Recursive Causal Reconstruction of Sparse Signal Sequences

Namrata Vaswani

Department of Electrical and Computer Engineering
Iowa State University
http://www.ece.iastate.edu/~namrata

Our Goal

- *Causally & recursively* reconstruct a time seq. of sparse signals
- with slowly changing sparsity patterns
- from a *small number* of linear projections at each time

- *"recursive": use only current measurements vector and previous reconstruction to get current reconstruction*

Our Goal

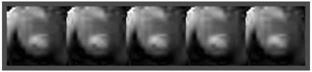
- *Causally & recursively* reconstruct a time seq. of sparse signals
- with slowly changing sparsity patterns
- from a *small number* of linear projections at each time
- *"recursive": use only current measurements vector and previous reconstruction to get current reconstruction*
- Applications
  - real-time dynamic MRI reconstruction
    - interventional radiology apps, e.g. MRI-guided surgery
    - fMRI-based study of neural activation patterns
  - single-pixel video imaging with a real-time video display, . . .

- Why causal?
  - needed for real-time applications
- Why causal & recursive?
  - much faster than causal & batch
    - $O(m^3)$ v/s $O(t^3\ m^3)$ at time t (m: signal length)
  - also much faster than offline & batch
- Why reduce the number of measurements needed?
  - data acquisition in MRI or single-pixel camera is sequential: fewer meas $\Rightarrow$ faster acquisition (needed for real-time)

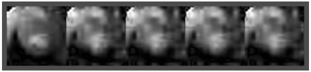
- Most existing work is either
  - for static sparse reconstruction or
  - or is offline & batch [Wakin et al'06(video)],[Gamper et al'08, Jung et al'09 (MRI)]
- Fails if applied to online problem with few measurements

Example: dynamic MRI recon. of a cardiac sequence

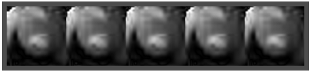
Original sequence



CS−reconstructed sequence



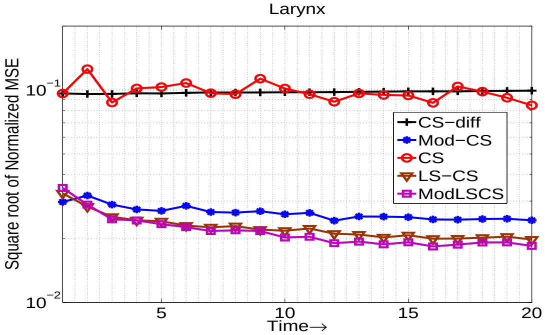
Modified−CS reconstructed sequence



using only 16% Fourier measurements at $t > 0$ (50% at $t = 0$), existing work (CS) gives large reconstruction error (10-12%), proposed approach (modified-CS) is very accurate

Example: dynamic MRI recon of a vocal tract sequence

videos: `http://www.ece.iastate.edu/~luwei/modcs/`



using only 19% Fourier measurements at all times,
existing work (CS, CS-diff) has large error

What is sparse reconstruction?

- Reconstruct a sparse signal x from $y := Ax$ (noiseless) or $y := Ax + w$ (noisy),
  - when A is a fat matrix
- Solved if one can find the sparsest vector satisfying $y = Ax$
  - and $spark(A) > 2|support(x)|$
- But, this has exponential complexity

What is sparse reconstruction?

- Reconstruct a sparse signal x from $y := Ax$ (noiseless) or $y := Ax + w$ (noisy),
  - when A is a fat matrix
- Solved if one can find the sparsest vector satisfying $y = Ax$
  - and $spark(A) > 2|support(x)|$
- But, this has exponential complexity
- Practical approaches (have polynomial complexity in m):
  - convex relaxation approaches, e.g. BP, BPDN, DS, ...
  - greedy methods, e.g. MP, OMP, CoSaMP, ...

What is sparse reconstruction?

- Reconstruct a sparse signal x from $y := Ax$ (noiseless) or $y := Ax + w$ (noisy),
  - when A is a fat matrix
- Solved if one can find the sparsest vector satisfying $y = Ax$
  - and $spark(A) > 2|support(x)|$
- But, this has exponential complexity
- Practical approaches (have polynomial complexity in m):
  - convex relaxation approaches, e.g. BP, BPDN, DS, ...
  - greedy methods, e.g. MP, OMP, CoSaMP, ...
- Compressed Sensing (CS) literature provides the missing theoretical guarantees for the practical approaches

Notation [Candes,Romberg,Tao'05]

- Notation:
  - $|T|$: cardinality of set T
  - $T^c = [1, 2, \ldots m] \setminus T$: complement of set T
  - $\|\beta\|_k$: ℓ_k norm of vector β, $\|\beta\|$: ℓ_2 norm
  - $\|A\|$: spectral matrix norm (induced 2-norm) of matrix A
  - β_T: sub-vector containing elements of β with indices in set T
  - A_T: sub-matrix containing columns of A with indices in set T
  - A': denotes the transpose of matrix A

Notation [Candes,Romberg,Tao'05]

- Notation:
  - $|T|$: cardinality of set T
  - $T^c = [1, 2, \dots m] \setminus T$: complement of set T
  - $\|\beta\|_k$: ℓ_k norm of vector β, $\|\beta\|$: ℓ_2 norm
  - $\|A\|$: spectral matrix norm (induced 2-norm) of matrix A
  - β_T: sub-vector containing elements of β with indices in set T
  - A_T: sub-matrix containing columns of A with indices in set T
  - A': denotes the transpose of matrix A
- *RIP constant, δ_S: smallest real number s.t. all eigenvalues of $A_T{}'A_T$ lie b/w $1 \pm \delta_S$ whenever $|T| \leq S$* [Candes,Romberg,Tao'05]

Notation [Candes,Romberg,Tao'05]

- Notation:
  - $|T|$: cardinality of set T
  - $T^c = [1, 2, \ldots m] \setminus T$: complement of set T
  - $\|\beta\|_k$: ℓ_k norm of vector β, $\|\beta\|$: ℓ_2 norm
  - $\|A\|$: spectral matrix norm (induced 2-norm) of matrix A
  - β_T: sub-vector containing elements of β with indices in set T
  - A_T: sub-matrix containing columns of A with indices in set T
  - A': denotes the transpose of matrix A
- *RIP constant, δ_S: smallest real number s.t. all eigenvalues of $A_T{}' A_T$ lie b/w $1 \pm \delta_S$ whenever $|T| \leq S$* [Candes,Romberg,Tao'05]
  - *$\delta_S < 1 \Leftrightarrow A$ satisfies the S-RIP*

Notation [Candes,Romberg,Tao'05]

- Notation:
  - $|T|$: cardinality of set T
  - $T^c = [1, 2, \dots m] \setminus T$: complement of set T
  - $\|\beta\|_k$: ℓ_k norm of vector β, $\|\beta\|$: ℓ_2 norm
  - $\|A\|$: spectral matrix norm (induced 2-norm) of matrix A
  - β_T: sub-vector containing elements of β with indices in set T
  - A_T: sub-matrix containing columns of A with indices in set T
  - A': denotes the transpose of matrix A
- *RIP constant, δ_S: smallest real number s.t. all eigenvalues of ${A_T}'A_T$ lie b/w $1 \pm \delta_S$ whenever $|T| \le S$* [Candes,Romberg,Tao'05]
  - *$\delta_S < 1 \Leftrightarrow A$ satisfies the S-RIP*
- *ROP constant, θ_{S_1,S_2}: smallest real number s.t. for disjoint sets, T_1, T_2 with $|T_1| \le S_1$, $|T_2| \le S_2$,*

$$|c_1' {A_{T_1}}' A_{T_2} c_2| \le \theta_{S_1,S_2} \|c_1\|_2 \; \|c_2\|_2$$ [Candes,Romberg,Tao'05]

Notation [Candes,Romberg,Tao'05]

- Notation:
  - $|T|$: cardinality of set T
  - $T^c = [1, 2, \dots m] \setminus T$: complement of set T
  - $\|\beta\|_k$: ℓ_k norm of vector β, $\|\beta\|$: ℓ_2 norm
  - $\|A\|$: spectral matrix norm (induced 2-norm) of matrix A
  - β_T: sub-vector containing elements of β with indices in set T
  - A_T: sub-matrix containing columns of A with indices in set T
  - A': denotes the transpose of matrix A
- *RIP constant, δ_S: smallest real number s.t. all eigenvalues of $A_T{}'A_T$ lie b/w $1 \pm \delta_S$ whenever $|T| \leq S$* [Candes,Romberg,Tao'05]
  - *$\delta_S < 1 \Leftrightarrow$ A satisfies the S-RIP*
- *ROP constant, θ_{S_1,S_2}: smallest real number s.t. for disjoint sets, T_1, T_2 with $|T_1| \leq S_1$, $|T_2| \leq S_2$,*

  $$|c_1' A_{T_1}{}' A_{T_2} c_2| \leq \theta_{S_1,S_2} \|c_1\|_2 \ \|c_2\|_2$$ [Candes,Romberg,Tao'05]

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

Compressive sensing [Candes,Romberg,Tao'05][Donoho'05]

- ℓ_1 min approaches
  - Basis pursuit (BP) [Chen,Donoho,Saunders'97]: $\min_\beta \|\beta\|_1 \ s.t. \ y = A\beta$
  - BP denoising (BPDN): $\min_\beta \|\beta\|_1 \ s.t. \ \|y - A\beta\|_2 \le \epsilon$
  - BPDN - unconst.: $\gamma \min_\beta \|\beta\|_1 + \|y - A\beta\|_2^2$
  - Dantzig selector (DS): $\min_\beta \|\beta\|_1 \ s.t. \ \|A'(y - A\beta)\|_\infty < \lambda$

