Sparse Penalties in Dynamical System Estimation

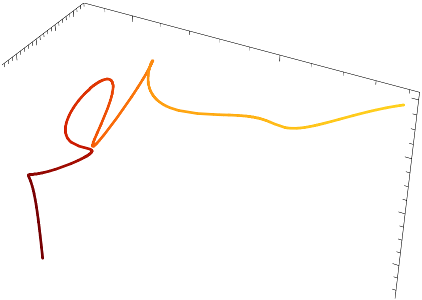
A. Charles, M. S. Asif, J. Romberg, C. Rozell

School of Electrical and Computer Engineering
Georgia Institute of Technology

CISS March 23, 2011

Dynamical Systems

Dynamical Systems



Systems Equations

- Dynamical system:

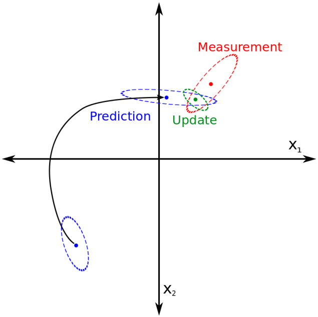
$$\boldsymbol{x}_n = f_n(\boldsymbol{x}_{n-1}) + \boldsymbol{\nu}_n$$

$$\boldsymbol{y}_n = \boldsymbol{G}_n \boldsymbol{x}_n + \boldsymbol{\epsilon}_n$$

- Estimate $\boldsymbol{x}_n$ at each iteration
- Like Kalman Filters, but with sparsity
- State, Innovations, or both can be sparse

Kalman Filter

- Linear and Gaussian assumptions $\rightarrow$ Kalman Filter is optimal
  - Estimates *best* $\boldsymbol{x}_n$ at time n
  - Uses only $\boldsymbol{y}_n$, $\hat{\boldsymbol{x}}_{n-1}$ and a parameter matrix (Cov[$\hat{\boldsymbol{x}}_{n-1}$])
  - Optimal AND Fast



Kalman as a Local Optimization

- Kalman really solves:

$$\{\hat{\boldsymbol{x}}_k\}_{k=0}^{n} = \arg\min_{\{\boldsymbol{x}_k\}} \left[\sum_{k=0}^{n} \|\boldsymbol{y}_k - \boldsymbol{G}_k \boldsymbol{x}_k\|_{\boldsymbol{Q}^{-1},2}^2 + \sum_{k=1}^{n} \|\boldsymbol{x}_k - \boldsymbol{F}_k \boldsymbol{x}_{k-1}\|_{\boldsymbol{R}^{-1},2}^2 \right]$$

- Where $\boldsymbol{Q} = \mathrm{Cov}[\epsilon]$ and $\boldsymbol{R} = \mathrm{Cov}[\boldsymbol{\nu}]$
- Calculates locally via

$$\hat{\boldsymbol{x}}_n = \arg\min_{\boldsymbol{x}} \left[\|\boldsymbol{y}_n - \boldsymbol{G}_n \boldsymbol{x}_n\|_{\boldsymbol{Q}^{-1},2}^2 + \|\boldsymbol{x} - \boldsymbol{F}_n \hat{\boldsymbol{x}}_{n-1}\|_{\boldsymbol{P}_{n|n-1}^{-1},2}^2 \right]$$

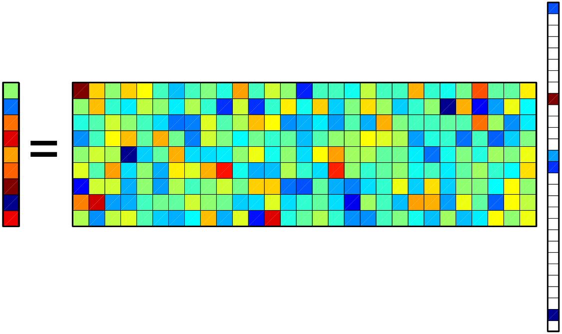
- Where $\boldsymbol{P}_{n|n-1} = \mathrm{Cov}[\boldsymbol{F}_n \hat{\boldsymbol{x}}_{n-1} + \boldsymbol{\nu}_n]$

Kalman Extensions

- No linearity/Gaussianity → Kalman optimality not guaranteed
- Extensions patch the process
  - Unscented/Extended Kalman → nonlinear processes
  - Robust Kalman → different noise statistics
- What about special structures?
- Can compressive sensing help with dynamic signal estimation?

