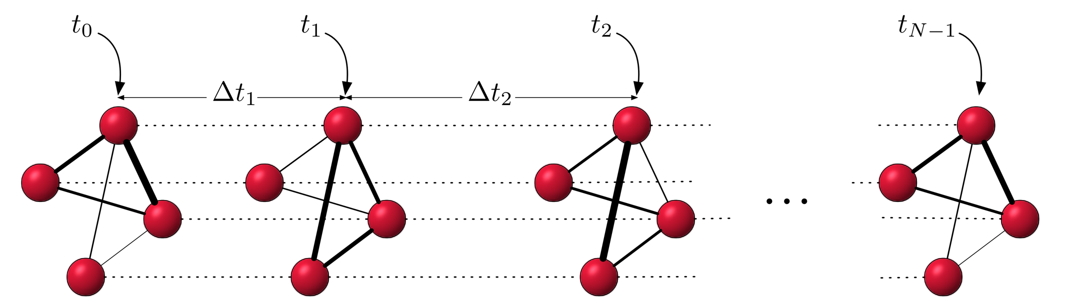
Fast Biological Network Reconstruction from High-dimensional Time-course Perturbation Data Using Sparse Multivariate Gaussian Processes

Sara Al-Sayed[⋆] and Heinz Koeppl[⋆†]

[⋆]Department of Electrical Engineering and [†]Department of Biology, Technische Universität Darmstadt, Germany

1. Introduction

Time-course data observed under the perturbation of biological systems contain rich information about the salient structure of interconnectivity among the entities of the network underlying the system.



Challenges: few noisy high-dimensional measurements at non-uniformly–spaced intervals; missing data; and computational complexity of inference, parameter estimation, and sparse structure search.

2. Approach

Linear Time-invariant Stochastic Differential Equation Model

Consider an undirected network of P entities described by adjacency matrix A, with trajectories $x_1(t), \ldots, x_P(t)$:

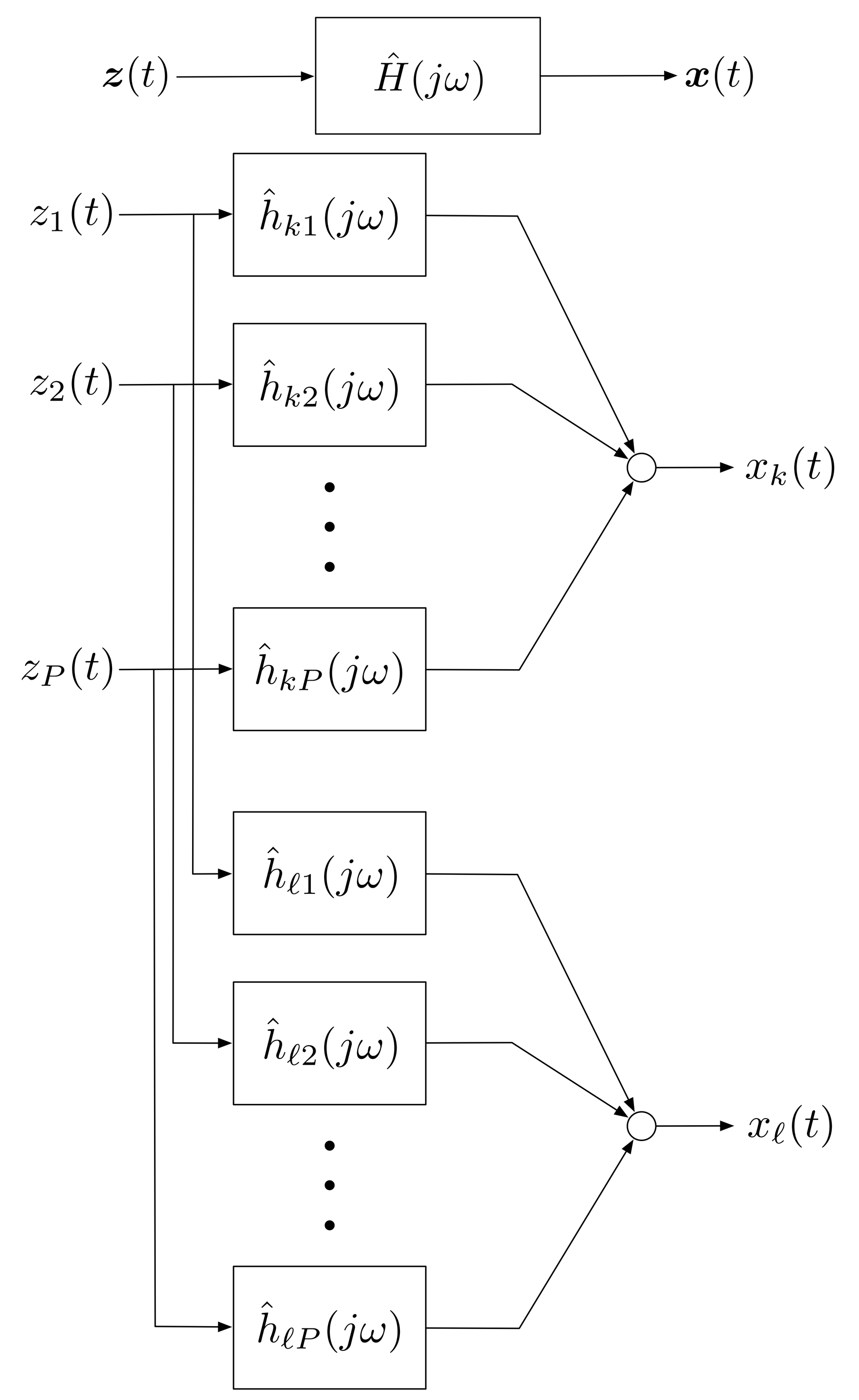
$$\frac{d^n \boldsymbol{x}(t)}{dt^n} + B_{n-1}\frac{d^{n-1}\boldsymbol{x}(t)}{dt^{n-1}} + \ldots + B_1\frac{d\boldsymbol{x}(t)}{dt} + B_0\boldsymbol{x}(t) = \boldsymbol{z}(t)$$