Compressive sensing [Candes,Romberg,Tao'05][Donoho'05]

- ℓ_1 min approaches
  - Basis pursuit (BP) [Chen,Donoho,Saunders'97]: $\min_\beta \|\beta\|_1 \ s.t. \ y = A\beta$
  - BP denoising (BPDN): $\min_\beta \|\beta\|_1 \ s.t. \ \|y - A\beta\|_2 \leq \epsilon$
  - BPDN - unconst.: $\gamma \min_\beta \|\beta\|_1 + \|y - A\beta\|_2^2$
  - Dantzig selector (DS): $\min_\beta \|\beta\|_1 \ s.t. \ \|A'(y - A\beta)\|_\infty < \lambda$
- If x is S-sparse and $\delta_{2S} + \theta_{S,2S} < 1$,
  - noiseless measurements: BP gives exact reconstruction
  - noisy meas.: DS or BPDN error can be bounded

    [Candes,Tao'06][Tropp'05]

Problem definition

- Measure

$$y_t = Ax_t \text{ (noise-free) or } \; y_t = Ax_t + w_t \text{ (noisy)}$$

  - $A = H\Phi$, H: measurement matrix, Φ: sparsity basis matrix
  - y_t: measurements $(n \times 1)$
  - x_t: sparsity basis coefficients $(m \times 1)$, $m > n$
  - N_t: support of x_t (set of indices of nonzero elements of x_t)
- Goal: recursively reconstruct x_t from $y_0, y_1, \ldots y_t$,
  - i.e. use only $\hat{x}_{t-1}$ and y_t for reconstructing x_t

Assumptions

- Measurement basis is “incoherent” w.r.t. sparsity basis
  - A satisfies S-RIP for $S > |N_t| + ?$
- x_t is sparse at each time with support denoted N_t

Assumptions

- Measurement basis is “incoherent” w.r.t. sparsity basis
  - A satisfies S-RIP for $S > |N_t| + ?$
- x_t is sparse at each time with support denoted N_t
- *Support changes slowly over time:*

$$|N_t \setminus N_{t-1}| \approx |N_{t-1} \setminus N_t| \ll |N_t|$$

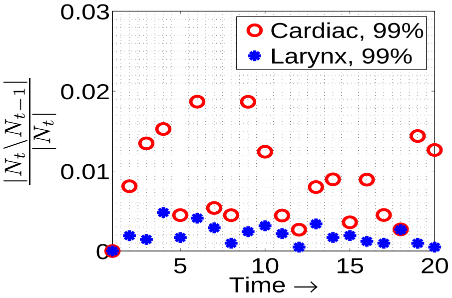
Assumptions

- Measurement basis is "incoherent" w.r.t. sparsity basis
  - A satisfies S-RIP for $S > |N_t| + ?$
- x_t is sparse at each time with support denoted N_t
- *Support changes slowly over time:*

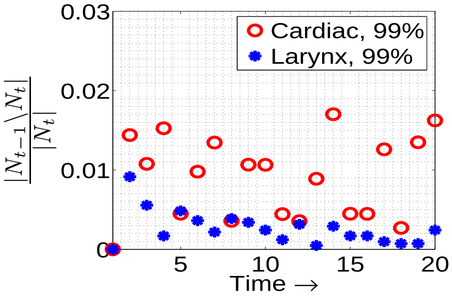
$$|N_t \setminus N_{t-1}| \approx |N_{t-1} \setminus N_t| \ll |N_t|$$

- *Usually nonzero elements of x_t also change slowly over time*

Slow support change in medical image sequences



(a) additions



(b) deletions

- N_t: 99%-energy support of the 2D-DWT of the image
- additions: $N_t \setminus N_{t-1}$, deletions: $N_{t-1} \setminus N_t$
- **maximum size of additions/deletions is less than** $0.02|N_t|$

Two Formulations

1. Only use “slow support change” assumption
   - $\Leftrightarrow$ **sparse reconstruction with partially known support**

Two Formulations

1. Only use "slow support change" assumption
   - $\Leftrightarrow$ **sparse reconstruction with partially known support**
   - two types of approaches
     - **LS-CS-residual (LS-CS)** [Vaswani, ICIP'08, IEEE Trans. SP (to appear)]
     - **Modified-CS (mod-CS)** [Vaswani, Lu, ISIT'09, IEEE Trans. SP (to appear)]

Two Formulations

1. Only use "slow support change" assumption
   - ⇔ **sparse reconstruction with partially known support**
   - two types of approaches
     - **LS-CS-residual (LS-CS)** [Vaswani, ICIP'08, IEEE Trans. SP (to appear)]
     - **Modified-CS (mod-CS)** [Vaswani, Lu, ISIT'09, IEEE Trans. SP (to appear)]
2. Also use "slow signal value change"
   - regularize both of the above approaches
     - KF-CS-residual (KF-CS) [Vaswani, ICIP'08, ICASSP'09]
     - KF-modCS (regularized mod-CS)

Two Formulations

1. Only use "slow support change" assumption
   - $\Leftrightarrow$ **sparse reconstruction with partially known support**
   - two types of approaches
     - **LS-CS-residual (LS-CS)** [Vaswani, ICIP'08, IEEE Trans. SP (to appear)]
     - **Modified-CS (mod-CS)** [Vaswani, Lu, ISIT'09, IEEE Trans. SP (to appear)]

2. Also use "slow signal value change"
   - regularize both of the above approaches
     - KF-CS-residual (KF-CS) [Vaswani, ICIP'08, ICASSP'09]
     - KF-modCS (regularized mod-CS)
   - significant performance improvement with fewer measurements

Two Formulations

1. Only use "slow support change" assumption
   - $\Leftrightarrow$ **sparse reconstruction with partially known support**
   - two types of approaches
     - **LS-CS-residual (LS-CS)** **[Vaswani, ICIP'08, IEEE Trans. SP (to appear)]**
     - **Modified-CS (mod-CS)** **[Vaswani, Lu, ISIT'09, IEEE Trans. SP (to appear)]**

2. Also use "slow signal value change"
   - regularize both of the above approaches
     - KF-CS-residual (KF-CS) [Vaswani, ICIP'08, ICASSP'09]
     - KF-modCS (regularized mod-CS)
   - significant performance improvement with fewer measurements
   - but difficult to analyze

Key Contributions [Vaswani, ICIP'08, Trans. SP (to appear)] [Vaswani, Lu, ISIT'09,Trans. SP (to appear)]

Assume: "slow support change"

Under *weaker sufficient conditions (fewer measurements)* than CS,

1. **Mod-CS achieves exact recon (noise-free case)**

Key Contributions [Vaswani, ICIP'08, Trans. SP (to appear)] [Vaswani, Lu, ISIT'09,Trans. SP (to appear)]

Assume: "slow support change"

Under *weaker sufficient conditions (fewer measurements)* than CS,

1. **Mod-CS achieves exact recon (noise-free case)**
2. Noisy case: error bound is significantly smaller than CS
   - if previous recon accurate enough

Key Contributions [Vaswani, ICIP'08, Trans. SP (to appear)] [Vaswani, Lu, ISIT'09,Trans. SP (to appear)]

Assume: "slow support change"

Under *weaker sufficient conditions (fewer measurements)* than CS,

1. **Mod-CS achieves exact recon (noise-free case)**
2. Noisy case: error bound is significantly smaller than CS
   - if previous recon accurate enough
   - holds for both Mod-CS & LS-CS

Key Contributions [Vaswani, ICIP'08, Trans. SP (to appear)] [Vaswani, Lu, ISIT'09, Trans. SP (to appear)]

Assume: "slow support change"

Under *weaker sufficient conditions (fewer measurements)* than CS,

1. **Mod-CS achieves exact recon (noise-free case)**
2. Noisy case: error bound is significantly smaller than CS
   - if previous recon accurate enough
   - holds for both Mod-CS & LS-CS
3. **Noisy case: error is "stable" over time**
   - time-invariant bound on support errors (misses/extras) & hence on recon errors; support error bound $\ll$ support size

Key Contributions [Vaswani, ICIP'08, Trans. SP (to appear)] [Vaswani, Lu, ISIT'09,Trans. SP (to appear)]

Assume: "slow support change"

Under *weaker sufficient conditions (fewer measurements)* than CS,

1. **Mod-CS achieves exact recon (noise-free case)**
2. Noisy case: error bound is significantly smaller than CS
   - if previous recon accurate enough
   - holds for both Mod-CS & LS-CS
3. **Noisy case: error is "stable" over time**
   - time-invariant bound on support errors (misses/extras) & hence on recon errors; support error bound $\ll$ support size
   - holds for both Mod-CS & LS-CS

Key Contributions [Vaswani, ICIP'08, Trans. SP (to appear)] [Vaswani, Lu, ISIT'09,Trans. SP (to appear)]

Assume: "slow support change"

Under *weaker sufficient conditions (fewer measurements)* than CS,

1. **Mod-CS achieves exact recon (noise-free case)**
2. Noisy case: error bound is significantly smaller than CS
   - if previous recon accurate enough
   - holds for both Mod-CS & LS-CS
3. **Noisy case: error is "stable" over time**
   - time-invariant bound on support errors (misses/extras) & hence on recon errors; support error bound $\ll$ support size
   - holds for both Mod-CS & LS-CS
4. Demonstrated all the above for recon'ing real image sequences (approx. sparse) from both partial Fourier (MRI) & Gaussian meas's

Related Work

- Batch CS [Wakin et al (video)],[Gamper et al,Jan'08 (MRI)],[Jung et al'09 (MRI)]
  - non-causal, very high reconstruction complexity

