Exact Low Tubal Rank Tensor Recovery from Gaussian Measurements

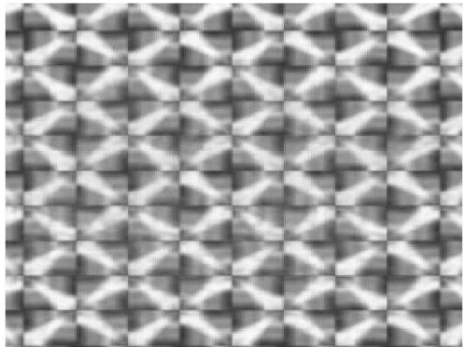
Canyi Lu[1], Jiashi Feng[2], Zhouchen Lin[3], Shuicheng Yan[2]

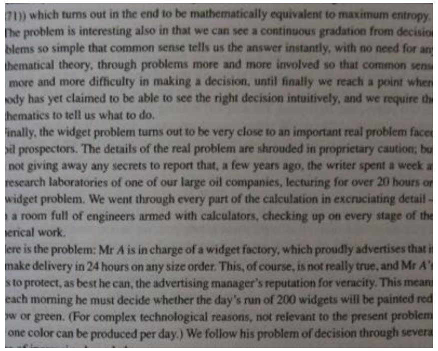
[1]Carnegie Mellon University

[2]National University of Singapore

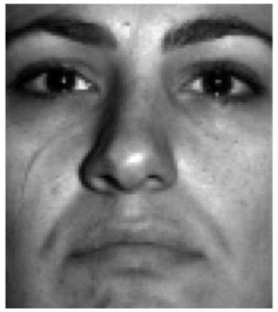
[3]Peking University

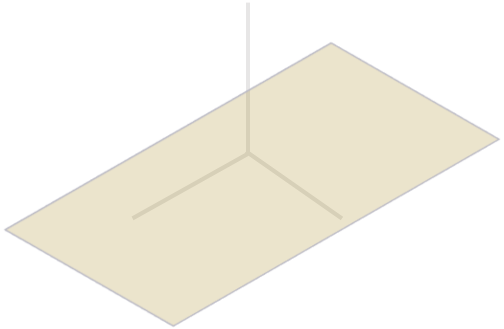
Low dimensional structures in visual data





71)) which turns out in the end to be mathematically equivalent to maximum entropy.
The problem is interesting also in that we can see a continuous gradation from decisio
blems so simple that common sense tells us the answer instantly, with no need for an
thematical theory, through problems more and more involved so that common sens
more and more difficulty in making a decision, until finally we reach a point wher
ody has yet claimed to be able to see the right decision intuitively, and we require th
hematics to tell us what to do.
Finally, the widget problem turns out to be very close to an important real problem face
oil prospectors. The details of the real problem are shrouded in proprietary caution; bu
not giving away any secrets to report that, a few years ago, the writer spent a week a
research laboratories of one of our large oil companies, lecturing for over 20 hours o
widget problem. We went through every part of the calculation in excruciating detail –
a room full of engineers armed with calculators, checking up on every stage of the
erical work.
Here is the problem: Mr A is in charge of a widget factory, which proudly advertises that i
make delivery in 24 hours on any size order. This, of course, is not really true, and Mr A'
s to protect, as best he can, the advertising manager's reputation for veracity. This mean
each morning he must decide whether the day's run of 200 widgets will be painted red
ow or green. (For complex technological reasons, not relevant to the present problem
one color can be produced per day.) We follow his problem of decision through severa





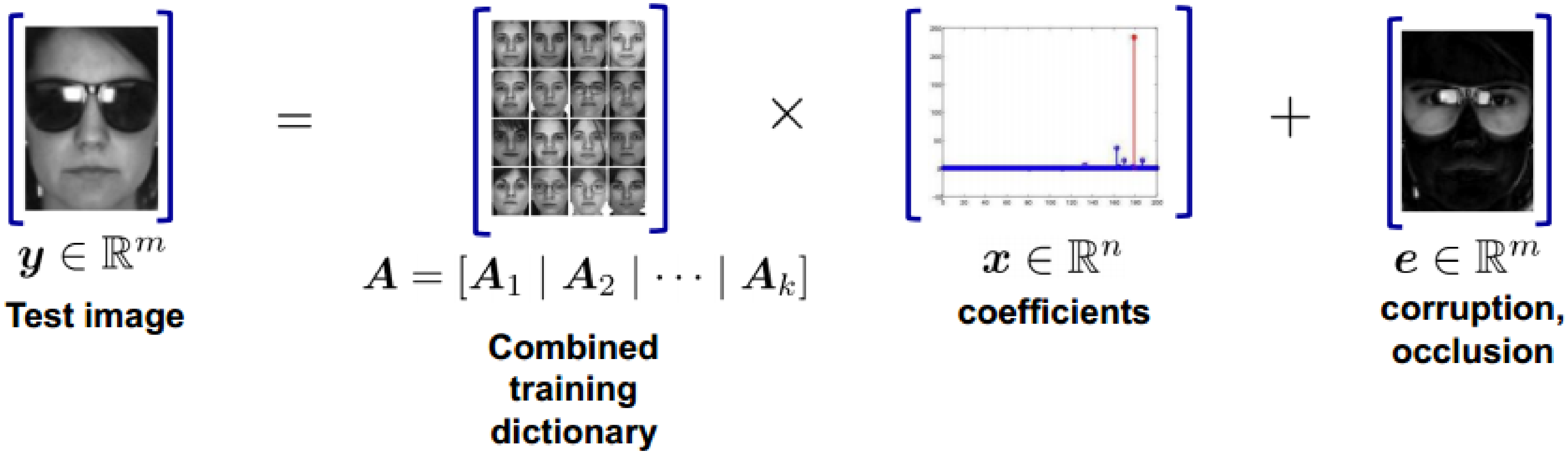
Learning by using the underlying low dimensional structure of data is important.

