Reconstructing the power grid dynamic model from sparse measurements

Andrey Lokhov

with Michael Chertkov, Deepjyoti Deka, Sidhant Misra, Marc Vuffray

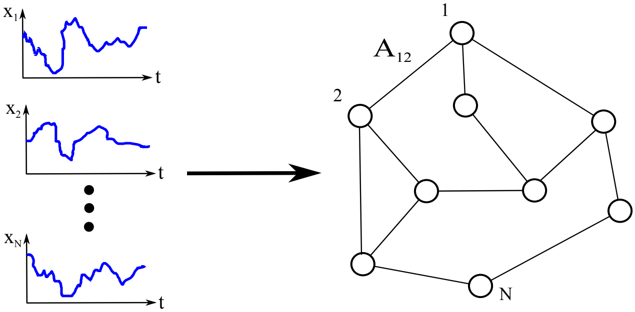
Los Alamos National Laboratory

Banff, Canada

Motivation: learning from correlated samples, or time series

Assume **linear dynamics**: $\dot{X}(t) = AX(t) + \xi(t)$, with $\xi(t)$ Gaussian noise

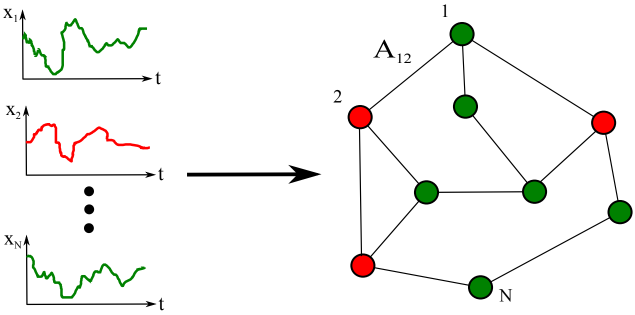
Given N time series, is it possible to reconstruct the structure and parameters of A?



Motivation: learning from correlated samples, or time series

Assume **linear dynamics**: $\dot{X}(t) = AX(t) + \xi(t)$, with $\xi(t)$ Gaussian noise

Given N time series, is it possible to reconstruct the structure and parameters of A?



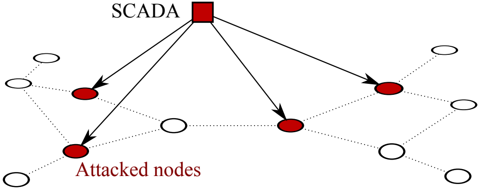
What happens if only $N_{\mathcal{O}} < N$ time series are observed? $X = X_{\mathcal{O}} \cup X_{\mathcal{H}}$

Example: graph-based anomaly detection in cyber-physical systems

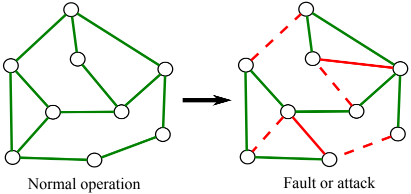
Task: detect and localize attacks on CPS using physical measurements

Smart Factories & Industry   Critical Infrastructures & Smart Grid

Self-Driving Cars and Avionics   Complex Transportation



Approach: assuming linearized dynamics, learn the normal graph and monitor changes



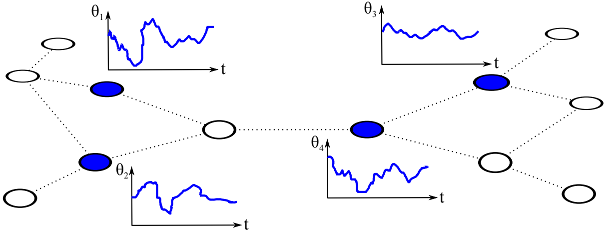
Setting: structure unknown, usually no hidden nodes

Example: reconstructing the power grid dynamics

State estimation and parameter learning in dynamics of the transmission power grid

$$\dot{\theta}_i = f_i, \qquad M_i \dot{f}_i + \tau_i f_i = p_i - \sum_{j \sim i} \beta_{ij}(\theta_i - \theta_j) + \xi_i(t).$$

Task: reconstruct parameters of generators and lines (evolve slowly, $\sim$ hours) and injections and consumptions (evolve rapidly, $\sim$ minutes) from sensor measurements



Setting: structure known, but hidden observations (sparsely located PMUs)

Reduction to the static problem?

