

Paola Bermolen, Sandrine Vaton, and Ilmari Juva. 2006. Search for optimality in traffic matrix estimation: A rational approach by Cramér-Rao lower bounds. In Proceedings of the 2nd Conference on Next Generation Internet Design and Engineering (NGI 2006), Valencia, Spain, 3-5 April 2006.

© 2006 IEEE

Reprinted with permission.

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of Helsinki University of Technology's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org.

By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

Search for Optimality in Traffic Matrix Estimation: A Rational Approach by Cramér-Rao Lower Bounds

Paola Bermolen¹, Sandrine Vaton², Ilmari Juva³

1: Universidad de la Republica, Uruguay, 2: ENST Bretagne, France, 3: Helsinki University of Technology, Finland
paola@fing.edu.uy sandrine.vaton@enst-bretagne.fr ilmari.juva@tkk.fi

Abstract—In this paper we study the problem of traffic matrix estimation. The problem is ill-posed and thus some additional information has to be brought in to obtain an estimate. One common approach is to use the second moment statistics through a functional mean-variance relationship. We derive analytically the Fisher information matrix under this framework and obtain the Cramér-Rao lower bound (CRLB) for the variance of an estimator of the traffic matrix. Applications for the use of the CRLB are then demonstrated. From the bounds we can directly obtain confidence intervals for maximum likelihood estimates. Another use for the CRLB is the possibility to evaluate the efficiency of an estimator against the lower bound. A third possible application is to utilize the bounds in an approach to find the best placement for direct measurements of OD flows, so that it is optimal with regard to the traffic matrix estimation problem.

I. INTRODUCTION

The traffic matrix gives the volume of traffic between each origin/destination (OD) pair in the network. While the knowledge of the traffic matrix is essential in network management and traffic engineering, it usually is not possible to measure it directly from the network. This would require Netflow or equivalent measurement devices running network-wide. This approach, however, has huge overhead because of the massive measurements, and is thus impractical in current IP networks.

The goal of traffic matrix estimation is to obtain an estimate for the traffic matrix using information which is readily available in the network: the link counts and the routing matrix.

In the network there are n OD pairs and m links. We denote the OD pair traffic volumes at time t by the n -vector \mathbf{x}_t , in which each element corresponds to an element of the traffic matrix, but the vector notation is used for computational reasons. Similarly, the link loads are denoted by m -vector \mathbf{y}_t . The $m \times n$ routing matrix

is denoted $\mathbf{A} = (a_{ij}) \in \mathcal{M}_{m \times n}$ such that

$$a_{ij} = \begin{cases} 1 & \text{if the OD pair } j \text{ is using link } i \\ 0 & \text{otherwise} \end{cases}$$

Typically, the routing matrix is assumed to be known, and we have several successive measurements of the link counts available, denoted by $\{y_1, \dots, y_T\}$. The basic relationship between the link counts and OD counts \mathbf{x}_t can be written as

$$\mathbf{y}_t = \mathbf{A}\mathbf{x}_t. \quad (1)$$

Should we know the OD counts, it would be straightforward to calculate the link counts. However the opposite is not true, because in any realistic size network there are more links than OD pairs ($n > m$), making the problem of solving the traffic matrix from link counts and routing matrix heavily under-constrained, and thus ill-posed. To overcome this ill-posedness, some type of additional information has to be brought in. Typically either a prior obtained by the gravity method or the second moment statistics of the link counts. Reviews of proposed methods can be found e.g. in [1], [2].

More complex estimation techniques naturally often yield more accurate results. In general, we can say that there is a trade-off between the computational complexity and the accuracy of the estimate. However, no matter how elaborate the technique, there is a bound for the accuracy of the estimate. This is due to the stochastic nature of the traffic process, which makes it impossible to obtain estimate accuracy below certain level.

The traffic volume can be considered a random variable. The Fisher information matrix gives the amount of information that the observed traffic volumes carry of the underlying parameter, namely the expected traffic volume. For any unbiased estimate, the Cramér-Rao lower bound (CRLB), which is the inverse of the Fisher information matrix, gives the limit of how small variances it is possible to obtain for an estimator.

In this paper we show how to calculate the Cramér-Rao bounds for the traffic matrix estimation problem. There are many benefits of obtaining an expression for the CRLB. In synthetic data situations we can obtain sample variances of an estimator by Monte Carlo simulations, and compare to the bound in order to evaluate how close to optimality the considered method is. In real data situation the CRLB can be used to get an idea of the uncertainty related to each OD pair estimate, since we can obtain a result for how large the variance is at least. As the Maximum likelihood estimate (MLE) is asymptotically efficient, its variance coincides with the Cramér-Rao lower bounds. Thus, we can obtain confidence intervals for the MLE directly from the bounds.

The rest of the paper is organized as follows. In section II-A we introduce the model and define the concepts of information matrix and Cramér-Rao lower bounds. In section II we derive an expression for the information matrix, which thus enables analytical expression for the CRLB. Section III gives examples of possible applications for the result.

II. CRAMÉR-RAO LOWER BOUND

In this section we develop an analytical expression for the Fisher information matrix, and thus for the Cramér-Rao lower bound for variance. First we introduce the model and the assumptions used, and define some expressions we need later. Then we consider the CRLB for a general multivariate Gaussian case, and then the specific problem of traffic matrix estimation with OD pairs following Gaussian distribution.

A. Preliminaries

In our model we assume that the OD pair traffic follows Gaussian distribution, that OD pairs are independent of each other, and also that successive measurements for each OD pair are independently and identically distributed. The expected value of OD pair counts x_t is denoted by the vector λ and the covariance matrix by Σ .

$$x_t \sim N(\lambda, \Sigma). \quad (2)$$

Furthermore, it is assumed that there is a functional relation between the mean and the variance with parameters Φ and c .

$$\Sigma = \Phi \text{diag}(\lambda^c), \quad (3)$$

In the sequel we use the notation

$$\Sigma' = \text{diag}(\lambda^c).$$

This is a typical assumption in traffic matrix estimation, which enables the use of maximum likelihood approach.

