector with finite mean and possibly dependent components. The vectorized kth order central moment of X is given by

$$\mu_k(X) = E\{\underbrace{\overline{X} \otimes \overline{X} \otimes \cdots \otimes \overline{X}}_{k \text{ times}}\}, \tag{2}$$

where k is any positive integer and $\otimes$ denotes the Kronecker product. When the length of X is L, the length of the vector $\mu_k(X)$ is L^k. For $Y = AX$, it follows from the identity

$$(A_1B_1) \otimes (A_2B_2) = (A_1 \otimes A_2)(B_1 \otimes B_2), \tag{3}$$

where A_1, A_2, B_1, B_2 are matrices of suitable dimensions [16, p. 408], that

$$\mu_k(Y) = \underbrace{(A \otimes A \otimes \cdots \otimes A)}_{k \text{ times}} \mu_k(X). \tag{4}$$

Expressions for the cumulants of the observed process in a state-space model were developed by Swami and Mendal [29]. The process in this article is a particular case for which simpler expressions hold. We provide an alternative derivation for our particular case in the Appendix. Let $K_k(X)$ denote the vectorized kth central cumulant of any real random vector X with finite mean. For $k = 1, 2, 3$, $K_k(X) = \mu_k(X)$. For $k = 4$, we have

$$\begin{aligned} K_4(X) = {} & \mu_4(X) - \mu_2(X) \otimes \mu_2(X) \\ & - \text{vec}\left\{R_{XX} \otimes R_{XX} + U_{L^2} \cdot \left(R_{XX} \otimes R_{XX}\right)\right\}, \end{aligned} \tag{5}$$

where R_{XX} is the covariance matrix of X and U_{L^2} is an $L^2 \times L^2$ permutation matrix defined in (A8). A permutation matrix is a square binary matrix that permutes the rows (or columns) of a matrix it is applied to. We also have

$$K_4(Y) = \left(A \otimes A \otimes A \otimes A\right) K_4(X). \tag{6}$$

When R_{XX} is a diagonal matrix, as is the case of interest in this article, it follows that $U_{L^2} \cdot \left(R_{XX} \otimes R_{XX}\right)$ is a symmetric matrix. We shall restrict our attention to cumulants of order $k \leq 4$ since estimation of higher-order cumulants is not practical.

Next, we make use of the assumption that the components $\{x_j\}$ of X are independent. This is a standard assumption in network tomography [34] that can be justified partially because source transmissions are typically independent of each other. Under the independence assumption, the central moments and central cumulants of X are conveniently and concisely expressed in terms of the Khatri–Rao product defined by

$$A \odot A := \left[a_1 \otimes a_1, a_2 \otimes a_2, \ldots, a_L \otimes a_L\right]. \tag{7}$$

It is shown in the Appendix that

$$\begin{aligned} \mu_2(Y) &= \left(A \odot A\right) \ \text{col}\left(E\left\{\bar{x}_1^2\right\}, E\left\{\bar{x}_2^2\right\}, \ldots, E\left\{\bar{x}_L^2\right\}\right), \\ \mu_3(Y) &= (A \odot A \odot A) \ \text{col}\left(E\left\{\bar{x}_1^3\right\}, E\left\{\bar{x}_2^3\right\}, \ldots, E\left\{\bar{x}_L^3\right\}\right), \end{aligned} \tag{8}$$

and

$$K_4(Y) = \left(A \odot A \odot A \odot A\right) \begin{pmatrix} \kappa_4\left(\bar{x}_1, \bar{x}_1, \bar{x}_1, \bar{x}_1\right) \\ \kappa_4\left(\bar{x}_2, \bar{x}_2, \bar{x}_2, \bar{x}_2\right) \\ \vdots \\ \kappa_4\left(\bar{x}_L, \bar{x}_L, \bar{x}_L, \bar{x}_L\right) \end{pmatrix}, \tag{9}$$

where

$$\kappa_4(\bar{x}_i, \bar{x}_j, \bar{x}_k, \bar{x}_l) = \begin{cases} E\{\bar{x}_i^4\} - 3(E\{\bar{x}_i^2\})^2, & i = j = k = l, \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

When $\{x_j\}$ are independent Poisson random variables, all cumulants of x_j equal its rate $E\{x_j\} = \lambda_j$, and we have from $m(Y) := E\{Y\} = AE\{X\}$ and (8)–(10) that

$$\begin{pmatrix} A \\ A \odot A \\ A \odot A \odot A \\ A \odot A \odot A \odot A \end{pmatrix} \lambda = \begin{pmatrix} m(Y) \\ \mu_2(Y) \\ \mu_3(Y) \\ K_4(Y) \end{pmatrix}. \tag{11}$$

This is the key cumulant matching equation for $r \leq 4$. In (11), let $\mathcal{A}_r$ denote the matrix of stacked Khatri–Rao products of A, and let $\eta_r(Y)$ denote the vector of stacked cumulants. Then the rate vector λ satisfies

$$\mathcal{A}_r \lambda = \eta_r(Y). \tag{12}$$

To estimate λ, $\eta_r(Y)$ is replaced by a vector of empirical estimates $\hat{\eta}_r(Y)$ and λ is estimated from

$$\boxed{\mathcal{A}_r \lambda = \hat{\eta}_r(Y).} \tag{13}$$

In this article, we consider least squares estimation and the minimum I-divergence iterative procedure (1). The vector $\hat{\eta}_r(Y)$ comprises minimum variance unbiased cumulant estimates given by the K-statistics which we discuss in the next section [14].

Vardi's second-order moment matching equation can be summarized as $\hat{\eta}_2(Y) = \mathcal{A}_2 \lambda$. With only two moments, the column rank of the matrix $\mathcal{A}_2$ may be too low to provide an accurate solution. Note that if the set of equation (13) is consistent, the theoretical cumulants are known, and r is sufficiently large so that $\mathcal{A}_r$ has full column rank, then λ can be estimated error free as the unique solution of (13). We demonstrate in Section 5 for the NSFnet [2] that $r = 3$ is sufficient to achieve full column rank, and that the theoretical performance is approachable when the cumulants are estimated from a sufficiently large number of realizations of Y.

3.2 | Error analysis

In this section, we assess the MSE in the least squares estimation of λ satisfying (12) from the cumulant matching equation (13). We assume for this analysis that r is sufficiently large so that the augmented matrix $\mathcal{A}_r$ has full column rank.

