Compressible State-Space Models: Observability, Estimation and Application to Signal Deconvolution

Abbas Kazemipour[1], Ji Liu[2], Min Wu[1], Patrick Kanold[2] and Behtash Babadi[1]

[1]Department of Electrical and Computer Engineering

[2]Department of Biology

University of Maryland, College Park

2016 IEEE Global Conference on Signal and Information Processing

UNIVERSITY OF MARYLAND

UMIACS University of Maryland Institute for Advanced Computer Studies

The Institute for Systems Research

Introduction: Compressive Sensing

- So, how much data do you need?

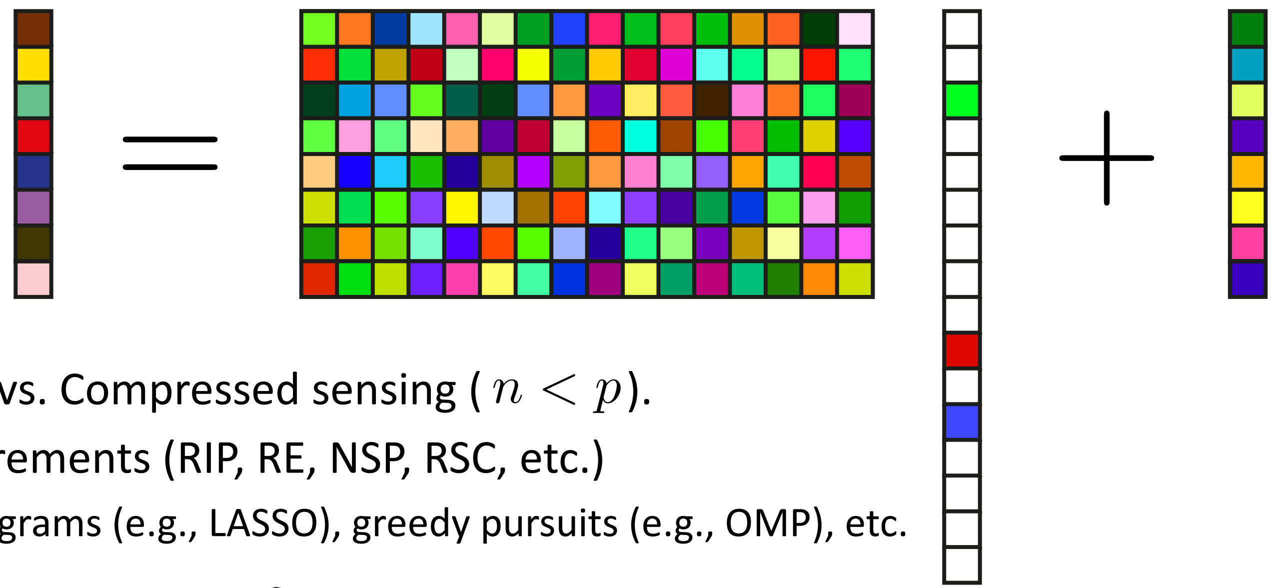
➢ Shannon-Nyquist sampling theorem (1928-1949)

➢ Compressed sensing (2005-present): beyond least squares.

Linear Model

$$s - \text{sparse}, \quad s \ll p$$

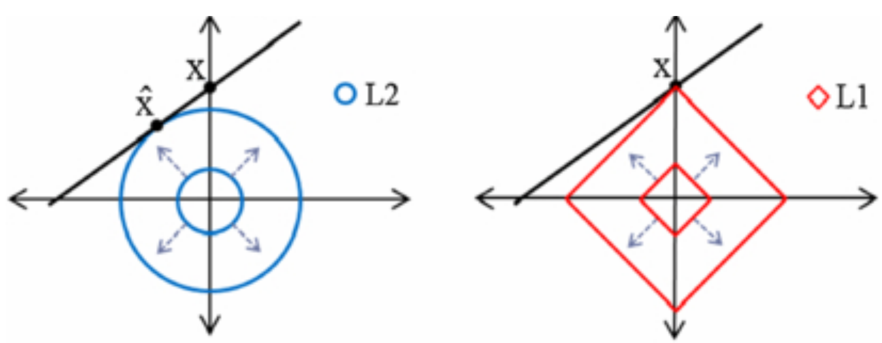
$$\mathbf{y} \in \mathbb{R}^n = \mathbf{X} \in \mathbb{R}^{n \times p} \; \boldsymbol{\theta} \in \mathbb{R}^p + \mathbf{z} \in \mathbb{R}^n$$



➢ Denoising ($n > p$) vs. Compressed sensing ($n < p$).

Randomized measurements (RIP, RE, NSP, RSC, etc.)

Recovery: convex programs (e.g., LASSO), greedy pursuits (e.g., OMP), etc.

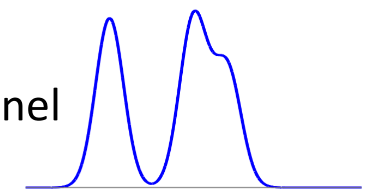


Guarantees: If $n = \mathcal{O}(s \log p)$,
then $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\| \leq \epsilon$ (stability).

