Real-time Dynamic MRI using Kalman Filtered Compressed Sensing

Chenlu Qiu, Wei Lu and Namrata Vaswani
Dept. of Electrical & Computer Engineering
Iowa State University
http://www.ece.iastate.edu/~namrata/

MR Imaging

- MRI measures the 2D Fourier transform of the image, which is "incoherent" w.r.t. the wavelet basis

- Medical images are approximately sparse (compressible) in the wavelet transform domain

- MR data acquisition is sequential, the scan time is reduced if fewer measurements are needed for accurate reconstruction.

M. Lustig: Compressed Sensing for Rapid MRI

Real-time dynamic MRI

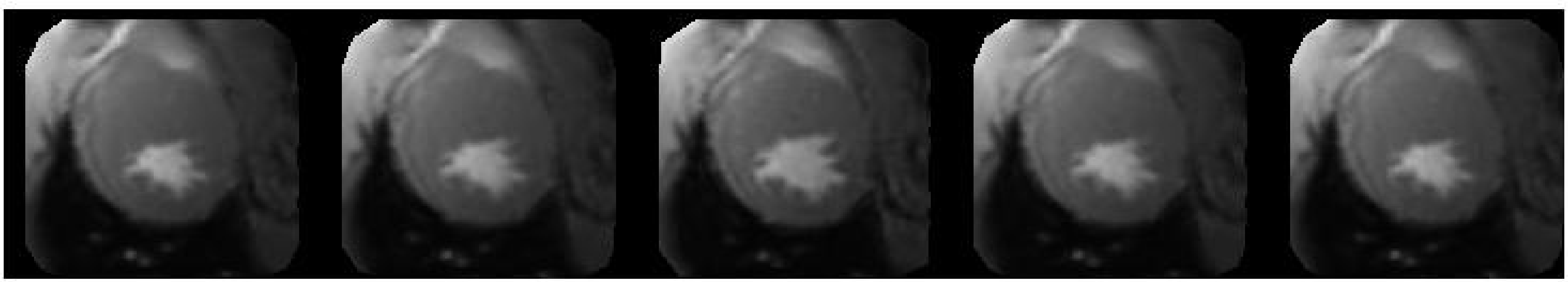
- Reduce scan time: use as few measurements as possible

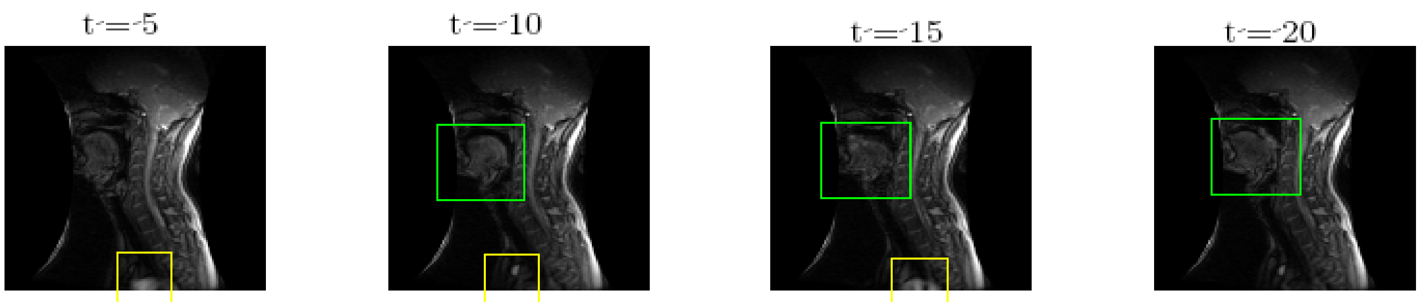
- Reduce reconstruction time: Reconstruct
  - Causally: using current and all past observations

  and
  - Recursively: use previous reconstruction and current observation to obtain current reconstruction)