Related Work

- Batch CS [Wakin et al (video)],[Gamper et al,Jan'08 (MRI)],[Jung et al'09 (MRI)]
  - non-causal, very high reconstruction complexity
- [Cevher et al, ECCV'08]: CS for background subtracted images (CS-diff)
  - $CS(y_t - y_{t-1})$: designed to only recon $x_t - x_{t-1}$
  - unstable if try to recon x_t from few measurements

Related Work

- Batch CS [Wakin et al (video)],[Gamper et al,Jan'08 (MRI)],[Jung et al'09 (MRI)]
  - non-causal, very high reconstruction complexity
- [Cevher et al, ECCV'08]: CS for background subtracted images (CS-diff)
  - $CS(y_t - y_{t-1})$: designed to only recon $x_t - x_{t-1}$
  - unstable if try to recon x_t from few measurements
- [von Borries et al,CAMSAP'07]: static CS with prior support knowledge
  - did not give any exact reconstruction guarantees or error bounds or experimental results

Parallel, later and not-so-related work

- Parallel work related to modified-CS [Vaswani,Lu, ISIT'09]
  - [Khajenejad et al, ISIT'09]: static CS with probabilistic prior on support

Parallel, later and not-so-related work

- Parallel work related to modified-CS [Vaswani,Lu, ISIT'09]
  - [Khajenejad et al, ISIT'09]: static CS with probabilistic prior on support
- Work related to KF-CS, LS-CS [Vaswani, ICIP'08]
  - [Angelosante,Giannakis, DSP'09]
    - focusses only on time-invariant support: restrictive
  - [Carmi et al, pseudo-measurement KF, IBM tech report'09]
    - modifies KF-CS [Vaswani, ICIP'08]

Parallel, later and not-so-related work

- Parallel work related to modified-CS [Vaswani,Lu, ISIT'09]
  - [Khajenejad et al, ISIT'09]: static CS with probabilistic prior on support
- Work related to KF-CS, LS-CS [Vaswani, ICIP'08]
  - [Angelosante,Giannakis, DSP'09]
    - focusses only on time-invariant support: restrictive
  - [Carmi et al, pseudo-measurement KF, IBM tech report'09]
    - modifies KF-CS [Vaswani, ICIP'08]
- **Our goals are very different from:**
  - homotopy methods e.g. [Asif,Romberg'09], [Rozell et al'07]
    - speed up optimization, do not reduce no. of meas's reqd.
  - reconstruct **one** signal recursively from seq. arriving meas's,
    - e.g. [Sequential CS,Maliotov et al,ICASSP'08], [Garrigues et al'08], [Asif,Romberg'08], [Angelosante,Giannakis,RLS-Lasso,ICASSP'09]
  - multiple measurements vector (MMV) problem

Outline

- Sparse reconstruction with partially known support
  - problem definition
  - LS-CS-residual and error bound
  - Modified-CS and exact reconstruction conditions
  - Stability over time
- Summary
- Ongoing work: KF-CS-residual, KF-mod-CS

Sparse reconstruction with partly known support

- Rewrite the support, N_t, as

$$N_t = T \cup \Delta \setminus \Delta_e$$

  - T: "known" part of the support at t, may have error
  - $\Delta_e := T \setminus N_t$: error in T, unknown
  - $\Delta := N_t \setminus T$: unknown part of support

Sparse reconstruction with partly known support

- Rewrite the support, N_t, as

$$N_t = T \cup \Delta \setminus \Delta_e$$

  - T: "known" part of the support at t, may have error
  - $\Delta_e := T \setminus N_t$: error in T, unknown
  - $\Delta := N_t \setminus T$: unknown part of support

- In our problem:
  - at $t > 0$, use $T = \hat{N}_{t-1}$

Sparse reconstruction with partly known support

- Rewrite the support, N_t, as

$$N_t = T \cup \Delta \setminus \Delta_e$$

  - T: "known" part of the support at t, may have error
  - $\Delta_e := T \setminus N_t$: error in T, unknown
  - $\Delta := N_t \setminus T$: unknown part of support

- In our problem:
  - at $t > 0$, use $T = \hat{N}_{t-1}$
  - if previous recon accurate enough and "slow support change",
    - $|\Delta_e|, |\Delta| \ll |N_t|$

Sparse reconstruction with partly known support

- Rewrite the support, N_t, as

$$N_t = T \cup \Delta \setminus \Delta_e$$

  - T: "known" part of the support at t, may have error
  - $\Delta_e := T \setminus N_t$: error in T, unknown
  - $\Delta := N_t \setminus T$: unknown part of support

- In our problem:
  - at $t > 0$, use $T = \hat{N}_{t-1}$
  - if previous recon accurate enough and "slow support change",
    - $|\Delta_e|, |\Delta| \ll |N_t|$
  - at $t = 0$, T = empty or use prior knowledge

- The problem is also of independent interest
- T may be available from prior knowledge
- Examples:
  1. piecewise smooth images with small black background
     - most approximation coefficients of its DWT are nonzero

- The problem is also of independent interest
- T may be available from prior knowledge
- Examples:
  1. piecewise smooth images with small black background
     - most approximation coefficients of its DWT are nonzero
  2. Fourier sparse signals/images: usually most low frequencies present

- The problem is also of independent interest
- T may be available from prior knowledge
- Examples:
  1. piecewise smooth images with small black background
     - most approximation coefficients of its DWT are nonzero
  2. Fourier sparse signals/images: usually most low frequencies present
  3. fMRI brain activation tracking: use initial frame support as "known part"

LS-CS-residual (LS-CS) [Vaswani, ICIP'08, ICASSP'09, Trans. SP (to appear)]

- Our problem: reconstruct x with support $N = T \cup \Delta \setminus \Delta_e$ from $y := Ax$ or from $y := Ax + w$, *when T is known*
- CS-residual idea:
  - compute an initial LS estimate assuming T is correct support

$$(\hat{x}_{\text{init}})_T = {A_T}^{\dagger} y$$
$$(\hat{x}_{\text{init}})_{T^c} = 0$$

LS-CS-residual (LS-CS) [Vaswani, ICIP'08, ICASSP'09, Trans. SP (to appear)]

- Our problem: reconstruct x with support $N = T \cup \Delta \setminus \Delta_e$ from $y := Ax$ or from $y := Ax + w$, *when T is known*
- CS-residual idea:
  - compute an initial LS estimate assuming T is correct support

$$(\hat{x}_{\text{init}})_T = {A_T}^\dagger y$$
$$(\hat{x}_{\text{init}})_{T^c} = 0$$

  - compute the observation residual

$$\tilde{y}_{\text{res}} = y - A\hat{x}_{\text{init}}$$

LS-CS-residual (LS-CS) [Vaswani, ICIP'08, ICASSP'09, Trans. SP (to appear)]

- Our problem: reconstruct x with support $N = T \cup \Delta \setminus \Delta_e$ from $y := Ax$ or from $y := Ax + w$, *when T is known*
- CS-residual idea:
  - compute an initial LS estimate assuming T is correct support

$$(\hat{x}_{\text{init}})_T = {A_T}^\dagger y$$
$$(\hat{x}_{\text{init}})_{T^c} = 0$$

  - compute the observation residual

$$\tilde{y}_{\text{res}} = y - A\hat{x}_{\text{init}}$$

  - CS on observation residual, add back $\hat{x}_{\text{init}}$

$$\hat{x}_{\text{CSres}} = CS(\tilde{y}_{\text{res}}) + \hat{x}_{\text{init}}$$

LS-CS-residual (LS-CS) [Vaswani, ICIP'08, ICASSP'09, Trans. SP (to appear)]

- Our problem: reconstruct x with support $N = T \cup \Delta \setminus \Delta_e$ from $y := Ax$ or from $y := Ax + w$, *when T is known*
- CS-residual idea:
  - compute an initial LS estimate assuming T is correct support

$$(\hat{x}_{\text{init}})_T = {A_T}^{\dagger} y$$
$$(\hat{x}_{\text{init}})_{T^c} = 0$$

  - compute the observation residual

$$\tilde{y}_{\text{res}} = y - A\hat{x}_{\text{init}}$$

  - CS on observation residual, add back $\hat{x}_{\text{init}}$

$$\hat{x}_{\text{CSres}} = CS(\tilde{y}_{\text{res}}) + \hat{x}_{\text{init}}$$

- Notice that $\tilde{y}_{\text{res}} = A\beta + w$, $\beta = x - \hat{x}_{\text{init}}$

Why CS-residual works better?

- Notice that $\tilde{y}_{\text{res}} = A\beta + w$, where $\beta = x - \hat{x}_{\text{init}}$ with

$$\begin{aligned}(\beta)_{(T\cup\Delta)^c} &= 0 \\ (\beta)_T &= (A_T{}'A_T)^{-1}A_T{}'(A_\Delta x_\Delta + w), \\ (\beta)_\Delta &= x_\Delta\end{aligned}$$

  - $|\Delta|$, $|\Delta_e|$ small $\Rightarrow \|A_T{}'A_\Delta\| \leq \theta_{|T|,|\Delta|}$ small

Why CS-residual works better?

- Notice that $\tilde{y}_{\text{res}} = A\beta + w$, where $\beta = x - \hat{x}_{\text{init}}$ with

$$\begin{aligned}(\beta)_{(T\cup\Delta)^c} &= 0 \\ (\beta)_T &= (A_T{}'A_T)^{-1}A_T{}'(A_\Delta x_\Delta + w), \\ (\beta)_\Delta &= x_\Delta\end{aligned}$$

  - $|\Delta|$, $|\Delta_e|$ small $\Rightarrow \|A_T{}'A_\Delta\| \le \theta_{|T|,|\Delta|}$ small
  - if noise also small $\Rightarrow$ second term in $(\beta)_T$ equation is small

Why CS-residual works better?

