Deconvolution of Mixing Time-Series on a Graph

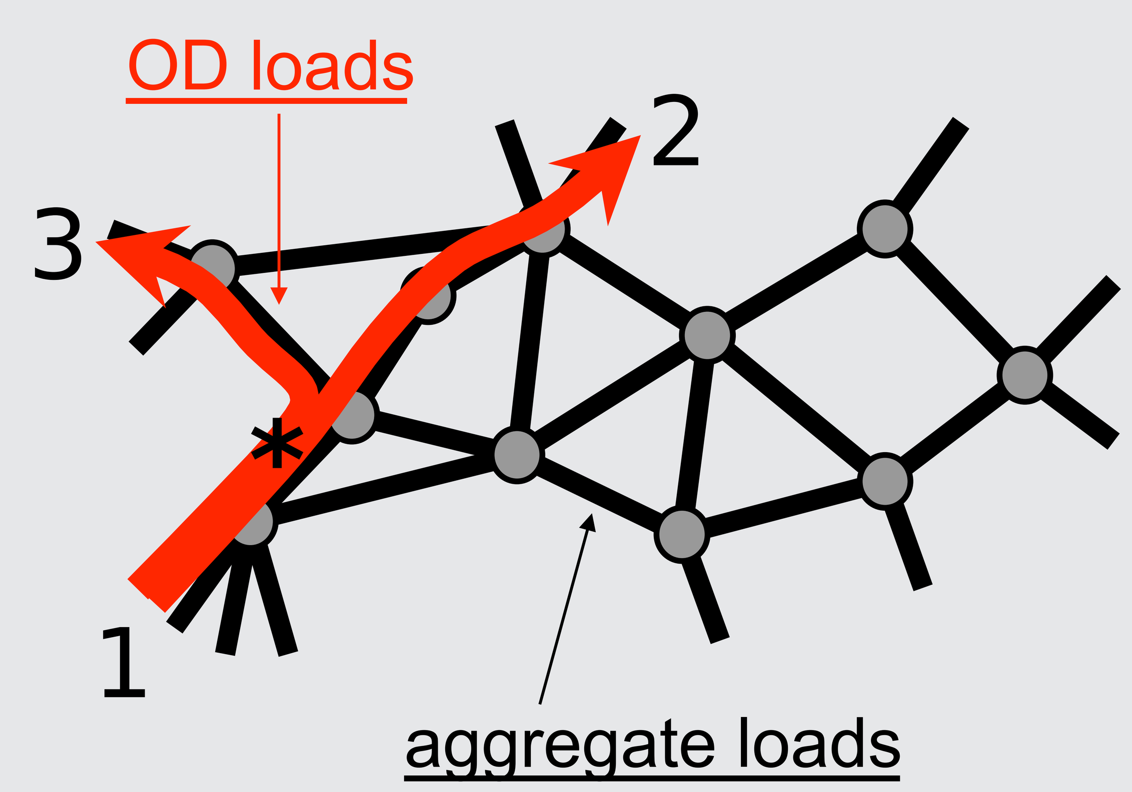
Alexander W. Blocker[1] & Edoardo M. Airoldi[1,2]

[1]Harvard University Department of Statistics, [2]Harvard FAS Center for Systems Biology

