Dictionary-Based Generalization of Robust PCA

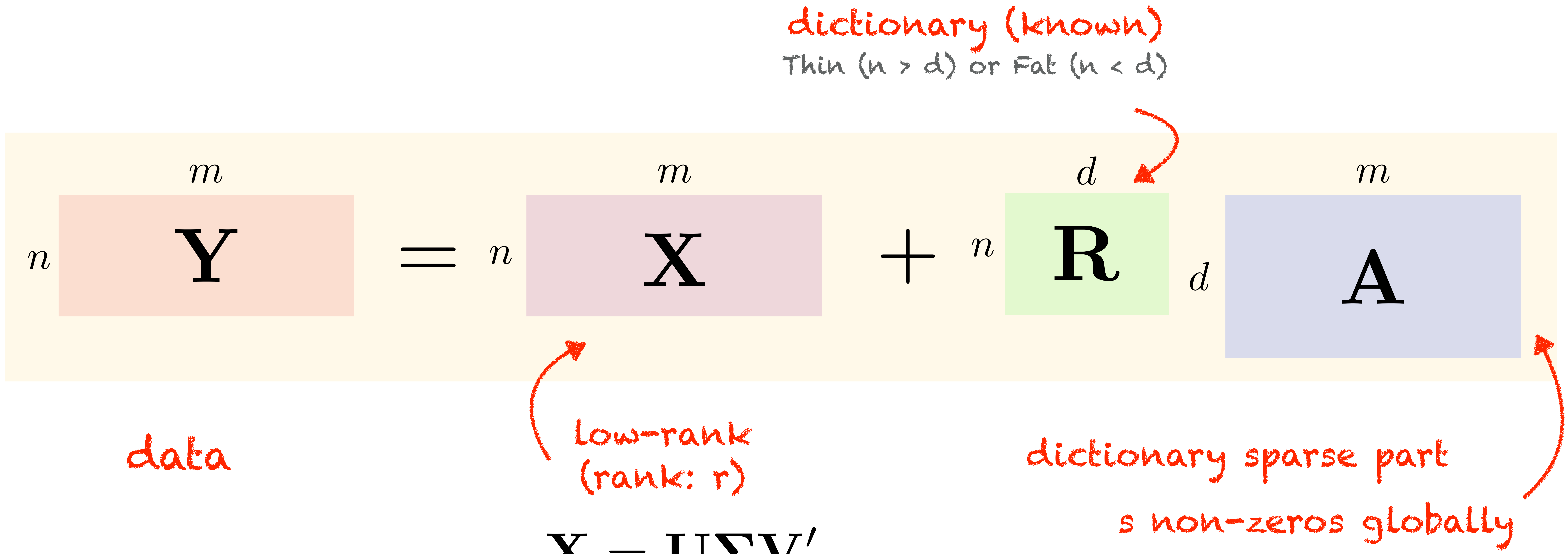
Sirisha Rambhatla, *Xingguo Li, and Jarvis Haupt*

University of Minnesota

Driven to Discover[SM]

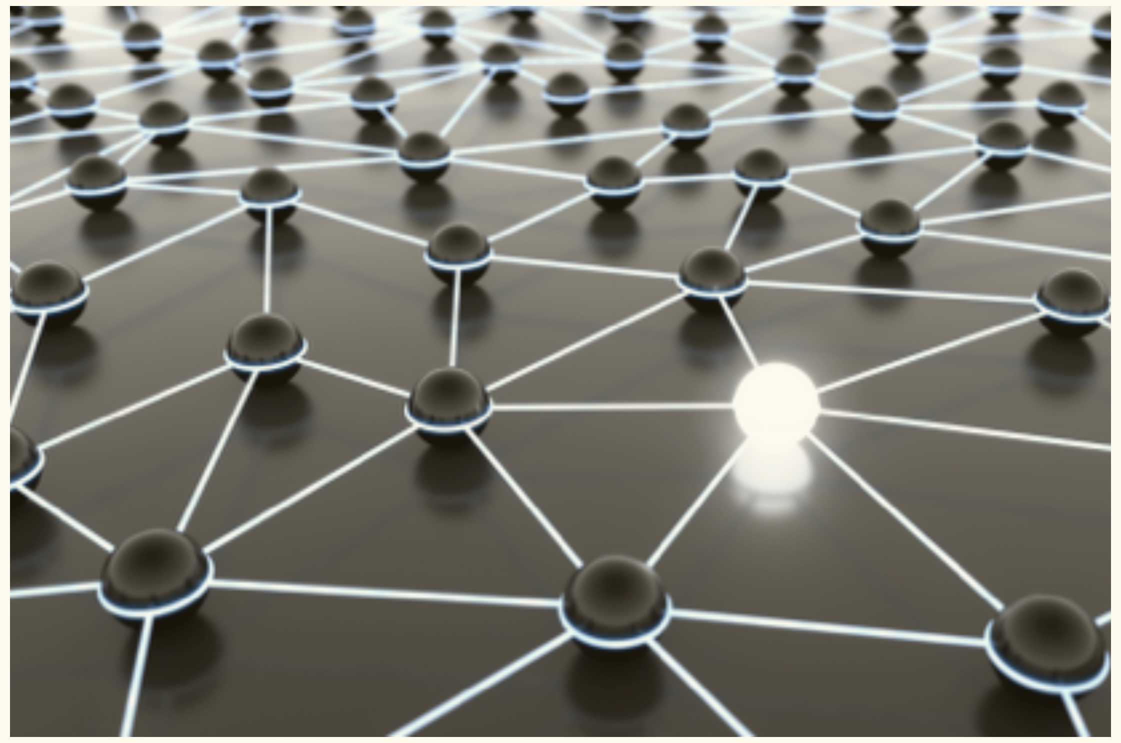
December 2016

What is this work about?