- Notice that $\tilde{y}_{\text{res}} = A\beta + w$, where $\beta = x - \hat{x}_{\text{init}}$ with

$$
\begin{aligned}
(\beta)_{(T\cup\Delta)^c} &= 0 \\
(\beta)_T &= ({A_T}' A_T)^{-1} {A_T}' (A_\Delta x_\Delta + w), \\
(\beta)_\Delta &= x_\Delta
\end{aligned}
$$

  - $|\Delta|, |\Delta_e|$ small $\Rightarrow \|{A_T}' A_\Delta\| \le \theta_{|T|,|\Delta|}$ small
  - if noise also small $\Rightarrow$ second term in $(\beta)_T$ equation is small
  - $|\Delta_e|$ small $\Rightarrow \|({A_T}' A_T)^{-1}\| \le (1 - \delta_{|T|})^{-1}$ not large

Why CS-residual works better?

- Notice that $\tilde{y}_{\text{res}} = A\beta + w$, where $\beta = x - \hat{x}_{\text{init}}$ with

$$\begin{aligned}(\beta)_{(T \cup \Delta)^c} &= 0 \\ (\beta)_T &= (A_T{}'A_T)^{-1}A_T{}'(A_\Delta x_\Delta + w), \\ (\beta)_\Delta &= x_\Delta\end{aligned}$$

  - $|\Delta|$, $|\Delta_e|$ small $\Rightarrow \|A_T{}'A_\Delta\| \leq \theta_{|T|,|\Delta|}$ small
  - if noise also small $\Rightarrow$ second term in $(\beta)_T$ equation is small
  - $|\Delta_e|$ small $\Rightarrow \|(A_T{}'A_T)^{-1}\| \leq (1-\delta_{|T|})^{-1}$ not large
  - all the above $\Rightarrow \|(\beta)_T\|$ small $\Rightarrow \beta$ is approx $|\Delta|$-sparse

Why CS-residual works better?

- Notice that $\tilde{y}_{\text{res}} = A\beta + w$, where $\beta = x - \hat{x}_{\text{init}}$ with

$$
\begin{aligned}
(\beta)_{(T\cup\Delta)^c} &= 0 \\
(\beta)_T &= (A_T{}'A_T)^{-1}A_T{}'(A_\Delta x_\Delta + w), \\
(\beta)_\Delta &= x_\Delta
\end{aligned}
$$

  - $|\Delta|$, $|\Delta_e|$ small $\Rightarrow \|A_T{}'A_\Delta\| \le \theta_{|T|,|\Delta|}$ small
  - if noise also small $\Rightarrow$ second term in $(\beta)_T$ equation is small
  - $|\Delta_e|$ small $\Rightarrow \|(A_T{}'A_T)^{-1}\| \le (1-\delta_{|T|})^{-1}$ not large
  - all the above $\Rightarrow \|(\beta)_T\|$ small $\Rightarrow$ β is approx $|\Delta|$-sparse
  - β **is "sparse-compressible"**

Why CS-residual works better?

- Notice that $\tilde{y}_{\text{res}} = A\beta + w$, where $\beta = x - \hat{x}_{\text{init}}$ with

$$\begin{aligned}(\beta)_{(T\cup\Delta)^c} &= 0\\ (\beta)_T &= (A_T{}'A_T)^{-1}A_T{}'(A_\Delta x_\Delta + w),\\ (\beta)_\Delta &= x_\Delta\end{aligned}$$

  - $|\Delta|$, $|\Delta_e|$ small $\Rightarrow \|A_T{}'A_\Delta\| \le \theta_{|T|,|\Delta|}$ small
  - if noise also small $\Rightarrow$ second term in $(\beta)_T$ equation is small
  - $|\Delta_e|$ small $\Rightarrow \|(A_T{}'A_T)^{-1}\| \le (1-\delta_{|T|})^{-1}$ not large
  - all the above $\Rightarrow \|(\beta)_T\|$ small $\Rightarrow \beta$ is approx $|\Delta|$-sparse
  - β **is "sparse-compressible"**
- $\Rightarrow$ CS-residual error much smaller than CS error, with fewer measurements,

Why CS-residual works better?

- Notice that $\tilde{y}_{\text{res}} = A\beta + w$, where $\beta = x - \hat{x}_{\text{init}}$ with

$$\begin{aligned}(\beta)_{(T\cup\Delta)^c} &= 0\\ (\beta)_T &= (A_T{}'A_T)^{-1}A_T{}'(A_\Delta x_\Delta + w),\\ (\beta)_\Delta &= x_\Delta\end{aligned}$$

  - $|\Delta|$, $|\Delta_e|$ small $\Rightarrow \|A_T{}'A_\Delta\| \le \theta_{|T|,|\Delta|}$ small
  - if noise also small $\Rightarrow$ second term in $(\beta)_T$ equation is small
  - $|\Delta_e|$ small $\Rightarrow \|(A_T{}'A_T)^{-1}\| \le (1-\delta_{|T|})^{-1}$ not large
  - all the above $\Rightarrow \|(\beta)_T\|$ small $\Rightarrow$ β is approx $|\Delta|$-sparse
  - **β is "sparse-compressible"**

- $\Rightarrow$ CS-residual error much smaller than CS error, with fewer measurements, when support errors are small

Reconstruction error bound [Vaswani, Trans. SP (to appear)]

- **Bound reconstruction error as a function of $|T|, |\Delta|$**
  - L1: obtain error bound for CS on sparse-compressible vectors
  - $(\beta)_T$ is "compressible" part of $\beta := x - \hat{x}_{\text{init}}$
  - bound $\|(\beta)_T\|_1$ and apply L1

Reconstruction error bound [Vaswani, Trans. SP (to appear)]

- **Bound reconstruction error as a function of $|T|, |\Delta|$**
  - L1: obtain error bound for CS on sparse-compressible vectors
  - $(\beta)_T$ is "compressible" part of $\beta := x - \hat{x}_{\text{init}}$
  - bound $\|(\beta)_T\|_1$ and apply L1
- **Comparison with CS:** if $|\Delta|, |\Delta_e| \ll |N|$
  - our bound holds under weaker sufficient cond's (fewer meas.)

Reconstruction error bound [Vaswani, Trans. SP (to appear)]

- **Bound reconstruction error as a function of $|T|, |\Delta|$**
  - L1: obtain error bound for CS on sparse-compressible vectors
  - $(\beta)_T$ is "compressible" part of $\beta := x - \hat{x}_{\text{init}}$
  - bound $\|(\beta)_T\|_1$ and apply L1
- **Comparison with CS:** if $|\Delta|, |\Delta_e| \ll |N|$
  - our bound holds under weaker sufficient cond's (fewer meas.)
  - under these sufficient conditions,
    - possible to obtain another CS error bound
    - can argue: our bound is smaller

recall: T: "known" support, Δ: unknown part of support, Δ_e: error in known part

Simulations: normalized MSE [Vaswani, Trans. SP (to appear)]

So far: only upper bound comparison, need simulations

Simulations: normalized MSE [Vaswani, Trans. SP (to appear)]

So far: only upper bound comparison, need simulations

- Simulation setup
  - signal length $m = 200$, $|N| = 20$, $|\Delta| = |\Delta_e| = 2$
  - $(x_N)_i$ i.i.d ± 1 w.p $1/2$
  - noise: zero mean Gaussian, vary σ^2 and n (no. of meas's)
  - compare with Dantzig selector (DS) with various choices of λ

Simulations: normalized MSE [Vaswani, Trans. SP (to appear)]

So far: only upper bound comparison, need simulations

- Simulation setup
  - signal length $m = 200$, $|N| = 20$, $|\Delta| = |\Delta_e| = 2$
  - $(x_N)_i$ i.i.d ± 1 w.p $1/2$
  - noise: zero mean Gaussian, vary σ^2 and n (no. of meas's)
  - compare with Dantzig selector (DS) with various choices of λ

	$n = 59$, $\sigma = 0.04$	$n = 59$, $\sigma = 0.09$	$n = 59$, $\sigma = 0.44$	$n = 100$, $\sigma = 0.04$
DS, $\lambda = 4\sigma$	0.6545	0.6759	0.9607	0.2622
DS, $\lambda = 0.4\sigma$	0.5375	0.5479	1.0525	**0.0209**
CS-residual, $\lambda = 4\sigma$	**0.0866**	**0.1069**	**0.1800**	0.0402
CS-residual w/ add-then-del	0.0044	0.0205	0.1793	0.0032

$n = 59$: **CS-residual error much smaller than CS error**

Support estimation [Vaswani, ICIP'08, Trans. SP (to appear)]

- Option 1: $\hat{N} = \{i : |(\hat{x}_{\text{CSres}})_i| > \alpha_a\}$

Support estimation [Vaswani, ICIP'08, Trans. SP (to appear)]

- Option 1: $\hat{N} = \{i : |(\hat{x}_{\text{CSres}})_i| > \alpha_a\}$
  - difficulty: CS output biased towards zero [Candes,Tao'06]
    - $\Rightarrow$ $\hat{x}_{\text{CSres}}$ biased towards zero along Δ, away from zero along T
    - $\Rightarrow$ need small threshold for addition, large for deletion

Support estimation [Vaswani, ICIP'08, Trans. SP (to appear)]

- Option 1: $\hat{N} = \{i : |(\hat{x}_{\text{CSres}})_i| > \alpha_a\}$
  - difficulty: CS output biased towards zero [Candes,Tao'06]
    - $\Rightarrow$ $\hat{x}_{\text{CSres}}$ biased towards zero along Δ, away from zero along T
    - $\Rightarrow$ need small threshold for addition, large for deletion
- Option 2: Add-then-Delete

$$\begin{aligned} T_{\text{add}} &= T \cup \{i : |(\hat{x}_{CSres})_i| > \alpha_{\text{add}}\} \\ \hat{x}_{\text{add}} &= LS(T_{\text{add}}, y_t) \\ \hat{N} &= T_{add} \setminus \{i : |(\hat{x}_{add})_i| \leq \alpha_{\text{del}}\} \end{aligned}$$