Introduction: Signal Deconvolution

➢ Unknown discrete events convolved with unknown kernel

kernel



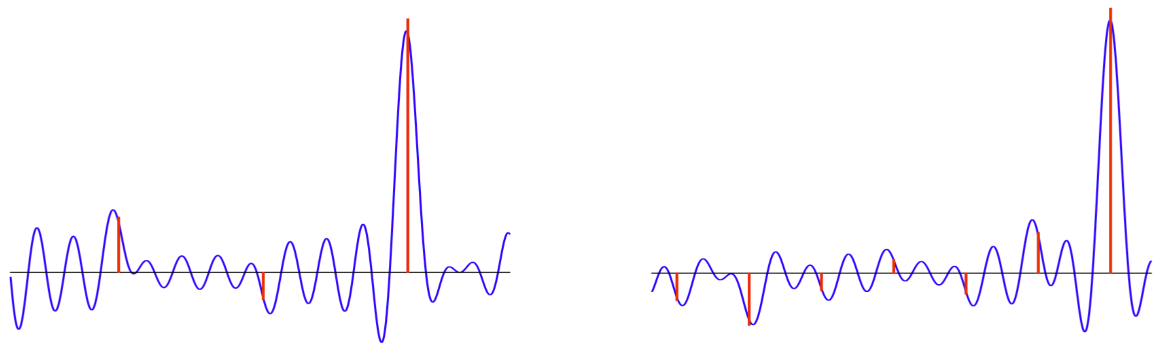


Figure from Candes and Fernandez-Granda (2012)

➢ Deconvolution Problem:

- Estimate signal and the kernel (blind)

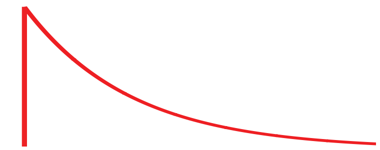
➢ Goals:

- Use partial knowledge about kernel (modeling)
- Fast and scalable solutions
- Performance guarantees using sparsity
- Event rate estimation
- Confidence in detected events
- Compressive measurements for higher rates?

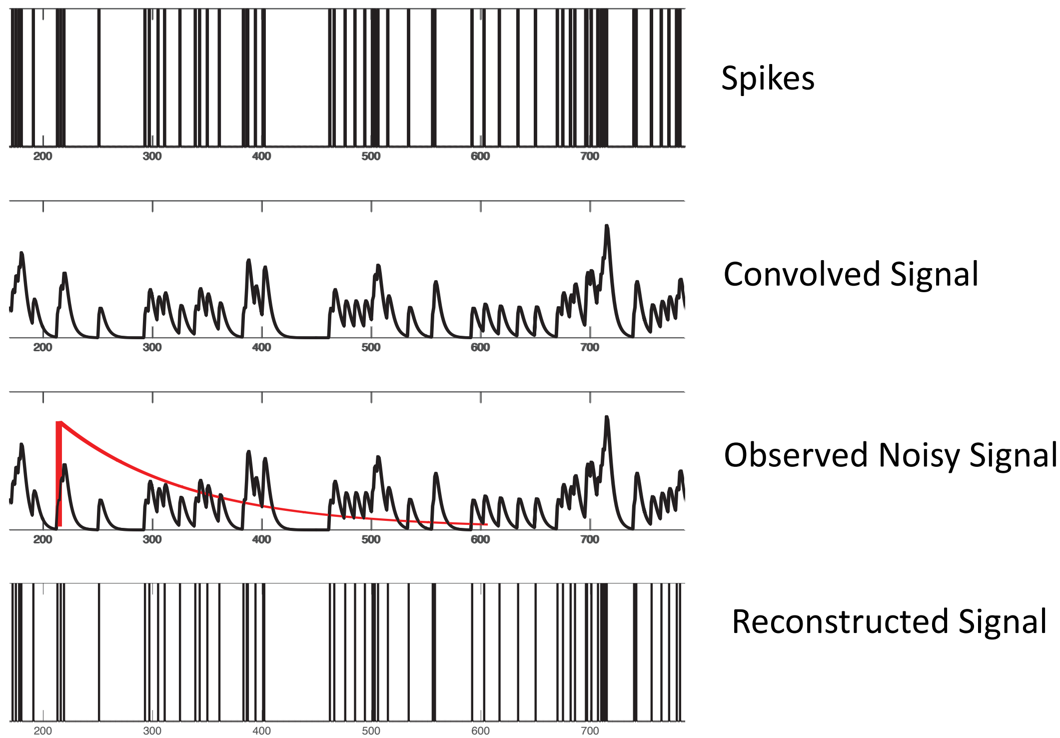
Image from Google

Example: Calcium Deconvolution

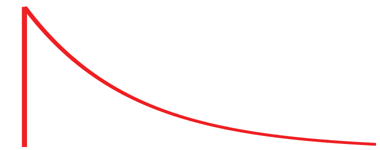
- Fast rise and slow decay in calcium due to spikes
- Autoregressive models: AR(1) or AR(2): $\mathbf{x}_t = \mathbf{\Theta} \mathbf{x}_{t-1} + \mathbf{w}_t$