For stable systems: explore Lyapunov equation for the stationary covariance matrix

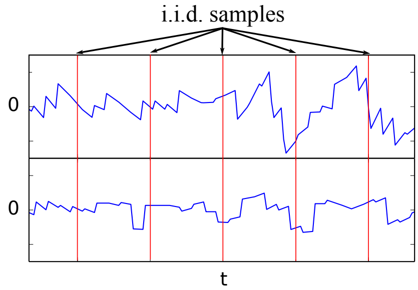
$$A\Sigma + \Sigma A^\top + I = 0$$

[Wang, Bialek, Turitsyn 2015], [Zare, Jovanović, Georgiou 2016]

Disadvantages: requires knowledge of some part of A, hard to generalize to hidden case

Subsampling independent samples:
use static Gaussian graphical model learning

Disadvantages: only stationary regime,
wasting samples (desiring $\sim \log N$ samples),
Σ has less information ($\text{supp}(A) \neq \text{supp}(\Sigma^{-1})$)



In what follows

For simplicity, consider **discrete-time** dynamics: $X_{t+1} = AX_t + \xi_t$, with ξ_t white noise*

✓ Complete observations on all nodes

(a) Known graph structure: least-squares objective

(b) Unknown graph: ℓ_1 and ℓ_0 regularizations

✓ Partially observed system

(a) Known graph: convex formulation, incomplete solution

(b) Unknown graph: sparsity and low-rank regularizations

(c) Non-convex EM-type algorithm

* **Remark: Intuitively and rigorously [Bento et al., 2010], in the case of continuous equations, there exists an optimal discretization step Δt**

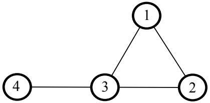
Complete observations: known graph structure

Assuming the uniform prior on A, $P(A \mid X, \xi) \propto \exp(-\sum_{t=1}^{T-1} \|X_{t+1} - AX_t\|^2 / 2\sigma^2)$

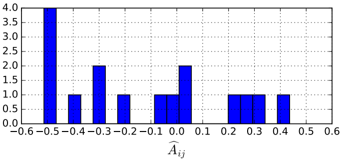
$$\hat{A}_{\text{MMSE}} = \hat{A}_{\text{MAP}} = \underset{A}{\operatorname{argmin}} \sum_{t=1}^{T-1} \|X_{t+1} - AX_t\|^2$$

For a sufficient number of samples $M \propto N$,

$$\hat{A} = \left(\sum_{t=1}^{T-1} X_{t+1} X_t^\top\right) \left(\sum_{t=1}^{T} X_t X_t^\top\right) = \Sigma_{t,t+1} \left(\Sigma_{t,t}\right)^{-1}$$



$$A = \begin{bmatrix} -0.5 & -0.2 & -0.3 & 0 \\ 0.2 & -0.5 & -0.3 & 0 \\ 0.3 & 0.3 & -0.5 & -0.4 \\ 0 & 0 & 0.4 & -0.5 \end{bmatrix}$$



Unknown graph and high-dimensional regime

Regularized least-squares: [Bento, Ibrahimi, Montanari 2010]

$$\hat{A} = \underset{A}{\operatorname{argmin}} \left(\sum_{t=1}^{T-1} \|X_{t+1} - AX_t\|^2 + \lambda \|A\|_1 \right)$$

Reconstructs graph structure with $M \propto \log N$ samples under incoherence condition and assumptions on $(\lambda_{\min}, \lambda_{\max})$ of covariance matrix

Open question: *similarly to the Gaussian GM selection, assumptions-free algorithm?*
Candidate: non-convex ℓ_0 sparsity constraint [Misra, Vuffray, AL, Chertkov 2017]

Partial observations: convex formulation

Likelihood of observations:

$$P(\tilde{A}_{\mathcal{O}} \mid X, \xi) = \int_{X_{\mathcal{H}}} dX_{\mathcal{H}} P(A \mid X, \xi), \qquad \tilde{A}_{\mathcal{O}} = A_{\mathcal{O}} - A_{\mathcal{O}\mathcal{H}} A_{\mathcal{H}}^{-1} A_{\mathcal{H}\mathcal{O}} \equiv A_{\mathcal{O}} + L$$

Partial observations: convex formulation

Likelihood of observations:

$$P(\tilde{A}_{\mathcal{O}} \mid X, \xi) = \int_{X_{\mathcal{H}}} dX_{\mathcal{H}} P(A \mid X, \xi), \qquad \tilde{A}_{\mathcal{O}} = A_{\mathcal{O}} - A_{\mathcal{O}\mathcal{H}} A_{\mathcal{H}}^{-1} A_{\mathcal{H}\mathcal{O}} \equiv A_{\mathcal{O}} + L$$

Leads to a convex "Lasso" type formulation for small $|\mathcal{H}|$:

$$(\hat{A}_{\mathcal{O}}, \hat{L}) = \underset{A_{\mathcal{O}}, L}{\operatorname{argmin}} \left[\sum_{t=1}^{T-1} \|X_{t+1}^{\mathcal{O}} - (A_{\mathcal{O}} + L) X_t^{\mathcal{O}}\|^2 + \lambda_1 \|A_{\mathcal{O}}\|_1 + \lambda_2 \|L\|_* \right]$$

Adaptation of [Giraud and Tsybakov 2012], [Jalali, Sanghavi 2012]

$M \propto \log N$ under incoherence assumption. If the graph is known, one could further attempt to decompose the matrix L into sparse factors, see e.g. [Witten, Tibshirani, Hastie 2009].