Compressive Sensing

- Compressive sensing: learning by using sparse vector structure

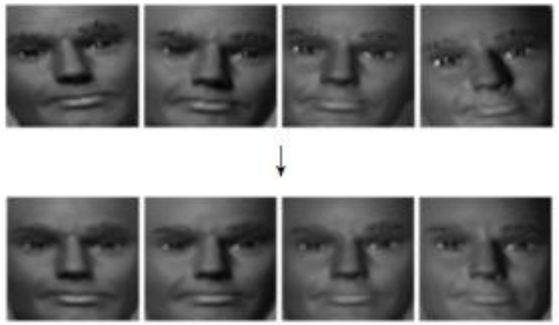
$$\min \|\boldsymbol{x}\|_1, \text{ s.t. } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$

- Face recognition (J. Wright, et al., TPAMI, 2009)



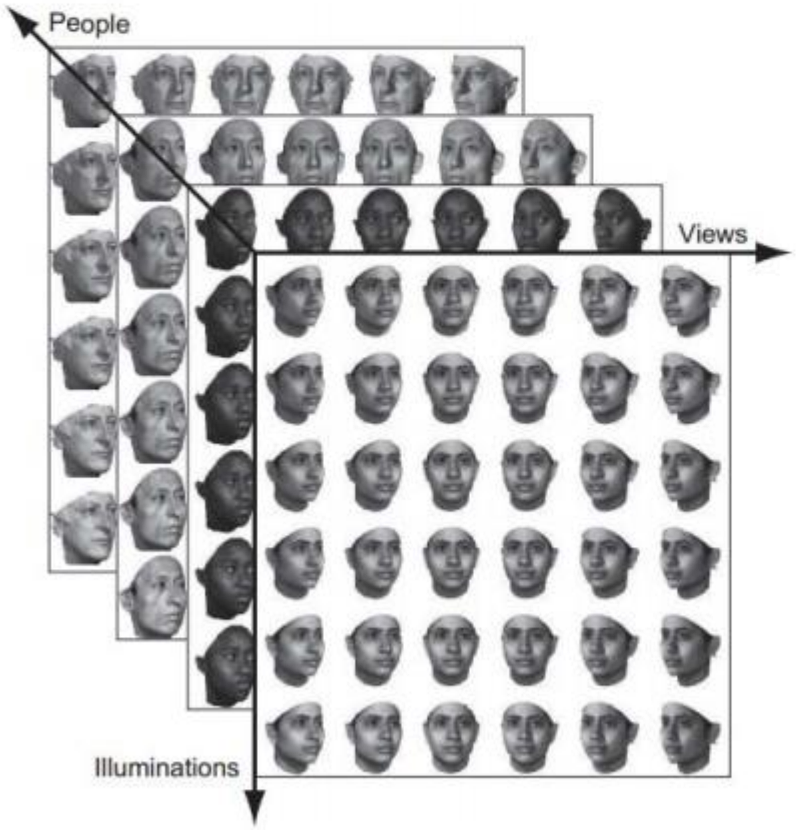
Low-rank Matrix Recovery

- Low-rank matrix: sparse singular values
- Low-rank structure is common in visual data
- Low-rank models, e.g., robust PCA, and matrix completion, have many applications
  - Background modeling
  - Removing shadows from face images
  - Image alignment
  - Many others…

