

Quantifying partial observability in network sensor location problems

Francesco Viti, University of Luxembourg

Marco Rinaldi, Katholieke Universiteit Leuven

Francesco Corman, Delft University of Technology

Introduction

Information in traffic networks is crucial for state estimation, model calibration, travelers' information services, traffic management, etc. In this view, location and positioning of sensors plays a fundamental role. This problem is traditionally referred to as Network Sensor Location Problem (NSLP). Traditionally, NSLP problems have been seen as an inverse problem wrt. the static origin-destination matrix estimation based on traffic counts, i.e. finding those locations and the minimum number of sensors necessary in order to minimize the uncertainty on the estimation. Therefore, the solutions depend essentially on the OD estimation technique adopted (GLS, Maximum Likelihood, etc.), sometimes on supply-side information (capacities) and on a-priori information of (part of) the flows in the network (OD, link or route flow samples). By exploiting the information available in terms of pre-selected locations (existing sensors or a-priori OD matrices) or by making educated assumptions over the possible distribution of the demand and flows, partial solutions could be calculated (see e.g. Gentili and Mirchandani, 2012 for an overview).

A new emerging class of NSLP problems concerns instead the exploitation of network topology information only. Deriving from other disciplines (telecommunications, electrical networks), these problems aim to identify the solutions which guarantee full network observability, i.e. find the minimum number and locations in the network such that all unobserved variables are derived by simple algebraic rules based on conservation of vehicles equations (see Viti and Corman, 2012, for a more detailed description). Since this class of problems is based only on basic network information and does not need any a-priori information on the flows, partial solutions haven't received great attention, since it is hard without any assumption on flows or estimation techniques to identify the amount of information not explained by the chosen sensors.

The problem that remains when adopting full observability solutions in real traffic networks is that very often the number of sensors needed is too large. One would then be interested in finding solutions for which a good deal of the variables in the network is still observable (or part of them) even with a subset of the full observability solution. However, a metric able to measure the loss of information in these partial observability solutions is lacking. This is the main focus of this paper.

Methodology

This work focuses on the development and testing of a new methodology and metric that, based solely on topological information and sound algebraic methods, is very responsive to the amount of information being captured by the observed variables. A measure of the maximum error on the unobserved link flows can be estimated and used to identify the set of links that will minimize this error in a partial observability setting.

The method is based on the pivoting process proposed recently by Castillo et al (2008), which provides a full observability solution where all selected links are linearly independent. Figure 1 provides an example of how a matrix is manipulated so that all remaining link flows on the rows

become function of only on the link flows indicated in the columns. One possible solution for this problem is also indicated in the figure (see Rinaldi et al., 2013 for a full elaboration of the case).

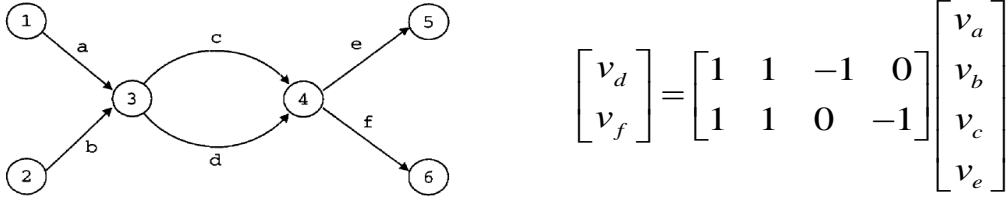


Figure 1: Simple toy network for which multiple pivoting solutions are available

Different solutions of the pivoting procedure are normally available, depending on the sequence of links selected to manipulate the link-route incidence matrix. This is an agreeable property in full observability solutions, but one may need to know which of these solutions is a more adequate choice when not all of the linearly independent links are actually equipped with a sensor.

Starting from a full observability solution \mathbf{P} , a partial observability solution \mathbf{P}' leads essentially to an underdetermined set of equations which exhibits a non-empty null space that can be represented by an algebraic base \mathbf{B}' . By relating this base vector to the full observability matrix \mathbf{P} we obtain a measure of the information lost when reducing the observed set from \mathbf{P} to \mathbf{P}' . We therefore define a metric indicating the *amount of information* pertaining to the set of variables corresponding to the highest possible reduction in solution space size, ultimately leading it to collapse into a single point, corresponding to full information. This metric is here defined as the Null-SPace metric:

$$NSP = \frac{\left\| \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{P} & \mathbf{0} \end{pmatrix}^T \mathbf{B}' \right\|_F}{\left\| \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{P} & \mathbf{0} \end{pmatrix} \right\|_F} \quad (1)$$

where \mathbf{I} is the identity matrix and $\|\cdot\|_F$ indicates the Frobenius or Hilbert-Schmidt norm. For a full elaboration of the method leading to this metric we refer to Rinaldi et al. (2013). By definition, this metric is defined in the range $[0,1]$ so it is easily interpretable as the amount of information ‘content’.

