Ilmari Juva, Sandrine Vaton, and Jorma Virtamo. 2006. Quick traffic matrix estimation based on link count covariances. In Proceedings of the 2006 IEEE International Conference on Communications (ICC 2006), Istanbul, Turkey, 11-15 June 2006, volume 2, pages 603-608.

# Quick Traffic Matrix Estimation Based on Link Count Covariances

Ilmari Juva[1], Sandrine Vaton[2], Jorma Virtamo[1]
1: Helsinki University of Technology, Finland, 2: ENST Bretagne, France
ilmari.juva@tkk.fi sandrine.vaton@enst-bretagne.fr jorma.virtamo@tkk.fi

***Abstract*— In this paper we consider the problem of traffic matrix estimation. As the problem is underconstrained, some additional information has to be brought in to obtain a solution. If we have a sequence of link count measurements available, a natural candidate is to use the link count sample covariance matrix under the assumption of a functional relationship between the mean and the variance of the traffic. We propose two computationally light-weight methods for traffic matrix estimation based on the covariance matrix, the projection method and constrained minimization method. The accuracy of these methods is compared with that of other methods using second order moment estimates by simulation under synthetic traffic scenarios.**
**Keywords: Traffic Matrix Estimation**

## I. Introduction

The traffic matrix, which gives volume of traffic between each origin/destination (OD) pair in the network, is a required input in many network management tasks. Such tasks include for instance traffic engineering and network capacity dimensioning. In many cases the knowledge on the underlying traffic volumes is assumed to be known. It is recognized that accurate demand matrices are crucial for traffic engineering. However, in reality, they are seldom available in current IP networks.

This raises the need to estimate the traffic matrix based on information readily available. Traffic matrix inference usually utilizes SNMP measurement data. While it has some limitations, the attractive feature of SNMP is that it is usually available everywhere in an IP network and is the only widespread tool to obtain link count data. We denote a set of link count measurements by $\boldsymbol{y}$. The routing matrix $\boldsymbol{A}$ is also readily available.

In traffic matrix estimation, the basic relationship between link counts $\boldsymbol{y}$ and origin-destination counts $\boldsymbol{x}$ can be written as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}, \tag{1}$$

where $\boldsymbol{A}$ is the routing matrix and $\boldsymbol{x}$ is the OD counts in vector form, i.e. each component represents a traffic of one OD pair. The above equation holds exactly and also the equation where we take the expectations from $\boldsymbol{y}$ and $\boldsymbol{x}$ holds. The expected valued of $\boldsymbol{x}$ is the traffic matrix $\boldsymbol{\lambda}$, and we get the first moment equation of traffic matrix estimation

$$\overline{\boldsymbol{y}} = \boldsymbol{A}\boldsymbol{\lambda}, \tag{2}$$

where $\overline{\boldsymbol{y}}$ denoted the sample mean of the link counts.

Since in any realistic network there are many more OD pairs than links, the problem of solving $\boldsymbol{\lambda}$ from $\boldsymbol{A}$ and $\boldsymbol{y}$ is strongly underdetermined. This means that accurate explicit solutions cannot be found, as there is an infinite number of solutions for $\boldsymbol{\lambda}$ that satisfy equation (2). To overcome this ill-posedness, some type of additional information has to be brought in to solve the problem. Reviews of the proposed methods can be found e.g. in [1], [2] and [3].

Typically, methods that need a prior distribution use the gravity model to obtain one. However, the gravity assumption that the traffic volume of an OD pair is proportional to the total traffic sent by the origin node and the total traffic terminating at the destination node, does not always hold. In [4] the authors study real traffic matrix of a North American backbone network and conclude that there are significant errors concerning the estimation of the largest OD pairs, which are the most important ones for traffic engineering purposes. Therefore, in this paper we propose another way of obtaining prior distribution based on the link count covariances and a functional mean-variance relationship. Based on this, we develop two computationally lightweight methods, similar in principle to the tomogravity method of [5], in the sense that they incorporate a starting point and link count measurements to obtain an estimate.

The rest of the paper is organized as follows. In section II we discuss methods relying on the link covariance as the additional information. In section III we show how to solve OD pair covariance matrix from the link count covariance matrix. Section IV and V present our quick estimation methods, the projection method and the constrained minimization method. In section VI we compare the performance of these methods to maximum likelihood estimation. And finally section VII concludes the paper.

## II. Methods based on link count covariances

Maximum likelihood estimation (MLE) uses the second moment statistic, the link count covariance, as the additional information that is needed to yield an estimate. It is also necessary to assume local stationarity for the measurements considered, and a distribution which the stochastic fluctuation of the traffic follows.

