Identifiability in Network Tomography

under Dependence

George Michailidis

Department of Statistics and EECS
The University of Michigan
gmichail@umich.edu, http://www.stat.lsa.umich.edu/∼gmichail

DIMACS DyDAn Working Group, March 2009

Joint work with Harsh Singhal
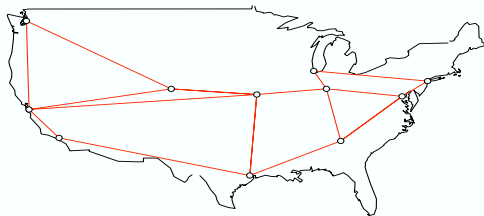
A High Level Look at the Problem

Consider the following linear inverse problem

$$Y = AX,$$

where Y is a $L \times 1$ vector of measurements, X a $J \times 1$ vector of unknown quantities of interest, and A a $L \times J$ known matrix, with $L << J$.

Interested in estimating X.

Consider a (computer) network comprised of nodes and wlog bidirectional links.

Network Flows

- A computer network carries packets, whose payload is expressed in bytes.
- A network flow contains all the packets originating at a node and destined for some other node in the network.
- Each flow can in principle traverse a set of paths connecting its origin and destination, which is determined by the routing policy, assumed known.
- The volume of traffic refers to either the number of packets and/or the number of bytes in a flow (or on a link) in a given time-interval.

Network Flow Volumes
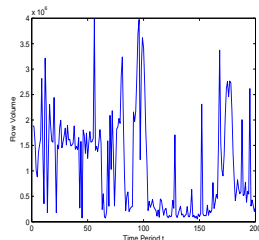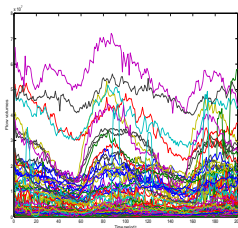
- Stochastic properties of flow volumes vary by the level of aggregation and time scales.
- Estimating traffic volumes is important for monitoring and provisioning such networks.

Aggregate Measurements
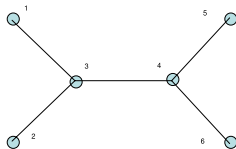
- Observations are made on edges which are a linear combination of the volumes corresponding to the flows passing through respective links.

-

$$Y_{(3,4)} = X_{(1,5)} + X_{(1,6)} + X_{(2,5)} + X_{(2,6)} + X_{(3,4)}$$

and

$$Y_{(4,3)} = X_{(5,1)} + X_{(6,1)} + X_{(5,2)} + X_{(6,2)} + X_{(4,3)}.$$

Aggregate Measurements

- In vector notation:

$$Y = AX,$$

where Y is a $L \times 1$ vector of observations on L edges, X is a $J \times 1$ vector of flow-volume variables associated with J flows and A is a $L \times J$ routing matrix where $[A]_{ij}$ indicates the fraction of the jth flow that traverses the ith link.

- If each origin-destination flow traverses exactly one path then A is binary.

- The matrix A is typically not full rank, as there are many more flows than links.

Some Interesting Extensions

- Multivariate Time Series
- Multimodal Measurements

Multivariate Time-Series Formulation

- Let $Y(t)$ denote the vector of observations on the links during measurement interval t.
- Let $X(t)$ be the (unobserved) vector of flow volumes in the same measurement interval.
- We will view $X(t)$ (and hence $Y(t)$) as random vectors satisfying some stochastic model.
- Thus, we have the following observation model:

$$Y(t) = AX(t), \quad t = 1, \cdots.$$

In this formulation the routing matrix A (typically not full rank) does not change over time.

- The distribution of $X(t)$ can be modeled at different levels of complexity from independent and identically (i.i.d.) Gaussian to long range dependent.

Multimodal Measurements

Suppose that

$$Y_P = AX_P, \quad Y_B = AX_B,$$

denote measurements on Packets and Bytes.

The two quantities can be related, for example, through a
compounding mechanism

$$X_B = \sum_{k=1}^{X_P} S_k,$$

where S_k denotes the payload in bytes of the $k-$th packet.

State assumptions and derive conditions
on the routing matrix A,
under which certain distributional parameters of X
are uniquely determined by the observable distribution of Y.

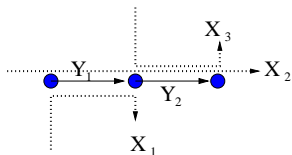# Identifiability: A Simple Example

FIGURE: Aggregate Volume Measurements

- Observations on links 1 and 2 are respectively given by

$$Y_1 = X_1 + X_2,$$
$$Y_2 = X_2 + X_3.$$

- Assume the flow volumes are uncorrelated.