- Use the fact that
  - sparsity pattern changes slowly over time, and
  - values of the current set of (significantly) nonzero wavelet coefficients also change slowly

Slowly Changing Sparsity





- **Approx. Sparsity.** Size of 99%-energy support set: less than 7% for the larynx sequence and less than 9% for the cardiac sequence.
- **Slow Change in Sparsity Pattern.** Maximum size of change in support: less than 2% of minimum sparsity size in both cases

Problem Definition

Recursively reconstruct a sparse vector, x_t, from the current observation, $y_t := Ax_t + w_t$, & all past observations, $y_{1:t-1}$

- $dim(y_t) = n < dim(x_t) = m$
- x_t is approx. S_t-sparse with approx. support set, N_t
- the support, N_t, changes slowly over time
- A is S_*-"approximately orthonormal" ($\delta_{S_*} < 1/2$) and $S_* > S_t$

- For MRI: $A = HFW'$ with
  - $H_{n \times m}$: random row selection matrix
  - $F_{m \times m}$: DFT matrix, $W_{m \times m}$: DWT matrix
  - "random sample kx-ky plane" or "random sample ky, full sample kx"
  - "unformly random sample" or "variable density undersampling"

RIP and ROP constants [Candes,Tao]

- Restricted Isometry constant, δ_S: smallest real number satisfying

$$(1-\delta_S)||c||_2^2 \leq ||A_T c||_2^2 \leq (1+\delta_S)||c||_2^2$$

  for all subsets T with $|T| \leq S$ and for all c

  - **Easy to see:** $||(A_T{}' A_T)^{-1}||_2 \leq 1/(1-\delta_{|T|})$

- Restricted Orthogonality constant, $\theta_{S,S'}$: smallest real number satisfying

$$|c_1{}' A_{T_1}{}' A_{T_2} c_2| \leq \theta_{S,S'} ||c_1||_2 ||c_2||_2$$

  for all disjoint sets T_1, T_2 with $|T_1| \leq S$, $|T_2| \leq S'$ and for all c_1, c_2

  - **Easy to see:** $||A_{T_1}{}' A_{T_2}||_2 \leq \theta_{|T_1|,|T_2|}$

Compressed Sensing [Candes, Romberg, Tao] [Donoho]

- CS (noiseless) [Candes, Romberg, Tao '05] [Donoho'05]: $\min_\beta ||\beta||_1$ s.t. $y_t = A\beta$
- **CS (noisy - Dantzig Selector)** [Candes, Tao '06]
- CS (noisy - Basis Pursuit Denoising (BPDN)) [Chen,Donoho] [Tropp'06]

$$\min_\beta \gamma||\beta||_1 + (1/2)||y_t - A\beta_t||_2^2$$

We use $\hat{x}_t = \mathbf{CS}(y_t)$ to denote the solution of above

- CS for MR image reconstruction [Lustig, Donoho, Pauly '07]

The Question

- Most existing work: Batch-CS on entire time sequence
  [Gamper et al '08 (dynamic MRI)], [Wakin et al (video)]
  - Offline and very slow, but uses few measurements
- Alternative: CS at each time separately (simple CS)
  - Causal and fast, but needs many more measurements
- The Question: How can we
  - improve simple CS by **using past observations**, and
  - how can we do it **recursively, i.e. by only using the previous signal estimate and the current observation?**

Finding a Recursive Solution

- Given $y_t := Ax_t + w_t$, x_t is sparse with support N_t, N_t changes slowly over time, A satisfies $\delta_{S_t} < 1/2$, $S_t := |N_t|$
- **If N_t known:** easy to compute a restricted-LS estimate

$$\hat{x}_t = \text{restrictedLS}(y_t, N_t) \quad := \quad (\hat{x}_t)_{N_t} = {A_{N_t}}^{\dagger} y_t, \ (\hat{x}_t)_{N_t^c} = 0$$