In [6], Vardi first proposed this kind of approach. The Poisson distribution is assumed, meaning that variance is equal to the mean, and the system becomes

$$\begin{pmatrix} \overline{\boldsymbol{y}} \\ \epsilon \boldsymbol{S}^{(y)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{A} \\ \epsilon \boldsymbol{B} \end{pmatrix} \boldsymbol{\lambda}, \tag{3}$$


*This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.*

where, as is explained in more detail in section III, $\boldsymbol{S}^{(y)}$ is the sample link covariance matrix and $\boldsymbol{B}$ is the matrix of element-wise products of rows of $\boldsymbol{A}$. Coefficient $\epsilon \in (0,1]$ defines how much weight is given to the second moment estimate in the final solution, and $\boldsymbol{\lambda}$ is the estimator for the mean of $\boldsymbol{x}$. This is a linear inverse positive, or LININPOS, problem and can be solved by numerical likelihood methods, such as the EM-algorithm. The solution obtained this way minimizes the Kullback-Leibler distance between the observed moments and theoretical values. If we instead minimize (3) in least square sense, the solution is easily obtained in closed form.

Vardi's method, however, does not give very accurate estimates, as was discovered in [2] and [4]. This is due to the fact that the Poisson assumption is not accurate in current IP networks. Cao et al. [7] generalize the maximum likelihood approach by assuming a Gaussian traffic distribution and assuming that the variance is related to the mean through a power-law. While this MLE approach is efficient and theoretically justifiable, the size of the problem in traffic matrix estimation requires the use of iterative numerical methods, such as the Expectation Maximization algorithm, which is computationally quite heavy.

The MLE relies on the fact that the system of first and second order link count statistics together make the system identifiable with regard to the first order OD-pair statistics, i.e. we are able to find solution for the likelihood equations if there exists a functional relationship between the mean and the variance of OD-pair traffic. The commonly used relation is the power-law relation

$$\boldsymbol{\Sigma} = \phi \cdot \operatorname{diag}\{\boldsymbol{\lambda}^c\}, \tag{4}$$

where $\boldsymbol{\Sigma}$ is a diagonal matrix, because we assume independence between OD pairs.

But, in fact, the second order statistic for OD-pairs is identifiable based solely on the second order statistic of the link counts, as long as we assume independence among OD-pairs and a sensible routing scheme. This result is proven by Soule et al. [8]. Since we can analytically solve the variance of the OD-pairs by least squares method, and the power-law relation between variance and mean is assumed, we can then solve the traffic matrix from our variance estimate.

The benefit is that this does not call for numerical methods, and is thus extremely quick to calculate. The problem with this approach is that it does not take into account the first moment equation 2, which is a stronger condition as opposed to the mean-variance relation which is only an assumption. Our methods incorporate this information into the solution obtained through estimation of the variance yet maintaining the computational simplicity of the model.

## III. Solving OD-pair covariance matrix from link counts

Let us denote the number of links by $J$ and the number of OD-pairs by $N$. Then the vector form of traffic matrix $\boldsymbol{x}$ has the dimension $(N \times 1)$, link loads $\boldsymbol{y}$ has the dimension $(J \times 1)$.

First, let us define $\boldsymbol{S}^{(y)}$ as a $\frac{1}{2}J(J+1)$-vector containing diagonal and upper triangle elements of the link covariance matrix $\boldsymbol{\Sigma}^{(y)}$. Define $\boldsymbol{S}^{(x)}$ as an $N$-vector containing the diagonal elements of the OD-pair covariance matrix $\boldsymbol{\Sigma}^{(x)}$. $\boldsymbol{A}$ is the $(J \times N)$ routing matrix, whose element $A_{i,j}$ is 1 if OD pair $x_j$ uses link $y_i$, and 0 otherwise. Then define a $(\frac{1}{2}J(J+1) \times N)$ matrix $\boldsymbol{B}$ that relates vector $\boldsymbol{S}^{(y)}$ to vector $\boldsymbol{S}^{(x)}$. A row of $\boldsymbol{B}$ is indexed by a compound index $(ij)$ where $i = 1, \ldots, J;\ j = i, \ldots, J$, meaning that the index runs through $\frac{1}{2}J(J+1)$ values,

$$B_{(ij),k} = A_{i,k}A_{j,k} \qquad i = 1, \ldots, J;\ j = i, \ldots, J \quad k = 1, \ldots, N.$$

In vector form this reads,

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{A}_1 \star \boldsymbol{A}_1 \\ \boldsymbol{A}_1 \star \boldsymbol{A}_2 \\ \vdots \\ \boldsymbol{A}_1 \star \boldsymbol{A}_c \\ \boldsymbol{A}_2 \star \boldsymbol{A}_2 \\ \boldsymbol{A}_2 \star \boldsymbol{A}_3 \\ \vdots \\ \boldsymbol{A}_c \star \boldsymbol{A}_c \end{pmatrix},$$

where $\boldsymbol{A}_i$ denotes the $i$th row of $\boldsymbol{A}$, and the componentwise product is denoted with the star ($\star$). Now the rows of $\boldsymbol{B}$ indicate the elements of $\boldsymbol{x}$ contributing to covariance between links $i$ and $j$.

