Online Subspace Estimation and Tracking from Incomplete and Corrupted Data

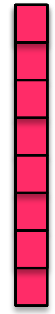
THE UNIVERSITY of WISCONSIN MADISON

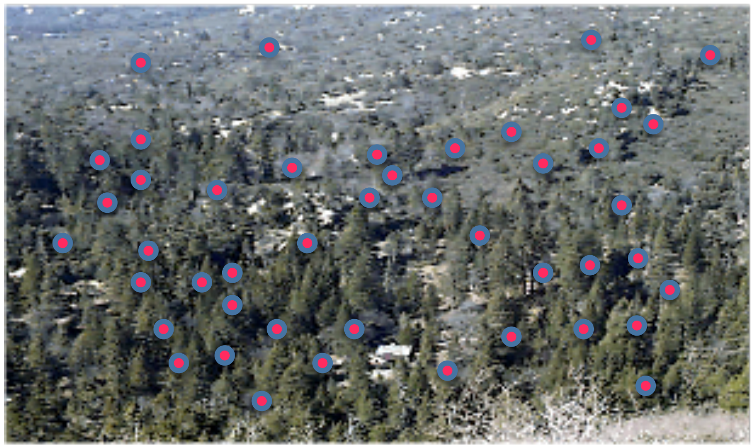
Laura Balzano
sunbeam@ece.wisc.edu

work with **Robert Nowak, Benjamin Recht,**
Jun He (Nanjing UIST),
and **Arthur Szlam (CUNY)**

Subspace Representations

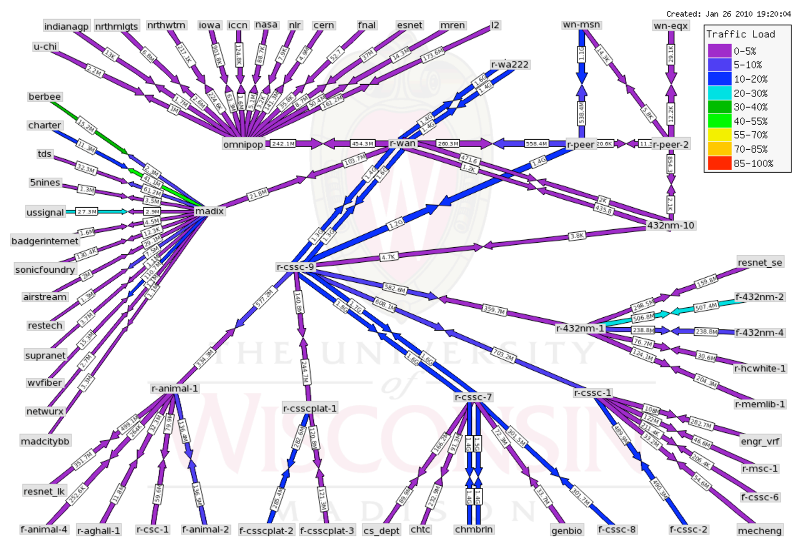
Monitor/sense with n nodes





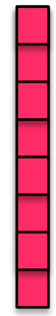
$v \in \mathbb{R}^n$ is a snapshot of the system state (e.g., temperature at each node)

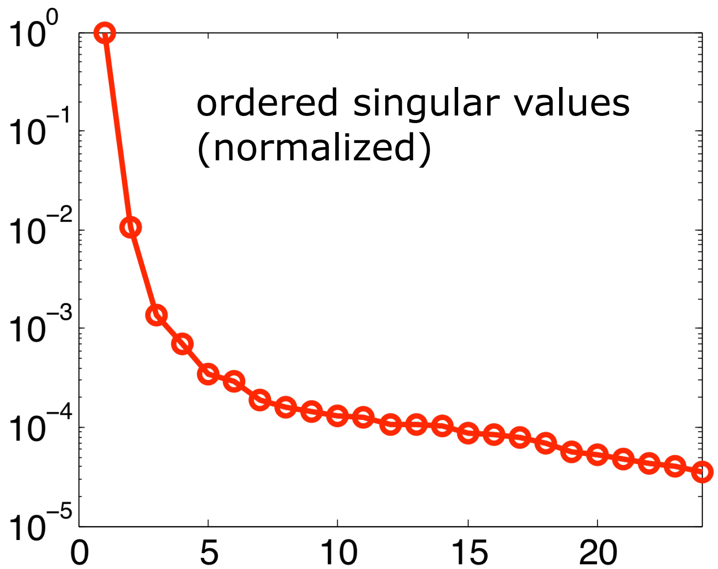
$v \in \mathbb{R}^n$ is a snapshot of the system state (e.g., traffic rates at each monitor)



