T-SVD Based Non-convex Tensor Completion and Robust Principal Component Analysis

Tao Li, Jinwen Ma

Department of Information Science, School of Mathematical Sciences and LMAM
Peking University, Beijing, China

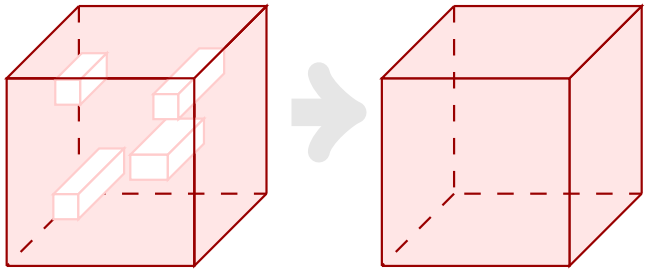
International Conference on Pattern Recognition
2021.1

Tensor Recovery

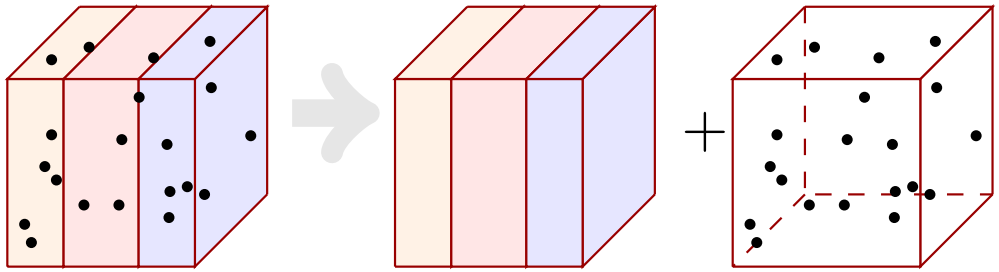
- Tensors, or multi-way arrays, have been extensively used in computer vision, signal processing and machine learning.
- Due to technical reasons, tensors in most applications are incomplete or polluted. Generically, recovering a tensor from corrupted observations is an inverse problem, which is ill-posed without prior knowledge.
- Prior knowledge: low rank.
- But what is rank? Based on T-SVD.

Tensor Recovery

- We mainly consider two tensor recovery problems: tensor completion and tensor robust principal component analysis (TRPCA).
- Tensor completion: estimating the missing values in tensors from partially observed data.
- TRPCA: decomposing a tensor into a low-rank tensor and sparse tensor.



Incomplete tensor Complete tensor



Corrupted tensor Low-rank tensor Sparse tensor

Non-convex Penalties

- As in the matrix case, the choices of rank surrogate function and sparsity measure substantially influence the final results.
- The nuclear norm of a matrix is equivalent to the ℓ_1-norm of its singular value. However, ℓ_1-norm over-penalizes large entries of vectors.
- Smoothly clipped absolute deviation (SCAD) penalty and minimax concave plus (MCP) penalty were proposed as ideal penalty functions.
- This inspires us that nuclear norm based tensor rank surrogate functions and ℓ_1-norm based tensor sparsity measure may suffer from a similar problem.
- To alleviate such phenomena, we propose to use non-convex penalties (SCAD and MCP) instead of an ℓ_1-norm in tensor nuclear norm and tensor sparsity measures.

Solving the Optimization Problem

- The introduction of non-convex penalties makes optimization problems even harder to solve.
- Once we replace ℓ_1-norm by SCAD or MCP, the problem is not convex anymore.
- We apply the majorization minimization algorithm to solve the non-convex optimization problems and analyze the theoretical properties of these algorithms.

Tensor Sigunlar Value Decomposition

Theorem (**T-SVD**)

Suppose $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. *Then there exists tensors* $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}, \mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ *and* $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ *such that* $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*$. *Furthermore,* $\mathcal{U}$ *and* $\mathcal{V}$ *are orthogonal, while* $\mathcal{S}$ *is f-diagonal.*

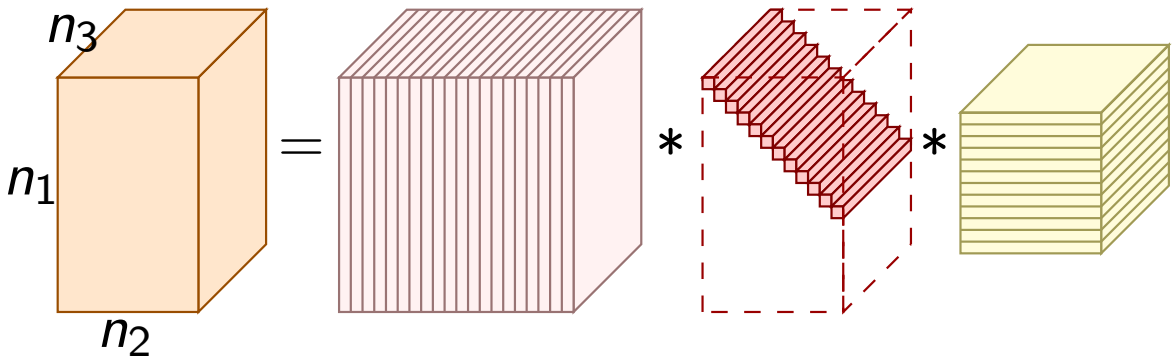


Figure: An illustration of the t-SVD of an $n_1 \times n_2 \times n_3$ tensor.

