Compressive acquisition of linear dynamical systems

Amirafshar Moshtaghpour

May 2015

ISP Group

icteam

UCL Université catholique de Louvain

Outline

- Background
- CS-LDS Architecture
- Estimating the state sequence
- Estimating the observation matrix
- Conclusion

Outline

- Background
- CS-LDS Architecture
- Estimating the state sequence
- Estimating the observation matrix
- Conclusion

Background

Compressed Sensing (CS)

- Original signal: $\mathbf{y} \in \mathbb{R}^N$
- K-sparse signal: $\mathbf{s} \in \mathbb{R}^N$
  - $\mathbf{y} = \Psi\mathbf{s}$
  - $\mathbf{s}$ has at most K non-zero elements
- Measurement matrix: $\Phi \in \mathbb{R}^{M \times N}$
  - $K < M \ll N$

$$\mathbf{z} = \Phi\mathbf{y} + \mathbf{e}$$

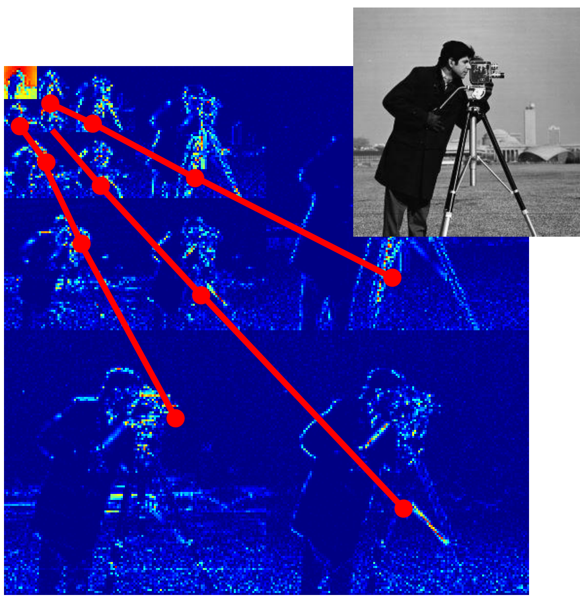
- Measurement vector: $\mathbf{z} \in \mathbb{R}^M$
- Measurement noise: $\mathbf{e} \in \mathbb{R}^M$

One possibility to recover $\mathbf{y}$
$\Phi \sim i.i.d$ Gaussian
$M = 4K \log\frac{N}{K}$

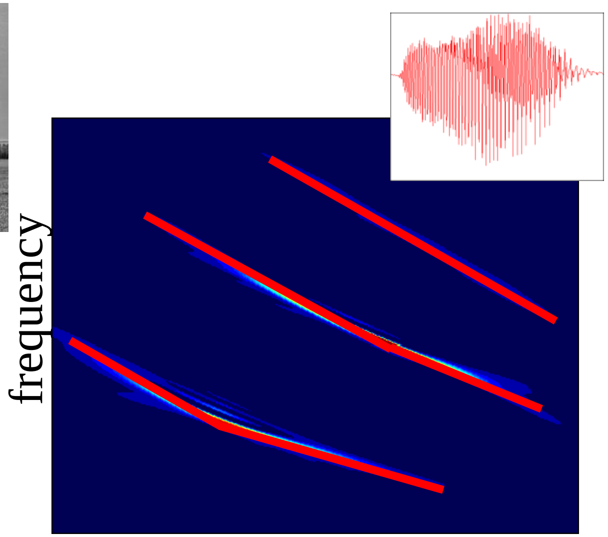
Background

Compressed Sensing (CS)

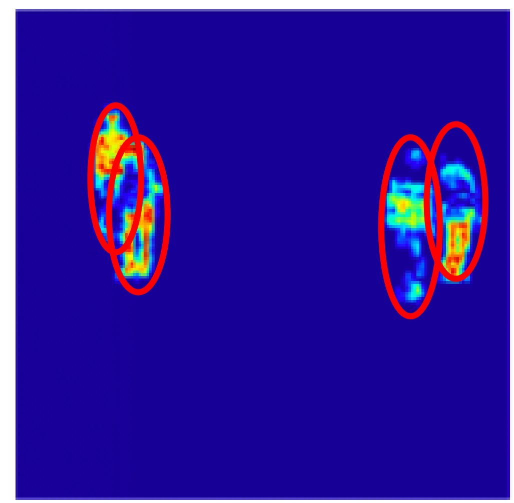
- Sparse Signals
- Structured-Sparse Signals



wavelet



time

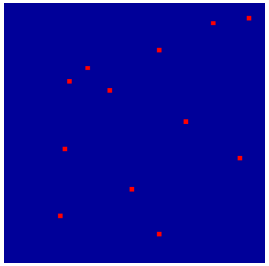


background subtracted image

Background

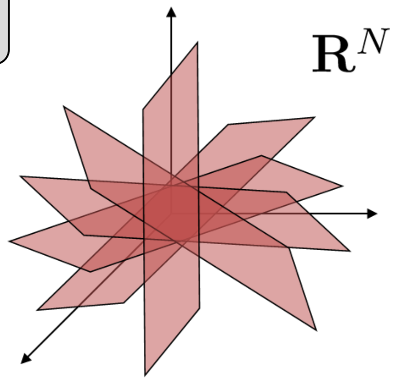
Compressed Sensing (CS)