Partial observations: convex formulation

Likelihood of observations:

$$P(\tilde{A}_{\mathcal{O}} \mid X, \xi) = \int_{X_{\mathcal{H}}} dX_{\mathcal{H}} P(A \mid X, \xi), \qquad \tilde{A}_{\mathcal{O}} = A_{\mathcal{O}} - A_{\mathcal{O}\mathcal{H}} A_{\mathcal{H}}^{-1} A_{\mathcal{H}\mathcal{O}} \equiv A_{\mathcal{O}} + L$$

Leads to a convex "Lasso" type formulation for small $|\mathcal{H}|$:

$$(\hat{A}_{\mathcal{O}}, \hat{L}) = \operatorname*{argmin}_{A_{\mathcal{O}}, L} \left[\sum_{t=1}^{T-1} \|X_{t+1}^{\mathcal{O}} - (A_{\mathcal{O}} + L) X_t^{\mathcal{O}}\|^2 + \lambda_1 \|A_{\mathcal{O}}\|_1 + \lambda_2 \|L\|_* \right]$$

Adaptation of [Giraud and Tsybakov 2012], [Jalali, Sanghavi 2012]

$M \propto \log N$ under incoherence assumption. If the graph is known, one could further attempt to decompose the matrix L into sparse factors, see e.g. [Witten, Tibshirani, Hastie 2009].

Open question: *is it possible to devise assumptions-free algorithm?*
Candidate: non-convex explicit rank constraint $\text{rank}(L) \leq |\mathcal{H}|$ [Yuan & Lauritzen, Meinshausen 2012] together with an ℓ_0 sparsity constraint [Misra, Vuffray, AL, Chertkov 2017]

Partial observations: alternative convex formulation

Likelihood $P(A \mid X, \xi)$ can be rewritten in the static form over the trajectories:

$$P(A \mid X, \xi) \propto \sqrt{\det B} \exp\left(-\vec{X}^\top B \vec{X}\right), \qquad \vec{X} \equiv [X_{t=1}, \ldots, X_{t=T}],$$

$$B = \begin{bmatrix} A^\top A & A & & & & \\ A^\top & 1 + A^\top A & A & & & \\ & A^\top & 1 + A^\top A & A & & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \\ & & & 1 + A^\top A & A & \\ & & & A^\top & 1 & \end{bmatrix}$$

Partial observations: alternative convex formulation

Likelihood $P(A \mid X, \xi)$ can be rewritten in the static form over the trajectories:

$$P(A \mid X, \xi) \propto \sqrt{\det B} \exp\left(-\vec{X}^\top B \vec{X}\right), \qquad \vec{X} \equiv [X_{t=1}, \ldots, X_{t=T}],$$

$$B = \begin{bmatrix} A^\top A & A & & & \\ A^\top & 1 + A^\top A & A & & \\ & A^\top & 1 + A^\top A & A & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & & & 1 + A^\top A & A \\ & & & A^\top & 1 \end{bmatrix}$$

Likelihood of observations:

$$P(\tilde{B}_\mathcal{O} \mid X, \xi) = \int_{X_\mathcal{H}} dX_\mathcal{H} P(B \mid X, \xi), \qquad \tilde{B}_\mathcal{O} = B_\mathcal{O} - B_{\mathcal{O}\mathcal{H}} B_\mathcal{H}^{-1} B_{\mathcal{H}\mathcal{O}}$$

Leads to a "Graph Lasso" type convex formulation for $B_\mathcal{O} - L \succ 0$ and $L \succeq 0$:

$$(\hat{B}_\mathcal{O}, \hat{L}) = \underset{B_\mathcal{O}, L}{\operatorname{argmin}} \left[\operatorname{tr}(\Sigma(B_\mathcal{O} - L)) - \log\det(B_\mathcal{O} - L) + \lambda_1 \|B_\mathcal{O}\|_1 + \lambda_2 \operatorname{tr}(L) \right]$$

Adaptation of [Chandrasekaran, Parillo and Willsky 2012]

Partial observations: Expectation-Maximization approach

Given initial guess $A^{(s=0)}$, iterate until convergence:

Expectation: compute $Q(A, A^{(s)}) = \mathbb{E}\left[P(A \mid X_{\mathcal{O}} \cup X_{\mathcal{H}}, \xi) \mid X_{\mathcal{O}}, A^{(s)}\right]$

Maximization: update $A^{(s+1)} = \underset{A}{\operatorname{argmax}}\, Q(A, A^{(s)})$

The closest reference [Shumway, Stoffer 1982]

Not widely considered (hard to analyse), but natural choice if the graph is known

Path forward

- ✓ Theoretical analysis of the algorithms
- ✓ Establishing best algorithms in practice (using modern solvers, EM)
- ✓ Application to the power grid and cyberphysical data sets