Support estimation [Vaswani, ICIP'08, Trans. SP (to appear)]

- Option 1: $\hat{N} = \{i : |(\hat{x}_{\text{CSres}})_i| > \alpha_a\}$
  - difficulty: CS output biased towards zero [Candes,Tao'06]
    - $\Rightarrow$ $\hat{x}_{\text{CSres}}$ biased towards zero along Δ, away from zero along T
    - $\Rightarrow$ need small threshold for addition, large for deletion
- Option 2: Add-then-Delete

$$\begin{aligned} T_{\text{add}} &= T \cup \{i : |(\hat{x}_{CSres})_i| > \alpha_{\text{add}}\} \\ \hat{x}_{\text{add}} &= LS(T_{\text{add}}, y_t) \\ \hat{N} &= T_{add} \setminus \{i : |(\hat{x}_{add})_i| \leq \alpha_{\text{del}}\} \end{aligned}$$

  - Advantage:
    - use α_{add} as small as possible: ensure LS step error small
    - if LS estimate accurate $\Rightarrow$ (a) deletion better, (b) α_{del} can be larger $\Rightarrow$ huge improvement in recon error

Support estimation [Vaswani, ICIP'08, Trans. SP (to appear)]

- Option 1: $\hat{N} = \{i : |(\hat{x}_{\text{CSres}})_i| > \alpha_a\}$
  - difficulty: CS output biased towards zero [Candes,Tao'06]
    - $\Rightarrow$ $\hat{x}_{\text{CSres}}$ biased towards zero along Δ, away from zero along T
    - $\Rightarrow$ need small threshold for addition, large for deletion
- Option 2: Add-then-Delete

$$
\begin{aligned}
T_{\text{add}} &= T \cup \{i : |(\hat{x}_{CSres})_i| > \alpha_{\text{add}}\} \\
\hat{x}_{\text{add}} &= LS(T_{\text{add}}, y_t) \\
\hat{N} &= T_{add} \setminus \{i : |(\hat{x}_{add})_i| \leq \alpha_{\text{del}}\}
\end{aligned}
$$

  - Advantage:
    - use α_{add} as small as possible: ensure LS step error small
    - if LS estimate accurate $\Rightarrow$ (a) deletion better, (b) α_{del} can be larger $\Rightarrow$ huge improvement in recon error
  - Similar idea also introduced in [Needell-Tropp,CoSaMP'08]

LS-CS-residual (LS-CS) algo [Vaswani, Trans. SP (to appear)]

At each time t,

- *Initial LS.*
  - compute $\hat{x}_{t,\text{init}} = \text{LS}(T, y_t)$
  - compute residual, $\tilde{y}_{t,\text{res}} = y_t - A\hat{x}_{t,\text{init}}$
- *CS-residual.*
  - compute $\hat{x}_{t,\text{CSres}} = \text{CS}(\tilde{y}_{t,\text{res}}) + \hat{x}_{t,\text{init}}$
- *Support Additions and LS.*
  - compute $\tilde{T}_{\text{add}} = T \cup \text{threshold}(\hat{x}_{t,\text{CSres}}, \alpha_{\text{add}})$
  - compute $\hat{x}_{t,\text{add}} = \text{LS}(\tilde{T}_{\text{add}}, y_t)$
- *Support Deletions and LS.*
  - compute $\hat{N}_t = \tilde{T}_{\text{add}} \setminus \text{threshold}(\hat{x}_{t,\text{add}}, \alpha_{\text{del}})$
  - compute $\hat{x}_t = \text{LS}(\hat{N}_t, y_t)$

Exact reconstruction from fewer noiseless measurements?

- Consider noise-free measurements, i.e. $y := Ax$.
- Can CS-residual achieve exact reconstruction using fewer measurements?

Exact reconstruction from fewer noiseless measurements?

- Consider noise-free measurements, i.e. $y := Ax$.
- Can CS-residual achieve exact reconstruction using fewer measurements?
- Answer: NO
  - No. of meas. needed for exact recon depends on support size
  - CS-residual reconstructs $\beta := x_t - \hat{x}_{t,\text{init}}$ from the LS residual
  - Support of β is $T \cup \Delta$ and $|T \cup \Delta| \geq |N|$ (support size of x)
- Need something else...

recall: T: "known" support, Δ: unknown part of support, Δ_e: error in known part

Modified-CS: noiseless measurements [Vaswani,Lu, ISIT'09, Trans. SP (to appear)]

- Our problem: reconstruct a sparse x with support $N = T \cup \Delta \setminus \Delta_e$ from $y := Ax$, *when T is known*

Modified-CS: noiseless measurements [Vaswani,Lu, ISIT'09, Trans. SP (to appear)]

- Our problem: reconstruct a sparse x with support $N = T \cup \Delta \setminus \Delta_e$ from $y := Ax$, *when T is known*
- Our idea: find a vector that is **sparsest outside of** T and satisfies the data constraint, i.e.

$$\min_{\beta} \|(\beta)_{T^c}\|_0 \ s.t. \ y = A\beta$$

Modified-CS: noiseless measurements [Vaswani,Lu, ISIT'09, Trans. SP (to appear)]

- Our problem: reconstruct a sparse x with support $N = T \cup \Delta \setminus \Delta_e$ from $y := Ax$, *when T is known*
- Our idea: find a vector that is **sparsest outside of T** and satisfies the data constraint, i.e.

$$\min_{\beta} \|(\beta)_{T^c}\|_0 \ s.t. \ y = A\beta$$

  - **Exact recon if** $\delta_{|N|+|\Delta_e|+|\Delta|} < 1$

Modified-CS: noiseless measurements [Vaswani,Lu, ISIT'09, Trans. SP (to appear)]

- Our problem: reconstruct a sparse x with support $N = T \cup \Delta \setminus \Delta_e$ from $y := Ax$, *when T is known*
- Our idea: find a vector that is **sparsest outside of** T and satisfies the data constraint, i.e.

$$\min_{\beta} \|(\beta)_{T^c}\|_0 \ s.t. \ y = A\beta$$

  - **Exact recon if** $\delta_{|N|+|\Delta_e|+|\Delta|} < 1$
  - ℓ_0-**CS** needs $\delta_{2|N|} < 1$

Modified-CS: noiseless measurements [Vaswani,Lu, ISIT'09, Trans. SP (to appear)]

- Our problem: reconstruct a sparse x with support $N = T \cup \Delta \setminus \Delta_e$ from $y := Ax$, *when T is known*
- Our idea: find a vector that is **sparsest outside of T** and satisfies the data constraint, i.e.

$$\min_\beta \|(\beta)_{T^c}\|_0 \ s.t. \ y = A\beta$$

  - **Exact recon if** $\delta_{|N|+|\Delta_e|+|\Delta|} < 1$
  - ℓ_0-**CS** needs $\delta_{2|N|} < 1$

- Replace ℓ_0 norm by ℓ_1 norm: get a convex problem:

$$\min_\beta \|(\beta)_{T^c}\|_1 \ s.t. \ y = A\beta \quad \textbf{(modified-CS)}$$

recall: T: known part of support, Δ: unknown part, Δ_e: error in known part

Exact recon with modified-CS [Vaswani,Lu, ISIT'09, Trans. SP (to appear)]

$$\min_{\beta} \|\beta_{T^c}\|_1 \; s.t. \; y = A\beta \quad \text{(modified-CS)}$$

Theorem

x is the unique minimizer of (modified-CS) if $\delta_{|T|+|\Delta|} < 1$ and

$$(\theta_{|\Delta|,|\Delta|} + \delta_{2|\Delta|} + \theta_{|\Delta|,2|\Delta|}) + (\delta_{|T|} + \theta^2_{|\Delta|,|T|} + 2\theta^2_{2|\Delta|,|T|}) < 1$$

recall: $T = N \cup \Delta_e \setminus \Delta$, T: known part of support, Δ: unknown part, Δ_e: error in known part

Comparing the sufficient conditions

- *Modified-CS needs* $\delta_{|T|+|\Delta|} < 1$ and

$$Mcond := (\delta_{2|\Delta|} + \theta_{|\Delta|,|\Delta|} + \theta_{|\Delta|,2|\Delta|}) + (\delta_{|T|} + \theta^2_{|\Delta|,|T|} + 2\theta^2_{2|\Delta|,|T|}) < 1$$

- *CS needs* [Decoding by LP, Candes,Tao'05]:

$$Ccond := \delta_{2|N|} + \theta_{|N|,|N|} + \theta_{|N|,2|N|} < 1$$

Comparing the sufficient conditions

- *Modified-CS needs* $\delta_{|T|+|\Delta|} < 1$ and

$$Mcond := (\delta_{2|\Delta|} + \theta_{|\Delta|,|\Delta|} + \theta_{|\Delta|,2|\Delta|}) + (\delta_{|T|} + \theta^2_{|\Delta|,|T|} + 2\theta^2_{2|\Delta|,|T|}) < 1$$

- *CS needs* [Decoding by LP, Candes,Tao'05]:

$$Ccond := \delta_{2|N|} + \theta_{|N|,|N|} + \theta_{|N|,2|N|} < 1$$

- If $|\Delta| \approx |\Delta_e| \ll |N|$ (typical for medical image seq's),

$$Mcond < Ccond$$

  - *the difference* $(Ccond - Mcond)$ *is larger when* n *is smaller*
  - *e.g. if* $n < 2|N|$, $Ccond > 1$, *but* $Mcond < 1$ *can hold*

recall: n is the number of measurements, T: known part of support, Δ: unknown part, Δ_e: error in known part