Subspace Representations

Monitor/sense with n nodes

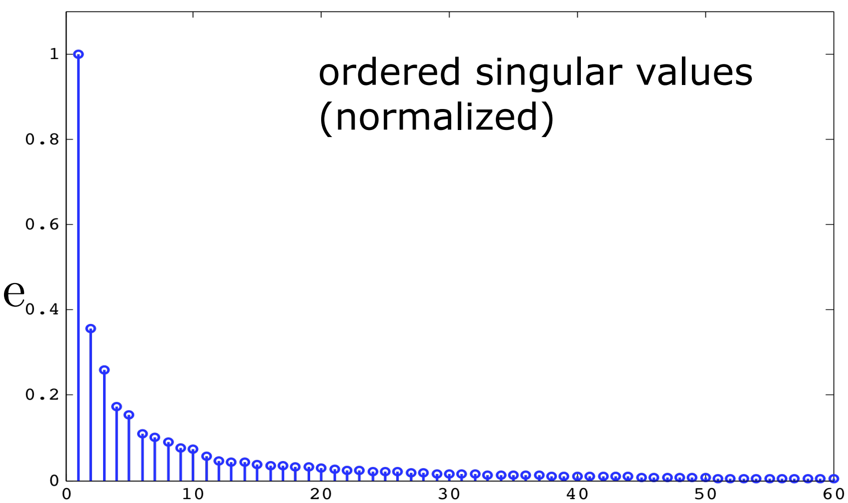




Temperature data from UCLA Sensornet

$v \in \mathbb{R}^n$ is a snapshot of the system state (e.g., temperature at each node)

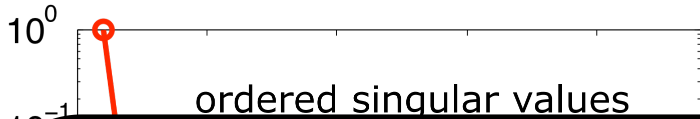
$v \in \mathbb{R}^n$ is a snapshot of the system state (e.g., traffic rates at each monitor)



Byte Count data from UW network

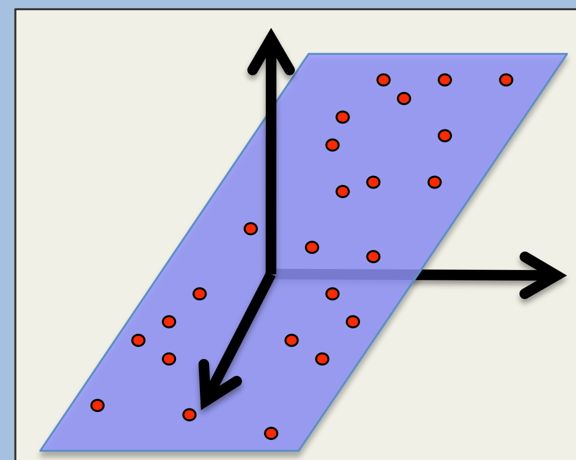
Subspace Representations

Monitor/sense with n nodes



Each snapshot lies near a low-dimensional subspace

$$S \subset \mathbb{R}^n$$



$v \in \mathbb{R}^n$ is a snapshot of the system state

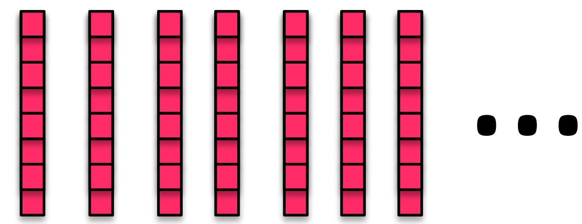
Using the **subspace as a model** for the data, we can leverage these dependencies for detection, estimation and prediction.

Subspace Representations: Imaging

- For each frame we have n pixels.
- The background of a collection of frames lies in a low-dimensional subspace, possibly time-varying.