- Naïve strategy 1:
  Template matching

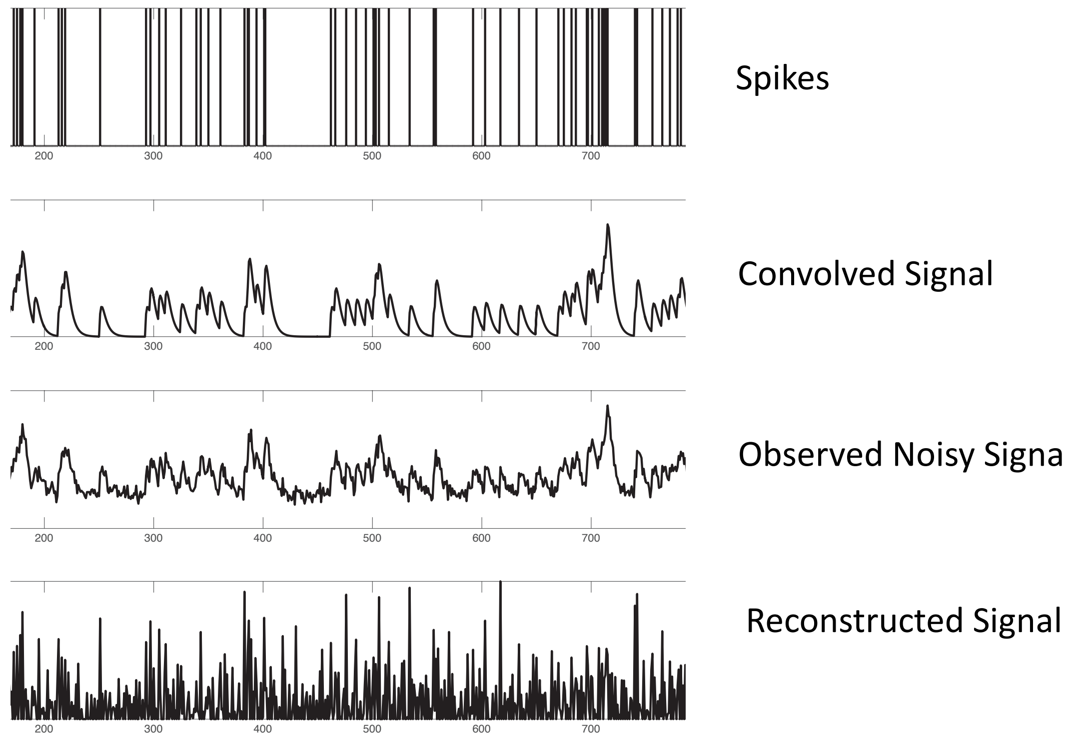


Example: Calcium Deconvolution



- Fast rise and slow decay in calcium due to spikes
- Autoregressive models: AR(1) or AR(2): $\mathbf{x}_t = \mathbf{\Theta}\mathbf{x}_{t-1} + \mathbf{w}_t$

- Naïve strategy 1: Template matching
- Naïve strategy 2: Inverse filtering

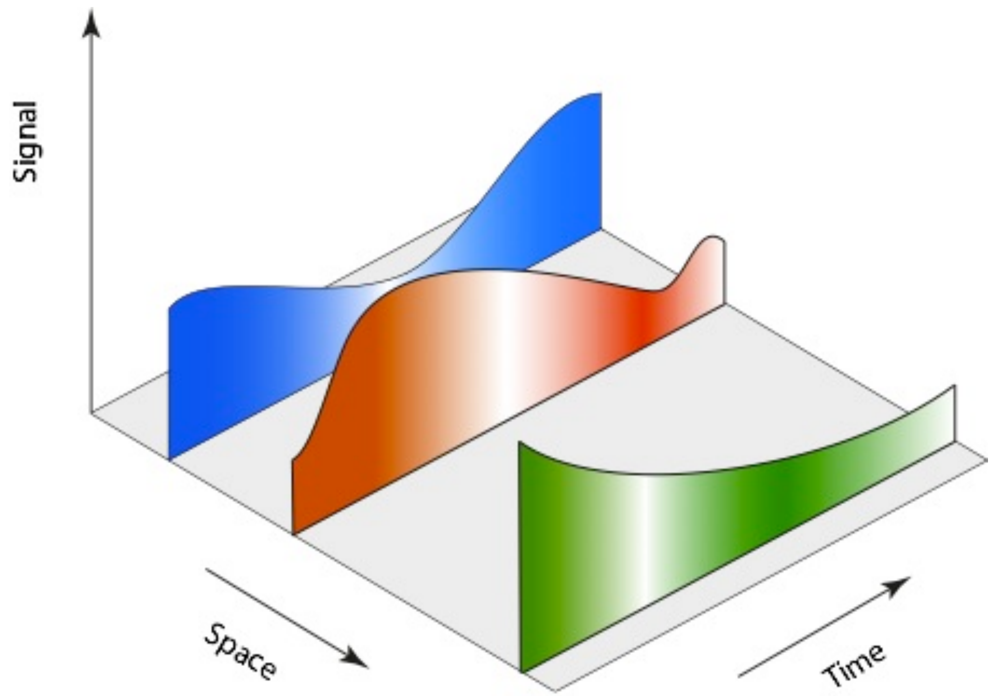


- Other methods:

- Greedy methods (Vogelstein et al., 2010), Supervised learning (Theis et al., 2015)
- Particle filtering (Vogelstein et al., 2009) , MCMC methods (Pnevmatikakis et al., 2013)
- **State-Space Models:** Nonnegative deconvolution (Vogelstein et al., 2010), Pnevmatikakis et al., 2016

Spatiotemporal structure not used or slow

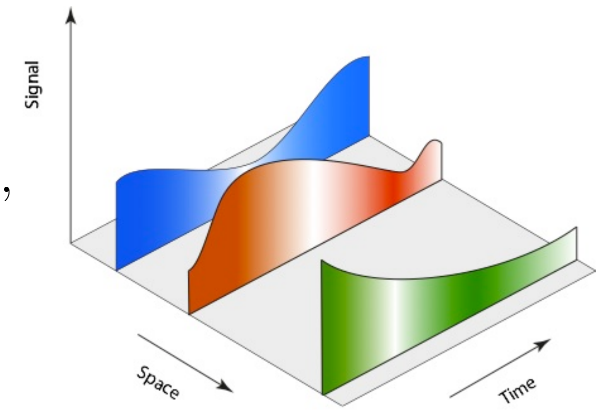
Compressible State-Space Models



Compressible State-Space (CSS) Models

- Compressible State-Space Models

innovations



State dynamics: $\mathbf{x}_t = \mathbf{\Theta}\mathbf{x}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{L}\left(\mathbf{0}, \lambda \mathbf{I}\right)$,