The measured link covariance matrix can be written as

$$\boldsymbol{\Sigma}^{(y)} = \sum_k \sigma_k^2 \boldsymbol{a}_k \boldsymbol{a}_k^{\mathrm{T}}, \tag{5}$$

where $\boldsymbol{a}_i$ is the $i$th column of $\boldsymbol{A}$. In component form we have

$$\boldsymbol{\Sigma}_{i,j}^{(y)} = \sum_k \sigma_k^2 A_{i,k} A_{j,k}. \tag{6}$$

Using vector notation, the equation becomes

$$\boldsymbol{S}^{(y)} = \boldsymbol{B}\boldsymbol{S}^{(x)}. \tag{7}$$

This is in fact quite similar to (3) in the case where $\epsilon$ would be set very large, leading to the part $\boldsymbol{S}^{(y)} = \boldsymbol{B}\boldsymbol{\lambda}$ to dominate the equation. We just have a more general power-law relation instead of the Poisson assumption, so we cannot now just replace $\boldsymbol{S}^{(x)}$ with $\boldsymbol{\lambda}$.

Typically $\frac{1}{2}J(J+1) > N$ and equation (7) is overdetermined. The least square estimate (LSE) solution (see e.g. [9]), to the equation is

$$\boldsymbol{S}^{(x)} = (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{S}^{(y)}. \tag{8}$$

## IV. Projection method

Now that we have an estimate for the variances of each OD-pair, it is trivial to find an estimate of the mean by using the mean-variance relation (4).

$$\boldsymbol{\lambda}_0 = (\phi^{-1}\boldsymbol{S}^{(x)})^{\frac{1}{c}}. \tag{9}$$

The problem with this estimate is, that it does not require the solution to satisfy the link count equation (2), which is a stronger condition than the second moment relation. The preliminary estimate $\boldsymbol{\lambda}_0$ can be improved by projecting the result to the surface that satisfies the first moment condition. This yields our estimate

$$\boldsymbol{\lambda} = \boldsymbol{\lambda}_0 + \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})^{-1}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\lambda}_0). \tag{10}$$

Compared to the maximum likelihood approach, we do the moment estimation sequentially: First obtaining an estimate for the covariance matrix and then solving for the mean. This does not yield quite as accurate estimates as MLE, but it is many times faster.

The projection might yield negative values for smaller OD pairs, as no positivity constraint is imposed in order to keep the method as light-weight as possible. We simply substitute the negative estimates by zero, concluding that these OD pairs are negligibly small.

The projection method works for any fixed parameters $\phi$ and $c$. In fact, we can try to estimate these parameters by requiring that $\boldsymbol{\lambda}_0$ comes as close as possible to satisfying (2), i.e. that they minimize

$$\begin{aligned} f(\phi, c) &= (\overline{\boldsymbol{y}} - \boldsymbol{A}\boldsymbol{\lambda}_0)^{\mathrm{T}}(\overline{\boldsymbol{y}} - \boldsymbol{A}\boldsymbol{\lambda}_0) \qquad (11)\\ &= (\overline{\boldsymbol{y}} - \boldsymbol{A}(\phi^{-1}\boldsymbol{S})^{\frac{1}{c}})^{\mathrm{T}}(\overline{\boldsymbol{y}} - \boldsymbol{A}(\phi^{-1}\boldsymbol{S})^{\frac{1}{c}}). \end{aligned}$$

The values of $\phi$ and $c$ that realize the minimum, can now be used in equation (9) to yield the estimate $\boldsymbol{\lambda}$.

### A. *Estimating the parameters $\phi$ and $c$*

In Cao et al. [7], the EM-algorithm is run after preselecting a convenient value for the exponent parameter $c$ in the power law relation (4), while $\phi$ remains a parameter that the algorithm optimizes. The authors point out that convergence is guaranteed for the algorithm only for integer values of $c$, namely 1 or 2. However, Gunnar et al. [4] in their study of the Global Crossing data find out that the correct values for $c$ in those particular networks are 1.5 and 1.6 for the European and North American core-networks respectively. Thus being limited to integer values in the solution makes sense for only computational reasons. The projection method, on the other hand, works for any preselected $c$. And, in fact, we can relax $c$ to be a free parameter, though this means that we will no longer be able to obtain a closed form solution.