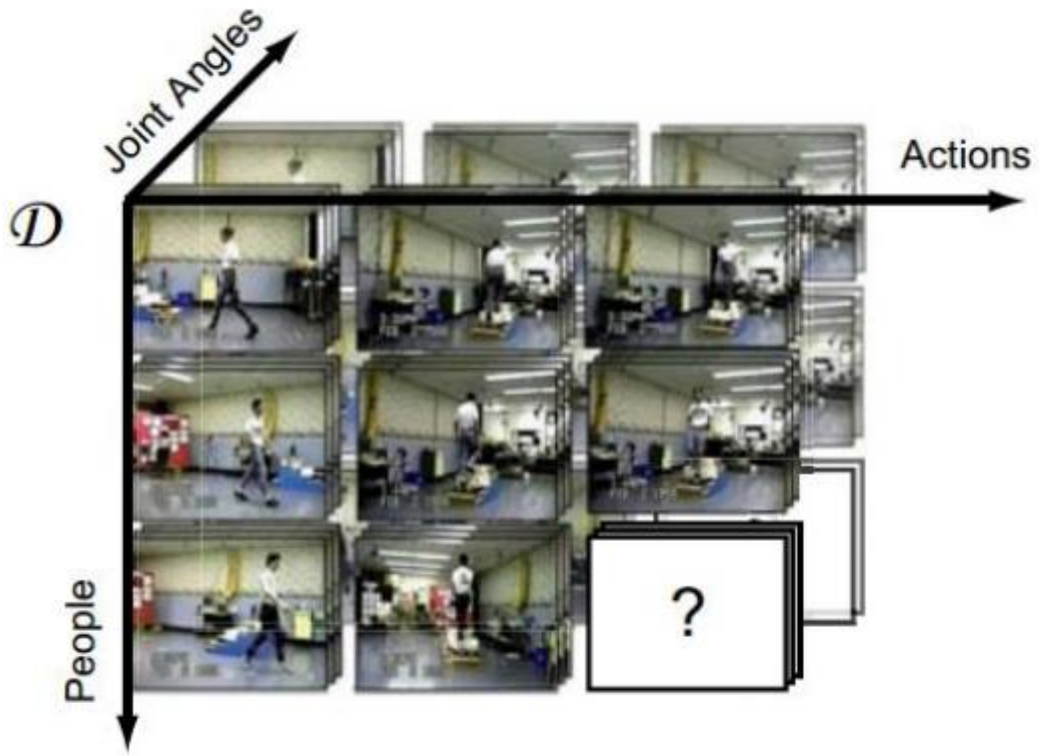


E. J. Cand` es, X. D. Li, Y. Ma, and J. Wright. Robust principal component analysis? Journal of the ACM, 2011

Multi-dimensional Data: Tensor

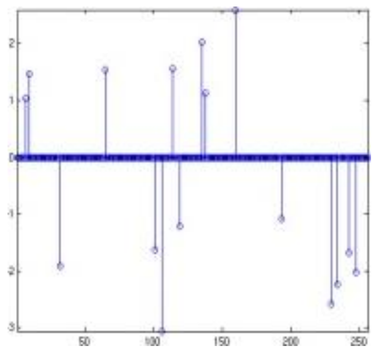


(a) Face images

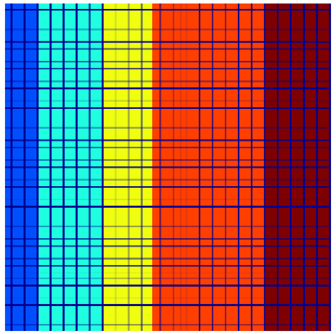


(b) Videos

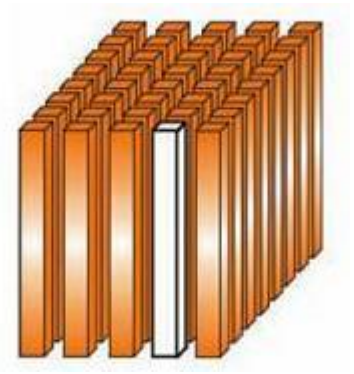
Structured Sparsity



Sparse vector



Low rank matrix



Low rank tensor

This work

Low-rank Tensor Learning Is Challenging

- The tensor rank and tensor nuclear norm are not well defined
  - Tensor CP-rank and its convex envelop are NP-hard to compute

$$\text{rank}_{cp}(\mathcal{X}) = \min_{R}\{R \mid \mathcal{X} = \sum_{i=1}^{R} \mathbf{a}_r^{(1)} \circ \mathbf{a}_r^{(2)} \circ \cdots \circ \mathbf{a}_r^{(N)}\}.$$

  - Tucker rank and Sum of Nuclear Norm (SNN)

$$\text{rank}_{\text{tc}}(\mathcal{X}) = \left(\text{rank}\left(\boldsymbol{X}^{(1)}\right), \text{rank}\left(\boldsymbol{X}^{(2)}\right), \cdots, \text{rank}\left(\boldsymbol{X}^{(k)}\right)\right) \qquad \sum_{i=1}^{k} \|\boldsymbol{X}^{(i)}\|_*$$

    - SNN is a loose convex surrogate of Tucker rank

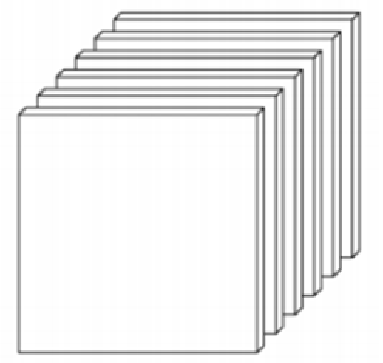
- Recently, we propose a new tensor nuclear norm induced by tensor-tensor product for low tubal rank recovery

Canyi Lu,et al.. Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization. CVPR. 2016.

