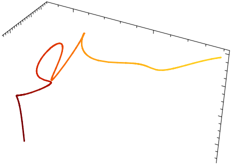
A Hierarchical Re-weighted-ℓ_1 Approach for Dynamic Sparse Signal Estimation

Adam Charles, Christopher Rozell

School of Electrical and Computer Engineering
Georgia Institute of Technology

SPARS11 June 28, 2011

Dynamic Signals



- Temporally evolving signals: Dynamic MRI, Video, Audio, Radar

Systems Equations

- Evolution and measurement equations:

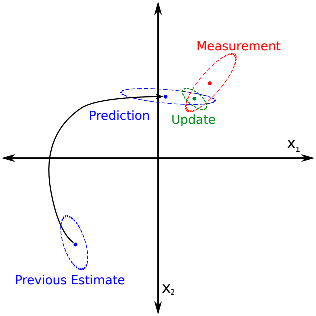
$$\boldsymbol{x}_n = f_n(\boldsymbol{x}_{n-1}) + \boldsymbol{\nu}_n$$

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

- Estimate $\boldsymbol{x}_n$ at each iteration

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 ($\text{Cov}[\hat{\boldsymbol{x}}_{n-1}]$)
  - 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 $\boldsymbol{Q} = \text{Cov}[\boldsymbol{\epsilon}]$ and $\boldsymbol{R} = \text{Cov}[\boldsymbol{\nu}]$
- 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 $\boldsymbol{P}_{n|n-1} = \text{Cov}[\boldsymbol{F}_n \hat{\boldsymbol{x}}_{n-1} + \boldsymbol{\nu}_n]$

Sparsity: The Static Case

- Static model:

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

Sparsity: The Static Case

- Static model:

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

- Regularize with ℓ_1

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

- Clean optimization framework using ℓ_1

Sparsity: The Static Case

- Static model:

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

- Regularize with ℓ_1

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

- Clean optimization framework using ℓ_1
- Leverage this framework for dynamics

(Asif et al. 2010, Charles et. al. 2011)

Where is the Sparsity?

- 3 models:

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

Where is the Sparsity?

- 3 models:

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

- Model 1: Sparse states

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

Where is the Sparsity?

- 3 models:

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

- Model 2: Sparse innovations

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

Where is the Sparsity?

- 3 models:

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

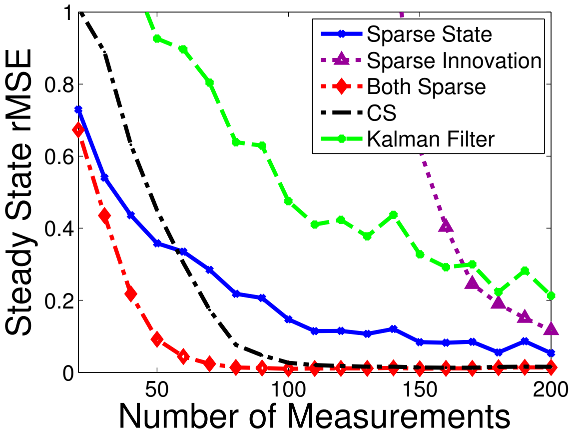
- Model 3: Both sparse states and innovations

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

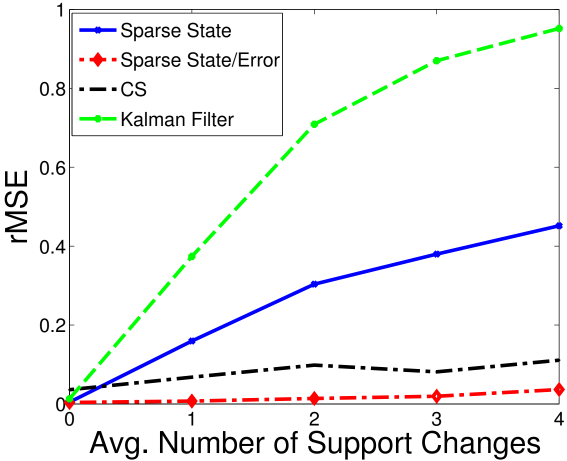
Simulation Model

- 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
- Both sparse states and innovations test: $\boldsymbol{x}_n$ is $K = 20$ sparse, sweep M and number of support changes P