Minimization of (11) with respect to $\phi$ is a simple quadratic problem. So we can easily find the minimizing value $\widehat{\phi}(c)$. Now we can either use a preselected value for $c$ to yield the optimal $\phi$ value, or insert $\widehat{\phi}(c)$ back to (11), which yields

$$f(\widehat{\phi}(c), c) = (\overline{\boldsymbol{y}} - \boldsymbol{A}(\widehat{\phi}(c)^{-1}\boldsymbol{S})^{\frac{1}{c}})^{\mathrm{T}}(\overline{\boldsymbol{y}} - \boldsymbol{A}(\widehat{\phi}(c)^{-1}\boldsymbol{S})^{\frac{1}{c}}). \tag{12}$$

Now we have a simple one parameter numerical optimization to find the optimal value of $c$. Expression (12) as a function of $c$ is depicted in Figure 1. The figure is based on a set of synthetic data that was generated by using the parameter value $c = 1.5$. Figure 2 shows a histogram of estimated values for the parameter $c$.

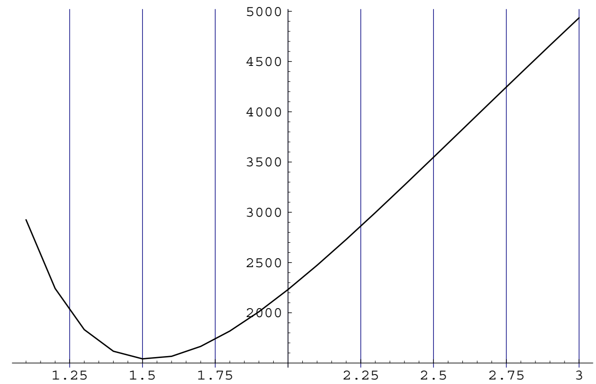


Fig. 1. Values of the objective function (12) as a function of parameter $c$.

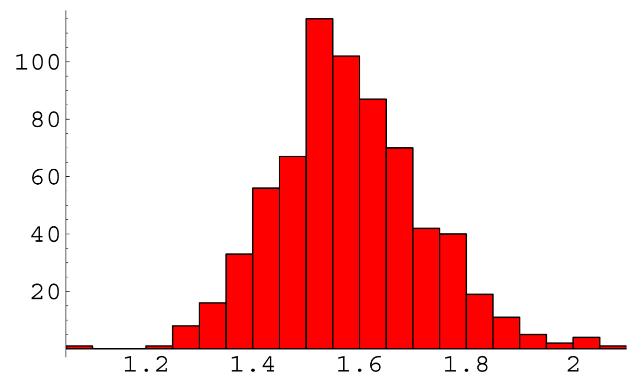


Fig. 2. Estimated c-values for synthetic data sets of size 500, generated with setting $c = 1.5$.

## V. Constrained minimization

Another approach is to require the condition $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{\lambda}$ to be satisfied from the outset, and try to satisfy the mean-variance relation in the least square sense. In general, this has to be solved numerically. However, in the special case of $c = 1$ an explicit solution can be derived.

This approach is equivalent to Vardi's method, if we set $\epsilon$ very small so that the first moment is the dominant factor in the estimation, with the exception that we treat $\phi$ as a parameter to be optimized, whereas in (3) it is fixed as 1 by the Poisson assumption.

We get a constrained minimization problem

$$\begin{aligned} \min_{\boldsymbol{\lambda}, \phi} \quad & \|\boldsymbol{S}^{(y)} - \boldsymbol{B}\phi\boldsymbol{\lambda}^c\| \qquad (13)\\ \text{subject to} \quad & \boldsymbol{y} = \boldsymbol{A}\boldsymbol{\lambda}. \end{aligned}$$

Introducing a vector of Lagrange multipliers $\boldsymbol{\alpha}$, the objective function to be minimized can be written as

$$\begin{aligned} & f(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \phi) \\ & = (\boldsymbol{S}^{(y)} - \phi\boldsymbol{B}\boldsymbol{\lambda})^{\mathrm{T}}(\boldsymbol{S}^{(y)} - \phi\boldsymbol{B}\boldsymbol{\lambda}) + 2\boldsymbol{\alpha}^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\lambda}) \\ & = \phi^2\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{\lambda} - 2\phi{\boldsymbol{S}^{(y)}}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{\lambda} - 2\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\lambda} \\ & \quad + {\boldsymbol{S}^{(y)}}^{\mathrm{T}}\boldsymbol{S}^{(y)} + 2\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{y}. \qquad (14) \end{aligned}$$