Problem

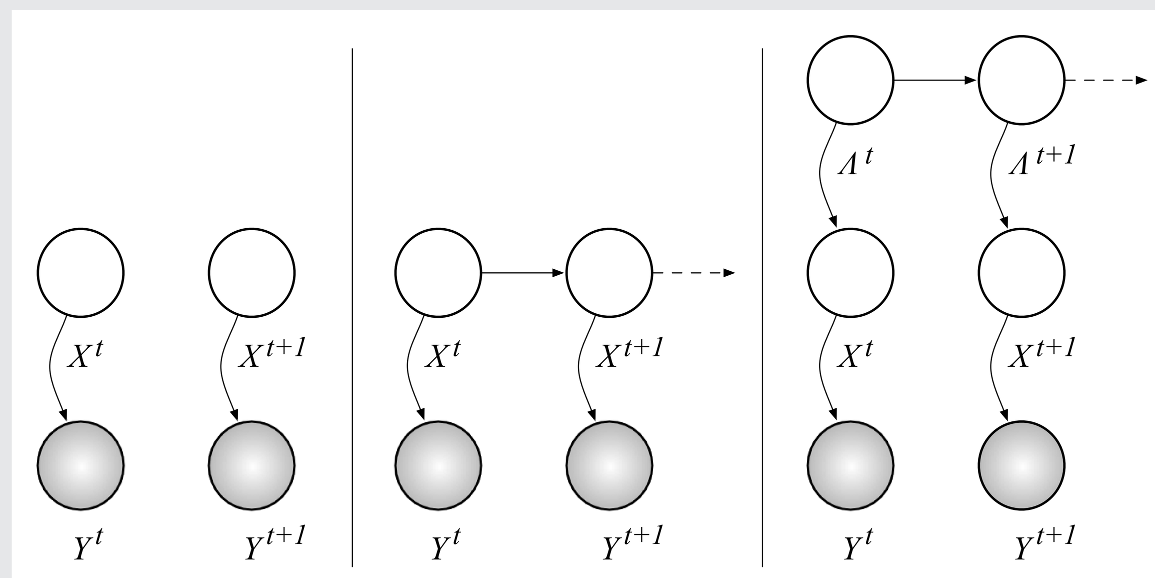
Interested in making inference on latent time series from indirect measurements — often low-dimensional projections resulting from mixing or aggregation

- Examples include positron emission tomography, super-resolution, and network traffic monitoring
- Inference in such settings requires solving a sequence of ill-posed inverse problems, $\boldsymbol{y}_t = A\boldsymbol{x}_t$, where the projection mechanism provides information on A
- We consider problems in which A specifies mixing on a graph of times series that are bursty and sparse.



Methods — Overview

- Sequence of ill-posed inverse problems $\boldsymbol{y}_t = A\boldsymbol{x}_t$
- Multilevel dynamic model with informative regularization
- Calibrating regularization with simpler method — data-driven, two-phase approach
- Efficient SMC inference on tightly-constrained spaces
- State-of-the-art performance on real networks & in large-scale simulation studies



Notation

- Latent origin-destination flows are $(x_{1,t}, \ldots, x_{m,t}) = \boldsymbol{x}_t$
- Observed aggregate flows are $(y_{1,t}, \ldots, y_{n,t}) = \boldsymbol{y}_t$
- Routing matrix A is $m \times n$; $\boldsymbol{y}_t = A\boldsymbol{x}_t$

Probability Model

Latent Intensities

- Introduce time-varying intensity $\lambda_{i,t}$ for each OD flow $x_{i,t}$
- Intensity evolves over time according to

$$\log \lambda_{i,t} = \rho \log \lambda_{i,t-1} + \varepsilon_{i,t}$$

- Process leads to bursty flows (few tightly cluster peaks) that are not sparse

Independent Variation

- Posit truncated normal model for $x_{i,t}|\lambda_{i,t}$

$$x_{i,t}|\lambda_{i,t}, \phi_t \sim \mathrm{TruncN}_{(0,\infty)}\left(\lambda_{i,t},\ \lambda_{i,t}^{\tau}(e^{\phi_t} - 1)\right)$$

- Structure approximates sparsity & retains computational stability by decoupling mass near 0 from bursty dynamic behavior

Posterior Updates & Multimodality

- Previous concerns about multimodality of posterior for this class of problem (e.g. Tebaldi and West 1998)
- Established that broad class of posterior updates lead to weakly unimodal (quasiconcave) posteriors for OD flows in continuous case
- Formally, quasiconcave predictive distribution $\rightarrow$ quasiconcave posterior on OD flows; former holds for many procedures

Regularization

Calibrate regularization parameters for generative model using a simpler model — analogous to approach of Clogg et al. (1991).