Through the mean-variance relation, the link covariances are used to bring in the extra information needed to yield an estimate for the traffic matrix.

The link counts are denoted by random vector \mathbf{Y} that has probability density function (pdf) $p(\mathbf{y}; \Psi)$, where Ψ is the vector containing the unknown parameters. Due to (1) the link counts have expected value $A\lambda$ and covariance matrix $A\Sigma A^t$. Thus, using (3), we can write

$$\mathbf{y}_t \sim N(A\lambda, \Phi A\Sigma' A^t). \quad (4)$$

The parameters Ψ of the model are the n elements of λ and the parameters relating the mean to the variance, that is Φ and c , of which c is treated as a preset constant as in [3].

$$\Psi = (\lambda_1, \lambda_2, \dots, \lambda_n, \Phi). \quad (5)$$

The likelihood function for Ψ , formed from the observed data \mathbf{y} is

$$L(\Psi) = p(\mathbf{y}; \Psi). \quad (6)$$

The log-likelihood function is denoted by

$$l(\Psi) = \log L(\Psi). \quad (7)$$

The gradient vector of the log-likelihood is given by the score statistic

$$S(\mathbf{y}; \Psi) = \frac{\partial l(\Psi)}{\partial \Psi}. \quad (8)$$

We assume that the pdf $p(\mathbf{y}; \Psi)$, where $\Psi = (\Psi_1, \Psi_2, \dots, \Psi_d)^t$, satisfies the regularity conditions

$$E\left(\frac{\partial l(\mathbf{y}; \Psi)}{\partial \Psi}\right) = 0 \quad \forall \Psi, \quad (9)$$

where the expectation is taken with respect to $p(\mathbf{y}; \Psi)$. These regularity condition are satisfied if it is possible to exchange the differentiation with the expectation. Using this, we can state the following

Proposition 1: Under the regularity conditions of (9), the expected (Fisher) information matrix $\mathcal{I}(\Psi)$ is given by

$$\mathcal{I}(\Psi) = E_{\Psi}(S(\mathbf{Y}; \Psi)S^t(\mathbf{Y}; \Psi))$$

Theorem 1: Under the regularity conditions of (9), the covariance matrix of any unbiased estimator Ψ^* satisfies

$$C_{\Psi^*} - \mathcal{I}^{-1}(\Psi) \geq 0, \quad (10)$$

where “ ≥ 0 ” is interpreted as meaning that the matrix is positive semidefinite, and $\mathcal{I}(\Psi)$ is the Fisher information matrix evaluated at the true value of Ψ .

The above theorem gives the Cramér-Rao lower bound. It states that C_{Ψ^*} , the variance/covariance matrix of any unbiased estimator cannot be lower than the inverse of the Fisher information matrix.

B. Information matrix for the general Gaussian case

We will set the traffic matrix framework aside for a moment and calculate the CRLB for the general Gaussian case. To avoid confusion, we use different symbols for mean and variance in this case, than those introduced for the traffic matrix estimation problem in the previous section. The incomplete data is a multivariate Gaussian with mean $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\Psi})$ and covariance matrix $\mathbf{C} = \mathbf{C}(\boldsymbol{\Psi})$. The probability density function is

$$p(\mathbf{y}; \boldsymbol{\Psi}) = \frac{1}{(2\pi)^{m/2} \det \mathbf{C}(\boldsymbol{\Psi})^{1/2}} \cdot \exp \left\{ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\Psi}))^t \mathbf{C}(\boldsymbol{\Psi})^{-1} (\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\Psi})) \right\}. \quad (11)$$

And it follows that the log-likelihood is

$$l(\mathbf{y}; \boldsymbol{\Psi}) = -\log(2\pi)^{m/2} - \frac{1}{2} \log \det(\mathbf{C}(\boldsymbol{\Psi})) - \frac{1}{2} (\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\Psi}))^t \mathbf{C}(\boldsymbol{\Psi})^{-1} (\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\Psi})). \quad (12)$$

An element of the information matrix can be written as

$$\mathcal{I}(\boldsymbol{\Psi})_{ij} = E \left[\frac{\partial l(\mathbf{y}; \boldsymbol{\Psi})}{\partial \Psi_i} \frac{\partial l(\mathbf{y}; \boldsymbol{\Psi})}{\partial \Psi_j} \right] \quad (13)$$

Proposition 2: The analytical expression for the information matrix is

$$\begin{aligned} \mathcal{I}(\boldsymbol{\Psi})_{ij} &= \frac{\partial \boldsymbol{\mu}(\boldsymbol{\Psi})^t}{\partial \Psi_i} \mathbf{C}^{-1}(\boldsymbol{\Psi}) \frac{\partial \boldsymbol{\mu}(\boldsymbol{\Psi})}{\partial \Psi_j} \\ &+ \frac{1}{2} \text{tr} \left(\mathbf{C}^{-1}(\boldsymbol{\Psi}) \frac{\partial \mathbf{C}(\boldsymbol{\Psi})}{\partial \Psi_i} \mathbf{C}^{-1}(\boldsymbol{\Psi}) \frac{\partial \mathbf{C}(\boldsymbol{\Psi})}{\partial \Psi_j} \right) \end{aligned} \quad (14)$$

The derivation of this expression is given in the appendix.