Algorithm

By starting from the smallest number of linearly independent variables that guarantees full observability (matrix \mathbf{P}), every link selected in the solution should contain a ‘unique’ piece of information necessary to explain one of the unknown link variables. The proposed metric is able to capture this information content. A global optimal partial observability solution given N links could be therefore found by looking for the N links among the linearly independent ones. Unfortunately this is an unrealistic operation given the numerous combinations that exist, especially in large networks, and also considering that the choice of the initial pivot may hinder globally optimal solutions reachability. (we skip this demonstration for sake of simplicity).

A greedy algorithm which selects links in sequence may be adopted to overcome the scalability issue. However, due to the linear relations between dependent and independent links, the choice of selecting one link influences the subsequent choices. Thus, given N links to equip with a sensor, if selecting one sensor in a partial solution may be equivalent to another link (e.g., links a or b in Figure 1) this solution may determine which other links to select, and the final partial solution may be suboptimal with respect to the alternative link choice.

It can be shown (we are working on the formal demonstration) that by using the metric in Equation (1) coupled with an enumeration of all possible full observability solutions, an optimal solution (i.e. the minimum value of the NSP metric given N sensors) is reached either by starting from an empty set of sensors and then sequentially adding sensors in the network, or by starting from the full

observability solution and removing in sequence the links until N are left. This agreeable property makes the problem more tractable even for large networks, as the only permutation operation to perform is to generate all possible full observability solutions.

Example

The method has been tested on various small networks confirming that optimal solutions are reached given a pivoting solution. Figure 2 shows the result on a large network (a simplified representation of the road network of Rotterdam, the Netherlands, left picture). Figure 2 (right) shows the NSP values for an increasing number of sensors selected, in comparison to a much simpler metric which simply takes independent link variables starting from those contained more frequently in the system of equations to calculate the linearly dependent links (Viti and Corman, 2013). As one can easily observe, the greedy algorithm using the NSP metric is able to identify links with the largest information content for any value of N with respect to the simpler metric. Another interesting finding is that the optimal solution found for $\text{NSP} = 0.5$ tends to get as much information as possible at motorway nodes, which intuitively contain the largest information in the network.

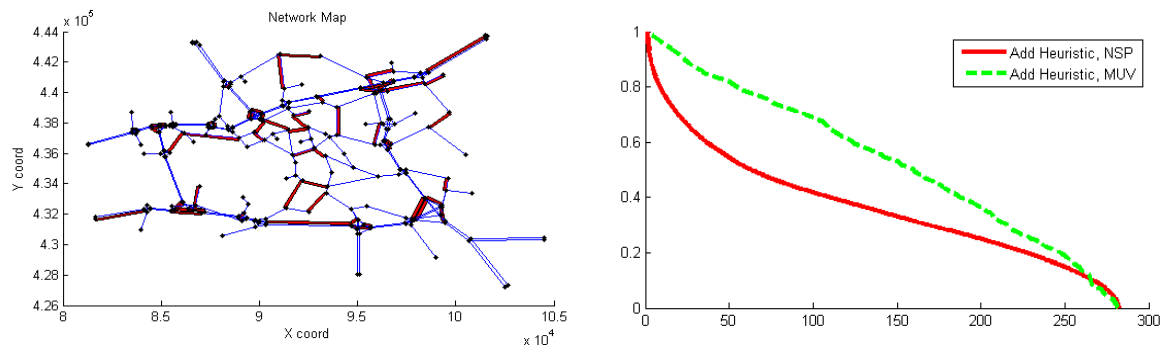


Figure 2: Greedy algorithm applied to install an increasing number of sensors in the network (right) and solution that results in $\text{NSP}=0.50$ visualized on the network (left)

Conclusions

In this paper the problem of measuring the amount of information gained in observing a link flow in a network is presented. The introduced new metric allows one to quantify the quality of a partial observability solution, which is a very agreeable knowledge for many possible applications, for instance to know on beforehand what could be the uncertainty due to incomplete network coverage on OD matrix or state estimation given the set of sensors installed in the network, or where should extra sensors be positioned in order to complement existing sensors, etc. Research is currently on going to explore the properties of this new methodology on different network layouts and types and using different modifications of the metric. Part of these new findings will be presented at the symposium.

References

- Castillo, E., Conejo A., Menendez, J.M., Jimenez, P. (2008a) The Observability Problem in Traffic Network Models. *Computer Aided Civil and Infrastructure Engineering*, Vol. 23, pp. 208–222.
- Gentili M., Mirchandani P. (2012) Survey of Models to Locate Sensors to Estimate Traffic Flows. *Transportation Research Record*, Vol. 2243, pp. 108-116.
- Rinaldi M., Corman F., Viti F. (2013). A Null-Space metric for the analysis of partial observability in sensor location problems. In *Proceedings of the TRB Annual Meeting*, Washington D.C.
- Viti F., Corman F. (2012). An Approach to the Sensor Location Problem for Measuring Network Variability. *Proceedings of the 5th INSTR Conference*, December 2012, Hong Kong, China.