- $\boldsymbol{x}(t) \triangleq [x_1(t), \ldots, x_P(t)]^T$.
- $\boldsymbol{z}(t) \triangleq [z_1(t), \ldots, z_P(t)]^T$, mutually independent zero-mean white Gaussian noise processes with powers $\left\{\sigma^2_{z,p}\right\}$.
- $B_{n'} = \left[b^{(k\ell)}_{n'}\right]$, $n' = 0, 1, \ldots, n-1$; $k, \ell = 1, \ldots, P$: coupling matrices. **Assumption:** $\mathrm{supp}\,(B_{n'}) = \mathrm{supp}\,(A+I)\ \forall n'$.

State-space Representation:

$$\frac{d\tilde{\boldsymbol{x}}(t)}{dt} = \tilde{B}\tilde{\boldsymbol{x}}(t) + L\boldsymbol{z}(t)$$

Multivariate Gaussian Process Covariance Functions



For decoupled entities, Matérn covariance function:

$$\hat{h}_{k\ell}(j\omega) = \begin{cases} \frac{1}{b_0^{(k\ell)} + b_1^{(k\ell)}(j\omega) + \ldots + b_{n-1}^{(k\ell)}(j\omega)^{n-1} + (j\omega)^n} \overset{!}{=} \frac{1}{(\lambda + j\omega)^n} & , \ \ell = k \\ 0 & , \ \text{o.w.} \end{cases}$$