Compressive Sensing and Sparse Coding

Compressive Sensing and Sparse Coding



Formalities

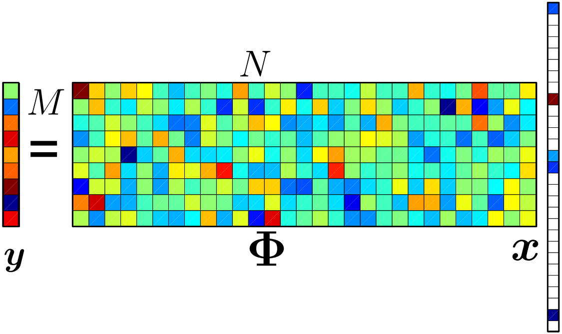
- Measurement: $\boldsymbol{y} \in \mathbb{R}^M$
- can write as

$$\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{\epsilon}$$

- $\boldsymbol{x} \in \mathbb{R}^N$ (typically $M < N$)
- Overcomplete so choice in $\boldsymbol{x}$: Choose sparse $\boldsymbol{x}$

Less Formal

- Can visualize as:



Finding the Coefficients

- Optimize ℓ_0 regularized least squares

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \left[\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_0 \right]$$

- Wonderful result: can optimize ℓ_1 regularized least squares instead

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \left[\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1 \right]$$

- Now it's convex! (fast solvers)

Ties to Dynamic Signal Estimation

- Gives a framework where sparsity knowledge can be leveraged
- Just add ℓ_1
- Can this help sparse dynamics signals?

Previous Work

- Mostly based around separating support from values

  (Vaswani 2008, Ziniel et al. 2010)

  - ℓ_1 should get both simultaneously
- Some attempts to modify Kalman directly

  (Carmi et al. 2010)

  - Not the same problem: solution may look very different
- Smoothing algorithms

  (Angelosante et al. 2010)
- Homotopy updates for fast estimation

  (Asif et al. 2010)

  - Deals with one possibility of sparsity

Problem Statement

- Use compressive sensing ideas to reduce M for dynamical systems
  - *Where* can sparsity be leveraged
  - *How* can we leverage it (State? Innovations?)
- Want: a framework where ℓ_1 norms can represent sparse priors

Where is the Sparsity?

- 3 models:

$$\boldsymbol{x}_n = f_n(\boldsymbol{x}_{n-1}) + \boldsymbol{\nu}_n$$

- Model 1: Sparse states
- Model 2: Sparse innovations
- Model 3: Both sparse states and innovations
- No longer globally optimal

State Sparsity Term

- If the state is sparse: regularize with ℓ_1

$$\hat{\boldsymbol{x}}_n = \arg\min_{\boldsymbol{x}} \left[\|\boldsymbol{y}_n - \boldsymbol{G}_n \boldsymbol{x}_n\|_2^2 + \lambda_1 \|\boldsymbol{x}\|_1 + \lambda_2 \|\boldsymbol{x} - f_n(\hat{\boldsymbol{x}}_{n-1})\|_2^2 \right]$$

- Tracking a sparse number of objects

Innovations Sparsity Term

- If the Innovations is sparse: change second norm to ℓ_1

$$\hat{\boldsymbol{x}}_n = \arg\min_{\boldsymbol{x}} \left[\|\boldsymbol{y}_n - \boldsymbol{G}_n \boldsymbol{x}_n\|_2^2 + \lambda \|\boldsymbol{x} - f_n(\hat{\boldsymbol{x}}_{n-1})\|_1 \right]$$

- Tracking many objects with sparse appearances/disappearances

Sparsity Everywhere

- If both are sparse: regularize with two ℓ_1 norms

$$\hat{\boldsymbol{x}}_n = \arg\min_{\boldsymbol{x}} \left[\|\boldsymbol{y}_n - \boldsymbol{G}_n \boldsymbol{x}_n\|_2^2 + \lambda_1 \|\boldsymbol{x}\|_1 + \lambda_2 \|\boldsymbol{x} - f_n(\hat{\boldsymbol{x}}_{n-1})\|_1 \right]$$