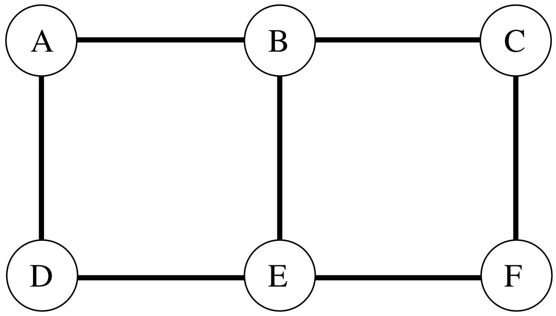


Fig. 3. Six node Test topology

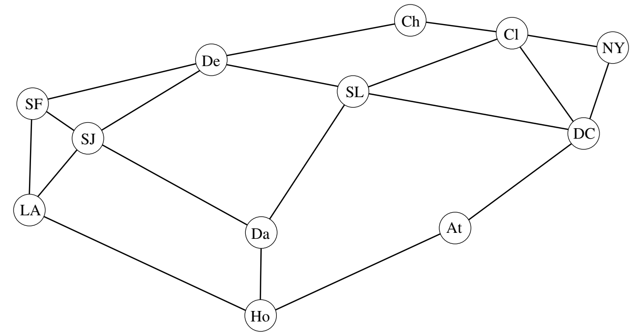


Fig. 4. Twelve node backbone test topology

The above expression is quadratic in $\boldsymbol{\lambda}$, and the minimum with respect to $\boldsymbol{\lambda}$ can easily be found,

$$\boldsymbol{\lambda} = \phi^{-2}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\alpha} + \phi\boldsymbol{B}^{\mathrm{T}}\boldsymbol{S}^{(y)}). \tag{15}$$

The Lagrange multipliers $\boldsymbol{\alpha}$ are then determined such that the constraints are satisfied:

$$\boldsymbol{y} = \boldsymbol{A}\phi^{-2}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\alpha} + \phi\boldsymbol{B}^{\mathrm{T}}\boldsymbol{S}^{(y)}), \tag{16}$$

from which

$$\begin{aligned}\boldsymbol{\alpha} = (\phi^{-2}\boldsymbol{A}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{A}^{\mathrm{T}})^{-1} \cdot \\ \cdot (\boldsymbol{y} - \phi^{-1}\boldsymbol{A}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{S}^{(y)}).\end{aligned} \tag{17}$$

Minimizing $f(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \phi)$ with respect to $\phi$ yields

$$\phi = (\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{\lambda})^{-1}{\boldsymbol{S}^{(y)}}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{\lambda}. \tag{18}$$

Substitution of (17) into (15) gives $\boldsymbol{\lambda}$ as a function of $\phi$

$$\boldsymbol{\lambda} = \boldsymbol{K}\boldsymbol{y} - \phi^{-1}\big(\boldsymbol{K}\boldsymbol{A}(\boldsymbol{B}^{T}\boldsymbol{B})^{-1}\boldsymbol{B}^{T}\boldsymbol{S}^{(y)} + \boldsymbol{B}^{T}\boldsymbol{S}^{(y)}\big),$$

where we use the notation

$$\boldsymbol{K} = (\boldsymbol{B}^{T}\boldsymbol{B})^{-1}\boldsymbol{A}^{T}(\boldsymbol{A}(\boldsymbol{B}^{T}\boldsymbol{B})^{-1}\boldsymbol{A}^{T})^{-1}.$$

Substituting $\boldsymbol{\lambda}$ further in (18) yields an quadratic equation for $\phi$, which is easily solvable. This solution can be then substituted back to (17) and (15) to obtain the explicit expression for $\boldsymbol{\lambda}$.

## VI. Comparison with the MLE method

The accuracy of the quick methods are evaluated by comparing them against Maximum likelihood estimation. In the following subsection we present the Maximum likelihood estimation used. In the subsequent sections the results of accuracy on synthetic data test cases is presented.

### A. Maximum Likelihood Estimation

We follow the approach of Cao et al. [7] in using the Expectation Maximization (EM) algorithm. For a review see also [2].

The log-likelihood for estimating $\boldsymbol{\lambda}$ is given as

$$\begin{aligned} l(\boldsymbol{\theta}|\boldsymbol{Y}) = -\frac{\tau}{2}\log|\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\mathrm{T}}| - \\ -\frac{1}{2}\sum_{t=1}^{\tau}(\boldsymbol{y}_t - \boldsymbol{A}\boldsymbol{\lambda})^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\mathrm{T}})^{-1}(\boldsymbol{y}_t - \boldsymbol{A}\boldsymbol{\lambda}),\end{aligned} \tag{19}$$

where $\tau$ is the number of measurements and $\boldsymbol{y}_t$ is the link count vector for measurement $t$. In Cao et al. $c$ is assumed to be constant and the parameters of the model are thus

$$\boldsymbol{\theta} = (\phi, \boldsymbol{\lambda}).$$

We can write $\boldsymbol{\Sigma}$ as a function of $\boldsymbol{\theta}$ according to (4). Now the problem can be solved numerically with the EM-algorithm. The complete data log-likelihood is of the form

$$l(\boldsymbol{\theta}|\boldsymbol{X}) = -\frac{\tau}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{t=1}^{\tau}(\boldsymbol{x}_t - \boldsymbol{\lambda})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_t - \boldsymbol{\lambda}).$$