K-sparse signals comprise a particular set of *K*-dim subspaces

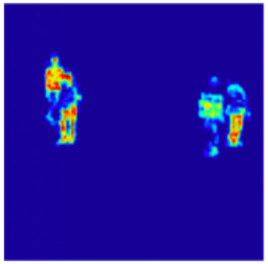


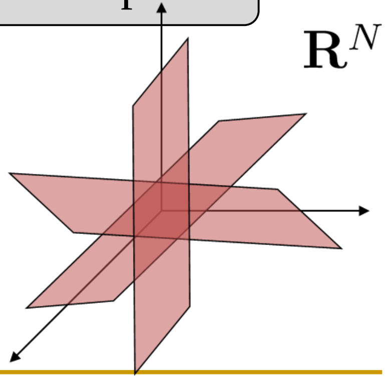
$$\|signal\|_0 \leq K$$

union of *K*-dimensional subspaces



A *K*-sparse **signal model** comprises a particular (reduced) set of *K*-dim subspaces





Background

Compressed Sensing (CS) [4, 5]

$a = \text{Model-CoSAMP}(\Phi, \mathbf{u}, s)$ ← $\mathbf{a} = \text{CoSaMP}(\Phi, \mathbf{u}, s)$

Input: Sampling Matrix Φ, measurement vector $\mathbf{u}$, sparsity level s

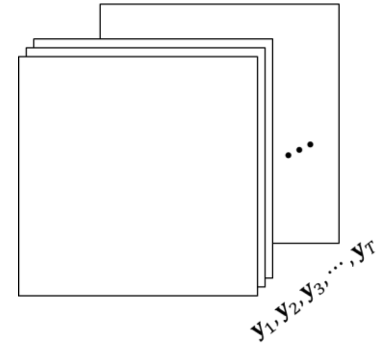
Output: An s-sparse approximation $\mathbf{a}$ of the target signal

$\mathbf{a}^0 \leftarrow \mathbf{0}$	Trivial initial approximation
$\mathbf{v} \leftarrow \mathbf{u}$	Current samples = input samples
$k \leftarrow 0$	Iteration index
repeat	
$k \leftarrow k + 1$	
$\mathbf{y} \leftarrow \Phi^* \mathbf{v}$	Form signal proxy
$\Omega \leftarrow \text{supp}(\mathbf{y}_{2s})$	Identify large components
$T \leftarrow \Omega \cup \text{supp}(\mathbf{a}^{k-1})$	Merge supports
$\mathbf{b}\|_T \leftarrow \Phi^\dagger \mathbf{u}$	Signal estimation by least-square
$\mathbf{b}\|_{T^c} \leftarrow \mathbf{0}$	
$\mathbf{a}^k \leftarrow \mathbf{b}_s$	Prune to obtain next approximation
$\mathbf{v} \leftarrow \mathbf{u} - \Phi \mathbf{a}^k$	Update current samples
until halting criterion true	

$\Omega \leftarrow \text{supp}(\mathbb{M}_2(y, s))$ ← $\Omega \leftarrow \text{supp}(\mathbf{y}_{2s})$

$a^k \leftarrow \mathbb{M}(b, s)$ ← $\mathbf{a}^k \leftarrow \mathbf{b}_s$

Background



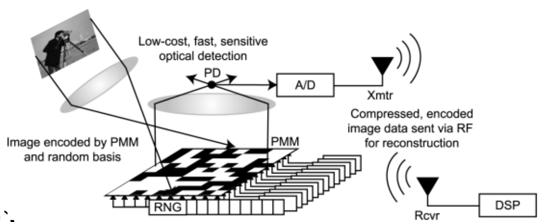
Video compressive sensing

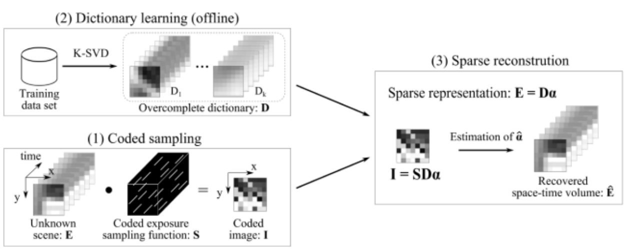
- $\mathbf{y}_t$: the image of a scene at time t
- $\mathbf{Y} = \mathbf{y}_{1:T} = [\mathbf{y}_1, \cdots, \mathbf{y}_T]$: video of the scene from time 1 to T

Goal: to recover $\mathbf{y}_{1:T}$ given $\mathbf{z}_{1:T}$

$$\mathbf{z}_t = \Phi_t \mathbf{y}_t$$

1. Single Pixel Camera (SPC)
   - Duarte *et al*, 2008
2. Programmable Pixel Camera (P2C)
   - Hitomi *et al*, 2011
   - Reddy *et al*, 2011
   - Veeraraghavan *et al*, 2011





Background

Linear Dynamical System (LDS)

- **Dynamical system:** Change of some variables (*state variables*)
  - **Continuous vs Discrete**
  - **Linear vs Non-linear**

Discrete-time LDS:

$$\mathbf{x}_{t+1} = \mathbf{A}_t\mathbf{x}_t + \mathbf{B}_t\mathbf{u}_t$$
$$\mathbf{y}_t = \mathbf{C}_t\mathbf{x}_t + \mathbf{D}_t\mathbf{u}_t$$

TI autonomous discrete-time LDS:

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t$$
$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t$$

$t \in \mathbb{R}$: time
$\mathbf{x} \in \mathbb{R}^d$: state vector (variables)
$\mathbf{u} \in \mathbb{R}^m$: input vector
$\mathbf{y} \in \mathbb{R}^N$: observation (output) vector $\neq$ measurement vector
$\mathbf{A} \in \mathbb{R}^{d\times d}$: state transition (dynamic) matrix
$\mathbf{B} \in \mathbb{R}^{d\times m}$: input matrix
$\mathbf{C} \in \mathbb{R}^{N\times d}$: observation (output or sensor) matrix
$\mathbf{D} \in \mathbb{R}^{N\times m}$: feed-through matrix

Ambiguity !!!