Sparsity in Signal and Innovations



Robustness to Support Changes



First order is Incomplete

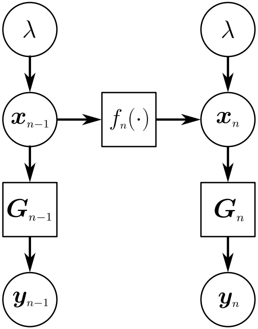
- Use Kalman idea: propagating higher order moments
- Reweighted approaches have already proven useful in static cases

$$\{\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}\} = \arg\min_{\boldsymbol{\lambda}, \boldsymbol{x}} \left[\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + \sum_{k=1}^{N} \boldsymbol{\lambda}(k)|\boldsymbol{x}(k)| \right]$$

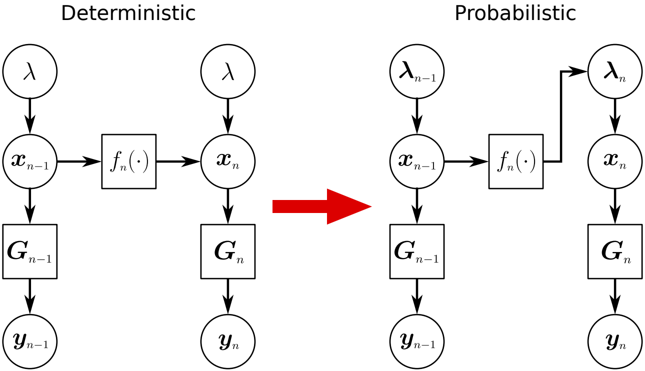
- Pass *prior* information on the tradeoff variables

Graphical Structure

Deterministic



Graphical Structure



The Re-Weighted Algorithm

- At time *n*, for *t* iterating:

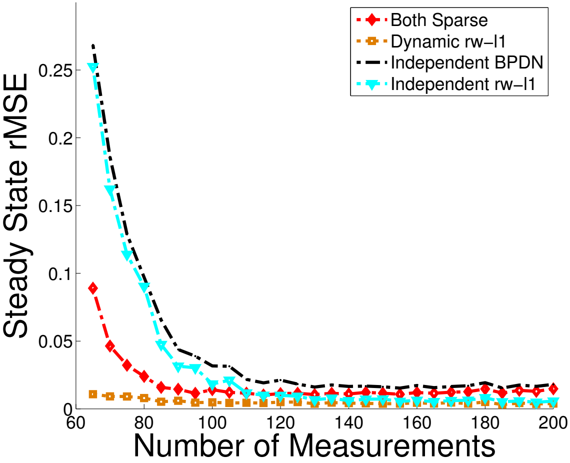
$$\hat{\boldsymbol{\lambda}}_n^t(k) = \frac{2}{|\hat{\boldsymbol{x}}_n^{t-1}(k)| + |f_n(\hat{\boldsymbol{x}}_{n-1})| + \beta}$$

$$\hat{\boldsymbol{x}}_n^t(k) = \arg\min_{\boldsymbol{x}} \left[\|\boldsymbol{y}_n - \boldsymbol{G}_n\boldsymbol{x}\|_2^2 + \sum_k \hat{\boldsymbol{\lambda}}_n^t(k)|\boldsymbol{x}(k)| \right]$$

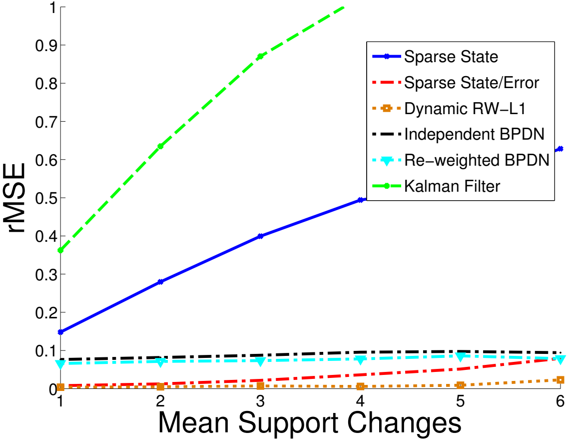
- Modulating *prior* on the tradeoff variables

Simulations Revisited

- Better Error/Iteration with re-weighted schemes



- Re-weighted schemes also more robust to innovations density



Summary and Continuation

- Empirical benefits using higher order moments: error and robustness
- Sub-optimal algorithms outperform optimal algorithms with model mismatch
- Show convergence
- Apply to real datasets (e.g. Dynamic MRI)

Questions

?

acharles6@gatech.edu