C. Information matrix for Gaussian origin-destination pairs

We will now return to the traffic matrix estimation problem, and use the result obtained above for the general Gaussian case. We have previously defined the vector of link counts \mathbf{y}_t as multivariate gaussian distribution with mean

$$\boldsymbol{\mu}(\boldsymbol{\Psi}) = \mathbf{A}\boldsymbol{\lambda}, \quad (15)$$

and covariance matrix

$$\mathbf{C}(\boldsymbol{\Psi}) = \Phi \mathbf{A} \boldsymbol{\Sigma}' \mathbf{A}^t, \quad (16)$$

where notation defined in section II-A is used. Thus the probability density function is

$$p(\mathbf{y}_t) = \frac{1}{(2\pi)^{m/2} \det(\Phi \mathbf{A} \boldsymbol{\Sigma}' \mathbf{A}^t)^{1/2}} \cdot \exp \left\{ -\frac{1}{2} (\mathbf{y}_t - \mathbf{A}\boldsymbol{\lambda})^t (\Phi \mathbf{A} \boldsymbol{\Sigma}' \mathbf{A}^t)^{-1} (\mathbf{y}_t - \mathbf{A}\boldsymbol{\lambda}) \right\}. \quad (17)$$

Since consecutive measurement samples of the link counts are considered independent from each other, the pdf can be written in product form

$$p(\mathbf{y}) = \prod_{t=1}^T p(\mathbf{y}_t). \quad (18)$$

It follows directly that the log-likelihood can be written as a sum

$$l(\mathbf{y}) = \sum_{t=1}^T l(\mathbf{y}_t), \quad (19)$$

and as the information matrix is the same for each time period, because of the iid property, we can write

$$\mathcal{I}(\boldsymbol{\Psi}) = T \mathcal{I}_t(\boldsymbol{\Psi}), \quad (20)$$

where

$$\mathcal{I}_t(\boldsymbol{\Psi}) = E_{\boldsymbol{\Psi}} (\mathbf{S}(Y_t; \boldsymbol{\Psi}) \mathbf{S}^t(Y_t; \boldsymbol{\Psi})), \quad (21)$$

and $\mathbf{S}(Y_t; \boldsymbol{\Psi})$ is the score statistic of the incomplete data, defined in equation (8).

An element of the information matrix for the general case was given in (14). Thus, to obtain this for the specific case in question here, we have to calculate the expressions

$$\frac{\partial \boldsymbol{\mu}(\boldsymbol{\Psi})}{\partial \Psi_i} \quad \text{and} \quad \frac{\partial \mathbf{C}(\boldsymbol{\Psi})}{\partial \Psi_i} \quad (22)$$

for the traffic matrix estimation problem.

Inserting the expressions from (15) and (16) into (22), we obtain for $i = 1, \dots, n$ the derivative of the mean as

$$\frac{\partial \boldsymbol{\mu}(\boldsymbol{\Psi})}{\partial \Psi_i} = \frac{\partial \mathbf{A}\boldsymbol{\lambda}}{\partial \lambda_i} = \mathbf{A}^i, \quad (23)$$

where \mathbf{A}^i is the i th column of \mathbf{A} , and for the covariance matrix

$$\frac{\partial \mathbf{C}(\boldsymbol{\Psi})}{\partial \Psi_i} = \Phi \frac{\partial \mathbf{A} \boldsymbol{\Sigma}' \mathbf{A}^t}{\partial \lambda_i} = \Phi \mathbf{A} \frac{\partial \boldsymbol{\Sigma}'}{\partial \lambda_i} \mathbf{A}^t = \Phi c \lambda_i^{c-1} \mathbf{A}^i \mathbf{A}^{i^t}. \quad (24)$$

For the case $i = n + 1$, the differentiation is done with regard to the parameter $\Psi_{n+1} = \Phi$. This yields

$$\frac{\partial \boldsymbol{\mu}(\boldsymbol{\Psi})}{\partial \Psi_i} = \frac{\partial \mathbf{A}\boldsymbol{\lambda}}{\partial \Phi} = 0 \quad (25)$$

and

$$\frac{\partial \mathbf{C}(\boldsymbol{\Psi})}{\partial \Psi_i} = \frac{\partial (\Phi \mathbf{A} \boldsymbol{\Sigma}' \mathbf{A}^t)}{\partial \Phi} = \mathbf{A} \boldsymbol{\Sigma}' \mathbf{A}^t. \quad (26)$$

The information matrix has the following form

$$\mathcal{I}_t(\boldsymbol{\Psi}) = \left(\begin{array}{c|c} I_1 & I_2 \\ \hline I_3 & I_4 \end{array} \right), \quad (27)$$

where I_1 is a $n \times n$ matrix, I_2 is a column vector of length n , I_3 is a row vector of the same length, and I_4 is a scalar. To simplify the notation we introduce the matrix

$$\mathbf{W} = \mathbf{A}^t (\mathbf{A} \boldsymbol{\Sigma}' \mathbf{A}^t)^{-1} \mathbf{A}, \quad (28)$$

which has the elements

$$w_{ij} = \mathbf{A}^{j^t} (\mathbf{A} \boldsymbol{\Sigma} \mathbf{A}^t)^{-1} \mathbf{A}^i. \quad (29)$$

Starting from the expression obtained in (14) and using the derivatives of $\boldsymbol{\mu}(\boldsymbol{\Psi})$ and $\mathbf{C}(\boldsymbol{\Psi})$ derived above, we can now calculate the expressions for the elements of the information matrix. For $i, j = 1, \dots, n$,

