Numerical Algorithms for Dynamic Traffic Demand Estimation Between Zones in a Network

Wai-Ki Ching*, Stefan Scholtes† & Shu-Qin Zhang‡

Abstract

We present numerical methods for dynamic traffic demand estimation between N zones in a network, where the zones are disjoint subsets of nodes of the network. Traffic is assumed to be generated or absorbed only in the zones and nowhere else in the network. Traffic volumes between zones over a fixed period of time are modelled as independent random variables with unknown means which we wish to estimate. For each zone, the volume of all incoming and outgoing traffic is counted on a regular basis but no information about the origin or destination of the observed

traffic is used. We suggest procedures for a regular update of estimates of the $N(N-1)$ mean traffic demands between the zones on the basis of an incoming stream of the $2N$ traffic counts. The procedures are based on an exponential smoothing scheme and are reminiscent of the expectation maximization (EM) algorithm if smoothing is removed. Fast and reliable numerical algorithms, based on the conjugate gradient method, are presented for normal as well as for Poisson traffic demands. The Poisson case is linked with entropy maximization. Computational tests based on simulated data demonstrate both the numerical and statistical efficiency of the procedures.

*Department of Mathematics, University of Hong Kong, Pokfulam Road, Hong Kong, wkc@maths.hku.hk

†Judge Institute of Management, University of Cambridge, Cambridge CB2 1AG, England, s.scholtes@jims.cam.ac.uk

‡Department of Mathematics, University of Hong Kong, Pokfulam Road, Hong Kong

Keywords: Traffic Demand Estimation; Traffic Network; Newton's Method; Conjugate Gradient Method; Entropy Maximization.

1 Introduction

The estimation of inter-zonal traffic demands from traffic counts on links in a network is an important problem in transportation planning and has been studied intensively in the transportation literature [2, 3, 4, 5, 18, 19, 21, 22, 24, 25, 32, 33]. In this context, traffic demands can refer to the number of passengers or vehicles or to tonnage of cargo that are processed from their origins to their destinations through a network of roads, or even to the number of data packets that are sent through an electronic network such as the internet. A key question is how to estimate the unobserved demands for traffic between various nodes, or zones, in the network from the observed traffic flows on the links of the network.

Traditional urban traffic demand estimation is based on detailed demographic data, such as travelers' income and household size, and employs a complex four-step procedure, focusing on trip generation, trip distribution, modal split and trip assignment, see for instance [24, pp. 11-16] and [11, pp. 17-32]. In parallel to these socio-economic modelling efforts, researchers have investigated the problem of estimating traffic demands on the mere basis of counted traffic on some of the links in the network. The present paper contributes to this literature. The approach is particularly appealing since the data collection can be automated through the use of induction loops. Link-flow based traffic demand estimation is typically classified into two categories, depending on whether or not congestion effects are taken into account. Cascetta and Nguyen [4, 5] and Ortuzar and Willumsen [25] review various methods for uncongested demand estimation, including maximum likelihood [27], maximum entropy [28], least square [3] and Bayesian [19] approaches. Estimation in the presence of congestion has been studied, among others, by Chen and Florian [6], Fisk [14, 15], Maher and Zhang [20], and Yang et al. [32, 33].

The contribution of our paper lies in the following aspects: First, we suggest a model which assumes that we can decide a priori on which links traffic is counted. This is in

contrast to the existing literature, which assumes that the links on which traffic is counted are exogenously determined and not under our control. Secondly, we suggest an updating procedure, reminiscent to forecasting, which uses regularly arriving data to update mean demand estimations on-line. Finally, we provide fast and reliable numerical algorithms for normal and Poisson demands and test their numerical and statistical performance.

2 The Traffic Network Model

We assume that a given traffic network is partitioned into N zones, each zone being a set of nodes of the traffic network. The zones are assumed to be disjoint but do not necessarily cover the whole network. We make the following assumptions:

(A1) For each zone, we assume that all the incoming and outgoing traffic can be observed and is counted regularly;

(A2) The traffic generated or absorbed anywhere in the network outside of the N zones is negligible;

(A3) The mean demand between any two zones in the network is strictly positive.

Assumption (A1) requires that we can decide a priori where the traffic is counted: At all entering and exiting arcs of each zone. Assumption (A3) is used for notational convenience. If we know that demand between certain zones is zero, then the model can be straight-forwardly adapted by excluding this zone-pair from consideration in the estimation procedures.

It is convenient to represent this set-up by a star-shaped meta-network, as depicted in Figure 1, whose nodes are the zones and a hub which is connected to the zones. The hub represents the possibly empty set of nodes of the real network which are not contained in any of the zones, whilst the arcs to and from the hub carry the total counted traffic out of or into the zones. Since traffic counts in our model are aggregated at the zone levels, congestion effects are less relevant for demand estimation in our model. Indeed, in a star

network any two zones are connected by a unique path and therefore network users, at the level where we observe them, cannot switch from a slow congested path to a quicker less congested path. In addition to their use for zone-based demand estimation in general traffic networks, star networks can be directly employed as models of traffic junctions. The counting of incoming and outgoing traffic can be automated by installing induction loops. One example of this type of network model is the traffic flow in a linear network, such as a motorway, where counts can be made on access and exit roads. Automated and regularly up-dated estimations of mean traffic demands over various time periods can then be usefully employed, e.g., in the design of time-varying traffic light or speed controls.

We model traffic demands from zone i to zone j as independent random variables whose mean we wish to estimate. We assume that the total amount of traffic entering or leaving any one of the zones is counted on a regular basis, say every day during a fixed time interval, but that there is no information available on the destinations of outgoing or origins or incoming traffic. The fixed time interval can be a full day of a particular peak hour. We also assume that all vehicles which enter the network (leaving their origin) during the fixed time interval also exit (arriving at their destination) during the fixed time interval. We will use the following notations:

(i) $Z_{ij}^{(k)}$ is the *unobserved* amount of traffic from zone i to zone j in the kth period, i.e., $Z_{ij}^{(k)}$ is sampled from a distribution with mean λ_{ij} which we wish to estimate.

(ii) $O_i^{(k)}$ is the *observed* total amount of traffic leaving zone i during the kth period, i.e.,

$$O_i^{(k)} = \sum_{j=1}^{N} Z_{ij}^{(k)}.$$

(iii) $I_j^{(k)}$ is the *observed* total amount of traffic entering zone j during the kth period, i.e.,

$$I_j^{(k)} = \sum_{i=1}^{N} Z_{ij}^{(k)}.$$

If we could observe $Z_{ij}^{(k)}$ over D periods then the sample mean

$$\bar{Z}_{ij} = \frac{1}{D} \sum_{k=1}^{D} Z_{ij}^{(k)} \tag{1}$$

would be the natural estimator for λ_{ij}. In our setting, however, we do not observe $Z_{ij}^{(k)}$ but only the total incoming and outgoing traffic $I_j^{(k)}$ and $O_i^{(k)}$ respectively. The question therefore arises which quantities should be used in place of the unobserved traffic $Z_{ij}^{(k)}$ in (1)? If the vector $\lambda = (\lambda_{ij})$ of mean traffic demands was known then it would be sensible to replace the unobserved vector of traffic between nodes by its conditional expectation

$$E_\lambda[\mathbf{Z} \mid \sum_{i=1}^{N} Z_{ij}^{(k)} = I_j^{(k)}, \sum_{j=1}^{N} Z_{ij}^{(k)} = O_i^{(k)}]. \tag{2}$$

The computation of this conditional expectation is difficult for general distributions but straight-forward for normally distributed demands as we will see in Section 3.

Alternatively, following Decartes' maxim "*when it is not in our power to determine what is true, we ought to act according to what is most probable*", we may replace the unobserved between-zones traffic by the most likely between-zones traffic subject to consistency with the traffic counts. This assumes that we know the family of density functions $p_{ij}(., \lambda_{ij})$ of the traffic demands between zones i and j in parametric form, depending on the mean λ_{ij} that we wish to estimate, and leads to the optimization problem

$$\begin{aligned} \max_{\mathbf{z}} \quad & p(\mathbf{z}, \lambda) = \prod_{i=1}^{N} \prod_{j=1}^{N} p_{ij}(z_{ij}^{(k)}, \lambda_{ij}) \\ \text{subject to} \quad & \sum_{i=1}^{N} z_{ij}^{(k)} = I_j^{(k)}, \quad j = 1, 2, \ldots, N, \\ & \sum_{j=1}^{N} z_{ij}^{(k)} = O_i^{(k)}, \quad i = 1, 2, \ldots, N, \\ & z_{ij}^{(k)} \geq 0 \quad i, j = 1, 2, \ldots, N. \end{aligned} \tag{3}$$

We will see later in Section 4 if $p_{ij}(\cdot)$ is the Poisson distribution then the optimization problem can be approximated by the entropy maximization problem. The Poisson distribution with mean λ can be approximated by the normal distribution $N(\lambda, \lambda)$ when λ is large [13]. In this case we expect problem (1) with normal demands and problem (3) with Poisson demands to give very similar demand estimation. This is indeed observed in our numerical experiments.