Comparison with the best sufficient conditions for CS

- CS gives exact recon if

$$\delta_{2|N|} < \sqrt{2} - 1 \quad \text{or} \quad \delta_{2|N|} + \delta_{3|N|} < 1$$ [Candes'08, Candes-Tao'06]

- Modified-CS gives exact recon if

$$2\delta_{2|\Delta|} + \delta_{3|\Delta|} + \delta_{|N|+|\Delta_e|-|\Delta|} + \delta^2_{|N|+|\Delta_e|} + 2\delta^2_{|N|+|\Delta_e|+|\Delta|} < 1$$

- use $\delta_{ck} \leq c\delta_{2k}$ [CoSaMP'08]
- If $|\Delta| = |\Delta_e| = 0.02|N|$ (typical in medical sequences),
  - **sufficient condition for CS:**

$$\delta_{2|\Delta|} < 1/241.5$$

  - **sufficient condition for modified-CS:**

$$\delta_{2|\Delta|} < 1/132.5$$

Simulations: probability of exact reconstruction

Simulation setup:

- signal length, $m = 256$, support size $s = |N| = 0.1m$
- use random-Gaussian A, varied n, $|\Delta|$ and $|\Delta_e|$
- for each choice, Monte Carlo averaged over N, $(x)_N$, Δ, Δ_e
- we say "works" (gives exact recon) if $\|x - \hat{x}\|_2 < 10^{-5}\|x\|_2$

Simulations: probability of exact reconstruction

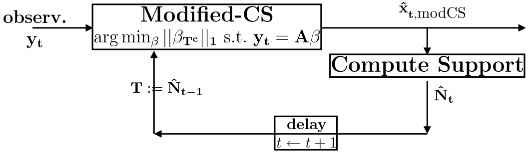
Simulation setup:

- signal length, $m = 256$, support size $s = |N| = 0.1m$
- use random-Gaussian A, varied n, $|\Delta|$ and $|\Delta_e|$
- for each choice, Monte Carlo averaged over N, $(x)_N$, Δ, Δ_e
- we say "works" (gives exact recon) if $\|x - \hat{x}\|_2 < 10^{-5}\|x\|_2$

n	mod-CS $\lvert\Delta\rvert, \lvert\Delta_e\rvert \le 0.08\lvert N\rvert$	mod-CS $\lvert\Delta\rvert, \lvert\Delta_e\rvert \le 0.20\lvert N\rvert$	CS $(\Delta = N, \Delta_e = \emptyset)$
19%	**0.998**	0.68	**0**
25%	1	0.99	0.002
40%	1	1	0.98

recall: n is number of measurements, Δ: unknown part of support, Δ_e: error in known part

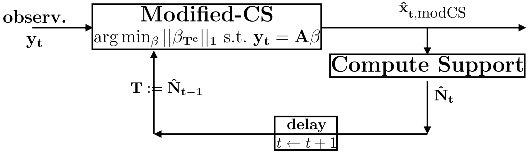
Modified-CS for time sequences



Initial time ($t = 0$):

- use T_0 from prior knowledge, e.g. wavelet approx. coeff's
- typically need more measurements at $t = 0$

Modified-CS for time sequences



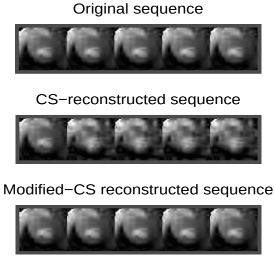
Initial time ($t = 0$):

- use T_0 from prior knowledge, e.g. wavelet approx. coeff's
- typically need more measurements at $t = 0$

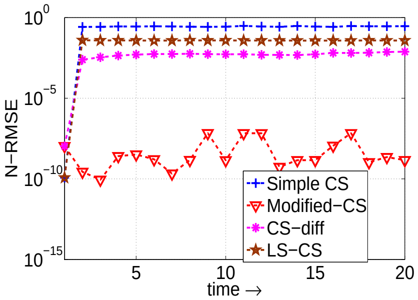
Stability: (trivial in the noise-free case)

- error stable at zero if $Mcond < 1$ at $t = 0$ and at all $t > 0$

Exact recon of a sparsified cardiac sequence



(c) $n_0 = 50\%$, $n = 16\%$



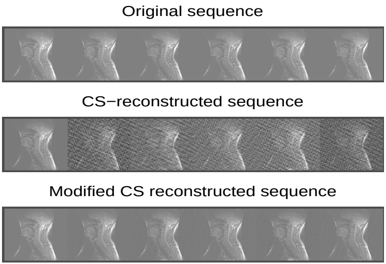
(d) $n_0 = 50\%$, $n = 16\%$

support size $\sim 10\%$

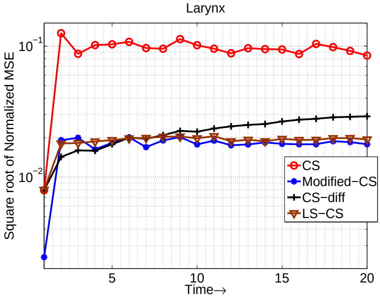
using $n = 16\%$ MRI measurements at $t > 0$, $n_0 = 50\%$ at $t = 0$.

modified-CS gives exact recon (NRMSE $\sim 10^{-8}$), others do not

Small error recon of a true larynx sequence



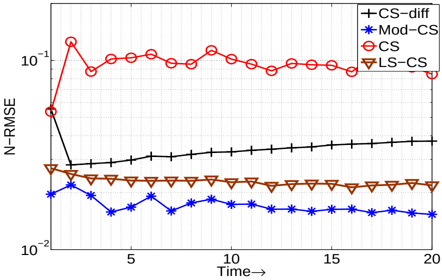
(e) $n_0 = 50\%,\ n = 19\%$



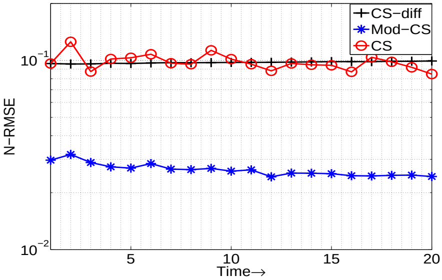
(f) $n_0 = 50\%,\ n = 19\%$

using $n = 19\%$ MRI measurements at $t > 0$, $n_0 = 50\%$ at $t = 0$

Small error recon of a true larynx sequence



(g) $n_0 = 20\%,\ n = 19\%$



(h) $n_0 = 19\%,\ n = 19\%$

reducing n_0 (no. of measurements at $t = 0$)

Noisy measurements

- Mod-CS(noisy): relax the data constraint, e.g.

$$\min_{\beta} \|\beta_{T^c}\|_1 \text{ s.t. } \|y_t - A\beta\|_2 \leq \epsilon$$

  - use add-then-delete for support estimation

Noisy measurements

- Mod-CS(noisy): relax the data constraint, e.g.

$$\min_{\beta} \|\beta_{T^c}\|_1 \text{ s.t. } \|y_t - A\beta\|_2 \leq \epsilon$$

  - use add-then-delete for support estimation

- Noisy meas's: Mod-CS or LS-CS-residual or Mod-CS-residual (ongoing work)

Noisy measurements

- Mod-CS(noisy): relax the data constraint, e.g.

$$\min_{\beta} \|\beta_{T^c}\|_1 \text{ s.t. } \|y_t - A\beta\|_2 \leq \epsilon$$

  - use add-then-delete for support estimation

- Noisy meas's: Mod-CS or LS-CS-residual or Mod-CS-residual (ongoing work)

- Easy to bound error as a function of $|T|, |\Delta|$ [Lu,Vaswani,ICASSP'10], [Jacques,Arxiv'09]
  - **but $|T|, |\Delta|$ depend on accuracy of previous recon**
  - **bound may keep increasing over time – limited use**

Stability [Vaswani, Trans. SP (to appear)]

- A bound that may keep increasing over time – limited use

Stability [Vaswani, Trans. SP (to appear)]

- A bound that may keep increasing over time – limited use
- Need conditions under which a time-invariant bound holds
  - **i.e. need conditions for "stability" over time**

Stability [Vaswani, Trans. SP (to appear)]

- A bound that may keep increasing over time – limited use
- Need conditions under which a time-invariant bound holds
  - **i.e. need conditions for "stability" over time**
- **Approach:**
  - *obtain a time-invariant bound on the support errors (extras & misses)*
  - *argue: bound small compared to support size*
  - directly implies time-invariant and small bound on recon error

Stability of modified-CS and LS-CS: setting

- Bounded measurement noise.
  - Why? -
    - Gaussian noise: error bounds at t hold with “large” probability
    - stability: need the bounds to hold for all t: will hold w.p. zero

Stability of modified-CS and LS-CS: setting

- Bounded measurement noise.
  - Why? -
    - Gaussian noise: error bounds at t hold with "large" probability
    - stability: need the bounds to hold for all t: will hold w.p. zero
- Two signal models with
  - two different support change models
    1. S_a elements added and deleted *"every-so-often"*
    2. S_a elements added and deleted *at every time*

Stability of modified-CS and LS-CS: setting

- Bounded measurement noise.
  - Why? -
    - Gaussian noise: error bounds at t hold with "large" probability
    - stability: need the bounds to hold for all t: will hold w.p. zero
- Two signal models with
  - two different support change models
    1. S_a elements added and deleted *"every-so-often"*
    2. S_a elements added and deleted *at every time*
  - almost constant signal power,
  - slowly increasing/decreasing coeff. magnitudes