Subspace Identification: Introduction

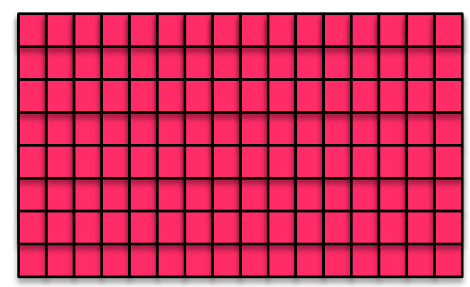
Suppose we receive a sequence of length-n vectors that lie in a d-dimensional subspace S:



$$v_1, v_2, \ldots, v_t, \ldots, \in S \subset \mathbb{R}^n$$

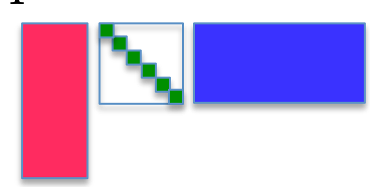
And then we collect T of these vectors into a matrix,

$$X = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \ldots & v_T \\ | & | & & | \end{bmatrix}$$



If S is static, we can identify it as the column space of this matrix by performing the SVD:

$$X = U \Sigma V^T \ .$$

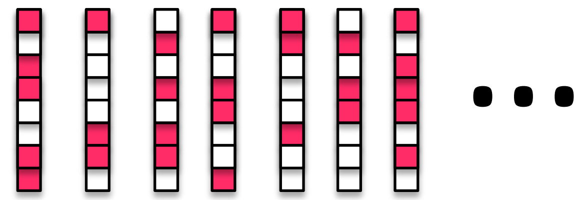


The orthogonal columns of U span the subspace S.

Subspace Identification: Introduction

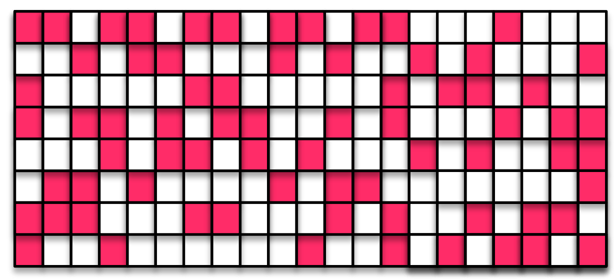
Suppose we receive a sequence of incomplete length-n vectors that lie in a d-dimensional subspace S, and $\Omega_t \subset \{1, \ldots, n\}$ refers to the observed indices:

$$v_{\Omega_1}, v_{\Omega_2}, \ldots, v_{\Omega_t}, \ldots, \in S \subset \mathbb{R}^n$$



And then we collect T of these vectors into a matrix:

$$X = \begin{bmatrix} | & | & & | \\ v_{\Omega_1} & v_{\Omega_2} & \ldots & v_{\Omega_T} \\ | & | & & | \end{bmatrix}$$



~~If S is static, we can identify it as the column space of this matrix by performing the SVD:~~

$$X = U \Sigma V^T \; .$$

The orthogonal columns of U span the subspace S.

Related Work: LMS subspace tracking

- Subspace tracking (with complete data) was approached with LMS methods in the 80s and 90s
  - Yang 1995, Projection Approximation Subspace Tracking; proof Delmas Cardoso 1998
  - Comon, Golub survey 1990

$$\|v - P_S v\|_2^2$$

- Incremental gradient methods are getting attention for their speed and convergence guarantees
  - Bertsekas, Tsitsiklis 2000