The least squares approach offers a further alternative. Assuming that the distribution of between-zone traffic clusters about the mean, one may take the estimate z to be as close

as possible to the mean λ subject to consistency with the observations. This leads to an optimization problem of the form

$$\begin{array}{ll} \min\limits_{\mathbf{z}} & d(z,\lambda) \\ \text{subject to} & \sum\limits_{i=1}^{N} z_{ij}^{(k)} = I_j^{(k)}, \quad j = 1,2,\ldots,N, \\ & \sum\limits_{j=1}^{N} z_{ij}^{(k)} = O_i^{(k)}, \quad i = 1,2,\ldots,N, \\ & z_{ij}^{(k)} \geq 0 \quad i,j = 1,2,\ldots,N, \end{array} \tag{4}$$

where $d(.,.)$ is a suitable distance function. If $d(.,.)$ is the Euclidean distance between z and λ (4) is an ordinary least square problem with linear constraints. Alternative generalized least square approaches can be found in [2, 3, 8, 19, 21, 22]. Whilst we focus in this paper on the maximum likelihood (maximum entropy) approach and the development of fast algorithms for solving (1) and (3), the algorithm for solving (3) can be easily modified to solve the above least squares problem. We compare the maximum entropy approach with the least square approach in Section 6.

In (1) we assumed the between-zones traffic $Z_{ij}^{(k)}$ to be observed and estimated the mean demands λ_{ij}, whereas in (2), (3) and (4) we assumed mean demands λ_{ij} to be known and calculated surrogates for the unobserved traffic $Z_{ij}^{(k)}$. The two views, although individually relying on unrealistic assumptions, can be naturally combined to an iterative procedure. Starting with initial estimates $\lambda_{ij}^{(0)}$ for mean demands, we may successively update $\lambda_{ij}^{(k-1)}$, e.g., using the well-known exponential smoothing formula [1] with parameter α:

$$\lambda_{ij}^{(k)} = \lambda_{ij}^{(k-1)} + \alpha(z_{ij}^{(k)} - \lambda_{ij}^{(k-1)}), \tag{5}$$

where $z_{ij}^{(k)}$ is obtained through (2), (3) or (4) with the unknown means λ_{ij} replaced by the current estimates $\lambda_{ij}^{(k-1)}$. This estimation method is essentially a forecasting procedure and allows for fluctuations of mean demands over time. It can be extended to account for trends in mean demands λ_{ij}, e.g., by employing a two-parameter exponential smoothing technique. The approach based on (3) is a generalization of the Expectation Maximization (EM) algorithm of [9] which updates $\lambda_{ij}^{(k)}$ by $z_{ij}^{(k)}$, i.e. uses the smoothing parameter $\alpha = 1$ in (5). For an introduction to the EM algorithm, we refer readers to [12]. Our smoothing approach should not be confused with the smoothed EM algorithm of [26] which smoothes out spikes in the reconstruction of a full density function.

2.1 Notational Convention

It will be useful in the sequel to represent the equations

$$\begin{aligned} \sum_{i=1}^{N} Z_{ij}^{(k)} &= I_j^{(k)}, \quad j = 1, \ldots, N \\ \sum_{j=1}^{N} Z_{ij}^{(k)} &= O_i^{(k)}, \quad i = 1, \ldots, N \end{aligned} \tag{6}$$

in a concise matrix form $A_N \mathbf{x} = \mathbf{b}$. To this end we use the following notation:

$$\begin{cases} (x_1, x_2, \ldots, x_{N-1}) & = & (Z_{1,N}, Z_{1,N-1}, \ldots, Z_{1,2}); \\ (x_N, x_{N+1}, \ldots, x_{2(N-1)}) & = & (Z_{2,1}, Z_{2,N}, Z_{2,N-1}, \ldots, Z_{2,3}); \\ (x_{2N-1}, x_{2N}, \ldots, x_{3(N-1)}) & = & (Z_{3,2}, Z_{3,1}, Z_{3,N}, \ldots, Z_{3,4}); \\ & \vdots & \\ (x_{(N-1)^2+1}, x_{(N-1)^2+2}, \ldots, x_{N(N-1)}) & = & (Z_{N,N-1}, Z_{N,N-2}, \ldots, Z_{N,1}). \end{cases} \tag{7}$$

Given this reordering of demands, the $2N \times N(N-1)$ arc-path incidence matrix A_N of the star network with N zones takes the form

$$A_N = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}, \tag{8}$$

where B_1 and B_2 are $N \times N(N-1)$ matrices of the form

$$B_1 = \underbrace{\begin{pmatrix} \mathbf{1}_{N-1} & \mathbf{0}_{N-1} & \ldots & \ldots & \mathbf{0}_{N-1} \\ \mathbf{0}_{N-1} & \mathbf{1}_{N-1} & \mathbf{0}_{N-1} & \ldots & \mathbf{0}_{N-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0}_{N-1} & \mathbf{0}_{N-1} & \ldots & \mathbf{0}_{N-1} & \mathbf{1}_{N-1} \end{pmatrix}}_{N}$$

and

$$B_2 = (\underbrace{\mathbf{Id}_N, \mathbf{Id}_N, \ldots, \mathbf{Id}_N}_{N-1}),$$

respectively. Here $\mathbf{1}_{N-1}$ is the $1 \times (N-1)$ row vector with all entries being 1, $\mathbf{0}_{N-1}$ is the $1 \times (N-1)$ row vector with all entries being 0 and $\mathbf{Id}_N$ is the $N \times N$ identity matrix.

Finally, we store the observed outflows O_i and inflows I_j suitably order in the column vector

$$\mathbf{b} = (O_1, O_2, \ldots, O_N, I_N, I_{N-1}, \ldots, I_1)^t. \tag{9}$$

2.2 Organization of the Paper

The rest of the paper is organized as follows. In Section 3, we explain how (2) can be approximated if the traffic demands are approximately normal. Then we turn our attention to the optimization problems (3) and (4). In Section 4 we analyze the inherent degeneracy in the constraints of the optimization problems and suggest a revised formulation of the objective function of (3) for the special case of Poisson demands that makes the problem amenable to optimization procedures. We will see that the revised formulation is related to entropy maximization, thus linking this popular estimation approach with the Poisson distribution. In Section 5, we present a solution procedure for the dual of the revised formulation of the maximization problem (3) for the case of Poisson demands. This procedure combines Newton's method with the conjugate gradient method. We argue theoretically why this combination is particularly adequate for the present problem. The simulation examples in Section 6 demonstrate the numerical and statistical efficiency of the proposed estimation technique. In Section 7, we outline possible extensions of the technique to take advantage of sparsity in the network structure. Finally, concluding remarks are given in Section 8 to summerize the paper.

3 Normal Demands

In this section we assume that traffic demands are normally distributed with means λ_{ij} and variances σ_{ij}^2. We begin with the assumption that the variances are known; unknown variances pose no additional problem as we will explain later.

The advantage of the normal distribution is that the conditional expectation (2) can be computed straight-forwardly. The following lemma is the basis for this approximation.

Lemma 1 *If the components of a vector X are independent standard normal variables then the distribution of X conditional on $C\mathbf{X} = \mathbf{c}$ is normal and the conditional expectation $E[\mathbf{X} \mid C\mathbf{X} = \mathbf{c}]$ is the smallest element of the hyperplane $C\mathbf{X} = \mathbf{c}$ in the Euclidean norm.*

Proof: We assume without loss of generality that C has full row rank. Notice that $\mathbf{X}$

and $\mathbf{Y} = C\mathbf{X}$ are both Gaussians with vanishing means and covariance matrix

$$\operatorname{cov}\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathrm{Id} & C^t \\ C & CC^t \end{bmatrix}.$$

The lemma is thus a direct consequence of the fact that the distribution of the Gaussian $\mathbf{X}$ conditional on the Gaussian $\mathbf{Y}$ is again Gaussian with

$$E[\mathbf{X}|\mathbf{Y}] = C^t(CC^t)^{-1}\mathbf{Y} \quad \text{and} \quad \operatorname{cov}(\mathbf{X}|\mathbf{Y}) = \mathrm{Id} - C^t(CC^t)^{-1}C,$$

cf. e.g. Lemma 11.1 of [29]. □

We can apply the latter result to the standardizations $X_{ij} = \frac{Z_{ij}-\lambda_{ij}}{\sigma_{ij}}$ of the variables Z_{ij} to obtain the conditional expectation

$$E_\lambda[\mathbf{Z} \mid A_N\mathbf{Z} = \mathbf{b}] = \lambda + \operatorname{diag}(\sigma)E[\mathbf{X} \mid A_N\operatorname{diag}(\sigma)\mathbf{X} = -A_N\lambda + \mathbf{b}], \tag{10}$$

where $\operatorname{diag}(\sigma)$ is the diagonal matrix with diagonal entries σ_{ij} and $\mathbf{b}$ is given in (9). The optimality conditions for the least distance problem in the lemma are

$$\begin{aligned} \mathbf{x} - \operatorname{diag}(\sigma)A_N^t\mathbf{y} &= \mathbf{0} \\ A_N\operatorname{diag}(\sigma)\mathbf{x} &= -A_N\lambda + \mathbf{b}. \end{aligned}$$

Elimination of $\mathbf{x}$ yields a square system of $2N$ equations

$$A_N\operatorname{diag}(\sigma^2)A_N^t\mathbf{y} = -A_N\lambda + \mathbf{b}, \tag{11}$$

where σ^2 is obtained from σ by squaring all components. Given the solution $\mathbf{y}$, (10) yields

$$E_\lambda[\mathbf{Z} \mid A_N\mathbf{Z} = \mathbf{b}] = \lambda + \operatorname{diag}(\sigma)x = \lambda + \operatorname{diag}(\sigma^2)A_N^t\mathbf{y}. \tag{12}$$

The main computational work is the solution of the linear system (11). We will argue later that the conjugate gradient method is particularly appropriate for such systems, see Proposition 2 below. Given the old estimation λ and the new observed traffic flows $\mathbf{b}$, one may update the mean demands between zones by the smoothing scheme (5) and (12) via solving (11).