K-statistics of scalar processes were developed in [14, p. 281]. For vector processes and $k = 1, 2, 3$, let $\hat{\mu}_k(Y)$ and $\hat{\mu}_k(X)$ denote the K-statistics of $\mu_k(Y)$ and $\mu_k(X)$, respectively. Similarly, let $\hat{K}_4(Y)$ and $\hat{K}_4(X)$ denote the K-statistics of $K_4(Y)$ and $K_4(X)$, respectively. Let $\{X_n, n = 1, 2, \dots, N\}$ denote a sequence of independent identically distributed source–destination traffic flows defined similarly to X. Each X_n comprises L independent Poisson random variables with rate $E\{X_n\} = \lambda$. Let $Y_n = AX_n$. Define

$$\tilde{Y}_n := Y_n - \hat{m}(Y), \tag{14}$$

where

$$\hat{m}(Y) := \frac{1}{N}\sum_{n=1}^{N} Y_n, \tag{15}$$

and the empirical cumulant estimate for $k > 1$ is

$$\tilde{\mu}_k(Y) := \frac{1}{N}\sum_{n=1}^{N} \underbrace{\tilde{Y}_n \otimes \tilde{Y}_n \otimes \cdots \otimes \tilde{Y}_n}_{k \text{ times}}. \tag{16}$$

Similar definitions follow for the X process when Y in (14)–(16) is replaced by X. The K-statistic for $k = 1$ is given by

$$\hat{m}(Y) = \frac{1}{N}\sum_{n=1}^{N} AX_n = A\hat{m}(X), \tag{17}$$

for $k = 2$,

$$\begin{aligned}\hat{\mu}_2(Y) &= \frac{N}{N-1}\tilde{\mu}_2(Y) \\ &= \left(A \otimes A\right)\hat{\mu}_2(X),\end{aligned} \tag{18}$$

for $k = 3$,

$$\begin{aligned}\hat{\mu}_3(Y) &= \frac{N^2}{(N-1)(N-2)}\tilde{\mu}_3(Y) \\ &= \left(A \otimes A \otimes A\right)\hat{\mu}_3(X),\end{aligned} \tag{19}$$

and for $k = 4$,

$$\begin{aligned}\hat{K}_4(Y) &:= \frac{N^2}{(N-1)(N-2)(N-3)}\left[(N+1)\tilde{\mu}_4(Y) - 3(N-1)\tilde{\mu}_2(Y) \otimes \tilde{\mu}_2(Y)\right] \\ &= \left(A \otimes A \otimes A \otimes A\right)\hat{K}_4(X),\end{aligned} \tag{20}$$

where we have used (3). Similar relations are expected to hold for higher-order cumulants. For $r \leq 3$ and sufficiently large N, the K-statistic $\hat{\mu}_r(Y)$ and the empirical cumulant $\tilde{\mu}_r(Y)$ are essentially the same.

Let $\hat{\eta}_r(X)$ and $\hat{\eta}_r(Y)$ denote vectors of stacked K-statistics of order smaller than or equal to r corresponding to X and Y, respectively. The least squares estimate of λ in (13) is given by

$$\hat{\lambda}_r = \left(\mathcal{A}_r^\top \mathcal{A}_r\right)^{-1} \mathcal{A}_r^\top\ \hat{\eta}_r(Y), \tag{21}$$

where $\cdot^T$ denotes matrix transpose. Let

$$\epsilon_r(Y) = \hat{\eta}_r(Y) - \eta_r(Y) \tag{22}$$

denote the error vector in estimating $\eta_r(Y)$. From (12),

$$\hat{\eta}_r(Y) = \mathcal{A}_r\lambda + \epsilon_r(Y). \tag{23}$$

Substituting (23) into (21) yields

$$\begin{aligned}\hat{\lambda}_r &= \left(\mathcal{A}_r^\top \mathcal{A}_r\right)^{-1} \mathcal{A}_r^\top \left[\mathcal{A}_r\lambda + \epsilon_r(Y)\right] \\ &= \lambda + \left(\mathcal{A}_r^\top \mathcal{A}_r\right)^{-1} \mathcal{A}_r^\top \epsilon_r(Y).\end{aligned} \tag{24}$$

Let $H_r = \left(\mathcal{A}_r^\top \mathcal{A}_r\right)^{-1} \mathcal{A}_r^\top$. The rate estimation error is given by

$$\hat{\lambda}_r - \lambda = H_r\epsilon_r(Y), \tag{25}$$

and the MSE is given by

$$\begin{aligned}\overline{\epsilon_r^2}(\lambda) &= E\left\{||\hat{\lambda}_r - \lambda||^2\right\}\\ &= E\left\{||H_r\epsilon_r(Y)||^2\right\}\\ &\leq \rho_{\max}\left(H_r^\top H_r\right)E\{||\epsilon_r(Y)||^2\},\end{aligned} \tag{26}$$

where the inequality follows from the Rayleigh quotient theorem[1] [16, Theorem 8.1.4], and $\rho_{\max}(H_r^\top H_r)$ denotes the maximal eigenvalue of the Hermitian matrix $H_r^\top H_r$. From the theory of singular value decomposition, we recall that for any matrix H_r,

$$\rho_{\max}\left(H_r H_r^\top\right) = \rho_{\max}\left(H_r^\top H_r\right). \tag{27}$$

Applying this to (26) gives

$$\begin{aligned}\overline{\epsilon_r^2}(\lambda) &\leq \rho_{\max}\left(\left(\left(\mathcal{A}_r^\top\mathcal{A}_r\right)^{-1}\mathcal{A}_r^\top\right)\left(\left(\mathcal{A}_r^\top\mathcal{A}_r\right)^{-1}\mathcal{A}_r^\top\right)^\top\right)E\left\{||\epsilon_r(Y)||^2\right\}\\ &= \rho_{\max}\left(\left(\mathcal{A}_r^\top\mathcal{A}_r\right)^{-1}\right)E\left\{||\epsilon_r(Y)||^2\right\}\\ &= \frac{1}{\rho_{\min}\left(\mathcal{A}_r^\top\mathcal{A}_r\right)}E\left\{||\epsilon_r(Y)||^2\right\}.\end{aligned} \tag{28}$$

Let $\mathcal{B}_r$ denote a block diagonal matrix with the kth diagonal block being $\mathcal{A}_k$. There are r diagonal blocks in total. Let

$$\epsilon_r(X) = \hat{\eta}_r(X) - \eta_r(X). \tag{29}$$

From (22), (17)–(20), and (11),

$$\epsilon_r(Y) = \mathcal{B}_r\epsilon_r(X). \tag{30}$$

It follows from the Rayleigh quotient [16, Theorem 8.1.4] that

$$||\epsilon_r(Y)||^2 \leq ||\mathcal{B}_r||^2 \cdot ||\epsilon_r(X)||^2, \tag{31}$$

where $||\mathcal{B}_r||$ equals the maximal singular value of $\mathcal{B}_r$, and

$$||\mathcal{B}_r||^2 = \rho_{\max}\left(\mathcal{B}_r^\top\mathcal{B}_r\right). \tag{32}$$

The kth diagonal block of $\mathcal{B}_r^\top\mathcal{B}_r$ is given by

$$\begin{aligned}&\left(A\otimes A\otimes\cdots\otimes A\right)^\top\left(A\otimes A\otimes\cdots\otimes A\right)\\ &\quad= A^\top A\otimes A^\top A\otimes\cdots\otimes A^\top A\\ &\quad=: \left(A^\top A\right)^{\otimes k}.\end{aligned} \tag{33}$$

Since the eigenvalues of the Kronecker product of two matrices equal the Kronecker product of the vectors of eigenvalues of each matrix [16, p. 412], we conclude that the eigenvalues of $\left(A^\top A\right)^{\otimes p}$ for any positive integer p are all possible products of length p of the eigenvalues of $A^\top A$. In particular, since $A^\top A$ is positive semi-definite,

$$\rho_{\max}\left(\left(A^\top A\right)^{\otimes p}\right) = \left[\rho_{\max}\left(A^\top A\right)\right]^p. \tag{34}$$