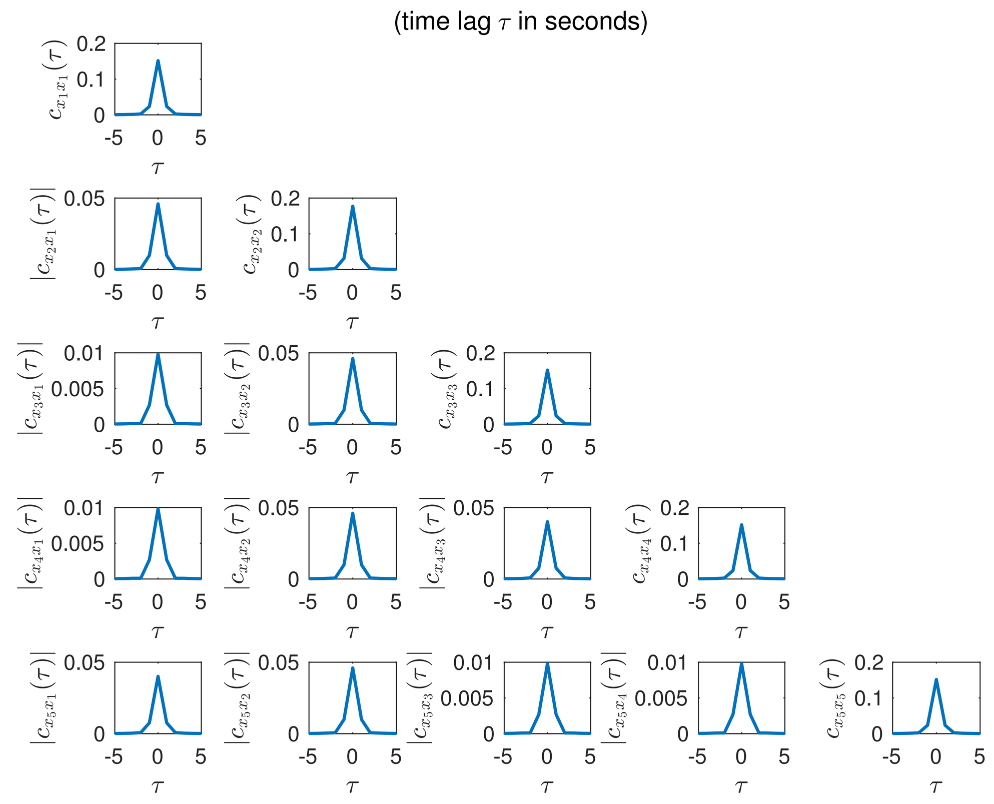
where $\lambda \triangleq \frac{\sqrt{2\nu}}{l}$; $\nu \triangleq n - \frac{1}{2}$, l: smoothness and length-scale parameters, respectively. Consider then the following design:

$$\hat{H}(j\omega) = \left(\begin{bmatrix} (\lambda + j\omega)^n & (\lambda + j\omega)^{n-1} & \cdots \\ (\lambda + j\omega)^{n-1} & \ddots & \\ \vdots & & (\lambda + j\omega)^n \end{bmatrix} \odot (A + I)\right)^{-1}$$

with $\sigma_z^2 = \frac{2\sigma^2\sqrt{\pi}\lambda^{2n-1}\Gamma(n)}{P\Gamma(2\nu)}\ \forall p$; σ^2: magnitude parameter.

Example:

$$A = \begin{bmatrix} 0&1&0&0&1 \\ 1&0&1&1&1 \\ 0&1&0&1&0 \\ 0&1&1&0&0 \\ 1&1&0&0&0 \end{bmatrix}, \quad n = 2, \lambda = 5, \sigma^2 = 1$$



Weighted Interactions

Consider $\Omega = [w_{k\ell}]$ that is symmetric, strictly diagonally dominant, with $w_{k\ell} = 1$ if $\ell = k$, and 0 if entities k, ℓ are disconnected.

$\Longrightarrow$ If the system is stable under $\tilde{B}$, it is also stable under $\tilde{B}_\Omega \triangleq \tilde{B} \odot (\Omega \otimes \mathbb{1}_{n\times n})$. The corresponding trajectories are then described by

$$\frac{d^n\boldsymbol{x}(t)}{dt^n} + \Omega \odot \sum_{n'=0}^{n-1} B_{n'}\frac{d^{n'}\boldsymbol{x}(t)}{dt^{n'}} = \boldsymbol{z}(t)$$

Discrete-time State-space Model

For measurement time t_i, $i = 0, \ldots, N-1$, $\boldsymbol{x}_i \equiv \tilde{\boldsymbol{x}}(t_i)$.

