EWGT2013 – 16th Meeting of the EURO Working Group on Transportation

Static OD estimation minimizing the relative error and the GEH index

Aleix Ruiz de Villa\textsuperscript{a,}*\textsuperscript{,} Jordi Casas\textsuperscript{a}, Martijn Breen\textsuperscript{a}, Josep Perarnau\textsuperscript{a}

\textsuperscript{a}TSS – Transport Simulation Systems, Passeig gràcia 12, Barcelona 08007, Spain

Abstract

One of the essential quantities in macro traffic simulation is the ‘who goes where’ OD matrix. This is usually estimated through an optimization problem of minimizing the squared differences between observed and simulated traffic counts. The choice of the mean square error loss is due to its good statistical properties and the fact that the related quadratic optimization problem is easy to deal with. However, in many applications following the recommendations of validation criteria suggested in many guidelines published by different road administrations, such as FHWA in USA, ARRB in Australia or Highways Agency in UK, the quality of an static OD adjustment is assessed with other quantities such as the relative error or the GEH index. In this paper we estimate OD matrices minimizing directly the mean geh or mean relative error.

Keywords: static OD estimation; geh; absolute deviation loss.

1. Introduction

The static Origin-Destination (OD) estimation or adjustment matrix is a classical problem applied and solved in Transport Planning. This class of problems can take on many different formats, depending upon what is considered to be known and what assumptions are made to derive the missing parameters. The description of generic problem is described below.

Given an OD matrix $Q$, denote by $v_a$ the volume on a link $a$ as a result of a particular traffic assignment with OD matrix $Q$. In practice, we are often interested in the reverse problem, that is, given a set of links $(a)_{a \in A}$ and observed volumes in such set $(v'_a)_{a \in A}$, find an OD matrix $Q$ such that the assigned volumes $v_a$ are as close as

\* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .

E-mail address: aleix.ruizdevilla@aimsun.com
possible to the measured ones. Usually an initial matrix \( Q_0 \), regarding prior information about traffic behavior (obtained for instance by a survey), is required and the objective is to modify \( Q_0 \) to fit current data. This process is called matrix adjustment or estimation. When traffic flows are modeled time-independently the matrix adjustment process is called static. Otherwise it is called dynamic. See Bera and Rao (2011) and the references therein for a summary of the main techniques used to approach this problem. In this paper, we will focus on bi-level static based techniques. Spiess (1990) solved this problem using a gradient based approach with minor modifications to ensure preservation of unused paths. In Yang, Sasaki, Iida and Asakura (1994) and Yang (1994) proposed heuristic approaches to deal with the bi-level mathematical programming. A Gauss-Seidel type coordinate decent approach can be found in (Florian & Chen, 1995).

Traditionally square error loss has been chosen to measure disagreement between assigned and measured volumes. The reason is that it is a widely studied loss (with a central role in statistics) and it is easy to use (it is differentiable and easy to calculate). Thus, the matrix adjustment problem can be stated as

\[
\min_Q \sum_{a \in A} (v_a - v_a^0)^2.
\]

The GEH Statistic is an index used in traffic modeling to compare two sets of traffic volumes. The GEH formula gets its name from Geoffrey E. Havers (UK Highways Agency, 1992). Although its mathematical form is similar to a chi-squared test, it is not a true statistical test. Rather, it is an empirical formula that has proven useful for a variety of traffic analysis purposes. This index is commonly used as validation criteria in many applications following the recommendations suggested in many guidelines published by different road administrations, such as FHWA in USA, ARRB in Australia or Highways Agency in UK. For traffic modeling the different guidelines has the following consensus: a GEH index of less than 5.0 in a measurement point is considered a good match between the modeled and observed volumes, and 85% of the volumes in a traffic model should have a GEH less than 5.0 for all measurement points.

The GEH of two quantities, usually measured and predicted or simulated nonnegative data, is defined as

\[
Geh(y, \hat{y}) = \frac{2(|y - \hat{y}|)^2}{y + \hat{y}} = \frac{|y - \hat{y}|}{\sqrt{(y + \hat{y})/2}}
\]

The main purpose of this paper is solving the optimization problem

\[
\min_Q \sum_{a \in A} Geh(v_a, v_a^0).
\]