Furthermore, since $\mathcal{B}_r^\top\mathcal{B}_r$ is block diagonal with blocks of increasing size, its eigenvalues are given by the union of the sets of eigenvalues associated with each block. Thus,

$$\rho_{\max}\left(\mathcal{B}_r^\top\mathcal{B}_r\right) = \max_{p:p\geq 1}\left[\rho_{\max}\left(A^\top A\right)\right]^p. \tag{35}$$

When $\rho_{\max}\left(A^\top A\right) > 1$, the maximum in (35) is achieved for $p = r$. Otherwise, when $\rho_{\max}\left(A^\top A\right) < 1$, the maximum in (35) is achieved for $p = 1$. Hence, from (31),

$$E\left\{||\epsilon_r(Y)||^2\right\} \leq \max\left\{\rho_{\max}\left(A^\top A\right), \left[\rho_{\max}\left(A^\top A\right)\right]^r\right\}\cdot E\left\{||\epsilon_r(X)||^2\right\}. \tag{36}$$

From (28) and (36), we obtain

$$\overline{\epsilon_r^2}(\lambda) \leq \frac{\max\left\{\rho_{\max}\left(A^\top A\right), \left[\rho_{\max}\left(A^\top A\right)\right]^r\right\}}{\rho_{\min}\left(\mathcal{A}_r^\top\mathcal{A}_r\right)}E\left\{||\epsilon_r(X)||^2\right\}. \tag{37}$$

[1] The Rayleigh quotient of a Hermitian matrix H is defined as $u^\top Hu/u^\top u$ where u is a vector of suitable dimension.

Evaluating the MSE $E\{||\epsilon_r(X)||^2\}$ is quite involved. For $r = 2, 3$, and sufficiently large N, the cross terms in $\hat{\mu}_r(X)$ will be approximately zero, and hence

$$E\{||\epsilon_r(X)||^2\} \approx \sum_{l=1}^{L} E\{||\epsilon_r(x_l)||^2\}, \tag{38}$$

where $E\{||\epsilon_r(x_l)||^2\}$ is the MSE associated with the K-statistics of the lth component of X. It is also the variance of $\hat{\mu}_r(x_l)$. Since the $\{X_n\}$ vectors are assumed independent, we can infer $E\{||\epsilon_r(x_l)||^2\}$ from the variance of the K-statistics of an IID sequence of scalar random variables given by [14, p. 291],

$$\begin{aligned} \text{var}\ (\hat{\mu}_2(x_l)) &= \frac{1}{N}K_4(x_l) + \frac{1}{N-1}2\mu_2^2(x_l), \\ \text{var}\ (\hat{\mu}_3(x_l)) &= \frac{1}{N}K_6(x_l) + \frac{9}{N-1}K_4(x_l)\mu_2(x_l) \\ &\quad + \frac{9}{N-1}\mu_3^2(x_l) + \frac{6N}{(N-1)(N-2)}\mu_2^3(x_l). \end{aligned} \tag{39}$$

When the IID random variables are Poisson with $E\{X_l\} = \lambda_l$, all cumulants equal λ_l, and hence,

$$\begin{aligned} E\{||\epsilon_2(x_l)||^2\} &= \text{var}\,(\hat{\mu}_2(x_l)) = \frac{1}{N}\lambda_l + \frac{1}{N-1}2\lambda_l^2 \\ &\approx \frac{1}{N}\left(\lambda_l + 2\lambda_l^2\right), \end{aligned} \tag{40}$$

$$\begin{aligned} E\{||\epsilon_3(x_l)||^2\} &= \text{var}\,(\hat{\mu}_3(x_l)) = \frac{1}{N}\lambda_l + \frac{9}{N-1}2\lambda_l^2 \\ &\quad + \frac{6N}{(N-1)(N-2)}\lambda_l^3 \approx \frac{1}{N}\left(\lambda_l + 18\lambda_l^2 + 6\lambda_l^3\right). \end{aligned} \tag{41}$$

The upper bound on $\overline{\epsilon_r^2}(\lambda)$ for $r = 2, 3$, follows from (37), (38), (40), and (41).

The bound (37) is loose when the condition number of the matrix $\mathcal{A}_r^*\mathcal{A}_r$ is large. It does have qualitative value as it shows that estimation of moments becomes increasingly harder as the order increases. Since usually $\lambda_l > 1$ in network tomography, the error in estimating $\mu_3(x_l)$ is much larger than that in estimating $\mu_2(x_l)$.

3.3 | Implementation

In this section, we address several aspects related to the implementation of the cumulant matching approach. The Khatri–Rao product can be expressed in terms of the rows of A instead of its columns as in (7). Let α_i denote the ith row of A, $i = 1, \ldots, M$. Then $A = \text{stack}\{\alpha_i : i = 1, \ldots, M\}$ where stack refers to the stacking of row vectors of equal length to form a matrix. We have

$$\begin{aligned} A \odot A &= \text{stack}\left\{\alpha_i \circ \alpha_j;\ i,j = 1, \ldots, M\right\}, \\ A \odot A \odot A &= \text{stack}\left\{\alpha_i \circ \alpha_j \circ \alpha_k;\ i,j,k = 1, \ldots, M\right\}, \end{aligned} \tag{42}$$

and so forth, where $\circ$ denotes the Schur–Hadamard product and lexicographic ordering of the indices (i,j,k) is assumed. Using this formulation, it is easy to see that $\mathcal{A}_r$ contains duplicate rows. For example, α_i in A and $\alpha_i \circ \alpha_i$ in $A \odot A$ are duplicates since the elements of A belong to $\{0, 1\}$. Additionally, $\mathcal{A}_r$ may contain null rows as is easy to see from (7). Thus, equations in (11) corresponding to duplicate and null rows in $\mathcal{A}_r$ can be removed. We shall always opt to remove duplicate rows that correspond to higher-order cumulants rather than to lower-order cumulants since higher-order cumulants are harder to estimate. To simplify notation we shall henceforth assume that $\mathcal{A}_r$ and the right-hand side (RHS) vector $\eta_r(Y)$ of (11) are given in their reduced form. Similarly, we shall consider (12) and (13) as given in their reduced form.