Process Equation:

$$\boldsymbol{x}_{i+1} = F_i\boldsymbol{x}_i + \boldsymbol{q}_i$$

- $F_i = \exp\left(\tilde{B}_\Omega \Delta t_{i+1}\right)$, $\Delta t_{i+1} \triangleq t_{i+1} - t_i$.
- temporally independent process noise $\boldsymbol{q}_i \sim \mathcal{N}(\boldsymbol{0}, Q_i)$:

$$Q_i = \int_0^{\Delta t_{i+1}} \exp\left[\tilde{B}_\Omega(\Delta t_{i+1} - \tau)\right] L\Sigma_z L^T \exp\left[\tilde{B}_\Omega(\Delta t_{i+1} - \tau)\right]^T d\tau$$

with $\Sigma_z = \sigma_z^2 I$.

- $\boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{0}, \Pi_\infty)$, where Π_∞ solves the continuous Lyapunov equation of the continuous-time process model.

Measurement Equation:

$$\boldsymbol{y}_i = H\boldsymbol{x}_i + \boldsymbol{\varepsilon}_i$$

where H selects $\left\{x_p(t_i)\right\}$ from $\tilde{\boldsymbol{x}}(t_i)$, and $\{\boldsymbol{\varepsilon}_i\}$ are i.i.d. measurement noise, $\sim \mathcal{N}(\boldsymbol{0}, \sigma_\varepsilon^2 I)$, independent of process noise.

Structure Scoring

Structure candidates are scored according to their *a posteriori* probability in terms of the conditional data likelihood:

$$p\left(\boldsymbol{y}_{0:N-1} \,|\, A, \hat{\Theta}\right) = \Pi_{i=0}^{N-1}\mathcal{N}(\boldsymbol{y}_i; H\hat{\boldsymbol{x}}_{i|i-1}, HP_{i|i-1}H^T + \sigma_\varepsilon^2 I)$$

where $\hat{\boldsymbol{x}}_{i|i-1}$ and $P_{i|i-1}$ are discrete-time Kalman filter *a posteriori* ith state mean and covariance matrices, respectively, given the structure candidate A and the estimated (hyper)parameters Θ.

Unknown (hyper)parameters Θ (interaction weights, noise variances, and covariance function parameters) are estimated jointly with the Kalman filter procedure in either a maximum-likelihood or a maximum *a posteriori* sense.

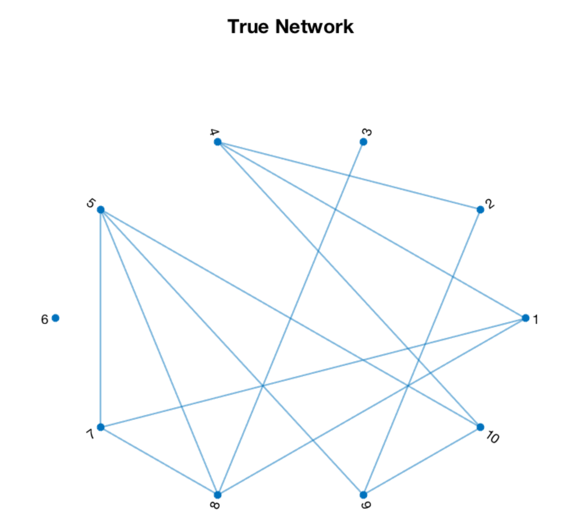
Structure Search and Prior Network Information

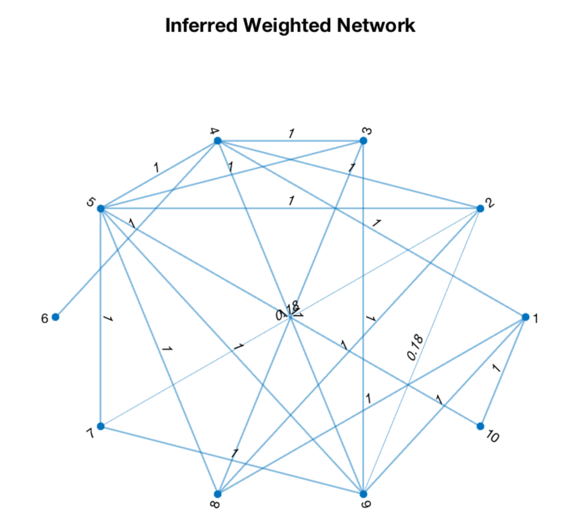
- Greedy (local hill-climbing) structure search algorithm variants are used to find a locally optimal structure.
- The availability of a sparsity prior and/or a prior network drawn from literature databases can help inform the choice of initial structures, as well as guide the neighbor search.