Definition (**Tensor nuclear norm**)

Let $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*$ *be the t-SVD of* $\mathcal{A}$, *the nuclear norm of* $\mathcal{A}$ *is defined as* $\|\mathcal{A}\|_* = \sum_i \mathcal{S}(i,i,1) = \frac{1}{n_3} \sum_k \overline{\mathcal{S}}(i,i,k)$.

Non-convex Penalties

Definition (**SCAD**)

For some $\gamma > 2$ and $\lambda > 0$, the SCAD function is given by

$$\varphi_{\lambda,\gamma}^{\mathrm{SCAD}}(t) = \begin{cases} \lambda|t| & \mathrm{if}|t| \leq \lambda, \\ \frac{\gamma\lambda|t| - 0.5(t^2+\lambda^2)}{\gamma-1} & \mathrm{if}\lambda < |t| < \gamma\lambda, \\ \frac{\gamma+1}{2}\lambda^2 & \mathrm{if}|t| > \gamma\lambda. \end{cases} \tag{1}$$

Definition (**MCP**)

For some $\gamma > 1$ and $\lambda > 0$, the MCP function is given by

$$\varphi_{\lambda,\gamma}^{\mathrm{MCP}}(t) = \begin{cases} \lambda|t| - \frac{t^2}{2\gamma} & \mathrm{if}|t| < \gamma\lambda, \\ \frac{\gamma\lambda^2}{2} & \mathrm{if}|t| \geq \gamma\lambda. \end{cases} \tag{2}$$

A Novel Tensor Sparsity Measure

The novel tensor sparsity measure is defined as

$$\Phi_{\lambda,\gamma}(\mathcal{A}) = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\varphi_{\lambda,\gamma}(\mathcal{A}_{ijk}). \tag{3}$$

Proposition

For $\mathcal{A} \in \mathbb{R}^{n_1\times n_2\times n_3}$, $\Phi_{\lambda,\gamma}(\mathcal{A})$ satisfies:

(i) $\Phi_{\lambda,\gamma}(\mathcal{A}) \geq 0$ *with the equality holds iff* $\mathcal{A} = 0$*;*

(ii) $\Phi_{\lambda,\gamma}(\mathcal{A})$ *is concave with respect to* $|\mathcal{A}|$*;*

(iii) $\Phi_{\lambda,\gamma}(\mathcal{A})$ *is increasing in* γ*,* $\Phi_{\lambda,\gamma}(\mathcal{A}) \leq \lambda\|\mathcal{A}\|_1$ *and* $\lim_{\gamma\to\infty}\Phi_{\lambda,\gamma}(\mathcal{A}) = \lambda\|\mathcal{A}\|_1$.

A Novel Tensor Rank Penalty

Suppose $\mathcal{A}$ has t-SVD $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*$, we define the $\gamma-$norm of $\mathcal{A}$ as

$$\|\mathcal{A}\|_\gamma = \frac{1}{n_3} \sum_{i,k} \varphi_{1,\gamma}(\overline{\mathcal{S}}(i,i,k)). \tag{4}$$

Proposition

For $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, suppose $\mathcal{A}$ has t-SVD $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*$, then $\|\mathcal{A}\|_\gamma$ satisfies:

(i) $\|\mathcal{A}\|_\gamma \geq 0$ *with equality holds iff* $\mathcal{A} = 0$;

(ii) $\|\mathcal{A}\|_\gamma$ *is increasing in* γ, $\|\mathcal{A}\|_\gamma \leq \|\mathcal{A}\|_*$ *and* $\lim_{\gamma \to \infty} \|\mathcal{A}\|_\gamma = \|\mathcal{A}\|_*$;

(iii) $\|\mathcal{A}\|_\gamma$ *is concave with respect to* $\{\overline{\mathcal{S}}(i,i,k)\}_{i,k}$;

(iv) $\|\mathcal{A}\|_\gamma$ *is orthogonal invariant, i.e., for any orthogonal tensors* $\mathcal{P} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$, $\mathcal{Q} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$, *we have* $\|\mathcal{P} * \mathcal{A} * \mathcal{Q}\|_\gamma = \|\mathcal{A}\|_\gamma$.