It is straight-forward to extend the procedure based on (5) to include an estimation of σ_{ij}^2 if the variance is unknown. Indeed, if the between zone traffics $Z_{ij}^{(k)}$ were observed then it would be appropriate to replace the unknown variance σ_{ij}^2 by the sample variance

$$S_{ij} = \frac{1}{D-1}\sum_{i=1}^{D}(Z_{ij}^{(k)} - \bar{Z}_{ij})^2.$$

The update formula (5) can therefore be extended to include a variance estimate

$$\sigma_{ij}^{(k)} = \sigma_{ij}^{(k-1)} + \alpha'(z_{ij}^{(k)} - \lambda_{ij}^{(k)})^2,$$

where $\lambda_{ij}^{(k)}$ is obtained from (5) and α' is a smoothing parameter.

4 Poisson Demands

The computation of the conditional expectation (2) does not seem to be as straight-forward for Poisson demands. We therefore resort to replacing the unobserved between zones traffic by the solution of the optimization problem (3). For the Poisson distribution the problem turns into

$$\begin{array}{ll} \max\limits_{\mathbf{x}} & \prod\limits_{i=1}^{n} \frac{\lambda_i^{x_i}}{x_i!} e^{-\lambda_i} \\ \text{subject to} & A_N\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}, \end{array} \tag{13}$$

where A_N is the $2N \times N(N-1)$ in (8), $n = N(N-1)$. By passing to the logarithm of the objective function we obtain the equivalent optimization problem

$$\begin{array}{ll} \max\limits_{\mathbf{x}} & \sum\limits_{i=1}^{n}(x_i \log \lambda_i - \log x_i!) \\ \text{subject to} & A_N\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{array} \tag{14}$$

Notice that the constraints $A_N\mathbf{x} = \mathbf{b}$ are degenerate in the sense that the matrix A_N is rank deficient (Rank$(A_N A_N^t)$= $2N-1$). Moreover, the ratio of the largest and the second smallest eigenvalue of the matrix $A_N A_N^t$ is bounded above independently of the matrix size N. The following lemma specifies these observations, which will be useful in the sequel.

Lemma 2 *The eigenvalues of the matrix* $(A_N A_N^t)$ *are* 0 *and* $2(N-1)$*, both with multiplicity* 1*, as well as* $N-2$ *and* N*, both with multiplicity* $N-1$*.*
Moreover $\sigma_{2N}(A_N A_N^t)/\sigma_2(A_N A_N^t) \leq 4$ *for* $N \geq 3$*. Here* $\sigma_i(H)$ *is the* i*th largest eigenvalue of a positive semi-definite matrix* H*.*

Proof: We can rewrite

$$A_N A_N^t = (N-1)\mathrm{Id}_{2N} - \underbrace{\begin{pmatrix} 0 & 0 & \dots & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & \dots & 0 & 0 \end{pmatrix}}_{=S} + \underbrace{\begin{pmatrix} 0 & \dots & 0 & 1 & \dots & 1 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & 1 & \dots & 1 \\ 1 & \dots & 1 & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ \underbrace{1 \quad \dots \quad 1}_{N} & & & \underbrace{0 \quad \dots \quad 0}_{N} & & \end{pmatrix}}_{=T}.$$

We note that S and T are symmetric matrices so they are diagonalizable. Moreover S and T commute, i.e. $(ST = TS)$, and therefore they are simultaneously diagonalizable, see Horn and Johnson [17]. Thus there exists an invertible P such that

$$P^{-1}SP = D_1 \quad \text{and} \quad P^{-1}TP = D_2$$

where D_1 and D_2 are diagonal matrices containing the eigenvalues of S and T respectively. Therefore we have

$$P^{-1}A_N A_N^t P = (N-1)\mathrm{Id} - D_1 + D_2.$$

The eigenvalues of S are 1 (multiplicity N) and -1 (multiplicity N). We note that $T\mathbf{v}_1 = -N\mathbf{v}_1$ and $T\mathbf{v}_2 = N\mathbf{v}_2$ where

$$\mathbf{v}_1 = \frac{1}{\sqrt{2N}}(\underbrace{1,1,\dots,1}_{N},\underbrace{-1,-1,\dots,-1}_{N})^t \quad \text{and} \quad \mathbf{v}_2 = \frac{1}{\sqrt{2N}}(1,1,\dots,1)^t.$$

Thus the eigenvalues of the rank 2 matrix T are N, $-N$ and 0, the latter with multiplicity $2N-2$. Direct calculation shows that

$$A_N A_N^t \mathbf{v}_1 = \mathbf{0} \quad \text{and} \quad (A_N A_N^t - 2(N-1)I)\mathbf{v}_2 = \mathbf{0},$$

hence 0 and $2(N-1)$ are two eigenvalues of $A_N A_N^t$. We conclude that the eigenvalues of $A_N A_N^t$ are 0, $N-2$ (multiplicity $N-1$), N (multiplicity $N-1$) and $2(N-1)$. It follows

that $\text{Rank}(A_N A_N^t)$ is $2N-1$. Finally, $\sigma_{2N}(A_N A_N^t) = 2(N-1)$ and $\sigma_2(A_N A_N^t) = (N-2)$. Hence for $N \geq 3$ we have

$$\frac{\sigma_{2N}(A_N A_N^t)}{\sigma_2(A_N A_N^t)} = \frac{2(N-1)}{N-2} \leq 4.$$

□

4.1 Stirling's Approximation and Entropy Maximization

To make the optimization problem (14) amenable to numerical optimization procedures we need to find a suitable approximation for the factorial function in the objective. The obvious candidate is Stirling's approximation

$$M! \approx \sqrt{2\pi} e^{-M} M^{M+\frac{1}{2}}$$

which leads to

$$\log M! \approx M \log M - M + O(\log M), \tag{15}$$

where $O(\log M) = (\log M + \log 2\pi)/2$. It is well-known that Stirling's approximation is surprisingly accurate. Even if we neglect the $O(\log M)$ term in (15) the relative error for $M = 100$ is only

$$\frac{\log 100! - (100 \log 100 - 100)}{\log 100!} \approx 0.9\%.$$

If we use Stirling's approximation to replace the factorial function then the maximization problem (14) becomes

$$\begin{aligned} \max_{\mathbf{x}} \quad & \sum_{i=1}^{n} -x_i \left(\log \left(\frac{x_i}{\lambda_i} \right) - 1 \right) \\ \text{subject to} \quad & A_N \mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned} \tag{16}$$

Estimating demand through the optimization problem (16) is called Maximum Entropy Estimation (MEE). This approach was first suggested by Wilson [30, 31] and has been used by Jornsten and Nguyen [18] and Van Zuylen and Willumsen [28] for traffic demand estimation, see also [32, 33]. We remark that the MEE approach in [28] is very similar to the one described in this paper except that we don't need to assume a constant proportional assignment of links to trips. Therefore we don't need to involve an assignment method in our algorithm and we can update the origin-destination matrix (demand

means) by using the old origin-destination matrix only.

The efficient method for (16) that we will develop in the next section is relevant for maximum entropy estimation which is typically applied without explicit distributional assumptions on traffic demands. Notice, though that our derivation of the MEE optimization problem (16) shows that, at least for sufficiently large means, MEE is equivalent to maximizing the likelihood of independent Poisson processes.

5 Solving the Maximization Problem

The optimization problem (16) has $N(N-1)$ variables and $2N$ equality constraints. There exists a unique optimal solution in the interior of the nonnegative orthant as the objective function is strictly concave and tends to $-\infty$ as x tends to ∞ or to the boundary of the nonnegative orthant. We suggest to solve the dual rather than the primal problem because the former has only $2N$ variables as opposed to the $N(N-1)$ variables of the primal.