Stability of modified-CS and LS-CS: setting

- Bounded measurement noise.
  - Why? -
    - Gaussian noise: error bounds at t hold with "large" probability
    - stability: need the bounds to hold for all t: will hold w.p. zero
- Two signal models with
  - two different support change models
    1. S_a elements added and deleted *"every-so-often"*
    2. S_a elements added and deleted *at every time*
  - almost constant signal power,
  - slowly increasing/decreasing coeff. magnitudes
- In case of 1.: perfect support estimation possible after a small delay

Stability of modified-CS and LS-CS: summary [Vaswani, Trans. SP (to appear)]

- For a given n (no. of meas.) and noise level,
  1. if use enough measurements for accurate recon at $t = 0$
  2. if
     - the support is small enough, and
     - the support changes slowly enough,
  3. if the nonzero coefficients increase/decrease fast enough, and
  4. if addition & deletion thresholds are appropriately set,

  then

  - support errors (no. of extras and misses) are bounded

Stability of modified-CS and LS-CS: summary [Vaswani, Trans. SP (to appear)]

- For a given n (no. of meas.) and noise level,
  1. if use enough measurements for accurate recon at $t = 0$
  2. if
     - the support is small enough, and
     - the support changes slowly enough,
  3. if the nonzero coefficients increase/decrease fast enough, and
  4. if addition & deletion thresholds are appropriately set,

  then
  - support errors (no. of extras and misses) are bounded
- **Can argue: our sufficient conditions allow larger support sizes, for a given n, than CS results**

Result 1: support changes every-so-often [Vaswani, Trans. SP (to appear)]

- Signal model:
  - S_a additions (removals) to (from) support every d frames
  - support size is always either S_0 or $S_0 - S_a$
  - the magnitude of the i^{th} new coeff increases at rate a_i for d time units and then becomes constant
  - similar model for coeff decrease

Result 1: support changes every-so-often [Vaswani, Trans. SP (to appear)]

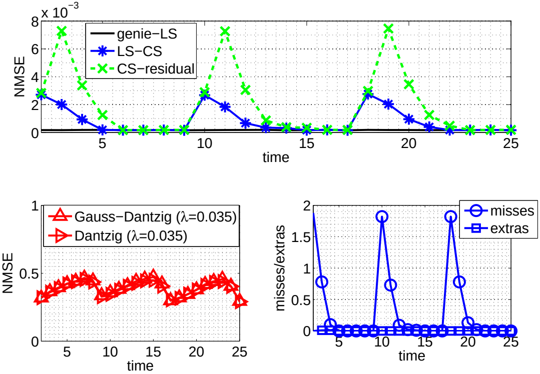
- Signal model:
  - S_a additions (removals) to (from) support every d frames
  - support size is always either S_0 or $S_0 - S_a$
  - the magnitude of the i^{th} new coeff increases at rate a_i for d time units and then becomes constant
  - similar model for coeff decrease
- If "conditions" hold, then
  - **at all times, misses and extras are bounded:**

$$|N_t \setminus \hat{N}_t| \leq S_a, \text{ and } |\hat{N}_t \setminus N_t| \leq 2S_a + 4$$

  - **within a short delay, $S_a + 2$, after a new addition, $\hat{N}_t = N_t$**
  - **remains this way until next addition time**

substituting $d_0 = 2$ in [Vaswani, LS-CS-residual, TSP (to appear)]

Simulations: verifying stability



- $m = 200$, support size, $S_0 = 20$
- $S_a = 2$ additions/removals every $d = 8$ frames
- 29.5% measurements at $t > 0$, noise $\sim unif(-c, c)$, $c = 0.05$

Proof strategy: key lemmas [Vaswani, Trans. SP (to appear)]

Obtain sufficient conditions so that

Proof strategy: key lemmas [Vaswani, Trans. SP (to appear)]

Obtain sufficient conditions so that

- (few false adds) there are S_a or less false adds per unit time

Proof strategy: key lemmas [Vaswani, Trans. SP (to appear)]

Obtain sufficient conditions so that

- (few false adds) there are S_a or less false adds per unit time
- *(detection)* all current coeff's with magnitude more than b **get detected**

Proof strategy: key lemmas [Vaswani, Trans. SP (to appear)]

Obtain sufficient conditions so that

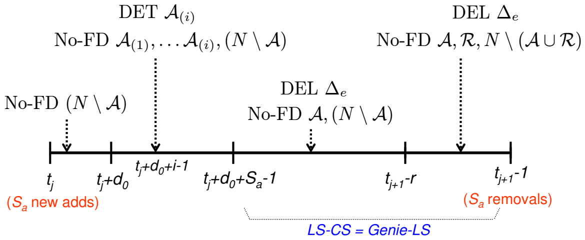
- (few false adds) there are S_a or less false adds per unit time
- *(detection)* all current coeff's with magnitude more than b **get detected**
- *(no false deletion)* these and previously added large coeff's **do not get falsely deleted**

Proof strategy: key lemmas [Vaswani, Trans. SP (to appear)]

Obtain sufficient conditions so that

- (few false adds) there are S_a or less false adds per unit time
- *(detection)* all current coeff's with magnitude more than b **get detected**
- *(no false deletion)* these and previously added large coeff's **do not get falsely deleted**
- *(true deletion)* all extras in support estimate (zero coeff's) **do get deleted**

Proof strategy: induction step [Vaswani, Trans. SP (to appear)]



DET: detection No-FD: no false deletion DEL: true deletion

- t_j : j^{th} support addition time, $t_{j+1} - 1$: support removal time
- $\mathcal{A}$: set added at t_j (increasing coeff's),
  - $\mathcal{A}_{(i)}$: i^{th} largest element of $\mathcal{A}$
- $N \setminus \mathcal{A}$: previously added set (constant coeff's)

Result 2: support changes at every time

- Signal model 2:
  - S_a additions and S_a removals **at each time**
  - support size constant at S_0
  - every new coeff's magnitude increases at rate r until it reaches a max value M
  - similar model for coeff decrease
- Noise, $\|w\| \leq \epsilon$
- If
  1. accurate recon at initial time,
  2. $\delta_{S_0+4S_a} < 0.414$ and $\theta_{S_0+2S_a,S_a} < 1/\sqrt{18S_a}$.
     - if LS error bound equal in all directions: only need $\theta_{S_0+2S_a,S_a} < \sqrt{S_0/18S_a}$
  3. $r > f_{incr}(S_0, S_a, \epsilon, \alpha_{add}, \alpha_{del})$,
  4. α_{add}, α_{del} large enough,

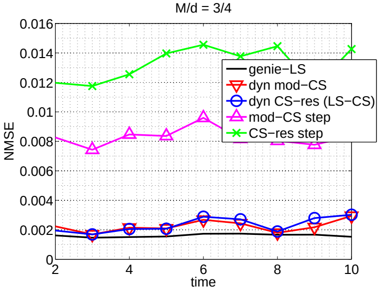
then $|N_t \setminus \hat{N}_t| \leq 2S_a$ **and** $|\hat{N}_t \setminus N_t| = 0$

Result 2: support changes at every time

- Signal model 2:
  - S_a additions and S_a removals **at each time**
  - support size constant at S_0
  - every new coeff's magnitude increases at rate r until it reaches a max value M
  - similar model for coeff decrease
- Noise, $\|w\| \leq \epsilon$
- If
  1. accurate recon at initial time,
  2. $\delta_{S_0+4S_a} < 0.414$ and $\theta_{S_0+2S_a,S_a} < 1/\sqrt{18S_a}$.
     - if LS error bound equal in all directions: only need $\theta_{S_0+2S_a,S_a} < \sqrt{S_0/18S_a}$
  3. $r > f_{incr}(S_0, S_a, \epsilon, \alpha_{add}, \alpha_{del})$,
  4. α_{add}, α_{del} large enough,

  then $|N_t \setminus \hat{N}_t| \leq 2S_a$ **and** $|\hat{N}_t \setminus N_t| = 0$
- Compare: CS needs $\delta_{2S_0} < 0.414$

Simulations: verifying stability



(i) $r = 0.75$

- $m = 200$, $S_0 = 20$, additions/removals, $S_a = 2$ at each time
- 29.5% measurements at $t > 0$, noise $\sim unif(-c, c)$, $c = 0.1266$
- *CS error 22-30% in all cases*

Summary

$T := \hat{N}_{t-1}, \quad \mu_T := (\hat{x}_{t-1})_T$

- CS on observation residual
  - initial estimate: compute using $LS(y_t, T)$ or use previous recon or use KF'ed estimate

Summary

$T := \hat{N}_{t-1}, \quad \mu_T := (\hat{x}_{t-1})_T$

- CS on observation residual
  - initial estimate: compute using $LS(y_t, T)$ or use previous recon or use KF'ed estimate
  - compute observation residual, do CS on residual, add back initial estimate

Summary

$T := \hat{N}_{t-1}, \quad \mu_T := (\hat{x}_{t-1})_T$

- CS on observation residual
  - initial estimate: compute using $LS(y_t, T)$ or use previous recon or use KF'ed estimate
  - compute observation residual, do CS on residual, add back initial estimate
- Modified-CS
  - search for a signal that is sparsest outside of T and satisfies the data constraint

Summary

$T := \hat{N}_{t-1}, \quad \mu_T := (\hat{x}_{t-1})_T$

- CS on observation residual
  - initial estimate: compute using $LS(y_t, T)$ or use previous recon or use KF'ed estimate
  - compute observation residual, do CS on residual, add back initial estimate
- Modified-CS
  - search for a signal that is sparsest outside of T and satisfies the data constraint
  - noisy meas.'s: can combine with CS-residual idea
  - or can combine with KF (or regularized LS) idea

$$\min_{\beta} \|(\beta)_{T^c}\|_1 + \gamma \|y_t - A\beta\|_2^2 + \lambda \|(\beta)_T - \mu_T\|_2^2$$