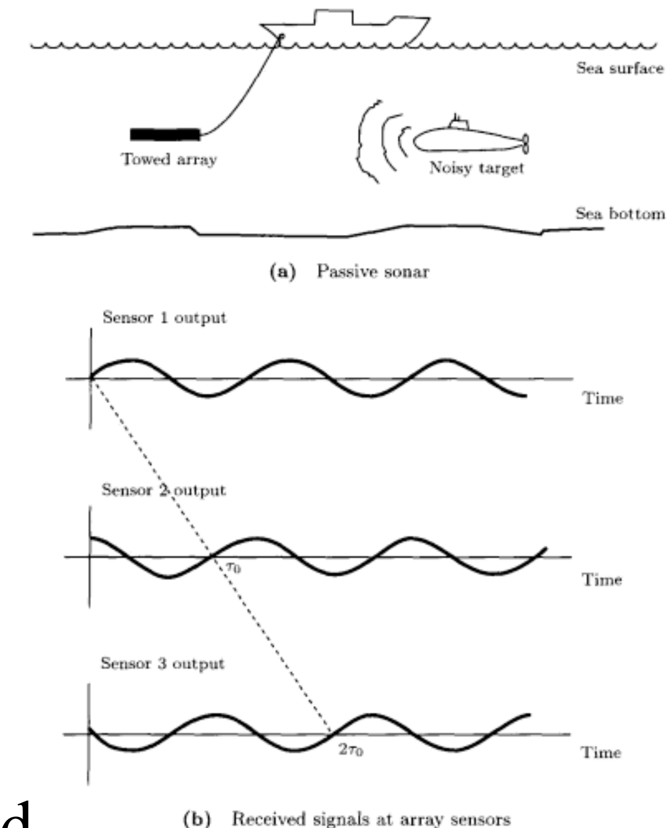


Figure 1.2 Passive sonar system

Figure from Stephen Kay, Fundamentals of Statistical Signal Processing Volume I: Estimation, p3.

Residual with Incomplete Data

U is an $n \times d$ orthogonal matrix whose columns span the d-dimensional subspace S.
U_Ω denotes the submatrix with rows indicated by Ω, where $\Omega \subset \{1, \ldots, n\}$ is the subset of indices observed.

Full-data Residual

$$P_S = U(U^T U)^{-1} U^T:$$

$$v_\perp = v - P_S v$$

Incomplete-data Residual

$$\text{Let } P_{S_\Omega} = U_\Omega \left(U_\Omega^T U_\Omega\right)^{-1} U_\Omega^T.$$

$$v_\perp = v_\Omega - P_{S_\Omega} v_\Omega$$

Theorem: Incomplete Data Residual Norm

S is a known $d < n$ dimensional subspace of $\mathbb{R}^n$ with coherence $\mu(S)$.

v_Ω is our observation and we wish to estimate $\|v_\perp\|_2^2 = \|v_\Omega - P_{S_\Omega} v_\Omega\|_2^2$

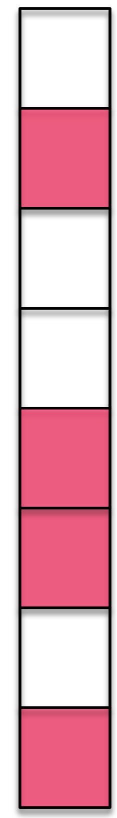
Theorem: If $|\Omega| = O(\mu(S) d \log d)$ and Ω is chosen uniformly with replacement, then with high probability and ignoring constant factors,

$$\frac{|\Omega| - d\mu(S)}{n} \|v - P_S v\|_2^2 \leq \|v_\Omega - P_{S_\Omega} v_\Omega\|_2^2 \leq \frac{|\Omega|}{n} \|v - P_S v\|_2^2$$

Subspace Tracking

Suppose we receive a sequence of incomplete vectors that lie in a d-dimensional subspace S:

$$v_{\Omega_1}, v_{\Omega_2}, \ldots, v_{\Omega_t}, \ldots$$



Given S_t and v_{Ω_t}, how do we generate S_{t+1}?

Choose S_{t+1} to decrease the error $\|v_\Omega - P_{S_\Omega} v_\Omega\|_2^2$.