5.1 The Dual Problem

We may discard the non-negativity constraints of problem (16) because the objective function approaches $-\infty$ as x tends to the boundary of the nonnegative orthant. The dual problem of (16) is therefore of the type

$$\min_{\mathbf{y}} \max_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}), \tag{17}$$

where $\mathbf{y}$ is the multiplier and $L(.,.)$ is the Lagrangian function

$$L(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} (x_i \log \lambda_i - x_i \log x_i + x_i) + \mathbf{y}^t(\mathbf{b} - A\mathbf{x}). \tag{18}$$

Here and in the remainder of the section we write A_N as A since N is fixed throughout. The optimal solution $\mathbf{x}^*(\mathbf{y})$ of the inner maximization problem of (17) solves the equations

$$\nabla_{x_i} L(\mathbf{x}, \mathbf{y}) = \log \lambda_i - \log x_i - \sum_{j=1}^{m} \mathbf{y}^t A_{.i} = 0, \quad i = 1, 2, \ldots, n$$

and is thus of the form

$$\mathbf{x}_i^*(\mathbf{y}) = \lambda_i e^{-\mathbf{y}^t A_{.i}}, \quad i = 1, 2, \ldots, n, \tag{19}$$

where $A_{.i}$ is the ith column of the matrix A. After substituting $\mathbf{x}^*(\mathbf{y})$ back into (17) the dual problem simplifies to

$$\min_{\mathbf{y}}\{\sum_{i=1}^{n}\lambda_i e^{-\mathbf{y}^t A_{.i}} + \mathbf{y}^t\mathbf{b}\}. \tag{20}$$

The solution of the primal problem (16) is obtained from the solution of the dual problem (20) through (19). We have thus transformed a constrained maximization problem with $N(N-1)$ variables into an unconstrained minimization problem with $2N$ variables.

5.2 Newton's Method for the Dual Problem

A variety of numerical methods are available for solving (20), see e.g. Nocedal and Wright [23]. In the following we will explain how Newton's method in conjunction with the conjugate gradient method can be used and give a theoretical justification for the efficiency of the conjugate gradient method in this setting. To this end we denote by

$$f(\mathbf{y}) = \sum_{i=1}^{n}\lambda_i e^{-\mathbf{y}^t A_{.i}} + \mathbf{y}^t\mathbf{b} \tag{21}$$

the function to be minimized. The gradient and Hessian of f are of the form

$$\begin{aligned} \nabla f(\mathbf{y}) &= -A\mathbf{x}^*(\mathbf{y}) + \mathbf{b} \\ \nabla^2 f(\mathbf{y}) &= A\cdot \mathrm{diag}(\mathbf{x}^*(\mathbf{y}))\cdot A^t, \end{aligned} \tag{22}$$

where $\mathbf{x}_i^*(\mathbf{y})$ is as defined in (19) and $\mathrm{diag}(\mathbf{x}^*(\mathbf{y}))$ is the diagonal matrix with diagonal entries $\mathbf{x}_i^*(\mathbf{y})$.

Consider the vector

$$\mathbf{v}_1 = \frac{1}{\sqrt{2N}}(\underbrace{1,1,\ldots,1}_{N},\underbrace{-1,-1,\ldots,-1}_{N})^t.$$

It is obvious that

$$\mathbf{v}_1^t A = 0 \tag{23}$$

$$\mathbf{v}_1^t \mathbf{b} = 0, \tag{24}$$

where the latter equation is a consequence of the balance equation $\sum_i^N I_i = \sum_i^N O_i$. Hence $f(\mathbf{y}+\mathbf{z}) = f(\mathbf{y})$ for every $\mathbf{y}\in\mathbf{v}_1^\perp$ and every $\mathbf{z}$ in the linear span of $\mathbf{v}_1$ and it therefore suffices to find a minimizer of f in $\mathbf{v}_1^\perp$.

Lemma 3 *The null space of $\nabla^2 f(\mathbf{y})$ is the linear span of $\mathbf{v}_1$ and $\nabla^2 f(\mathbf{y})$ is positive definite on $\mathbf{v}_1^\perp$.*

Proof: By our assumption (A3) in Section 2, all the demands λ_i are positive. Therefore the Hessian

$$\nabla^2 f(\mathbf{y}) = A \cdot \mathrm{diag}(\mathbf{x}^*(\mathbf{y})) \cdot A^t$$

is obviously positive semi-definite because $\mathbf{x}_i^*(\mathbf{y}) = \lambda_i e^{-\mathbf{y}^t A_{\cdot i}} \geq 0$. If $\mathbf{z}^t \nabla^2 f(\mathbf{y}) \mathbf{z} = 0$ then

$$0 = \mathbf{z}^t \nabla^2 f(\mathbf{y}) \mathbf{z} \geq \min\{\mathbf{x}_i^*(\mathbf{y})\} \mathbf{z}^t A A^t \mathbf{z} \geq 0$$

and therefore $\mathbf{z}^t A A^t \mathbf{z} = 0$. Lemma 2 thus implies that $\mathbf{z} = \alpha \mathbf{v}_1$. □

We conclude that f is strictly convex on $\mathbf{v}_1^\perp$ and will now proceed to applying Newton's method to minimizing f on the latter subspace. Newton's method produces a sequence of points $\mathbf{y}_k$ according to the iteration $\mathbf{y}_k = \mathbf{y}_{k-1} + \mathbf{p}_k$, where the Newton step $\mathbf{p}_k$ is the solution of the equation

$$\nabla^2 f(\mathbf{y}_k) \mathbf{p}_k = -\nabla f(\mathbf{y}_k). \tag{25}$$

Notice that, by the above lemma, $\nabla^2 f(\mathbf{y}_k)$ is a one-to-one mapping of $\mathbf{v}_1^\perp$ onto itself. Hence, since $\nabla f(\mathbf{y}_k) \in \mathbf{v}_1^\perp$ by (22), (23) and (24), there exists a unique $\mathbf{p}_k \in \mathbf{v}_1^\perp$ such that (25) holds. Newton's method for minimizing f on $\mathbf{v}_1^\perp$ is thus well defined. If we start with $\mathbf{y}_0 \in \mathbf{v}_1^\perp$ the Newton sequence will remain in this subspace. Moreover, since the Hessian $\nabla^2 f(\mathbf{y})$ is non-singular on this subspace, the Newton sequence converges locally at a quadratic rate. To enforce global convergence one may wish to resort to line search or trust region techniques, cf. e.g. [7, 10]. We did not find this necessary in our computational experiments for which we used the following standard procedure:

Newton's Method (see Nocedal and Wright [23]):
Choose starting point $\mathbf{y}_0 \in \mathbf{v}_1^\perp$
$k = 1;$
while $||\nabla f(\mathbf{y}_k)||_2 > tolerance$
find $\mathbf{p}_k \in \mathbf{v}_1^\perp$ with $\nabla^2 f(\mathbf{y}_{k-1}) \mathbf{p}_k = -\nabla f(\mathbf{y}_{k-1})$;
set $\mathbf{y}_k = \mathbf{y}_{k-1} + \mathbf{p}_k$;
$k = k + 1;$

end.

A possible starting point is e.g. $\mathbf{y}_0 = \frac{1}{\sqrt{2N}}(1, 1, \ldots, 1)^t$. Since the estimation procedure outlined in the foregoing section needs the solutions of a sequence of such optimization problems with a constant $\mathbf{v}_1$ it seems sensible to use the optimal solution of the $(k-1)$th problem as starting point for the kth problem in the sequence. The main computation in an iteration of Newton's method is the solution of the linear system $\nabla^2 f(\mathbf{y}_{k-1})\mathbf{p}_k = -\nabla f(\mathbf{y}_{k-1})$ in $\mathbf{v}_1^\perp$. The positive definiteness of $\nabla^2 f(\mathbf{y})$ on $\mathbf{v}_1^\perp$ suggests the use of the conjugate gradient method, which is an iterative method. By making use of the sparse structure of the matrix A, we don't need to store the iteration matrix $\nabla^2 f(\mathbf{y}_{k-1})$, which is an advantage over a direct method. Moreover, the conjugate gradient method converges very fast because the iteration matrix has a favorable spectral property which we will explain in Propositions 1 and 2. The numerical efficiency is also apparent in our computational examples in Section 6.

Conjugate Gradient Method (see Golub and van Loan [16]):

Choose starting point $\mathbf{y}_0 \in \mathbf{v}_1^\perp$

$\mathbf{r}_0 = \mathbf{b} - H\mathbf{y}_0$;

$k = 1$;

$\mathbf{p}_1 = \mathbf{r}_0$;

$\alpha_1 = \mathbf{r}_0^t\mathbf{r}_0/\mathbf{p}_1^t H\mathbf{p}_1$;

$\mathbf{y}_1 = \mathbf{y}_0 + \alpha_1\mathbf{p}_1$;

$\mathbf{r}_1 = \mathbf{r}_0 - \alpha_1 H\mathbf{p}_1$;

while $||\mathbf{r}_k||_2 > tolerance$,

  $k = k + 1$;

  $\beta_k = \mathbf{r}_{k-1}^t\mathbf{r}_{k-1}/\mathbf{r}_{k-2}^t\mathbf{r}_{k-2}$;

  $\mathbf{p}_k = \mathbf{r}_{k-1} + \beta_k\mathbf{p}_{k-1}$;

  $\alpha_k = \mathbf{r}_{k-1}^t\mathbf{r}_{k-1}/\mathbf{p}_k^t H\mathbf{p}_k$;

  $\mathbf{y}_k = \mathbf{y}_{k-1} + \alpha_k\mathbf{p}_k$;

  $\mathbf{r}_k = \mathbf{r}_{k-1} - \alpha_k H\mathbf{p}_k$;

end;

$\mathbf{y} = \mathbf{y}_k$.