- **If N_t unknown:** an option is to estimate it by thresholding CS output

$$\hat{N}_t \quad = \quad \text{threshold}(\text{CS}(y_t)) \quad \text{threshold}(x) := \{i : x_i^2 > \alpha\}$$

and then do the same thing

  - **But: not using past observations: large error**

CS-residual idea [Vaswani, ICIP'08]

- Let $T := \hat{N}_{t-1}$ (estimated support at $t-1$) and $\Delta := N_t \setminus \hat{N}_{t-1}$
- Assume that the undetected set, Δ, is small, i.e.
  - the support changes slowly, and
  - the support at $t-1$ is well estimated
- Use $T := \hat{N}_{t-1}$ to compute restricted LS estimate, & observation residual

$$\begin{aligned} (\hat{x}_{t,\text{init}})_T &= \text{restrictedLS}(y_t, T) \\ y_{t,\text{res}} &= y_t - A\hat{x}_{t,\text{init}} \end{aligned}$$

- **CS-residual:** $\hat{x}_t = \hat{x}_{t,\text{init}} + \text{CS}(y_{t,\text{res}})$
  - $y_{t,\mathbf{res}}$ **is a noisy measurement of an approx. $|\Delta|$ sparse vector**

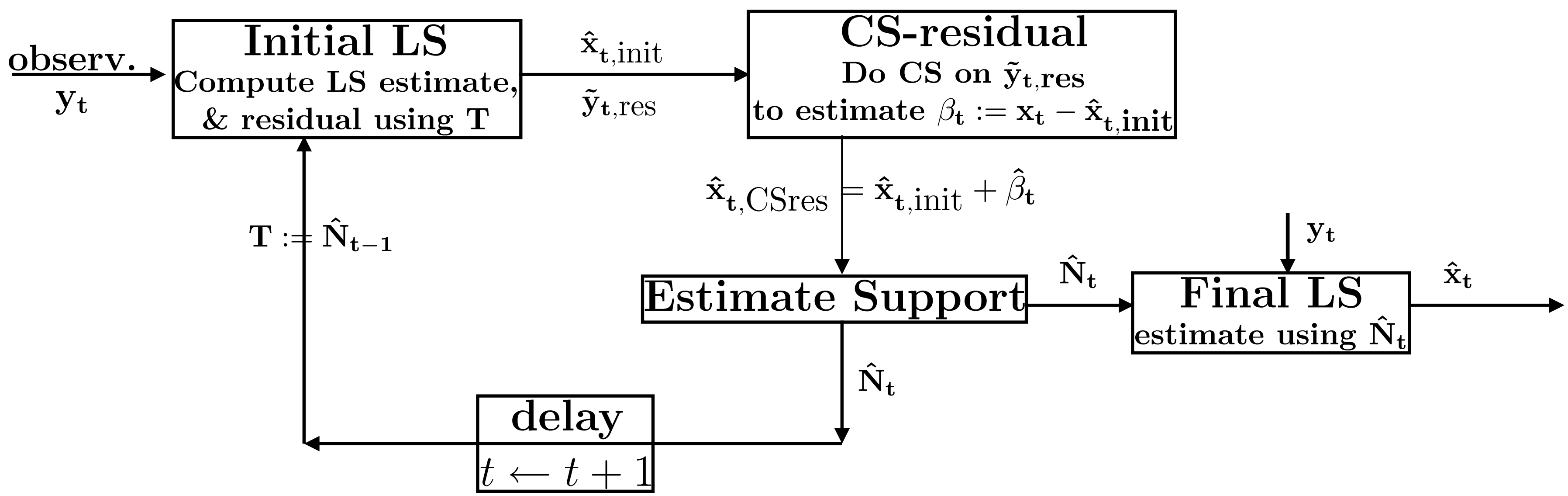
Why CS-residual works?

