Recursive Sparse Recovery and Applications in Dynamic Imaging

Namrata Vaswani

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

(portions joint work with Wei Lu and Chenlu Qiu)

Talk Outline

Background on Sparse Recovery

Recursive Reconstruction of Sparse Signal Sequences (RecSparsRec)

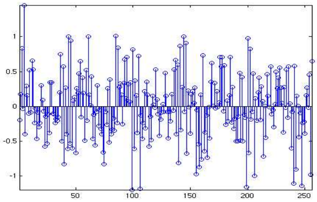
- The problem, motivation and applications, key ideas
- Modified-CS: noise-free case and exact recovery result
- Modified-CS: noisy case and time-invariant error bounds (stability)

Rec Robust PCA $\Leftrightarrow$ RecSparsRec in Large but Correlated Noise

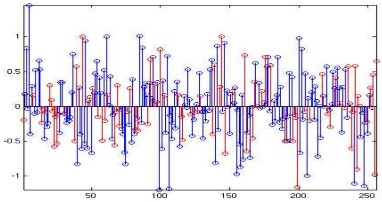
- Video Surveillance – Background subtraction application

Sparse Recovery: the question

- Can I recover a 256-length signal from only 80 samples?



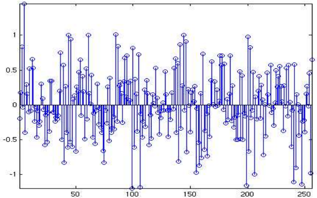
(a) the unknown signal



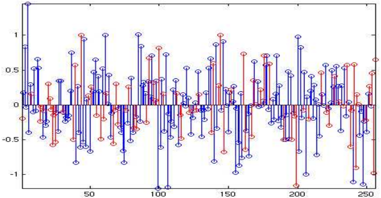
(b) its 80 time samples (red)

Sparse Recovery: the question

- Can I recover a 256-length signal from only 80 samples?



(c) the unknown signal



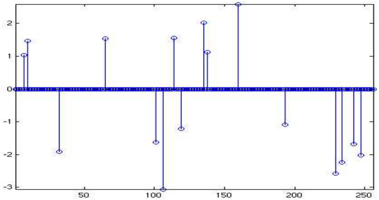
(d) its 80 time samples (red)

- Under certain situations: YES!
  - if it is bandlimited – use Nyquist
  - or if it is a weighted sum of only a few sinusoids – use sparsity

Example taken from L1-Magic webpage of Candes,Romberg,Tao

Sparse Recovery: the answer

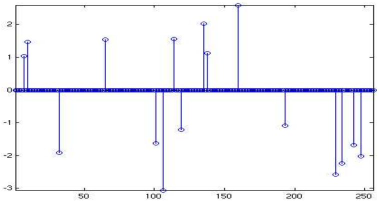
- This signal satisfies the latter – it is Fourier sparse



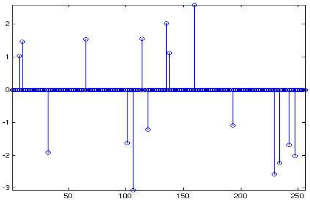
(e) DFT of original signal

Sparse Recovery: the answer

- This signal satisfies the latter – it is Fourier sparse



(g) DFT of original signal

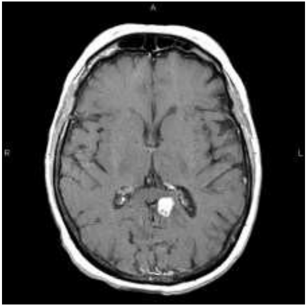


(h) recovered DFT: exact!

- We used its Fourier sparsity and ℓ_1 minimization to recover its DFT exactly!
  - one-to-one mapping between a signal and its DFT

Example taken from L1-Magic webpage of Candes,Romberg,Tao

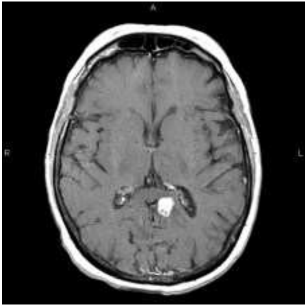
Sparse (or Compressible) Signals



a brain image:
wavelet compressible

- Sparse vector: only a few nonzero elements
- Compressible vector: approx sparse vector (most energy lies in only a few elements)
- Sparse (compressible) signal: either the signal or a linear transform of it is sparse (compress.)

Sparse (or Compressible) Signals



a brain image:
wavelet compressible

- Sparse vector: only a few nonzero elements
- Compressible vector: approx sparse vector (most energy lies in only a few elements)
- Sparse (compressible) signal: either the signal or a linear transform of it is sparse (compress.)
- Support: set of indices of the nonzero (non-negligible) elements of the vector,
  - e.g. 99%-energy support: set containing indices of the largest elements that make up 99% of the total energy

Sparse recovery [Mallat et al'93],[Chen,Donoho'95],[Candes,Romberg,Tao'05],[Donoho'05]

- Reconstruct a sparse signal x, with support N, from $y := Ax$,
  - when A has more columns than rows (underdetermined sys)

Sparse recovery [Mallat et al'93],[Chen,Donoho'95],[Candes,Romberg,Tao'05],[Donoho'05]

- Reconstruct a sparse signal x, with support N, from $y := Ax$,
  - when A has more columns than rows (underdetermined sys)
- Solved if we can find the sparsest vector satisfying $y = A\beta$, i.e.

$$\min_{\beta} \underbrace{\|\beta\|_0}_{\#\text{ of nonzero elements}} \quad \text{subject to } y = A\beta$$

Sparse recovery [Mallat et al'93],[Chen,Donoho'95],[Candes,Romberg,Tao'05],[Donoho'05]

- Reconstruct a sparse signal x, with support N, from $y := Ax$,
  - when A has more columns than rows (underdetermined sys)
- Solved if we can find the sparsest vector satisfying $y = A\beta$, i.e.

$$\min_{\beta} \underbrace{\|\beta\|_0}_{\#\text{ of nonzero elements}} \quad \text{subject to } y = A\beta$$

  - and any $S = 2|N|$ columns of A are linearly independent

Sparse recovery [Mallat et al'93],[Chen,Donoho'95],[Candes,Romberg,Tao'05],[Donoho'05]

- Reconstruct a sparse signal x, with support N, from $y := Ax$,
  - when A has more columns than rows (underdetermined sys)
- Solved if we can find the sparsest vector satisfying $y = A\beta$, i.e.

$$\min_{\beta} \underbrace{\|\beta\|_0}_{\#\text{ of nonzero elements}} \quad \text{subject to } y = A\beta$$