LDS $(\mathbf{A}, \mathbf{C}, \mathbf{x}) \equiv$ LDS $(\mathbf{L}^{-1}\mathbf{A}\mathbf{L}, \mathbf{C}\mathbf{L}, \mathbf{L}^{-1}\mathbf{x})$

For any invertible matrix $\mathbf{L} \in \mathbb{R}^{d\times d}$

Background

<u>Linear Dynamical System (LDS)</u>

- A matrix **H** is called *Hankel matrix* if the entries on the anti-diagonals be the same, i.e. $H_{i,j} = H_{i-1,j+1}$

Given $\mathbf{h} \in \mathbb{R}^N \rightarrow$ build $\mathbf{H} \in \mathbb{R}^{L\times K}$

Hankel matrix

$$\mathbf{H} = \begin{bmatrix} h_1 & h_2 & \cdots & h_K \\ h_2 & h_3 & \cdots & h_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ h_L & h_{L+1} & \cdots & h_N \end{bmatrix}$$

$K = N - L + 1$

Given $\mathbf{Y} = \mathbf{y}_{1:T} \in \mathbb{R}^{N\times T} \rightarrow$ build $\mathbf{H} \in \mathbb{R}^{LN\times K}$

Block-Hankel matrix

$$\mathbf{H} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_K \\ \mathbf{y}_2 & \mathbf{y}_3 & \cdots & \mathbf{y}_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_L & \mathbf{y}_{L+1} & \cdots & \mathbf{y}_T \end{bmatrix}$$

$K = T - L + 1$

Background

Linear Dynamical System (LDS)

$$\mathbf{H} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_K \\ \mathbf{y}_2 & \mathbf{y}_3 & \cdots & \mathbf{y}_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_L & \mathbf{y}_{L+1} & \cdots & \mathbf{y}_T \end{bmatrix} = \begin{bmatrix} \mathbf{C}\mathbf{x}_1 & \mathbf{C}\mathbf{x}_2 & \cdots & \mathbf{C}\mathbf{x}_K \\ \mathbf{C}\mathbf{A}\mathbf{x}_1 & \mathbf{C}\mathbf{A}\mathbf{x}_2 & \cdots & \mathbf{C}\mathbf{A}\mathbf{x}_K \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}\mathbf{A}^{L-1}\mathbf{x}_1 & \mathbf{C}\mathbf{A}^{L-1}\mathbf{x}_2 & \cdots & \mathbf{C}\mathbf{A}^{L-1}\mathbf{x}_K \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A} \\ \vdots \\ \mathbf{C}\mathbf{A}^{L-1} \end{bmatrix} \times [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_K]$$

$$= \mathbf{O}(C, A)\mathbf{C}(\mathbf{x}),$$

$$\mathbf{H} = \mathbf{U}_d \mathbf{S}_d \mathbf{V}_d^T,$$

$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t$
↓
$\mathbf{x}_1 = \mathbf{x}_1$
$\mathbf{x}_2 = \mathbf{A}\mathbf{x}_1$
$\vdots$
$\mathbf{x}_K = \mathbf{A}^{K-1}\mathbf{x}_1$

$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t$
↓
$\mathbf{y}_1 = \mathbf{C}\mathbf{x}_1$
$\mathbf{y}_2 = \mathbf{C}\mathbf{x}_2 = \mathbf{C}\mathbf{A}\mathbf{x}_1$
$\vdots$
$\mathbf{y}_L = \mathbf{C}\mathbf{x}_L = \boldsymbol{C}\mathbf{A}^{L-1}\mathbf{x}_1$

$\mathbf{O} \in \mathbb{R}^{LN\times d}$: Observability matrix
$\mathbf{C}(\mathbf{x}) \in \mathbb{R}^{d\times K}$: Controllability matrix

$$\mathbf{O}(\mathbf{C}, \mathbf{A}) = \mathbf{U}_d$$
$$[\mathbf{x}_{1:T-L+1}] = \mathbf{S}_d \mathbf{V}_d^T$$

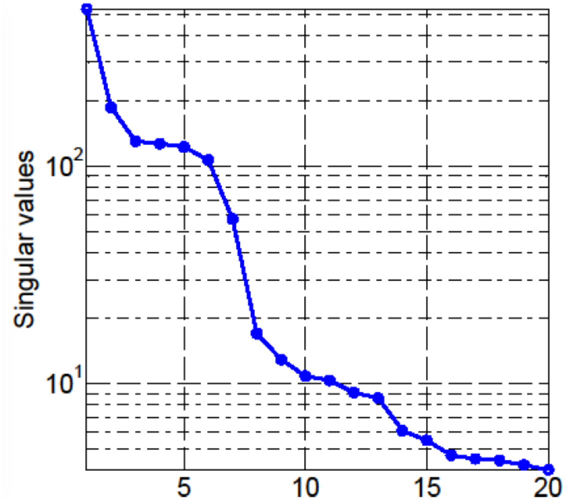
LDS $(\mathbf{A}, \mathbf{C}, \mathbf{x}) \equiv$ LDS $(\mathbf{L}^{-1}\mathbf{A}\mathbf{L}, \mathbf{C}\mathbf{L}, \mathbf{L}^{-1}\mathbf{x})$