$$\begin{aligned} (I_1)_{i,j} &= \frac{\partial \boldsymbol{\mu}(\boldsymbol{\Psi})^t}{\partial \lambda_i} \mathbf{C}(\boldsymbol{\Psi})^{-1} \frac{\partial \boldsymbol{\mu}(\boldsymbol{\Psi})}{\partial \lambda_j} \\ &+ \frac{1}{2} \text{tr} \left(\mathbf{C}(\boldsymbol{\Psi})^{-1} \frac{\partial \mathbf{C}(\boldsymbol{\Psi})}{\partial \lambda_i} \mathbf{C}(\boldsymbol{\Psi})^{-1} \frac{\partial \mathbf{C}(\boldsymbol{\Psi})}{\partial \lambda_j} \right) \\ &= \frac{1}{\Phi} \mathbf{A}^{i^t} (\mathbf{A} \boldsymbol{\Sigma}' \mathbf{A}^t)^{-1} \mathbf{A}^j + \frac{c^2 \lambda_i^{c-1} \lambda_j^{c-1}}{2} \\ &\cdot \text{tr} \left((\mathbf{A} \boldsymbol{\Sigma}' \mathbf{A}^t)^{-1} \mathbf{A}^i \mathbf{A}^{i^t} (\mathbf{A} \boldsymbol{\Sigma}' \mathbf{A}^t)^{-1} \mathbf{A}^j \mathbf{A}^{j^t} \right) \\ &= \frac{1}{\Phi} \mathbf{A}^{i^t} (\mathbf{A} \boldsymbol{\Sigma}' \mathbf{A}^t)^{-1} \mathbf{A}^j + \frac{c^2 \lambda_i^{c-1} \lambda_j^{c-1}}{2} \\ &\cdot \left(\mathbf{A}^{j^t} (\mathbf{A} \boldsymbol{\Sigma}' \mathbf{A}^t)^{-1} \mathbf{A}^i \mathbf{A}^{i^t} (\mathbf{A} \boldsymbol{\Sigma}' \mathbf{A}^t)^{-1} \mathbf{A}^j \right) \\ &= w_{ij} + \frac{c^2 \lambda_i^{c-1} \lambda_j^{c-1}}{2} \phi^2 w_{ij}^2. \end{aligned} \quad (30)$$

For $i = 1, \dots, n$

$$\begin{aligned} (I_2)_{i,(n+1)} &= \frac{1}{2} \text{tr} \left(\mathbf{C}(\boldsymbol{\Psi})^{-1} \frac{\partial \mathbf{C}(\boldsymbol{\Psi})}{\partial \Phi} \mathbf{C}(\boldsymbol{\Psi})^{-1} \frac{\partial \mathbf{C}(\boldsymbol{\Psi})}{\partial \lambda_i} \right) \\ &= \frac{c \lambda_i^{c-1}}{2} \text{tr} \left(\frac{1}{\Phi} (\mathbf{A} \boldsymbol{\Sigma}' \mathbf{A}^t)^{-1} \mathbf{A} \boldsymbol{\Sigma}' \mathbf{A}^t \frac{1}{\Phi} (\mathbf{A} \boldsymbol{\Sigma}' \mathbf{A}^t)^{-1} \Phi \mathbf{A}^i \mathbf{A}^{i^t} \right) \\ &= \frac{c \lambda_i^{c-1}}{2\Phi} \text{tr} \left((\mathbf{A} \boldsymbol{\Sigma}' \mathbf{A}^t)^{-1} \mathbf{A}^i \mathbf{A}^{i^t} \right) \\ &= \frac{c \lambda_i^{c-1}}{2\Phi} \left(\mathbf{A}^{i^t} (\mathbf{A} \boldsymbol{\Sigma}' \mathbf{A}^t)^{-1} \mathbf{A}^i \right) \\ &= \frac{c \lambda_i^{c-1}}{2} w_{ii} \end{aligned} \quad (31)$$

Analogously, for $j = 1, \dots, n$

$$(I_3)_{(n+1),j} = \frac{c \lambda_j^{c-1}}{2} w_{jj} \quad (32)$$

And finally,

$$\begin{aligned} (I_4)_{(n+1),(n+1)} &= \frac{1}{2} \text{tr} \left(\mathbf{C}(\boldsymbol{\Psi})^{-1} \frac{\partial \mathbf{C}(\boldsymbol{\Psi})}{\partial \Phi} \mathbf{C}(\boldsymbol{\Psi})^{-1} \frac{\partial \mathbf{C}(\boldsymbol{\Psi})}{\partial \Phi} \right) \\ &= \frac{1}{2} \text{tr} \left(\frac{1}{\Phi} (\mathbf{A} \boldsymbol{\Sigma}' \mathbf{A}^t)^{-1} \mathbf{A} \boldsymbol{\Sigma}' \mathbf{A}^t \frac{1}{\Phi} (\mathbf{A} \boldsymbol{\Sigma}' \mathbf{A}^t)^{-1} \mathbf{A} \boldsymbol{\Sigma}' \mathbf{A}^t \right) \\ &= \frac{1}{2\Phi^2} \text{tr} (\mathbf{Id}_{m \times m}) = \frac{m}{2\Phi^2}. \end{aligned} \quad (33)$$

We have now obtained an analytical expression for the Fisher information matrix of the traffic matrix estimation problem. The Cramér-Rao lower bound for the variance of an estimator is then just \mathcal{I}^{-1} , where the CRLB for

variances of the parameters are the diagonal elements. In the next section we will demonstrate the benefits of obtaining the expression for the CRLB.

III. APPLICATIONS

A. Evaluation of estimation techniques

Based on synthetic data evaluation studies ([1], [4]) of traffic matrix estimation methods, it would seem that the most effective methods are the ones using the maximum likelihood approach, most notably the method by Cao et al. [3]. The problem with this approach is that even with numerical methods such as the EM algorithm [5] the method does not scale well to realistic size networks. Thus, computationally lighter approaches, such as [6], [7], that trade accuracy for computational lightness have been proposed. However, as the likelihood method scales poorly, it is difficult to make comparisons about the tradeoff between estimation accuracy and computation time in realistic size situations. Indeed, in both [6] and [7], the comparisons between the proposed methods and the full likelihood method is performed only in a small topology.