Observation model: $\mathbf{y}_t = \mathbf{A}_t\mathbf{x}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \sigma^2 I)$

- sparse in time

- $\mathbf{x}_t - \mathbf{\Theta}\mathbf{x}_{t-1}$ is s_t -sparse

n_t measurements at time t

- Examples:

Foopsi (Vogelstein et al., 2010)

$$\mathbf{x}_t = \theta \mathbf{x}_{t-1} + \mathbf{n}_t, \quad (\mathbf{n}_t)_i \sim \text{Poi}(\lambda)$$
$$\mathbf{y}_t = \alpha \mathbf{x}_t + \boldsymbol{\beta} + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}) \; .$$

(Pnevmatikakis et al., 2016)

$$\mathbf{x}_t = \theta_1 \mathbf{x}_{t-1} + \theta_2 \mathbf{x}_{t-2} + \mathbf{w}_t$$
$$\mathbf{y}_t = \mathbf{x}_t + \mathbf{v}_t$$

Subject to: $\mathbf{x}_t \geq 0$, $\mathbf{w}_t \geq 0$

- MEG data, fMRI, Heartbeats, Video denoising, Estimation of time-varying networks

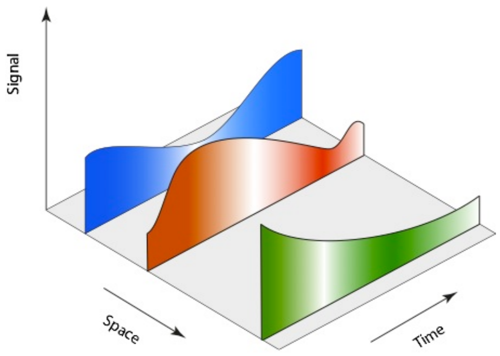
- How many measurements?

- Performance Guarantees of the ℓ_1 -regularized estimator
- Can we have fast solvers?

Compressible State-Space (CSS) Models

- Laplace State-Space Models

Laplace innovations



State dynamics: $\mathbf{x}_t = \mathbf{\Theta}\mathbf{x}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{L}\left(\mathbf{0}, \lambda \mathbf{I}\right)$,

Observation model: $\mathbf{y}_t = \mathbf{A}_t\mathbf{x}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \sigma^2 I)$

- Estimator (inducing sparsity): MAP estimator for Laplace distribution

- CSS:

$$\{(\widehat{\mathbf{x}}_t)_{t=1}^T, \mathbf{\Theta}\} = \underset{\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_T,\mathbf{\Theta}}{\operatorname{argmin}} \quad \lambda \sum_{t=1}^{T} \frac{\|\mathbf{x}_t - \mathbf{\Theta}\mathbf{x}_{t-1}\|_1}{\sqrt{s_t}} + \frac{1}{n_t}\frac{\|\mathbf{y}_t - \mathbf{A}_t\mathbf{x}_t\|_2^2}{2\sigma^2}.$$

- General solutions: batch mode.

- Computationally demanding in modern applications.

Fast Iterative Solution to CSS Model (FCSS)

- Expectation Maximization (EM) algorithm:

- E-step: Iterative Reweighted Least Squares (IRLS): go from ℓ_1 to ℓ_2 .

- CSS:

$$\{(\widehat{\mathbf{x}}_t)_{t=1}^T, \mathbf{\Theta}\} = \underset{\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_T,\mathbf{\Theta}}{\operatorname{argmin}} \quad \lambda \sum_{t=1}^{T} \frac{\|\mathbf{x}_t - \mathbf{\Theta}\mathbf{x}_{t-1}\|_1}{\sqrt{s_t}} + \frac{1}{n_t} \frac{\|\mathbf{y}_t - \mathbf{A}_t\mathbf{x}_t\|_2^2}{2\sigma^2}.$$

Iteratively Re-weighted Least Squares (IRLS)

- Consider the following minimization problem:

$$\min_{x_1, x_2} \quad |x_1| + |x_2| \qquad \text{s.t.} \quad x_2 - \tfrac{1}{2}x_1 = 2$$

- Re-write as follows:

$$\min_{x_1, x_2} \quad \frac{x_1^2}{|x_1|} + \frac{x_2^2}{|x_2|} \qquad \text{s.t.} \quad x_2 - \tfrac{1}{2}x_1 = 2$$

- Substitute the denominator by a guess: $\widehat{x}_1 = -0.8, \quad \widehat{x}_2 = 1.6$

