Understanding delay propagation through latent variable models
(see [1] for details)

Guillaume Dalle, Yohann De Castro
CERMICS – SNCF

“Decisions, Algorithms, Geometry” Seminar
October 14th 2021
Motivation
Overview

Delay propagation model

Minimax lower bound

Sparse estimator

Numerical results

Perspectives
Overview

Delay propagation model

Minimax lower bound

Sparse estimator

Numerical results

Perspectives
Context of the study

- Goal: a delay prediction model for the Paris suburban network
- Central challenge with many applications:
  - tactical scheduling
  - real-time regulation
  - passenger information
- Problem: we cannot consider trains individually because they interact through shared resources (infrastructure, driver, rolling stock, etc.)
What is delay propagation?

- Primary delay + resource conflicts $\Rightarrow$ propagated delays
- Standard methods assume knowledge of the conflicts to build an event dependency graph and propagate delays.
- Inadequate here because:
  - lack of data on microscopic resource use
  - unpredictable modifications in real time
- Our solution: implicit interactions carried by the railway network
A two-layer propagation model

- A directed graph $G = (V, E)$ for the railway lines and stations
- Latent congestion values $X_{t,e}$ for each time step $t$:

$$X_{t,e_2} = \sum_{e_2 \sim e_1} \theta_{e_1,e_2} X_{t-1,e_1} + \mathcal{N}(0, \sigma^2)$$

  new congestion value

  propagation from nearby edges

  noise

- Observed arrival times $A_{k,v}$ for each train $k$:

$$\underbrace{(A_{k,v} - A_{k,v}^{th})}_{\text{delay at } v} - \underbrace{(A_{k,u} - A_{k,u}^{th})}_{\text{delay at } u} = X_{\lfloor A_{k,u}/\Delta t\rfloor,(u,v)} + \mathcal{N}(0, \omega^2)$$

  congestion

  noise
Abstract formulation

- Latent congestion $X$: Vector AutoRegressive (VAR) process

$$X_t = \theta X_{t-1} + \varepsilon \quad \text{where} \quad \varepsilon \sim \mathcal{N}(0, \Sigma)$$

- Observed delay variation $Y = (A_{k,v} - A^{th}_{k,v}) - (A_{k,u} - A^{th}_{k,u})$: random projection $\Pi$ of $X$, where each component $(t,e)$ is present on average $p$ times

$$Y = \Pi X + \eta \quad \text{where} \quad \eta \sim \mathcal{N}(0, \omega^2 I)$$
Formalizing our assumptions

- Delay propagation is local: the transition matrix $\theta$ is $s$-sparse (row-wise) and bounded

$$\theta \in \Theta_s = \{ \theta \in \mathbb{R}^{D \times D} : \|\theta\|_2 \leq \vartheta < 1 \quad \text{and} \quad \forall i, \|\theta_i, \cdot\|_0 \leq s \}$$

- Traffic is dense: every edge is crossed often enough ($p \ll 1$)
- Temporal & spatial correlations are allowed in the projection $\Pi$
- Congestion plays a significant role:

$$\text{sp}(\Sigma) \subset [\sigma^2_{\min}, \sigma^2_{\max}] \quad \text{with} \quad \sigma_{\max} \sim \sigma_{\min} \gg \omega$$
Two main questions

1. How hard is it to estimate the transition matrix $\theta$ on a specific network?

2. Does this simple model give rise to interesting insights when applied to open railway data?
Overview

Delay propagation model

Minimax lower bound

Sparse estimator

Numerical results

Perspectives
Bounding the performance of any estimator

- We study the worst-case probability of error for the best possible estimator:

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{P}_{\theta}(\|\hat{\theta} - \theta\|_{\infty} \geq \delta)$$

- Using Fano’s method, we reduce the problem to finding a set of parameters $\theta_i$ for $0 \leq i \leq M$ that are both
  1. sufficiently far apart: $\max_{i \neq j} \|\theta_i - \theta_j\|_{\infty} \geq \alpha$
  2. hard to distinguish in terms of induced distributions:

$$\frac{1}{M+1} \sum_{i=1}^{M} \text{KL}(\mathbb{P}_{\theta_i} \| \mathbb{P}_{\theta_0}) \leq \beta$$
Our negative result

**Theorem: Minimax lower bound**

Let us define

$$\gamma_{\ell} = (1 - \vartheta)^{3/2} \frac{\sigma_{\min}^2 + \omega^2}{\sigma_{\max}^2}.$$

Then there is a universal constant $c$ such that

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta_s} \mathbb{P}_{\theta} \left( \|\hat{\theta} - \theta\|_{\infty} \geq c \frac{\gamma_{\ell}s}{p\sqrt{T}} \right) \geq \frac{1}{2}.$$
Key ingredients

**Lemma: Covariance decomposition**

The covariance matrix of $Y$ given $\Pi$ decomposes as the sum of a constant term and a residual: $\text{Cov}_\theta[Y|\Pi] = Q_\Pi + R_\Pi(\theta)$.