  - and any $S = 2|N|$ columns of A are linearly independent
  - but combinatorial search – $O(m^{|N|})$ complexity
- Practical approaches (polynomial complexity in m)
  - convex relaxation approaches [Chen,Donoho'95], ..., [Candes,Tao'06],...: ℓ_1 minimization
    - replace ℓ_0 norm by ℓ_1 norm – convex problem
  - greedy methods [Mallat,Zhang'93], [Pati et al'93], [Dai,Milenkovic'09], [Needell,Tropp'09]

Sparse recovery and Compressive Sensing

- Compressed Sensing (CS) literature [Candes,Romberg,Tao'05], [Donoho'05]
  - provides exact reconstruction conditions and error bounds for the practical approaches – much stronger results than earlier ones based on mutual coherence

Sparse recovery and Compressive Sensing

- Compressed Sensing (CS) literature [Candes,Romberg,Tao'05], [Donoho'05]
  - provides exact reconstruction conditions and error bounds for the practical approaches – much stronger results than earlier ones based on mutual coherence
- Restricted Isometry Constant (RIC), $\delta_s(A)$ [Candes,Romberg,Tao,T-IT'05]
  - quantifies approx orthogonality of any s-column sub-matrix of an $n \times m$ matrix A

Sparse recovery and Compressive Sensing

- Compressed Sensing (CS) literature [Candes,Romberg,Tao'05], [Donoho'05]
  - provides exact reconstruction conditions and error bounds for the practical approaches – much stronger results than earlier ones based on mutual coherence
- Restricted Isometry Constant (RIC), $\delta_s(A)$ [Candes,Romberg,Tao,T-IT'05]
  - quantifies approx orthogonality of any s-column sub-matrix of an $n \times m$ matrix A
  - $\delta_s(A)$ is the smallest real number s.t. singular values of any s-column sub-matrix of A lie in $[\sqrt{1-\delta_s}, \sqrt{1+\delta_s}]$

Sparse recovery and Compressive Sensing

- Compressed Sensing (CS) literature [Candes,Romberg,Tao'05], [Donoho'05]
  - provides exact reconstruction conditions and error bounds for the practical approaches – much stronger results than earlier ones based on mutual coherence
- Restricted Isometry Constant (RIC), $\delta_s(A)$ [Candes,Romberg,Tao,T-IT'05]
  - quantifies approx orthogonality of any s-column sub-matrix of an $n \times m$ matrix A
  - $\delta_s(A)$ is the smallest real number s.t. singular values of any s-column sub-matrix of A lie in $[\sqrt{1-\delta_s}, \sqrt{1+\delta_s}]$
  - non-increasing function of n (# of measurements)
  - random Gaussian matrices: if $n = O(S \log m)$, $\delta_S(A) < b < 1$ holds w.h.p. (prob $\to 0$ as $m \to \infty$)
    - similar results for Rademacher and partial Fourier matrices

Sparse recovery and Compressive Sensing

- Compressed Sensing (CS) literature [Candes,Romberg,Tao'05], [Donoho'05]
  - provides exact reconstruction conditions and error bounds for the practical approaches – much stronger results than earlier ones based on mutual coherence
- Restricted Isometry Constant (RIC), $\delta_s(A)$ [Candes,Romberg,Tao,T-IT'05]
  - quantifies approx orthogonality of any s-column sub-matrix of an $n \times m$ matrix A
  - $\delta_s(A)$ is the smallest real number s.t. singular values of any s-column sub-matrix of A lie in $[\sqrt{1-\delta_s}, \sqrt{1+\delta_s}]$
  - non-increasing function of n (# of measurements)
  - random Gaussian matrices: if $n = O(S \log m)$, $\delta_S(A) < b < 1$ holds w.h.p. (prob $\to 0$ as $m \to \infty$)
    - similar results for Rademacher and partial Fourier matrices
- this talk: sparse recovery ⇔ CS ⇔ ℓ_1 minimization

Recursive Sparse Recovery [Vaswani,ICIP'08][1]

- Recursive approaches for causally reconstructing a time sequence of sparse signals
- from a greatly reduced number of measurements at each time.
- "recursive": use only current measurement vector and the previous reconstructed signal to reconstruct the current signal

[1] N. Vaswani, Kalman Filtered Compressed Sensing, ICIP, 2008

Recursive Sparse Recovery [Vaswani,ICIP'08][1]

- Recursive approaches for causally reconstructing a time sequence of sparse signals
- from a greatly reduced number of measurements at each time.

- "recursive": use only current measurement vector and the previous reconstructed signal to reconstruct the current signal

- Sparsity patterns can change with time, but the changes are gradual

- Existing work: mostly batch CS approaches – expensive

[1] N. Vaswani, Kalman Filtered Compressed Sensing, ICIP, 2008

Potential Applications

- Dynamic medical imaging for real-time apps, e.g.
  - MRI-guided interventional radiology, MRI-guided surgery,
  - real-time functional MRI
- Video surveillance or denoising or fMRI based active region detection
  - track one or more moving objects/regions when the background scene itself is changing – foreground is sparse

Why "reduced" measurements?

- Projection Imaging, e.g. MRI or CT or single-pixel camera
  - Fourier transform or Radon transform or random-projections of the region-of-interest acquired sequentially
    - Fewer measurements $\Rightarrow$ faster scanning – needed for real-time imaging for fast changing phenomena

Why "reduced" measurements?

- Projection Imaging, e.g. MRI or CT or single-pixel camera
  - Fourier transform or Radon transform or random-projections of the region-of-interest acquired sequentially
    - Fewer measurements ⇒ faster scanning – needed for real-time imaging for fast changing phenomena
- Computer Vision
  - The full image is acquired in one go,

Why "reduced" measurements?

- Projection Imaging, e.g. MRI or CT or single-pixel camera
  - Fourier transform or Radon transform or random-projections of the region-of-interest acquired sequentially
    - Fewer measurements ⇒ faster scanning – needed for real-time imaging for fast changing phenomena
- Computer Vision
  - The full image is acquired in one go, but it can have more than one layers, e.g. foreground and background
    - both change, how can I estimate both?

