Online Structured Signals' Recovery and Applications in Bio-Imaging

Namrata Vaswani

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

(joint work with Wei Lu, Chenlu Qiu, Brian Lois, Ian Atkinson, Leslie Hogben)

June 5, 2014

Talk Outline

Background

Recursive/Online Recovery of Sparse Signal Sequences

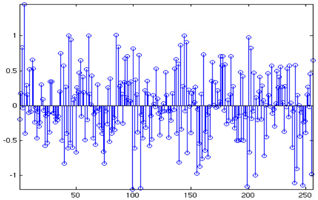
- The Problem
- Modified-CS and exact recovery result
- Noisy Modified-CS and error stability (over time)

Online Sparse + Low-rank Matrix Recovery

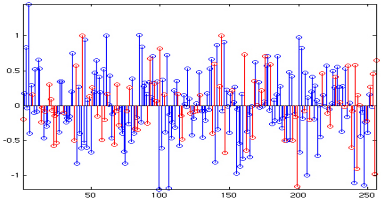
- The Problem
- Proposed online Robust PCA solution: ReProCS
- Performance guarantees
- Proof Outline

Structured Signal Recovery: The question

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



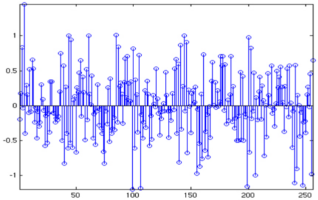
(a) the unknown signal



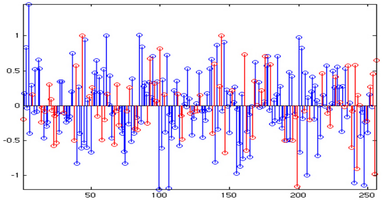
(b) its 80 time samples (red)

Structured Signal Recovery: The question

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



(c) the unknown signal

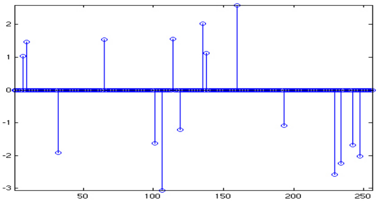


(d) its 80 time samples (red)

- If the signal has some structure: YES, e.g.,
  - if it is bandlimited – use a low-pass filter
  - if it is a weighted sum of only a few sinusoids – use sparsity

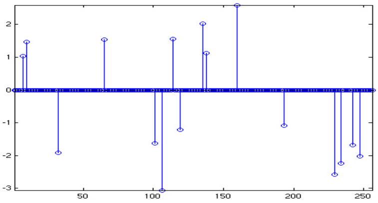
Example taken from L1-Magic webpage of Candes,Romberg,Tao

- This signal is Fourier sparse

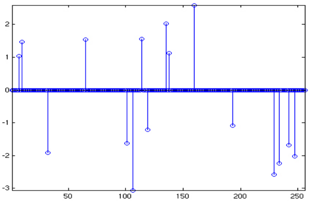


(e) DFT of original signal

- This signal is Fourier sparse



(g) DFT of original signal

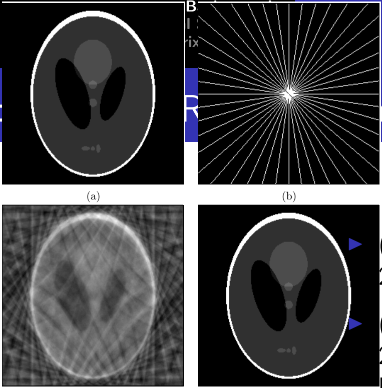


(h) recovered DFT: exact!

- Use its 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

Example 2: … R … maging (MRI)



(a) (b) (c) (d)

- (a) Shepp-Logan phantom: 256 $\times$ 256 image
- (b) MR imaging pattern: 256-point DFT along 22 radial lines
- (c) Inverse-DFT (min ℓ_2 norm soln)
- (d) Basis Pursuit soln (min ℓ_1 norm soln)

Example taken from [Candes,Romberg,Tao,T-IT, Feb 2006]

Sparse Recovery [Mallat et al'93], [Feng,Bresler'96], [Gordinsky,Rao'97], [Chen,Donoho'98]

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

Sparse Recovery [Mallat et al'93], [Feng,Bresler'96], [Gordinsky,Rao'97], [Chen,Donoho'98]

- Reconstruct a sparse vector x, with support size s, 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.

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

Sparse Recovery [Mallat et al'93], [Feng,Bresler'96], [Gordinsky,Rao'97], [Chen,Donoho'98]

- Reconstruct a sparse vector x, with support size s, 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 set of $2s$ columns of A are linearly independent

Sparse Recovery [Mallat et al'93], [Feng,Bresler'96], [Gordinsky,Rao'97], [Chen,Donoho'98]

- Reconstruct a sparse vector x, with support size s, 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.

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

  - and any set of $2s$ columns of A are linearly independent
  - but exponential complexity – $O(m^s)$
- Practical (polynomial complexity) approaches
  - convex relaxation approaches
    - ℓ_1 minimization: replace ℓ_0 norm by ℓ_1 norm
  - greedy methods [Mallat,Zhang'93], [Pati et al'93], [Dai,Milenkovic'09], [Needell,Tropp'09]
  - many more ...

Compressive Sensing (CS) [Feng,Bresler'96], [Gordinsky,Rao'97], [Candes,Romberg,Tao'05], [Donoho'05]

- Compressive Sense (CS):
  - since most images are (approx) sparse, just "sense" less
    - e.g., medical images are often wavelet sparse
  - recover image from measurements using sparse recovery
- Applications: projection imaging, e.g., MRI, CT, astronomy,

Compressive Sensing (CS) [Feng,Bresler'96], [Gordinsky,Rao'97], [Candes,Romberg,Tao'05], [Donoho'05]

- Compressive Sense (CS):
  - since most images are (approx) sparse, just "sense" less
    - e.g., medical images are often wavelet sparse
  - recover image from measurements using sparse recovery
- Applications: projection imaging, e.g., MRI, CT, astronomy,
- CS literature
  - much stronger performance guarantees for ℓ_1 minimization than earlier work
  - introduced a weaker notion of "incoherence" b/w measurement and sparsity basis

Compressive Sensing (CS) [Feng,Bresler'96], [Gordinsky,Rao'97], [Candes,Romberg,Tao'05], [Donoho'05]