The asymptotic efficiency of the MLE is a well known results, see e.g. [5]. That is to say, The asymptotic covariance matrix of the MLE is equal to the inverse of the expected information matrix.

$$\sqrt{n}(\hat{\Psi}_n - \Psi) \rightarrow N(0, I_t(\Psi)^{-1}).$$

From simulations with synthetic traffic matrices we can obtain sample variances for the considered methods. Then calculating the CRLB, it is possible to compare them to the bound, and thus to the variance of MLE. This way we can evaluate how much less accurate the methods are than the full MLE, without having to run the full likelihood method.

B. Optimal location for direct measurements

Consider that on one link of the network we could deploy a measurement device able to obtain direct measurements of the OD counts. For instance NetFlow [8] is capable of collecting these types of flow level measurements. We consider here link-wise measurements, but the basic technique is the same even if measurements are available router-wise for each link adjacent to the router where the measurement device is used.

By direct OD flow measurements we can obtain the actual traffic volumes of each OD pair traversing the measured link, instead of just the total link load available by SNMP measurements, leading to more accurate estimates of the traffic matrix. By directly observing some

TABLE I
BEST PLACEMENTS FOR A SINGLE LINK MEASUREMENT

link	18r	18	15r	9r	14	10	...	2
Avg. Var	0.58	0.72	0.72	0.73	0.73	0.74	...	0.99

TABLE II
BEST PLACEMENTS FOR TWO LINK MEASUREMENTS

link1	18r	18r	18r	18r	18r	18r	18r	18r	18r	18	18	18r
link2	18	9r	15	10	14r	19	4r	12	5r	15r	14	3r
Avg. Var	0.36	0.39	0.40	0.40	0.40	0.43	0.44	0.46	0.47	0.47	0.48	0.48

of the OD pairs, not only do we get accurate information about these OD pairs, but also the estimates for other OD pairs are more accurate, since the situation is now less underconstrained, due to the extra measurements.

To incorporate the direct measurements of some OD flows to the traffic matrix estimation framework we propose a model that creates a new linear system. This can be interpreted as a virtual topology, where the link where the direct measurements are made is replaced by several virtual links, such that each OD pair using the link would have its own virtual link. This enables us to incorporate the direct OD pair measurements without changing the basic situation. For example, the network in Figure 1 has three links (AB, BC, BD) and four OD pairs ($x_{AC}, x_{AD}, x_{BC}, x_{BD}$). The routing matrix is

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}.$$

The link BD has two OD flows, x_{AD} and x_{BD} , using it. In the virtual topology the link is replaced by virtual links BD_1 and BD_2 . Now x_{AD} is the only OD flow traversing link BD_1 and x_{BD} is the only OD flow traversing link BD_2 . Thus the direct measurements can

be incorporated into the estimate through the usual inference techniques from equation

$$y' = A'x,$$

where y' and A' are the link loads and the routing matrix of the virtual topology. The routing matrix changes so that the last row of A , that corresponds to link BD , is divided into two rows corresponding to virtual links BD_1 and BD_2 .

$$A' = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Using the equations derived earlier in this paper we can calculate the Cramér-Rao lower bounds for variances of the OD pairs for this virtual topology. If we have maximum likelihood estimates for the OD pairs, we can calculate the variance of the estimates by plugging in the MLE into the CRLB equations. As some of the OD pairs are directly observed, their variances are very small. They are not zero, however, since we are observing the OD pair traffic loads x ,

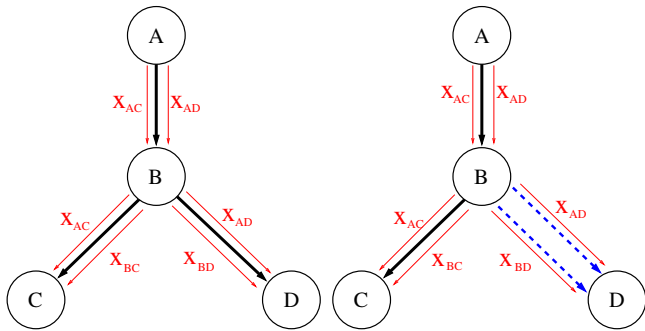


Fig. 1. Left: Example topology. Right: Link BD is replaced by virtual links BD_1 and BD_2

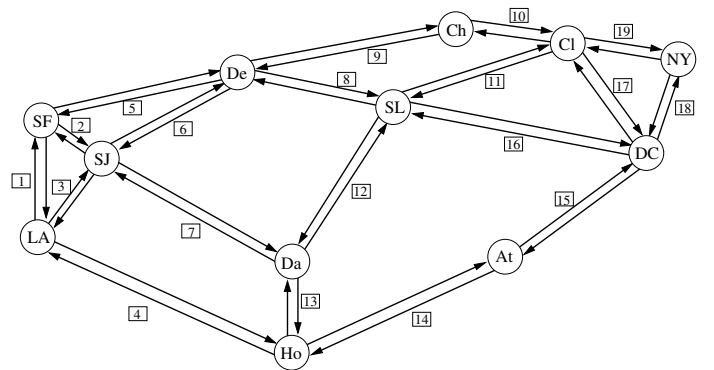


Fig. 2. Fictional US Backbone topology

a stochastic variable, whose expected value λ we are trying to estimate.

Changing the location of the measurement device, creates different virtual topologies that lead to different OD pairs being directly observed.

Also, as this is all analytical calculations, it is rather quick to calculate all two link combinations, to find out how to best place two measurement devices.

Comparing the OD pair variances of the virtual topology to the OD pair variances of the original topology, the accuracy gained from measurements on a given link can be evaluated. Any number of criterions may be used, but for the sake of example we use the ratio of the average of the OD pair variances

$$\frac{\text{tr}(\mathcal{I}(\mathbf{A}')^{-1})/n}{\text{tr}(\mathcal{I}(\mathbf{A})^{-1})/n} \quad (34)$$

in the sequel, as this gives an good indication about how much a measurement is able to reduce the variances.

Repeating the above procedure for each link, we can compare the results each measurement would yield for the average variance, and thus be able to find the optimal location for measurement, that is, the link yielding the lowest average variance.