Why "causal" and "recursive"

- Why causal?
  - reconstruct as soon as get data for current frame – desirable for real-time (or at most allow small buffering)

Why "causal" and "recursive"

- Why causal?
  - reconstruct as soon as get data for current frame – desirable for real-time (or at most allow small buffering)
- Why recursive?
  - one way to ensure computational and storage complexity is comparable to CS for one image (simple CS)
  - much faster and lower on memory than both causal and offline implementations of batch CS
  - recursive CS at time t v/s causal batch CS at time t
    - time: $O(1)$ v/s $O(t^3)$
    - memory: $O(1)$ v/s $O(t)$
    - $O(1)$: time taken or memory reqd for CS for one image

Problem Formulation [Vaswani,ICIP'08] (KF-CS)

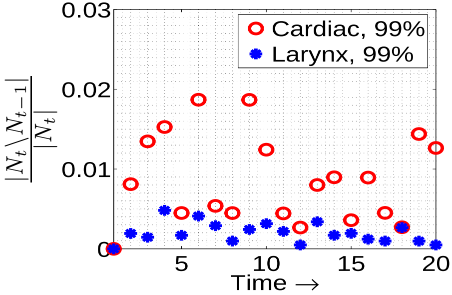
- Measure

$$y_t := Ax_t + w_t$$

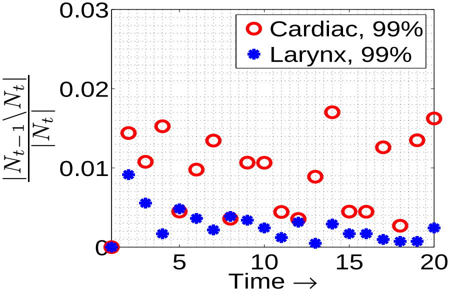
  - $A = H\Phi$ (given): $n \times m$, $n < m$
    - H: measurement matrix, Φ: sparsity basis matrix
    - e.g. in MRI: H = partial Fourier, Φ = inverse wavelet
  - y_t: measurements (given)
  - x_t: sparsity basis vector
  - N_t: support set of x_t (set of indices of nonzero elements of x_t)
- Goal: recursively 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:
  - support set of x_t, N_t, changes slowly over time
  - also use slow signal value change where valid

Slow sparsity pattern change in medical image sequences [Qiu, Lu, Vaswani,ICASSP'09]

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



(a) slow support changes (adds)

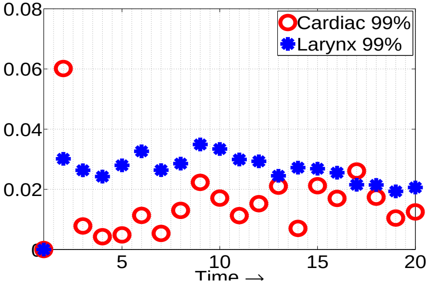


(b) slow support changes (removals)

- N_t: 99%-energy support set of x_t, where
- x_t: wavelet transform of cardiac or larynx image at time t
- Notice: all support changes are less than 2% of support size

Slow signal value change in medical seq's (common tracking assumption) [Lu,Vaswani,ArXiv]

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



- Plot of $\frac{\|(x_t - x_{t-1})\|_2}{\|(x_t)\|_2}$ against time, t
- x_t: wavelet transform of cardiac or larynx image at time t
- Notice: almost all changes are less than 4%

Questions we answer

1. How to solve RecSparsRec *while not increasing reconstruction algorithm speed or memory requirement w.r.t. simple CS?*

Questions we answer

1. How to solve RecSparsRec *while not increasing reconstruction algorithm speed or memory requirement w.r.t. simple CS?*
2. When does it achieve exact recovery?
3. Is it provably stable over time and under what conditions?
   - (critical question for a recursive approach)
   - are the conditions required weaker than those for simple CS?
4. How much better do our algorithms do compared to existing work for real experimental data?

Questions we answer

1. How to solve RecSparsRec *while not increasing reconstruction algorithm speed or memory requirement w.r.t. simple CS?*
2. When does it achieve exact recovery?
3. Is it provably stable over time and under what conditions?
   - (critical question for a recursive approach)
   - are the conditions required weaker than those for simple CS?
4. How much better do our algorithms do compared to existing work for real experimental data?
5. RecSparsRec in large but correlated noise

Related Work

- Simple CS (CS done at each time separately)
- CS-diff (CS on difference meas's) [Cevher et al,ECCV'08]: works only if
  - first frame reconstructed very accurately, and
  - difference signal sparser or signal values change very slowly
- Kalman Filtered CS (KF-CS) & LS-CS [Vaswani,ICIP'08,T-SP'10]
  - defined RecSparsRec problem; proposed an efficient solution
- Modified-CS [Vaswani,Lu, ISIT'09]: this talk

Related Work

- Simple CS (CS done at each time separately)
- CS-diff (CS on difference meas's) [Cevher et al,ECCV'08]: works only if
  - first frame reconstructed very accurately, and
  - difference signal sparser or signal values change very slowly
- Kalman Filtered CS (KF-CS) & LS-CS [Vaswani,ICIP'08,T-SP'10]
  - defined RecSparsRec problem; proposed an efficient solution
- Modified-CS [Vaswani,Lu, ISIT'09]: this talk
- Work with different goals than ours
  - homotopy methods: speed up optimization but not reduce n [Asif,Romberg'08,09]
  - recover *one* signal recursively as more meas's come in [Sanghavi et al,'08], [Angelosante et al'09], [Asif,Romberg'09], [Ghaoui et al'09]
  - batch methods: much slower, need a lot more memory [Wakin et al'06(video)],[Gamper et al'08 (MRI)], [Angelosante et al'09 (dyn Lasso)]

Least Squares CS and Kalman Filtered CS [Vaswani,ICIP'08][2], [Vaswani,IEEE Trans. SP,Aug'10][3]

At each time t,

- Let $T = \hat{N}_{t-1}$ be previous support estimate
- Compute LS (or KF) estimate assuming T is current support
  - LS estimate: $(\mu)_T = {A_T}^\dagger y_t$, $(\mu)_{T^c} = 0$
- CS on Residual
  - CS-residual: $\hat{\beta} = \arg\min \|\beta\|_1$ s.t. $\|(y_t - A\mu) - A\beta\|_2 \leq \epsilon$
  - Compute $\hat{x}_t = \hat{\beta} + \mu$
- Estimate support $\hat{N}_t = \{i : |(\hat{x}_t)_i| > \alpha\}$
- Final LS (or KF) using $\hat{N}_t$

[2] N. Vaswani, Kalman Filtered Compressed Sensing, ICIP, 2008

[3] N. Vaswani, "LS-CS-residual (LS-CS): Compressive Sensing on the Least Squares Residual", IEEE Trans. Sig. Proc., Aug. 2010.

Least Squares CS and Kalman Filtered CS [Vaswani,ICIP'08][4], [Vaswani,IEEE Trans. SP,Aug'10][5]

- Have same complexity and memory requirement as simple-CS
  - but accurate recovery with much fewer noisy measurements
- Proved LS-CS error "stability" (time-invariant error bound) under mild assumptions [Vaswani,TSP,Aug'10]:
- BUT: could not achieve exact recovery with fewer measurements

[4] N. Vaswani, Kalman Filtered Compressed Sensing, ICIP, 2008

[5] N. Vaswani, "LS-CS-residual (LS-CS): Compressive Sensing on the Least Squares Residual", IEEE Trans. Sig. Proc., Aug. 2010.