- Compressive Sense (CS):
  - since most images are (approx) sparse, just "sense" less
    - e.g., medical images are often wavelet sparse
  - recover image from measurements using sparse recovery
- Applications: projection imaging, e.g., MRI, CT, astronomy,
- CS literature
  - much stronger performance guarantees for ℓ_1 minimization than earlier work
  - introduced a weaker notion of "incoherence" b/w measurement and sparsity basis
- Sparse Recovery $\Leftrightarrow$ CS $\Leftrightarrow$ ℓ_1 minimization

Structured Signals' Recovery

- Sparse recovery is one example
- Other examples
  - Block sparse signals' recovery
  - Low-rank matrix completion: recover a low-rank matrix from a subset of its entries [Recht et al,2009],...
  - Sparse matrix plus low-rank matrix recovery / robust PCA: recover S, L from $M := S + L$ or from undersampled measurements [Candes et al,2011,Chandrasekharan et al,2011],...
- Applications: MRI, Netflix problem, foreground-background separation in video or functional MRI, ...

Our Work: The question

- How to use the above ideas for dynamic medical imaging?
  - e.g., dynamic MRI, functional MRI, dynamic CT
- Option 1: batch methods
  - treat the entire sequence as one spatio-temporal structured signal that is recovered jointly
  - need few measurements, but slow and memory-intensive
- Option 2: simple CS
  - recover each image in the sequence independently
  - fast and memory-efficient, but will need more measurements

Our Work: The question

- How to use the above ideas for dynamic medical imaging?
  - e.g., dynamic MRI, functional MRI, dynamic CT
- Option 1: batch methods
  - treat the entire sequence as one spatio-temporal structured signal that is recovered jointly
  - need few measurements, but slow and memory-intensive
- Option 2: simple CS
  - recover each image in the sequence independently
  - fast and memory-efficient, but will need more measurements
- Option 3: design recursive algorithms (our work)
  - use previous recovered signal(s) and current measurements' vector to recover current signal

Problem 1: Recursive Recovery of Sparse Signal Sequences [Vaswani,ICIP'08][1]

- Given measurements

$$y_t := Ax_t + w_t, \quad \|w_t\|_2 \le \epsilon, \quad t = 0, 1, 2, \ldots$$

  - $A = H\Phi$ (given): $n \times m$, $n < m$
    - H: measurement matrix, Φ: sparsity basis matrix
    - e.g., in MRI: $H =$ partial Fourier, $\Phi =$ inverse wavelet
  - y_t: measurements (given)
  - x_t: sparsity basis vector
  - N_t: support set of x_t
  - w_t: noise ($\epsilon = 0$: noise-free, $\epsilon \ll \|x_t\|$: small noise)
- Goal: recursively reconstruct x_t from $y_0, y_1, \ldots y_t$,
  - i.e. use only y_t and $\hat{x}_{t-1}$ for recovering x_t

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

Key attributes we look for

1. measurements'-efficient: **always needed**
   - use less meas's than simple CS solutions for a given accuracy
   - *simple CS: recover each sparse signal separately at each time*
2. fast and memory-efficient: **always needed**
   - computational & memory complexity $\sim$ simple CS solutions
3. causal: needed for real-time applications
4. meaningful performance guarantees: desirable
   - provably exact recovery with fewer meas's: noise-free case
   - time-invariant error bounds: noisy case

Batch methods don't satisfy 2. and 3., sometimes not 1. either

Existing Work

- In 2008: almost none
- Only batch methods
  - [Wakin et al'06(video)],[Gamper et al'08 (MRI)]: exploit Fourier sparsity along time axis
  - multiple measurements' vectors (MMV) approaches: assume support does not change with time
- Limitations
  - not causal
  - slow and memory-intensive even for offline apps
  - above assumptions may not hold

Our solution approach [Vaswani,ICIP'08][2]

- Exploit practically valid assumptions to get fast and measurements'-efficient recursive algorithms
- Slow support change: (recall $N_t = \text{support}(x_t)$)

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

  - introduced in [Vaswani,ICIP'08] , verified in [Qiu, Lu, Vaswani,ICASSP'09]

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

Our solution approach [Vaswani,ICIP'08][2]

- Exploit practically valid assumptions to get fast and measurements'-efficient recursive algorithms
- Slow support change: (recall $N_t = \text{support}(x_t)$)

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

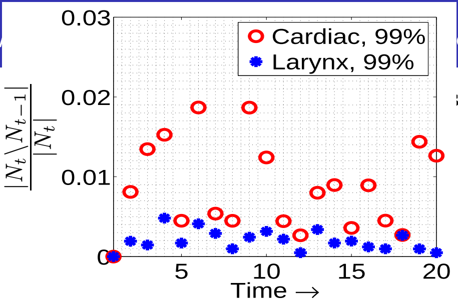
  - introduced in [Vaswani,ICIP'08] , verified in [Qiu, Lu, Vaswani,ICASSP'09]
- Slow signal value change (use when valid):

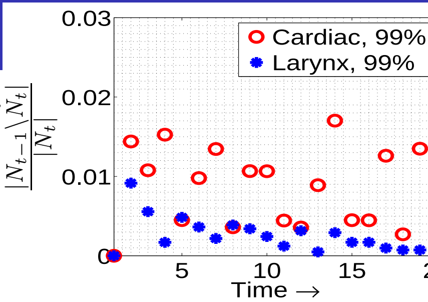
$$\|(x_t - x_{t-1})\|_2 \ll \|(x_t)\|_2$$

  - commonly used in all tracking algo's, adaptive filtering, etc

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

Slow





(a) slow support changes (adds)

(b) slow support changes (removals)

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

First recursive solutions and their limitation

- Kalman filtered CS (KF-CS) and Least Squares CS (LS-CS)
  [Vaswani,ICIP'08], [Vaswani, ICASSP'09, Trans-SP,Aug'10][3]
  - causal; fast and memory efficient; and measurements'-efficient for accurate recovery;
  - could get time-invariant error bounds under mild assumptions for LS-CS

[3] N. Vaswani, "Kalman Filtered Compressed Sensing", ICIP 2008
N. Vaswani, "LS-CS-residual (LS-CS): Compressive Sensing on the Least Squares Residual", IEEE Trans. Sig. Proc., Aug. 2010