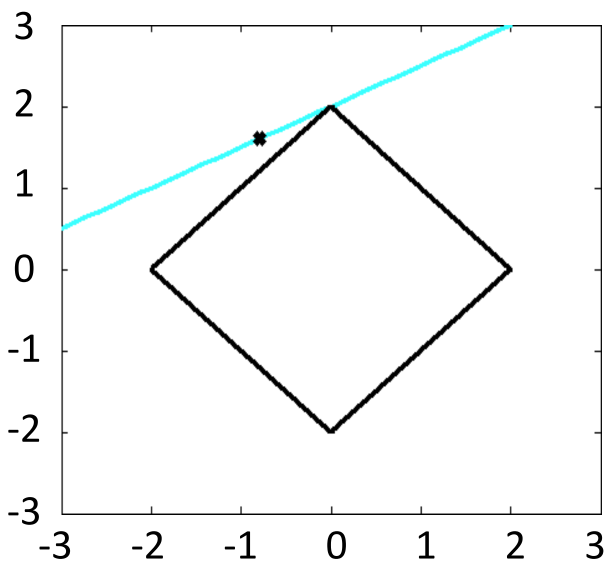
$$\min_{x_1, x_2} \quad \frac{x_1^2}{0.8} + \frac{x_2^2}{1.6} \qquad \text{s.t.} \quad x_2 - \tfrac{1}{2}x_1 = 2$$

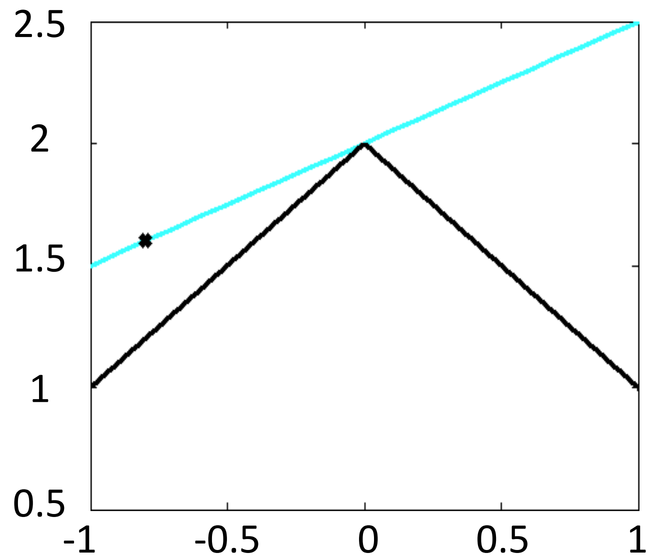
Closed form solution exists!

Ellipse

Line

EM algorithm



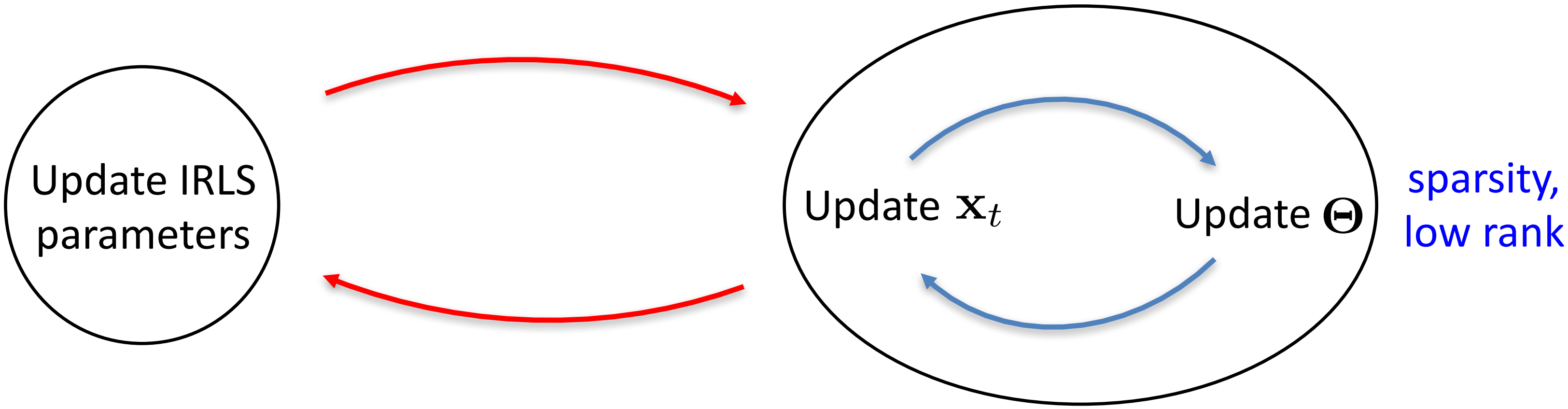


Fast Iterative Solution to CSS Model (FCSS)

- CSS: $$\{(\widehat{\mathbf{x}}_t)_{t=1}^T, \boldsymbol{\Theta}\} = \underset{\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_T,\boldsymbol{\Theta}}{\operatorname{argmin}} \quad \lambda \sum_{t=1}^{T} \frac{\|\mathbf{x}_t - \boldsymbol{\Theta}\mathbf{x}_{t-1}\|_1}{\sqrt{s_t}} + \frac{1}{n_t} \frac{\|\mathbf{y}_t - \mathbf{A}_t\mathbf{x}_t\|_2^2}{2\sigma^2}.$$

➢ Expectation Maximization (EM) algorithm:

- E-step: Iterative Reweighted Least Squares (IRLS): go from ℓ_1 to ℓ_2 .