Least Squares CS and Kalman Filtered CS [Vaswani,ICIP'08][4], [Vaswani,IEEE Trans. SP,Aug'10][5]

- Have same complexity and memory requirement as simple-CS
  - but accurate recovery with much fewer noisy measurements
- Proved LS-CS error "stability" (time-invariant error bound) under mild assumptions [Vaswani,TSP,Aug'10]:
  1. support changes every-so-often and delay b/w support change times is large enough;
  2. support change size, S_a, and support size, S_0, small enough (for a given A);
  3. newly added elements' either added at a large-enough value or their value increases at least at a certain rate, r
- BUT: could not achieve exact recovery with fewer measurements

[4] N. Vaswani, Kalman Filtered Compressed Sensing, ICIP, 2008

[5] N. Vaswani, "LS-CS-residual (LS-CS): Compressive Sensing on the Least Squares Residual", IEEE Trans. Sig. Proc., Aug. 2010.

CS with partially known support [Vaswani,Lu, ISIT'09, T-SP, Sept'10][6]

- Reconstruct a sparse signal, x, with support, N, from $y := Ax$
  - given partial and possibly erroneous support knowledge: T

[6] N. Vaswani and W. Lu, "Modified-CS: Modifying Compressive Sensing for Problems with Partially Known Support", IEEE Trans. Sig. Proc., Sept. 2010. (shorter version in ISIT'09)

CS with partially known support [Vaswani,Lu, ISIT'09, T-SP, Sept'10][6]

- Reconstruct a sparse signal, x, with support, N, from $y := Ax$
  - given partial and possibly erroneous support knowledge: T
- Rewrite the true support, N, as

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

  - T: erroneous support estimate (use $T = \hat{N}_{t-1}$ at time t)
  - $\Delta := N \setminus T$: errors (misses) in T – unknown
  - $\Delta_e := T \setminus N$: errors (extras) in T – unknown

[6] N. Vaswani and W. Lu, "Modified-CS: Modifying Compressive Sensing for Problems with Partially Known Support", IEEE Trans. Sig. Proc., Sept. 2010. (shorter version in ISIT'09)

Modified-CS idea

- If Δ_e empty: above ⇔ find signal that is sparsest outside T

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

  - the unknowns are Δ, $(\beta)_\Delta$ **and** $(\beta)_T$

Modified-CS idea

- If Δ_e empty: above ⇔ find signal that is sparsest outside T

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

  - the unknowns are Δ, $(\beta)_\Delta$ **and** $(\beta)_T$

- Same solution also works if Δ_e is not empty but small

Modified-CS idea

- If Δ_e empty: above ⇔ find signal that is sparsest outside T

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

  - the unknowns are Δ, $(\beta)_\Delta$ **and** $(\beta)_T$

- Same solution also works if Δ_e is not empty but small

- Exact recovery: if every set of $(|T| + 2|\Delta|) = (|N| + |\Delta_e| + |\Delta|)$ columns of A are linearly independent

- Compare: ℓ_0-CS needs this to hold for every set of $2|N|$ columns

- Under slow support change, $|\Delta| \ll |N|$ and $|\Delta_e| \ll |N|$

Modified-CS [Vaswani,Lu, ISIT'09, T-SP,Sept'10][7]

- Modified-CS

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

  - we obtained exact recon cond's for Modified-CS; argued they are weaker than CS

- Other related parallel/later work:
  - [vonBorries et al, TSP'09]: no exact recon conditions or expts
  - [Khajenejad et al, ISIT'09]: probab. prior on support, studies exact recon
  - Later: [Jacques, Elsev.Sig.Proc'10]: error bounds for noisy mod-CS

[7] N. Vaswani and W. Lu, "Modified-CS: Modifying Compressive Sensing for Problems with Partially Known Support", IEEE Trans. Sig. Proc., Sept. 2010. (shorter version in ISIT'09)

Exact reconstruction result [Vaswani,Lu, ISIT'09, T-SP,Sept.'10]

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

Theorem (simplified condition)

x is the unique minimizer of (modified-CS) if

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

Exact reconstruction result [Vaswani,Lu, ISIT'09, T-SP,Sept.'10]

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

Theorem (simplified condition)

x is the unique minimizer of (modified-CS) if

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

- *δ_S: RIP constant – smallest real number s.t. singular values of any S-column sub-matrix of A lie in $[\sqrt{1-\delta_S}, \sqrt{1+\delta_S}]$* [Candes,Tao,T-IT'05]
  - non-increasing function of n (# of measurements)

recall: $\Delta := N \setminus T$: misses in T, $\Delta_e := T \setminus N$: extras in T

Proof Outline [Vaswani,Lu, ISIT'09, T-SP,Sept.'10][8]

Use overall approach of [Candes,Tao,Decoding by LP,T-IT,Dec'05]

- Obtain conditions on the Lagrange multiplier, w, to ensure that x is a *unique* minimizer
- Find sufficient conditions under which such a w can be found
  - key lemma: create a w that satisfies most reqd conditions
  - apply this lemma recursively to get a final w that satisfies *all* reqd conditions.

[8] N. Vaswani and W. Lu, "Modified-CS: Modifying Compressive Sensing for Problems with Partially Known Support", IEEE Trans. Sig. Proc., Sept. 2010. (shorter version in ISIT'09)

Comparison with best sufficient cond's for CS

- CS gives exact reconstruction if [Candes'08, Candes-Tao'06]

$$\delta_{2|N|} < \sqrt{2} - 1 \quad \text{or} \quad \delta_{2|N|} + \delta_{3|N|} < 1$$

- Modified-CS gives exact reconstruction if

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

- If $|\Delta| = |\Delta_e| = 0.02|N|$ (typical in medical sequences),
  - **sufficient condition for CS** to achieve exact recovery:

$$\delta_{0.04|N|} < 0.004$$

  - **sufficient condition for Mod-CS** to achieve exact recovery:

$$\delta_{0.04|N|} < 0.008$$

- **Mod-CS sufficient condition is weaker (needs fewer meas's)**

recall: $\Delta := N \setminus T, \quad \Delta_e := T \setminus N$

Simulations: exact reconstruction probability

Simulation setup:

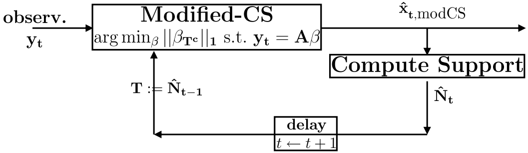
- signal length, $m = 256$, supp size, $|N| = 0.1m$
- supp error sizes, $|\Delta| = |\Delta_e| = 0.08|N|$
- used random-Gaussian A, varied n
- we say "works" (gives exact recon) if $\|x - \hat{x}\|_2 < 10^{-5}\|x\|_2$

Conclusions:

- With 19% measurements:
  - mod-CS "works" w.p. 99.8%, CS "works" w.p. 0
- With 25% measurements:
  - mod-CS "works" w.p. 100%, CS "works" w.p. 0.2%
- CS needs 40% measurements to "work" w.p. 98%

recall: Δ: errors (misses) in T, Δ_e: errors (extras) in T

Modified-CS for time sequences



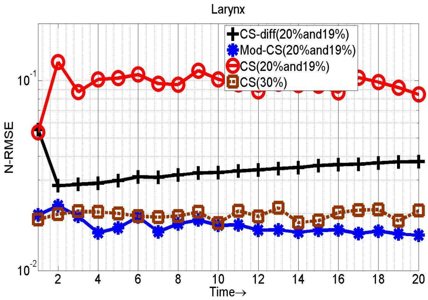
Support Estimation: use thresholding

$$\hat{N}_t := \{i : |(\hat{x}_{t,\text{modCS}})_i| > \alpha\}$$

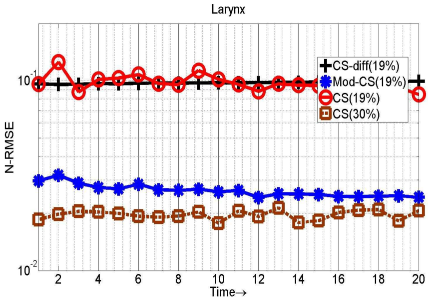
Initial time ($t = 0$):

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

Simulated MRI of an actual larynx (vocal tract) sequence: noise-free case



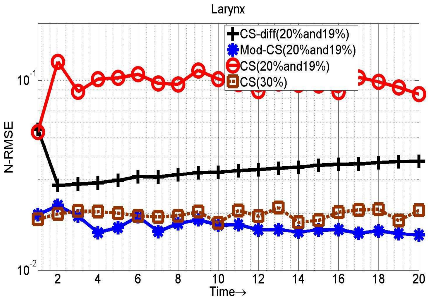
(c) $n_0 = 20\%$, $n = 19\%$



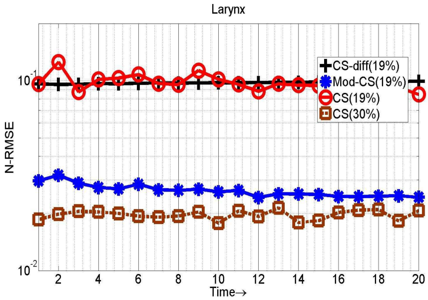
(d) $n_0 = 19\%$, $n = 19\%$

- A real image sequence: only compressible (approx sparse)
- With only $n = 19\%$ MRI meas's, Mod-CS error is small and stable at 2-3%, CS-diff error is unstable or large, simple-CS error is large

Simulated MRI of an actual larynx (vocal tract) sequence: noise-free case

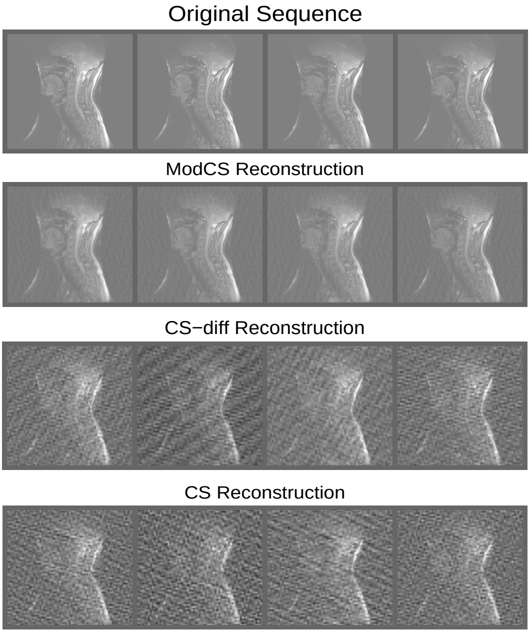


(e) $n_0 = 20\%$, $n = 19\%$



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

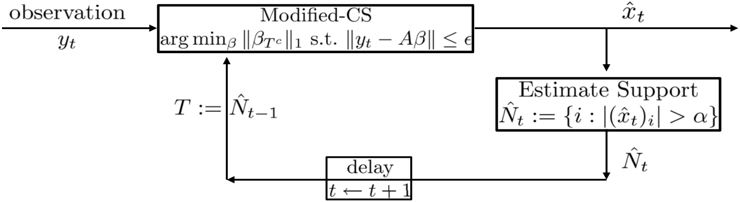
- A real image sequence: only compressible (approx sparse)
- With only $n = 19\%$ MRI meas's, Mod-CS error is small and stable at 2-3%, CS-diff error is unstable or large, simple-CS error is large
  - simple CS needs $n = 30\%$ to achieve small error



A larynx sequence (not sparsified)

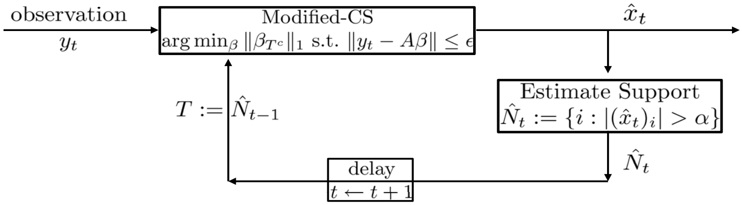
- 99%-support size $\sim$ 7%,
- supp change $\sim$ 2%
- using only 19% MRI measurements at all times
- simple CS needs $n = 30\%$ for same error
- http://www.ece.iastate.

Modified-CS for noisy measurements



- Difficulty:
  - along T^c: solution is biased towards zero
  - along T: no cost and only data constraint – solution can be biased away from zero

Modified-CS for noisy measurements



- Difficulty:
  - along T^c: solution is biased towards zero
  - along T: no cost and only data constraint – solution can be biased away from zero
  - the misses' set $\Delta_t \subset T^c$, while the extras' set, $\Delta_{e,t} \subset T$
    - need α small to add Δ_t, need α large to delete $\Delta_{e,t}$

(recall: $\Delta_t := N_t \setminus T = N_t \setminus \hat{N}_{t-1}$, $\Delta_{e,t} := T \setminus N_t = \hat{N}_{t-1} \setminus N_t$)

Possible Solutions

- Solution 1: improved support estimation (Add-LS-Del)
- Solution 2: use "slow signal value change" to constrain $(\beta)_T$

  [Lu, Vaswani, Trans.SP, Jan'12], [Raisali, Vaswani, CISS'11]

$$\arg\min_{\beta} \|(\beta)_{T^c}\|_1 \text{ s.t. } \|y_t - A\beta\|_2 \leq \epsilon, \quad \|\beta_T - \mu_T\|_2 \leq \gamma$$

  - with $\mu := \hat{x}_{t-1}$, $T := \hat{N}_{t-1}$ (Reg-Mod-CS – ongoing work)
  - useful if signal value change is "slow enough"

Modified-CS with Add-LS-Del (improved support support estimation)[9]