GROUSE

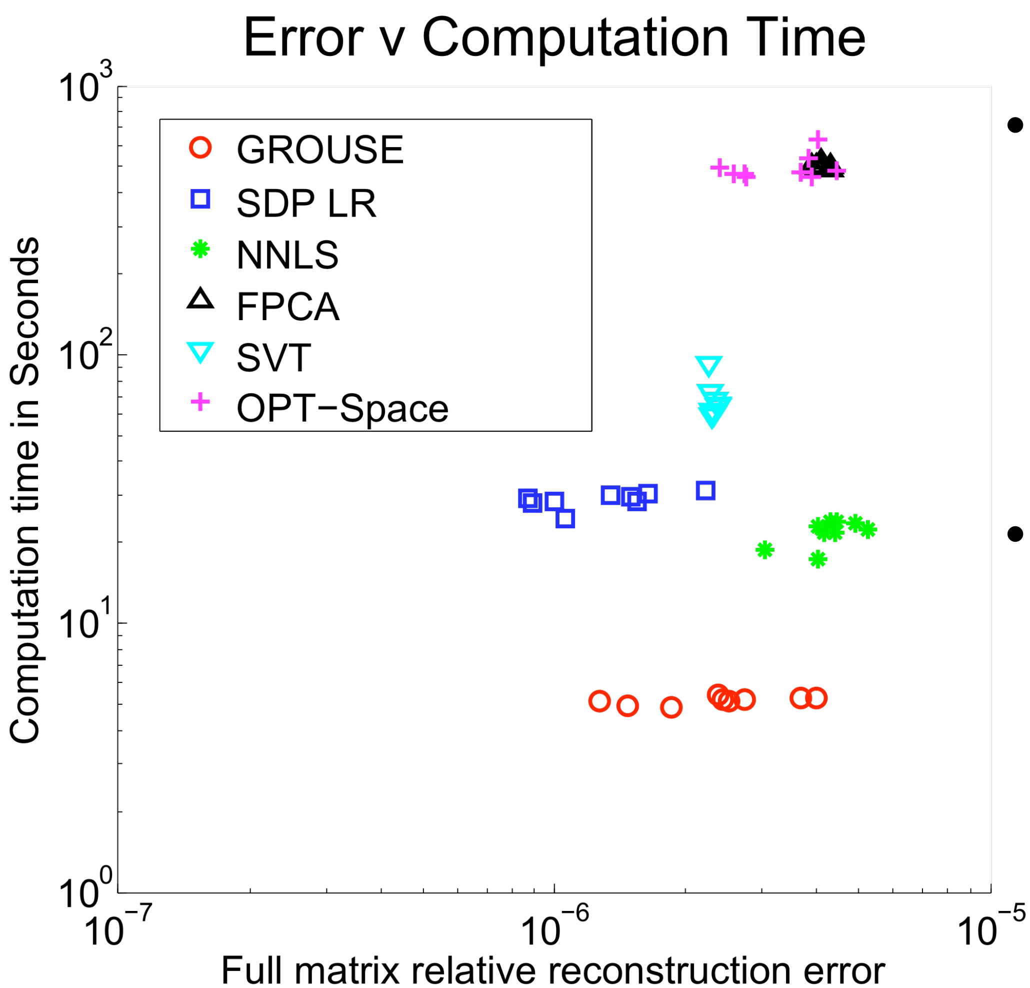
- Given step size η_t, subspace basis $U_t \in \mathbb{R}^{n\times d}$, observations v_{Ω_t}
- Calculate Weights:
  $w = \arg\min_a \|U_{\Omega_t} a - v_{\Omega_t}\|_2^2$
- Predict full vector: $v_{\|} = U_t w$
- Compute Residual on observed entries: $v_\perp = v_{\Omega_t} - (v_{\|})_{\Omega_t}$ and zero-pad.
- Update subspace:

$$U_{t+1} = U_t + \left(\sin(\sigma\eta_t)\frac{v_\perp}{\|v_\perp\|} + (\cos(\sigma\eta_t) - 1)\frac{v_{\|}}{\|v_{\|}\|}\right)\frac{w^T}{\|w\|}$$

where $\sigma = \|v_\perp\|\|v_{\|}\|$

- One iteration involves a projection and an outer product.
- The algorithm is *simple* and *fast*.

GROUSE Performance: Matrix Completion



- To use GROUSE, we pass over the columns of the matrix a few times in random order, doing an update for every column.
- We compared against other state of the art MC algorithms on reconstruction error and computation time.