**Lemma: KL divergence between close Gaussians**

Let $\Delta$ be a symmetric matrix such that $\lambda_{\min}(\Delta) > -1$, and let $M$ be a matrix such that $MM' > 0$. Then

$$\text{KL} \left[ \mathcal{N}(\mu, M(I+\Delta)M') \| \mathcal{N}(\mu, MM') \right] \leq \frac{\|\Delta\|_F^2}{2(1 + \lambda_{\min}(\Delta))}.$$
Overview

Delay propagation model

Minimax lower bound

Sparse estimator

Numerical results

Perspectives
Autocovariance estimation

- The autocovariance of a VAR process satisfies
  \[
  \Gamma_h(\theta) = \text{Cov}_\theta[X_{t+h}, X_t] = \theta^h \Gamma_0(\theta)
  \]  
  (YW)

- We use a corrected plugin estimator:
  \[
  \hat{\Gamma}_h := \frac{1}{S(h)} \odot \left[ \frac{1}{T-h} \sum_{t=1}^{T-h} \left( \Pi_{t+h}^{+} Y_{t+h} \right) \left( \Pi_t^{+} Y_t \right)' \right] - C(h).
  \]
The Dantzig selector

- We look for a sparse approximate solution to \((YW)\). The first option would be the LASSO

\[
\hat{\theta} := \arg\min_{M \in \mathbb{R}^{D^2}} \| M \hat{\Gamma}_0 - \hat{\Gamma}_1 \|_2^2 + \lambda \| \text{vec}(M) \|_1
\]

- We use the Dantzig selector instead (computationally cheaper and theoretically easier)

\[
\hat{\theta} := \arg\min_{M \in \mathbb{R}^{D^2}} \| \text{vec}(M) \|_1 \quad \text{s.t.} \quad \| M \hat{\Gamma}_0 - \hat{\Gamma}_1 \|_{\max} \leq \lambda \quad (LP)
\]
Our positive result

**Theorem: Sparse estimator convergence**

Let us define

\[
\gamma_u(\theta) = \frac{\|\theta\|_\infty + 1}{(1 - \|\theta\|_2)^2} \frac{\sigma_{\text{max}}^2 + \omega^2}{\|\Gamma_0(\theta)^{-1}\|_1^{-1}}.
\]

Then there is a universal constant \(c\) such that, for the optimal value of \(\lambda\), with high probability,

\[
\|\hat{\theta} - \theta\|_\infty \leq c \frac{\gamma_u(\theta)s}{p\sqrt{T}} \sqrt{\log(D/\delta)}
\]
Key ingredients

**Lemma: Discrete concentration**

Let \( Z_t = \pi_{t+h,e_1} \pi_{t,e_2} \), where \( \pi_{t,e} \) is the number of activations of component \((t,e)\) in the observations. Then

\[
P\left( \left| \frac{1}{T-h} \sum_{t=1}^{T-h} Z_t - \mathbb{E}[Z] \right| \geq u \mathbb{E}[Z] \right) \leq c_1 \exp(-c_2 u^2 \mathbb{E}[Z]).
\]
Key ingredients (2)

**Lemma: Conditional Gaussian concentration**

Let $A$ be a random matrix satisfying $\|A\|_2 \leq M_2$ and $\|A\|_F^2 \leq M_F^2$ with probability $1 - \delta$. If $X \sim \mathcal{N}(0, I)$ and $X \perp A$, then

$$
\mathbb{P} \left( |X'AX - \mathbb{E}[X'AX]| \geq u \right) \leq \delta + 2 \exp \left( -c \min \left\{ \frac{u^2}{M_F^2}, \frac{u}{M_2} \right\} \right) \\
+ \mathbb{P} \left( |\text{Tr}(A - \mathbb{E}[A])| \geq u/2 \right)
$$
Comments on the bounds

- Coherent w.r.t. the role of dimension parameters $s, p, T$
- Influence of the noise $(1 + \omega^2/\sigma^2)$ related to Fisher information
- Correlated projections (especially Markov) introduce additional challenges:
  - need for custom concentration inequalities
  - mismatch between lower and upper bound for Markov sampling
## Overview

- Delay propagation model
- Minimax lower bound
- Sparse estimator
- Numerical results
- Perspectives

Understanding delay propagation through latent variable models
Role of the dimension

Figure 1: Influence of $D$ and $T$ on the error
Role of the noise

Figure 2: Influence of $\sigma$ and $\omega$ on the error
Role of the sparsity

Figure 3: Influence of $D$ on the error with fixed $s$
Data preprocessing

Figure 4: Map of the Zürich tram network generated from consecutive events
Results on Zürich tramway data

Figure 5: Effect of regularization and time discretization interval on some features of the estimate $\hat{\theta}$
Overview

Delay propagation model

Minimax lower bound

Sparse estimator

Numerical results

Perspectives
Improving the model

- Capture the congestion with a Gaussian Process to
  - Add nonlinearities
  - Switch to continuous time
- Consider external features when predicting arrival times
- Propose an EM-based learning procedure
Putting predictions to good use


Goal: define an efficient policy to minimize delays in spite of stochastic accidents.

Figure 6: An example of Flatland environment
Bibliography I