Lemma 4 *(Golub and van Loan [16, p. 525]) When the conjugate gradient method is applied to solve $H\mathbf{x} = \mathbf{b}$ where H a symmetric positive definite square matrix, then we have*

$$\frac{||\mathbf{x}_i - \mathbf{x}||}{||\mathbf{x}_0 - \mathbf{x}||} \leq 2\left(\frac{\sqrt{\tau}-1}{\sqrt{\tau}+1}\right)^i.$$

Here $\mathbf{x}_i$ is the approximate solution obtained at the ith iteration and $\tau = \sigma_{\max}(H)/\sigma_{\min}(H)$, the condition number of H.

A possible starting point for the conjugate gradient method is again

$$\mathbf{y}_0 = \frac{1}{\sqrt{2N}}(1, 1, \ldots, 1)^t \in \mathbf{v}_1^{\perp}.$$

Starting with $\mathbf{y}_0 \in \mathbf{v}_1^{\perp}$ the conjugate gradient method converges in at most $2N-1$ steps. However, convergence can occur earlier depending on the chosen tolerance and the condition number of $\nabla^2 f(\mathbf{y})$ on $\mathbf{v}_1^{\perp}$ by Lemma 4. Since the Hessian matrix is symmetric positive definite on $\mathbf{v}_1^{\perp}$, the convergence rate of the conjugate gradient method depends on the *effective condition number*

$$\tau_{effective} = \frac{\sigma_{2N}(\nabla^2 f(\mathbf{y}))}{\sigma_2(\nabla^2 f(\mathbf{y}))},$$

where $\sigma_i(\nabla^2 f(\mathbf{y}))$ is the ith smallest eigenvalues of $\nabla^2 f(\mathbf{y})$. The following proposition gives a bound on the effective condition number in terms of bounds on traffic demands but independent of the number of zones N.

Proposition 1 *If L is the smallest and U the largest mean traffic demand λ_i then*

$$(N-2)Le^{-2||\mathbf{y}||_\infty} \leq \sigma_2(\nabla^2 f(\mathbf{y})) \leq \sigma_{2N}(\nabla^2 f(\mathbf{y})) \leq 2(N-1)Ue^{2||\mathbf{y}||_\infty}.$$

Hence the effective condition number of the matrix $\nabla^2 f(\mathbf{y})$ is bounded above by $3Ue^{4||\mathbf{y}||_\infty}/L$.

Proof: In view of the definition of $\mathbf{x}^*(\mathbf{y})$ in equation (19) we have

$$Le^{-2||\mathbf{y}||_\infty} \leq \mathbf{x}_i^*(\mathbf{y}) \leq Ue^{2||\mathbf{y}||_\infty} \quad , i = 1, 2, \ldots, 2N. \tag{26}$$

Hence for all $\mathbf{z} \in \mathbf{v}_1^{\perp}$ we obtain

$$(Le^{-2||\mathbf{y}||_\infty})\frac{\mathbf{z}^t A A^t \mathbf{z}}{\mathbf{z}^t \mathbf{z}} \leq \frac{\mathbf{z}^t A \cdot \mathrm{diag}(\mathbf{x}^*(\mathbf{y})) \cdot A^t \mathbf{z}}{\mathbf{z}^t \mathbf{z}} \leq (Ue^{2||\mathbf{y}||_\infty})\frac{\mathbf{z}^t A A^t \mathbf{z}}{\mathbf{z}^t \mathbf{z}}.$$

Thus Lemma 2 yields

$$Le^{-2||\mathbf{y}||_\infty}(N-2) \leq \sigma_2(\nabla^2 f(\mathbf{y})) \leq \sigma_{2N}(\nabla^2 f(\mathbf{y})) \leq 2Ue^{2||\mathbf{y}||_\infty}(N-1).$$

We conclude that the effective condition number of the Hessian matrix satisfies

$$\frac{\sigma_{2N}(\nabla^2 f(\mathbf{y}))}{\sigma_2(\nabla^2 f(\mathbf{y}))} \leq \frac{2Ue^{2||\mathbf{y}||_\infty}(N-1)}{Le^{-2||\mathbf{y}||_\infty}(N-2)} \leq \frac{3Ue^{4||\mathbf{y}||_\infty}}{L}.$$

□

We conclude this section with three remarks.

Remark 1: Note that the effective condition number in Proposition 1 depends on $||\mathbf{y}||_\infty$ and recall that in the iterations of Newton's method we have

$$\mathbf{y}_k = \mathbf{y}_{k-1} + \mathbf{p}_k \quad \text{and} \quad \nabla^2 f(\mathbf{y}_{k-1})\mathbf{p}_k = -\nabla f(\mathbf{y}_{k-1}) = \mathbf{b} - A_N \mathbf{x}^*(\mathbf{y}_{k-1}).$$

We write

$$\mathbf{p}_k = \nabla^2 f(\mathbf{y}_{k-1})^{-1}(\mathbf{b} - A_N \mathbf{x}^*(\mathbf{y}_{k-1}))$$

where $\nabla^2 f(\mathbf{y}_{k-1})^{-1}$ is valid for $\nabla f(\mathbf{y}_{k-1}) \in \mathbf{v}_1^\perp$. Then

$$\begin{aligned} ||\mathbf{p}_k||_2 &\leq \frac{e^{2||\mathbf{y}_{k-1}||_\infty}}{L(N-2)}\left(||\mathbf{b}||_2 + 2(N-1)Ue^{2||\mathbf{y}_{k-1}||_\infty}\right) \\ &\leq \frac{e^{2||\mathbf{y}_{k-1}||_\infty}}{L}\left(\frac{||\mathbf{b}||_2}{N-2} + 4Ue^{2||\mathbf{y}_{k-1}||_\infty}\right) \end{aligned}$$

because

$$||A_N \mathbf{x}^*(\mathbf{y}_{k-1})||_2 \leq 2(N-1)Ue^{2||\mathbf{y}_{k-1}||_\infty}.$$

Since $||\mathbf{p}_k||_\infty \leq ||\mathbf{p}_k||_2$, we have

$$||\mathbf{y}_k||_\infty \leq ||\mathbf{y}_{k-1}||_\infty + \frac{e^{2||\mathbf{y}_{k-1}||_\infty}}{L}\left(\frac{||\mathbf{b}||_2}{N-2} + 4Ue^{2||\mathbf{y}_{k-1}||_\infty}\right). \tag{27}$$

Recall that $\mathbf{b}$, defined in (9), is a random vector in the smoothing scheme. In Lemma 5 we will show that $E(||\mathbf{b}||_2) \leq 2\sqrt{\Lambda}$ where Λ is the expected total traffic in the network. Thus if the expected total traffic in the network grows at a rate less than $O(N^2)$ (in fact, $\Lambda \leq \max_{i,j}\{\lambda_{ij}\} \cdot N^2$) when the number of zones N increases, then $E(||\mathbf{b}||_2/(N-2)) \leq \gamma$, for certain $\gamma > 0$ independent of N. In this case, the expectation of the right-hand-side

of (27) is independent of N. In this case, we expect that the conjugate gradient method will converge fast independently of the network size, when applied to solving system (25). This effect can be observed in the numerical examples in Section 6 below.

Lemma 5 *If traffic demands between zones are independent and follow Poisson distributions with means $\lambda_{ij}(i,j=1,2,\ldots,N)$ and $\lambda_{ii}=0$ for $i=1,2,\ldots,N$, then*

$$E(||\mathbf{b}||_2) \leq 2\sqrt{\Lambda},$$

where

$$\Lambda = \sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_{ij}$$

is the expected total traffic in the network.

Proof: Note that

$$E(||\mathbf{b}||_2)^2 \leq E(||\mathbf{b}||_2^2) = E(\sum_{i=1}^{N}(O_i^2+I_i^2)),$$

where O_i and I_i is the observed outflow and inflow of zone i. Note also that

$$E(O_i) = \sum_{j=1}^{N}\lambda_{ij}, \quad E(I_i) = \sum_{j=1}^{N}\lambda_{ji}, \quad Var(O_i) = \sum_{j=1}^{N}\lambda_{ij}, \quad Var(I_i) = \sum_{j=1}^{N}\lambda_{ji},$$

because the demands between zones are assumed to be independent and to follow Poisson distributions with means λ_{ij} and variances λ_{ij}. Thus, using of the well-known formula $E(x^2) = Var(x) + E(x)^2$, we arrive at

$$E(\sum_{i=1}^{N}(O_i^2+I_i^2)) = \sum_{i=1}^{N}\Big(E(O_i^2)+E(I_i^2)\Big) = 4\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_{ij} = 4\Lambda,$$

which proves the result. □

Remark 2: Instead of using the proposed subspace method, one may suggest to remove the degeneracy in the system of linear equations directly, e.g. by deleting the first row of the matrix A. One can then apply Newton's method since the new matrix $\tilde{A}\tilde{A}^t$ is symmetric and positive definite. The drawback of this approach is that the spectral properties of $\tilde{A}\tilde{A}^t$ are not as favorable for the conjugate gradient method as those of AA^t. The numerical results in the following section indicate that the alternative approach is indeed less efficient.