For example, consider the fictional US backbone topology in Figure 2, where the links have been enumerated such that we refer to the direction with the number next to it by that number, and indicate the opposite direction by lower case r. So for instance the link from LA to SF is 1 and the link in the reverse direction from SF to LA is then 1r.

We consider a traffic distribution generated by the gravity model based on the population of the cities in question. There are a few larger OD pairs, especially the ones between LA and NY, as well as from those two to the middle sized cities. The best location for the measurements is not solely dependent on the location of the link, but needs to capture as many of the bigger flows as possible. It turns out that the best placement by far is link 18r. The next best locations include the other links from the same *NY – LA* route. Table I shows the optimal links and the ratio of average variances divided by the average variances of the original case.

Also the best two link combinations are dominated by the aforementioned link 18r. Best combination is to have both links between *NY* and *DC* to capture the big flows between *NY* and *LA* in both directions. The best placements are shown in Table II

Selecting the combination of two links from the single link calculations would in these example cases yield the optimal placements, but not always as the second best

link might be capturing some of the same large OD pairs as the first link, making it good location for a single measurement point but not very reasonable for a second point if the large OD pair is already measured by the first location. This is the case, for instance, between best and third best links in the gravity case, which combination would not be very effective choice for two measurement locations. Doing the selection sequentially, on the other hand, would remove this problem and yield in most cases optimal solution reducing the running time of the calculations from m^2 to $2m$, where m is the number of links.

IV. CONCLUSION

In this paper we derived an analytical expression for the Fisher information matrix in the traffic matrix estimation framework. The result was used to yield the Cramér-Rao lower bound for the variance of an estimator in the situation where we assume a functional mean-variance relationship for origin-destination flows in the network. We demonstrated why this result is extremely useful in various ways. We can obtain variances, and thus confidence intervals for the maximum likelihood estimate directly from the Cramér-Rao lower bounds. This means that we can identify the OD pairs whose estimates have large uncertainties. If the estimated traffic matrix is used, for instance, in load balancing, it should prove beneficial to know for which OD pairs the estimate might not be accurate. The CRLB can be used also in evaluation of estimation techniques, as we can compare the variance of the evaluated estimator to the lower bound to see how effective it is. A third use for the bounds is demonstrated in section III-B, where we show how to utilize the result in finding the optimal place for direct measurements to reduce the average error of a traffic matrix estimate as much as possible.

APPENDIX

A. Derivation of the general gaussian case information matrix

An element of the information matrix can be written as

$$\mathcal{I}(\Psi)_{ij} = E \left[\frac{\partial l(\mathbf{y}; \Psi)}{\partial \Psi_i} \frac{\partial l(\mathbf{y}; \Psi)}{\partial \Psi_j} \right]. \quad (35)$$

We will use the following identities

$$\frac{\partial \log \det(\mathbf{C}(\Psi))}{\partial \Psi_i} = \text{tr} \left(\mathbf{C}(\Psi)^{-1} \frac{\partial \mathbf{C}(\Psi)}{\partial \Psi_i} \right), \quad (36)$$

and

$$\frac{\partial C(\Psi)^{-1}}{\partial \Psi_i} = -C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_i} C(\Psi)^{-1}. \quad (37)$$

Now we need to calculate the first order derivatives of the log-likelihood.

$$\begin{aligned} \frac{\partial l(\mathbf{y}; \Psi)}{\partial \Psi_i} &= -\frac{1}{2} \frac{\partial \log \det(C(\Psi))}{\partial \Psi_i} \\ &\quad - \frac{1}{2} \frac{\partial}{\partial \Psi_i} [(\mathbf{y} - \mu(\Psi))^t C(\Psi)^{-1} (\mathbf{y} - \mu(\Psi))]. \end{aligned} \quad (38)$$

The first term is

$$-\frac{1}{2} \frac{\partial \log \det(C(\Psi))}{\partial \Psi_i} = -\frac{1}{2} \text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_i} \right).$$

We then consider the second term:

$$\begin{aligned} &\frac{\partial}{\partial \Psi_i} [(\mathbf{y} - \mu(\Psi))^t C(\Psi)^{-1} (\mathbf{y} - \mu(\Psi))] \\ &= \frac{\partial (\mathbf{y} - \mu(\Psi))^t}{\partial \Psi_i} C(\Psi)^{-1} (\mathbf{y} - \mu(\Psi)) \\ &\quad + (\mathbf{y} - \mu(\Psi))^t \frac{\partial C(\Psi)^{-1}}{\partial \Psi_i} (\mathbf{y} - \mu(\Psi)) \\ &\quad + (\mathbf{y} - \mu(\Psi))^t C(\Psi)^{-1} \frac{\partial (\mathbf{y} - \mu(\Psi))}{\partial \Psi_i} \\ &= -\frac{\partial \mu(\Psi)^t}{\partial \Psi_i} C(\Psi)^{-1} (\mathbf{y} - \mu(\Psi)) \\ &\quad + (\mathbf{y} - \mu(\Psi))^t \frac{\partial C(\Psi)^{-1}}{\partial \Psi_i} (\mathbf{y} - \mu(\Psi)) \\ &\quad - (\mathbf{y} - \mu(\Psi))^t C(\Psi)^{-1} \frac{\partial \mu(\Psi)}{\partial \Psi_i}. \end{aligned} \quad (39)$$