First recursive solutions and their limitation

- Kalman filtered CS (KF-CS) and Least Squares CS (LS-CS)
  [Vaswani,ICIP'08], [Vaswani, ICASSP'09, Trans-SP,Aug'10][3]
  - causal; fast and memory efficient; and measurements'-efficient for accurate recovery;
  - could get time-invariant error bounds under mild assumptions for LS-CS
  - **But, neither was measurements'-efficient for exact recovery**

[3] N. Vaswani, "Kalman Filtered Compressed Sensing", ICIP 2008
N. Vaswani, "LS-CS-residual (LS-CS): Compressive Sensing on the Least Squares Residual", IEEE Trans. Sig. Proc., Aug. 2010

First recursive solutions and their limitation

- Kalman filtered CS (KF-CS) and Least Squares CS (LS-CS)
  [Vaswani,ICIP'08], [Vaswani, ICASSP'09, Trans-SP,Aug'10][3]
  - causal; fast and memory efficient; and measurements'-efficient for accurate recovery;
  - could get time-invariant error bounds under mild assumptions for LS-CS
  - **But, neither was measurements'-efficient for exact recovery**
- Other parallel, somewhat related work:
  - CS-diff [Cevher-et-al,ECCV'08]: meas-efficient only if difference signal sparser (not valid mostly); CS-time-varying [Angelosante et al,ICASSP'09]: not fast (batch and causal); homotopies for dynamic-ℓ_1 [Asif-Romberg'09]: not measurements-efficient

[3] N. Vaswani, "Kalman Filtered Compressed Sensing", ICIP 2008
N. Vaswani, "LS-CS-residual (LS-CS): Compressive Sensing on the Least Squares Residual", IEEE Trans. Sig. Proc., Aug. 2010

Sparse recovery with partially known support [Vaswani,Lu, ISIT'09, T-SP, Sept'10][4]

- To get measurements'-efficient exact recovery:
  - reformulate problem as sparse rec with partially known support

[4] 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)

Sparse recovery with partially known support [Vaswani,Lu, ISIT'09, T-SP, Sept'10][4]

- To get measurements'-efficient exact recovery:
  - reformulate problem as sparse rec with partially known support
- New problem: recover x with support, N, from $y := Ax$
  - given partial and possibly erroneous support knowledge: T

[4] 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)

Sparse recovery with partially known support [Vaswani,Lu, ISIT'09, T-SP, Sept'10][4]

- To get measurements'-efficient exact recovery:
  - reformulate problem as sparse rec with partially known support
- New problem: recover x with support, N, from $y := Ax$
  - given partial and possibly erroneous support knowledge: T
- Rewrite the true support of x, 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

[4] 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

- Given T, find x from $y := Ax$. True support, $N = T \cup \Delta \setminus \Delta_e$.
- If Δ_e empty: above $\Leftrightarrow$ 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

- Given T, find x from $y := Ax$. True support, $N = T \cup \Delta \setminus \Delta_e$.
- If Δ_e empty: above $\Leftrightarrow$ 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

- Given T, find x from $y := Ax$. True support, $N = T \cup \Delta \setminus \Delta_e$.
- If Δ_e empty: above $\Leftrightarrow$ 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 condition: every set of $(|N| + |\Delta_e| + |\Delta|)$ columns of A are linearly independent
  - Compare: simple-ℓ_0 needs this to hold for every set of $2|N|$ columns of A
  - Slow support change $\Rightarrow |\Delta| \ll |N|$ and $|\Delta_e| \ll |N|$: modified-ℓ_0 condition weaker

Modified-CS [Vaswani,Lu, ISIT'09, T-SP,Sept'10][5]

- Modified-CS

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

- Other related parallel work:
  - [Khajenejad et al, ISIT'09]: probab. prior on support, studies exact recon for weighted ℓ_1
  - [vonBorries et al, TSP'09]: no exact recon conditions or expts

[5] N. Vaswani and W. Lu, "Modified-CS: Modifying Compressive Sensing for Problems with Partially Known Support", IEEE Trans. Sig. Proc., Sept. 2010.

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]

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]

Proof Outline: motivated by [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 conditions; apply

Comparison

- CS (ℓ_1 min) gives exact recon if [Candes'08, Candes-Tao'06]

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

- 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)**

Simulations: exact reconstruction probability

Simulation setup:

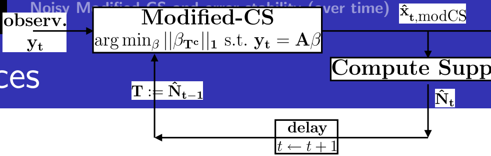
- signal length, $m = 256$, supp size, $|N| = 0.1m$
- supp error sizes, $|\Delta| = |\Delta_e| = 0.08|N|$
- used random-Gaussian A, varied n
- we say "works" (gives exact recon) if $\|x - \hat{x}\|_2 < 10^{-5}\|x\|_2$

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



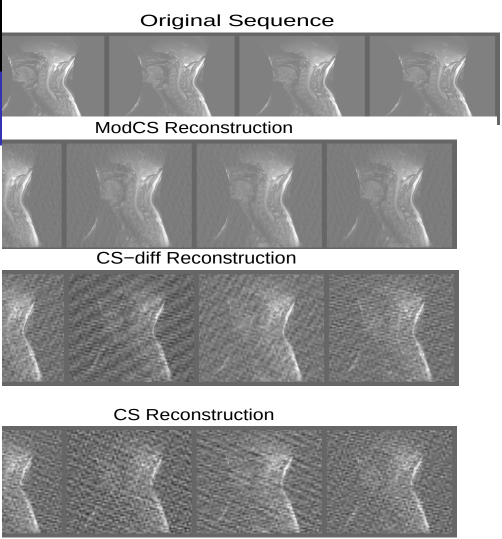
Support Estimation: use thresholding

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

Initial time ($t = 0$):

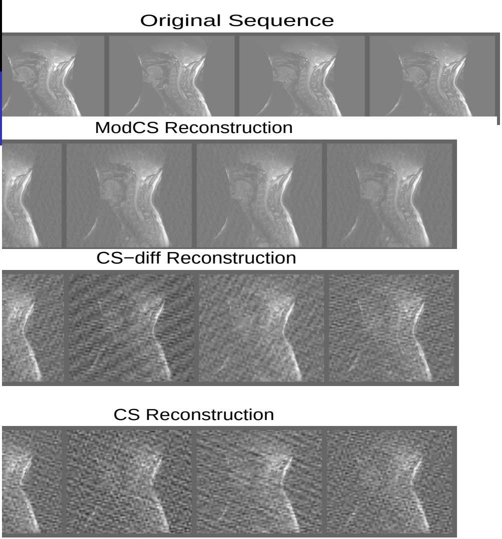
- use T_0 from prior knowledge, e.g. wavelet approximation coeff's
- may need more measurements at $t = 0$

ng example)