Remark 3: Finally, we remark that the conjugate gradient method is equally suited to solve (11). Indeed, the following analogy to Proposition 1 holds and can be proved in the same way.

Proposition 2 *If L is the smallest and U is the largest mean traffic demand λ_i then*

$$(N-2)L \leq \sigma_2(A \cdot \mathrm{diag}(\lambda) \cdot A^t) \leq \sigma_{2N}(A \cdot \mathrm{diag}(\lambda) \cdot A^t) \leq 2(N-1)U$$

Hence the effective condition number of the matrix $A \cdot \mathrm{diag}(\lambda) \cdot A^t$ is bounded above by $3U/L$, independent of N.

6 Numerical Experiments

We will next report on some numerical tests which demonstrate the fast convergence of the proposed optimization algorithm as well as the statistical efficiency of the suggested estimation method. In all tests we used as the stopping criterion for Newton's method

$$||\nabla f(\mathbf{y}_k)||_2 < 10^{-7}$$

whereas the conjugate gradient method was stopped if

$$||\mathbf{r}_k||_2 < 10^{-14}.$$

For the test results reported below we used star networks of varying size with two arbitrarily chosen patterns of the initial traffic demands λ_{ij} and the traffic counts b_k are given as follows:

(a) $\lambda_{ij} = 10000 + 1000(-1)^{i+j}$ and

$$\begin{cases} b_i & = & \sum_j(\lambda_{ij} + (-1)^{i+j}100(i+j)), i = 1, 2, \ldots, N \\ b_{2N-i+1} & = & \sum_j(\lambda_{ji} + (-1)^{i+j}100(i+j)), i = 1, 2, \ldots, N \end{cases}$$

(b) $\lambda_{ij} = 10000 + (-1)^{i+j}1000(i+j)(i+j+1)^{-1}$ and

$$\begin{cases} b_i & = & \sum_j(\lambda_{ij} + (-1)^{i+j}1000(i+j)), i = 1, 2, \ldots, N \\ b_{2N-i+1} & = & \sum_j(\lambda_{ji} + (-1)^{i+j}1000(i+j)), i = 1, 2, \ldots, N \end{cases}$$

6.1 Numerical Efficiency of the Optimization Procedure

Tables 1 and 2 contain numerical efficiency performance measures for the two examples. Here, N is the number of zones in the star networks, $n = N(N-1)$ is the number of parameters which are to be estimated, $W_i (i = 1, 2)$ is the number of iterations of Newton's method, $T_i (i = 1, 2)$ is the total number of iterations of the conjugate gradient method and $C_i (i = 1, 2)$ is the average number of iterations of the conjugate gradient method, $C_i = T_i/W_i$. Here $i = 1$ represents our proposed method and $i = 2$ represents the method where the incidence matrix is regularized by deleting the first equation in $A_N\mathbf{x} = \mathbf{b}$ as indicated at the end of the last section.

Both, the number of iterations W_1 for Newton's method and the average number of iterations for the conjugate gradient method are moderate and indeed roughly constant for the problem sizes and classes considered here. The computational costs of the algorithms are mainly due to matrix-vector multiplications of the form $(ADA^t)\mathbf{x}$ which take $O(N(N-1))$ operations. The comparison of the two methods shows that, as expected, the conjugate gradient method is less efficient for the second approach based on a straightforward regularization of the constraints. We have run several other examples with similar results.

6.2 Successive Updates of Demand Estimates

We tested the statistical efficiency of the estimation procedure as explained in the introductory section:

Choose initial demand estimates $\lambda_i^{(0)}$
For $k = 1 : D$

> Generate unobserved traffic $\mathbf{z}^{(k)}$ between zones, either by solving (16) or through (12), based on observed traffic counts $\mathbf{b}^{(k)}$ and $\lambda = \lambda^{(k-1)}$.
>
> Update λ by the exponential smoothing formula $\lambda_i^{(k)} = \lambda_i^{(k-1)} + \alpha(\mathbf{z}_i^{(k)} - \lambda_i^{(k-1)})$

end.

We recall that for large mean demands eg. ($\lambda_i > 100$), the Poisson distribution is well approximated by the normal distribution $N(\lambda_i, \lambda_i)$. In this case the procedure based on (12) is preferable because the computation only involves solving one linear system (11) in each iteration and we will only report the numerical results generated by this method. As a measure of statistical performance we used the absolute relative error:

$$E_{i,D} = \frac{|\lambda_i^{(D)} - \lambda_i|}{\lambda_i}.$$

Here, D is the total number of simulation runs (the length of the period). Tables 3-5 present simulation results for N-zone star networks with $N = 3, 6, 12, 24$ and 48 and different smoothing factors α including the popular $\alpha = 0.1$, the case of EM algorithm $\alpha = 1$ and the optimal smoothing factor (correct to two decimal places). The demand from zone i to zone j ($i \neq j$) is assumed to follow a Poisson distribution with mean

$$\lambda_{ij} = 30000 + \frac{1000i(-1)^{i-j}}{i+j}. \tag{28}$$

The number of observation periods was chosen to be $D = 30$. Tables 3-5 report the average relative errors $\frac{1}{n}\sum_{i=1}^{n} E_{i,30}$ for the following three cases:

Case 1. Initial mean demands $\lambda_i^{(0)}$ are the correct demands.

Case 2. Initial mean demands underestimate the correct demands by 3000.

Case 3. Initial mean demands overestimate the correct demands by 3000.

The average relative errors are below 0.5 % for all tested network sizes and initial estimates; only a slight dependence of the performance on the initial estimates was observed. The rate at which it decreases depended again somewhat on the quality of the initial estimate. We first observe that in Case 1, when the initial demands are correct, the smaller the smoothing factor the better the result. In fact, the optimal smoothing factor in this case is zero. In Case 2 and 3, the initial demands deviate substantially from the true demands. In these cases, the numerical results suggest that any smoothing factor between 0.1 and 1 gives a reasonably good result. Although we have no recipe for choosing the optimal smoothing factor, the numerical results seem to indicate that $\alpha = 0.1$ is a sensible choice. In Figure 2, we present the average absolute errors of the smoothing scheme when

$\alpha = 0.1$ and $N = 48$ in $K = 30$ smoothing steps. For Case 2 and 3, although the initial errors are large, the smoothing process cuts down the errors to less than 1% in 30 iterations.

Finally we also compare the MEE approach based on (3) with (4) based on the L_2-norm, the least square approach (which in fact is a quadratic programming problem) by assuming Erlangian (l phases) traffic demands with mean given in (28). In Figure 3, we present the average absolute errors of the smoothing scheme when $\alpha = 0.1$ and $N = 48$ after 30 smoothing steps for different values of l. We only present Case 3 because similar patterns are observed for the other cases. For small values of $l(l \leq 2)$, the L_2-norm approach is better. For larger values of l, the MEE approach is better and the average errors are decreasing.

7 Exploitation of Specific Network Structure

In this final section, we briefly illustrate how the efficiency of the proposed estimation methods can be enhanced by exploiting specific network structures. We focus on the case of a two-star network as depicted in Figure 4.

For the first star network A with hub H, the zones $1, 2, \ldots, N$ are connected to H and for the second star network A' with hub H' the zones $1', 2', \ldots, N'$ are connected to H'. To travel from zone i of network A to zone j' of network A', one has to pass through both hubs H and H' and vice versa. The two-star network can be thought of as a single-star network if the hubs H and H' are treated as a single hub. This results in a star network with $N + N'$ zones. It is more efficient, however, to treat the two networks separately initially and then combine the results. In this approach, the computational cost can be further reduced by using parallel computing techniques. For the star network A, one may regard the star network A' as the $(N+1)$th zone of the star network A, represented by hub H'. Similarly for star network A', one may regard the star network A as the $(N' + 1)$th zone of the star network A', represented by hub H. The traffic demand among the zones in each of the networks can be estimated by our proposed method. Moreover, at the same time one can also estimate the mean demand from a zone to the other network i.e.

$$\{Z_{1H'}, \ldots, Z_{NH'}, Z_{1'H}, \ldots, Z_{N'H}\}$$

and also the demand moving into a zone from the other network i.e.

$$\{Z_{H'1}, \ldots, Z_{H'N}, Z_{H1'}, \ldots, Z_{HN'}\}.$$

Here $Z_{iH'}$ is the estimated mean demand from zone i of network A to network A', $Z_{j'H}$ is the estimated mean demand from zone j' of network A' to network A, $Z_{H'i}$ is the estimated mean demand from network A' to zone i of network A and $Z_{Hj'}$ is the estimated mean demand from network A to zone j' of network A'. The demand from zone i in network A to zone j' in network A' can then be estimated by

$$Z_{iH'} \times \frac{Z_{Hj'}}{Z_{H1'} + Z_{H2'} + \ldots + Z_{HN'}}.$$

Similarly the demand from zone j' in network A' to zone i in network A can be estimated by

$$Z_{j'H} \times \frac{Z_{H'i}}{Z_{H'1} + Z_{H'2} + \ldots + Z_{H'N}}.$$

This aggregation method can be extended to networks that are composed of a series of separated star networks with the hub being linked by a tree so that between any two hubs there is only one path joining them.