For any invertible matrix $\mathbf{L} \in \mathbb{R}^{d\times d}$

Background

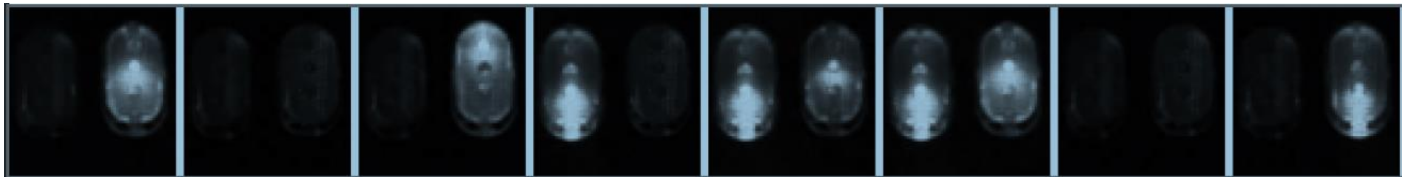
<u>LDS model for video sequences</u>

- Challenges for video sequences:
  - Ephemeral nature of videos
  - High-dimensional signals

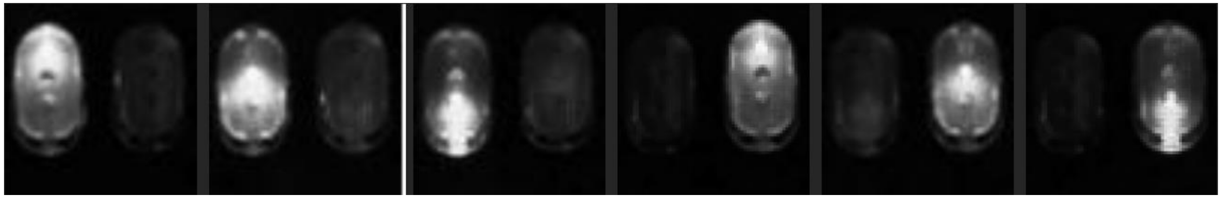


(b) Top 20 singular values of the data matrix

Few frames



Six basis frames



All frames can be estimated using linear combinations of SIX images

Outline

- Background
- CS-LDS Architecture
- Estimating the state sequence
- Estimating the observation matrix
- Conclusion

CS-LDS Architecture

Authors: A. C. Sankaranarayanan, P. K. Turaga, R. Chellappa, and R. G. Baraniuk, 2013

Goal: to build a CS framework, implementable on the SPC, for videos that are modeled as LDS.

- We seek to recovery $\mathbf{C}$ and $\mathbf{x}_{1:T}$, given compressive measurements of the form

$$\mathbf{z}_t = \Phi_t \mathbf{y}_t = \Phi_t \mathbf{C} \mathbf{x}_t$$

  - $\mathbf{z}_t \in \mathbb{R}^M, \Phi_t \in \mathbb{R}^{M \times N}$
  - Bilinear unknowns → non-convex optimization problem

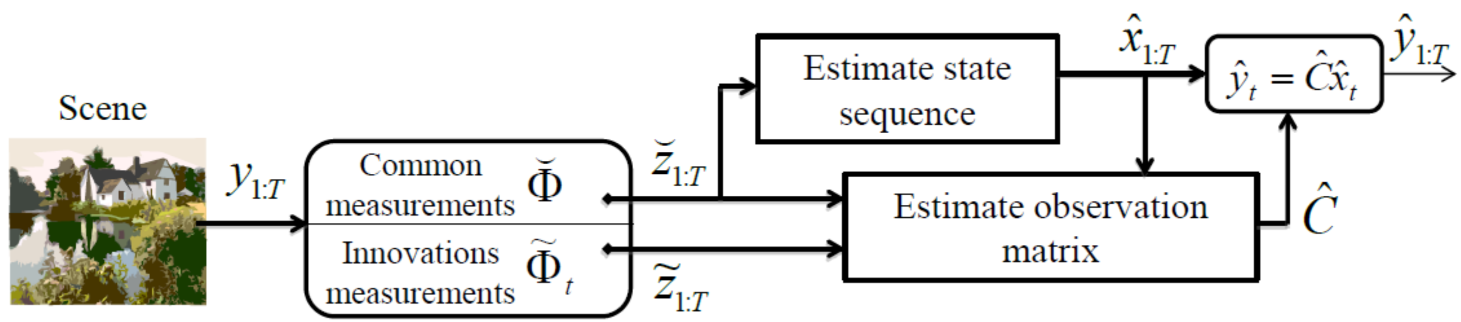
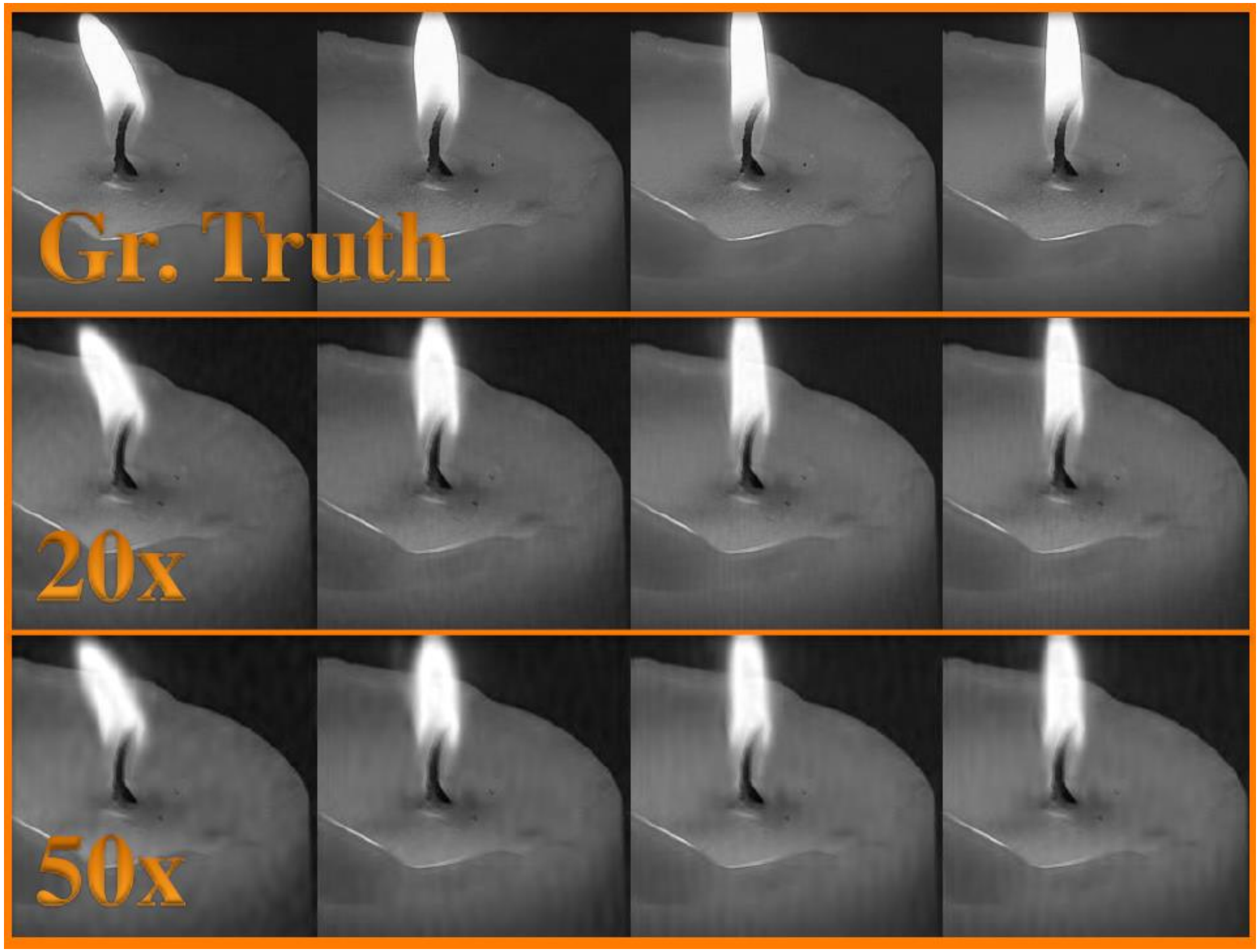