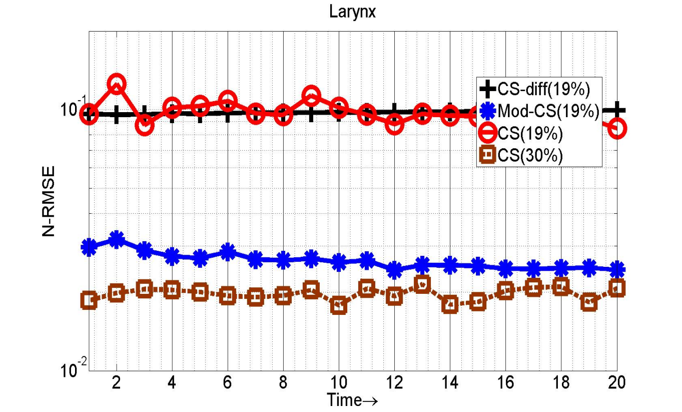
- Recovering a larynx sequence from only 19% simulated MRI measurements
- Proposed algorithm: Modified-CS. Here CS $\Leftrightarrow \ell_1$ min

ng example)



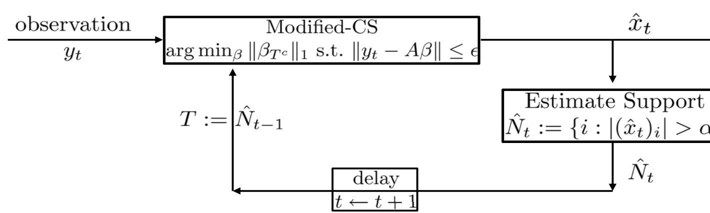
- Recovering a larynx sequence from only 19% simulated MRI measurements
- Proposed algorithm: Modified-CS. Here CS $\Leftrightarrow \ell_1$ min
- Modified-CS NRMSE was 3%. Simple ℓ_1-min NRMSE was 10%. It needed $n = 30\%$ meas's to get 3% error.

Dynamic MRI (larynx imaging example)



- With only $n = 19\%$ measurements, modified-CS error is small and stable below 3%
- Simple ℓ_1 needs $n = 30\%$ for same error

Modified-CS fo



- Difficulty with one step support estimation:
  - 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, $\Delta_{e,t} \subset T$
- Partial solutions: mod-CS-add-LS-del, regularized-mod-CS

Error stability over time (time-invariant error bounds) [Vaswani,T-SP, Aug'10] [6]

- Easy to bound the reconstruction error at a given time, t, but
  - the result depends on the support errors $|\Delta_t|$, $|\Delta_{e,t}|$, and these may increase over time
    (recall: $\Delta_t := N_t \setminus \hat{N}_{t-1}, \quad \Delta_{e,t} := \hat{N}_{t-1} \setminus N_t$)

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

Error stability over time (time-invariant error bounds) [Vaswani,T-SP, Aug'10] [6]

- Easy to bound the reconstruction error at a given time, t, but
  - the result depends on the support errors $|\Delta_t|$, $|\Delta_{e,t}|$, and these may increase over time
    (recall: $\Delta_t := N_t \setminus \hat{N}_{t-1}, \quad \Delta_{e,t} := \hat{N}_{t-1} \setminus N_t$)
- Key question for a recursive algorithm: when can we get a time-invariant and small bound on the error?
  - our work provides answers for modified-CS and modified-CS-add-LS-del

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

Main idea [Vaswani,Allerton'10], [Zhan,Vaswani,ISIT'13]

- Bound modified-CS error at time, t, in terms of $|\Delta_t|$, $|\Delta_{e,t}|$
  - require: number of support changes bounded by $S_a \ll S$ where S is upper bound on $|N_t|$
- Ensure: within a finite delay d_0, all newly added elements detected; all decreasing elements get deleted from $\hat{N}_t$
  - require: either every newly added support element is added at a large enough value or added small, but increases to a large enough value within a finite delay;
  - and decreasing elements become zero within a finite delay

Signal Change Model

1. $S_{a,t}$ additions and $S_{r,t}$ removals from support at time t
2. a new element j gets added at an initial magnitude $a_{j,t}$ and its magnitude increases at rate $r_{j,\tau}$ (at time τ) for the next $d_{j,t} \geq d_{\min}$ time units
3. $S_{d,t}$ elements out of the "large elements" set $\mathcal{L}_t$[7] leave the set and begin to decrease at time t
4. elements in $\mathcal{L}_t$ either remain in $\mathcal{L}_{t+1}$ (while increasing /decreasing /constant) or decrease enough to leave it
5. all decreasing elements that have left $\mathcal{L}_t$ get removed from support in at most b time units
6. $0 \leq S_{a,t} \leq S_a$, $0 \leq S_{r,t} \leq S_a$, $0 \leq S_{d,t} \leq S_a$, $|N_t| \leq S$.

[7] $\mathcal{L}_t := \{i \notin \cup^t \ldots \Delta \ldots : |(x_t)_i| > \ell\}$

Theorem (Modified-CS)

Assume signal model. If there exists a $d_0 \le d_{\min}$ s.t.

1. support estimation threshold, $\alpha = 7.50\epsilon$
2. RIP condition: $\delta_{S+kS_a}(A_t) \le 0.207$, $k := 3(\frac{(b+1)}{2} + d_0 + 1)$
3. new elements' initial mag or mag incr rate large enough

$$\min\{\ell, \min_{j:\mathbf{t_j}\neq\emptyset} \min_{t\in\mathbf{t_j}}(a_{j,t} + \sum_{\tau=t+1}^{t+d_0} r_{j,\tau})\} > \alpha + 7.50\epsilon,$$

4. $t = 0$: $\delta_{2S}(A_0) \le (\sqrt{2}-1)/2$ and enough large elements

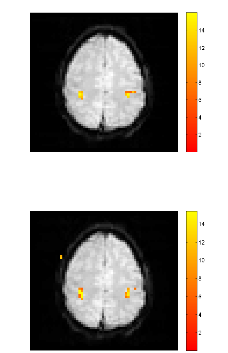
then, for all times, t,

1. $|\tilde{\Delta}_t| \le k' S_a$, $|\tilde{\Delta}_{e,t}| = 0$, (with $k' := \frac{(b+1)}{2} + d_0$)
2. and $\|x_t - \hat{x}_{t,modcs}\|_2 \le 7.50\epsilon$

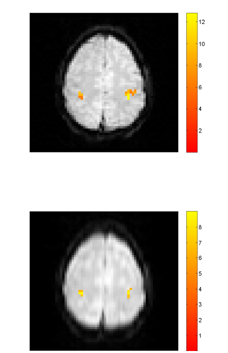
Time-invariant Error Bounds: Summary

- Support recovery error is bounded by a small and time-invariant value (small w.r.t. support size). Same true for recons error
- Results need weaker RIP conditions than simple ℓ_1
  - modified-CS needs $\delta_{S+kS_a}(A) \leq 0.2$, ℓ_1 needs $\delta_{2S}(A) \leq 0.2$,
- Other assumptions needed
  1. support threshold(s) appropriately set
  2. support size below S, support change size below S_a
  3. for any new element that is added to the support, either its initial magnitude is large enough or for the first few time instants, its magnitude increases at a large enough rate;
  4. a decreasing element decreases to zero within a short delay
  5. stronger RIP assumptions at $t = 0$

App based on detection



fu



dified-(

k-t-FOCUSS

simple ℓ_1

- Activation maps
- Used modified-CS for reconstructing the fMRI sequence; standard tools for active region detection
- Actual MRI scanner data; retrospective undersampling w/ $n_0 = 100\%$, $n = 30\%$,
- *Ongoing joint work with Dr. Ian Atkinson (UIC)*

Background: Robust PCA / Sparse + Low-Rank Recovery

- Most practical data are approximately low-dimensional. PCA: recovers the low-dim subspace of the data
- Robust PCA: PCA in presence of outliers. Many useful heuristics in older work, e.g., RSL [De la Torre et al,2003], others
- Recent work of Candes et al posed this as: separate a low-rank matrix L & a sparse matrix X from

$$Y := X + L$$

Background: Robust PCA / Sparse + Low-Rank Recovery

- Most practical data are approximately low-dimensional. PCA: recovers the low-dim subspace of the data
- Robust PCA: PCA in presence of outliers. Many useful heuristics in older work, e.g., RSL [De la Torre et al,2003], others
- Recent work of Candes et al posed this as: separate a low-rank matrix L & a sparse matrix X from

$$Y := X + L$$

- PCP (convex opt sol): [Candes et al, Chandrasekharan et al, 2011]

$$\min_{\tilde{X},\tilde{L}} \|\tilde{L}\|_* + \lambda\|\tilde{X}\|_1 \text{ s.t. } Y = \tilde{X} + \tilde{L}$$

- If (a) left and right singular vectors of L dense enough; (b) rank of L small; (c) support of X generated uniformly at random; then PCP gives exact recovery w.h.p.

The need for a recursive / online solution

- Causal: needed for video surveillance, Netflix problem, ...
- Fast and memory efficient compared to batch solutions
- Exploit temporal dependencies in the dataset; sometimes no practical (polynomial complexity) way to do this in a batch fashion w/o putting (Bayesian) priors

Recursive Sparse + Low-Rank Recovery [Qiu,Vaswani,Allerton'10,'11][8]

The Problem:

- Given sequentially arriving data vectors

$$y_t := x_t + \ell_t, \quad t = 1, 2, \ldots$$

  - x_t's are sparse vectors,
  - $T_t :=$ support(x_t) changes over time (not constant),
  - ℓ_t's lie in a fixed or "slowly changing" low-dimensional subspace,
  - ℓ_t's are dense,
- and given an estimate of span$([\ell_1, \ell_2, \ldots \ell_{t_0}])$,

[8] 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

Recursive Sparse + Low-Rank Recovery [Qiu,Vaswani,Allerton'10,'11][8]

The Problem:

- Given sequentially arriving data vectors

$$y_t := x_t + \ell_t, \quad t = 1, 2, \ldots$$

  - x_t's are sparse vectors,
  - $T_t := \text{support}(x_t)$ changes over time (not constant),
  - ℓ_t's lie in a fixed or "slowly changing" low-dimensional subspace,
  - ℓ_t's are dense,
- and given an estimate of span($[\ell_1, \ell_2, \ldots \ell_{t_0}]$),
- Goal: recursively recover x_t and ℓ_t at all $t > t_0$.

[8] 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

- Various interpretations:
  - Recursive sparse recovery in large but structured noise
    - large noise: $\|\ell_t\|$ can be much larger than $\|x_t\|$
    - extension to the $y_t := Ax_t + B\ell_t$ is easy
  - Recursive robust PCA:
    - x_t is outlier, recover ℓ_t and span$([\ell_1, \ell_2, \ldots \ell_t])$
  - Recursive matrix completion: simpler special case of above
- Applications: fg and bg extraction in video (e.g. for surveillance apps), brain activity detection in fMRI, dynamic Netflix problem, ...
- **Almost all existing work with performance guarantees: batch solutions.**

Our Solution: Recursive Projected CS (ReProCS) [Qiu,Vaswani,Allerton'10,Allerton'11][9]

Recall that $y_t := x_t + \ell_t$

Our Solution: Recursive Projected CS (ReProCS) [Qiu,Vaswani,Allerton'10,Allerton'11][9]

Recall that $y_t := x_t + \ell_t$

Initialize: SVD on training background data to compute $\hat{P}_0$

For $t > t_0$

- Projection: compute $\tilde{y}_t := \Phi_{(t)} y_t$, where $\Phi_{(t)} := I - \hat{P}_{(t-1)} \hat{P}'_{(t-1)}$
  - then $\tilde{y}_t = \Phi_{(t)} x_t + \beta_t$, $\beta_t := \Phi_{(t)} \ell_t$ is small "noise"

Our Solution: Recursive Projected CS (ReProCS)[Qiu,Vaswani,Allerton'10,Allerton'11][9]

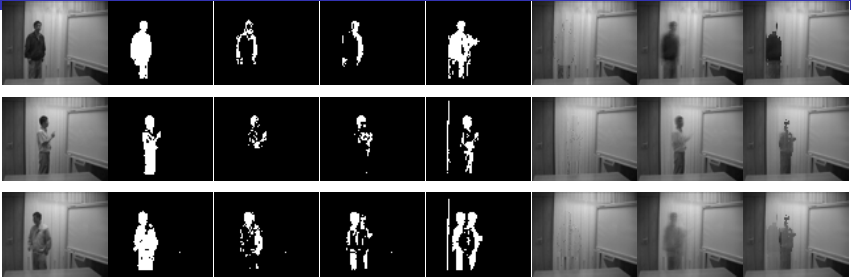
Recall that $y_t := x_t + \ell_t$

Initialize: SVD on training background data to compute $\hat{P}_0$

For $t > t_0$

- Projection: compute $\tilde{y}_t := \Phi_{(t)} y_t$, where $\Phi_{(t)} := I - \hat{P}_{(t-1)} \hat{P}'_{(t-1)}$
  - then $\tilde{y}_t = \Phi_{(t)} x_t + \beta_t$, $\beta_t := \Phi_{(t)} \ell_t$ is small "noise"
- Sparse Recovery: ℓ_1 min + support estimation + LS: get $\hat{x}_t$
- Get $\hat{\ell}_t = y_t - \hat{x}_t$
- Subspace update: use the last α $\hat{\ell}_t$'s to update $\hat{P}_{(t)}$: projection-PCA or its practical version
  - simple PCA not work: $e_t := \hat{\ell}_t - \ell_t = x_t - \hat{x}_t$ correlated with ℓ_t

Application: Video Layering (Fg and Bg extraction)



original video	ReProCS fg	PCP fg	RSL fg	GRASTA fg	ReProCS bg	PCP bg	RSL bg

- Separating fg and bg layers in a real video seq: bg is window curtains moving due to wind
- Proposed algorithm: ReProCS; RSL: [de la Torre et al, 2003], GRASTA: [Balzano et al,CVPR 2012]

Quantifying Denseness of span(B) [Qiu,Vaswani, ISIT'2013, ICASSP 2013][10]

- Define the denseness coefficient for a matrix/vector B as

$$\kappa_s(B) = \kappa_s(\text{span}(B)) := \max_{|T| \le s} \|{I_T}' Q(B)\|_2$$

where $Q(B)$ is an ortho basis for span(B)

  - intuition: if B is a vector, then $Q = B/\|B\|_2$, κ_s small means B is a dense vector

Lemma (relation to RIC)

Let $\Phi := I - Q(B)Q(B)'$. *Then* $\delta_s(\Phi) = \kappa_s(B)^2$.

[10] C. Qiu, N. Vaswani, B. Lois and L. Hogben, Recursive Robust PCA or Recursive Sparse Recovery in Large but Structured Noise, revised for IEEE Trans. IT, 2013, shorter versions in ISIT and ICASSP 2013

Model on ℓ_t

- $\ell_t = P_{(t)} a_t$ where $P_{(t)} = P_j$ for $t \in [t_j, t_{j+1} - 1]$,
  - P_j: tall $n \times r_j$ matrix with ortho col's that changes as

$$P_j = [P_{j-1}, \;\; P_{j,\text{new}}]$$

  - $r_j \ll n$, $r_j \ll (t_{j+1} - t_j)$, $0 \leq \text{rank}(P_{j,\text{new}}) \leq c$
  - a_t is a zero mean bounded random variable: $\|a_t\|_\infty \leq \gamma_*$
  - a_t's mutually independent over time
  - $j = 1, 2, \ldots J$ (total of J subspace change times)

Model on ℓ_t

- $\ell_t = P_{(t)} a_t$ where $P_{(t)} = P_j$ for $t \in [t_j, t_{j+1} - 1]$,
  - P_j: tall $n \times r_j$ matrix with ortho col's that changes as

$$P_j = [P_{j-1},\ \ P_{j,\text{new}}]$$

  - $r_j \ll n$, $r_j \ll (t_{j+1} - t_j)$, $0 \leq \text{rank}(P_{j,\text{new}}) \leq c$
  - a_t is a zero mean bounded random variable: $\|a_t\|_\infty \leq \gamma_*$
  - a_t's mutually independent over time
  - $j = 1, 2, \ldots J$ (total of J subspace change times)
- Define $f := \frac{\max_t \lambda_{\max}(\Lambda_t)}{\min_t \lambda_{\min}(\Lambda_t)}$ where $\Lambda_t := \text{Cov}(a_t)$

No bound needed on f or on γ_*: allow large but structured ℓ_t

Subspace update: Projection PCA

Assume $t_{j+1} - t_j > K\alpha$; recall: t_j: subspace change times

at $t = t_j + k\alpha$, compute $\hat{P}_{j,\text{new},k}$ as the c "top" left singular vectors of $(I - \hat{P}_{j-1}\hat{P}_{j-1}')[\hat{\ell}_{t_j+(k-1)\alpha}, \dots \hat{\ell}_{t_j+k\alpha-1}]$; update $\hat{P}_{(t)} = [\hat{P}_{j-1}, \hat{P}_{j,\text{new},k}]$

Theorem

Pick $\zeta \leq \min(\frac{10^{-4}}{(r_0+Jc)^2 f}, \frac{1}{(r_0+Jc)^3\gamma_*^2})$. *Assume the model on* ℓ_t*, algorithm parameters appropriately set &* $\|(I - \hat{P}_0\hat{P}_0')P_0\|_2 \leq r_0\zeta$. *If*

1. *slow subspace change holds:*
   - $\min_j(t_{j+1} - t_j) \geq K\alpha$ *and*
   - $\max_{t\in[t_j+(k-1)\alpha, t_j+k\alpha)} \|P_{j,new}'\ell_t\|_\infty \leq 1.2^{k-1}\gamma_{new}$ *with* γ_{new} *small enough*

Theorem

Pick $\zeta \leq \min(\frac{10^{-4}}{(r_0+Jc)^2 f}, \frac{1}{(r_0+Jc)^3 \gamma_*^2})$. *Assume the model on* ℓ_t, *algorithm parameters appropriately set &* $\|(I - \hat{P}_0 \hat{P}_0')P_0\|_2 \leq r_0 \zeta$. *If*

1. *slow subspace change holds:*
   - $\min_j (t_{j+1} - t_j) \geq K\alpha$ *and*
   - $\max_{t \in [t_j+(k-1)\alpha, t_j+k\alpha)} \|P_{j,new}' \ell_t\|_\infty \leq 1.2^{k-1} \gamma_{new}$ *with* γ_{new} *small enough*
2. *denseness of subspace holds:*

$$\max_j \kappa_{2s}(P_j) \leq 0.3, \; \textit{and} \; \max_j \kappa_s(P_{j,new}) \leq 0.15,$$

Theorem

Pick $\zeta \leq \min(\frac{10^{-4}}{(r_0+Jc)^2 f}, \frac{1}{(r_0+Jc)^3 \gamma_*^2})$. *Assume the model on* ℓ_t, *algorithm parameters appropriately set &* $\|(I - \hat{P}_0 \hat{P}_0')P_0\|_2 \leq r_0 \zeta$. *If*

1. *slow subspace change holds:*
   - $\min_j (t_{j+1} - t_j) \geq K\alpha$ *and*
   - $\max_{t \in [t_j+(k-1)\alpha, t_j+k\alpha)} \|P_{j,new}' \ell_t\|_\infty \leq 1.2^{k-1} \gamma_{new}$ *with* γ_{new} *small enough*
2. *denseness of subspace holds:*
   $$\max_j \kappa_{2s}(P_j) \leq 0.3, \text{ and } \max_j \kappa_s(P_{j,new}) \leq 0.15,$$
3. $c = 1$ *or condition number of* $Cov(a_{t,new})$ *below 1.4 at all times,*

Theorem

Pick $\zeta \leq \min(\frac{10^{-4}}{(r_0+Jc)^2 f}, \frac{1}{(r_0+Jc)^3\gamma_*^2})$. *Assume the model on* ℓ_t*, algorithm parameters appropriately set &* $\|(I - \hat{P}_0\hat{P}_0')P_0\|_2 \leq r_0\zeta$. *If*

1. *slow subspace change holds:*
   - $\min_j(t_{j+1} - t_j) \geq K\alpha$ *and*
   - $\max_{t\in[t_j+(k-1)\alpha, t_j+k\alpha)} \|P_{j,new}'\ell_t\|_\infty \leq 1.2^{k-1}\gamma_{new}$ *with* γ_{new} *small enough*
2. *denseness of subspace holds:*
$$\max_j \kappa_{2s}(P_j) \leq 0.3, \text{ and } \max_j \kappa_s(P_{j,new}) \leq 0.15,$$
3. $c = 1$ *or condition number of* $Cov(a_{t,new})$ *below 1.4 at all times,*
4. $\max_{k=1,2,\dots K} \kappa_s(D_{new,k}) \leq 0.153$,
   $D_{new,k} := (I - \hat{P}_{j-1}\hat{P}_{j-1}' - \hat{P}_{j,new,k}\hat{P}_{j,new,k}')P_{j,new}$,

then, w.p. $> 1 - n^{-10}$, $\hat{T}_t = T_t$ **and** $\|x_t - \hat{x}_t\|_2 \leq b \ll \|x_t\|_2$

Discussion: Main Limitations and Ongoing Work

- Result is not a correctness result because the $D_{\text{new},k}$ assumption depends on algorithm estimates
  - Ongoing work (to be submitted to NIPS 2014):
    - replaces this by an assumption on support change of x_t's, gives a correctness result

Discussion: Main Limitations and Ongoing Work

- Result is not a correctness result because the $D_{\text{new},k}$ assumption depends on algorithm estimates
  - Ongoing work (to be submitted to NIPS 2014):
    - replaces this by an assumption on support change of x_t's, gives a correctness result
- Result assumes independence of ℓ_t's over time
  - Ongoing work (ISIT 2014):
    - replaces this by an autoregressive model on the ℓ_t's

Discussion: Main Limitations and Ongoing Work

- Result is not a correctness result because the $D_{\text{new},k}$ assumption depends on algorithm estimates
  - Ongoing work (to be submitted to NIPS 2014):
    - replaces this by an assumption on support change of x_t's, gives a correctness result
- Result assumes independence of ℓ_t's over time
  - Ongoing work (ISIT 2014):
    - replaces this by an autoregressive model on the ℓ_t's
- Algorithm that is analyzed assumes knowledge of subspace change times and number of changed directions

Discussion: Contributions

- Among the first works to analyze the online (recursive) robust PCA problem
  - equivalently also among the first results for recursive sparse recovery in large but low-dimensional noise

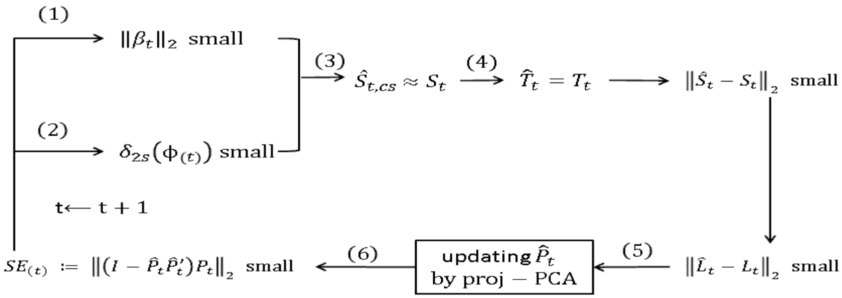
Discussion: Contributions

- Among the first works to analyze the online (recursive) robust PCA problem
  - equivalently also among the first results for recursive sparse recovery in large but low-dimensional noise
- New proof techniques needed to obtain the result
  - all existing robust PCA results are for batch approaches
  - all previous finite sample PCA results assume $e_t := \hat{\ell}_t - \ell_t$ is uncorrelated with ℓ_t

Discussion: Contributions

- Among the first works to analyze the online (recursive) robust PCA problem
  - equivalently also among the first results for recursive sparse recovery in large but low-dimensional noise
- New proof techniques needed to obtain the result
  - all existing robust PCA results are for batch approaches
  - all previous finite sample PCA results assume $e_t := \hat{\ell}_t - \ell_t$ is uncorrelated with ℓ_t
- Result 2: allows subspace removals, Advantage:
  - no bound needed on # of subspace changes, J, as long as $(t_{j+1} - t_j)$ increases in proportion to $\log J \Leftrightarrow$ no bound on $\mathrm{rank}(L)$

Proof Outline: Overall idea



In the figure: $S_t \equiv x_t$, $L_t \equiv \ell_t$

Let $\Phi := (I - \hat{P}_{(t-1)}\hat{P}'_{(t-1)})$, $\beta_t := \Phi \ell_t$ (noise seen by ℓ_1 step)

Proof Outline: Key steps

- Define subspace error, $\text{SE}(P, \hat{P}) := \|(I - \hat{P}\hat{P}')P\|_2$.
- Start with $\text{SE}(P_{j-1}, \hat{P}_{j-1}) \leq r_{j-1}\zeta \ll 1$.
- Key steps
  1. Analyze projected sparse recovery for $t \in [t_j, t_j + \alpha)$
  2. Analyze projection-PCA at $t = t_j + \alpha - 1$
  3. Repeat for each of the K projection-PCA intervals: show that $\text{SE}(P_{\text{new}}, \hat{P}_{\text{new},k}) \leq 0.6^k + 0.4c\zeta$
  4. Pick K so that $0.6^k + 0.4c\zeta \leq c\zeta$
- Thus, $\text{SE}(P_j, \hat{P}_j) \leq \text{SE}(P_{j-1}, \hat{P}_{j-1}) + \text{SE}(P_{\text{new}}, \hat{P}_{\text{new},K}) \leq r_{j-1}\zeta + c\zeta = r_j\zeta$

Proof Outline: Projected sparse recovery for $t \in [t_j, t_j + \alpha)$

1. Recall: $P_{(t)} = [P_{j-1}, P_{\text{new}}]$, $\hat{P}_{(t-1)} = \hat{P}_{j-1}$, $\tilde{y}_t := \Phi y_t = \Phi x_t + \beta_t$, where $\Phi := I - \hat{P}_{(t-1)} \hat{P}'_{(t-1)}$ and $\beta_t := \Phi \ell_t$
2. Using slow subspace change,
$$\|\beta_t\|_2 \le \sqrt{\zeta} + \sqrt{c}\gamma_{\text{new}}$$
3. Using denseness,
$$\delta_s(\Phi) = \kappa_s(\hat{P}_{j-1})^2 \le \kappa_s(P_{j-1})^2 + r\zeta \le 0.1$$
4. Thus, $\|\hat{x}_{t,cs} - x_t\| \le 7\sqrt{c}\gamma_{\text{new}}$
5. Appropriate support threshold & γ_{new} small $\Rightarrow$ $\hat{T}_t = T_t$
6. LS step: get exact expression for $e_t := x_t - \hat{x}_t = \hat{\ell}_t - \ell_t$
$$e_t = I_{T_t} [\Phi_{T_t}{}' \Phi_{T_t}]^{-1} I_{T_t}{}' \Phi \ell_t$$

Proof Outline: Projection-PCA at $t = t_j + \alpha - 1$

1. Bound $\text{SE}(P_{\text{new}}, \hat{P}_{\text{new},1})$ in terms of minimum eigenvalue of the signal subspace part of the true data matrix, $\sum_t \Phi_{j,0} \ell_t \ell_t' \Phi_0'$, and the maximum eigenvalue of the perturbation matrix, $\sum_t \Phi_0 (\ell_t \ell_t' - \hat{\ell}_t \hat{\ell}_t') \Phi_0'$
   - use $\sin\theta$ theorem: 1970s linear algebra result of Kahan,Davis
2. Get high probability bounds on each of the terms in this bound
   - use the matrix Hoeffding inequality: Tropp 2012
3. Simplify using denseness of $D_{\text{new}} := (I - \hat{P}_{j-1} \hat{P}_{j-1}') P_{\text{new}}$ to get $\text{SE}(P_{\text{new}}, \hat{P}_{\text{new},1}) \leq 0.6$
   - easy to see $\kappa_s(D_{\text{new}}) \leq 1.01 \kappa_s(P_{\text{new}}) + 0.01 \leq 0.153$

Proof Outline: k-th projection PCA interval

- $P_{(t)} = [P_{j-1}, P_{\text{new}}]$, $\hat{P}_{(t-1)} = [\hat{P}_{j-1}, \hat{P}_{\text{new},k-1}]$.
- Using slow subspace change,

$$\|\beta_t\|_2 \leq \sqrt{\zeta} + 0.6^{k-1}\sqrt{c}\gamma_{\text{new}}$$

- Smaller $\beta_t \Rightarrow$ smaller $e_t = x_t - \hat{x}_t = \hat{\ell}_t - \ell_t \Rightarrow$ smaller $\text{SE}(P_{\text{new}}, \hat{P}_{\text{new},k}) \Rightarrow$ even smaller β_t at next iteration
- Can show $\text{SE}(P_{\text{new}}, \hat{P}_{\text{new},k}) \leq 0.6^k + 0.4c\zeta$

Proof Outline: k-th projection PCA interval

- $P_{(t)} = [P_{j-1}, P_{\text{new}}]$, $\hat{P}_{(t-1)} = [\hat{P}_{j-1}, \hat{P}_{\text{new},k-1}]$.
- Using slow subspace change,

$$\|\beta_t\|_2 \leq \sqrt{\zeta} + 0.6^{k-1}\sqrt{c}\gamma_{\text{new}}$$

- Smaller β_t $\Rightarrow$ smaller $e_t = x_t - \hat{x}_t = \hat{\ell}_t - \ell_t$ $\Rightarrow$ smaller $\text{SE}(P_{\text{new}}, \hat{P}_{\text{new},k})$ $\Rightarrow$ even smaller β_t at next iteration
- Can show $\text{SE}(P_{\text{new}}, \hat{P}_{\text{new},k}) \leq 0.6^k + 0.4c\zeta$
- Pick K so that $\text{SE}(P_{\text{new}}, \hat{P}_{\text{new},k}) \leq c\zeta$
- Thus, $\text{SE}(P_j, \hat{P}_j) \leq \text{SE}(P_{j-1}, \hat{P}_{j-1}) + \text{SE}(P_{\text{new}}, \hat{P}_{\text{new},K}) \leq r_j\zeta$

Conclusions and Future Directions I

- Studied two recursive structured signals' recovery problems
  1. recursive sparse signals' recovery
  2. recursive sparse plus low-dimensional signals' recovery
- Problem 1: reformulate as sparse rec w/ partial support knowledge
  - Modified-CS has significantly improved recovery for proof-of-concept dynamic MRI expts
  - its exact recovery conditions weaker than those for simple ℓ_1
  - its error is bounded by a time-invariant and small value under mild assumptions in the noisy case
- Problem 2:

Conclusions and Future Directions II

- ReProCS has significantly improved performance compared w/ existing robust PCA solutions for difficult videos
- Obtained conditions for its exact support recovery w.h.p.

- Future Directions
  - Correctness result for ReProCS: ongoing
  - ReProCS for other "big-data" applications
  - ReProCS for fMRI based brain activity detection