- Notice that $y_{t,\text{res}} = A\beta_t + w_t$ and $\beta_t := x_t - \hat{x}_{t,\text{init}}$ satisfies

$$\begin{aligned}(\beta_t)_\Delta &= (x_t)_\Delta \\ (\beta_t)_T &= -{A_T}^\dagger(A_\Delta (x_t)_\Delta + w_t) \\ (\beta_t)_{(T\cup\Delta)^c} &= 0\end{aligned}$$

- If $|\Delta|$ small enough s.t. $||{A_T}' A_\Delta||_2 < \theta_{|T|,|\Delta|}$ small:
  - **β_t small along T, i.e. it is only $|\Delta|$-approx-sparse**
- CS error strongly depends on approx. sparsity size
  - **CS-residual: much smaller error** than CS on y_t (simple CS)

LS CS-residual (LS-CS)



- **Initial LS.** Compute $\hat{x}_{t,\mathrm{init}}$ & observation residual $y_{t,\mathrm{res}}$ using $T := \hat{N}_{t-1}$
- **CS-residual.** Compute $\hat{x}_{t,CSres} = \hat{x}_{t,\mathrm{init}} + CS(y_{t,\mathrm{res}})$
- **Estimate Support.** Compute $\hat{N}_t = \mathrm{threshold}(\hat{x}_{t,CSres})$
- **Final LS.** $\hat{x}_t = \mathrm{restrictedLS}(y_t, \hat{N}_t)$ often improves estimate [Candes, Tao '06]

Kalman filtered CS-residual (KF-CS)

[Vaswani, ICIP'08]

- So far only used $\hat{N}_{t-1}$ to improve accuracy of CS at t: did not use $\hat{x}_{t-1}$
- If a prior dynamic model for nonzero coefficients of x_t is available: do this by replacing initial LS by a KF for $(x_t)_T$
- A possible prior model: random-walk on $(x_t)_{N_t}$ starting with $x_0 = 0$

$$\begin{aligned} (x_t)_{N_{t-1}} &= (x_{t-1})_{N_{t-1}} + \mathcal{N}(0, \sigma_s^2 I) \\ (x_t)_{N_t \setminus N_{t-1}} &= \mathcal{N}(0, \sigma_s^2 I) \\ (x_t)_{N_t^c} &= 0 \end{aligned}$$

- **KF CS-residual:**
  - dimension-varying KF with current states' set being $T := \hat{N}_{t-1}$
  - compute $\hat{N}_t$ by thresholding output of CS on KF residual

KF-CS algorithm

Initialize: $\hat{N}_0 = \phi$, $\hat{x}_0 = 0$, $P_0 = 0$. For $t > 0$, do

- **Initial KF. KF on $(x_t)_{\hat{N}_{t-1}}$ and compute KF residual, $y_{t,\mathbf{res}}$**

$$\hat{x}_{t,\text{init}} = \text{KF}(I, \sigma_{sys}^2 I_{\hat{N}_{t-1}}, A_{\hat{N}_{t-1}}, \sigma^2 I)(y_t, \hat{x}_{t-1}, P_{t-1})$$

$$y_{t,\text{res}} = y_t - A\hat{x}_{t,\text{init}}$$

- **CS-residual.** Compute $\hat{x}_{t,CSres} = \text{CS}(\tilde{y}_t) + \hat{x}_{t,\text{init}}$
- **Estimate Support.** Compute $\hat{N}_t = \text{threshold}(\hat{x}_{t,CSres})$
  - zero out elements of deleted set, $\hat{N}_{t-1} \setminus \hat{N}_t$, from $\hat{x}_{t-1}$, P_{t-1}
- **Final KF. KF on $(x_t)_{\hat{N}_t}$**