$$\mathbf{X} = \mathbf{U\Sigma V}'$$

Applications



Traffic Anomalies

Mardani, Mateos and Giannakis, 2013

Fat dictionary

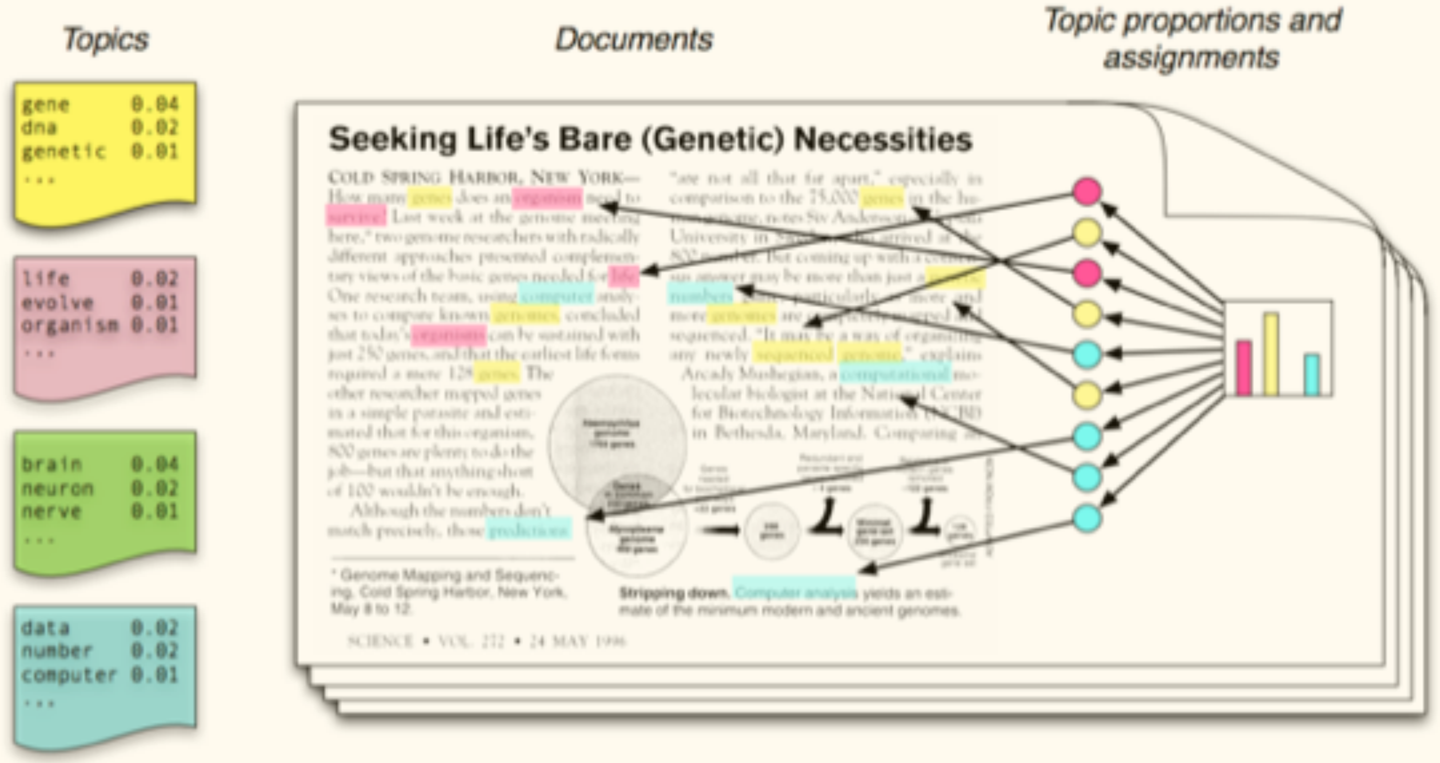
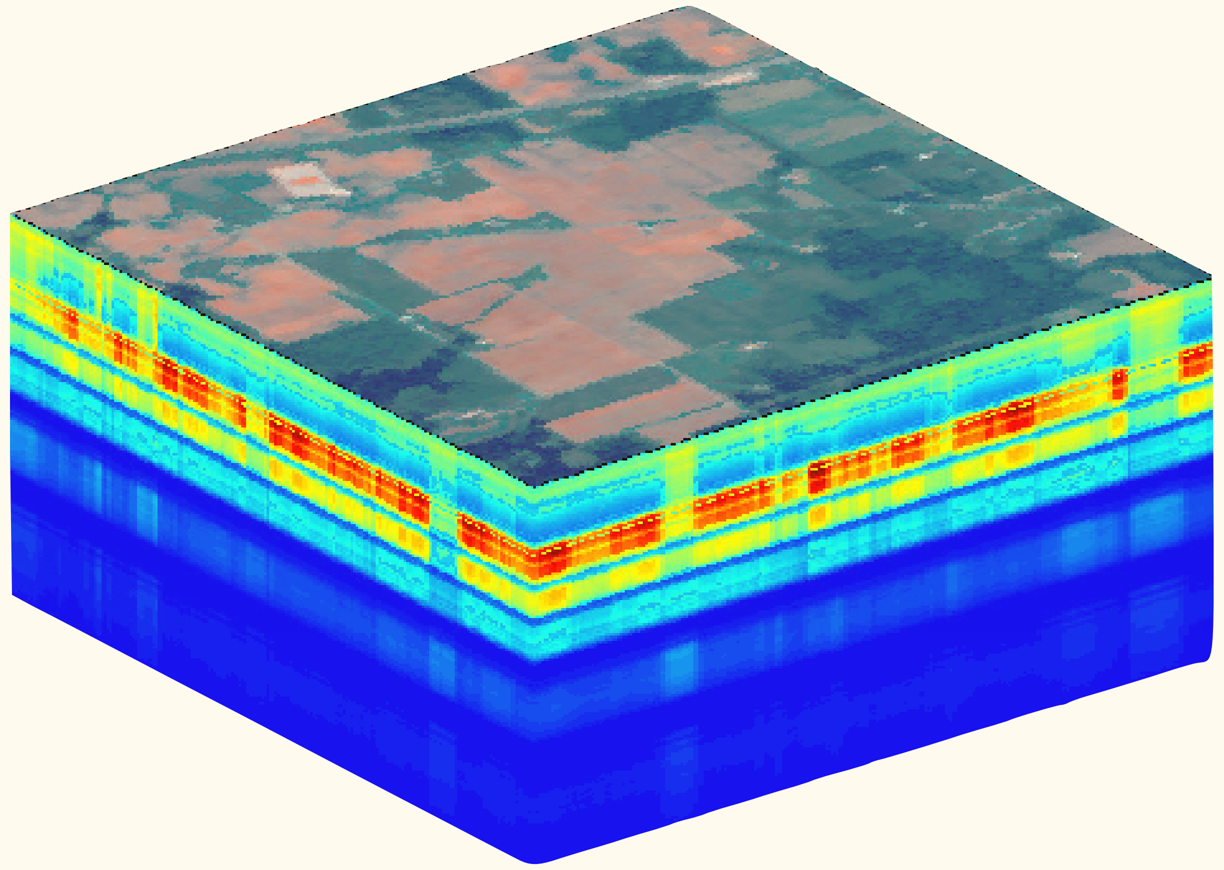


Figure source: Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77-84.