Notations

- Block circulant matrix of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$

$$\mathtt{bcirc}(\mathcal{A}) = \begin{bmatrix} \boldsymbol{A}^{(1)} & \boldsymbol{A}^{(n_3)} & \cdots & \boldsymbol{A}^{(2)} \\ \boldsymbol{A}^{(2)} & \boldsymbol{A}^{(1)} & \cdots & \boldsymbol{A}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{A}^{(n_3)} & \boldsymbol{A}^{(n_3-1)} & \cdots & \boldsymbol{A}^{(1)} \end{bmatrix}$$



frontal slices $\boldsymbol{A}^{(i)}$

- Two operators

$$\mathtt{unfold}(\mathcal{A}) = \begin{bmatrix} \boldsymbol{A}^{(1)} \\ \boldsymbol{A}^{(2)} \\ \vdots \\ \boldsymbol{A}^{(n_3)} \end{bmatrix}, \ \mathtt{fold}(\mathtt{unfold}(\mathcal{A})) = \mathcal{A}.$$

Tensor-Tensor Product

- Tensor-tensor product is a natural extension of matrix-matrix product.

Definition 1. ***(t-product)*** *[Kilmer and Martin, 2011] Let* $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ *and* $\mathcal{B} \in \mathbb{R}^{n_2 \times l \times n_3}$*. Then the t-product* $\mathcal{A} * \mathcal{B}$ *is defined to be a tensor* $\mathcal{C} \in \mathbb{R}^{n_1 \times l \times n_3}$*,*

$$\mathcal{C} = \mathcal{A} * \mathcal{B} = \mathtt{fold}(\mathtt{bcirc}(\mathcal{A}) \cdot \mathtt{unfold}(\mathcal{B})).$$

Definition 2. ***(Conjugate transpose)*** *[Lu* et al.*, 2016; 2018a] The conjugate transpose of a tensor* $\mathcal{A}$ *of size* $n_1 \times n_2 \times n_3$ *is the* $n_2 \times n_1 \times n_3$ *tensor* $\mathcal{A}^*$ *obtained by conjugate transposing each of the frontal slice and then reversing the order of transposed frontal slices 2 through* n_3*.*

Definition 3. ***(Identity tensor)*** *[Kilmer and Martin, 2011] The identity tensor* $\mathcal{I} \in \mathbb{R}^{n \times n \times n_3}$ *is the tensor whose first frontal slice is the* $n \times n$ *identity matrix, and other frontal slices are all zeros.*

Definition 4. ***(Orthogonal tensor)*** *[Kilmer and Martin, 2011] A tensor* $\mathcal{Q} \in \mathbb{R}^{n \times n \times n_3}$ *is orthogonal if it satisfies*

$$\mathcal{Q}^* * \mathcal{Q} = \mathcal{Q} * \mathcal{Q}^* = \mathcal{I}.$$

Definition 5. ***(F-diagonal Tensor)*** *[Kilmer and Martin, 2011] A tensor is called f-diagonal if each of its frontal slices is a diagonal matrix.*

Misha E Kilmer and Carla D Martin. Factorization strategies for third-order tensors. Linear Algebra and its Applications, 2011

Tensor-SVD

Theorem 1. ***(T-SVD)*** *[Lu* et al.*, 2018a; Kilmer and Martin, 2011] Let* $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$*. Then it can be factored as*

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^*,$$

where $\boldsymbol{\mathcal{U}} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$, $\boldsymbol{\mathcal{V}} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ *are orthogonal, and* $\boldsymbol{\mathcal{S}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ *is a f-diagonal tensor.*

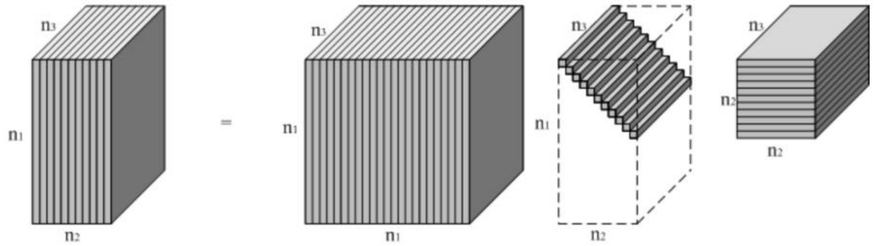


Fig. 2: The t-SVD of an $n_1 \times n_2 \times n_3$ tensor.

Definition 6. ***(Tensor tubal rank)*** *[Lu* et al.*, 2018a] For* $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$*, the tensor tubal rank, denoted as* $\text{rank}_t(\boldsymbol{\mathcal{A}})$*, is defined as the number of nonzero singular values of* $\boldsymbol{\mathcal{S}}$*, where* $\boldsymbol{\mathcal{S}}$ *is from the t-SVD of* $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^*$*. We can write*

$$\text{rank}_t(\boldsymbol{\mathcal{A}}) = \#\{i, \boldsymbol{\mathcal{S}}(i,i,1) \neq 0\} = \#\{i, \boldsymbol{\mathcal{S}}(i,i,:) \neq 0\}.$$

Definition 7. ***(Tensor nuclear norm)*** *[Lu* et al.*, 2018a] Let* $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^*$ *be the t-SVD of* $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$*. The tensor nuclear norm of* $\boldsymbol{\mathcal{A}}$ *is defined as the sum of the tensor singular values, i.e.,* $\|\boldsymbol{\mathcal{A}}\|_* = \sum_{i=1}^{r} \boldsymbol{\mathcal{S}}(i,i,1)$*, where* $r = \text{rank}_t(\boldsymbol{\mathcal{A}})$*.*

Canyi Lu,et al.. Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization. CVPR. 2016.

Problem I: Low-rank Tensor Recovery from Gaussian Measurements

- Given a linear map $\Phi : \mathbb{R}^{n_1 \times n_2 \times n_3} \rightarrow \mathbb{R}^m$ and the observations $\mathbf{y} = \Phi(\mathcal{M})$ for $\mathcal{M} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with tubal rank r.
- Goal: to recover the low-rank tensor $\mathcal{M}$ from the observations $\mathbf{y}$.
- Method: recovery by convex optimization

$$\hat{\mathcal{X}} = \arg\min_{\mathcal{X}} \|\mathcal{X}\|_*, \text{ s.t. } \mathbf{y} = \Phi(\mathcal{X}).$$

- Question: what is the number of measurements m required for exact recovery, i.e., $\hat{\mathcal{X}} = \mathcal{M}$?

Main Result: Low-rank Tensor Recovery from Gaussian Measurements

Theorem 4. *Let* $\Phi : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}^n$ *be a random map with i.i.d. zero-mean Gaussian entries having variance* $\frac{1}{m}$ *and* $\boldsymbol{\mathcal{M}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ *be a tensor of tubal rank* r*. Then, with high probability, we have:*

(1) ***exact recovery:*** $\hat{\boldsymbol{\mathcal{X}}} = \boldsymbol{\mathcal{M}}$, *where* $\hat{\boldsymbol{\mathcal{X}}}$ *is the unique optimum of (3), provided that* $m \geq 3r(n_1 + n_2 - r)n_3 + 1$*;*

(2) ***robust recovery:*** $\|\hat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{M}}\|_F \leq \frac{2\delta}{\epsilon}$, *where* $\hat{\boldsymbol{\mathcal{X}}}$ *is optimal to*

$$\hat{\boldsymbol{\mathcal{X}}} = \arg\min_{\boldsymbol{\mathcal{X}}} \|\boldsymbol{\mathcal{X}}\|_*, \ \ s.t. \ \|\mathbf{y} - \Phi(\boldsymbol{\mathcal{X}})\|_2 \leq \delta, \tag{7}$$

provided that $m \geq \frac{3r(n_1+n_2-r)n_3+3/2}{(1-\epsilon)^2}$.

- For Gaussian measurements, the recovery is exact by convex optimization.
- The required number of measurements is $O(r(n_1 + n_2 - r)n_3)$ which is order optimal.

Problem II: Low-rank Tensor Completion

- Given an incomplete tensor $\mathcal{M} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with tubal rank r.
- Goal: to recover the low-rank tensor $\mathcal{M}$ from partial observations $\mathcal{P}_\Omega(\mathcal{M})$
- Method: recovery by convex optimization

$$\min_{\mathcal{X}} \|\mathcal{X}\|_*, \text{ s.t. } \mathcal{P}_\Omega(\mathcal{X}) = \mathcal{P}_\Omega(\mathcal{M}),$$

- Question: any exact recovery guarantee by convex optimization?

Main Result: Low-rank Tensor Completion

Theorem 6. *Let* $\mathcal{M} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ *with* $\text{rank}_t(\mathcal{M}) = r$ *and the skinny t-SVD be* $\mathcal{M} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*$. *Suppose that the indices* $\boldsymbol{\Omega} \sim Ber(p)$ *and the tensor incoherence conditions (9)-(10) hold. There exist universal constants* $c_0, c_1, c_2 > 0$ *such that if*

$$p \geq \frac{c_0 \mu r \log^2(n_{(1)} n_3)}{n_{(2)} n_3},$$

then $\mathcal{M}$ *is the unique solution to (8) with probability at least* $1 - c_1(n_1 + n_2)^{-c_2}$.

- Exact recovery when the sampling complexity is of the order $O(rn_{(1)}n_3 \log^2(n_{(1)}n_3))$.

Experiment: recovery from Gaussian measurements

$$\hat{\mathcal{X}} = \arg\min_{\mathcal{X}} \|\mathcal{X}\|_*, \text{ s.t. } \mathbf{y} = \Phi(\mathcal{X}).$$

$r = \text{rank}_t(\mathcal{X}_0) = 0.2n$

n	$\text{rank}_t(\mathcal{X}_0)$	m	$\text{rank}_t(\hat{\mathcal{X}})$	$\frac{\|\hat{\mathcal{X}}-\mathcal{X}_0\|_F}{\|\mathcal{X}_0\|_F}$
10	2	541	2	1.2e−9
20	4	2161	4	1.6e−9
30	6	4861	6	1.5e−9

$r = \text{rank}_t(\mathcal{X}_0) = 0.3n$

n	$\text{rank}_t(\mathcal{X}_0)$	m	$\text{rank}_t(\hat{\mathcal{X}})$	$\frac{\|\hat{\mathcal{X}}-\mathcal{X}_0\|_F}{\|\mathcal{X}_0\|_F}$
10	3	766	3	1.6e−9
20	6	3061	6	1.2e−9
30	9	6886	9	1.2e−9

Table 1: Exact low tubal rank tensor recovery from Gaussian measurements with sufficient number of measurements.

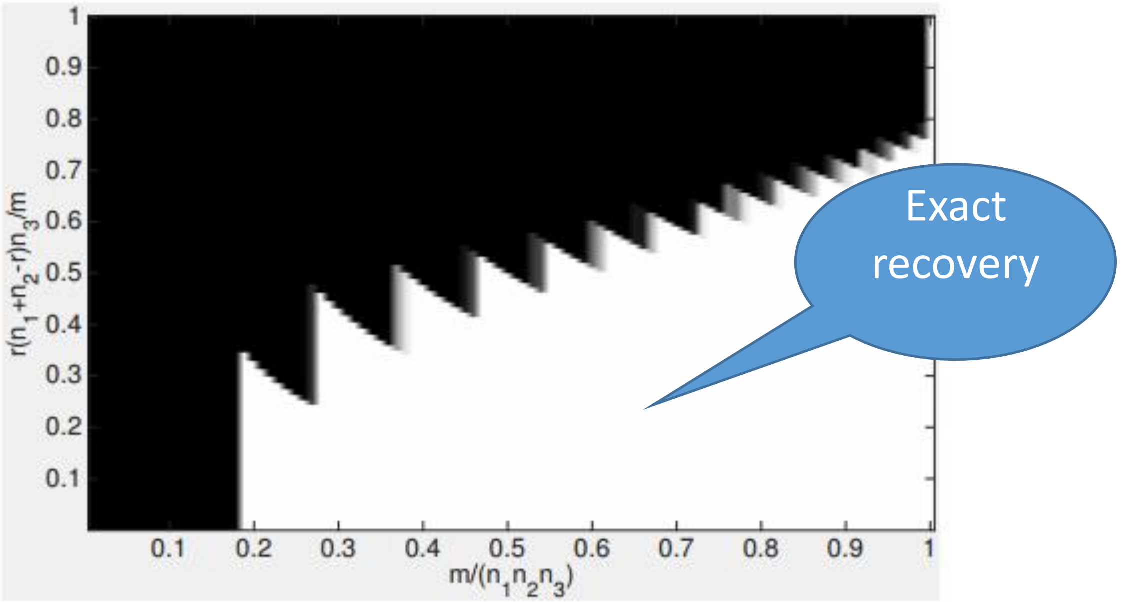


Figure 2: Phase transitions for low tubal rank tensor recovery from Gaussian measurements. Fraction of correct recoveries is across 10 trials, as a function of $\frac{r(n_1+n_2-r)n_3}{m}$ (y-axis) and sampling rate $\frac{m}{n_1 n_2 n_3}$. In this test, $n_1 = n_2 = 30$, $n_3 = 5$.

Experiment: low-rank tensor completion

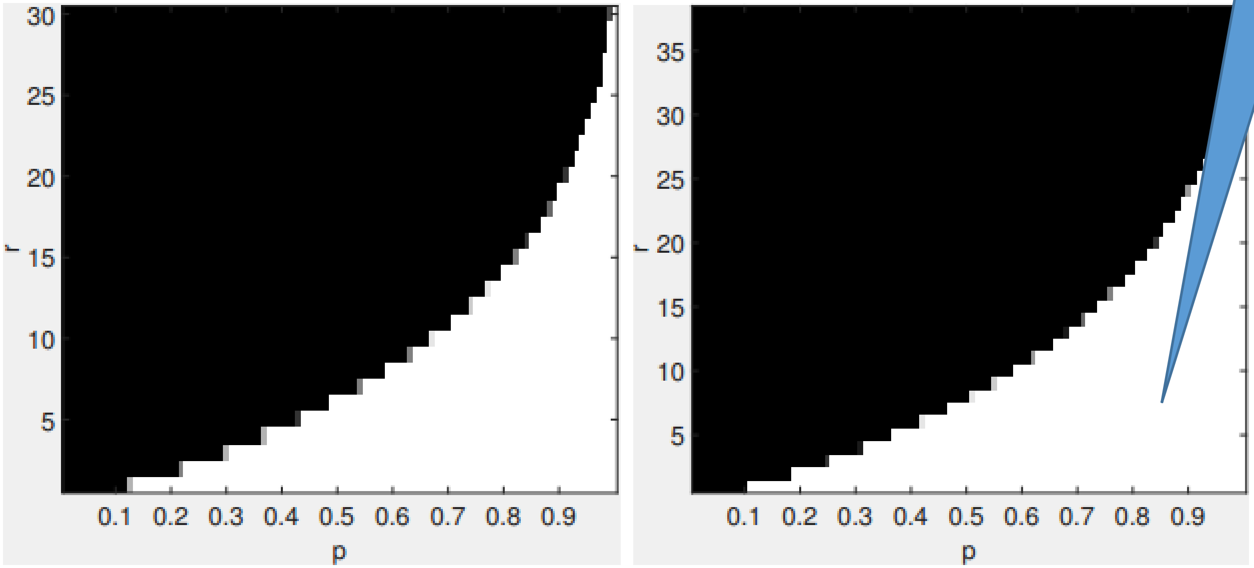
Exact recovery

$$\min_{\mathcal{X}} \|\mathcal{X}\|_*, \text{ s.t. } \mathcal{P}_\Omega(\mathcal{X}) = \mathcal{P}_\Omega(\mathcal{M}),$$

$\mathcal{X}_0 \in \mathbb{R}^{n\times n\times n}$, $r = \text{rank}_t(\mathcal{X}_0)$, $m = pn^3$, $d_r = r(2n - r)$

n	r	$\frac{m}{d_r}$	p	$\text{rank}_t(\hat{\mathcal{X}})$	$\frac{\|\hat{\mathcal{X}}-\mathcal{X}\|_F}{\|\mathcal{X}\|_F}$
50	3	4	0.47	3	3.9e−7
50	5	3	0.57	5	3.5e−7
50	10	2	0.72	10	4.1e−7
100	5	4	0.39	5	1.4e−6
100	10	3	0.57	10	9.2e−7
100	15	2	0.56	15	8.4e−7
200	5	4	0.20	5	4.2e−6
200	10	3	0.29	10	3.2e−6
200	20	2	0.38	20	3.1e−6
300	10	4	0.26	10	5.1e−6
300	20	3	0.39	20	4.2e−6
300	30	3	0.57	30	2.9e−6

Table 2: Exact tensor completion on random data.



(a) $n = 40$ (b) $n = 50$

Figure 3: Phase transitions for tensor completion. Fraction of correct recoveries is across 10 trials, as a function of tubal rank r (y-axis) and sampling rate p (x-axis). The results are shown for different sizes of $\mathcal{M} \in \mathbb{R}^{n\times n\times n}$: (a) $n = 40$; (b) $n = 50$.

Experiment: tensor completion for image recovery

(a) orignal image (b) observed image (c) LRMC (d) LRTC (e) TMac (f) TNN

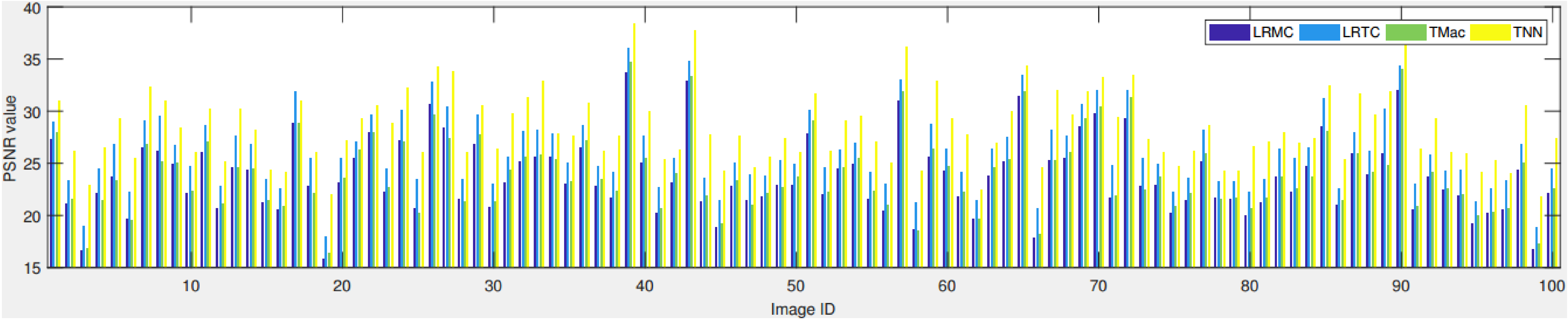
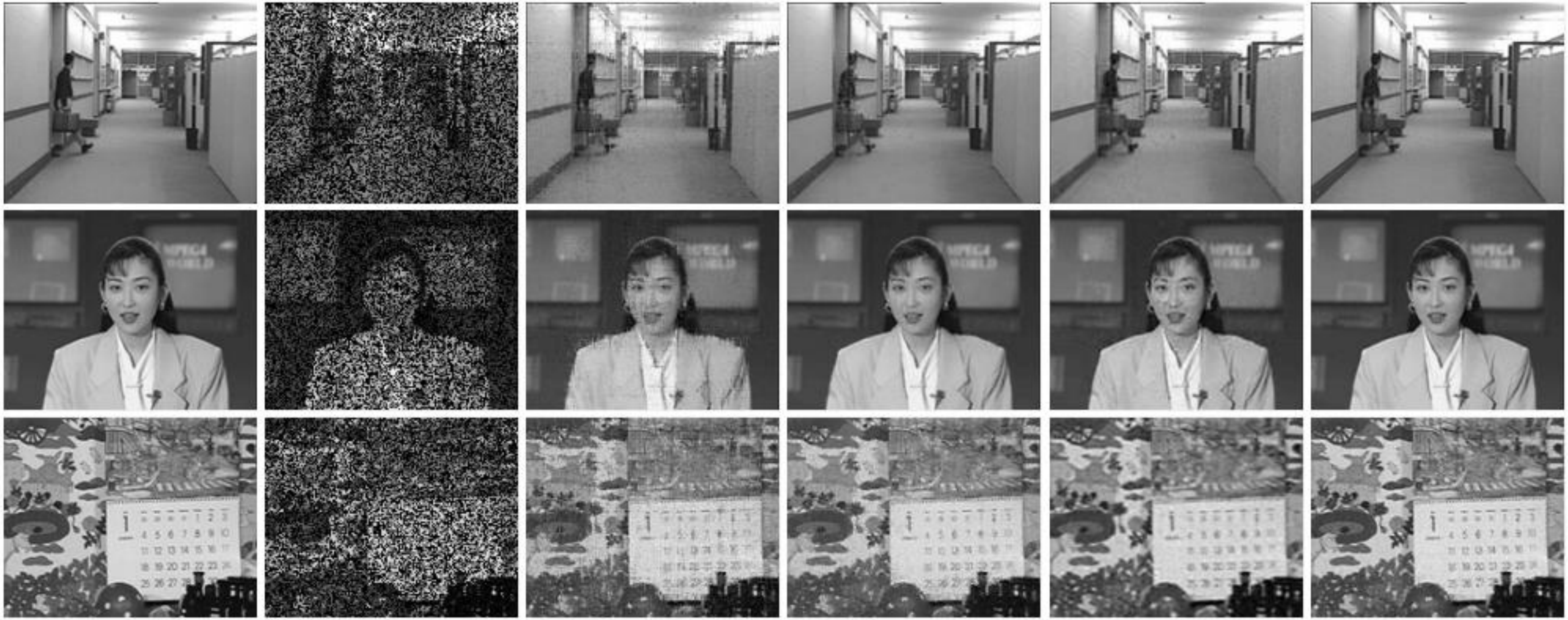


Figure 6: Comparison of the PSNR values obtained by using LRMC, LRTC, TMac and TNN. The rate of observed entries is $p = 0.3$.

Experiment: tensor completion for video recovery



(a) an example frame (b) observed frame (c) LRMC (d) LRTC (e) TMac (f) TNN

Figure 9: Examples for video recovery performance comparison. (a) Example frames from the sequences Coastguard, Hall, Akiyo and Mobile; (b) frames with partially observed entries (the rate is $p = 0.5$); (c)-(f) recovered frames by LRMC, LRTC, TMac and TNN, respectively.

Experiment: tensor completion for video recovery

Table 3: PSNR values of the compared methods.

ID	Videos	LRMC	LRTC	TMac	TNN
1	Highway	13.8	18.0	19.2	**20.8**
2	Coastguard	9.6	11.2	13.1	**17.5**
3	Hall	9.3	17.4	18.7	**22.0**
4	Carphone	10.9	16.7	18.3	**20.3**
5	Bridge (close)	10.5	17.8	17.6	**20.9**
6	News	8.6	15.4	16.7	**20.3**
7	Grandma	11.2	20.1	20.2	**25.7**
8	Suzie	14.5	17.4	19.9	**19.7**
9	Miss America	15.8	21.4	24.8	**25.7**
10	Container	8.4	17.8	17.3	**29.0**
11	Foreman	9.3	14.0	16.1	**18.6**
12	Mother-daughter	12.7	18.8	19.8	**22.9**
13	Silent	11.5	17.6	19.1	**22.9**
14	Akiyo	11.2	20.2	20.4	**27.0**
15	Claire	14.5	23.2	25.7	**27.4**

Conclusions

- Tensor nuclear norm is a recently proposed convex surrogate for the pursuit of tensor tubal rank induced by the tensor-tensor product
- Theoretical guarantee for low tubal rank tensor recovery from Gaussian measurements

$$\hat{\boldsymbol{\mathcal{X}}} = \arg\min_{\boldsymbol{\mathcal{X}}} \|\boldsymbol{\mathcal{X}}\|_*, \text{ s.t. } \mathbf{y} = \Phi(\boldsymbol{\mathcal{X}}).$$

- Theoretical guarantee for low tubal rank tensor completion

$$\min_{\boldsymbol{\mathcal{X}}} \|\boldsymbol{\mathcal{X}}\|_*, \text{ s.t. } \mathcal{P}_\Omega(\boldsymbol{\mathcal{X}}) = \mathcal{P}_\Omega(\boldsymbol{\mathcal{M}}),$$