- Assuming $\boldsymbol{x}_t$ follows Gaussian autoregressive process; amounts to a standard Gaussian state-space formulation:

$$\begin{aligned} \boldsymbol{x}_t &= F \cdot \boldsymbol{x}_{t-1} + Q \cdot \boldsymbol{1} + \boldsymbol{e}_t \\ \boldsymbol{y}_t &= A \cdot \boldsymbol{x}_t + \boldsymbol{\epsilon}_t \end{aligned}$$

- Estimate $\boldsymbol{Q}$ and $\mathrm{Cov}\, \boldsymbol{e}_t$; fix $F = \rho \boldsymbol{I}$; $\mathrm{Cov}\, \boldsymbol{\epsilon}_t = \sigma^2 \boldsymbol{I}$
- Assume $Q = \mathrm{diag}\,(\boldsymbol{\lambda})_t$ and $\mathrm{Cov}\, \boldsymbol{e}_t = \boldsymbol{\Sigma}_t = \phi\, \mathrm{diag}\,(\boldsymbol{\lambda})_t^{\tau}$ — similar to Cao et al. (2000), adding explicit temporal dependence
- Maximum likelihood estimation of $\boldsymbol{\lambda}$ & ϕ within local windows
- Marginal likelihood calculation via standard Kalman filter

Regularization parameters set based on results of this estimation

- IPFP & smooth estimated OD flows with running median for feasibility & stability; obtain $\hat{\boldsymbol{x}}_t$
- $\theta_{1\,i,t} = \log \hat{x}_{i,t} - \log \hat{x}_{i,t-1}$
- $\theta_{2\,i,t} = (1 - \rho^2)\ \log(1 + \hat{V}_{i,t}/\hat{x}_{i,t}^2)$

Estimation & Inference

Using sequential Monte Carlo for estimation of probability model; sample-resample-move (SIRM) algorithm, akin to Gilks and Berzuini (2001)

- Each iteration begins with SIR particle filter update
- Particles then moved according to reversible kernel (MCMC iterations) for diversity
- Using "random directions algorithm" (RDA) of Smith (1984) to sample on constrained spaces
  - Given observed $\boldsymbol{y}_t = A\boldsymbol{x}_t$ and appropriate pivot of A, have $A = [A_1 A_2]$ where A_1 is $n \times n$ and full-rank
  - Pivoting and decomposing $\boldsymbol{x}_t$ analogously, have $\boldsymbol{x}_{1,t} = A_1^{-1}(\boldsymbol{y}_t - A_2\boldsymbol{x}_{2,t})$
  - Decomposition allows for efficient random-walk within feasible region for $\boldsymbol{x}_t$; fast computation of bounds along arbitrary directions
- Informative regularization guides computation — large decrease in particles needed
- Careful sampling on constrained spaces is vital — naïve alternatives can require 10^9+ draws for one feasible particle in small cases

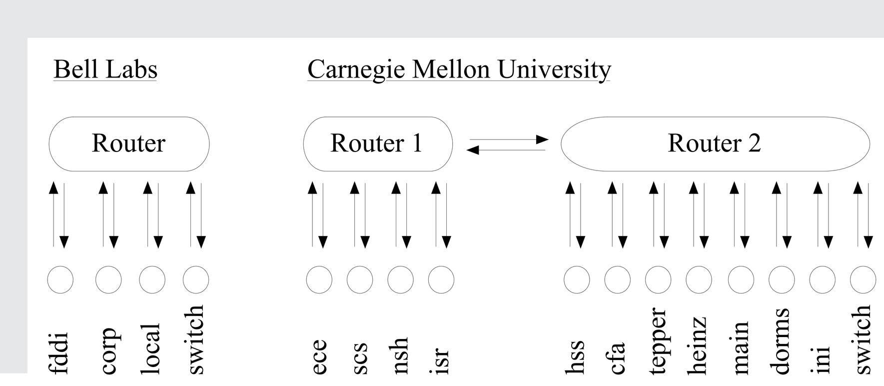
Results

Simulation Study

- Generated data from model on networks with latent of spaces between 2 and 9 dimensions
- Performed inference with informative and random-walk regularization
- Found consistent performance gains from informative regularization, especially in larger networks