For illustration purposes we have applied this modified method to a five-zone network in [3], see Figure 5. We note that in this network, there is only one directed path between any two zones. The traffic demands between zone i and zone j is assumed to be a Poisson process with mean D_{ij} as follows:

$$D = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{pmatrix} - & 100 & 300 & 400 & 100 \\ 50 & - & 125 & 125 & 50 \\ 300 & 100 & - & 700 & 700 \\ 500 & 250 & 500 & - & 1000 \\ 300 & 225 & 225 & 600 & - \end{pmatrix}.$$

The network can be decomposed into three 3-zone star networks as follows: (1,2,H), (H,3,H") and (H',4,5). We test our algorithm by using simulation with initial guess being

given by

$$D_0 = \begin{pmatrix} - & 500 & 500 & 500 & 500 \\ 500 & - & 500 & 500 & 500 \\ 500 & 500 & - & 500 & 500 \\ 500 & 500 & 500 & - & 500 \\ 500 & 500 & 500 & 500 & - \end{pmatrix}.$$

The average relative errors for the first 30 iterations are given in Figure 6. We see that the average errors decrease very quickly.

8 Concluding Remarks

We have studied the problem of updating estimates of traffic demands between zones in a network from observed traffic counts. The contribution of the paper is two-folded: First, we suggest an estimation model which makes an explicit assumption on the arcs where traffic is counted, at all entrances and exits of each zone, and uses a forecasting procedure to regularly update demand estimations based on arriving traffic counts. Second, we suggest numerical procedures for the estimation and argue theoretically and through numerical illustrations that these procedures are efficient, numerically and statistically. The first approach is based on replacing unobserved between-zones traffic by its expectation conditional on observed traffic counts, whereas the second method is based on replacing the unobserved between-zones traffic by the most likely traffic volume subject to consistency with the observed counts. The first approach is specified for normally distributed traffic demands, while the second is particularly useful for Poisson demands. The second approach is, through the use of Stirling's approximation, linked with the maximum entropy in the literature. Numerically, the first procedure is simpler since it relies on the solution of a single system of linear equations for each observation, whereas the second procedure is based on solving an unconstrained optimization problems when new data arrives.

Acknowledgments

Research partially supported by RGC Grant No. HKU 7126/02P and HKU CRCG Grant No. 10203919 and 10204436. We thank Richard Weber of Cambridge University for pointing us to Lemma 1 and Mike Maher of Napier University, Edinburgh, for his helpful comments on an earlier draft of the manuscript.

References

[1] Abraham B. and Ledolter J. (1983) *Statistical Methods for Forecasting*, Wiley, New York.

[2] Bell M.G.H. (1991) The real time estimation of origin-destination flows in the presence of platoon dispersion. *Transportation Research*, **25 B**, 115-125.

[3] Cascetta E. (1984) Estimation of trip matrices from traffic counts and survey data: a generalised least square estimator. *Transportation. Research*, **18B**, 289-299.

[4] Cascetta E. and Nguyen S. (1988) A unified framework for estimating or updating origin/destination matrices from traffic counts. *Transportation Research*, **22B**, 437-455.

[5] Cascetta E. and Nguyen S. (1997) Calibrating aggregated travel demand models with traffic counts: estimators and statistical performance. *Transportation Research*, **24B**, 271-293.

[6] Chen Y. and Florian M. (1988) Congested O-D trip demand adjustment problem: bi-level programming formulation and optimality conditions in: *Multilevel Optimisation: Algorithms and Applications* (eds. A. Migdalas, P.M. Pardalos and P. Varbrand). Kluwer Academic Publishers, Dordrecht.

[7] Conn A.R., Gould B.I.M., and Toint Ph.L. (forthcoming), *Trust-region Methods*. SIAM Publications, Philadelphia.

[8] Cremer M. and Keller M. (1987) A new class of dynamic methods for the identification of origin-destination flow. *Transportation Research*, **21B**, 117-132.

[9] Dempster A , Laird N. and Rubin D. (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc.* **39 B**, 1-38.

[10] Dennis J.E. and Schnabel R.B. (1983) *Numerical Methods for Unconstrained Optimisation and Nonlinear Equations*. Prentice Hall, Englewood Cliffs.

[11] Domencich T. and McFadden D. (1975) *Urban travel demand: A behavioral analysis*. North-Holland Publishing Company, New York.

[12] Duda R., Hart P. and Stork D. (2000) *Pattern Classifications*, Wiley, New York.

[13] Feller W. (1957) *An introduction to probability theory and its applications*. Vol. 1, John Wiley, New York.

[14] Fisk C. (1988) On combining maximum entropy trip matrix estimation with user optimal assignment. *Transportation Research*, **22B**, 66-79.

[15] C. Fisk (1989) Trip matrix estimation from link traffic counts: The congested network case. *Transportation Research*, **23B**, 331-356.

[16] Golub G. and van Loan C. (1993) *Matrix Computations*. The John Hopkins University Press, Baltimore.

[17] Horn R.A and Johnson C.R. (1985) *Matrix Analysis*. Cambridge University Press, Cambridge.

[18] Jornsten K. and Nguyen S. (1980) On the estimation of trip matrix from network data. Report NITH-MAT-R79-36, Linkoping Institute of Technology, Linkoping, Sweden.

[19] Maher M.J. (1983) Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach, *Transportation Research*, **17B**, 435-447.

[20] Maher M.J. and Zhang X. (1999) Algorithms for the solution of the congested trip matrix estimation problem, in: *Transportation and traffic theory* (ed. A. Ceder), Elsevier Science, 445-469.

[21] Nihan N.L. and Davis G.A. (1987) Recursive estimation of origin-destination matrices from input/output counts. *Transportation Research*, **21B**, 149-163.

[22] Nihan N.L. and Davis G.A. (1989) Application of prediction-error minimisation and maximum likelihood to estimate intersection O-D matrix from traffic counts. *Transportation Science*, **23**, 77-90.

[23] Nocedal J. and Wright S. (1999) *Numerical Optimisation*, Springer-Verlag, New York.

[24] Oppenhein N. (1995) *Urban Travel Demand Modeling*, Wiley, New York.

[25] Ortuzar J.D. and Willumsen L.G. (1994) *Modelling Transportation*, Wiley, New York.

[26] Silverman B., Jones M., Wilson J. and Nychka D. (1990) A smoothed EM approach to indirect estimation problems with particular reference to sterelogy and emission tomography. *J. R. Statist. Soc.*, **52 B**, 271-324.

[27] H. Speiss (1987) A maximum likelihood model for estimating origin-destination matrices. *Transportation Research*, **21B**, 395-412.

[28] Van Zuylen H. and Willumsen L. (1980) The most likely trip matrix estimated from traffic counts. *Transportation Research*, **14B**, 281-293.

[29] Weber R. (1999) *Optimisation and Control*, Lecture Notes, Statistical Laboratory, University of Cambridge, UK.

[30] Wilson A.G. (1970) *Entropy in Urban and Regional Modelling*. Pion, London.

[31] Wilson A.G. (1974) *Urban and Regional Models in Geography and Planning*. John Wiley and Sons, London

[32] Yang H., Sasaki T., Iida Y. and Asakura Y. (1992) Estimation of origin-destination matrices from link traffic counts on congested network. *Transportation Research*, **26B**, 417-434.

[33] Yang H., Iida Y. and Sasaki T. (1994) The equilibrium-based origin-destination estimation problem. *Transportation Research*, **28B**, 23-33.

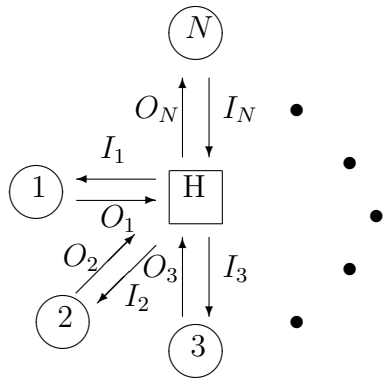

Figure 1: The Star Network.

Table 1: Number of Iterations for Convergence for (a).

N	n	W_1	T_1	C_1	W_2	T_2	C_2
25	600	5	46	9.20	5	59	11.80
50	2450	4	34	8.50	4	45	11.25
100	9900	4	34	8.50	4	45	11.25
200	39800	4	34	8.50	4	46	11.50

Table 2: Number of Iterations for Convergence for (b).

N	n	W_1	T_1	C_1	W_2	T_2	C_2
25	600	5	71	14.20	5	88	17.60
50	2450	4	49	12.25	4	62	15.50
100	9900	4	49	12.25	4	61	15.25
200	39800	4	46	11.50	4	63	15.75

Table 3: Average Absolute Errors for Different Smoothing Factors (Case 1).