FIG. 2. *Block diagram of the CS-LDS framework.*

CS-LDS Architecture

$\frac{N}{M} = 20$, SNR: 25.81 dB

$\frac{N}{M} = 50$, SNR: 24.09 dB



CS-LDS Architecture

$$\mathbf{z}_t = \begin{bmatrix} \check{\mathbf{z}}_t \\ \tilde{\mathbf{z}}_t \end{bmatrix} = \begin{bmatrix} \check{\Phi} \\ \tilde{\Phi}_t \end{bmatrix} \mathbf{y}_t, \longrightarrow \begin{array}{l} \check{\mathbf{z}}_t = \check{\Phi}\mathbf{C}\mathbf{x}_t, \\ \tilde{\mathbf{z}}_t = \tilde{\Phi}_t\mathbf{C}\mathbf{x}_t \end{array}$$

$\check{\mathbf{z}}_t \in \mathbb{R}^{\check{M}}$

$\tilde{\mathbf{z}}_t \in \mathbb{R}^{\tilde{M}}$

$M = \check{M} + \tilde{M}$

1. State sequence estimation:
   1. Build Hankel Matrix
   2. Compute SVD
   3. Compute estimated state sequences

Given $\check{\mathbf{Z}} = \check{\mathbf{z}}_{1:T}$

$\downarrow$

$$\mathbf{H} = \begin{bmatrix} \check{\mathbf{z}}_1 & \check{\mathbf{z}}_2 & \cdots & \check{\mathbf{z}}_{T-L+1} \\ \check{\mathbf{z}}_2 & \check{\mathbf{z}}_3 & \cdots & \check{\mathbf{z}}_{T-L} \\ \vdots & \vdots & \ddots & \vdots \\ \check{\mathbf{z}}_L & \check{\mathbf{z}}_{L+1} & \cdots & \check{\mathbf{z}}_T \end{bmatrix}$$

$\downarrow$

$$\mathbf{H} = \mathbf{U}_d \mathbf{S}_d \mathbf{V}_d^T$$

$\downarrow$

$$\hat{\mathbf{X}} = \hat{\mathbf{x}}_{1:T-L+1} = \mathbf{S}_d \mathbf{V}_d^T$$

CS-LDS Architecture

2. Observation matrix estimation:
   - **C** is time-invariant
   - Given **Z** and $\widehat{\mathbf{X}}$, recover **C**

$$\min_{\mathbf{c}_i} \sum_{i=1}^{d} \left\|\Psi^T \mathbf{c}_i\right\|_1 \quad \text{s.t} \quad \forall t, \left\|\mathbf{z}_t - \Phi_t \mathbf{C} \hat{\mathbf{x}}_t\right\|_2 \leq \epsilon,$$

   - Ψ is sparsifying basis for the columns of **C**

Outline

- Background
- CS-LDS Architecture
- Estimating the state sequence
- Estimating the observation matrix
- Conclusion

Estimating the state sequence

QS#1: What are the sufficient conditions for reliable estimation?