Identifiability: A Simple Example

-

$$v_Y = \begin{pmatrix} \text{Var}(Y_1) \\ \text{Var}(Y_2) \\ \text{Cov}(Y_1, Y_2) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \text{Var}(X_1) \\ \text{Var}(X_2) \\ \text{Var}(X_3) \end{pmatrix} \equiv Bv_X.$$

  Thus, v_Y that contains the variances and the covariance of (Y_1, Y_2), uniquely determines v_X that contains the variances of X_1, X_2 and X_3, since B is a matrix of full rank.

- Now, the matrix B is clearly a function of the routing matrix A given by

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

- It can therefore be seen that "identifiability" of variances of the X_i's is related to a matrix function of A being full rank when the X_i's are uncorrelated.

IDENTIFIABILITY: DEFINITION

- The distribution of a J-dimensional random vector X is identifiable up to mean
  - under model $\mathcal{M}$,
  - from observations of the form $Y = AX$,
  - if for $Y_1 = AX_1$ and $Y_2 = AX_2$, $\mathcal{L}(X_1), \mathcal{L}(X_2) \in \mathcal{M}$,
  - $Y_1 \stackrel{d}{=} Y_2$ (i.e. $\mathcal{L}(Y_1) = \mathcal{L}(Y_2)$) implies that $X_1 \stackrel{d}{=} X_2 + c$ (i.e. $\mathcal{L}(X_1) = \mathcal{L}(X_2 + c)$) for some constant $c \in \mathbb{R}^J$.
- Generally first moments are not identifiable in this setting.
- Similarly, a parameter, $\theta(\mathcal{L}(X))$ is said to be identifiable under model $\mathcal{M}$ if $Y_1 \stackrel{d}{=} Y_2$ (i.e. $\mathcal{L}(Y_1) = \mathcal{L}(Y_2)$) implies that $\theta(\mathcal{L}(X_1)) = \theta(\mathcal{L}(X_2))$.

Identifiability Results

- For the case of independent flow volumes three kinds of identifiability results are known.
- These are conditions on the routing matrix under which flow volume variances are identifiable (Cao et al., 2000), conditions on the routing matrix under which entire flow volume distributions are identifiable up to mean (Chen et al., 2007) and conditions on the *routing policy* or network structure that imply that the routing matrix satisfies the required properties for identifiability (SM, 2007).
- We establish similar results for particular models of flow volume dependence.
- The techniques are naturally more involved and the independence case can be recovered as a special case.
- These results seek to address the question of "how complex can the dependence structure of a linear inverse problem be and still be identifiable".

A Useful Technical Tool

For a $L \times J$ matrix $A = [a_1, \cdots, a_J]$
and $M \times J$ matrix $B = [b_1, \cdots, b_J]$,
the $LM \times J$ Khatri-Rao product $A \odot B$ is defined as
$[a_1 \otimes b_1, \cdots, a_J \otimes b_J]$
where $\otimes$ denotes the Kronecker product.

Rows in $A \odot B$ are element-wise products of a row in A and a row in B.

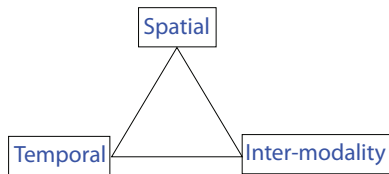### PROPOSITION

For X_j, $j = 1, \cdots, J$ independently distributed and whose characteristic functions are either analytic or possess no real roots, the distribution of X is identifiable up to mean from $Y = AX$, if $B = A \odot A$ has rank J.

- We will consider three types of dependence in flow volumes.

- Spatial dependence is the most challenging since it is in the spatial domain that the problem is ill-posed.

Models of Spatial Dependence

Given the ill-posedness nature of the problem, identifiability in the presence of dependence relies on some notion of 'sparsity' in the dependence structure.

Introduce three distinct, but related models, for which identifiability can be established.

A Covariance Model

Suppose that modeling dependence through second moments suffices.

Let $Cov(X) = V(\theta)$ with

$$V(\theta) = \theta_1 u_1 u_1' + \cdots + \theta_r u_r u_r'$$

with $U = [u_1, \cdots, u_r]$ assumed known.

Note that an arbitrary $J \times J$ covariance matrix can be modeled by using $U_J \equiv [I_J \ \ P]$, with P being a binary matrix of appropriate dimensions with distinct columns, each of which has exactly 2 non-zero entries.