N	Optimal α	Optimal Average Absolute Error	Average Absolute Error When $\alpha = 0.1$	Average Absolute Error When $\alpha = 1$
3	0.01	0.00%	0.00%	0.00%
6	0.01	0.10%	0.37%	0.39%
12	0.01	0.08%	0.29%	0.30%
24	0.01	0.08%	0.28%	0.29%
48	0.01	0.08%	0.28%	0.29%

Table 4: Average Absolute Errors for Different Smoothing Factors (Case 2).

N	Optimal α	Optimal Average Absolute Error	Average Absolute Error When $\alpha = 0.1$	Average Absolute Error When $\alpha = 1$
3	0.16	0.05%	0.39%	0.05%
6	0.13	0.36%	0.47%	0.38%
12	0.20	0.30%	0.42%	0.30%
24	0.19	0.31%	0.41%	0.31%
48	0.18	0.32%	0.43%	0.32%

Table 5: Average Absolute Errors for Different Smoothing Factors (Case 3).

N	Optimal α	Optimal Average Absolute Error	Average Absolute Error When $\alpha = 0.1$	Average Absolute Error When $\alpha = 1$
3	0.18	0.06%	0.47%	0.06%
6	0.30	0.48%	0.68%	0.48%
12	0.34	0.40%	0.59%	0.40%
24	0.30	0.39%	0.56%	0.39%
48	0.26	0.40%	0.55%	0.40%

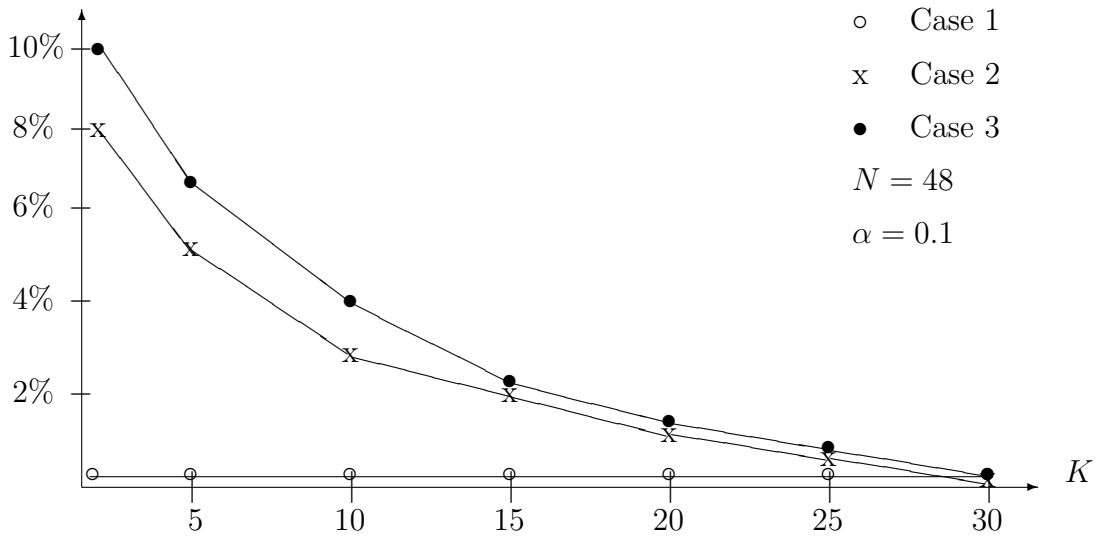

Figure 2: The Average Relative Errors.

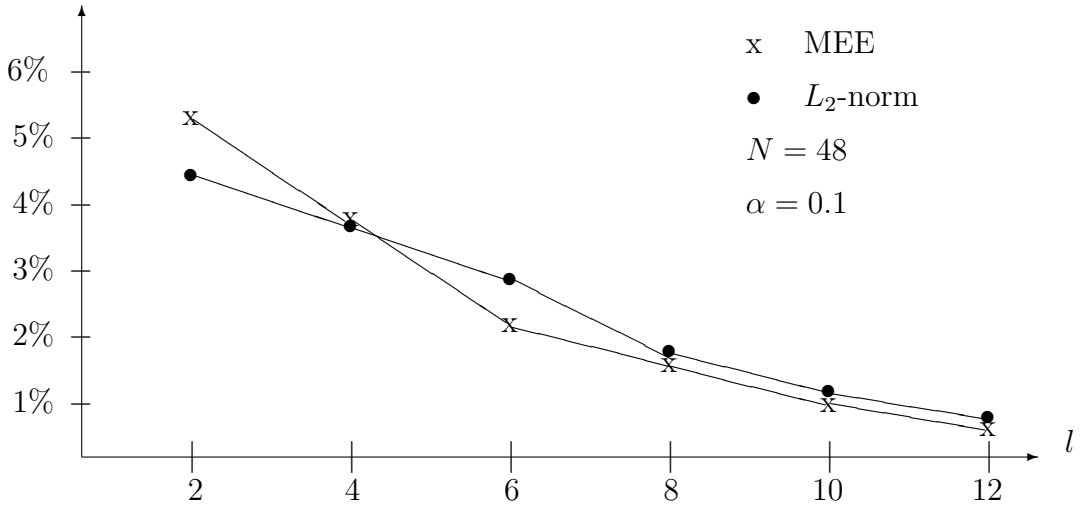

Figure 3: The Average Relative Errors.

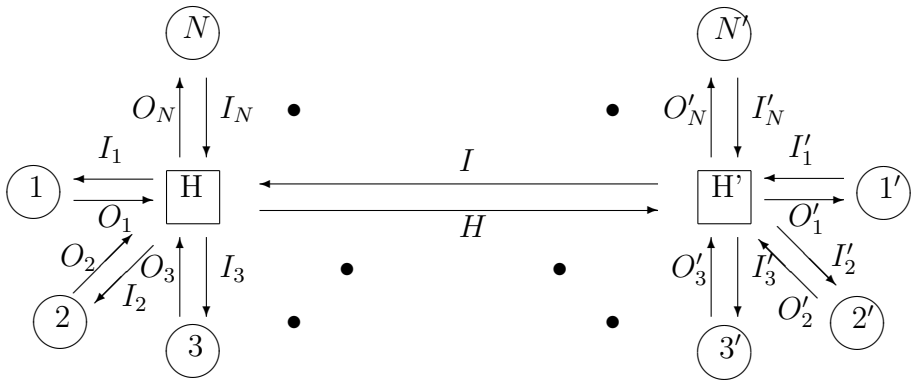

Figure 4: The Two-Star Network.

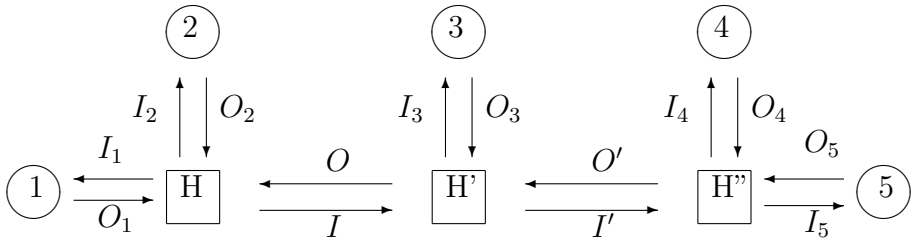

Figure 5: The Five-zone Network.

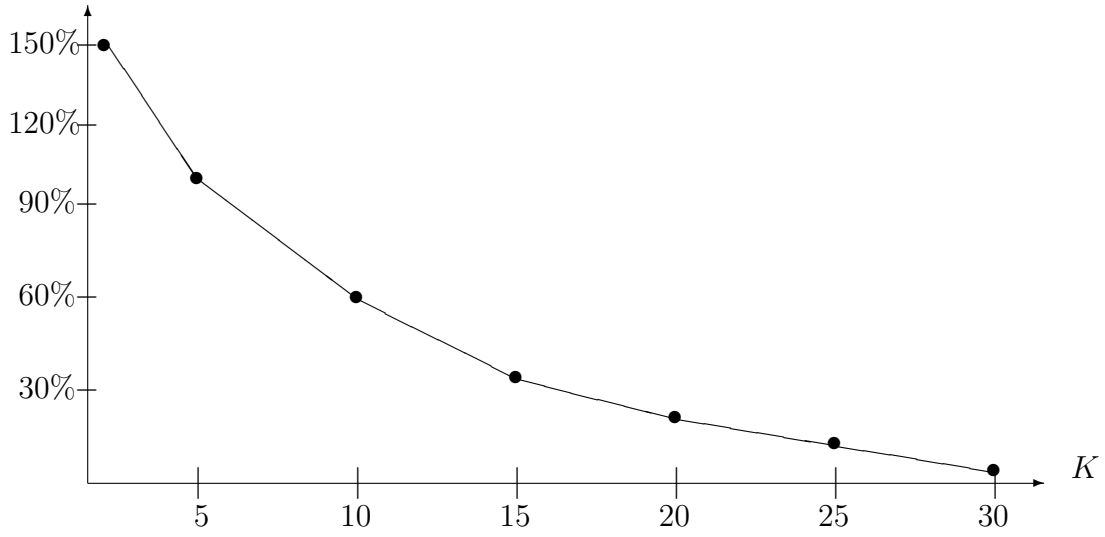

Figure 6: The Average Relative Errors.