Empirical Results

- Data from two real-world networks: Bell Labs (as in Cao et al. 2000) and CMU
- Compared locally IID method of Cao et al. (2000); MCMC method of Tebaldi and West (1998); our model with random-walk (naïve) regularization; our calibration model; & our model with informative regularization
- Our two-phase method showed excellent performance & stability with lower computational cost than MCMC



Performance Comparison

	BELL LABS			
Method	**L_2 Error**	**SE**	**L_1 Error**	**SE**
Locally IID model	104.59	5.54	160.24	6.53
Smoothed locally IID	104.25	5.52	157.87	6.48
Tebaldi & West (uniform prior)	76.60	4.91	173.94	7.49
Tebaldi & West (joint proposal)*	49.43	2.58	147.66	6.18
Dynamic multilevel model (naïve)	63.29	3.35	178.43	8.09
Calibration model (stage 1)	19.35	0.72	57.66	2.06
Dynamic multilevel model (stage 2)	19.93	0.87	58.20	2.39

	CMU			
Method	**L_2 Error**	**SE**	**L_1 Error**	**SE**
Locally IID model	592.49	9.91	1169.15	17.11
Smoothed locally IID	—	—	—	—
Tebaldi & West (uniform prior)	—	—	—	—
Tebaldi & West (joint proposal)*	167.94	4.42	712.37	14.68
Dynamic multilevel model (naïve)	311.21	6.25	1109.68	19.58
Calibration model (stage 1)	110.47	6.19	389.14	16.72
Dynamic multilevel model (stage 2)	93.42	5.20	334.74	13.64

Remarks

- Ill-posed inverse problems arise often in network applications
- Improving deconvolution performance by modeling key features of time series (burstiness, skewed marginals, etc.)
- Model-based regularization strategy improves upon non-informative priors
- Extensive validation of approach on both simulated and actual data
- Scalable technique – computation scales linearly in length of series and in $m - n$
- Two-phase approach to inference leverages power of probability modeling without many of its drawbacks

References

J. Cao, D. Davis, S. Van Der Viel, and B. Yu. Time-varying network tomography: router link data. *Journal of the American Statistical Association*, 95:1063–75, 2000.

C. C. Clogg et al. Multiple imputation of industry and occupation codes in census public-use samples using Bayesian logistic regression. *Journal of the American Statistical Association*, 86(413):68–78, 1991.

W.R. Gilks and C. Berzuini. Following a Moving Target-Monte Carlo Inference for Dynamic Bayesian Models. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 63(1):127–146, 2001.

X. L. Meng. Machine learning with human intelligence: Principled corner cutting (pc^2). Plenary Invited Talk, Annual Conference on Neural Information Processing Systems (NIPS), December 2010

R. L. Smith. Efficient monte carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research*, 32(6):pp. 1296–1308, 1984.

C. Tebaldi and M. West. Bayesian inference on network traffic using link count data. *Journal of the American Statistical Association*, 93(442):557–573, 1998.

Thanks to: AT&T, Facebook, Google, CDC. We would like to thank C Faloutsos, S Fienberg, E Xing, D Blei, A Goldenberg, A Zheng, Z Ghahramani, J Leskovec, XL Meng, D Rubin, J Liu, S Kou, and J Blitzstein. We would also like to thank our anonymous reviewers for their useful feedback.