It is interesting to compare the number of rows in the original and reduced $\mathcal{A}_r$. The number of rows in the original $\mathcal{A}_r$ equals the total number of empirical cumulants up to the rth order, which is given by

$$\overline{n}_r(M) = M + M^2 + M^3 + \cdots + M^r = \frac{M(M^r - 1)}{M - 1}. \tag{43}$$

The maximum number of distinct rows in the reduced $\mathcal{A}_r$ is counted as the sum of the maximum number of distinct rows contributed individually by A, $A \circ A$, $A \circ A \circ A$, and so forth, with the last contribution from the r-fold Khatri–Rao product of A with itself. The contribution from the ith term is given by $\binom{M}{i}$, $i = 1, \ldots, r$, which represents the number of unordered

TABLE 1 Number of unreduced ($\overline{n}_r(M)$) and reduced ($n_r(M)$) cumulant matching equations, and row rank of $\mathcal{A}_r$, for the 4×15 routing matrix of all nonzero binary 4-tuples

r	1	2	3	4
$\overline{n}_r(M)$	4	20	64	320
$n_r(M)$	4	10	14	15
Row rank of $\mathcal{A}_r$	4	10	14	15

combinations of i rows chosen without replacement from the given M rows. Thus, the maximum number of distinct rows in the reduced $\mathcal{A}_r$ is given by

$$n_r(M) = \binom{M}{1} + \binom{M}{2} + \binom{M}{3} + \cdots + \binom{M}{r}. \tag{44}$$

For example, consider a network with a 4×15 routing matrix A whose columns comprise all lexicographically ordered nonzero binary 4-tuples. Here, $M = 4$, and $\overline{n}_r(M)$ and $n_r(M)$ are shown in Table 1. Note that for this example, the number of reduced equations coincides with the row rank of $\mathcal{A}_r$ for each r. Furthermore, full column rank is achieved only when $r = 4$.

The set of equations (13) may contain infeasible equations that result when an empirical central cumulant has a negative value. While all traffic data are non-negative, the empirical cumulants are constructed from centralized data and hence may be negative. This situation leads to infeasible equations whereas the left-hand side of (13) is non-negative while the RHS of (13) is negative. Following Vardi [34], infeasible equations are removed. The number of equations, M_b, in (1) is given by $n_r(M)$ minus the number of removed infeasible equations.

We study two estimates of λ that approximate the cumulant matching equation (13). The first estimate follows from the iteration (1) which is initialized by a constant vector. The second estimate follows from least squares. Specifically, we use the unique Tikhonov regularized least squares solution for the inconsistent set of equations (13), when $\mathcal{A}_r$ is not necessarily full column rank. This estimator is given by [13, p. 51]

$$\hat{\lambda}_r = \left(\mathcal{A}_r^\top \mathcal{A}_r + \gamma I\right)^{-1} \mathcal{A}_r^\top \; \hat{\eta}_r(Y) \tag{45}$$

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

To mitigate the effects of the error introduced by the empirical cumulant estimates, while allowing the estimator of λ to benefit from the higher-order cumulants, the relative weights of the third and fourth order cumulants compared to the first and second order moments, can be reduced. This can be done by multiplying all equations in (45) with rows originating from $\mathcal{A}_3$ by some $0 < \epsilon_3 < 1$, and all equations with rows originating from $\mathcal{A}_4$ by some $0 < \epsilon_4 < 1$. This regularization approach was advocated by Vardi [34] in the context of his second-order moment matching approach. Note that the *constants* ϵ_3 and ϵ_4 are defined independently of the *functions* $\epsilon_r(\cdot)$, $r = 1, 2, 3, 4$ in Section 3.2. Following this approach, let the reduced and ϵ-weighted matrix $\mathcal{A}_r$ be denoted by $\mathcal{A}_{r,\epsilon}$, and let the reduced ϵ-weighted vector $\hat{\eta}_r(Y)$ be denoted by $\hat{\eta}_{r,\epsilon}(Y)$. Then, from (45), the rate vector λ is estimated from

$$\hat{\lambda}_{r,\epsilon} = \left(\mathcal{A}_{r,\epsilon}^\top \mathcal{A}_{r,\epsilon} + \gamma I\right)^{-1} \mathcal{A}_{r,\epsilon}^\top \; \hat{\eta}_{r,\epsilon}(Y). \tag{46}$$

Note that the estimator (46) is not guaranteed to be non-negative. A non-negative estimate of λ can be obtained by using non-negative least squares optimization [17]. In our numeric examples, we have arbitrarily substituted negative estimates with the value of 0.005. This approach resulted in substantially lower MSE compared to using the constrained optimization algorithm of [17, p. 161]. The performance of the algorithm remained essentially the same when other small values (e.g., 0.1 or 0.2) were used since occurrences of negative estimates are infrequent at our working point. See Table 5.

3.4 | Computational complexity

The computational effort in the cumulant matching approach consists of the effort to construct and solve the set of equations (13). The number of equations in this set is $M_b \le n_r(M)$, which is of order $O(M^r)$. Construction of the left-hand side of (13) requires at most $(M^2 + M^3 + \cdots + M^r)L$ operations. Construction of the RHS of (13) requires at most $(M^2 + M^3 + \cdots + M^r)N$ operations where N is the number vectors used to estimate each cumulant. Therefore, the complexity of constructing (13) can be written as $O(M^r(L + N))$.

In the iterative method, the complexity of evaluating (1) in each iteration is $O(M_b L) = O(M^r L)$. Incorporating the number of iterations R and the computational effort for constructing (13) as discussed above, the overall complexity of the iterative method can be written as $O(M^r LR + M^r(L + N)) = O(M^r(L + N))$, since R is a constant. We have used $R = 300$ in our experiments (see Section 5).

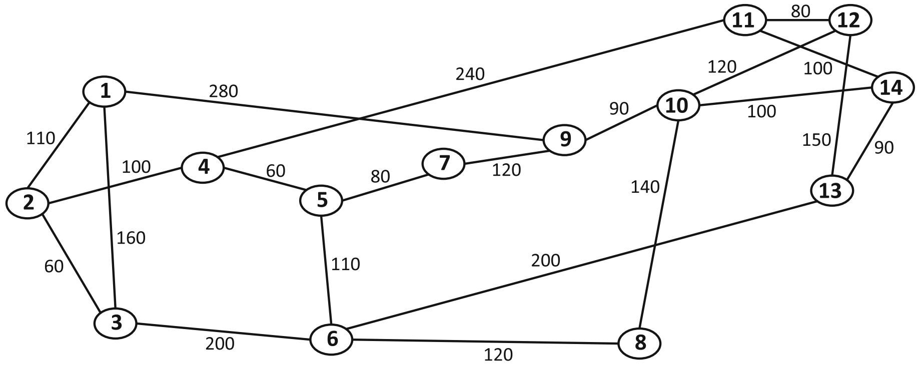

FIGURE 1 NSFnet topology with link weights as in [2, fig. 4]

The least squares estimator requires evaluation of (46). On the RHS of (46), we note that the matrix product $\left(\mathcal{A}_r^*\mathcal{A}_r + \gamma I\right)^{-1}\mathcal{A}_r^*$ can be precomputed. Taking this into account, the evaluation of (46) contributes a complexity of $O(M^r L)$. Therefore, the overall computational complexity of the least squares approach can be written as $O\left(M^r L + M^r(L+N)\right) = O\left(M^r(L+N)\right)$, which is the same as that of the iterative method. Hence, for both the iterative and least squares methods, the additional computational effort in going from $r = 2$ to $r = 3$ is roughly a factor of M. For fixed M and L, the complexity for both schemes is linear in N.

4 | RANDOM ROUTING

In a typical network, there are multiple paths connecting every source node with every destination node. When the network operates under a deterministic routing regime, a single fixed path is used for every source–destination pair. When the network operates under a random routing regime, a path from the source to destination nodes is selected according to some probability law. Vardi attributed Markovian routing for traffic on each source–destination pair. The accessible nodes and links for the given pair are represented by the states and transition probabilities of the Markov chain, respectively. The transition probabilities of the Markov chain determine the probability of each path with the same source–destination address.