- Modified-CS: set $T = \hat{N}_{t-1}$ and compute $\hat{x}_t$ as the solution of

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

- Support Add using a small threshold
  - use α_{add} just large enough s.t. well-conditioned $(A)_{T_{\text{add}}}$
- Compute LS estimate on T_{add}, call it $\hat{x}_{t,\text{add}}$
  - reduces bias and error if $T_{\text{add}} \approx N_t$ [Candes,Tao'06]
- Support Delete by thresholding on $\hat{x}_{t,\text{add}}$ w/ a larger threshold
  - $\hat{x}_{t,\text{add}}$ more accurate $\Rightarrow \alpha_{\text{del}}$ can be larger

[9] introduced in [Vaswani,ICIP'08,T-SP'10] & also in [Dai,Milenkovic'09], [Needell,Tropp'09] for static case

Modified-CS with Add-LS-Del (improved support support estimation)[9]

- Modified-CS: set $T = \hat{N}_{t-1}$ and compute $\hat{x}_t$ as the solution of

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

- Support Add using a small threshold
  - use α_{add} just large enough s.t. well-conditioned $(A)_{T_{\text{add}}}$
  - $T_{\text{add}} = T \cup \{i : |(\hat{x}_t)_i| > \alpha_{\text{add}}\}$
- Compute LS estimate on T_{add}, call it $\hat{x}_{t,\text{add}}$
  - reduces bias and error if $T_{\text{add}} \approx N_t$ [Candes,Tao'06]
- Support Delete by thresholding on $\hat{x}_{t,\text{add}}$ w/ a larger threshold
  - $\hat{x}_{t,\text{add}}$ more accurate $\Rightarrow$ α_{del} can be larger
  - $\hat{N}_t = T_{\text{add}} \setminus \{i : |(\hat{x}_{t,\text{add}})_i| \leq \alpha_{\text{del}}\}$

[9] introduced in [Vaswani,ICIP'08,T-SP'10] & also in [Dai,Milenkovic'09], [Needell,Tropp'09] for static case

Stability over time [Vaswani,T-SP, Aug'10] [10], [Vaswani,Allerton'10][11]

- Easy to bound the reconstruction error at a given time, t
  - the result depends on the support errors $|\Delta_t|$, $|\Delta_{e,t}|$
    (recall: $\Delta_t := N_t \setminus \hat{N}_{t-1}$, $\Delta_{e,t} := \hat{N}_{t-1} \setminus N_t$)
- **Key question for a recursive algorithm: when can we get a time-invariant and small bound on the error?**

[10] N. Vaswani, "LS-CS-residual (LS-CS): Compressive Sensing on the Least Squares Residual", IEEE Trans. Sig. Proc., Aug. 2010

[11] N. Vaswani, Stability (over time) of Modified-CS for Recursive Causal Sparse Reconstruction, Allerton 2010

Stability over time [Vaswani,T-SP, Aug'10] [10], [Vaswani,Allerton'10][11]

- Easy to bound the reconstruction error at a given time, t
  - the result depends on the support errors $|\Delta_t|$, $|\Delta_{e,t}|$
    (recall: $\Delta_t := N_t \setminus \hat{N}_{t-1}$, $\Delta_{e,t} := \hat{N}_{t-1} \setminus N_t$)
- **Key question for a recursive algorithm: when can we get a time-invariant and small bound on the error?**
- Solution approach: first obtain conditions under which time-invariant bounds on $|\Delta_t|$, $|\Delta_{e,t}|$ hold
  - direct corollary: time-invariant bound on the recon error

[10] N. Vaswani, "LS-CS-residual (LS-CS): Compressive Sensing on the Least Squares Residual", IEEE Trans. Sig. Proc., Aug. 2010

[11] N. Vaswani, Stability (over time) of Modified-CS for Recursive Causal Sparse Reconstruction, Allerton 2010

Signal change model, measurement model and our result

Signal change model:

- S_a additions and removals from the support **at each time**
- support size constant at S_0
- new elements added at a small value, r; magnitude increases at rate r per unit time, until it reaches a maximum magnitude dr
  - similarly for decrease before removal

Measurement model:

$$y_t = Ax_t + w_t, \quad \|w_t\|_2 \leq \epsilon$$

Signal change model, measurement model and our result

Signal change model:

- S_a additions and removals from the support **at each time**
- support size constant at S_0
- new elements added at a small value, r; magnitude increases at rate r per unit time, until it reaches a maximum magnitude dr
  - similarly for decrease before removal

Measurement model:

$$y_t = Ax_t + w_t, \quad \|w_t\|_2 \leq \epsilon$$

Our result: "stability" holds if

1. S_a and S_0 are small enough (for a given A),
   - ensures the error bound holds at all times
2. r is large enough
   - ensures newly added elements detected within a finite delay

Theorem (Modified-CS stability [Vaswani, Allerton'10])

If

1. *support estimation threshold,* $\alpha = 8.79\epsilon$
2. *support size, support change size* S_0, S_a *satisfy*
   - $\delta_{S_0+3S_a} < (\sqrt{2}-1)/2$ *(for a given A)*
3. *new element initial value and increase rate,* $r \geq 8.79\epsilon$,
4. *at initial time,* $t = 0$, n_0 *large enough s.t.* $\delta_{2S_0} < (\sqrt{2}-1)/2$

then, at all times, t,

- *final support errors,* $|\tilde{\Delta}_t| \leq 2S_a$ *and* $|\tilde{\Delta}_{e,t}| = 0$
- *initial support errors,* $|\Delta_t| \leq 2S_a$ *and* $|\Delta_{e,t}| \leq S_a$
- *and so recon error satisfies* $\|x_t - \hat{x}_{t,modcs}\|_2 \leq 8.79\epsilon$

recall: $\Delta_t := N_t \setminus \hat{N}_{t-1}$, $\Delta_{e,t} := \hat{N}_{t-1} \setminus N_t$, $\tilde{\Delta}_t := N_t \setminus \hat{N}_t$, $\tilde{\Delta}_{e,t} := \hat{N}_t \setminus N_t$

Proof Outline: use induction

Here, "bounded" ⇔ bounded by a time-invariant value

- Induction assumption:
  - final support errors (misses and extras) at $t-1$ bounded
- $+$ signal model $\Rightarrow$ predicted support errors at t bounded
- $+$ n large enough (or S_0 small enough) $\Rightarrow$ Mod-CS error bounded
- $+$ α large enough $\Rightarrow$ no extras
- $+$ r large enough $\Rightarrow$ all elements with mag $> 2r$ detected (bounded misses)
- $\Rightarrow$ final support errors (misses and extras) at t bounded

Theorem (Modified-CS-with-Add-LS-Del stability [Vaswani,Allerton'10])

Let $e := (x - \hat{x}_{add})_{T_{add}}$. *If*

$$\|e\|_\infty \leq (1/\sqrt{S_a})\ \|e\|_2,$$

1. *(addition and deletion thresholds)*
    - α_{add} *is large enough s.t. at most* S_a *false adds per unit time,*
    - $\alpha_{del} = \sqrt{\frac{2}{S_a}}\epsilon + 2\theta_{S_0+2S_a,S_a} r$,
2. *(support size, support change size)* S_0, S_a *satisfy*
    - $\delta_{S_0+3S_a} < (\sqrt{2}-1)/2$ *and* $\theta_{S_0+2S_a,S_a} < \frac{1}{4}$ *(for a given A),*
3. *(new coeff. increase rate)* $r \geq \max(G_1, G_2)$, *where*

$$G_1 := \frac{\alpha_{add} + 8.79\epsilon}{2},\ \ G_2 := \frac{\sqrt{2}\epsilon}{\sqrt{S_a}(1 - 2\theta_{S_0+2S_a,S_a})}$$

then, all the same conclusions hold.