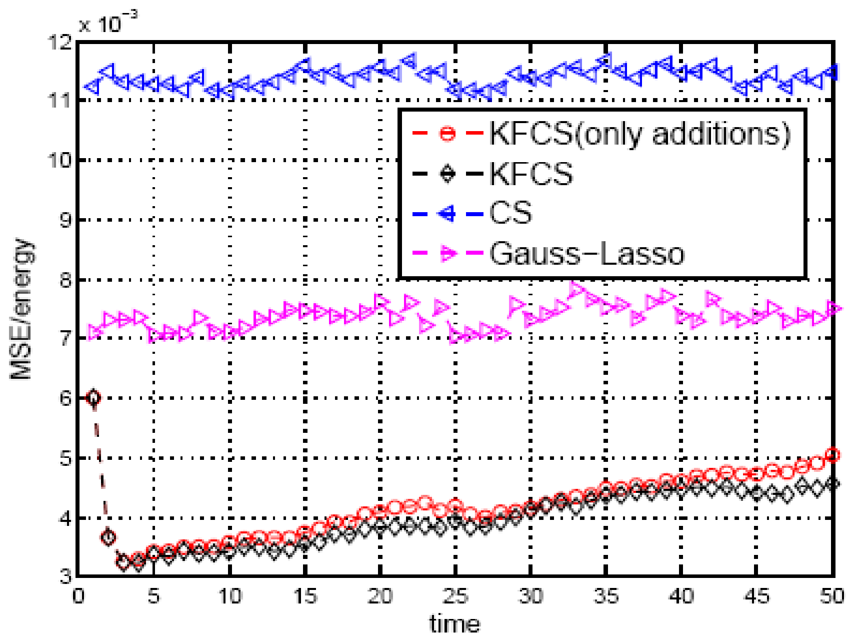
$$[\hat{x}_t, P_t] = \text{KF}(I, \sigma_{sys}^2 I_{\hat{N}_t}, A_{\hat{N}_t}, \sigma^2 I)(y_t, \hat{x}_{t-1}, P_{t-1})$$

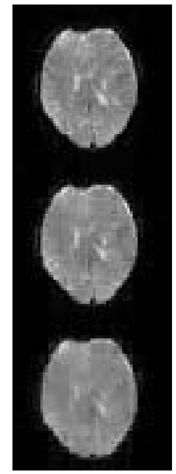
Initial KF. Let $T = \hat{N}_{t-1}$

$$\begin{aligned} P_{t|t-1} &= P_{t-1} + \hat{Q}_t, \quad \text{where} \;\; \hat{Q}_t := \sigma_s^2 I_T \\ K_t &= P_{t|t-1} A' (A P_{t|t-1} A' + \sigma^2 I)^{-1}, \quad P_t = (I - K_t A) P_{t|t-1} \\ \hat{x}_{t,\text{init}} &= (I - K_t A)\hat{x}_{t-1} + K_t y_t \\ y_{t,\text{res}} &= y_t - A\hat{x}_{t,\text{init}} \end{aligned}$$