Tebaldi and West argued that random routing can be viewed as deterministic routing in a super-network in which all possible paths for each source–destination address are listed [30]. This approach results in an expanded zero-one routing matrix with each column in the original routing matrix replaced by multiple columns representing the feasible paths for the source–destination pair. The Poisson traffic flow on a source–destination pair is thinned into multiple Poisson traffic flows with the multinomial thinning probabilities being the probabilities of the paths with the same source–destination address. Thus, the super-network and the original network operate under similar statistical (Poisson) models. The thinned Poisson rates can now be estimated as was originally done, for example, using cumulant matching, and each source–destination rate estimate can be obtained from the sum of the thinned rate estimates in that source–destination pair.

5 | NUMERICAL EXAMPLE

In this section, we demonstrate the performance of our approach and the gain realized by using higher-order empirical cumulants[2]. We study the NSFnet [2] whose topology is shown in Figure 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 subnetwork of interest of a larger network. The link weights in Figure 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 [36]. When $k = 1$, the number of source–destination paths equals the number of source–destination pairs and the routing matrix A is a 21×91 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

[2]Our experiments were run on ARGO, a research computing cluster provided by the Office of Research Computing at George Mason University.

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 cumulant matching.

The network with $k = 2$ and a 21×182 routing matrix A could also be seen as a super-network in the Tebaldi–West sense [30] 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 21×182 routing matrix A.

The arrival rates $\{\lambda_j\}$ in our experiment were generated randomly from the interval $[0, 4]$. The state of the random generator was reset in each run. For example, for the tth run we used the MATLAB command $s = \mathsf{rng}(t)$. In each of $T = 500$ simulation runs, a rate vector λ was generated, and 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 to generate the empirical cumulants using (15) when $r = 1$ and (16) when $r = 2, 3$. We experimented with N in the range of 10 000 to 500 000. We remark that for a 10 Gbps transmission link in the core of the Internet and packets of average size 1 KByte, the number of packets observed in one second would be approximately 1.25 million. Thus, $N = 500\,000$ samples can be collected in a reasonable amount of time for traffic rate network tomography analysis. The MSE in estimating the cumulants for the NSFnet with $M = 21$ are given in Table 2. Clearly, the MSE decreases monotonically with N for $r = 1, 2, 3$.

The empirical cumulants were used to estimate the rate vector in the current run. For the cumulants regularization we have used $\epsilon_2 = 1$ and $\epsilon_3 = 0.01$. Let $\lambda_t(i)$ and $\hat{\lambda}_t(i)$ denote, respectively, the ith component of λ and its estimate at the tth run where $i = 1, \ldots, L$ and $t = 1, \ldots, T$. For each estimate we evaluated the *normalized MSE* defined by

$$\xi_i^2 = \frac{\frac{1}{T}\sum_{t=1}^{T}\left(\lambda_t(i) - \hat{\lambda}_t(i)\right)^2}{\frac{1}{T}\sum_{t=1}^{T}\left(\lambda_t(i)\right)^2} \tag{47}$$

and the *averaged normalized MSE* defined by

$$\overline{\xi^2} = \frac{1}{L}\sum_{i=1}^{L}\xi_i^2. \tag{48}$$

The MSE in estimating λ_i is approximately $\xi_i^2 \cdot E\{\lambda^2(i)\}$ when T is sufficiently large.

Two rate estimators were used, the iterative estimator (1) and the least squares estimator (46). The iteration was initialized uniformly with all rates set to 0.1. It was terminated after 300 iterations. The least squares regularization factor was set to

TABLE 2 MSE in cumulant estimation for the NSFnet with $M = 21$ links

N	$r = 1$	$r = 2$	$r = 3$
10 000	0.0048	0.2578	14.9097
20 000	0.0025	0.1285	7.4569
50 000	0.0010	0.0515	2.9978
100 000	0.0005	0.0257	1.4989
200 000	0.0002	0.0131	0.7429
500 000	0.0001	0.0052	0.2993

TABLE 3 Averaged normalized MSE $\overline{\xi^2}$ in $L = 182$ source–destination path rate estimation of the NSFnet using iteration (1)

N	$r = 1$	$r = 2$	$r = 3$
10 000	0.2292	0.0993	0.1008
20 000	0.2292	0.0701	0.0694
50 000	0.2292	0.0502	0.0467
100 000	0.2292	0.0434	0.0381
200 000	0.2292	0.0399	0.0335
500 000	0.2292	0.0372	0.0298
Theoretical	0.2292	0.0353	0.0015
Rank ($\mathcal{A}_r$)	21	162	182

TABLE 4 Averaged normalized MSE $\overline{\xi^2}$ in $L = 182$ source–destination path rate estimation of the NSFnet using the least squares estimator (46)

N	*r* = 1	*r* = 2	*r* = 3
10 000	0.3037	0.0951	0.0950
20 000	0.3037	0.0637	0.0625
50 000	0.3037	0.0430	0.0407
100 000	0.3037	0.0358	0.0329
200 000	0.3037	0.0321	0.0288
500 000	0.3037	0.0293	0.0257
Theoretical	0.3037	0.0275	0.0000
Rank ($\mathcal{A}_r$)	21	162	182

TABLE 5 Percent of negative rate estimates in Table 4

N	*r* = 1	*r* = 2	*r* = 3
10 000	0.1890	4.1747	4.2275
20 000	0.1868	2.9813	3.0626
50 000	0.1890	1.9879	1.9901
100 000	0.1890	1.6385	1.5780
200 000	0.1890	1.3747	1.3242
500 000	0.1890	1.1956	1.1165
Theoretical	0.1890	0.9692	0.0022

$\gamma = 0.0005$. The empirical results are provided in Tables 3 and 4, respectively. Table 5 provides the percent of negative rate estimates in the least squares estimate.

6 | CONCLUDING REMARKS

We have developed a framework for higher-order cumulant 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 empirical cumulants of the link measurements to the rate vector λ via a matrix involving Khatri–Rao products of the routing matrix. We studied an iterative minimum I-divergence approach and least squares estimation of the rate vector from the cumulant matching equations. We have established an upper bound on the MSE in least squares estimation of λ. The bound is useful for full column rank matrices $\mathcal{A}_r$ with small condition number. We demonstrated the performance of the cumulant matching approach and compared it with the second-order moment matching system on the NSFnet. We demonstrated that supplementing Vardi's approach with the third-order empirical cumulant reduces its minimum averaged normalized MSE in rate estimation by 19.89% when $N = 500\,000$ samples and iteration (1) was used. For the same N, a more modest improvement of 12.29% was obtained when the least squares estimator was used. These figures follow from Tables 3 and 4, respectively.

In ongoing work, we are extending the higher-order cumulant moment matching technique to the mixed Poisson traffic model studied in [8]. Mixture distributions are overdispersed with variance larger than their mean, whereas for Poisson traffic flows, the mean equals the variance. Another relevant direction of research of interest lies in the development of better estimators for higher-order cumulants to reduce the number of observation samples required, see, for example, [6].

ACKNOWLEDGMENTS

We thank the associate editor and reviewers for their constructive comments. This work was supported by resources provided by the Office of Research Computing at George Mason University (https://orc.gmu.edu).