Proof Outline – 1 [Vaswani,Allerton'10] [12]

- Goal: ensure that within a finite delay d_0, all newly added elements get detected and all zeroed (removed) elements get deleted
  - simpler case: fix $d_0 = 2$
- Starting point
  - conditions and bound for Modified-CS error at t
    - simple modification of Candes' approach for CS
  - conditions and bound for LS step error at t – also easy
- Key lemmas: sufficient conditions to ensure that, at a given t,
  1. an undetected large-enough element *gets added*
  2. an existing large-enough element *does not get falsely deleted*
  3. a falsely detected zero element *does get deleted*

[12] N. Vaswani, "Stability (over time) of Modified-CS for Recursive Causal Sparse Reconstruction", Allerton 2010, submitted to IEEE Trans. Info. Th.

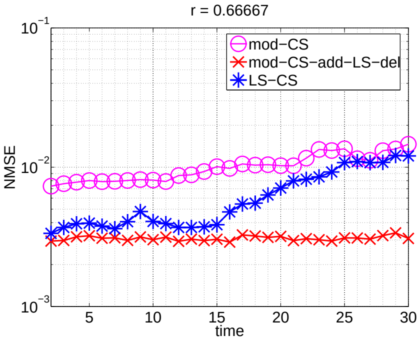
Proof Outline – 2: Induction step idea

- Assume $|\tilde{\Delta}_{t-1}| \le 2S_a$, $|\tilde{\Delta}_{e,t-1}| = 0$ (induction assumption)
- Above + signal model $\Rightarrow |\Delta_t|, |\Delta_{e,t}|$ bounded
  (recall: $\Delta_t := N_t \setminus \hat{N}_{t-1}$, $\Delta_{e,t} := \hat{N}_{t-1} \setminus N_t$, $\tilde{\Delta}_t := N_t \setminus \hat{N}_t$, $\tilde{\Delta}_{e,t} := \hat{N}_t \setminus N_t$)
- Above + S_0, S_a small enough $\Rightarrow$ Mod-CS error bounded at t
- Add step
  - above + signal model + r large enough $\Rightarrow$ elements with mag. $\ge 2r$ definitely get detected,
  - need α_{add} large enough s.t. few and bounded false adds
  - above two ensure support errors bounded after the add step
- LS and Delete step
  - above + S_0, S_a small enough $\Rightarrow$ LS step error bounded
  - above + signal model + r large enough $\Rightarrow$ only elements $< 2r$ may get falsely deleted ($|\tilde{\Delta}_t| \le 2S_a$)
  - above + α_{del} large enough $\Rightarrow$ all extras deleted ($|\tilde{\Delta}_{e,t}| = 0$)

Discussion

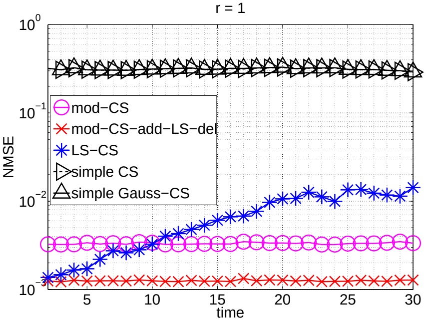
- Slow supp change $\Rightarrow S_a \ll S_0 \Rightarrow$ supp error bound, $2S_a$, is small compared to the supp size, S_0 (meaningful result)
- Modified-CS stability result – only needs $\delta_{S_0+2S_a} < (\sqrt{2}-1)/2$
  - needs weaker conditions on A than simple CS
    - Simple CS needs $\delta_{2S_0} < (\sqrt{2}-1)/2$ (for same error bound)
- Modified-CS-Add-LS-del stability result – needs
  - weaker conditions on A than CS (for same error bound)
  - weaker conditions on r than modified-CS
    - it needs $r \geq (\alpha_{\text{add}} + 8.79\epsilon)/2$ but modified-CS needs $r \geq 8.79\epsilon$
  - weaker conditions on both A and r compared to LS-CS result

Normalized mean squared error (NMSE) v/s time

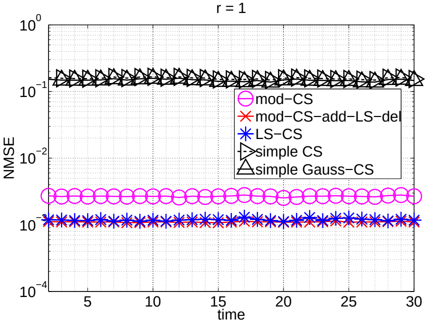


- *A*: random-Gaussian, $n \times m$, $n = 29.5\%$; noise: unif(-0.13,0.13);
- new elem's added at mag. $r = 0.67$; incr. at rate r, until reach $M = 2$
- $m = 200$, support size, $S_0 = 0.1m$, support change size, $S_a = 0.1S_0$
- ModCS-Add-LS-del stable, others are not

Normalized mean squared error (NMSE) v/s time



$r = 1, \ n = 29.5\%$



$r = 1, \ n = 32.5\%$

- $m = 200, \ S_0 = 0.1m, \ S_a = 0.1S_0, \ d = 3.$
- ModCS needs larger r, LS-CS needs larger r and larger n
- Simple-CS has large error even with $n = 32.5\%$

Video: Background subtraction

image sequence, $M_t = L_t + S_t$

background sequence, L_t: low rank, changing subspace

foreground sequence, F_t: sparse w/ correlated support changes

$N_t = \text{support}(F_t)$, $(S_t)_{N_t} = (F_t - L_t)_{N_t}$, $(S_t)_{N_t^c} = 0$

The Problem

- Measurement: $M_t := L_t + S_t$
  - S_t: sparse vector, with correlated support change over time
  - L_t: low dimensional vector (matrix $L := [L_{t-\tau}, \dots L_t]$ is low rank)
    - subspace in which L_t lies changes gradually over time
    - matrix P_t: its columns span the subspace in which L_t lies
- Given P_0, recursively recover S_t, L_t and the matrix P_t