Final KF. Do the above but with $T = \hat{N}_t$

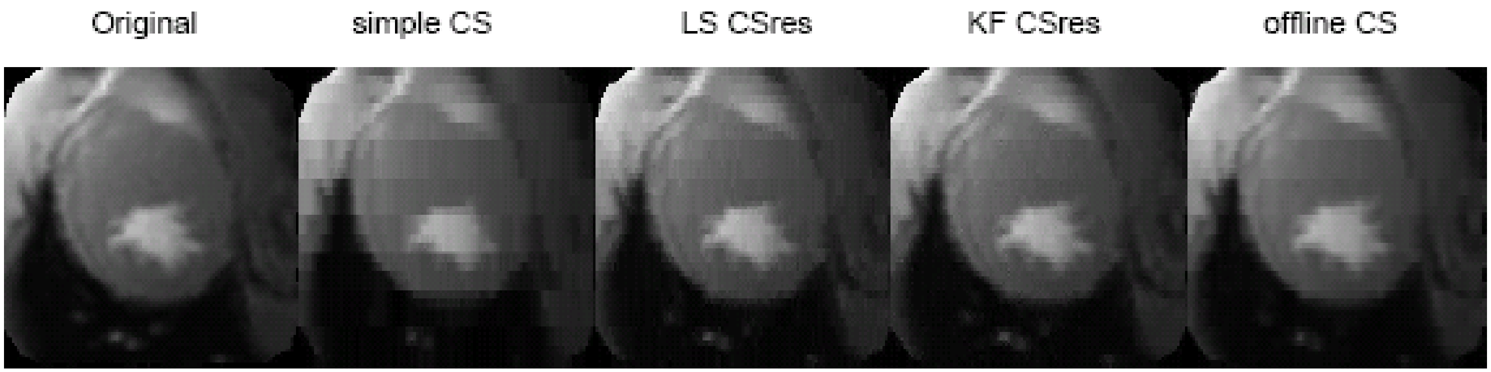
Brain fMRI sequence





- Variable density undersampling in kx-ky
- Use $\gamma = 2\,(2 \log m)^{1/2}\sigma$ as suggested in [Candes-?]
- $m = 4096$, $n = m/2$, $\sigma^2 = 25$