Robust Low-Rank Modeling (Robust PCA)

- Sparse + Low-Rank Model

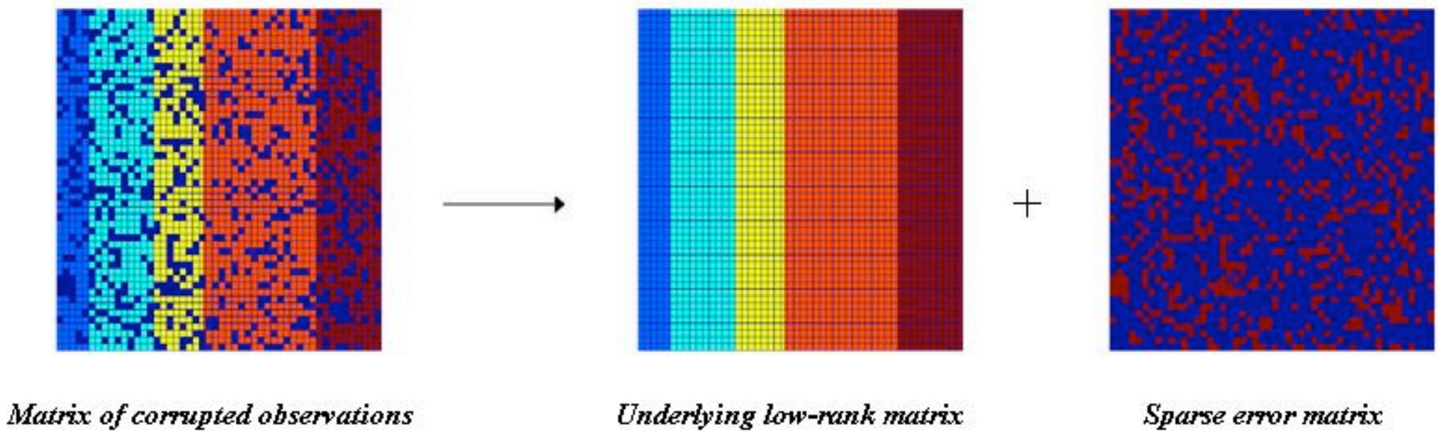


image from Nuit Blanche

- Several algorithms have been developed to find such a decomposition from a matrix observation
  - convex optimization and approximations

GROUSE to GRASTA with Jun He and Arthur Szlam

$$F_{grouse}(\mathcal{S};t) = \min_w \|U_{\Omega_t} w - v_{\Omega_t}\|_2^2$$

$$F_{grasta}(\mathcal{S};t) = \min_w \|U_{\Omega_t} w - v_{\Omega_t}\|_1$$

$$U_{t+1} = U_t + \left((\cos(\sigma\eta_t) - 1) U_t \frac{w_t}{\|w_t\|} + \sin(\sigma\eta_t) \frac{\Gamma}{\|\Gamma\|} \right) \frac{w_t^T}{\|w_t\|}$$

GRASTA Performance on RPCA
with Jun He

- Simulated 2000 x 2000 matrix, rank=5
- Compare GRASTA to Inexact Augmented Lagrange Multiplier Method for RPCA

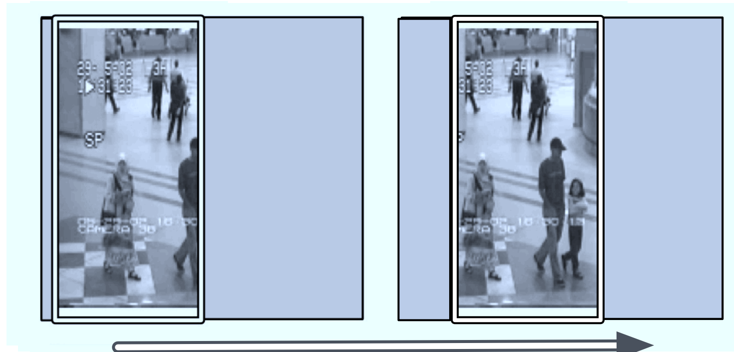
% outliers	noise var	GRASTA 100% sampled	GRASTA 30% sampled	IALM FOR RPCA
10%	0.5e-3	3.64e-4 / 58.31 sec	6.07e-4 / 20.79 sec	16.7e-4 / 93.16 sec
10%	1.0e-3	7.643-4 / 59.55 sec	12.3e-4 / 20.66 sec	36.4e-4 / 117.76 sec
30%	0.5e-3	6.13e-4 / 67.19 sec	12.6e-4 / 22.63 sec	26.4e-4 / 324.26 sec
30%	1.0e-3	9.87e-4 / 69.06 sec	19.3e-4 / 22.85 sec	56.2e-4 / 362.62 sec