Using

$$\begin{aligned} &(\mathbf{y} - \mu(\Psi))^t C(\Psi)^{-1} \frac{\partial \mu(\Psi)}{\partial \Psi_i} \\ &= \left(\frac{\partial \mu(\Psi)^t}{\partial \Psi_i} C(\Psi)^{-1} (\mathbf{y} - \mu(\Psi)) \right)^t, \end{aligned} \quad (40)$$

it follows that

$$\begin{aligned} &\frac{\partial}{\partial \Psi_i} [(\mathbf{y} - \mu(\Psi))^t C(\Psi)^{-1} (\mathbf{y} - \mu(\Psi))] \\ &= -2 \frac{\partial \mu(\Psi)^t}{\partial \Psi_i} C(\Psi)^{-1} (\mathbf{y} - \mu(\Psi)) \\ &\quad + (\mathbf{y} - \mu(\Psi))^t \frac{\partial C(\Psi)^{-1}}{\partial \Psi_i} (\mathbf{y} - \mu(\Psi)). \end{aligned} \quad (41)$$

Then,

$$\begin{aligned} \frac{\partial l(\mathbf{y}; \Psi)}{\partial \Psi_i} &= -\frac{1}{2} \text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_i} \right) \\ &\quad + \frac{\partial \mu(\Psi)^t}{\partial \Psi_i} C(\Psi)^{-1} (\mathbf{y} - \mu(\Psi)) \\ &\quad - \frac{1}{2} (\mathbf{y} - \mu(\Psi))^t \frac{\partial C(\Psi)^{-1}}{\partial \Psi_i} (\mathbf{y} - \mu(\Psi)). \end{aligned} \quad (42)$$

Having obtained an expression for the derivatives of the log-likelihood, we are now ready to calculate the information matrix, i.e.:

$$\mathcal{I}(\Psi)_{ij} = E \left[\frac{\partial l(\mathbf{y}; \Psi)}{\partial \Psi_i} \frac{\partial l(\mathbf{y}; \Psi)}{\partial \Psi_j} \right]. \quad (43)$$

Let us define $\mathbf{z} = (\mathbf{y} - \mu(\Psi))$ for a shorter notation.

$$\begin{aligned} &\frac{\partial l(\mathbf{y}; \Psi)}{\partial \Psi_i} \frac{\partial l(\mathbf{y}; \Psi)}{\partial \Psi_j} \\ &= \frac{1}{4} \text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_i} \right) \text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_j} \right) \end{aligned} \quad (44a)$$

$$- \frac{1}{2} \text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_i} \right) \frac{\partial \mu(\Psi)^t}{\partial \Psi_j} C(\Psi)^{-1} \mathbf{z} \quad (44b)$$

$$+ \frac{1}{4} \text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_i} \right) \mathbf{z}^t \frac{\partial C(\Psi)^{-1}}{\partial \Psi_j} \mathbf{z} \quad (44c)$$

$$- \frac{1}{2} \frac{\partial \mu(\Psi)^t}{\partial \Psi_i} C(\Psi)^{-1} \mathbf{z} \text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_j} \right) \quad (44d)$$

$$+ \frac{\partial \mu(\Psi)^t}{\partial \Psi_i} C(\Psi)^{-1} \mathbf{z} \frac{\partial \mu(\Psi)^t}{\partial \Psi_j} C(\Psi)^{-1} \mathbf{z} \quad (44e)$$

$$- \frac{1}{2} \frac{\partial \mu(\Psi)^t}{\partial \Psi_i} C(\Psi)^{-1} \mathbf{z} \mathbf{z}^t \frac{\partial C(\Psi)^{-1}}{\partial \Psi_j} \mathbf{z} \quad (44f)$$

$$+ \frac{1}{4} \mathbf{z}^t \frac{\partial C(\Psi)^{-1}}{\partial \Psi_i} \mathbf{z} \text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_j} \right) \quad (44g)$$

$$- \frac{1}{2} \mathbf{z}^t \frac{\partial C(\Psi)^{-1}}{\partial \Psi_i} \mathbf{z} \frac{\partial \mu(\Psi)^t}{\partial \Psi_j} C(\Psi)^{-1} \mathbf{z} \quad (44h)$$

$$+ \frac{1}{4} \mathbf{z}^t \frac{\partial C(\Psi)^{-1}}{\partial \Psi_i} \mathbf{z} \mathbf{z}^t \frac{\partial C(\Psi)^{-1}}{\partial \Psi_j} \mathbf{z}. \quad (44i)$$

Since

$$E(\mathbf{z}) = E(\mathbf{y} - \mu(\Psi)) = 0, \quad (45)$$

and all odd order moments are also zero, terms (44b), (44d), (44f) and (44h) are zero in the above equation.

For calculating the expectation

$$E \left(\frac{\partial l(\mathbf{y}; \Psi)}{\partial \Psi_i} \frac{\partial l(\mathbf{y}; \Psi)}{\partial \Psi_j} \right), \quad (46)$$

we need to consider the expectations of terms (44a), (44c), (44e), (44g) and (44i). Before we compute the expected value of (44c) let us first compute

$$E \left(\mathbf{z}^t \frac{\partial C(\Phi)^{-1}}{\partial \Phi_j} \mathbf{z} \right).$$

Since

$$E(u^t v) = \text{tr}(E(v u^t)),$$

we have that

$$\begin{aligned} E \left(\mathbf{z}^t \frac{\partial C(\Psi)^{-1}}{\partial \Psi_j} \mathbf{z} \right) &= \text{tr} \left(\frac{\partial C(\Psi)^{-1}}{\partial \Psi_j} E(\mathbf{z} \mathbf{z}^t) \right) \\ &= \text{tr} \left(\frac{\partial C(\Psi)^{-1}}{\partial \Psi_j} C(\Psi) \right) \\ &= -\text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_j} \right), \end{aligned} \quad (47)$$

where the last equality follows from using equation (37). So the expected value of (44c) is

$$\begin{aligned} & \frac{1}{4} \text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_i} \right) E \left(z^t \frac{\partial C(\Psi)^{-1}}{\partial \Psi_j} z \right) \\ &= -\frac{1}{4} \text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_i} \right) \text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_j} \right) \end{aligned} \quad (48)$$