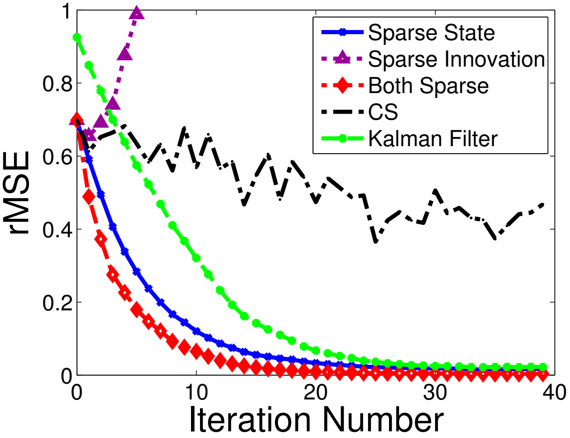
- Video/Dynamic MRI

Simulation Model

- Simulated data: $\boldsymbol{x}_n$ has length 500
- $\boldsymbol{F}_n$ is randomly selected permutation matrix with a random scaling (500x500 matrix)
- $\boldsymbol{G}_n$ is a random Gaussian matrix of size Mx500
- 3 Simulations:
  - Sparse states test: $\boldsymbol{x}_n$ is $K = 20$ sparse, $M = 30$ measurements per iteration
  - Sparse innovations test: $\boldsymbol{x}_n$ is dense, $\boldsymbol{\nu}_n$ is 30 sparse, $M = 200$ measurements per iteration
  - Both sparse states and innovations test: $\boldsymbol{x}_n$ is $K = 20$ sparse, sweep M and number of support changes P

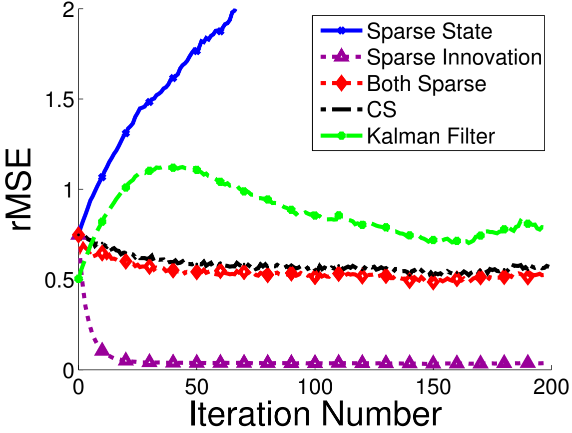
Sparsity in the State

- Sparse state/Gaussian innovations: short term gains



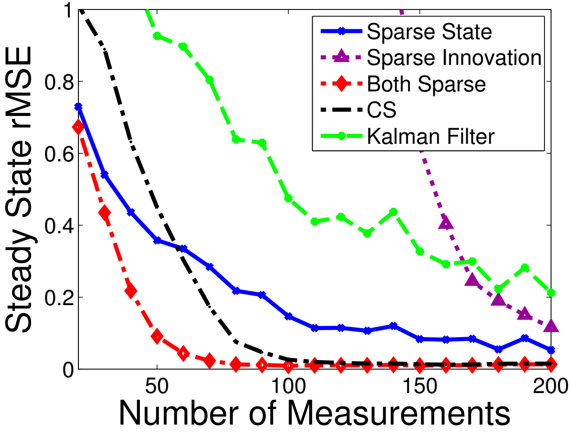
Sparsity in the Innovations

- Sparse Innovations: Tracks dense signals



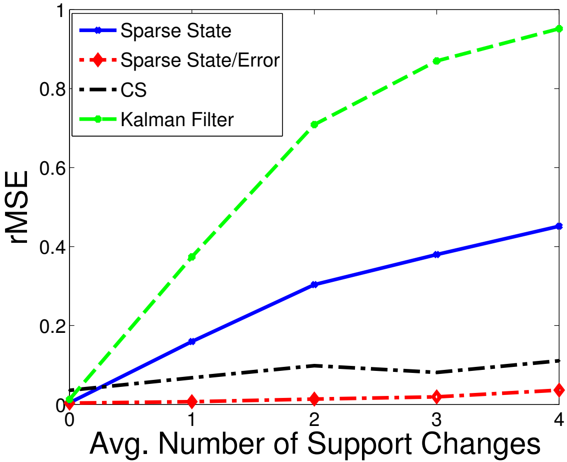
Sparsity in the Both

- Needs many less measurements/iteration



Robustness to Support Changes

- Robust to more changes in support



Summary

- Writing Dynamic state estimation as an optimization problem leads to a principled method of incorporating sparsity information
- Looked at 3 types of sparsity in dynamical system
- Determined a one-step update
- Simulated results show benefits in steady state rMSE, convergence time and rMSE per measurements

Future Work

- Show convergence
- Apply to real datasets (e.g. Dynamic MRI)
- Expand to propagate higher order statistics

Questions

?

acharles6@gatech.edu