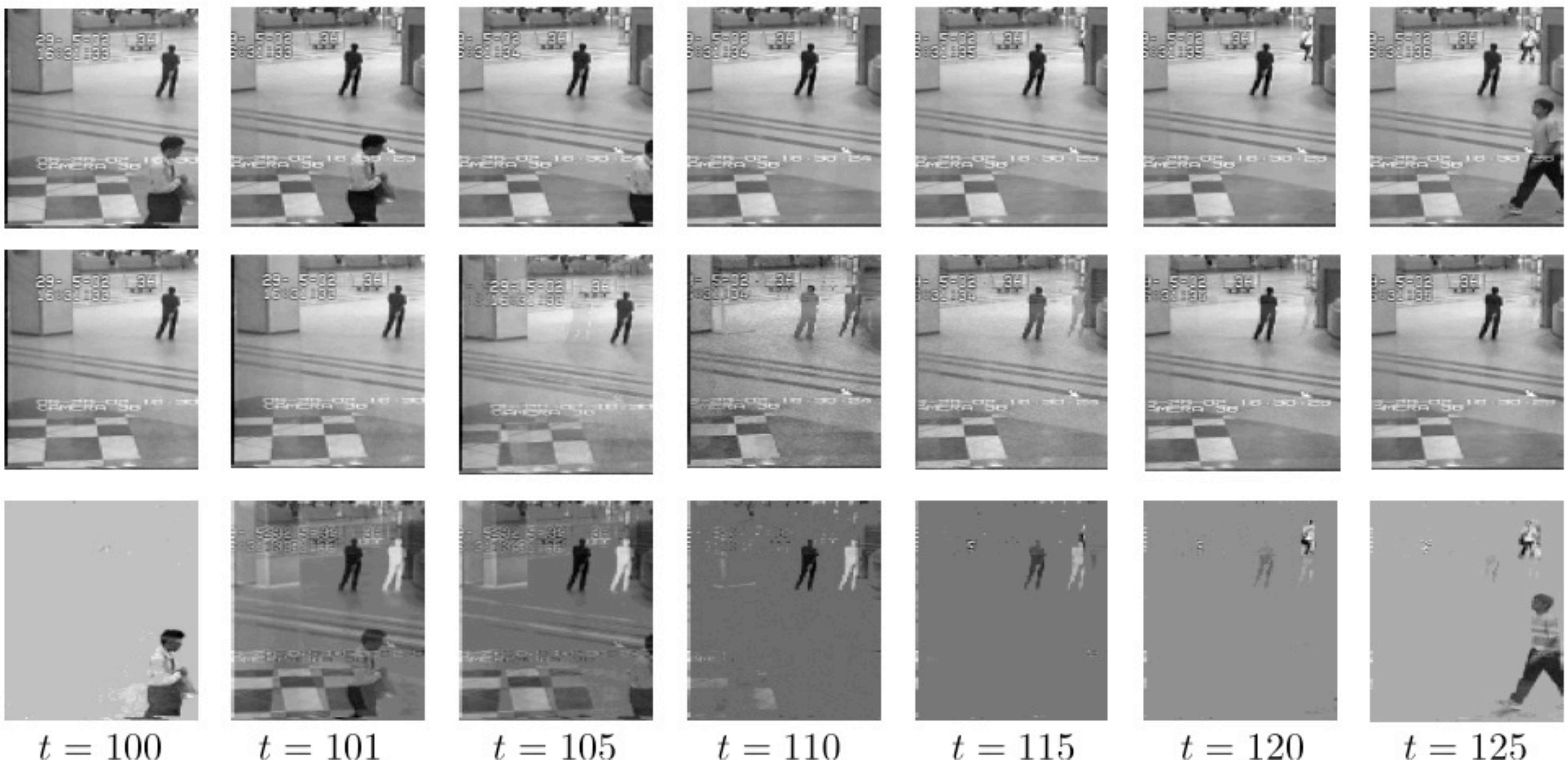
GRASTA Performance on Background Subtraction
with Jun He