The EM-equation is

$$\begin{aligned} &Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \mathrm{E}[l(\boldsymbol{\theta}|\boldsymbol{X})|\boldsymbol{Y}, \boldsymbol{\theta}^{(k)}] \\ &= \mathrm{E}[-\frac{\tau}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{t=1}^{\tau}(\boldsymbol{x}_t - \boldsymbol{\lambda})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_t - \boldsymbol{\lambda})|\boldsymbol{Y}, \boldsymbol{\theta}^{(k)}]\end{aligned}$$

Since

$$\begin{aligned} &\mathrm{E}[(\boldsymbol{x} - \boldsymbol{\lambda})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\lambda})] \\ &= \mathrm{E}[\mathrm{Tr}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\lambda})(\boldsymbol{x} - \boldsymbol{\lambda})^{\mathrm{T}}\}] \\ &= \mathrm{Tr}\{\boldsymbol{\Sigma}^{-1}\mathrm{E}[(\boldsymbol{x} - \boldsymbol{\lambda})(\boldsymbol{x} - \boldsymbol{\lambda})^{\mathrm{T}}]\} \\ &= \mathrm{Tr}\{\boldsymbol{\Sigma}^{-1}\mathrm{E}[((\boldsymbol{x} - \boldsymbol{m}) + (\boldsymbol{m} - \boldsymbol{\lambda}))((\boldsymbol{x} - \boldsymbol{m}) + (\boldsymbol{m} - \boldsymbol{\lambda}))^{\mathrm{T}}]\} \\ &= \mathrm{Tr}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{R} + (\boldsymbol{m} - \boldsymbol{\lambda})(\boldsymbol{m} - \boldsymbol{\lambda})^{\mathrm{T}})\} \\ &= \mathrm{Tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{R}\} + (\boldsymbol{m} - \boldsymbol{\lambda})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{m} - \boldsymbol{\lambda})\end{aligned}$$

we can write

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) &= -\frac{\tau}{2}(\log|\boldsymbol{\Sigma}| + \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{R}^{(k)})) \\ &\quad -\frac{1}{2}\sum_{t=1}^{\tau}(\boldsymbol{m}_t^{(k)} - \boldsymbol{\lambda})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{m}_t^{(k)} - \boldsymbol{\lambda}),\end{aligned}$$

where

$$\begin{aligned} \boldsymbol{m}_t^{(k)} &= \mathrm{E}[\boldsymbol{x}_t|\boldsymbol{y}_t, \boldsymbol{\theta}^{(k)}] \\ &= \boldsymbol{\lambda}^{(k)} + \boldsymbol{\Sigma}^{(k)}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{\Sigma}^{(k)}\boldsymbol{A}^{\mathrm{T}})^{-1}(\boldsymbol{y}_t - \boldsymbol{A}\boldsymbol{\lambda}) \\ \boldsymbol{R}^{(k)} &= \mathrm{Var}[\boldsymbol{x}_t|\boldsymbol{y}_t, \boldsymbol{\theta}^{(k)}] \\ &= \boldsymbol{\Sigma}^{(k)} - \boldsymbol{\Sigma}^{(k)}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{\Sigma}^{(k)}\boldsymbol{A}^{\mathrm{T}})^{-1}\boldsymbol{A}\boldsymbol{\Sigma}^{(k)}.\end{aligned}$$

According to [7], convergence to the maximum likelihood estimate is guaranteed in the special cases of $c = 1$ and $c = 2$.