Analogously for (44g),

$$\begin{aligned} & E \left(\frac{1}{4} z^t \frac{\partial C(\Psi)^{-1}}{\partial \Psi_i} z \right) \text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_j} \right) \\ &= -\frac{1}{4} \text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_i} \right) \text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_j} \right) \end{aligned} \quad (49)$$

Let us now compute the expected value of (44e).

$$\begin{aligned} & E \left(\frac{\partial \mu(\Psi)^t}{\partial \Psi_i} C(\Psi)^{-1} z \frac{\partial \mu(\Psi)^t}{\partial \Psi_j} C(\Psi)^{-1} z \right) \\ &= \frac{\partial \mu(\Psi)^t}{\partial \Psi_i} C(\Psi)^{-1} E(z z^t) C(\Psi)^{-1} \frac{\partial \mu(\Psi)}{\partial \Psi_j} \\ &= \frac{\partial \mu(\Psi)^t}{\partial \Psi_i} C(\Psi)^{-1} \frac{\partial \mu(\Psi)}{\partial \Psi_j}. \end{aligned} \quad (50)$$

Before we compute the expected value of (44i) let us remind that if D_1 and D_2 are symmetric then

$$E(y^t D_1 y y^t D_2 y) = \text{tr}(D_1 D) \text{tr}(D_2 D) + 2 \text{tr}(D_1 D D_2 D),$$

where $D = E(y y^t)$. Using that result it comes that the expected value of (44i) is

$$\begin{aligned} & E \left(\frac{1}{4} z^t \frac{\partial C(\Psi)^{-1}}{\partial \Psi_i} z z^t \frac{\partial C(\Psi)^{-1}}{\partial \Psi_j} z \right) \\ &= \frac{1}{4} \text{tr} \left(\frac{\partial C(\Psi)^{-1}}{\partial \Psi_i} C(\Psi) \right) \text{tr} \left(\frac{\partial C(\Psi)^{-1}}{\partial \Psi_j} C(\Psi) \right) \\ &+ \frac{1}{2} \text{tr} \left(\frac{\partial C(\Psi)^{-1}}{\partial \Psi_i} C(\Psi) \frac{\partial C(\Psi)^{-1}}{\partial \Psi_j} C(\Psi) \right) \\ &= \frac{1}{4} \text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_i} \right) \text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_j} \right) \\ &+ \frac{1}{2} \text{tr} \left(\frac{\partial C(\Psi)^{-1}}{\partial \Psi_i} C(\Psi) \frac{\partial C(\Psi)^{-1}}{\partial \Psi_j} C(\Psi) \right), \end{aligned} \quad (51)$$

where the final equality follows from using (37) on the first term.

Thus, noting that the expectation of (44a) is just the term itself and taking the terms (48),(50),(49), and (51),

we get

$$\begin{aligned} & E \left(\frac{\partial l(\mathbf{y}; \Psi)}{\partial \Psi_i} \frac{\partial l(\mathbf{y}; \Psi)}{\partial \Psi_j} \right) \\ &= \frac{1}{4} \text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_i} \right) \text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_j} \right) \\ &- \frac{1}{4} \text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_i} \right) \text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_j} \right) \\ &+ \frac{\partial \mu(\Psi)^t}{\partial \Psi_i} C(\Psi)^{-1} \frac{\partial \mu(\Psi)}{\partial \Psi_j} \\ &- \frac{1}{4} \text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_i} \right) \text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_j} \right) \\ &+ \frac{1}{4} \text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_i} \right) \text{tr} \left(C(\Psi)^{-1} \frac{\partial C(\Psi)}{\partial \Psi_j} \right) \\ &+ \frac{1}{2} \text{tr} \left(\frac{\partial C(\Psi)^{-1}}{\partial \Psi_i} C(\Psi) \frac{\partial C(\Psi)^{-1}}{\partial \Psi_j} C(\Psi) \right). \end{aligned} \quad (52)$$

And finally, as the first and second, as well as the fourth and fifth term above cancel each other out, we get

$$\begin{aligned} \mathcal{I}(\Psi)_{ij} &= \frac{\partial \mu(\Psi)^t}{\partial \Psi_i} C^{-1}(\Psi) \frac{\partial \mu(\Psi)}{\partial \Psi_j} \\ &+ \frac{1}{2} \text{tr} \left(C^{-1}(\Psi) \frac{\partial C(\Psi)}{\partial \Psi_i} C^{-1}(\Psi) \frac{\partial C(\Psi)}{\partial \Psi_j} \right) \end{aligned} \quad (53)$$

REFERENCES

- [1] A. Medina, N. Taft, K. Salamatian, S. Bhattacharyya, and C. Diot, "Traffic matrix estimation: Existing techniques and new directions", ACM SIGCOMM, Pittsburg, USA. August 2002.
- [2] S. Vaton, J.S. Bedo, A. Gravey, "Advanced methods for the estimation of the Origin Destination traffic matrix", Revue du 25ème anniversaire du GERAD, 2005.
- [3] J. Cao, D. Davis, S. V. Wiel, B. Yu, "Time-varying network tomography," Journal of the American Statistical Association, vol. 95, pp. 1063–1075, 2000.
- [4] A. Medina, K. Salamatian, N. Taft, I. Matta, Y. Tsang, C. Diot, "On the Convergence of Statistical Techniques for Inferring Network Traffic Demands", Technical Report, 2003.
- [5] G.J. McLachlan and T. Krishnan, "The EM Algorithm and Extensions", John Wiley and Sons, Inc, 1997.
- [6] J. Cao, S. V. Wiel, B. Yu, Z. Zhu, "A scalable method for estimating network traffic matrices," Bell Labs Technical Report, 2001.
- [7] G. Liang, B. Yu, "Pseudo Likelihood Estimation in Network Tomography", in Infocom 2003.
- [8] Cisco NetFlow, www.cisco.com/warp/public/732/Tech/netflow.