$t = 1$ $t = 600$ $t = 1200$

Dataset	Resolution	Total Frames	Training Time	Tracking and Separating Time	FPS
Airport Hall	144×176	3584	11.3 sec	20.9 sec	171.5
Shopping Mall	320×256	1286	33.9 sec	27.5 sec	46.8
Lobby	144×176	1546	3.9 sec	71.3 sec	21.7
Hall with Virtual Pan (1)	144×88	3584	3.8 sec	191.3 sec	18.7
Hall with Virtual Pan (2)	144×88	3584	3.7 sec	144.8 sec	24.8

GRASTA Performance on Background Subtraction
with Jun He





GRASTA demo

- Written by Arthur Szlam at CUNY in Open CV

For more information: http://sunbeam.ece.wisc.edu/

Incremental Gradient on the Grassmannian for Background and Foreground Separation in Subsampled Video
Jun He, Laura Balzano, and Arthur Szlam. To appear at CVPR, June 2012.

Online Robust Subspace Tracking from Partial Information
Jun He, Laura Balzano, Arthur Szlam, and John C.S. Lui. In preparation for Trans PAMI.
http://arxiv.org/abs/1109.3827

Online Identification and Tracking of Subspaces from Highly Incomplete Information
Laura Balzano, Robert Nowak, and Benjamin Recht. Allerton, September 2010.

High-Dimensional Matched Subspace Detection when Data are Missing
Laura Balzano, Benjamin Recht, and Robert Nowak. ISIT, June 2010.

THANK YOU!
Questions?

Theorem: Incomplete Data Residual Norm

Full Theorem: If $|\Omega| \geq \frac{8}{3}\mu(S) d \log(2d/\delta)$ and Ω is chosen uniformly with replacement, then with probability $1-4\delta$,

$$\frac{|\Omega|(1-\alpha) - d\mu(S)\frac{(1+\beta)^2}{(1-\gamma)}}{n}\|v - P_S v\|_2^2 \leq \|v_\Omega - P_{S_\Omega} v_\Omega\|_2^2 \leq (1+\alpha)\frac{|\Omega|}{n}\|v - P_S v\|_2^2$$

where we write $v = x + y,\ \ x \in S,\ \ y \in S_\perp$,

$\alpha = \sqrt{\frac{2\mu(y)^2}{|\Omega|}\log\left(\frac{1}{\delta}\right)}$, $\beta = \sqrt{2\mu(y)\log\left(\frac{1}{\delta}\right)}$, and $\gamma = \sqrt{\frac{8d\mu(S)}{3|\Omega|}\log\left(\frac{2d}{\delta}\right)}$.

Lemma 1: $\|y_\Omega\|_2^2 \geq (1-\alpha)\frac{|\Omega|}{n}\|y\|_2^2$ — McDiarmid's inequality with sum of RVs

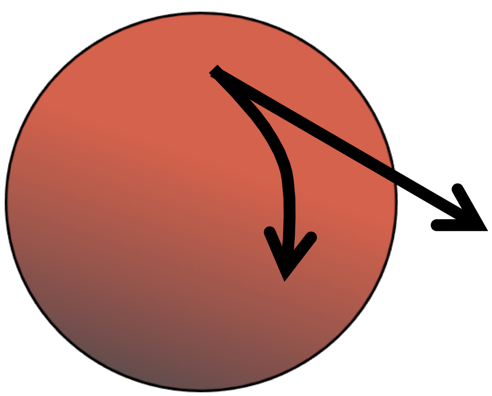
Lemma 2: $\|U_\Omega^T y_\Omega\|_2^2 \leq (1+\beta)^2 \frac{|\Omega|}{n}\frac{d\mu(S)}{n}\|y\|_2^2$ — McDiarmid's inequality.

Lemma 3: $\|\left(U_\Omega^T U_\Omega\right)^{-1}\|_2 \leq \frac{n}{(1-\gamma)|\Omega|}$ — Non-commutative Bernstein inequality

Descent on the Grassmannian

- Idea: Stochastic gradient descent to minimize the incomplete project residual one vector at a time. (Subspace Tracking in Signal Processing is done this way, using the complete-data residual.)
- Since we are estimating a subspace, we can perform gradient descent directly on the Grassman manifold G(n,d) and follow its geodesics.
    - (There are explicit formulas for a gradient descent step that follows the Grassmannian geodesic.)

A. Edelman, T. A. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.



1-d subspaces in R^{2}: