Locating Outliers in Large Matrices with Adaptive Compressive Sampling

Xingguo Li
Department of Electrical and Computer Engineering
University of Minnesota

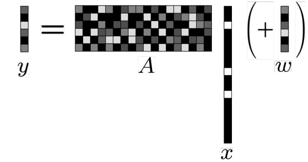
Xerox Research Centre Europe
June 30, 2016

University of Minnesota

Background and Motivation

sparsity-enabled inference

Compressed sensing:

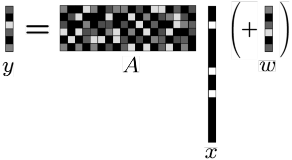


$$\arg\min_x \|x\|_1 \text{ s.t. } \|y - Ax\|_2 \leq \epsilon$$

$\Rightarrow$ infer sparse x from $\{y, A\}$ (Candes, Romberg, & Tao; Donoho; many, many others...)

sparsity-enabled inference

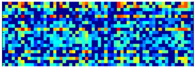
Compressed sensing:



$$\arg\min_x \|x\|_1 \text{ s.t. } \|y - Ax\|_2 \leq \epsilon$$

$\Rightarrow$ infer sparse x from $\{y, A\}$ (Candes, Romberg, & Tao; Donoho; many, many others...)

Its Variants: (e.g., matrix completion)



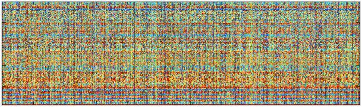
$$\arg\min_X \|X\|_* \text{ s.t. } \sum_{(i,j)\in\mathcal{S}} |Y_{i,j} - X_{i,j}|^2 \leq \epsilon$$

$\Rightarrow$ observe low-rank matrix at subset of loc's; recover by convex method(s)

(Candes & Recht; Keshavan, Montanari, & Oh; Candes & Plan; Negahban & Wainwright; Koltchinskii, Lounici, & Tsybakov; many, many others...)

problem: outliers in “big data” applications

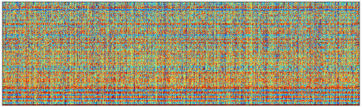
Malicious responses in survey data...



(data from personality-testing.info)

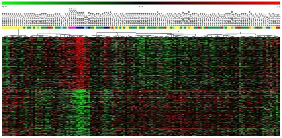
problem: outliers in "big data" applications

Malicious responses in survey data...



(data from personality-testing.info)

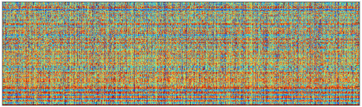
Corrupted genomics experiments...



(image from biomedcentral.com)

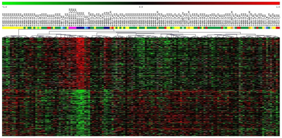
problem: outliers in “big data” applications

Malicious responses in survey data...



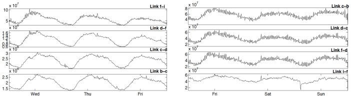
(data from personality-testing.info)

Corrupted genomics experiments...



(image from biomedcentral.com)

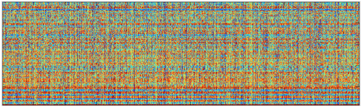
Unusual origin-desntination flows...



(Lakhina, Crovella, & Diot 2004;
Mardani, Mateos, & Giannakis 2013)

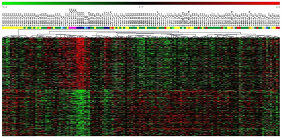
problem: outliers in “big data” applications

Malicious responses in survey data...



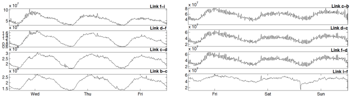
(data from personality-testing.info)

Corrupted genomics experiments...



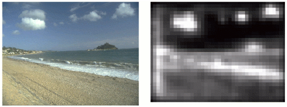
(image from biomedcentral.com)

Unusual origin-desntination flows...



(Lakhina, Crovella, & Diot 2004;
Mardani, Mateos, & Giannakis 2013)

Salient regions in high-res images...



(Itti, Koch, & Niebur 1998; many others...)

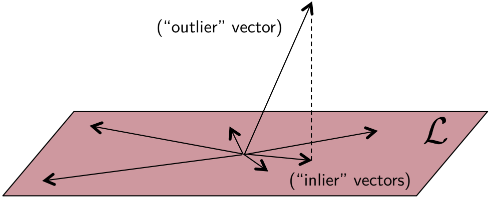
Challenge: Data points are *high-dimensional*, and numerous
Opportunity: Often we only want to locate data points that are anomalous
Question: Can sparse inference ideas help?

a structural model for outliers

Consider matrices $\mathbf{M} \in \mathbb{R}^{n_1 \times n_2}$ admitting a decomposition

$$\mathbf{M} = \underbrace{\mathbf{L}}_{\text{low rank}} + \underbrace{\mathbf{C}}_{\text{column sparse}}$$

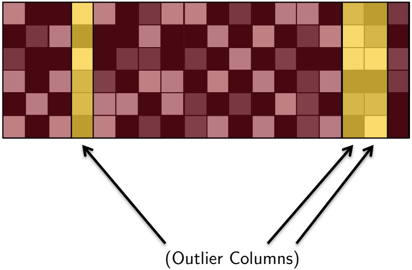
"Outliers" are vectors w/energy outside of (unknown) common subspace $\mathcal{L}$



a structural model for outliers

A *color-coded* matrix view:

$$\mathbf{M} = \mathbf{L} + \mathbf{C}$$



Define **outlier column support** to be set of indices of the outlier columns:

$$\mathcal{I}_{\mathbf{C}} := \{i \in \{1, 2, \ldots, n_2\} \ : \ \|\mathbf{P}_{\mathcal{L}^\perp}\mathbf{M}_{:,i}\|_2 > 0\}$$

prior work – recovery via convex demixing

Idea: Decompose $\mathbf{M}$ as sum of *low rank* and *column-sparse* components

(Outlier Pursuit; Xu, Caramanis, & Sanghavi 2012)

$$\left\{\widehat{\mathbf{L}}, \widehat{\mathbf{C}}\right\} = \underset{\boldsymbol{L},\boldsymbol{C}}{\operatorname{argmin}} \quad \|\boldsymbol{L}\|_* + \lambda\|\boldsymbol{C}\|_{1,2} \;\; \text{s.t. } \mathbf{M} = \boldsymbol{L} + \boldsymbol{C}$$

Here,

- $\|\mathbf{L}\|_* := \sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i$, where $\{\sigma_i\}$ are *singular values* of $\mathbf{L}$
- $\|\mathbf{C}\|_{1,2} := \sum_{i=1}^{n_2} \|\mathbf{C}_{:,i}\|_2$, where $\|\cdot\|_2$ is Euclidean norm
- $\lambda > 0$ is a regularization parameter

(Related work on robust subspace estimation: Lerman, McCoy, Tropp, & Zheng, 2012)

structural "identifiability" assumptions

Def'n: (Column Incoherence Property)

Matrix $\mathbf{L} \in \mathbb{R}^{n_1 \times n_2}$ with $n_{\mathbf{L}} \leq n_2$ nonzero columns, rank r, and compact SVD $\mathbf{L} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$ is said to satisfy the *column incoherence property* with parameter $\mu_{\mathbf{L}}$ if

$$\max_i \|\mathbf{V}^*\mathbf{e}_i\|_2^2 \leq \mu_{\mathbf{L}} \frac{r}{n_{\mathbf{L}}},$$

where $\{\mathbf{e}_i\}$ are basis vectors of the canonical basis for $\mathbb{R}^{n_2}$.

(small $\mu_{\mathbf{L}}$ precludes subspaces $\mathcal{L}$ defined by single col's of $\mathbf{L}$; an assumption that guarantees identifiability of $\{\mathbf{L}, \mathbf{C}\}$)

structural "identifiability" assumptions

Def'n: (Column Incoherence Property)

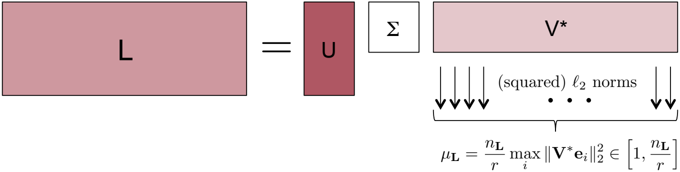
Matrix $\mathbf{L} \in \mathbb{R}^{n_1 \times n_2}$ with $n_{\mathbf{L}} \leq n_2$ nonzero columns, rank r, and compact SVD $\mathbf{L} = \mathbf{U}\Sigma\mathbf{V}^*$ is said to satisfy the *column incoherence property* with parameter $\mu_{\mathbf{L}}$ if

$$\max_i \|\mathbf{V}^* \mathbf{e}_i\|_2^2 \leq \mu_{\mathbf{L}} \frac{r}{n_{\mathbf{L}}},$$

where $\{\mathbf{e}_i\}$ are basis vectors of the canonical basis for $\mathbb{R}^{n_2}$.

(small $\mu_{\mathbf{L}}$ precludes subspaces $\mathcal{L}$ defined by single col's of $\mathbf{L}$*; an assumption that guarantees identifiability of* $\{\mathbf{L}, \mathbf{C}\}$*)*

Graphically:



existing recovery guarantees

Recovery result for Outlier Pursuit: (Xu, Caramanis, & Sanghavi 2012)

Suppose components $\mathbf{L}$ and $\mathbf{C}$ of $\mathbf{M}$ satisfy the structural conditions

- $\mathrm{rank}(\mathbf{L}) = r$,
- $\mathbf{L}$ satisfies the *column incoherence property* with parameter $\mu_{\mathbf{L}}$, and
- $|\mathcal{I}_{\mathbf{C}}| \leq \text{const.}\,(n_2/r\mu_{\mathbf{L}})$.

For any $k_{\mathrm{ub}} \geq k$, can identify a range of allowable $\lambda = \lambda(k_{\mathrm{ub}})$ s.t. any solutions $\{\widehat{\mathbf{L}}, \widehat{\mathbf{C}}\}$ of the *outlier pursuit* procedure satisfy span $\left(\widehat{\mathbf{L}}\right) = \mathcal{L}$, and $\widehat{\mathcal{I}}_{\widehat{\mathbf{C}}} \triangleq \{i : \|\widehat{\mathbf{C}}_{:,i}\|_2 > 0\} = \mathcal{I}_{\mathbf{C}}$.

Nice! But, outlier pursuit can be *computationally expensive* on large-scale data (req's iterative computation of SVD's of $n_1 \times n_2$ matrices)

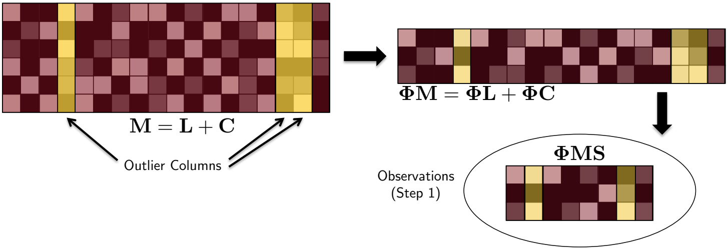
What if we seek only outlier locations?

Adaptive CS for Outlier Localization

a two-step approach (step 1)

Collect Measurements: $\mathbf{Y}_{(1)} := \mathbf{\Phi M S} = \mathbf{\Phi}(\mathbf{L} + \mathbf{C})\mathbf{S}$ where

- $\mathbf{\Phi} \in \mathbb{R}^{m \times n_1}$ is a (random) measurement matrix ($m < n$)
- For $\gamma \in (0, 1)$, $\mathbf{S}$ is a column sub matrix of identity with $\approx \gamma n_2$ columns *(rows sampled iid from a Bernoulli(γ) model)*

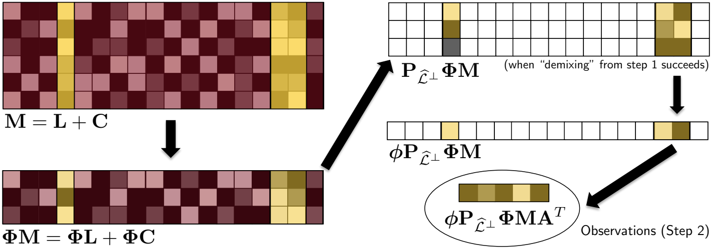


Apply *Outlier Pursuit* to "pocket-sized" data $\mathbf{\Phi MS}$ *(idea: identify span of $\mathbf{\Phi L}$)*

a two-step approach (step 2)

Collect measurements $\mathbf{y}_{(2)} := \phi \, \mathbf{P}_{\widehat{\mathcal{L}}_{(1)}^{\perp}} \, \mathbf{\Phi M A}^T$ where

- $\mathbf{\Phi} \in \mathbb{R}^{m \times n_1}$ is same (random) measurement matrix as in step 1,
- $\widehat{\mathcal{L}}_{(1)}$ is the linear subspace spanned by col's of $\widehat{\mathbf{L}}_{(1)}$ (learned in step 1)
- $\mathbf{P}_{\widehat{\mathcal{L}}_{(1)}}$ is orthogonal projector onto $\widehat{\mathcal{L}}_{(1)}$; $\mathbf{P}_{\widehat{\mathcal{L}}_{(1)}^{\perp}} \triangleq \mathbf{I} - \mathbf{P}_{\widehat{\mathcal{L}}_{(1)}}$
- $\phi \in \mathbb{R}^{1 \times m}$ a random vector, $\mathbf{A} \in \mathbb{R}^{p \times n_2}$ a random matrix



Solve $\widehat{\mathbf{c}} = \arg\min_{\mathbf{c}} \; \|\mathbf{c}\|_1$ s.t. $\mathbf{y}_{(2)} = \mathbf{c}\mathbf{A}^T$

($\text{support}(\widehat{\mathbf{c}}) \triangleq \{i : \widehat{c}_i \neq 0\}$, *becomes estimate for outlier locations*)

Performance Analysis

assumptions

Suppose components **L** and **C** of **M** satisfy the structural conditions

- $\mathrm{rank}(\mathbf{L}) = r$,
- **L** has $n_{\mathbf{L}} \leq n_2$ nonzero columns,
- **L** satisfies the *column incoherence property* with parameter $\mu_{\mathbf{L}}$, and
- $|\mathcal{I}_{\mathbf{C}}| = k \leq \frac{1}{3}\left(\frac{1}{1+121\ r\mu_{\mathbf{L}}}\right) n_2$.

assumptions

Suppose components **L** and **C** of **M** satisfy the structural conditions

- $\mathrm{rank}(\mathbf{L}) = r$,
- **L** has $n_{\mathbf{L}} \leq n_2$ nonzero columns,
- **L** satisfies the *column incoherence property* with parameter $\mu_{\mathbf{L}}$, and
- $|\mathcal{I}_{\mathbf{C}}| = k \leq \frac{1}{3}\left(\frac{1}{1+121\ r\mu_{\mathbf{L}}}\right) n_2$.

Take ϕ to have elements drawn iid from any continuous distribution.

assumptions

Suppose components **L** and **C** of **M** satisfy the structural conditions

- $\text{rank}(\mathbf{L}) = r$,
- **L** has $n_{\mathbf{L}} \leq n_2$ nonzero columns,
- **L** satisfies the *column incoherence property* with parameter $\mu_{\mathbf{L}}$, and
- $|\mathcal{I}_{\mathbf{C}}| = k \leq \frac{1}{3}\left(\frac{1}{1+121\ r\mu_{\mathbf{L}}}\right) n_2$.

Take ϕ to have elements drawn iid from any continuous distribution.

Take $\mathbf{\Phi}$ and $\boldsymbol{A}$ to satisfy the *Distributional Johnson-Lindenstrauss (JL) Property*:

Def'n: (Distributional Johnson Lindenstrauss (JL) Property)

An $m \times n$ matrix $\boldsymbol{B}$ is said to satisfy the *distributional JL property* if for any fixed $\mathbf{v} \in \mathbb{R}^n$ and any $\epsilon \in (0, 1)$,

$$\Pr\left(\ \left|\ \|\boldsymbol{B}\mathbf{v}\|_2^2 - \|\mathbf{v}\|_2^2\ \right| \geq \epsilon\|\mathbf{v}\|_2^2\ \right) \leq 2e^{-mf(\epsilon)},$$

where $f(\epsilon) > 0$ is a constant depending on ϵ, specific to the distribution of $\boldsymbol{B}$.

(e.g., $f(\epsilon) = \epsilon^2/4 - \epsilon^3/6$ for iid zero-mean $\mathcal{N}(0, 1/m)$ ensemble)

Then...

provable recovery

Theorem: (Xingguo Li & JH 2015, 2016)

Fix any $\delta \in (0, 1/3)$, and choose

$$\gamma \geq \max\left\{ \frac{200\log(\frac{6}{\delta})}{n_{\mathbf{L}}}, \frac{600(1+121r\mu_{\mathbf{L}})\log(\frac{6}{\delta})}{n_2}, \frac{10r\mu_{\mathbf{L}}\log(\frac{6r}{\delta})}{n_{\mathbf{L}}} \right\},$$

$$m \geq \frac{5(r+1)+\log(k)+\log(2/\delta)}{f(1/4)}, \quad p \geq \frac{11k+2k\log(n_2/k)+\log(2/\delta)}{f(1/4)}.$$

Take the outlier pursuit reg. parameter $\lambda = \frac{3}{7\sqrt{k_{\rm ub}}}$, where $k_{\rm ub}$ is any upper bound of k. Then w.p. $\geq 1-3\delta$, the support estimate produced by our method is correct, and the total number of observations is no greater than $\left(\frac{3}{2}\right)\gamma m n_2 + p$.

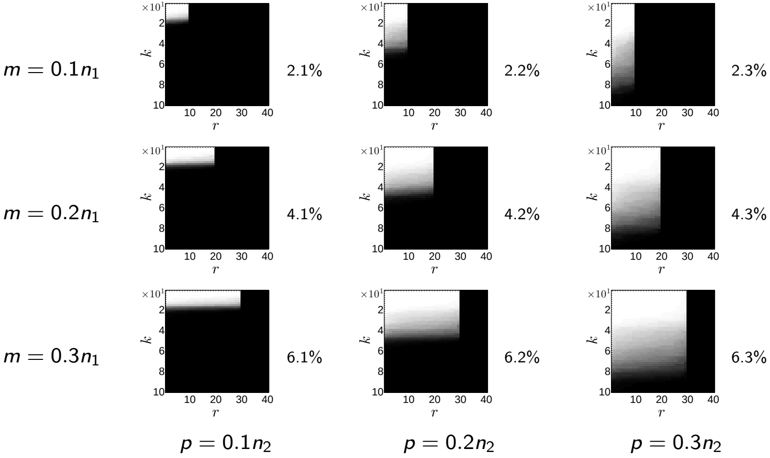
IEEE Trans. Sig. Proc. 63(7) pp. 1792-1807, April 2015
IEEE Workshop on Stat. Sig. Proc., 2016

Key Point: Localization from as few as $\underbrace{\mathcal{O}\left((r+\log k)(\mu_{\mathbf{L}} r \log r) + k\log(n_2/k)\right)}_{\mathcal{O}\left((\mu_{\mathbf{L}} r^2 + k)\,\cdot\,\text{polylog}(k,r,n_2) \right) \ll n_1 n_2}$ obs.

Experimental Results

phase transitions – synthetic (Gaussian) data

Outlier recovery phase transitions (white regions ↔ successful recovery).

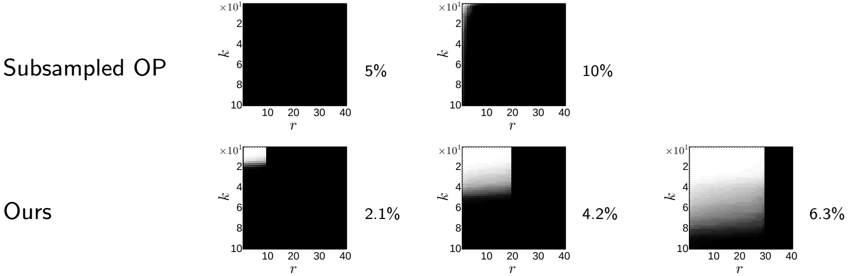


Top → bottom: increasing # rows in $\mathbf{\Phi}$ (recover w/increasing rank r of $\mathcal{L}$)
Left → right: increasing # cols in $\boldsymbol{A}^T$ (recover increasing # k of outliers)

comparisons w/existing subsampling methods

Compare with entry-wise subsampled variant of *outlier pursuit*

(Chen, Xu, Caramanis, & Sanghavi "Robust matrix completion with corrupted columns," ICML, 2011)



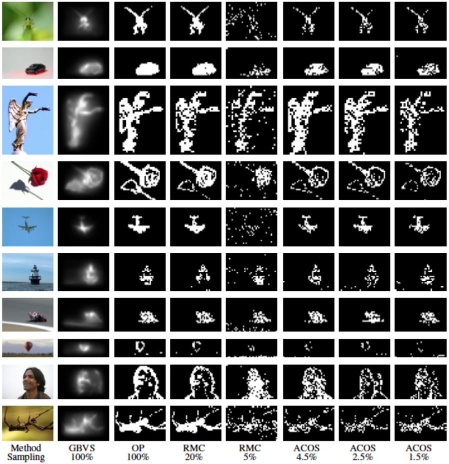
Take-away: Accurate outlier localization in a wider range of (r, k) w/ours, *but (to be fair!) we are using a "more favorable" sampling model*

an application in computer vision

Given an image $\mathbf{F} \in \mathbb{R}^{t_1 \times t_2}$,

- Decompose into n_2 non-overlapping $p_1 \times p_2$-pixel patches
- Vectorize patches into $n_1 \times 1$ column vectors, where $n_1 = p_1 p_2$
- Assemble column vectors into a matrix $\mathbf{M} \in \mathbb{R}^{n_1 \times n_2}$ (overall, $n_1 n_2 = t_1 t_2$)

Images from the Microsoft Research Salient Object Database



"GBVS" = Graph-based visual saliency [Harel et al. 2007]

"RMC" = Subsampled OP [Chen et al. 2011]

"ACOS" = Adapt. Compr. Outlier Sensing (ours)

an application in computer vision

Timing Comparison:

Method Sampling	GBVS 100%	OP 100%	RMC 20%	RMC 5%	ACOS 4.5%	ACOS 2.5%	ACOS 1.5%
Step 1	0.9926	2.9441	2.6324	2.7254	0.0533	0.0214	0.0105
	(0.2742)	(0.3854)	(0.3237)	(0.3660)	(0.0118)	(0.0056)	(0.0025)
Step 2	–	–	–	–	0.2010	0.2014	0.2065
	–	–	–	–	(0.0674)	(0.0692)	(0.0689)

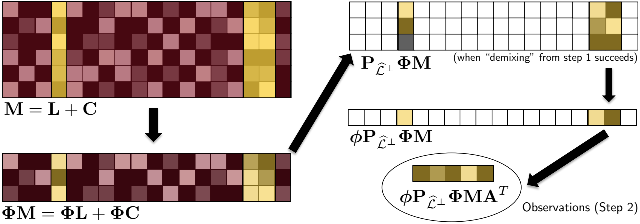
mean (st. dev.) in seconds; averaged over 1000 images from Microsoft Research Salient Object Database

Extensions

finding "group-structured" features

Recall the second step of our two-step approach...

Observations:

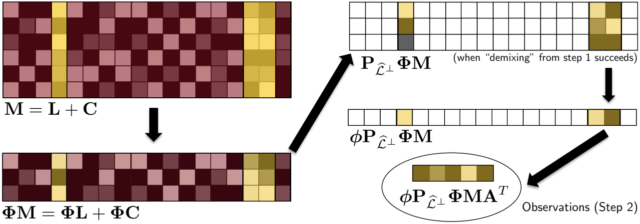


Inference: $\widehat{\mathbf{c}} = \text{argmin}_{\mathbf{c}\in\mathbb{R}^{n_2}} \quad \|\mathbf{c}\|_1 \;\; \text{s.t. } \mathbf{y}_{(2)} = \mathbf{c}\mathbf{A}^T$

finding "group-structured" features

Recall the second step of our two-step approach...

Observations:



Inference: $\widehat{\mathbf{c}} = \mathrm{argmin}_{\mathbf{c}\in\mathbb{R}^{n_2}} \quad \|\mathbf{c}\|_1 \;\; \text{s.t. } \mathbf{y}_{(2)} = \mathbf{c}\mathbf{A}^T$

$\Rightarrow$ Can instead seek "structured sparse" outliers, e.g.,

$$\widehat{\mathbf{c}} = \underset{\mathbf{c}\in\mathbb{R}^{n_2}}{\mathrm{argmin}} \sum_{j\in J} \|\mathbf{c}_j\|_2 \;\; \text{s.t. } \mathbf{y}_{(2)} = \mathbf{c}\mathbf{A}^T$$

($j \in J$ indexes groups partitioning $\{1, \ldots, n_2\}$, $\mathbf{c}_j$ is corresp. subvector of $\mathbf{c}$)

improved recovery results

Assume nonzero columns occur in groups of size $B = n_2/J$.

Under same structural assumptions, can locate outliers from as few as

$$m_{\text{tot}} = \mathcal{O}\left((r + \log k)(\mu_{\mathbf{L}} r \log r) + k + \frac{k}{\sqrt{B}} \log\left(\frac{n_2 - k}{B}\right)\right)$$

measurements. [Li & Haupt, GlobalSIP 2015 (**Best Student Paper Award**)]

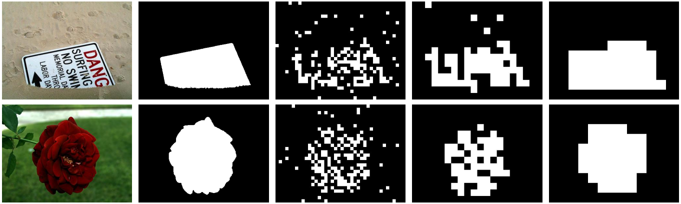
improved recovery results

Assume nonzero columns occur in groups of size $B = n_2/J$.

Under same structural assumptions, can locate outliers from as few as

$$m_{\text{tot}} = \mathcal{O}\left((r + \log k)(\mu_{\mathbf{L}} r \log r) + k + \frac{k}{\sqrt{B}} \log\left(\frac{n_2 - k}{B}\right)\right)$$

measurements. [Li & Haupt, GlobalSIP 2015 (**Best Student Paper Award**)]



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

Detection results with the grouping effect. (a) original images; (b) ground truth; detection result (c) w/o grouping ($B = 1$) and with grouping effects: (d) $B = 2$ and (e) $B = 3$. Sampling rate: 2.5% ($\gamma = 0.2$, $m = 0.1n_1$ and $p = 0.5n_2$).

locating salient features in "filtered" images...

Each step of our two-step process obtains linear measurements of the image pixels.

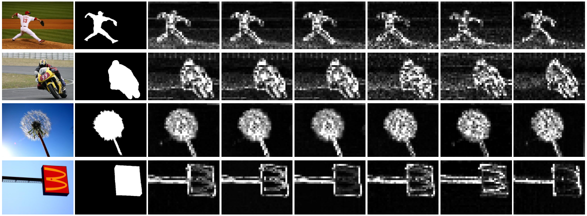
$\Rightarrow$ Can incorporate any linear "preprocessing" (e.g., *filtering*) into the overall measurement model; seek salient features of filtered image [Li & Haupt, GlobalSIP 2015].

locating salient features in “filtered” images...

Each step of our two-step process obtains linear measurements of the image pixels.

⇒ Can incorporate any linear “preprocessing” (e.g., *filtering*) into the overall measurement model; seek salient features of filtered image [Li & Haupt, GlobalSIP 2015].

Examples:



(a) (b) (c) (d) (e) (f) (g) (h)

Gray scale saliency map estimation. (a) original images; (b) ground truth; (c)-(e) RGB planes individually; filtered intensity images with (f) Laplacian of Gaussian filter, (g) Horizontal Edge filter and (h) Vertical Edge filter. Sampling rate: 4.5% ($\gamma = 0.2$, $m = 0.2n_1$, $p = 0.5n_2$, $n_1 = 100$ and $n_2 = 1200$).

(Shown: magnitudes of recovered **c** vector elements, reshaped into images.)

Summary

summary

Key Insight:

- Interpret outlier localization as generalized sparse "support recovery"
- Extend adaptive & compressive sensing ideas to outlier identification!

Main Results:

- For low-rank-plus-outlier matrix model, accurate outlier localization from $\mathcal{O}\left((\mu_{\mathbf{L}} r^2 + k) \cdot \text{polylog}(k, r, n_2) \right)$ obs. (fewer when exploiting group structure!)
- *Can find outliers using (storing/processing)* **much smaller** *data "footprint"!*
- Reminiscent of "standard" CS, with add'l $\mathcal{O}(r^2 \text{polylog}(k, r, n_2))$ term;
  $\rightarrow$ interpretation: sampling overhead to pay for not knowing "background"

Extensions to noisy & missing data settings & dictionary based outlier detection

Thanks for your attention!

Extra Slides

a "simplified" method

Collect Measurements: $\mathbf{Y} = \mathbf{\Phi}\mathbf{M}$ (w/ $\mathbf{\Phi}$ $m \times n$, random as above)

For column subsampling matrix $\mathbf{S}$ as above, let $\mathbf{Y}_{(1)} = \mathbf{YS}$ and solve

$$\{\widehat{\mathbf{L}}_{(1)}, \widehat{\mathbf{C}}_{(1)}\} = \operatorname*{argmin}_{\boldsymbol{L},\boldsymbol{C}} \|\boldsymbol{L}\|_* + \lambda\|\boldsymbol{C}\|_{1,2} \quad \text{s.t.} \quad \mathbf{Y}_{(1)} = \boldsymbol{L} + \boldsymbol{C}$$

Let $\widehat{\mathcal{L}}_{(1)}$ be span of $\widehat{\mathbf{L}}_{(1)}$, and form $\widehat{\mathbf{c}}$ with $\widehat{c}_i = \mathbf{1}_{\{\|\mathbf{P}_{\widehat{\mathcal{L}}_{(1)}^{\perp}} \mathbf{Y}_{:,i}\|_2 \neq 0\}}$ for $i = 1, \ldots, n_2$

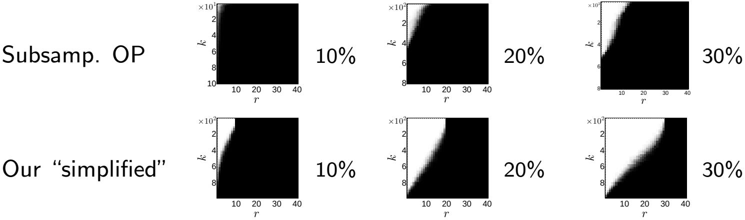
a "simplified" method

Collect Measurements: $\mathbf{Y} = \mathbf{\Phi M}$ (w/ $\mathbf{\Phi}$ $m \times n$, random as above)

For column subsampling matrix $\mathbf{S}$ as above, let $\mathbf{Y}_{(1)} = \mathbf{YS}$ and solve

$$\{\widehat{\mathbf{L}}_{(1)}, \widehat{\mathbf{C}}_{(1)}\} = \underset{\boldsymbol{L},\boldsymbol{C}}{\operatorname{argmin}} \|\boldsymbol{L}\|_* + \lambda \|\boldsymbol{C}\|_{1,2} \quad \text{s.t.} \quad \mathbf{Y}_{(1)} = \boldsymbol{L} + \boldsymbol{C}$$

Let $\widehat{\mathcal{L}}_{(1)}$ be span of $\widehat{\mathbf{L}}_{(1)}$, and form $\widehat{\mathbf{c}}$ with $\widehat{c}_i = \mathbf{1}_{\{\|\mathbf{P}_{\widehat{\mathcal{L}}_{(1)}^{\perp}} \mathbf{Y}_{:,i}\|_2 \neq 0\}}$ for $i = 1, \ldots, n_2$



Comparison between Subsampled OP and our "simplified" method for outlier recovery phase transitions plots (white regions $\leftrightarrow$ successful recovery).

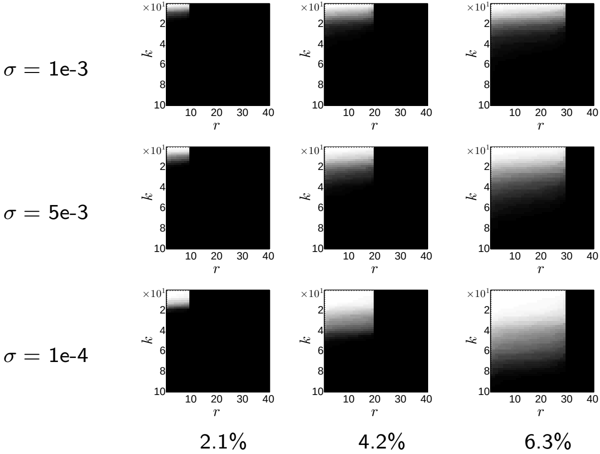
noise/modeling error

Noisy observations:

$$\mathbf{M} = \mathbf{L} + \mathbf{C} + \mathbf{N},$$

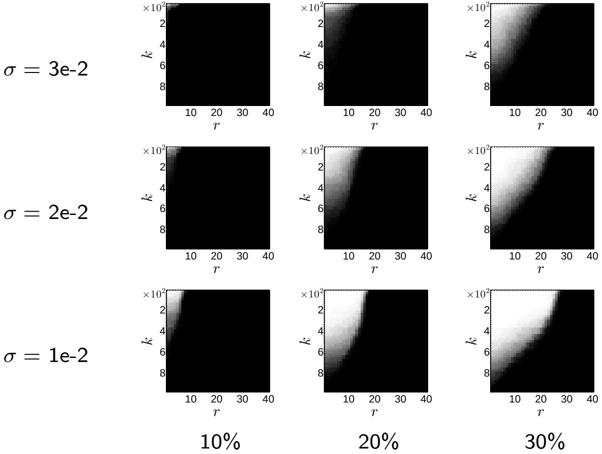
where $\mathbf{N}$ has i.i.d. $\mathcal{N}(0, \sigma^2)$ entries.

phase transitions: noisy case, our method



Outlier recovery phase transitions plots for our method with noise (white regions $\leftrightarrow$ successful recovery).

phase transitions: noisy case, our simplified method



Outlier recovery phase transitions plots for our simplified method with noise (white regions $\leftrightarrow$ successful recovery).

missing data

Given a subset $\mathbf{\Omega} \subseteq \{1, \ldots, n_1\} \times \{1, \ldots, n_2\}$, the available data is modeled as

$$\mathbf{P}_{\mathbf{\Omega}}(\mathbf{M}) = \mathbf{P}_{\mathbf{\Omega}}(\mathbf{L}) + \mathbf{P}_{\mathbf{\Omega}}(\mathbf{C}),$$

where $\mathbf{P}_{\mathbf{\Omega}}(\cdot)$ masks its argument at locations not in $\mathbf{\Omega}$.

missing data

Given a subset $\mathbf{\Omega} \subseteq \{1, \ldots, n_1\} \times \{1, \ldots, n_2\}$, the available data is modeled as

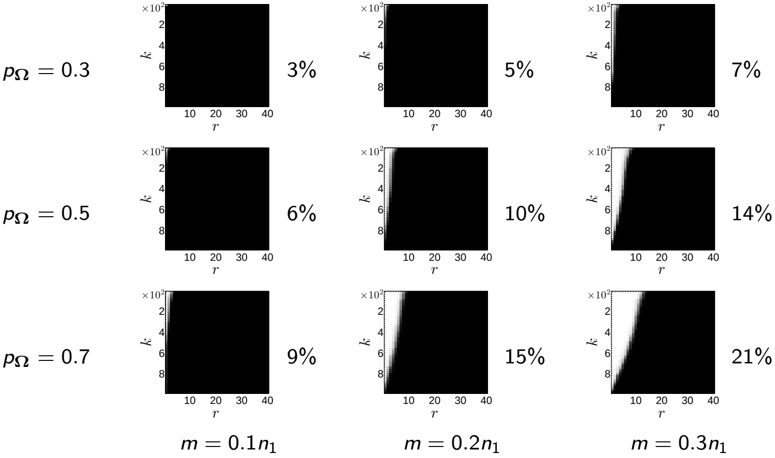
$$\mathbf{P}_{\mathbf{\Omega}}(\mathbf{M}) = \mathbf{P}_{\mathbf{\Omega}}(\mathbf{L}) + \mathbf{P}_{\mathbf{\Omega}}(\mathbf{C}),$$

where $\mathbf{P}_{\mathbf{\Omega}}(\cdot)$ masks its argument at locations not in $\mathbf{\Omega}$.

Modifications to our (simplified) method

- Choose $\mathbf{\Phi}$ to be a *row* subsampling matrix and observe $\mathbf{Y} = \mathbf{\Phi}\mathbf{P}_{\mathbf{\Omega}}(\mathbf{M})$
- Note subsampling operations "commute": $\mathbf{\Phi}\mathbf{P}_{\mathbf{\Omega}}(\mathbf{M}) = \mathbf{P}_{\mathbf{\Omega}_{\mathbf{\Phi}}}(\mathbf{\Phi}\mathbf{M})$,
- Step 1: let $\mathbf{Y}_{(1)} = \mathbf{Y}\mathbf{S}$ and solve
  $\{\widehat{\mathbf{L}}_{(1)}, \widehat{\mathbf{C}}_{(1)}\} = \mathrm{argmin}_{\boldsymbol{L},\boldsymbol{C}} \|\boldsymbol{L}\|_* + \lambda \|\boldsymbol{C}\|_{1,2}$ s.t. $\mathbf{Y}_{(1)} = \mathbf{P}_{\mathbf{\Omega}_{\mathbf{\Phi}}}(\boldsymbol{L} + \boldsymbol{C})$
  to learn subspace spanned by $\mathbf{\Phi}\mathbf{L}$
- Step 2: for each column $\mathbf{Y}_{:,j}$
  - let $\mathcal{I}_j$ be set of observed loc's
  - find subspace spanned by col's of *row-sampled* $(\widehat{\mathbf{L}}_{(1)})_{\mathcal{I}_j,:}$
  - project $\mathbf{Y}_{:,j}$ onto the orth. complement of that subspace
  - compute norm of resulting "residual" vector (nonzero $\leftrightarrow$ outlier column)

phase transitions



Outlier recovery phase transitions plots for models with missing data (white regions $\leftrightarrow$ successful recovery).