Topic Modeling

Thin dictionary

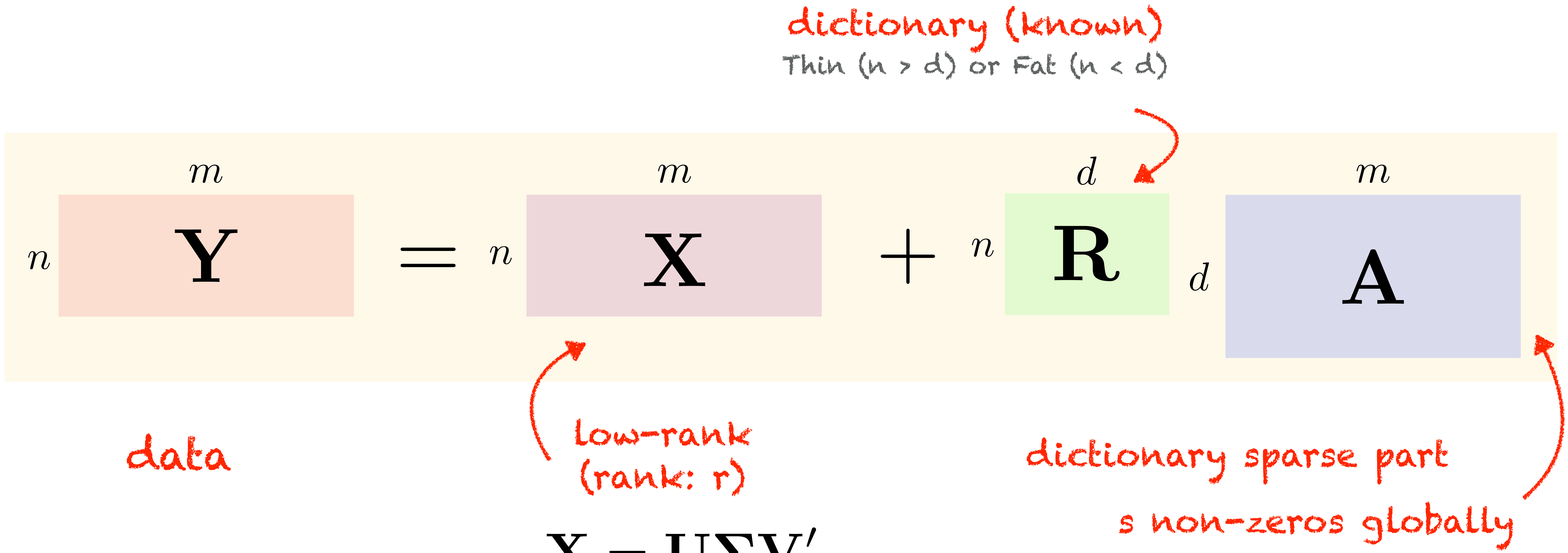


Hyper-Spectral Imaging

Thin dictionary

(More on this later)

Recall



$$\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}'$$

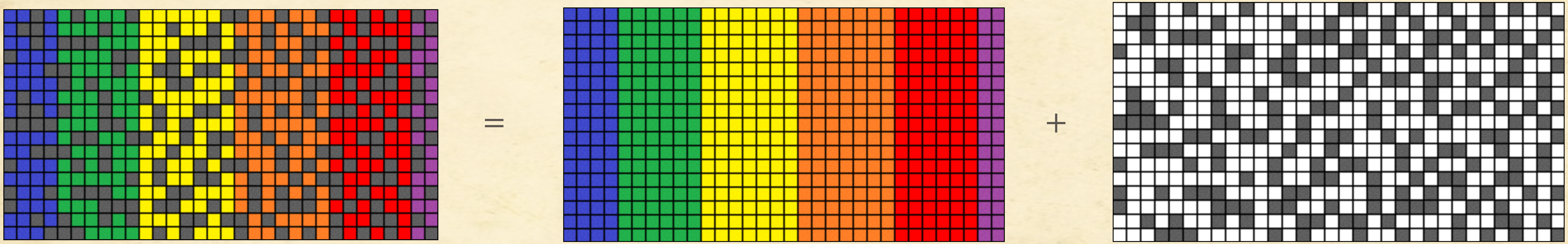
Optimization Problem

$$\underset{\mathbf{X},\mathbf{A}}{\text{minimize}} \quad \|\mathbf{X}\|_* + \lambda\|\mathbf{A}\|_1 \text{ s.t. } \mathbf{Y} = \mathbf{X_0} + \mathbf{R}\mathbf{A_0} \quad (1)$$

$\|.\|_* =$ nuclear norm and $\|.\|_1 = l_1-$norm of the vectorized matrix.

Prior Art

PCA



Centered Data
Y

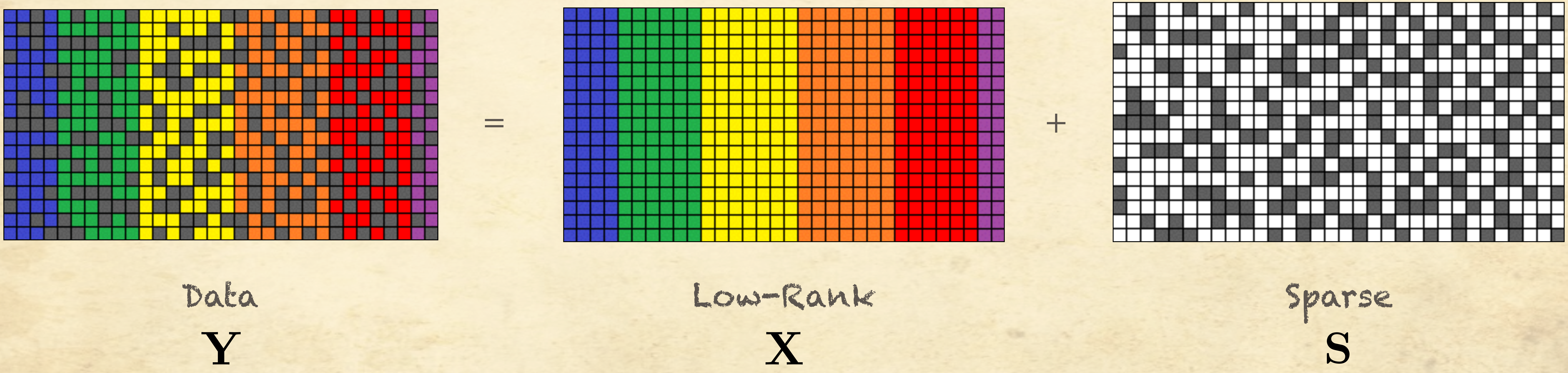
Low-Rank Approximation
X

Gaussian Noise
N

Pearson, 1901

Prior Art

Robust PCA



Candès, Li, Ma, and Wright, 2009

Chandrasekaran, Sanghavi, Parrilo, and Willsky, 2011

and many more..

Outlier Pursuit

Xu, Caramanis, and Sanghavi, 2010

Prior Art

Low-rank Plus Dictionary Sparse Decomposition

Mardani, Mateos, and Giannakis, 2013

Prior Art

Low-rank Plus Dictionary Sparse Decomposition

Mardani, Mateos, and Giannakis, 2013

Prior Art

Our Contributions

1

Establish recovery results for the
Thin dictionary case
with constraints on the global sparsity of A

2

Establish recovery results for the
Fat dictionary case
with constraints on the global sparsity of A
and k non-zeros per column

Remove orthogonality constraint on the rows of the dictionary

Our Contributions

1

Establish recovery results for the
Thin dictionary case
with constraints on the global
sparsity of A

2

Establish recovery results for the
Fat dictionary case
with constraints on the global sparsity of A
and k non-zeros per column

Remove orthogonality constraint on the
rows of the dictionary

This talk

In paper
(analogous to the thin case)

Theoretical Underpinnings

So, what exactly do we need for exact recovery of X and A?