Such a problem can become difficult because both the numerator and denominator depend on \( Q \). For this reason we consider the approximation

\[
\min_Q \sum_{a \in A} \frac{|v_a - v_a^0|}{\sqrt{v_a}} \tag{1}
\]

provided that \( \frac{v_a + v_a^0}{2} = v_a \). The main difficulty now is that the weighted absolute deviation loss

\[
g(v_a, v_a^0) = |v_a - v_a^0|
\]

is nondifferentiable, and thus classical gradient descent techniques cannot be directly implemented. Fitting linear models using absolute deviation loss, called median regression, is a well researched area. See (Koenker, 2005) for a comprehensive treatment on the subject. It is well known that one of the strengths of using absolute value loss is that the resulting models are more robust against outliers. Usually median regression models are solved transforming the problem into a linear programming one and applying techniques such as the simplex or interior...
point methods. Instead, in this paper, we follow a more flexible approach based on Staines and Barber (2012). First we find differentiable functions $g_\sigma \to g$ as $\sigma \to 0$. Then, for a selected sequence $\sigma_n$, we find $Q_n$ minimizing
\[
\min_{Q} \sum_{a \in A} \frac{1}{\sqrt{v^2_a}} g_{\sigma_n}(v_a, v^\sigma_a),
\]
using classical gradient descent techniques.

2. Smoothed Loss

To find the functions $g_\sigma$ described above we will follow the ideas Staines and Barber (2012), in our case under the point of view of approximations to the identity. Given a function $\eta: \mathbb{R} \to \mathbb{R}$ with $\int \eta = 1$ and $\sigma > 0$, consider $\eta_\sigma(s) = \frac{1}{\sigma} \eta(\frac{s}{\sigma})$ and the convolution
\[
g_\sigma(r) = g * \eta_\sigma(r) = \int g(r - s) \eta_\sigma(s) ds = \int g(s) \eta_\sigma(r - s) ds.
\]
The family $\{g_\sigma\}$ is called approximation to the identity. The reason is that if $g$ is continuous and compactly supported, then $g_\sigma(r) \to g(r)$ uniformly as $\sigma \to 0$. On the other hand, if $g$ is integrable ($\int |g| < \infty$), then $\int |g_\sigma - g| \to 0$ as $\sigma \to 0$. See Duoandikoetxea (2001) for more details. The rationale behind this is the following: as $\sigma$ decreases, all the mass of $\eta_\sigma$ tends to concentrate near 0. For small $\sigma$, the value of $g_\sigma(r)$ at the point $r$ is a weighted average of the points near $r$, so only points close to $r$ are taken into account. In the limit, in some sense, we are only considering the point $r$ itself.

The key point is that if $\eta$ is infinitely differentiable and $g$ integrable, the functions $\{g_\sigma\}$ are infinitely differentiable, which means that we are approximating integrable functions (not necessarily continuous) by infinitely differentiable functions!

Our choice for the function $\eta$ is the Gaussian kernel
\[
\phi(s) = \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}},
\]
and correspondingly
\[
\phi_\sigma(s) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{s^2}{2\sigma^2}}.
\]
In our case, the function $g(r) = |r|$ is neither compactly supported nor integrable. However, $g_\sigma(r) = g * \phi_\sigma(r)$ is well defined and infinitely differentiable. Denoting $\Phi$ the cumulative distribution function of $\phi$, it is not difficult to see that
\[
g_\sigma(r) = g * \eta_\sigma(r) = \frac{2\sigma}{\sqrt{2\pi}} e^{-\frac{r^2}{2\sigma^2}} + r \left(1 - 2\Phi\left(-\frac{r}{\sigma}\right)\right).
\]
In Fig. 1, we see the approximating sequence for values of sigma $\{0, 0.25, 0.5, 1, 2\}$. The functions $g_\sigma$ converge uniformly to $g$ as $\sigma \to 0$, thus suitable for our purpose. Their derivatives are given by
\[
g'_\sigma(r) = 1 - 2\Phi\left(-\frac{r}{\sigma}\right).
\]
3. The algorithm

We propose a gradient based algorithm, following the ideas of Spiess (1990). Observe that this algorithm can be used with any traffic assignment algorithm that given an OD matrix \( Q \) assigns volumes to the set of links \( (a)_{a \in A} \). First, let us introduce some notation. For each origin-destination pair \( (i, j) \), path \( k \in K_{ij} \) = {paths with origin \( i \) and destination \( j \)} and link \( a \):

- \( f_{ij} \) denotes the flow sent from \( i \) to \( j \), that is, \( Q = (f_{ij})_{ij} \).
- \( h^k_{ij} \) denotes the flow sent from \( i \) to \( j \) taking path \( k \in K_{ij} \) as a result of a specified traffic assignment algorithm.
- \( p^k_{ij} = h^k_{ij}/f_{ij} \) denotes path probabilities.
- The delta function \( \delta_{a,k} \) takes value 1 if the path \( k \) passes through the link \( a \) and 0 otherwise.
- For a given \( \sigma > 0 \),
  \[
  S = \sum_a \frac{1}{\sqrt{v_a^2}} g\sigma (v_a - v_a^0).
  \]