FUNDING INFORMATION

This work was supported in part by the U.S. National Science Foundation under Grant No. 1717033.

DATA AVAILABILITY STATEMENT

The code and data will be made publicly available if the manuscript is accepted for publication.

ORCID

Hanoch Lev-Ari https://orcid.org/0000-0003-1937-5702
Brian L. Mark https://orcid.org/0000-0002-4030-5592

REFERENCES

[1] A. Anandkumar, A. Hassidim, and J. Kelner, "*Topology discovery of sparse random graphs with few participants*," *ACM SIGMETRICS'11*, ACM, San Jose, CA, 2011.
[2] L. H. Bonani, *Modeling an optical network operating with hybrid-switching paradigms*, J. Microw. Optoelectron. Electromagn. Appl. **15** (2016), 275–292.
[3] R. Cáceres, N. G. Duffield, J. Horowitz, and D. F. Towsley, *Multicast-based inference of network internal loss characteristics*, IEEE Trans. Inf. Theo. **45** (1999), 2462–2480.
[4] J. Cao, D. Davis, S. Vander Wiel, and B. Yu, *Time-varying network tomography: Router link data*, J. Am. Stat. Assoc. **95** (2004), 1063–1075.
[5] R. Castro, M. Coates, G. Liang, R. Nowak, and B. Yu, *Network tomography: Recent developments*, J. Stat. Sci. **19** (2004), 499–517.
[6] L. H. Chan, K. Chen, C. Li, C. W. Wang, and C. Y. Lau, *On higher-order moment and cumulant estimation*, J. Stat. Comput. Simul. **15** (2016), 275–292.
[7] I. Csiszár, *Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems*, Ann. Stat. **19** (1991), 2032–2066.
[8] Y. Ephraim, J. Coblenz, B.L. Mark, and H. Lev-Ari, *Mixed Poisson traffic rate network tomography*, Proc. Conf. Inf. Sci. Syst. (CISS). Baltimore, MD, 2021.
[9] B. Eriksson, G. Dasarathy, P. Barford, and R. Nowak, *Toward the practical use of network tomography for Internet topology discovery*, Proc. IEEE INFOCOM, San Diego, CA, 2010, pp. 1–9.
[10] N. Etemadi Rad, Y. Ephraim, and B. L. Mark, *Delay network tomography using a partially observable bivariate Markov chain*, IEEE/ACM Trans. Net **25** (2017), 126–138.
[11] M. L. Hazelton, *Network tomography for integer-valued traffic*, Ann. Appl. Stat. **9** (2015), 474–506.
[12] 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 **24** (2006), 2182–2195.
[13] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear estimation*, 1st ed., Prentice Hall, Hoboken, NJ, 2000.
[14] M. G. Kendall and A. Stuart, *The advanced theory of statistics*, Hafner Publishing Company, New York, NY, 1969.
[15] R. Kumar and J. Kaur, *Practical beacon placement for link monitoring using network tomography*, IEEE J. Sel. Areas Commu **24** (2006), 2196–2209.
[16] P. Lancaster and M. Tismenetsky, *The theory of matrices*, 2nd ed., Academic Press, Orlando, FL, 1985.
[17] C. L. Lawson and R. J. Hanson, *Solving least squares problems*, Prentice-Hall, Hoboken, NJ, 1974.
[18] G. Liang and B. Yu, *Maximum pseudo likelihood estimation in network tomography*, IEEE Trans. Signal Process. **51** (2003), 2043–2053.
[19] F. Lo Presti, N. G. Duffield, J. Horowitz, and D. Towsley, *Multicast-based inference of network-internal delay distributions*, IEEE/ACM Trans. Net. **10** (2002), 761–775.
[20] L. Ma, T. He, K. K. Leung, D. Towsley, and A. Swami, *Efficient identification of additive link metrics via network tomography*, Proc. IEEE 33rd Int. Conf. Distrib. Comput. Syst., Philadelphia, PA, 2013, pp. 581–590.
[21] M. Mardani and G. B. Giannakis, *Estimating traffic and anomaly maps via network tomography*, IEEE/ACM Trans. Net. **24** (2016), 1533–1547.
[22] J. Ni, H. Xie, S. Tatikonda, and Y. R. Yang, *Efficient and dynamic routing topology inference from end-to-end measurements*, IEEE/ACM Trans. Net. **18** (2010), 123–135.
[23] P. Qin, B. Dai, B. Huang, G. Xu, and K. Wu, *A survey on network tomography with network coding*, IEEE Commun. Surv. Tut. **16** (2014), 1981–1995.
[24] M. G. Rabbat, M. J. Coates, and R. D. Nowak, *Multiple-source internet tomography*, IEEE J. Sel. Areas Commun. **24** (2006), 2221–2234.
[25] W. H. Richardson, *Bayesian-based iterative method of image restoration*, J. Opt. Soc. Am. **62** (1972), 55–59.
[26] L. A. Shepp and Y. Vardi, *Maximum likelihood reconstruction for emission tomography*, IEEE Trans Med Imaging **1** (1982), 113–122.
[27] H. Singhal and G. Michailidis, *Identifiability of flow distributions from link measurements with applications to computer networks*, Inverse Prob. **23** (2007), 1821–1849.
[28] D. L. Snyder, T. J. Schulz, and J. A. O'Sullivan, *Deblurring subject to nonnegativity constraints*, IEEE Trans. Sig. Proc. **40** (1992), 1143–1150.
[29] A. Swami and J. M. Mendel, *Time and lag recursive computation of cumulants from a state-space model*, IEEE Trans. Automat. Contr. **35** (1990), 4–17.
[30] C. Tebaldi and M. West, *Bayesian inference on network traffic using link count data (with discussion)*, J. Am. Stat. Assoc. **93** (1998), 557–576.
[31] Y. Tsang, M. J. Coates, and R. Nowak, *Network delay tomography*, IEEE Trans. Signal Process. **51** (2003), 2125–2136.
[32] R. J. Vanderbei and J. Iannone, *An EM approach to OD matrix estimation* Technical report SOR 94-04, Princeton University, 1994.
[33] Y. Vardi, "*Applications of the EM algorithm to linear inverse problems with positivity constraints*," *Image models (and their speech cousins)*, S. E. Levinson and L. A. Shepp (eds.), Springer, New York, 1996, pp. 183–198.
[34] Y. Vardi, *Network tomography: Estimating source-destination traffic intensities from link data*, J. Amer. Stat. Assoc. **91** (1996), 365–377.
[35] H. Yao, S. Jaggi, and M. Chen, *Passive network tomography for erroneous networks: A network coding approach*, IEEE Trans. Inf. Theo **58** (2012), 5922–5940.
[36] J. Y. Yen, *Finding the K shortest loopless paths in a network*, Manage. Sci. **17** (1971), 712–716.