Definition: (Observability of an LDS) An LDS is observable if the current state can be estimated from a finite number of observations (for any possible state sequence).

Lemma: Observable LDS$(\mathbf{A}, \mathbf{C}) \Leftrightarrow$ the observability matrix $\mathbf{O}(\mathbf{A}, \mathbf{C})$ is full rank.

Remark: $N \gg d \rightarrow$ LDS$(\mathbf{A}, \mathbf{C})$ is observable with high probability

Lemma: for $N > d$, the LDS$(\mathbf{A}, \breve{\Phi}\mathbf{C})$ is observable with high probability, if

- $\breve{M} \geq d$
- Entries of $\breve{\Phi}$ are i.i.d samples of a sub-Gaussian distribution.

Sum up: Then we can estimate state sequences by factorizing the block-Hankel matrix.

Estimating the state sequence

QS#2: How about $\breve{M} = 1$? (one common measurement for each video sequence)

Theorem: $\breve{M} = 1$ and the elements of $\breve{\Phi} \in \mathbb{R}^{1\times N}$ be i.i.d from a sub-Gaussian distribution. With high probability $\mathbf{O}(\mathbf{A}, \Phi\mathbf{C})$ is full rank if

- The state transition matrix is diagonalizable,
- Its eigenvalues and eigenvectors are unique.

QS#3: How about $\breve{M} < 1$? (missing measurements in some time instants)

- We obtain common measurements at some time instants $\mathcal{I} \subset \{1, \cdots, T\}$
- We have knowledge of $\check{\mathbf{z}}_i, i \in \mathcal{I}$
- Incomplete knowledge of the block-Hankel matrix

Matrix completion: $\min \operatorname{rank}(\mathbf{H}(\check{\mathbf{z}}_i))$ s.t. $i \in \mathcal{I}$

- Non-convex

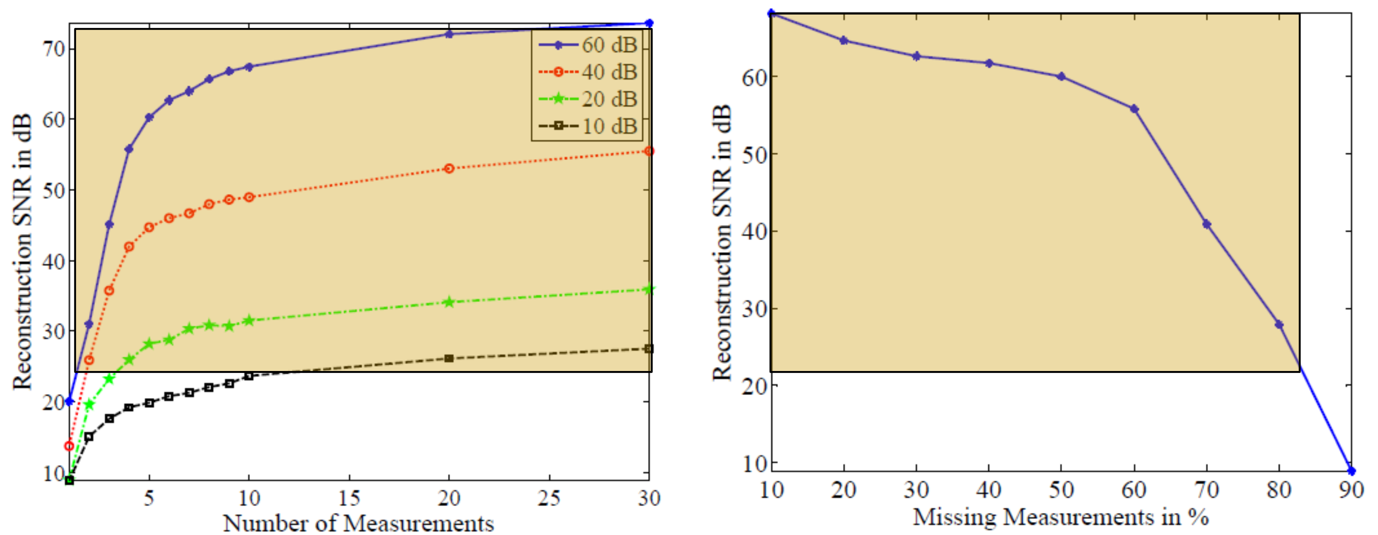
Solution: (Nuclear norm) $\min \|\mathbf{H}(\check{\mathbf{z}}_i)\|_*$ s.t. $i \in \mathcal{I}$

Estimating the state sequence

Accuracy of state sequence estimation from common measurements

- $T = 500, d = 10$
- Reconstruction SNR = $10\log_{10}\left(\frac{\sum_{t=1}^{T}\|\mathbf{y}_t\|_2^2}{\sum_{t=1}^{T}\|\mathbf{y}_t - \widehat{\mathbf{y}}_t\|_2^2}\right)$

Frobenius norm



(a) $\breve{M} \geqslant 1$ (b) $\breve{M} < 1$

Outline

- Background
- CS-LDS Architecture
- Estimating the state sequence
- <u>Estimating the observation matrix</u>
- Conclusion

Estimating the observation matrix

- Images are sparse in some domains like Wavelet and DCT.
- Smooth changes in sequential frames
  - The motion is spatially correlated.
  - The supports of frames are highly overlapping.
  - The columns of **C** captures dominant motion patterns.
  - **C** can be interpreted as a basis for the frames of the video.
  - The columns of **C** are sparse in the same domain.