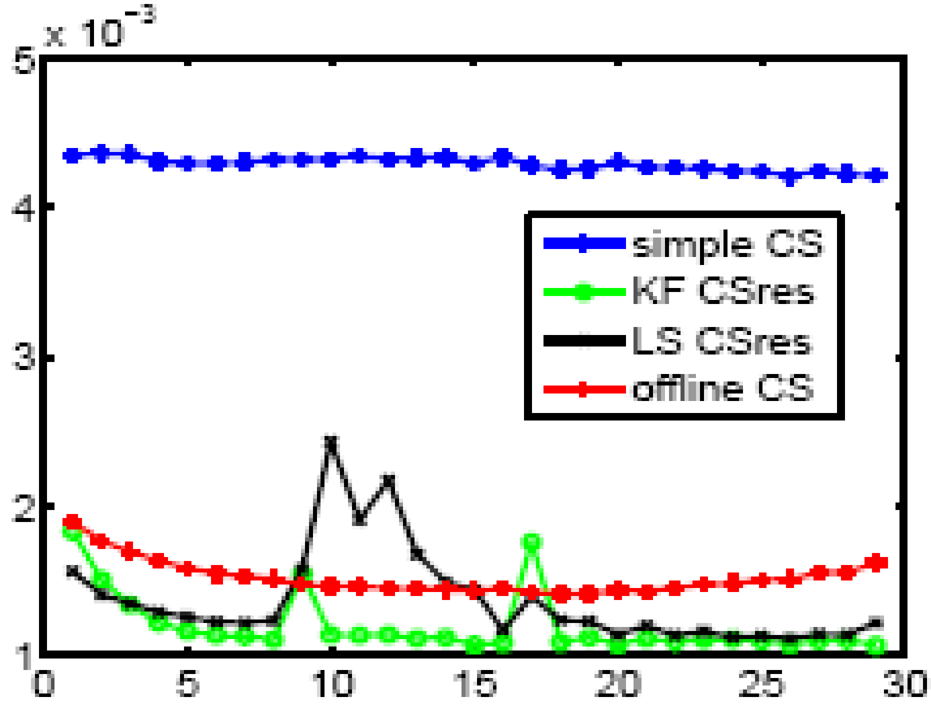
Cardiac sequence



Cardiac sequence: reconstructed last frame

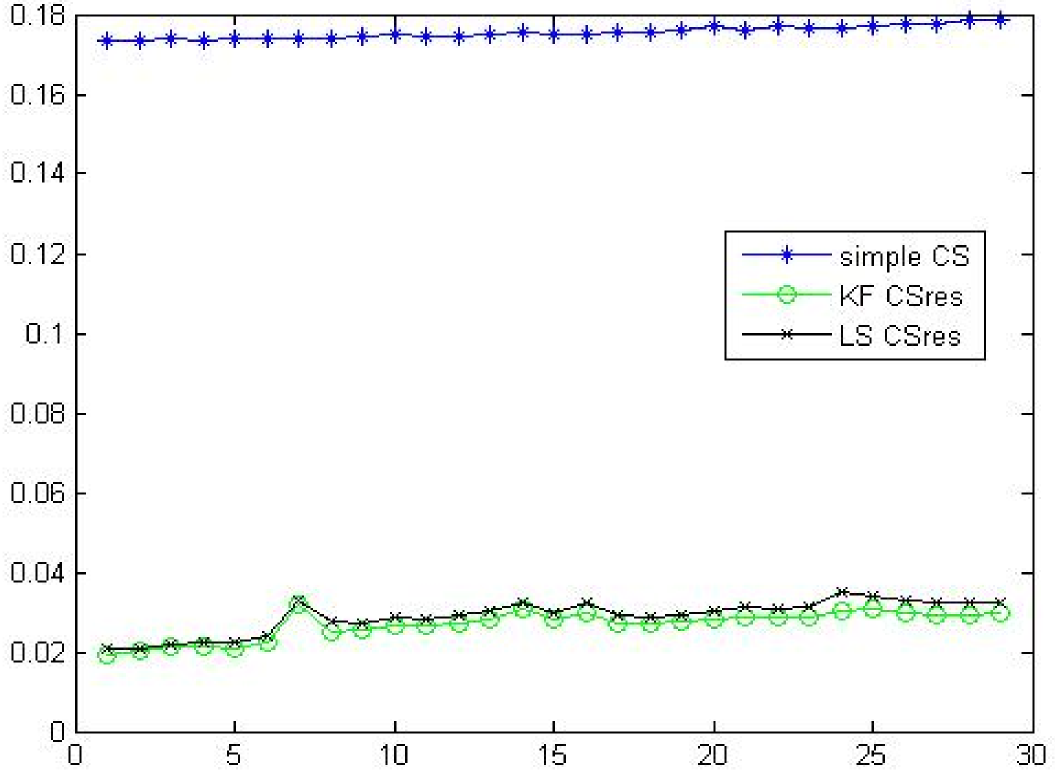
- variable density undersampling in ky, full resolution in kx
- select γ using a heuristic motivated by the error bound of [Tropp'06]
- m = 128 (one column at a time), n = m/2, $\sigma^2 = 25$

Cardiac sequence



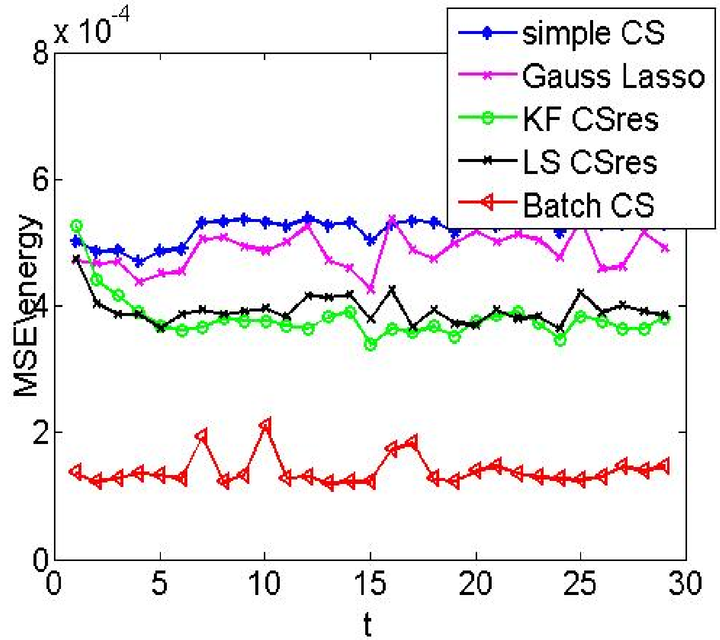
- variable density undersampling in ky, full resolution in kx
- select γ using a heuristic motivated by the error bound of [Tropp'06]
- m = 128 (one column at a time), n = m/4, σ^2 = 25