Summary

$T := \hat{N}_{t-1}, \quad \mu_T := (\hat{x}_{t-1})_T$

- CS on observation residual
  - initial estimate: compute using $LS(y_t, T)$ or use previous recon or use KF'ed estimate
  - compute observation residual, do CS on residual, add back initial estimate
- Modified-CS
  - search for a signal that is sparsest outside of T and satisfies the data constraint
  - noisy meas.'s: can combine with CS-residual idea
  - or can combine with KF (or regularized LS) idea

$$\min_\beta \|(\beta)_{T^c}\|_1 + \gamma \|y_t - A\beta\|_2^2 + \lambda \|(\beta)_T - \mu_T\|_2^2$$

- Support estimation in either case
  - add-LS-delete is better than simple thresholding

- If support changes slowly enough,
- under much weaker sufficient conditions than CS,
  - modified-CS gives exact reconstruction
    - its stability proof is trivial
  - noisy meas's: LS-CS & modified-CS error is stable
  - noisy meas's: both error bounds *smaller* than CS bound

- If support changes slowly enough,
- under much weaker sufficient conditions than CS,
  - modified-CS gives exact reconstruction
    - its stability proof is trivial
  - noisy meas's: LS-CS & modified-CS error is stable
  - noisy meas's: both error bounds *smaller* than CS bound
- For dynamic MRI and video reconstruction,
  - significant improvement over CS, Gauss-CS and CS-diff
  - only slightly worse than batch methods (batch-CS, k-t-focuss)

Key References

- **Namrata Vaswani, LS-CS-residual (LS-CS): Compressive Sensing on the Least Squares Residual, To Appear in IEEE Trans. Signal Processing**
  - Namrata Vaswani, *Kalman Filtered Compressed Sensing*, IEEE Intl. Conf. Image Proc. (ICIP), 2008
  - Namrata Vaswani, *Analyzing Least Squares and Kalman Filtered Compressed Sensing*, IEEE Intl. Conf. Acous. Speech. Sig. Proc. (ICASSP), 2009
- **Namrata Vaswani and Wei Lu, Modified-CS: Modifying Compressive Sensing for Problems with Partially Known Support, To Appear in IEEE Trans. Signal Processing**
  - Namrata Vaswani and Wei Lu, *Modified-CS: Modifying Compressive Sensing for Problems with Partially Known Support*, IEEE Intl. Symp. Info. Theory (ISIT), 2009

Acknowledgements

- Research supported by NSF grants
  - CCF-0917015 and ECCS-0725849
- Ian Atkinson (UIC): functional MRI
- Graduate students
  - **Wei Lu**
  - Chenlu Qiu
  - Taoran Li
  - Samarjit Das
  - Fardad Raisali

Slow support and signal value change [Vaswani, ICIP'08, ICASSP'09]

- $\Leftrightarrow$ **Track a time sequence of signals with slowly changing "principal" directions (in a given sparsity basis) and slowly changing principal coefficient values**
  1. can we accurately detect the changes?
  2. can we compute/approximate the causal MMSE estimate?

Slow support and signal value change [Vaswani, ICIP'08, ICASSP'09]

- **$\Leftrightarrow$ Track a time sequence of signals with slowly changing "principal" directions (in a given sparsity basis) and slowly changing principal coefficient values**
  1. can we accurately detect the changes?
  2. can we compute/approximate the causal MMSE estimate?
- **Proposed solution 1: Kalman filtered CS-residual (KF-CS)**
  - KF on the coefficients in the current "principal" subspace

Slow support and signal value change [Vaswani, ICIP'08, ICASSP'09]

- **⇔ Track a time sequence of signals with slowly changing "principal" directions (in a given sparsity basis) and slowly changing principal coefficient values**
  1. can we accurately detect the changes?
  2. can we compute/approximate the causal MMSE estimate?
- **Proposed solution 1: Kalman filtered CS-residual (KF-CS)**
  - KF on the coefficients in the current "principal" subspace
  - CS on KF residual to add new "principal" directions (if any)
    - replacing "CS on LS residual" by "CS on KF residual"

Slow support and signal value change [Vaswani, ICIP'08, ICASSP'09]

- **$\Leftrightarrow$ Track a time sequence of signals with slowly changing "principal" directions (in a given sparsity basis) and slowly changing principal coefficient values**
  1. can we accurately detect the changes?
  2. can we compute/approximate the causal MMSE estimate?

- **Proposed solution 1: Kalman filtered CS-residual (KF-CS)**
  - KF on the coefficients in the current "principal" subspace
  - CS on KF residual to add new "principal" directions (if any)
    - replacing "CS on LS residual" by "CS on KF residual"
  - delete directions which have near-zero coefficients

Slow support and signal value change [Vaswani, ICIP'08, ICASSP'09]

- **⇔ Track a time sequence of signals with slowly changing "principal" directions (in a given sparsity basis) and slowly changing principal coefficient values**
  1. can we accurately detect the changes?
  2. can we compute/approximate the causal MMSE estimate?
- **Proposed solution 1: Kalman filtered CS-residual (KF-CS)**
  - KF on the coefficients in the current "principal" subspace
  - CS on KF residual to add new "principal" directions (if any)
    - replacing "CS on LS residual" by "CS on KF residual"
  - delete directions which have near-zero coefficients
- *KF-CS error is stable (so 1. and 2. hold): under strong assumptions*

Proposed solution 2: Regularized Modified-CS

- Most practical apps: the (significantly) nonzero elements of x_t also change slowly
- To also use this fact, we can solve

$$\min_{\beta} \|(\beta)_{T^c}\|_1 + \gamma \|(\beta)_T - \mu_T\|_2^2 \text{ s.t. } y_t = A\beta$$

with $T = \hat{N}_{t-1}, \quad \mu_T = (\hat{x}_{t-1})_T, \quad \mu_{T^c} = 0$

Proposed solution 2: Regularized Modified-CS

- Most practical apps: the (significantly) nonzero elements of x_t also change slowly
- To also use this fact, we can solve

$$\min_\beta \|(\beta)_{T^c}\|_1 + \gamma \|(\beta)_T - \mu_T\|_2^2 \text{ s.t. } y_t = A\beta$$

with $T = \hat{N}_{t-1}, \quad \mu_T = (\hat{x}_{t-1})_T, \quad \mu_{T^c} = 0$

- Above computes a *causal MAP estimate* if
  - posterior at $t-1$ approx by a Dirac delta at μ
  - $(x_t)_T$ are i.i.d. Gaussian with mean μ_T and variance σ^2,
  - $(x_t)_{T^c}$ are i.i.d. Laplacian with mean zero and scale b,
  - *and we set* $\gamma = b/2\sigma^2$

Regularized Modified CS (RegModCS): simulations

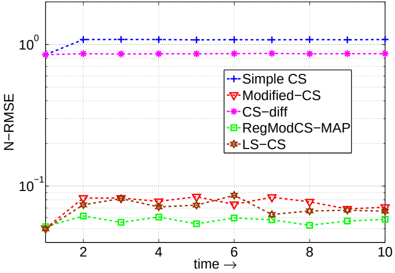
$$\min_{\beta} \|(\beta)_{T^c}\|_1 + \gamma\|y_t - A\beta\|_2^2 + \lambda\|(\beta)_T - \mu_T\|_2^2$$

with $T = \hat{N}_{t-1}, \quad \mu_T = (\hat{x}_{t-1})_T$

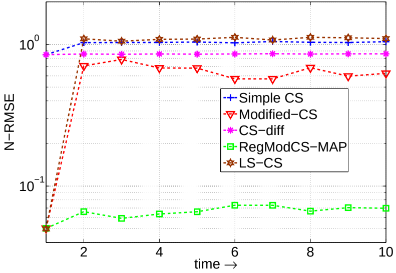
Regularized Modified CS (RegModCS): simulations

$$\min_{\beta} \|(\beta)_{T^c}\|_1 + \gamma \|y_t - A\beta\|_2^2 + \lambda \|(\beta)_T - \mu_T\|_2^2$$

with $T = \hat{N}_{t-1}, \quad \mu_T = (\hat{x}_{t-1})_T$



(l) $n_0 = n = 19\%$



(m) $n_0 = 19\%, \ n = 6\%$

Ongoing/Future Work

Most current work: non-Bayesian, only recursive approaches

Ongoing/Future Work

Most current work: non-Bayesian, only recursive approaches

Open questions:

- **Partly Bayesian and Fully Bayesian models:**
  - **KF-CS, reg-mod-CS: "stability" under mild assumptions?**
    - much more difficult: dependence on prev. recon signal values

Ongoing/Future Work

Most current work: non-Bayesian, only recursive approaches

Open questions:

- **Partly Bayesian and Fully Bayesian models:**
  - **KF-CS, reg-mod-CS: "stability" under mild assumptions?**
    - much more difficult: dependence on prev. recon signal values
  - **Fully Bayesian case: open...**

Ongoing/Future Work

Most current work: non-Bayesian, only recursive approaches

Open questions:

- **Partly Bayesian and Fully Bayesian models:**
  - **KF-CS, reg-mod-CS: "stability" under mild assumptions?**
    - much more difficult: dependence on prev. recon signal values
  - **Fully Bayesian case: open...**
- **"Optimal" causal estimate?**

Ongoing/Future Work

Most current work: non-Bayesian, only recursive approaches

Open questions:

- **Partly Bayesian and Fully Bayesian models:**
  - **KF-CS, reg-mod-CS: "stability" under mild assumptions?**
    - much more difficult: dependence on prev. recon signal values
  - **Fully Bayesian case: open...**
- **"Optimal" causal estimate?**
- **Real dynamic MRI problems** (ongoing: fMRI)