In tomography applications, an interesting model will be based on a block diagonal $Cov(X)$ comprised of m blocks of size k. It can easily be modeled by

$$U = I_m \otimes U_k$$

Analogously, one can construct an increasing family of such models that capture hierarchical notions of spatial dependence.

Independent Components Model

Let $X = UZ$, with $Z_1, \cdots, Z_r$ being independent random variables with U assumed known as above.

The use of arbitrary distributions for $Z_1, \cdots, Z_r$ allows us to model dependencies of the distribution and not just through covariances.

When second moments of $Z_1, \cdots, Z_r$ exist, the covariance of X is given by $V(\theta)$. However the coefficients $\theta_1, \cdots, \theta_r$ are restricted to be positive and equal to the variances of $Z_1, \cdots, Z_r$.

Latent Variables Model

A latent variable model corresponding to a covariance model (defined by matrix U) is given by $X = CZ$, where $Z_1, \cdots, Z_J$ are independent random variables and $C \in \mathcal{C}(U)$.

If $C \in \mathcal{C}(U)$, then C is a lower triangular matrix with all diagonal entries equal to 1, such that for every vector $d \in \mathbb{R}_+^J$ there is a vector $\theta \in \mathbb{R}^r$ satisfying $CDiag(d)C' = V(\theta) \geq 0$.

When $V(\theta)$ is positive definite, then the Cholesky decomposition gives the corresponding unique coefficient matrix C.

A necessary condition for $\mathcal{C}(U)$ to be non-trivial is if U has rank J.

When U corresponds to a block diagonal covariance matrix then $\mathcal{C}(U)$ contains every matrix obtained from the Cholesky decomposition of $V(\theta)$ (for all $V(\theta) > 0$).

Main Result

Proposition

Given $Y = AX$, if $AU \odot AU$ has rank r then

1. *If $Cov(X)$ exists and is equal to $V(\theta)$ given by the **Covariance Model** then θ is identifiable from $Cov(Y)$.*

2. *If X satisfies the **Independent Components Model** with $Z_1, \cdots, Z_r$ such that either their characteristic functions are all analytic or all have no real roots, then the distributions of $Z_1, \cdots, Z_r$ are identifiable up to mean from the distribution of Y.*

3. *If U has rank J and X satisfies the **Latent Variable Model** with $Z_1, \cdots, Z_J$ all non-normal random variables such that either their characteristic functions are all analytic or all have no real roots, then the matrix C and distributions of $Z_1, \cdots, Z_J$ are identifiable up to mean from the distribution of Y.*
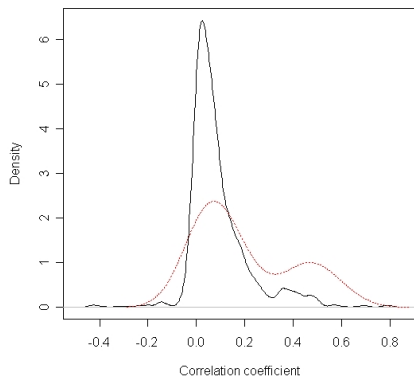
Some empirical facts:
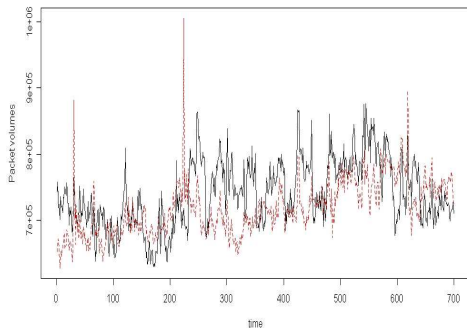
FIGURE: Densities of observed correlations: Forward-reverse (dashed) and the rest (solid)

Forward and Reverse Flows

- In real computer networks, a large part of the traffic is connection oriented.
- For example, traffic flows transported using the TCP protocol, or connections involving Internet (Voice over IP) telephony, lead to packets being exchanged between the two endpoints.
- Therefore, volumes of flow from node n_1 to node n_2 and vice-versa, are correlated.
- One of these flows is labeled as a forward flow and the other as a reverse flow and form a flow pair. It is reasonable to assume that flow pairs are independent with possible dependence between forward and reverse flows of a flow pair.

Independent Connections Model

-
$$Y^{(p)} = AX^{(p)},$$

  where $A = (A_F, A_R)$ and $X^{(p)} = (X_F^{(p)'}, X_R^{(p)'})'$.