3. Results and Discussion

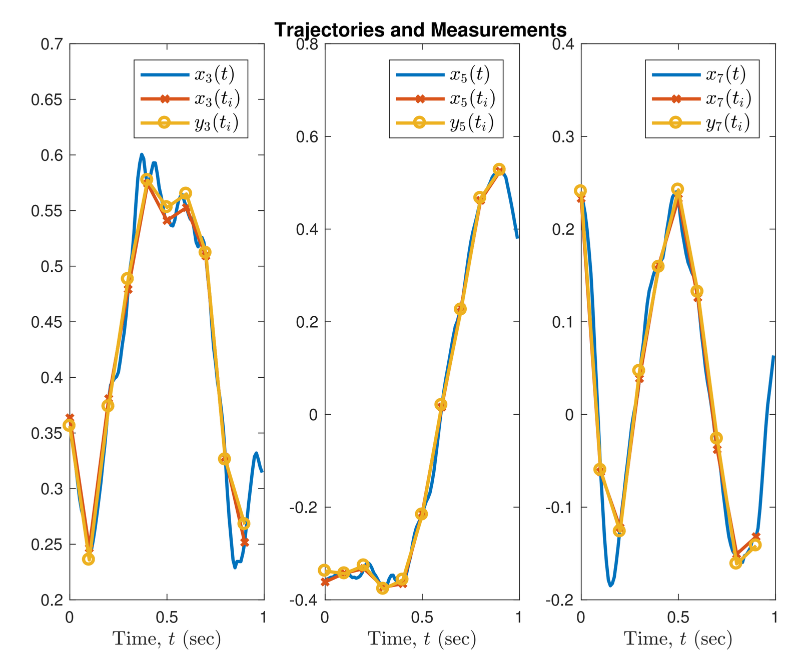
Synthetic Experiment

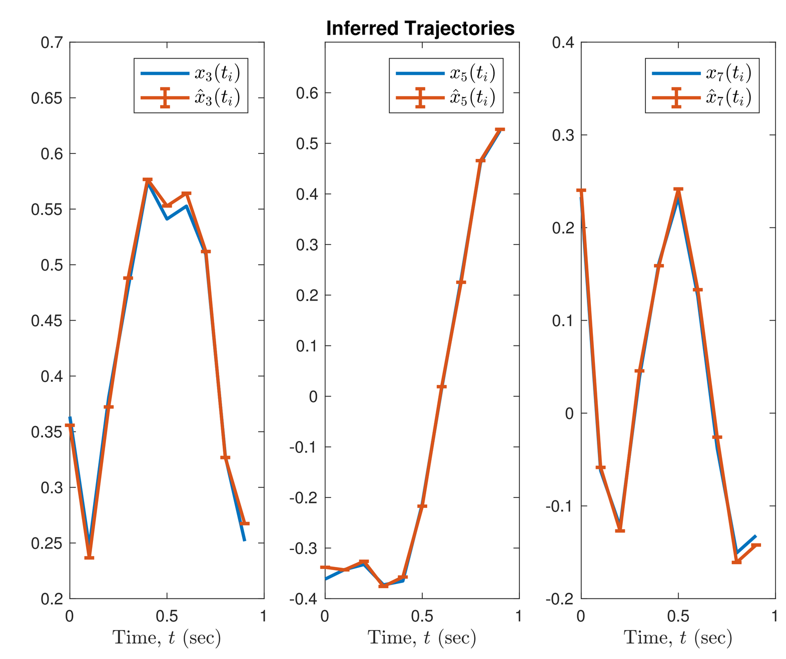
- network size $P = 10$
- stochastic differential equation degree $n = 2$
- smoothness parameter $\lambda = 5$, magnitude parameter $\sigma^2 = 1$
- true network structure: 0.3-edge probability
- binary interactions
- measurement noise variance $\sigma_\varepsilon^2 = 10^{-4}$
- greedy structure search algorithm with 0.3-edge probability initialization





Threshold ρ	TP	FP
$0 \le \rho < 0.18$	0.69	0.38
$0.18 \le \rho < 1$	0.62	0.34





This work was funded by the European Union's Horizon 2020 research and innovation programme under grant agreement 668858.