Volumes on links can be calculated by

\[
v_a = \sum_{ij} \sum_{k \in K_{ij}} h^k_{ij} \delta_{a,k} = \sum_{ij} f_{ij} \sum_{k \in K_{ij}} p^k_{ij} \delta_{a,k}.
\]

Assuming that path probabilities are locally constant, we approximate
\[ \frac{\partial v_a}{f_{ij}} \approx \sum_{k \in K_{ij}} p_{ik}^k \delta_{a,k} \]

and the gradient \( \nabla S = \left( \frac{\partial S}{f_{ij}} \right) \) is approximated by

\[ \frac{\partial S}{f_{ij}} = \sum_a \frac{1}{\sqrt{v_a}} \frac{\partial v_a}{f_{ij}} g_a(v_a - v_a^o) \approx \sum_{a,k \in K_{ij}} \frac{p_{ik}^k g_a}{\sqrt{v_a}} g_a(v_a - v_a^o). \]  

(3)

The gradient descent iteration is modified in order to preserve zeros in flows \( f_{ij} \), see (Spiess 1990), obtaining the new OD matrix by

\[ f_{ij}^N = f_{ij}(1 - \alpha \frac{\partial S}{f_{ij}}), \]

where the step size \( \alpha \) satisfies \( 0 \leq \alpha \leq 1 / \frac{\partial S}{f_{ij}} \) for all \( i, j \) with \( \frac{\partial S}{f_{ij}} \geq 0 \). Let \( v_a^N \) be the resulting volumes of a new traffic assignment with OD matrix \( Q^N = (f_{ij}^N)_{ij} \), and

\[ \tilde{v}_a^N = \sum_{i,j} f_{ij}^N \sum_{k \in K_{ij}} p_{ik}^k \delta_{a,k} = v_a - \alpha \sum_{i,j} f_{ij} \frac{\partial S}{f_{ij}} \sum_{k \in K_{ij}} p_{ik}^k \delta_{a,k} = v_a - \alpha c_a \]

the estimated volumes assuming that path probabilities are locally constant. For small \( \sigma \),

\[ \sum_a \frac{1}{\sqrt{v_a}} g_a (v_a^N - v_a^o) \approx \sum_a \frac{1}{\sqrt{v_a}} g_a (\tilde{v}_a^N - v_a^o) \approx \sum_a \frac{1}{\sqrt{v_a}} |v_a^N - v_a^o| = \sum_a \frac{1}{\sqrt{v_a}} |v_a - v_a^o - \alpha c_a|. \]  

(4)

Thus, \( \alpha \) can be estimated solving

\[ \min_\alpha \sum_a \frac{1}{\sqrt{v_a}} |v_a - v_a^o - \alpha c_a| \]  

s. t. \( 0 \leq \alpha \leq 1 / \frac{\partial S}{f_{ij}} \),

using standard univariate quantile regression.

In general, the choice of the gradient and step size \( \alpha \) suggested above doesn’t provide a descent direction (a decrease of the loss with the current iteration is not guaranteed). The derivative of \(|x|\) is 1 if \( x > 0 \) and \(-1\) if \( x < 0 \), thus, unlike the L2 loss, it does not provide any information on how close the minimum is. Therefore, this method has to be complemented with some line search algorithm (see for instance Nocedal and Wright (2006)). However, the value \( \alpha \) above can be used as an initial point for such algorithms.

---

Fig. 2. Test network with detectors
4. Results

We tested our method on a network of a medium sized city of Spain. The implementation has been done in a combination of Aimsun (2013) and R (2012).

4.1. Network

The network that has been used to perform the experiments is a medium urban network with a highway stretch north of the city centre. The network comprises a total road length of 631 km and the different zones are modeled with 57 centroids. The network detection consists of 362 count detectors which are located both in the urban centre as on the highway stretch. In the Fig. 2 the distribution of the detectors is shown in which the green dots represent the count detectors.

The assignment period is the evening peak hour in which a total of 35000 cars are assigned to the road network. With this demand a synthetic dataset has been created which will be used to validate the O/D estimation procedure.

The original demand matrix is perturbed in order to be used as a starting point for the O/D estimation procedure. Each cell is multiplied by a random normal distribution with mean 0.9 and standard deviation 0.2. In the Fig. 3 the trip values of the original matrix are plotted against the perturbed trip values to show the degree of perturbation of the original matrix

![Fig. 3. Perturbed cells](image)

The assignment method that has been used for this experiment is the Frank-Wolfe algorithm, which is available in the traffic simulation package Aimsun (2013). The network achieves an equilibrium status with a relative gap below 1% after 50 iterations as shown in the figure below. From the assignment result the path proportions are extracted to be used in the O/D estimation.