*This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.*

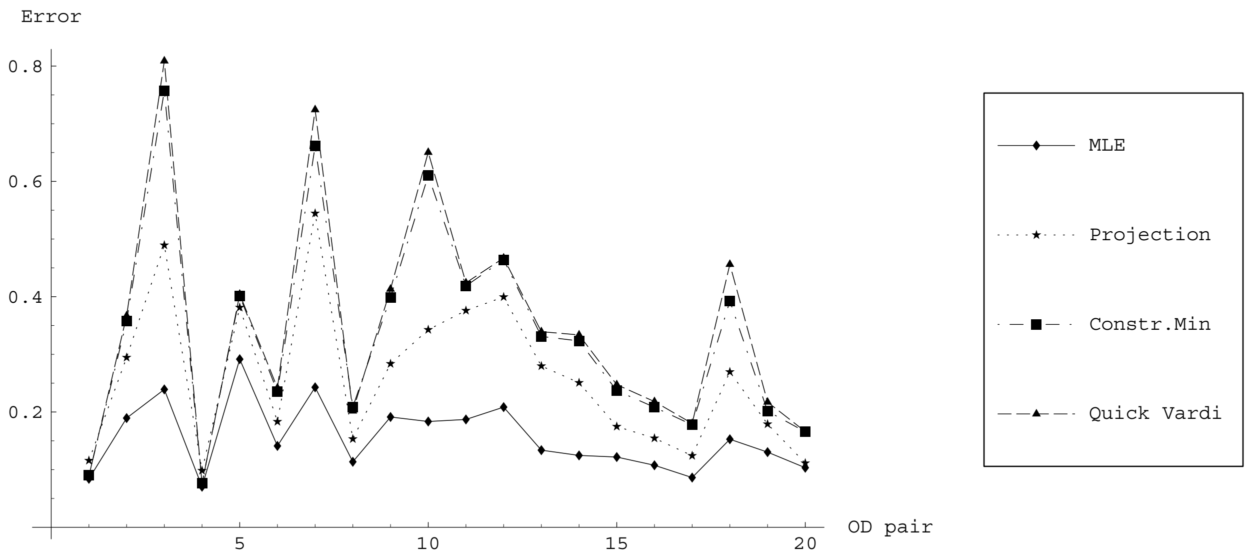


Fig. 5. Errors for OD pairs in 6-node topology in ascending order of traffic amount for case $c = 1$.

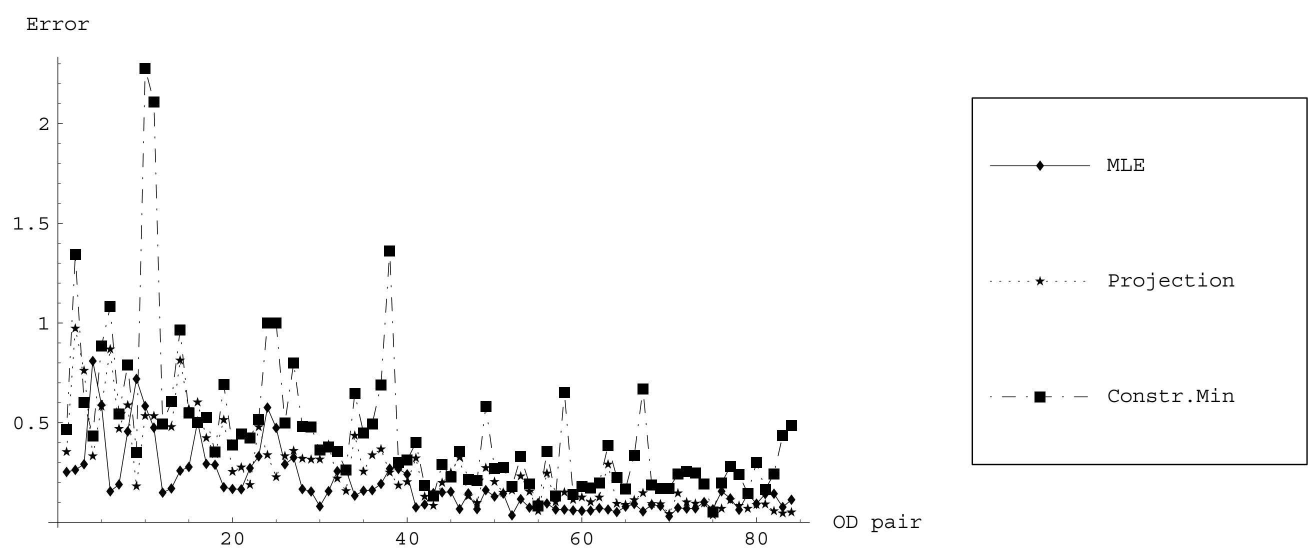


Fig. 6. Errors for largest OD pairs in 12-node topology in ascending order of traffic amount

### *B. Results*

For the evaluation of the methods we use two topologies. A small six-node topology, shown in Figure 3, has 14 one-way links, two links between each connected pair of node. Assuming traffic from each node to all other nodes, there are 30 OD pairs in the network. In the more realistic size fictitious backbone topology shown in Figure 4, there are 12 nodes, 38 links, and 132 OD pairs. For both topologies, we generate synthetic Gaussian data sets, where the power-law holds. Sample size is set to 500 measurements for each simulation.

*1) A Simple six node topology:* In our synthetic OD pair traffic the traffic varies so that the largest OD pairs are ten times as large as the smallest ones.

Figure 5 shows the results for the maximum likelihood estimates, projection method, the constrained minimization, and Vardi's method solved with the least square method, which we call here Quick Vardi. The synthetic data used for the evaluations is generated with parameters $c = 1$, $\phi = 1$. This is equivalent to the Poisson assumption made in Vardi's method.