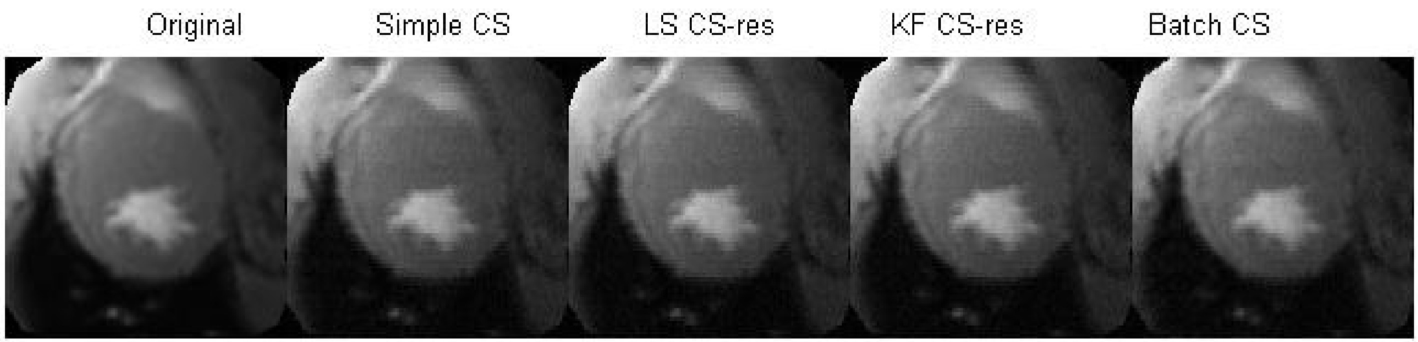
Cardiac sequence



- Uniformly random sample in ky, full resolution in kx
- Use best possible γ for each method
  - (γ = 0.05 for CS, γ = 20 for KF-CS, LS-CS)
- m = 128 (one column at a time), n = m/2, σ^2 = 25

Cardiac





- Variable density undersampling in ky, full resolution in kx
- Use best possible γ for each method
  - ($\gamma = 0.05$ for all)
- $m = 128$ (one column at a time), $n = m/2$, $\sigma^2 = 25$

Related Work

- Our Kalman filtered CS work first appeared in ICIP'08
- **Works not using current observation to compute residual**
  - k-t FOCUSS [Jung,Ye, ISBI'08]
  - Locally Competitive Algorithms for sparse coding [Rozell et al, ICIP'07]
- Very recent work
  - Recursive Lasso [Angelosante, Giannakis, ICASSP'09]
  - Dynamic l1 minimization [Asif, Romberg, CISS'09]
  - Analyzing LS and KF CS [Vaswani, ICASSP'09]
  - Modified-CS [Vaswani, Lu, ISIT'09]

Ongoing/Future Work

- Comparisons, use real MR scanner data, volume seqeunce reconstruction
- Bounding reconstruction error, Studying stability of LS and KF CS-residual
- **Modified-CS** [Vaswani, Lu, ISIT'09]. $\hat{x}_t$ is the solution of

$$\min_\beta ||\beta_{T^c}||_1 \text{ s.t. } y_t = A\beta$$

  - an approach for provably exact reconstruction from noiseless measurements using partly known support, $T := \hat{N}_{t-1}$
  - **exact reconstruction if** $\delta_{|T|+2|\Delta|} < 1/5$ **(much weaker than CS)**

- Combine Modified-CS with CS-residual for noisy/compressible cases

$$\min_\beta \gamma ||\beta_{T^c}||_1 + (1/2)||y_t - A\beta||_2^2$$

$$\min_\beta \gamma ||\beta_{T^c}||_1 + ||\beta_T - (\hat{x}_{t-1})_T||^2_{P_{t|t-1}} + (1/2\sigma^2)||y_t - A\beta||_2^2$$