1 Thin vs. Fat

Frame Condition

For all vectors $v \in \mathbb{R}^d$,

$$\mathbf{F}_L \|\mathbf{v}\|_2^2 \leq \|\mathbf{R}\mathbf{v}\|_2^2 \leq \mathbf{F}_U \|\mathbf{v}\|_2^2$$

$$0 < \mathbf{F}_L \leq \mathbf{F}_U$$

Thin ($n > d$)

Restricted Isometry Property

For all k-sparse vectors $v \in \mathbb{R}^d$,

$$(1-\delta)\|\mathbf{v}\|_2^2 \leq \|\mathbf{R}\mathbf{v}\|_2^2 \leq (1+\delta)\|\mathbf{v}\|_2^2$$

Fat ($n < d$)

Recall

dictionary (known)

Thin ($n > d$) or Fat ($n < d$)

$$\underset{n \times m}{\mathbf{Y}} = \underset{n \times m}{\mathbf{X}} + \underset{n \times d}{\mathbf{R}} \; \underset{d \times m}{\mathbf{A}}$$

data

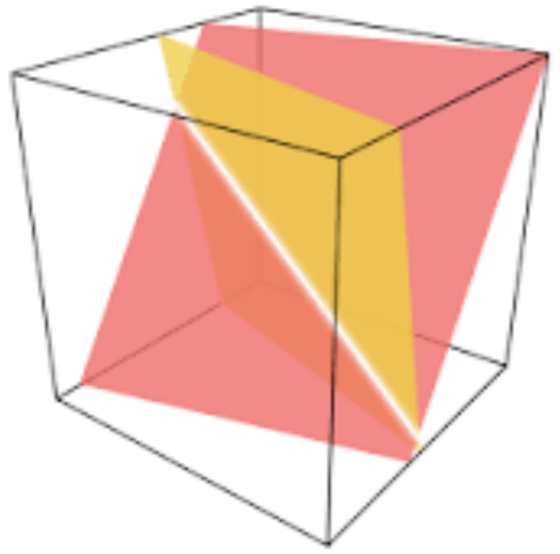
low-rank (rank: r)

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}'$$

dictionary sparse part

s non-zeros globally

2 Subspaces



$$\mathbf{X} = \mathbf{U\Sigma V'}$$

$\mathbf{\Phi}$	$\mathbf{\Omega}$	$\mathbf{\Omega_R}$	$\mathbf{P_U}$	$\mathbf{P_V}$
$\mathbf{UW_1} + \mathbf{W_2V'}$, $\mathbf{W_1}, \mathbf{W_2} \in \mathbb{R}^{\mathbf{n \times r}}$	$\mathbf{H} \in \mathbb{R}^{d \times m}$ Same support as $\mathbf{A}_0$	$\mathbf{Z} = \mathbf{RH}$, $\mathbf{H} \in \mathbf{\Omega}$	Projection Matrix for the column space of X	Projection Matrix for the row space of X
Low-rank X	Sparse Coefficient A	Dictionary Sparse RA	Column space of X U	Row space of X V

Theoretical Underpinnings

$\mathbf{X} = \mathbf{U\Sigma V'}$

$\mathbf{\Phi}$: Low-rank X

$\mathbf{\Omega}_R$: Dict. Sparse RA

3 Parameters

μ	$\gamma_{\mathbf{UR}}$	$\gamma_{\mathbf{V}}$	ξ
$\max_{\mathbf{Z}\in\mathbf{\Omega}_R\setminus\{\mathbf{0}_{d\times m}\}} \frac{\\|\mathcal{P}_{\Phi}(\mathbf{Z})\\|_F}{\\|\mathbf{Z}\\|_F}$	$\max_i \frac{\\|\mathbf{P}_U\mathbf{R}\mathbf{e}_i\\|^2}{\\|\mathbf{R}\mathbf{e}_i\\|^2}$	$\max_i \\|\mathbf{P}_V\mathbf{e}_i\\|^2$	$\\|\mathbf{R}'\mathbf{U}\mathbf{V}'\\|_\infty$
How close is RA to the low-rank part X?	Does the column space of X resemble the dictionary?	Is the row space of X sparse?	The max inner-product of R and UV'.

$\mathcal{P}(.)$ denotes projection, $\|.\| =$ spectral-norm and $\|.\|_\infty =$ max. element

4

Optimality Conditions

$$\underset{\mathbf{X},\mathbf{A}}{\text{minimize}} \quad \|\mathbf{X}\|_* + \lambda\|\mathbf{A}\|_1 \quad \text{s.t.} \quad \mathbf{Y} = \mathbf{X} + \mathbf{R}\mathbf{A} \qquad (1)$$

$$\mathcal{L}(\mathbf{X}, \mathbf{A}, \mathbf{\Lambda}) = \|\mathbf{X}\|_* + \lambda\|\mathbf{A}\|_1 + \langle \mathbf{\Lambda}, \mathbf{Y} - \mathbf{X} - \mathbf{R}\mathbf{A} \rangle$$

First-order optimality

$$\mathbf{\Lambda} \in \partial_X \|\mathbf{X}\|_* \big|_{\mathbf{X}=\mathbf{X_0}} \quad \mathbf{R}'\mathbf{\Lambda} \in \lambda \partial_A \|\mathbf{A}\|_1 \big|_{\mathbf{A}=\mathbf{A_0}}$$

$\|.\|_* =$ nuclear-norm and $\|.\|_1 =$ l_1-norm

4 Optimality Conditions

$$\underset{\mathbf{X},\mathbf{A}}{\text{minimize}} \quad \|\mathbf{X}\|_* + \lambda\|\mathbf{A}\|_1 \quad \text{s.t.} \quad \mathbf{Y} = \mathbf{X} + \mathbf{R}\mathbf{A} \qquad (1)$$

$$\mathcal{L}(\mathbf{X}, \mathbf{A}, \mathbf{\Lambda}) = \|\mathbf{X}\|_* + \lambda\|\mathbf{A}\|_1 + \langle \mathbf{\Lambda}, \mathbf{Y} - \mathbf{X} - \mathbf{R}\mathbf{A} \rangle$$

First-order optimality

$$\mathbf{\Lambda} \in \partial_X \|\mathbf{X}\|_* \big|_{\mathbf{X}=\mathbf{X}_0} \quad \mathbf{R}'\mathbf{\Lambda} \in \lambda\partial_A \|\mathbf{A}\|_1 \big|_{\mathbf{A}=\mathbf{A}_0}$$

$\|.\|_* =$ nuclear-norm and $\|.\|_1 =$ l_1-norm

4 Optimality Conditions

$$\underset{\mathbf{X},\mathbf{A}}{\text{minimize}} \quad \|\mathbf{X}\|_* + \lambda\|\mathbf{A}\|_1 \quad \text{s.t.} \quad \mathbf{Y} = \mathbf{X} + \mathbf{R}\mathbf{A} \qquad (1)$$

$$\mathcal{L}(\mathbf{X}, \mathbf{A}, \mathbf{\Lambda}) = \|\mathbf{X}\|_* + \lambda\|\mathbf{A}\|_1 + \langle \mathbf{\Lambda}, \mathbf{Y} - \mathbf{X} - \mathbf{R}\mathbf{A} \rangle$$

First-order optimality

$$\mathbf{\Lambda} \in \partial_X \|\mathbf{X}\|_* \big|_{\mathbf{X}=\mathbf{X_0}} \quad \mathbf{R}'\mathbf{\Lambda} \in \lambda \partial_A \|\mathbf{A}\|_1 \big|_{\mathbf{A}=\mathbf{A_0}}$$

$\|.\|_* =$ nuclear-norm and $\|.\|_1 =$ l_1-norm

4 Dual Certificate

First-order optimality

$$\mathbf{\Lambda} \in \mathbf{UV}' \ + \ \mathbf{W}, \ \|\mathbf{W}\| \leq 1, \ \mathcal{P}_{\Phi}(\mathbf{W}) = \mathbf{0}_{n \times m}$$

$$\mathbf{R}'\mathbf{\Lambda} \in \lambda \text{sign}(\mathbf{A}_0) \ + \ \lambda \mathbf{F}, \ \|\mathbf{F}\|_{\infty} \leq 1, \ \mathcal{P}_{\Omega}(\mathbf{F}) = \mathbf{0}_{d \times m}$$

Lemma 1 : (from Lemma 2 in Mardani et. al and Thm. 3 in Xu et. al.): *If there exists a dual certificate* $\mathbf{\Gamma} \in \mathbb{R}^{n \times m}$ *satisfying*

$$\text{C1} : \mathcal{P}_{\Phi}(\mathbf{\Gamma}) = \mathbf{UV}' \qquad \text{C2} : \mathcal{P}_{\Omega}(\mathbf{R}'\mathbf{\Gamma}) = \lambda \text{sign}(\mathbf{A}_0)$$

$$\text{C3} : \|\mathcal{P}_{\Phi^{\perp}}(\mathbf{\Gamma})\| \ < \ 1 \qquad \text{C4} : \|\mathcal{P}_{\Omega^{\perp}}(\mathbf{R}'\mathbf{\Gamma})\|_{\infty} \ < \ \lambda$$

then the pair $\{\mathbf{X}_0, \ \mathbf{A}_0\}$ *is the unique solution of eq (1).*

$(.)^{\perp}$ denotes the orthogonal complement.

4 Dual Certificate

First-order optimality →

$$\mathbf{\Lambda} \in \mathbf{U}\mathbf{V}' \; + \; \mathbf{W}, \; \|\mathbf{W}\| \leq 1, \; \mathcal{P}_{\Phi}(\mathbf{W}) = \mathbf{0}_{n\times m}$$

$$\mathbf{R}'\mathbf{\Lambda} \in \lambda \mathrm{sign}(\mathbf{A}_0) \; + \; \lambda\mathbf{F}, \; \|\mathbf{F}\|_{\infty} \leq 1, \; \mathcal{P}_{\Omega}(\mathbf{F}) = \mathbf{0}_{d\times m}$$

Lemma 1 : (from Lemma 2 in Mardani et. al and Thm. 3 in Xu et. al.): *If there exists a dual certificate* $\mathbf{\Gamma} \in \mathbb{R}^{n\times m}$ *satisfying*

$$\begin{array}{ll} \mathrm{C1}: \mathcal{P}_{\Phi}(\mathbf{\Gamma}) = \mathbf{U}\mathbf{V}' & \mathrm{C2}: \mathcal{P}_{\Omega}(\mathbf{R}'\mathbf{\Gamma}) = \lambda \mathrm{sign}(\mathbf{A}_0) \\ \mathrm{C3}: \|\mathcal{P}_{\Phi^{\perp}}(\mathbf{\Gamma})\| \; < \; 1 & \mathrm{C4}: \|\mathcal{P}_{\Omega^{\perp}}(\mathbf{R}'\mathbf{\Gamma})\|_{\infty} \; < \; \lambda \end{array}$$

then the pair $\{\mathbf{X}_0, \; \mathbf{A}_0\}$ *is the unique solution of eq (1).*

$(.)^{\perp}$ denotes the orthogonal complement.

5

Analyzing the Dual Certificate

$$\underset{\mathbf{X},\mathbf{A}}{\text{minimize}} \quad \|\mathbf{X}\|_* + \lambda\|\mathbf{A}\|_1 \quad \text{s.t.} \quad \mathbf{Y} = \mathbf{X} + \mathbf{R}\mathbf{A}$$

Definition D.1.

$$\lambda_{\min} := \frac{1+C}{1-C}\,\xi$$

$$C := \frac{c_t}{\mathbf{F}_L(1-\mu)^2 - c_t} \quad , \quad \text{where } \mathbf{F}_L \le \frac{1}{(1-\mu)^2}$$

$$c_t := \frac{\mathbf{F}_U}{2}\left[(1+2\gamma_{UR})(\min(s,d)+s\gamma_V)+2s\gamma_V\right] - \frac{\mathbf{F}_L}{2}\left[\min(s,d)+s\gamma_V\right]$$

Definition D.2.

$$\lambda_{\max} := \frac{1}{\sqrt{s}}\left(\sqrt{\mathbf{F}_L}(\,1-\mu\,) - \sqrt{r\mathbf{F}_U}\mu\right)$$

Assumption A.1.

$$\lambda_{\max} \ge \lambda_{\min}$$

Assumption A.2.

$$s_{\max} := \frac{(1-\mu)^2}{2}\,\frac{m}{r}$$

$$\gamma_{UR} \quad \le \quad \begin{cases} \frac{(1-\mu)^2 - 2s\gamma_V}{2s(1+\gamma_V)}, & \text{for } s \le \min\,(d, s_{\max}) \\ \frac{(1-\mu)^2 - 2s\gamma_V}{2(d+s\gamma_V)}, & \text{for } d < s \le s_{\max} \end{cases}$$

5 Analyzing the Dual Certificate

$$\underset{\mathbf{X},\mathbf{A}}{\text{minimize}} \quad \|\mathbf{X}\|_* + \lambda\|\mathbf{A}\|_1 \quad \text{s.t.} \quad \mathbf{Y} = \mathbf{X} + \mathbf{R}\mathbf{A}$$

Definition D.1.

$$\lambda_{\min} := \frac{1+C}{1-C}\ \xi$$

Characterizes $\lambda_{\min}$

$$C := \ldots \frac{\ldots}{\mathbf{F}_L(1-\mu)^2 - c_t}, \text{ where } \mathbf{F}_L = \frac{1}{(1-\mu)^2}$$

$$c_t := \frac{\mathbf{F}_U}{2}[(1+2\gamma_{UR})(\min(s,d) + s\gamma_V) + 2s\gamma_V] - \frac{\mathbf{F}_L}{2}[\min(s,d) + s\gamma_V]$$

Definition D.2.

Characterizes $\lambda_{\max}$

$$\lambda_{\max} := \frac{1}{\sqrt{s}}\left(\sqrt{\mathbf{F}_L}(1-\mu) - \sqrt{r\mathbf{F}_U}\,\mu\right)$$

Assumption A.1.

Existence of λ

$$\lambda_{\max} \geq \lambda_{\min}$$

Assumption A.2.

$$s_{\max} := \frac{(1-\mu)^2}{2}\frac{m}{r}$$

Interplay of parameters

$$\gamma_{UR} \leq \begin{cases} \frac{\ldots}{2s(1+\gamma_V)}, & \text{for } s \leq \min(d, s_{\max}) \\ \frac{(1-\mu)^2 - 2s\gamma_V}{2(d+s\gamma_V)}, & \text{for } d < s \leq s_{\max} \end{cases}$$

Interplay of parameters

μ

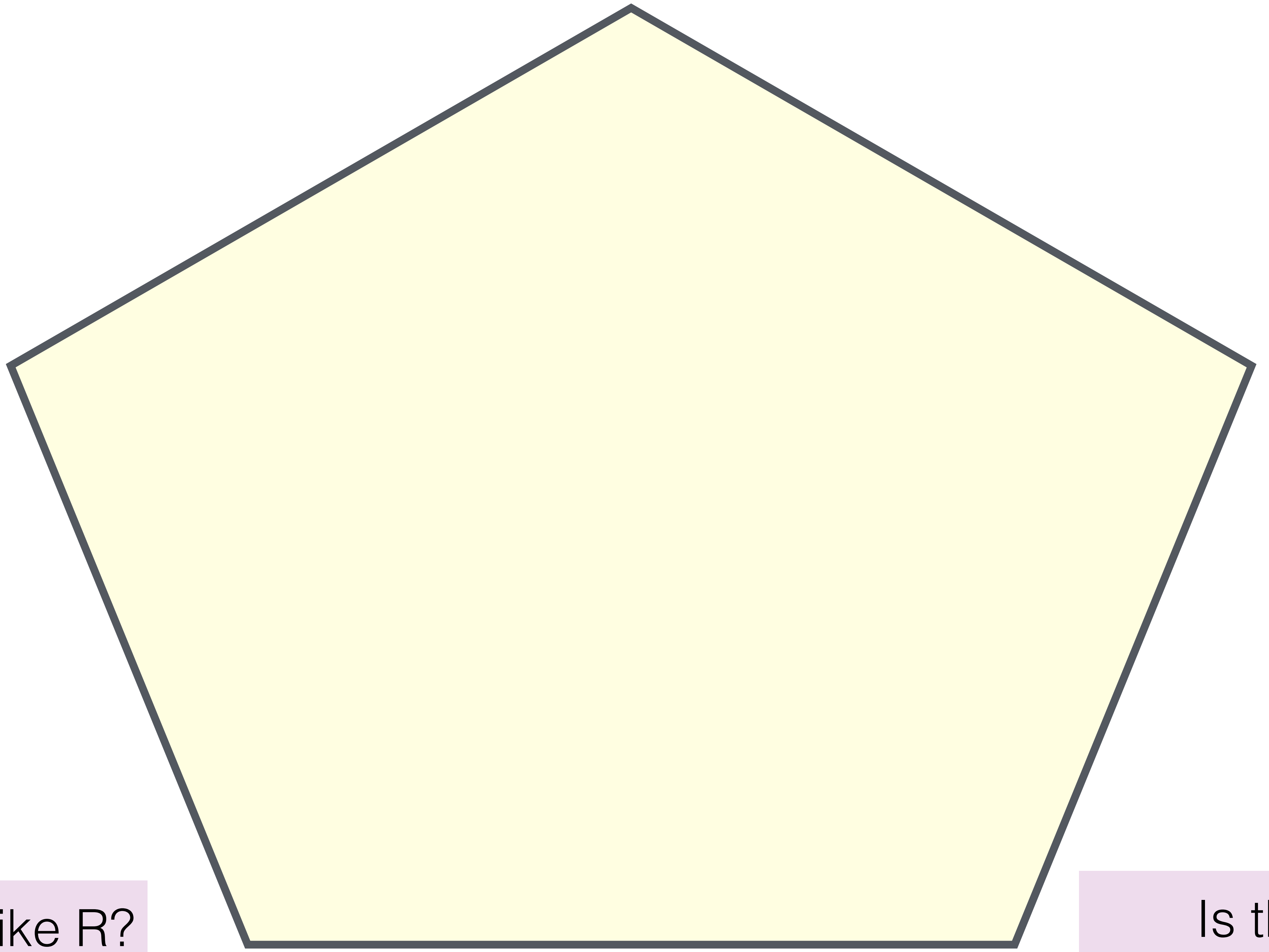
How close is RA to X?

ξ

Measure of coherence between R and UV

s

Global sparsity of A



$\gamma_{\mathbf{UR}}$

Is the column space of X acting like R?

$\gamma_{\mathbf{V}}$

Is the row space of R sparse?

Main Result (Thin Case)

Theorem 1 - Consider a superposition $\mathbf{Y} = \mathbf{X_0} + \mathbf{RA_0}$, of a low-rank matrix $\mathbf{X_0} \in \mathbb{R}^{\mathbf{n \times m}}$ of rank r, and a dictionary sparse component $\mathbf{RA_0}$, wherein the dictionary $\mathbf{R} \in \mathbb{R}^{n \times d}$ with $d \leq n$ obeys the frame condition with frame bounds $[\mathbf{F}_L, \mathbf{F}_U]$ and the sparse coefficient matrix $\mathbf{A_0} \in \mathbb{R}^{d \times m}$ has at most s non-zeros, i.e., $\|\mathbf{A}_0\|_0 = s$, with parameters $\gamma_{UR},\ \xi,\ \gamma_V \in [r/m,\ 1]$ and $\mu \in [0, 1]$.

Then, if the assumptions A.1. and A.2. hold for any $\lambda \in [\ \lambda_{\min}, \lambda_{\max}]$, then solving the optimization problem shown in eq.(1) will exactly recover matrices $\mathbf{X_0}$ and $\mathbf{A_0}$.

Phase transition in rank and sparsity for $s \leq s_{\max}$

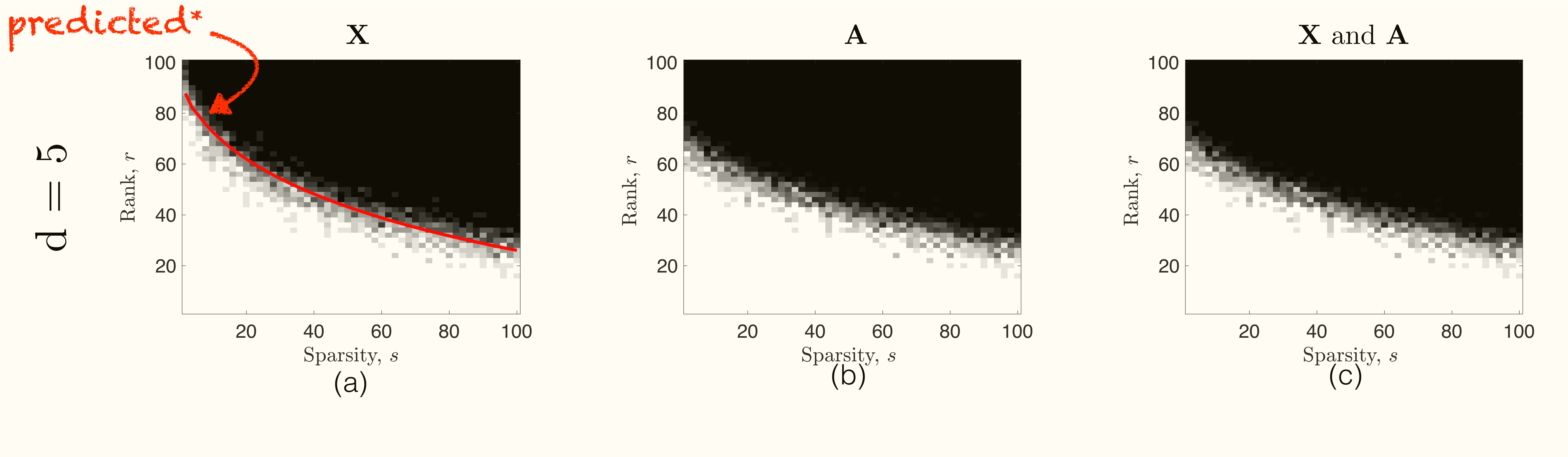


Fig.1 - Recovery for varying ranks of X and sparsity of A for the thin case with d = 5. Average recovery across 10 trials, n = m = 100, success (in white) is determined by $\|\mathbf{X}-\hat{\mathbf{X}}\|_F/\|\mathbf{X}\|_F \leq 0.02$ and $\|\mathbf{A}-\hat{\mathbf{A}}\|_F/\|\mathbf{A}\|_F \leq 0.02$. We use the accelerated proximal gradient algorithm outlined in Mardani, Mateos, and Giannakis, 2013.

$$r \leq \left(\sqrt{\frac{\mathbf{F}_L}{\mathbf{F}_U}} \frac{1-\mu}{\mu} - \frac{\xi}{\sqrt{\mathbf{F}_U}\mu} \frac{1+C}{1-C} \sqrt{s} \right)^2$$

* the parameters are manually tuned.

Phase transition in rank and sparsity for $s > s_{\max}$

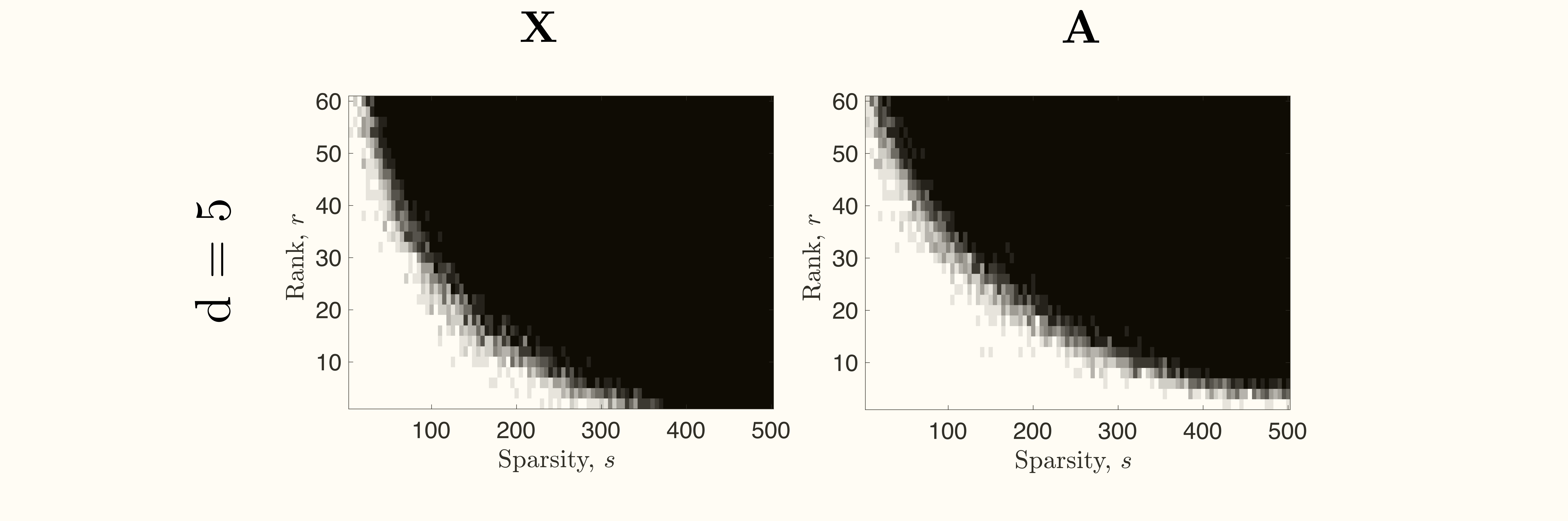
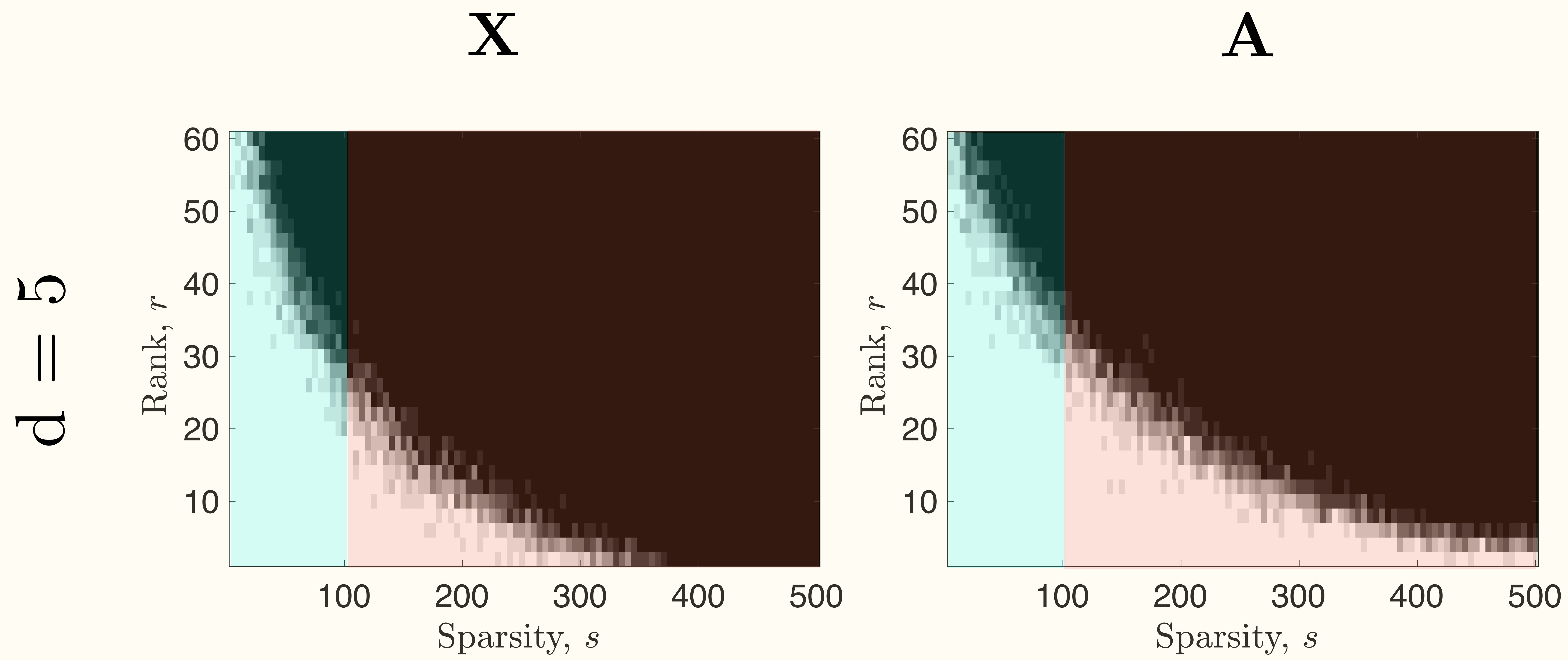


Fig.1 - Recovery for varying ranks of X and sparsity of A for the thin case with d = 5. Average recovery across 10 trials, n = m = 100, success (in white) is determined by $\|\mathbf{X}-\hat{\mathbf{X}}\|_F/\|\mathbf{X}\|_F \leq 0.02$ and $\|\mathbf{A}-\hat{\mathbf{A}}\|_F/\|\mathbf{A}\|_F \leq 0.02$. We use the accelerated proximal gradient algorithm outlined in Mardani, Mateos, and Giannakis, 2013.

Phase transition in rank and sparsity for $s > s_{\max}$



This Work
Deterministic analysis
(Location of non-zeros)

Future Work
Randomized analysis
(Location of non-zeros)

Fig.2 - Recovery for varying ranks of X and sparsity of A for the thin case with d = 5 for all sparsity levels. Average recovery across 10 trials, n = m = 100, success (in white) is determined by $\|\mathbf{X}-\hat{\mathbf{X}}\|_F/\|\mathbf{X}\|_F \leq 0.02$ and $\|\mathbf{A}-\hat{\mathbf{A}}\|_F/\|\mathbf{A}\|_F \leq 0.02$. We use the accelerated proximal gradient algorithm outlined in Mardani, Mateos, and Giannakis, 2013.

Application

Target Identification in Hyper-spectral Imaging

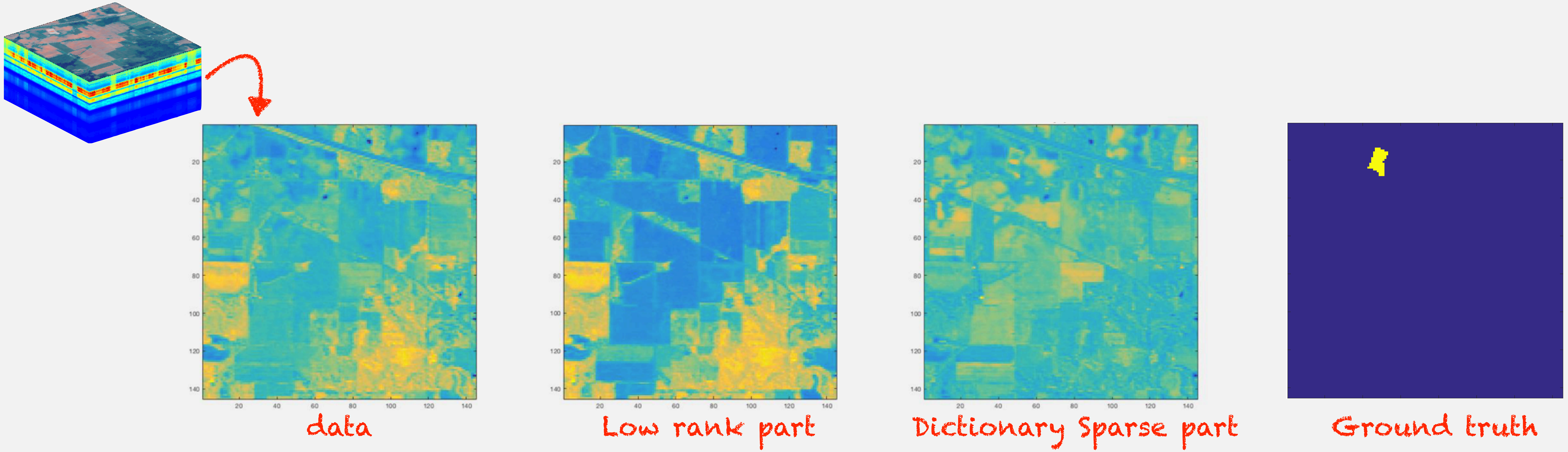


Fig.2 - Identifying the Stone-steel towers in the Indian pines hyper-spectral image data. The video shows the result of demixing of the 50th spectral band into a low rank part and a dictionary sparse part for across the range of λs .

Future Work & Conclusions

What's next?

Future Work

1

Extend the results to higher sparsity levels by assuming a random distribution on the locations of non-zeros of coefficient matrix A.

2

Analyze the problem for the noisy case Can we hope for support recovery in the presence of noise?

Conclusions

- We analyze a dictionary-based generalization of the robust PCA problem, wherein the known dictionary R can be thin or fat.
- In the thin case, we assume that the dictionary obeys the frame conditions, while in the fat case it obeys RIP of order k.
- We relax some of the constraints required by the prior art, namely orthogonality of rows of R and sparsity of rows of A for the fat case to provide a unified analysis.
- The predicted trend is confirmed by the experimental results in the form of phase transitions in rank and sparsity.

References

- [Pearson et.al.,1901] K. Pearson "On Lines and Planes of Closest Fit to Systems of Points in Space”, Philosophical Magazine. 2 (11): 559–572.

- [Candès et. al., 2009] E. J. Candès, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?,” *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 11, 2011.

- [Chandrasekaran, Sanghavi, Parrilo, and Willsky, 2011] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, “Rank-sparsity incoherence for matrix decomposition,” SIAM Journal on Optimization, vol. 21, no. 2, pp. 572–596, 2011.

- [Xu et. al., 2010] H. Xu, C. Caramanis, and S. Sanghavi, “Robust PCA via outlier pursuit,” in *Advances in Neural Information Processing Systems*, 2010, pp. 2496–2504.

- [Mardani et. al., 2013] M. Mardani, G. Mateos, and G. B. Giannakis, “Recovery of low-rank plus compressed sparse matrices with appli- cation to unveiling traffic anomalies,” *IEEE Transactions on Information Theory*, vol. 59, no. 8, pp. 5186–5205, 2013.

Clipart/Images References

- Topic modeling image - http://bigdata.ices.utexas.edu/project/scalable-topic-modeling/
- Network Traffic Anomaly Image - http://www.pdr-team.ch/businesskunden/services/
- Time travel image - http://www.slate.com/articles/health_and_science/science/2009/08/timetraveling_for_dummies.html
- Old paper background - http://wallpapercave.com/wp/4c8xmGs.jpg
- Subspaces image - https://upload.wikimedia.org/wikipedia/commons/thumb/d/d6/Intersecting_Planes_2.svg/220px-Intersecting_Planes_2.svg.png
- “The best” clipart - https://school.discoveryeducation.com/clipart/clip/certifct.html
- Parameters - http://www.free-icons-download.net/images/parameters-icon-64936.png

Thank You!

Questions and comments are welcome.

The End