$$\min_{\mathbf{c}_i} \sum_{i=1}^{d} \left\| \Psi^T \mathbf{c}_i \right\|_1 \quad \text{s.t} \quad \forall t, \left\| \mathbf{z}_t - \Phi_t \mathbf{C} \hat{\mathbf{x}}_t \right\|_2 \leq \epsilon,$$

  - Insufficient for recovering **C**
    - $\hat{\mathbf{x}}_t \approx \mathbf{L}^{-1}\mathbf{x}_t$

LDS $(\mathbf{A}, \mathbf{C}, \mathbf{x}) \equiv$ LDS $(\mathbf{L}^{-1}\mathbf{A}\mathbf{L}, \mathbf{C}\mathbf{L}, \mathbf{L}^{-1}\mathbf{x})$

For any invertible matrix $\mathbf{L} \in \mathbb{R}^{d \times d}$

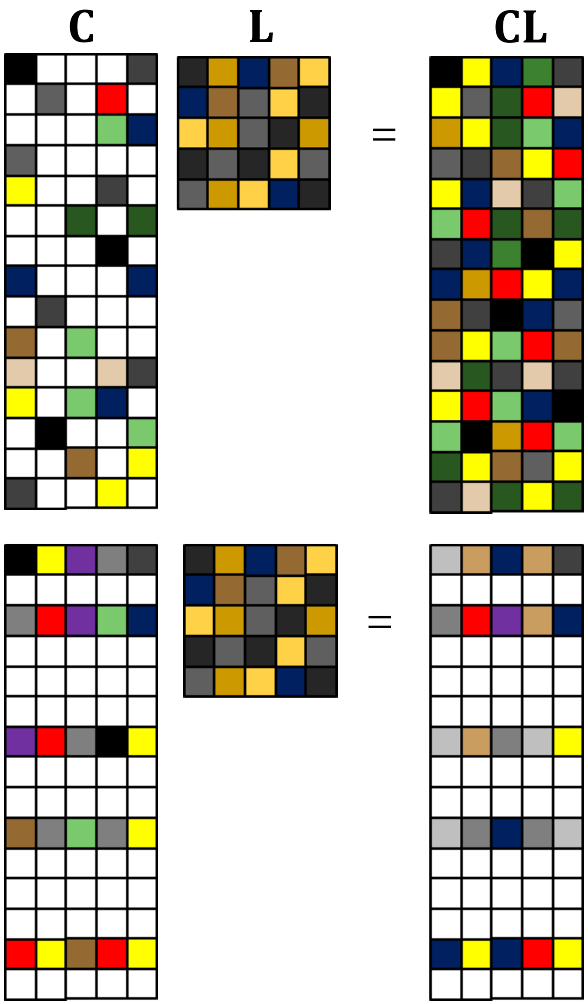
Estimating the observation matrix

- suppose **C** is canonical sparse: $\Psi = \mathbf{I}$ (wlog)
- *Worst case:* **disjoint** sparsity pattern
- *Best case:* **same** sparsity pattern
- Recovering **C** using column group sparsity

$$(P_{\ell_2-\ell_1}) \ \min \sum_{i=1}^{N} \|\mathbf{s}_i\|_2 \quad \text{s.t } C = \Psi S,\ \forall t, \|\mathbf{z}_t - \Phi_t C \hat{\mathbf{x}}_t\|_2 \leqslant \epsilon,$$

- *Solver*: Model-based CoSaMP
- Value of $\widetilde{M}$:

$$\widetilde{M}T = 4dK\log(N/K) \implies \widetilde{M} = 4\frac{dK}{T}\log(N/K)$$



Estimating the observation matrix

Model-based CoSaMP

Algorithm 1: $\widehat{C}$ = Model-based CoSAMP $(\Psi, K, \mathbf{z}_t, \widehat{\mathbf{x}}_t, \Phi_t, t = 1, \ldots, T)$

Notation:

supp$(vec; K)$ returns the support of K largest elements of vec

$A_{|\Omega,\cdot}$ represents the submatrix of A with rows indexed by Ω and all columns.

$A_{|\cdot,\Omega}$ represents the submatrix of A with columns indexed by Ω and all rows.

Initialization

$\forall t, \Theta_t \leftarrow \Phi_t \Psi$

$\forall t, \mathbf{v}_t \leftarrow \mathbf{0} \in \mathbb{R}^M$

$\Omega_{\text{old}} \leftarrow \phi$

while *(stopping conditions are not met)* **do**

Compute signal proxy:

$R = \sum_t \Theta_t^T \mathbf{v}_t \widehat{\mathbf{x}}_t^T$

Compute energy in each row:

$k \in [1, \ldots, N], \mathbf{r}(k) = \sum_{i=1}^{d} R^2(k, i)$

Support identification and merger:

$\Omega \leftarrow \Omega_{\text{old}} \bigcup \text{supp}(\mathbf{r}; 2K)$

Least squares estimation:

Find $A \in \mathbb{R}^{|\Omega| \times d}$ that minimizes $\sum_t \|\mathbf{z}_t - (\Theta_t)_{|\cdot,\Omega} A \widehat{\mathbf{x}}_t\|_2$

$B_{|\Omega,\cdot} \leftarrow A, \; B_{|\Omega^c,\cdot} \leftarrow 0$

Pruning support:

$k \in [1, \ldots, N], \mathbf{b}(k) = \sum_{i=1}^{d} B^2(k, i)$

$\Omega \leftarrow \text{supp}(\mathbf{b}; K), \; S_{|\Omega,\cdot} \leftarrow B_{|\Omega,\cdot}, \; S_{|\Omega^c,\cdot} \leftarrow 0$