How to cite this article: H. Lev-Ari, Y. Ephraim, and B. L. Mark, *Traffic rate network tomography with higher-order cumulants*, Networks.. (2022), 1–15. https://doi.org/10.1002/net.22127

APPENDIX A: CUMULANTS OF LINEAR MAPS

Let e_i be an $L \times 1$ vector with a 1 in the ith component and zeros elsewhere. The vector $\overline{X}$ can be written as $\overline{X} = \sum_{i=1}^{L} \overline{x}_i e_i$. Substituting in (2), and applying the distributive property of the Kronecker product, we obtain

$$\mu_4(X) = \sum_{i,j,k,l} E\left\{\overline{x}_i\overline{x}_j\overline{x}_k\overline{x}_l\right\}\left(e_i \otimes e_j \otimes e_k \otimes e_l\right). \tag{A1}$$

Similarly, the fourth-order cumulant of X is given by

$$K_4(X) = \sum_{i,j,k,l} \kappa_4\left(\overline{x}_i, \overline{x}_j, \overline{x}_k, \overline{x}_l\right)\left(e_i \otimes e_j \otimes e_k \otimes e_l\right), \tag{A2}$$

where

$$\begin{aligned}\kappa_4\left(\overline{x}_i, \overline{x}_j, \overline{x}_k, \overline{x}_l\right) = {} & E\left\{\overline{x}_i\overline{x}_j\overline{x}_k\overline{x}_l\right\} - E\left\{\overline{x}_i\overline{x}_j\right\}E\left\{\overline{x}_k\overline{x}_l\right\} \\ & - E\left\{\overline{x}_i\overline{x}_k\right\}E\left\{\overline{x}_j\overline{x}_l\right\} - E\left\{\overline{x}_i\overline{x}_l\right\}E\left\{\overline{x}_j\overline{x}_k\right\}.\end{aligned} \tag{A3}$$

Substituting (A3) into (A2) we obtain

$$\begin{aligned}K_4(X) & \overset{\text{(i)}}{=} \sum_{i,j,k,l} E\left\{\overline{x}_i\overline{x}_j\overline{x}_k\overline{x}_l\right\}\left(e_i \otimes e_j \otimes e_k \otimes e_l\right) \\ & \overset{\text{(ii)}}{-} \sum_{i,j,k,l} E\left\{\overline{x}_i\overline{x}_j\right\}E\left\{\overline{x}_k\overline{x}_l\right\}\left(e_i \otimes e_j \otimes e_k \otimes e_l\right) \\ & \overset{\text{(iii)}}{-} \sum_{i,j,k,l} E\left\{\overline{x}_i\overline{x}_k\right\}E\left\{\overline{x}_j\overline{x}_l\right\}\left(e_i \otimes e_j \otimes e_k \otimes e_l\right) \\ & \overset{\text{(iv)}}{-} \sum_{i,j,k,l} E\left\{\overline{x}_i\overline{x}_l\right\}E\left\{\overline{x}_j\overline{x}_k\right\}\left(e_i \otimes e_j \otimes e_k \otimes e_l\right).\end{aligned} \tag{A4}$$

Row (i) coincides with (A1) and hence it equals $\mu_4(X)$. The expression in row (ii) can be written as

$$\begin{aligned}& \sum_{i,j} E\left\{\overline{x}_i\overline{x}_j\right\}\left(e_i \otimes e_j\right) \otimes \sum_{k,l} E\left\{\overline{x}_k\overline{x}_l\right\}\left(e_k \otimes e_l\right) \\ & \quad = \mu_2(X) \otimes \mu_2(X).\end{aligned} \tag{A5}$$

For row (iii),

$$\begin{aligned}\left(e_i \otimes e_j \otimes e_k \otimes e_l\right) & = \text{vec}\left\{\left(e_k \otimes e_l\right)\left(e_i \otimes e_j\right)^\top\right\} \\ & = \text{vec}\left\{\left(e_k \otimes e_l\right)\left(e_i^\top \otimes e_j^\top\right)\right\} \\ & = \text{vec}\left\{\left(e_k e_i^\top\right) \otimes \left(e_l e_j^\top\right)\right\},\end{aligned} \tag{A6}$$

and hence

$$\begin{aligned}& \sum_{i,j,k,l} E\left\{\overline{x}_i\overline{x}_k\right\}E\left\{\overline{x}_j\overline{x}_l\right\}\left(e_i \otimes e_j \otimes e_k \otimes e_l\right) \\ & \quad = \text{vec} \sum_{i,j,k,l} E\left\{\overline{x}_i\overline{x}_k\right\}E\left\{\overline{x}_j\overline{x}_l\right\}\left(e_k e_i^\top \otimes e_l e_j^\top\right) \\ & \quad = \text{vec}\left\{R_{XX} \otimes R_{XX}\right\}.\end{aligned} \tag{A7}$$

For row (iv), we utilize the $L^2 \times L^2$ permutation matrix

$$U_{L^2} = \sum_{i=1}^{L}\sum_{j=1}^{L}\left(e_i e_j^\top\right) \otimes \left(e_j e_i^\top\right) \tag{A8}$$

to obtain

$$\begin{aligned}\left(e_i \otimes e_j \otimes e_k \otimes e_l\right) & = \left(e_i \otimes e_j\right) \otimes \left(U_{L^2} \cdot \left(e_l \otimes e_k\right)\right) \\ & = \text{vec}\left\{U_{L^2} \cdot \left(e_l \otimes e_k\right) \cdot \left(e_i \otimes e_j\right)^\top\right\} \\ & = \text{vec}\left\{U_{L^2} \cdot \left(e_l \otimes e_k\right) \cdot \left(e_i^\top \otimes e_j^\top\right)\right\} \\ & = \text{vec}\left\{U_{L^2} \cdot \left(e_l e_i^\top\right) \otimes \left(e_k e_j^\top\right)\right\}.\end{aligned} \tag{A9}$$

Hence, row (iv) is given by

$$\operatorname{vec}\left\{U_{L^2}\sum_{i,j,k,l}E\{\overline{x}_i\overline{x}_l\}E\{\overline{x}_j\overline{x}_k\}\left(e_le_i^\top\right)\otimes\left(e_ke_j^\top\right)\right\}$$
$$=\operatorname{vec}\left\{U_{L^2}\left(R_{XX}\otimes R_{XX}\right)\right\}. \tag{A10}$$

Substituting these results into (A4) yields (5).

We next express $K_4(Y)$ in terms of $K_4(X)$ for $Y = AX$. By substituting X with Y in (5), we obtain

$$K_4(Y) = \mu_4(Y) - \mu_2(Y)\otimes\mu_2(Y)$$
$$-\operatorname{vec}\left\{R_{YY}\otimes R_{YY} + U_{M^2}\cdot\left(R_{YY}\otimes R_{YY}\right)\right\}. \tag{A11}$$

The first term in the RHS of (A11) is given by (4). The second term can be expressed using (3) and (4) as

$$\mu_2(Y)\otimes\mu_2(Y) = \left[\left(A\otimes A\right)\mu_2(X)\right]\otimes\left[\left(A\otimes A\right)\mu_2(X)\right]$$
$$=\left(A\otimes A\otimes A\otimes A\right)\left(\mu_2(X)\otimes\mu_2(X)\right). \tag{A12}$$