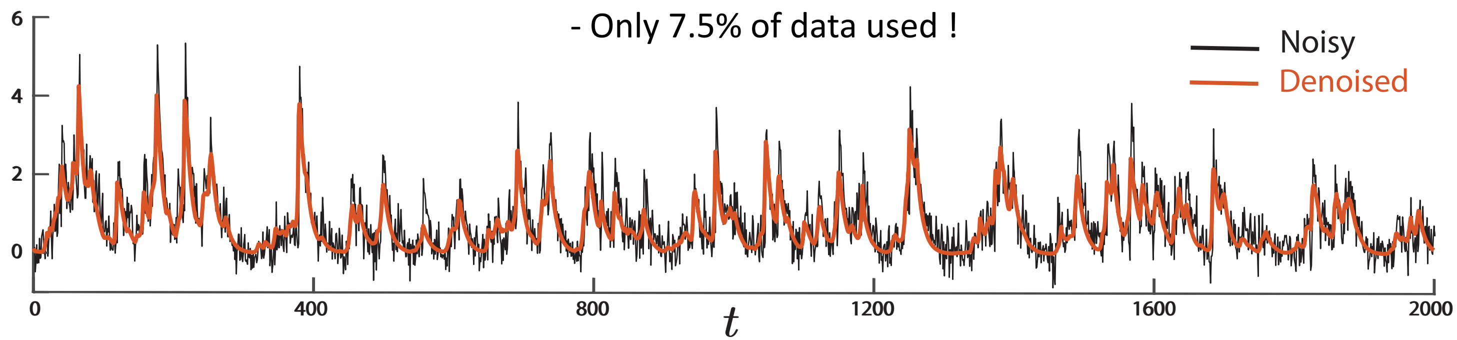
from ℓ_1 to ℓ_2

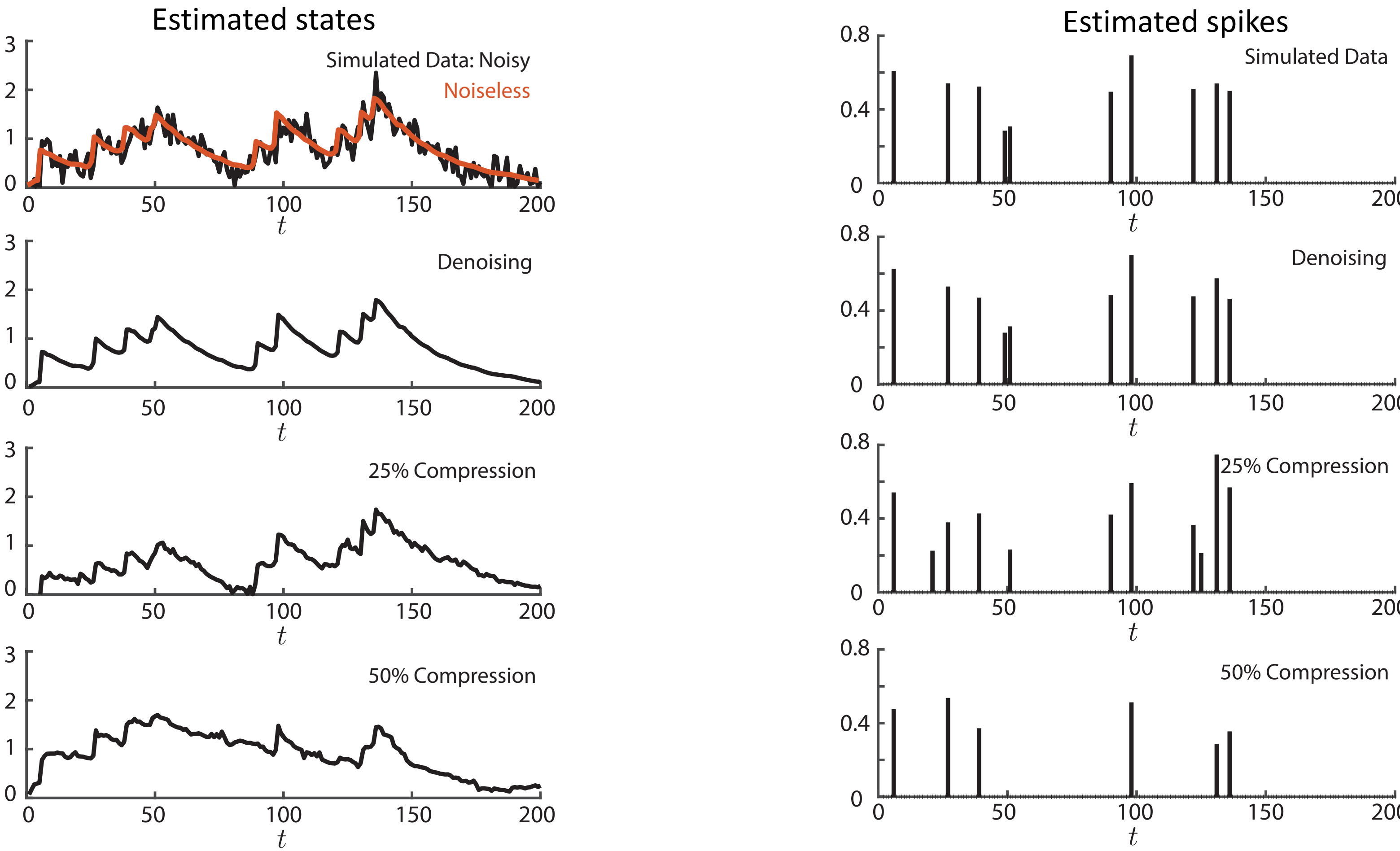
Fast solution: Fixed Interval Smoother

- Can be generalized to other state-space models.