4.2. Algorithm’s specifications

In our numerical experiments we have used a simple line search algorithm; we define \( \alpha_k = (n - k)/n \) with \( n \) fixed (\( n = 20 \) in our case). We start with \( k = 0 \), and perform traffic assignments with OD matrices
for each \(0 < k \leq \eta\) until we find a descent direction. We have used the decreasing sequence \(\sigma_n = \gamma^n \sigma_0\) with \(\gamma = 0.8\) and \(\sigma_0 = 0.5\).

4.3. Experiments

The performance of the proposed method has been compared with the traditional least squares estimation of Spiess (1990). In the following figures we distinguish between the optimization method (called Minimization: geh and L2) and the evaluated loss or objective function (mean geh “Geh”, mean square error “L2” and the percentage of measurement points with a GEH index < 5 “%geh<5”). Two types of perturbations have been considered. On the one hand, in applications, since we do not know the OD matrix that generates the observed data, an estimated initial OD matrix is obtained from external sources, such as a survey. Thus, the original matrix in our experiments has been perturbed as explained above with some random noise. On the other hand, usually observed data is perturbed due to measurement errors and many other causes. For this reason we have perturbed the assigned volumes from the original matrix in two ways. In the first one, all of the assigned volumes from the original matrix have been perturbed with random gaussian distributions with factors of 5\% and 15\%. The second set of experiments recreates the effect of outliers: the volume of four relevant detectors (with high flow) has been increased in a 30\%. The results are shown in Fig. 4. The x-axis of the panel is used for the type of perturbation (none, Gaussian and outliers), while the y-axis is used for the evaluated loss.

Generally speaking, with the “Geh” and “L2” we see that minimizing a quantity, forces the other also to decrease (and increase the “%geh<5”). However, both minimizations are not equivalent; as expected, we get lower L2 mean loss when using a L2 minimization and vice versa. Moreover there is a gap that cannot be avoided using a crossed loss choice (for instance using a L2 minimization and evaluating with the geh). In the case of the 15\% perturbation, after 80 iterations we get a mean geh of 2.34 and 2.47 with geh and L2 minimizations respectively, thus leading only to a 5\% difference between the algorithms. On the other hand we get a 50\% difference in L2 loss between methods (mean square error of 8244 and 5507 with geh and L2 minimizations respectively). In the outliers case, we achieve an 80\% difference with the geh evaluation (mean geh of 0.6 vs 1.10 geh and L2 minimizations respectively), while only a 25\% difference with the L2 evaluation (mean square error of 1789 and 1425 with geh and L2 minimizations respectively). Observe that the “%geh<5” follows “the best” in each case: L2 minimization gets a higher “%geh<5” in the 15\% perturbation case, while the Geh minimization gets a higher “%geh<5” in the presence of outliers.
5. Conclusions

We have introduced a methodology to adjust OD matrices minimizing the mean geh between observed and assigned flows. The main difficult is to deal with the absolute loss because of its nondifferentiability. We have solved this problem providing a sequence of differentiable loss functions tending to the absolute loss and solving the related minimization problems using differentiable techniques. One of the strengths of this method is that ways of addressing problem (2) can be borrowed from a well researched area such as optimization of differentiable functions; inheriting not only the efficiency but also the flexibility. In particular computational time could be improved using newton or quasi-newton approaches.
Numerical experiments suggest that minimizing the mean geh is more suitable in presence of outliers. When using squared loss, greater errors have relatively more effect in the estimation process than when absolute loss is used. This is the analog to linear regression, as mentioned at the introduction.

We also see that maximizing the “%geh<5”, although tightly related with minimizing the mean geh or the L2, is not equivalent to any of the two. If we wanted to directly maximize the “%geh<5”, other algorithms should be developed. In fact, this is a combinatorial problem and seems, at least for the authors, much more difficult than the one developed in this paper.

Minor modifications of the algorithm can lead to minimizing the relative error

$$\min_Q \sum_{a \in A} \frac{|v_a - v_a^0|}{v_a^0}$$

or the absolute error

$$\min_Q \sum_{a \in A} |v_a - v_a^0|.$$ 

That is, in the first case it is sufficient to replace weights \( \left( \frac{1}{\sqrt{v_a^0}} \right) \) in equations (3), (4) and (5) by weights \( \left( \frac{1}{v_a^0} \right) \). In the second, such weights should be removed.

References