For the third term in (A11), we use (3) and the well-known identity [16, p. 410]

$$\operatorname{vec}\left\{ARB^\top\right\} = (B\otimes A)\operatorname{vec}\{R\} \tag{A13}$$

to obtain

$$\operatorname{vec}\ \{R_{YY}\otimes R_{YY}\} = \operatorname{vec}\left\{\left(AR_{XX}A^\top\right)\otimes\left(AR_{XX}A^\top\right)\right\}$$
$$=\operatorname{vec}\left\{\left(A\otimes A\right)\left(R_{XX}\otimes R_{XX}\right)\left(A\otimes A\right)^\top\right\}$$
$$=\left(A\otimes A\otimes A\otimes A\right)\operatorname{vec}\left\{R_{XX}\otimes R_{XX}\right\}. \tag{A14}$$

For the last term in (A11), we use (3) and the relation

$$U_{M^2}\left(A\otimes A\right) = \left(A\otimes A\right)U_{L^2} \tag{A15}$$

to obtain

$$U_{M^2}\left(R_{YY}\otimes R_{YY}\right) = U_{M^2}\left(A\otimes A\right)\left(R_{XX}\otimes R_{XX}\right)\left(A\otimes A\right)^\top$$
$$=\left(A\otimes A\right)U_{L^2}\left(R_{XX}\otimes R_{XX}\right)\left(A\otimes A\right)^\top. \tag{A16}$$

Hence, from (A13),

$$\operatorname{vec}\ \left\{U_{M^2}\left(R_{YY}\otimes R_{YY}\right)\right\}$$
$$=\left(A\otimes A\otimes A\otimes A\right)\operatorname{vec}\left\{U_{L^2}\left(R_{XX}\otimes R_{XX}\right)\right\}. \tag{A17}$$

Combining these results we obtain (6).

Suppose now that the components of X are independent random variables. Then,

$$E\left\{\overline{x}_{l_1}\cdot\overline{x}_{l_2}\cdots\overline{x}_{l_k}\right\} = \begin{cases} E\left\{\overline{x}_l^k\right\}, & l_1 = l_2 = \cdots = l_k = l,\\ 0, & \text{otherwise},\end{cases} \tag{A18}$$

and

$$\mu_3\left(X\right) = E\left\{\overline{X}_1\otimes\overline{X}_1\otimes\overline{X}_1\right\}$$
$$=\sum_{i,j,k}E\left\{\overline{x}_i\overline{x}_j\overline{x}_k\right\}\left(e_i\otimes e_j\otimes e_k\right)$$
$$=\sum_{i}E\left\{\overline{x}_i^3\right\}\left(e_i\otimes e_i\otimes e_i\right). \tag{A19}$$

The central moment $\mu_3(Y)$ follows from (4). Expressing

$$A = \sum_{i=1}^{L}a_ie_i^\top, \tag{A20}$$

we have

$$A \otimes A \otimes A = \sum_{i,j,k} \left(a_i e_i^\top\right) \otimes \left(a_j e_j^\top\right) \otimes \left(a_k e_k^\top\right)$$
$$= \sum_{i,j,k} \left(a_i \otimes a_j \otimes a_k\right)\left(e_i \otimes e_j \otimes e_k\right)^\top. \tag{A21}$$

Substituting (A19) and (A21) in (4), and using the orthonormality of the $L^3 \times 1$ vectors $\left\{e_i \otimes e_j \otimes e_k\right\}$, we obtain

$$\mu_3(Y) = \sum_{i=1}^{L} E\left\{\overline{x}_i^3\right\}\left(a_i \otimes a_i \otimes a_i\right)$$
$$= (A \odot A \odot A) \operatorname{col}\left(E\left\{\overline{x}_1^3\right\}, E\left\{\overline{x}_2^3\right\}, \dots, E\left\{\overline{x}_L^3\right\}\right). \tag{A22}$$

It can similarly be shown that

$$\mu_2(Y) = (A \odot A) \operatorname{col}\left(E\left\{\overline{x}_1^2\right\}, E\left\{\overline{x}_2^2\right\}, \dots, E\left\{\overline{x}_L^2\right\}\right). \tag{A23}$$

When the components of X are independent random variables

$$\kappa_4\left(\overline{x}_i, \overline{x}_j, \overline{x}_k, \overline{x}_l\right) = \delta_{ij}\delta_{jk}\delta_{kl}\kappa_4\left(\overline{x}_i, \overline{x}_i, \overline{x}_i, \overline{x}_i\right), \tag{A24}$$

and

$$\kappa_4\left(\overline{x}_i, \overline{x}_j, \overline{x}_k, \overline{x}_l\right) = \begin{cases} E\left\{\overline{x}_i^4\right\} - 3\left(E\left\{\overline{x}_i^2\right\}\right)^2, & i = j = k = l, \\ 0, & \text{otherwise.} \end{cases} \tag{A25}$$

Substituting (A25) into (A4) yields

$$K_4(X) = \sum_{i,j,k,l} \kappa_4\left(\overline{x}_i, \overline{x}_i, \overline{x}_i, \overline{x}_i\right)\delta_{ij}\delta_{jk}\delta_{kl}\left(e_i \otimes e_j \otimes e_k \otimes e_l\right)$$
$$= \sum_{p} \kappa_4\left(\overline{x}_p, \overline{x}_p, \overline{x}_p, \overline{x}_p\right)\left(e_p \otimes e_p \otimes e_p \otimes e_p\right). \tag{A26}$$

To find $K_4(Y)$, we use (6) where

$$A \otimes A \otimes A \otimes A = \sum_{i,j,k,l} \left(a_i \otimes a_j \otimes a_k \otimes a_l\right)\left(e_i \otimes e_j \otimes e_k \otimes e_l\right)^\top \tag{A27}$$

follows from (A20) and (A21), and $K_4(X)$ is given in (A26). Using orthonormality of the $L^4 \times 1$ vectors $\left\{e_i \otimes e_j \otimes e_k \otimes e_l\right\}$ as in (A22), yields

$$K_4(Y) = \sum_{p=1}^{L} \kappa_4\left(\overline{x}_p, \overline{x}_p, \overline{x}_p, \overline{x}_p\right)\left(a_p \otimes a_p \otimes a_p \otimes a_p\right)$$
$$= \left(A \odot A \odot A \odot A\right)\begin{pmatrix} \kappa_4\left(\overline{x}_1, \overline{x}_1, \overline{x}_1, \overline{x}_1\right) \\ \kappa_4\left(\overline{x}_2, \overline{x}_2, \overline{x}_2, \overline{x}_2\right) \\ \vdots \\ \kappa_4\left(\overline{x}_L, \overline{x}_L, \overline{x}_L, \overline{x}_L\right) \end{pmatrix}, \tag{A28}$$

which is (9). We conjecture that the same relation holds for all higher-order cumulants.

When the components of X are independent Poisson random variables with $E\{X\} = \lambda$,

$$E\left\{\overline{x}_i^2\right\} = E\left\{\overline{x}_i^3\right\} = \lambda_i,$$
$$E\left\{\overline{x}_i^4\right\} = \lambda_i + 3\lambda_i^2,$$
$$\kappa_4\left(\overline{x}_i, \overline{x}_i, \overline{x}_i, \overline{x}_i\right) = \left(\lambda_i + 3\lambda_i^2\right) - 3\lambda_i^2 = \lambda_i, \tag{A29}$$

and (11) follows.