A simulated example

$\mathbf{\Theta} = 0.95\mathbf{I}, p = 200, s_1 = 20, n_1 = 75, s_t = 4, n_t = 15, (t \geq 2)$



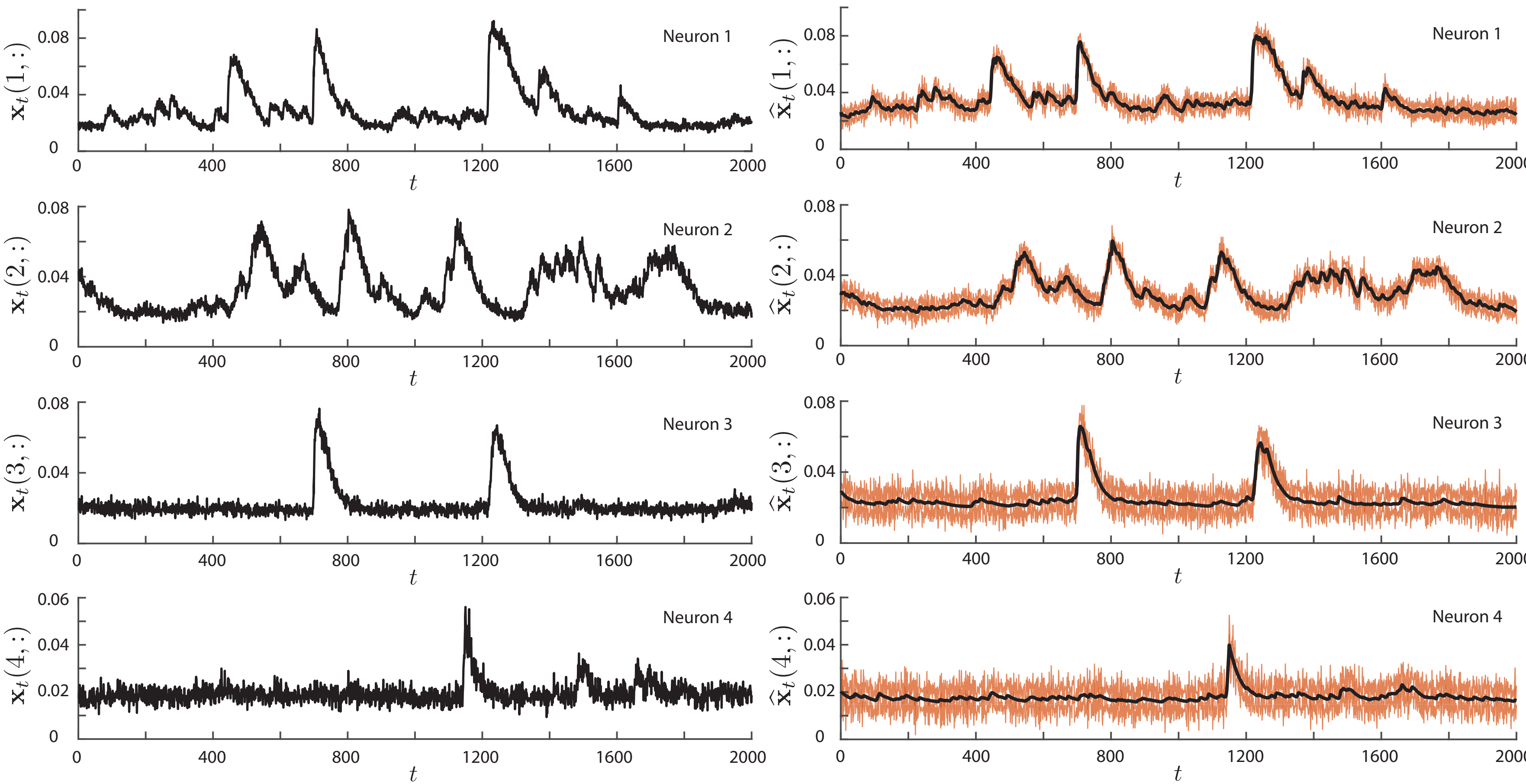


Application to Calcium Deconvolution

- In vivo recordings for 108 ROI's
  (Ji Liu and Patrick Kanold)

$p = 108, n_t = 72$ Random projections

Recorded vs. reconstructed signal



Compression: $n/p = 2/3$

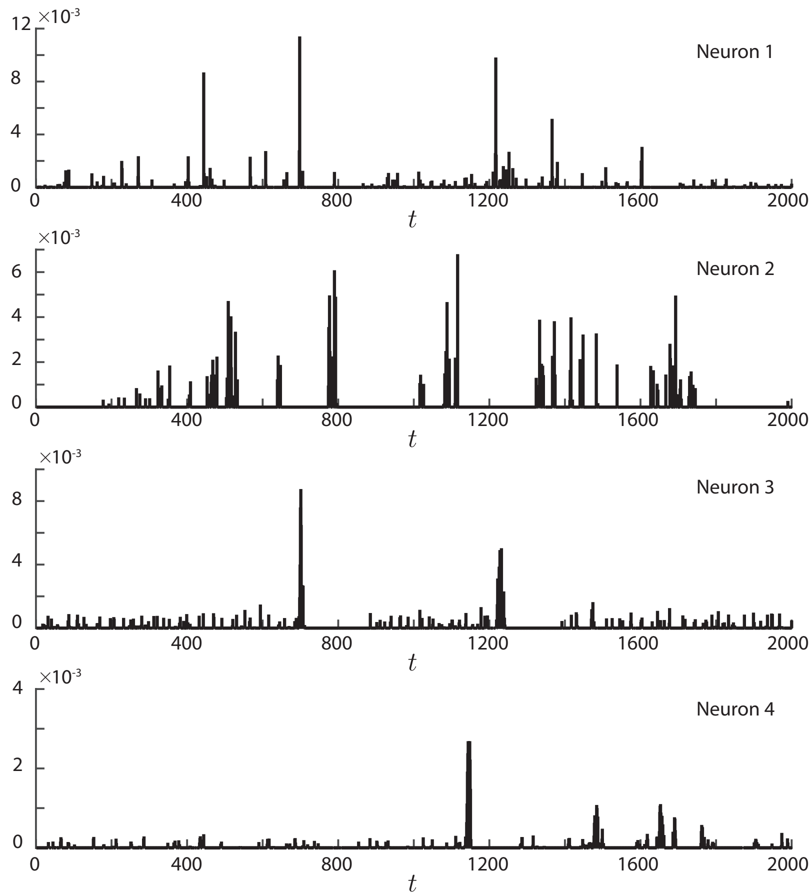
Van de geer et al., (2014)