The Problem

- Measurement: $M_t := L_t + S_t$
  - S_t: sparse vector, with correlated support change over time
  - L_t: low dimensional vector (matrix $L := [L_{t-\tau}, \dots L_t]$ is low rank)
    - subspace in which L_t lies changes gradually over time
    - matrix P_t: its columns span the subspace in which L_t lies
- Given P_0, recursively recover S_t, L_t and the matrix P_t
- Recursive Robust PCA:
  - S_t: corruption (outlier), L_t: signal, P_t: its PC matrix
- RecSparsRec in Large but Low-dimensional Noise:
  - S_t: sparse signal, L_t: corruption (low dimensional noise)
  - our solutions apply even if $M_t = \Psi S_t + L_t$, Ψ: fat matrix

Motivation and Applications

- Existing work [Candes,Wright,Ma,Li], [DeLaTorre,Black], ...
  - simple thresholding (recovers only S_t)
  - detect outliers and either downweight them, e.g. RSL, or fill in using heuristics
  - PCP - recover L, S from $M = L + S$ (S: sparse but not low rank, L: low rank but not sparse)
- Need an approach that can
  - handle correlated S_t's (PCP cannot)
  - can handle fairly large support-sized S_t's (RSL, PCP cannot)
  - recover small magnitude S_t's (RSL, thresh cannot)
  - work in real-time
- Applications: recover sparse signals (most natural signals) in large but spatially correlated noise (most natural noise sources)
  - video/audio denoising, fMRI based active region detection, sensor nets, ...

ReProCS: Recursive Projected CS [Qiu,Vaswani,Allerton'10, Allerton'11][13]

$$M_t = S_t + L_t, \quad L_t = P_t a_t$$

- Update $\hat{P}_t$ every-so-often: recursive PCA
- Project M_t into space perp to $\hat{P}_t$: get y_t
- Recover S_t from y_t: noisy sparse recovery
- Compute $\hat{L}_t := M_t - \hat{S}_t$

[13] C. Qiu and N. Vaswani, Real-time Robust Principal Components' Pursuit, Allerton, 2010
C. Qiu and N. Vaswani, Recursive Sparse Recovery in Large but Correlated Noise, Allerton 2011

ReProCS: Recursive Projected CS [Qiu,Vaswani,Allerton'10, Allerton'11][13]

$$M_t = S_t + L_t, \quad L_t = P_t a_t$$

- Update $\hat{P}_t$ every-so-often: recursive PCA
  - $\hat{P}_t = \text{recursive-PCA}(\hat{P}_{t-1}, [\hat{L}_{t-\tau}, \dots \hat{L}_{t-1}])$
- Project M_t into space perp to $\hat{P}_t$: get y_t
  - $y_t := (\hat{P}_{t,\perp})' M_t = (\hat{P}_{t,\perp})' S_t + \beta_t, \quad \beta_t$: small noise
- Recover S_t from y_t: noisy sparse recovery
  - $\hat{S}_t = \arg\min_b \|b\|_1 \text{ s.t. } \|y_t - (\hat{P}_{t,\perp})' b\|_2 \le \epsilon$
- Compute $\hat{L}_t := M_t - \hat{S}_t$

[13] C. Qiu and N. Vaswani, Real-time Robust Principal Components' Pursuit, Allerton, 2010
C. Qiu and N. Vaswani, Recursive Sparse Recovery in Large but Correlated Noise, Allerton 2011

Support-predicted Modified-CS in ReProCS [Qiu,Vaswani,ISIT'11][14]

- If $r := rank(\hat{P}_t)$ small enough for a given $s := |\text{support}(S_t)|$
  - $\frac{s}{n-r}$ large enough for CS to work

[14] C. Qiu and N. Vaswani, Support-Predicted Modified-CS for Recursive Robust Principal Components' Pursuit, ISIT, 2011

Support-predicted Modified-CS in ReProCS [Qiu,Vaswani,ISIT'11][14]

- If $r := rank(\hat{P}_t)$ small enough for a given $s := |\text{support}(S_t)|$
  - $\frac{s}{n-r}$ large enough for CS to work
- If s too large or r too large: need Modified-CS

[14] C. Qiu and N. Vaswani, Support-Predicted Modified-CS for Recursive Robust Principal Components' Pursuit, ISIT, 2011

Support-predicted Modified-CS in ReProCS [Qiu,Vaswani,ISIT'11][14]

- If $r := rank(\hat{P}_t)$ small enough for a given $s := |\text{support}(S_t)|$
  - $\frac{s}{n-r}$ large enough for CS to work
- If s too large or r too large: need Modified-CS
- Video: support changes over time much more
  - e.g. 10x10 block: one pixel motion – supp change of 10
  - $T = \hat{N}_{t-1}$ is not a good approx to N_t
- Support-predicted Modified-CS idea:
  - use $T = \text{model-predict}(\hat{N}_{t-1})$ in Mod-CS
  - use $\hat{N}_t$ to update correlation model parameters

[14] C. Qiu and N. Vaswani, Support-Predicted Modified-CS for Recursive Robust Principal Components' Pursuit, ISIT, 2011

Experiments

- Chenlu Qiu's webpage
- ReProCS magic (S_t invisible in video, its support large, is correlated)
- ReProCS overlay (real bgnd, foregnd somewhat visible but overlay)
- ReProCS (modCS) overlay (very large support of S_t: ReProCS fails)

Acknowledgements

- This talk is mostly based on joint work with my Ph.D. students Wei Lu and Chenlu Qiu
- Research support: NSF grants
  - CCF-1117125 (Recursive Robust PCA)
  - CCF-0917015 (Recursive Reconstruction of Sparse Signal Sequences)
  - ECCS-0725849 (Change Detection in Nonlinear Systems and Applications in Shape Analysis)
- The fMRI work is in collaboration with Dr. Ian Atkinson at UIC Center for MR Research

Ongoing and Future Work

- RecSparsRec in Large but Correlated Noise – Rec Robust PCA
- Regularized ModCS and Kalman filtered ModCS (KalMoCS)
  - open q – when is KalMoCS stable w.r.t. a genie-aided KF?

Ongoing and Future Work

- RecSparsRec in Large but Correlated Noise – Rec Robust PCA
- Regularized ModCS and Kalman filtered ModCS (KalMoCS)
  - open q – when is KalMoCS stable w.r.t. a genie-aided KF?
- Functional MRI (fMRI) applications [Lu, Li, Atkinson, Vaswani, ICIP'11]
- Computer Vision
  - ReProCS and applications in video [Qiu, Vaswani, Allerton'10, ISIT'11]
  - Large dimensional visual tracking – use ideas from RecSparsRec