- If second moments exist, then the covariance matrix of $X^{(p)}$ is of the form

$$\Sigma_X = \left(\begin{array}{cc} \text{Diag}(\delta_{FF}) & \text{Diag}(\delta_{FR}) \\ \text{Diag}(\delta_{FR}) & \text{Diag}(\delta_{RR}) \end{array} \right),$$

  where $\delta_{FF}, \delta_{FR}, \delta_{RR}$ correspond to the variances of $X_F^{(p)}$, covariances of $X_F^{(p)}$ and $X_R^{(p)}$ and variances of $X_R^{(p)}$, respectively.

- Can be extended to include time and byte modality.

Independent Connections Model

The independent connections model is obtained using

$$U = \begin{pmatrix} I_{J/2 \times J/2} & 0 & I_{J/2 \times J/2} \\ 0 & I_{J/2 \times J/2} & I_{J/2 \times J/2} \end{pmatrix}$$

In the case of Independent Connections Model $AU \odot AU$ having rank $r = 3J/2$ is equivalent to
$\overline{B}_c \equiv [A_F \odot A_F, A_F \odot A_R + A_R \odot A_F, A_R \odot A_R]$ having rank $r = 3J/2$.

This can be shown to be the case under reasonable conditions on routing and network structure.

Spatio-Temporal Dependence

- The covariance model can be easily extended to handle spatio-temporal dependence.
- For example: For the Independent Connections Model identifiability of the corresponding (spatio-temporal) covariance model, assuming only within flow temporal dependence, requires

$$\overline{B} = [A_F \odot A_F, A_R \odot A_R, A_F \odot A_R, A_R \odot A_F]$$

and

$$\overline{B}_c = [A_F \odot A_F, A_R \odot A_R, A_F \odot A_R + A_R \odot A_F]$$

to be full rank.

Multimodal Tomography

Suppose $Y_P = AX_P$ and $Y_B = AX_B$ with $(X_{Pj}, X_{Bj})'$, distributed independently for $j = 1, \cdots, J$.

Proposition

If

$$B = A \odot A$$

has rank J and the joint characteristic functions of $(X_{Pj}, X_{Bj})'$ are either analytic or have no roots in $\mathbb{R}^2$ for all $j = 1, \cdots, J$. Then the distribution of $(X_P', X_B')'$ is identifiable from (Y_P', Y_B') up to a mean ambiguity.

The above proposition can be easily extended to multimodal tomography and time dependence (treating time as another modality).

Multimodal Tomography: Compound Model

Suppose

1. $Prob(X_P \in \mathbb{N}) = 1$ and

2. The distribution of X_P is non-trivial i.e. there is no $n \in \mathbb{N}$ such that $Prob(X_P = n) = 1$ and

3.

$$X_B = \sum_{i=1}^{X_P} S_i$$

   where $X_P, S_1, S_2, \cdots$ are distributed independently and $S_1, S_2, \cdots$ are distributed identically and

4. The distribution of S_1 is non-trivial i.e. there is no $s \in \mathbb{R}$ such that $Prob(S_1 = s) = 1$.

IDENTIFIABILITY OF COMPOUND MODEL

- Under the conditions of the previous proposition and additionally the above assumptions of a *compound model* on each $(X_{Pj}, X_{Bj})'$ for $j = 1, \cdots, J$, the distribution of $(X_P', X_B')'$ is fully identifiable.

- This is the only model where the mean is also identifiable, "without making parametric/moment-relation assumptions".

Sufficient Conditions on Routing for Identifiability

- Recall matrices

$$\overline{B} = [A_F \odot A_F, A_R \odot A_R, A_F \odot A_R, A_R \odot A_F]$$

and

$$\overline{B}_c = [A_F \odot A_F, A_R \odot A_R, A_F \odot A_R + A_R \odot A_F]$$

and

$$B = A \odot A$$

- Require full-rankness of above for identifiability.
- Under what conditions on the routing discipline and network structure do we obtain full-rankness?

SUFFICIENT CONDITIONS ON ROUTING FOR IDENTIFIABILITY (SM07)

PROPOSITION

1. *Under balanced minimum weight routing on a symmetric graph, the matrix $\overline{B}$ (and hence $\overline{B}_c$) is full rank.*

2. *The matrix $\overline{B}$ is full rank for hierarchical networks.*

3. *For a directed acyclic graph, the matrix B is full rank.*

Concluding Remarks

- The above results show that under reasonable models of dependence second and higher order moments are estimable from the data.
- This can be leveraged for estimation of first moments of X, by imposing an appropriate relationship with higher order moments.
- Are there other more general models of spatio-temporal dependence?
- For the generic $Y = AX$ problem with A not being full rank and where X satisfies one of the posited dependence models, how can we design A so that the distribution of X becomes identifiable (up to mean shifts).