Form new estimate of C:

$\widehat{C} \leftarrow \Psi S$

Update residue:

$\forall t, \mathbf{v}_t \leftarrow \mathbf{z}_t - \Theta_t S \widehat{\mathbf{x}}_t$

$\Omega_{\text{old}} \leftarrow \Omega$

end

Group sparsity

2

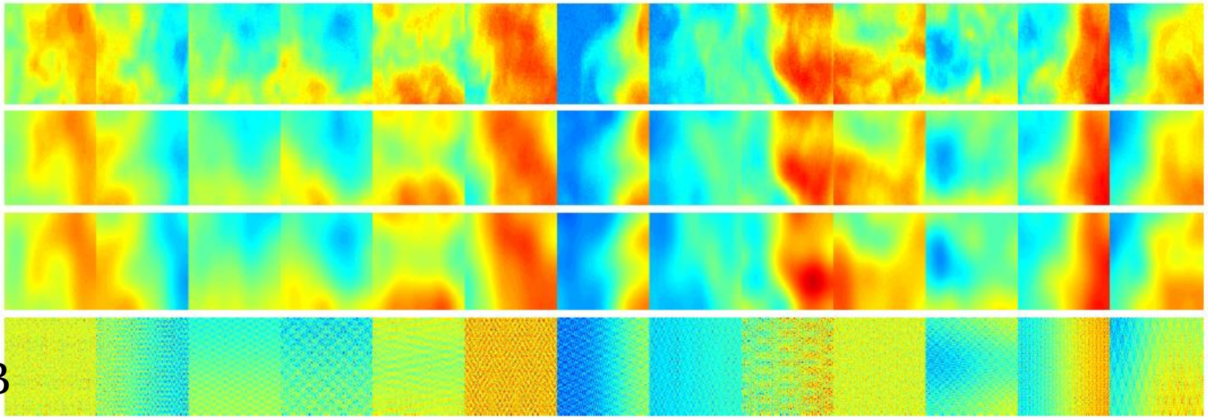
Estimating the observation matrix

Ground truth

Oracle LDS: 24.97 dB

CS-LDS: 22.08 dB

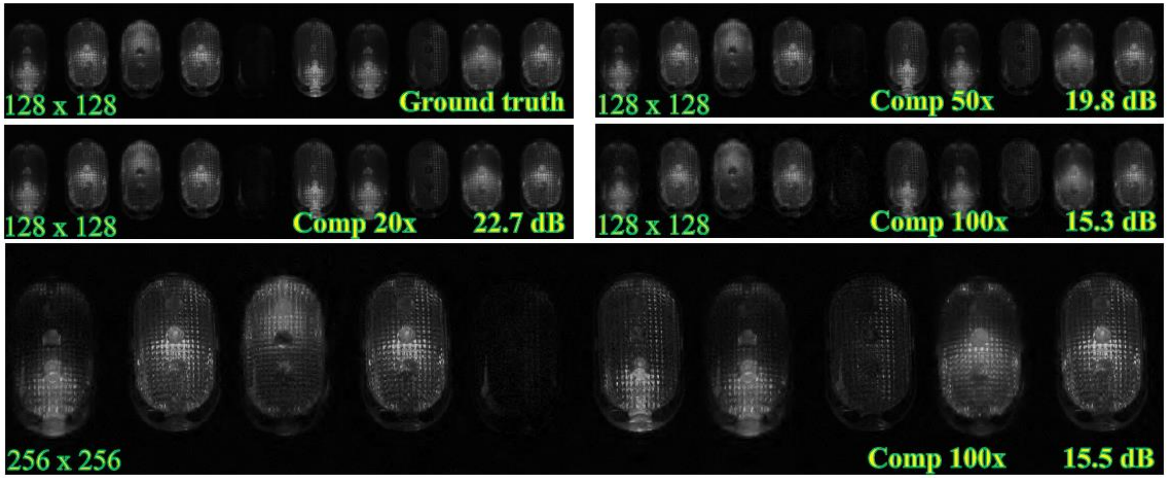
Frame-to-Frame CS: 11.75 dB



$\frac{N}{M} = 234$ for all methods

Oracle LDS:
No CS (Nyquist sampling) + knowledge of d

Sparsity: DCT, Wavelet
Meas.: Noiselet, Gaussian



Outline

- Background
- CS-LDS Architecture
- Estimating the state sequence
- Estimating the observation matrix
- Conclusion

Conclusion

- Not efficient to use conventional CS for video sequences
  - Ephemeral nature
  - High-dimensional
- Model video sequences as
  - Low-dimensional dynamic parameters (the state sequences)
  - High-dimensional static parameters (the observation matrix)
- Solution included
  - SVD
  - Convex optimization

Main References

[1] A. C. Sankaranarayanan, P. K. Turaga, R. Chellappa, and R. G. Baraniuk, "Compressive acquisition of linear dynamical systems," *SIAM Journal on Imaging Sciences* 6, No. 4, pp. 2109-2133, 2013.

[2] A. C. Sankaranarayanan, P. K. Turaga, K. Pavan, R. G. Baraniuk, and R. Chellappa, "Compressive acquisition of dynamic scenes," in *European Conference on Computer Vision*, Sep. 2010.

[3] [Online]: CS-LDS, www.ece.rice.edu/~as48/research/cslds.

[4] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on Information Theory*, Vol. 56, No. 4, pp. 1982-2001, 2010.

[5] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, Vol. 26, No. 23, pp. 301-323, 2009.

Thanks for Your Attention.

Any Question?