Application to Calcium Deconvolution

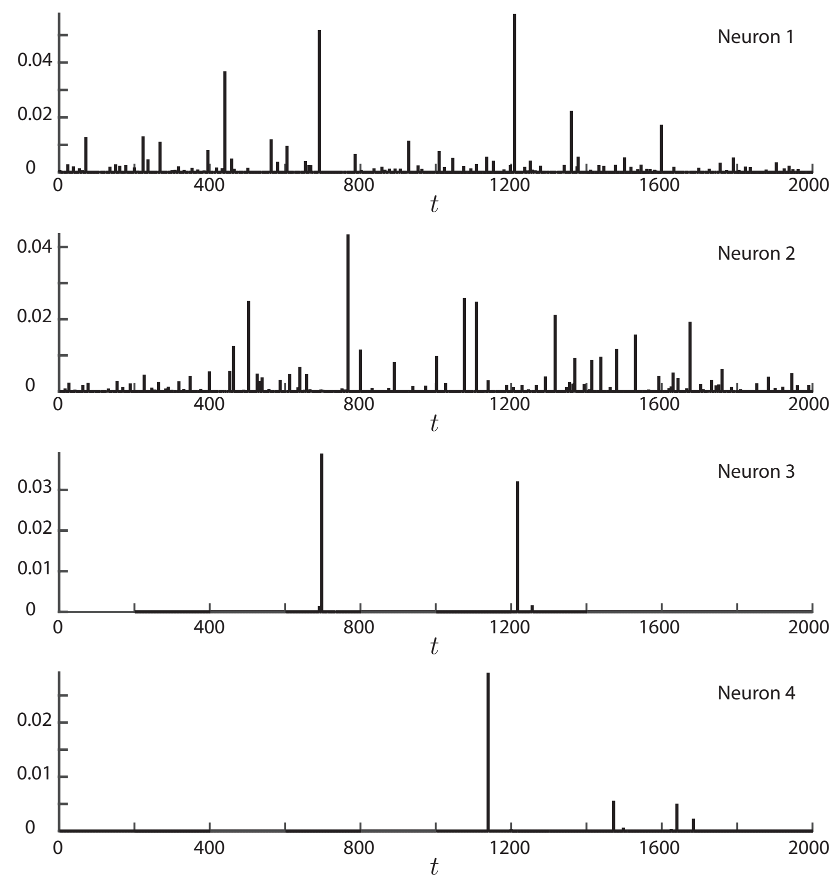
- In vivo recordings for 108 ROI's
  (Ji Liu and Patrick Kanold)

$p = 108, n_t = 72$ Random projections

Reconstructed spikes



foopsi algorithm (Vogelstein et. al. 2010)



FCSS

Raw Data

Theoretical results

Kazemipour et al., 2016

A matrix $\mathbf{A} \in \mathbb{R}^{n \times p}$ satisfies the Restricted Isometry Property (RIP) of order s with constant δ_s if for any s–sparse vector $\mathbf{x} \in \mathbb{R}^p$ we have

$$(1 - \delta_s)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta_s)\|\mathbf{x}\|_2^2$$

- Mild condition: dynamics with convergent transition matrix

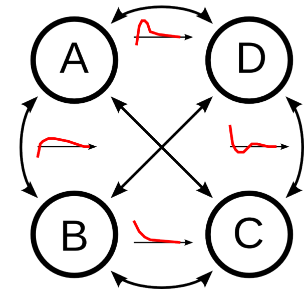
Theorem 1. If $\tilde{\mathbf{A}}_t = \sqrt{\frac{n_t}{n_1}}\mathbf{A}_t$, satisfies RIP of order $4s_t$, with $\delta_{4s_t} < 1/3$, then there exists a constant c_{Θ} such that any solution to the CSS problem satisfies:

$$\sum_{t=1}^{T} \|\mathbf{x}_t - \widehat{\mathbf{x}}_t\|_2 \leq 12.55 c_{\Theta} T \epsilon$$

- Only require $n_t = \mathcal{O}(s_t \log p)$ measurements.

Summary

- Laplace state-space models
- Advantages:
  - Fast solver
  - Spatiotemporal structure: can enforce sparsity, low rank …
  - Theoretical performance guarantees
  - Precise confidence bounds for events
  - Allows compressive measurements! Robust to noise
- Superresolution properties also known: Candes and Fernandez-Granda (2012)
- Can generalize to other stationary processes



FCSS code available on github
kaazemi/FCSS

Fast and Stable Signal Deconvolution via Compressible State-Space Models
BioRxiv/2016/092643

Acknowledgement

Min Wu

Behtash Babadi

Ji Liu

Patrick Kanold