The OD pairs are presented in ascending order based on the traffic volume, so that the smaller OD pairs are on the left and the largest on the right. We see that, as expected, the MLE performs better on average, but not overwhelmingly so. The mean relative errors are $15\%$, $26\%$, $34\%$ and $35\%$ for the MLE, the projection method, constrained minimization and Quick Vardi respectively.

*2) A 12 node backbone topology:* In this example case we use synthetic data generated with the parameter value $c = 1.5$. The traffic volumes for the OD pairs vary so that the largest are approximately hundred times as large as the smallest ones. This creates great difficulties for the quick methods regarding the estimation of the smaller OD pairs. The estimates of the projection method for the smallest OD pairs are far off the real traffic volumes. Due to the fact that the estimates for some of the smallest OD pairs have errors of several hundred percent, the mean relative error is also affected greatly by these, and is $59\%$ for the projection method and $110\%$ for the constrained optimization, while it is $29\%$ for the MLE. The mean error for the Quick Vardi method is several hundred percent, so it is not considered here.

However, the most important thing is to estimate the largest OD pairs. If we concentrate only on the largest OD pairs that comprise $90\%$ of total traffic in volume, the projection method is more competitive. The errors for these OD pairs are shown in Figure 6. The errors are $27\%$ for the projection method, $46\%$ for the constrained minimization and $19\%$ for the MLE.

## VII. Conclusion and Future Work

This paper presented ways to obtain estimate for traffic matrix by explicit calculations utilizing the link count covariance matrix. We illustrated how one can obtain the OD pair traffic variance estimates from empirical link count covariance matrix, and developed computationally light weight methods, the projection method and the constrained minimization method, to obtain an estimate for the traffic matrix based on the link count covariance matrix, in a way that would still be consistent with the link counts.

The constrained minimization method was recognized, in fact, to be a special case of Vardi's method. We give an explicit solution for it in the case $c = 1$ and also obtain an estimate for the second parameter $\phi$ in the mean-variance relation. For the projection method we have an even simpler and quicker to compute solution. Also in this case we get estimates for the parameters $c$ and $\phi$.

We evaluated the accuracy of the methods in a simulation study by comparing them against the maximum likelihood solution by Cao et al., and found that they perform reasonably well, considering that they are much quicker and simpler to calculate than the MLE, which requires the use of an iterative numerical method, namely the EM-algorithm. In the worst case, the errors in the estimate of a traffic matrix element for the largest components given by the quick method were three times as large as those by the MLE method, in many cases they difference was smaller. As for the running time, the difference between the MLE method and quick methods was big. With our non-optimized Mathematica code running the MLE method took of the order of tens of minutes, while the quick methods yielded the result in a few seconds.

In this paper all comparisons were made with synthetic data. Evaluation with real data would be very important to assess the true effectiveness of the methods. For now, we have used in our evaluations a sample size of $500$, which may be rather large in comparison to what is available in reality. Accuracy of the estimated covariance matrix with various sample sizes should be studied, as well as the effect the measurement inaccuracies have on the subsequent traffic matrix estimates.

## Acknowledgments

The work for this paper was carried out as a part of project Fancy funded by Academy of Finland, of the Specific Joint Research Project on Traffic matrix estimation funded by Euro-NGI, and of the research project on the traffic matrix funded by "region Bretagne" (operation A3C942, program 1042).

## 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] A. Soule, A. Lakhina, N. Taft, K. Papagiannaki, K. Salamatian, A. Nucci, M. Crovella, and C. Diot, Traffic matrices: Balancing measurements, inference and modeling, in SIGMETRICS'05, Banff, Canada, June 2005.

[4] A. Gunnar, M. Johansson, and T. Telkamp, "Traffic matrix estimation on a large IP backbone – A comparison on real data", IMC'04, Taormina, Italy, 2004.

[5] Y. Zhang, M. Roughan, N. Duffield, A. Greenberg, "Fast Accurate Computation of Large-Scale IP Traffic Matrices from Link Loads", ACM Sigmetrics 2003.

[6] Y. Vardi, "Network tomography: estimating source-destination traffic intensities from link data", Journal of the American Statistical Association, vol. 91, pp. 365–377, 1996.

[7] 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.

[8] A. Soule, A. Nucci, R. Cruz, E. Leonardi, N. Taft, "How to Identify and Estimate the Largest Traffic Matrix Elements in a Dynamic Environment" SIGMETRICS / Performance'04, New York, USA, pp. 73–84, 2004.

[9] David G. Luenberger, "Optimization by Vector Space Methods", New York NY, John Wiley & Sons, 1969.