Majorization Functions

Theorem

We can view $\Phi_{\lambda,\mu}(\mathcal{X})$ as a function of $|\mathcal{X}|$, and $\|\mathcal{X}\|_\gamma$ as a function of $\{\overline{\mathcal{S}}(i,i,k)\}_{i,k}$. For any $\mathcal{X}^{\text{old}}$, let

$$\begin{aligned} Q_{\lambda,\gamma}(\mathcal{X}|\mathcal{X}^{\text{old}}) &= \Phi_{\lambda,\gamma}(\mathcal{X}^{\text{old}}) + \sum_{i,j,k} \varphi'_{\lambda,\gamma}(|\mathcal{X}^{\text{old}}_{ijk}|)(|\mathcal{X}_{ijk}| - |\mathcal{X}^{old}_{ijk}|), \\ Q_\gamma(\mathcal{X}|\mathcal{X}^{\text{old}}) &= \|\mathcal{X}^{\text{old}}\|_\gamma + \frac{1}{n_3}\sum_{i,k} \varphi'_{1,\gamma}(\overline{\mathcal{S}}^{\text{old}}_{iik})(\overline{\mathcal{S}}_{iik} - \overline{\mathcal{S}}^{\text{old}}_{iik}), \end{aligned} \tag{5}$$

then

$$\begin{aligned} Q_{\lambda,\gamma}(\mathcal{X}^{\text{old}}|\mathcal{X}^{\text{old}}) = \Phi_{\lambda,\gamma}(\mathcal{X}^{\text{old}}), \Phi_{\lambda,\gamma}(\mathcal{X}) \le Q_{\lambda,\gamma}(\mathcal{X}|\mathcal{X}^{\text{old}}), \\ Q_\gamma(\mathcal{X}^{\text{old}}|\mathcal{X}^{\text{old}}) = \|\mathcal{X}^{\text{old}}\|_\gamma, \|\mathcal{X}\|_\gamma \le Q_\gamma(\mathcal{X}|\mathcal{X}^{\text{old}}). \end{aligned} \tag{6}$$

Generalized Soft Thresholding

Definition (Generalized soft thresholding)

Suppose $\mathcal{X}, \mathcal{W} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, *the generalized soft thresholding operator is defined as*

$$[\mathcal{T}_{\mathcal{W}}(\mathcal{X})]_{ijk} = \mathcal{T}_{\mathcal{W}_{ijk}}(\mathcal{X}_{ijk}). \tag{7}$$

Theorem

For $\forall \mu > 0$, *let* $\mathcal{W}_{ijk} = \varphi'_{\lambda,\gamma}(|\mathcal{X}^{\text{old}}_{ijk}|)/\mu$, *then*

$$\mathcal{T}_{\mathcal{W}}(\mathcal{Y}) = \arg\min_{\mathcal{X}} Q_{\lambda,\gamma}(\mathcal{X}|\mathcal{X}^{\text{old}}) + \frac{\mu}{2}\|\mathcal{X} - \mathcal{Y}\|_F^2. \tag{8}$$

Generalized Tensor Singular Value Thresholding

Definition (**Generalized t-SVT**)

Suppose a 3-way tensor $\mathcal{Y}$ has t-SVD $\mathcal{Y} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*$, $\mathcal{W}$ is a tensor with the same shape of $\mathcal{Y}$, the generalized tensor singular value thresholding operator is defined as

$$\mathcal{D}_{\mathcal{W}}(\mathcal{Y}) = \mathcal{U} * \tilde{\mathcal{S}} * \mathcal{V}^*, \tag{9}$$

where $\tilde{\mathcal{S}} = \texttt{ifft}(\mathcal{T}_{\mathcal{W}}(\mathcal{S}), 3)$.

Theorem

For $\forall \mu > 0$, let $\mathcal{W}_{ijk} = \delta_i^j \varphi'_{1,\gamma}(\overline{\mathcal{S}}_{iik}^{\text{old}})/\mu$ where δ_i^j is the Kronecker symbol, then

$$\mathcal{D}_{\mathcal{W}}(\mathcal{Y}) = \arg\min_{\mathcal{X}} Q_\gamma(\mathcal{X}|\mathcal{X}^{\text{old}}) + \frac{\mu}{2}\|\mathcal{X} - \mathcal{Y}\|_F^2. \tag{10}$$

Non-convex Tensor Completion

- Given a partially observed tensor $\mathcal{O} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, tensor completion task aims to recover the full tensor $\mathcal{X}$ which coincides with $\mathcal{O}$ in the observed positions. Suppose the observed positions are indexed by Ω, *i.e.*, $\Omega_{ijk} = 1$ denotes the $(i, j, k)-$th element is observed while $\Omega_{ijk} = 0$ denotes the $(i, j, k)-$th element is unknown. Based on low rank assumption, tensor completion can be modeled as

$$\min_{\mathcal{X}} \operatorname{rank}(\mathcal{X}) \quad \text{s.t. } \mathcal{O}_\Omega = \mathcal{X}_\Omega. \tag{11}$$

- Use the proposed tensor $\gamma-$norm to replace "rank" :

$$\min_{\mathcal{X}} \|\mathcal{X}\|_\gamma \quad \text{s.t. } \mathcal{O}_\Omega = \mathcal{X}_\Omega. \tag{12}$$

- Majorization Minimization:

$$\min_{\mathcal{X}} Q_\gamma(\mathcal{X}|\mathcal{X}^{\text{old}}) \quad \text{s.t. } \mathcal{O}_\Omega = \mathcal{X}_\Omega. \tag{13}$$

Non-convex Tensor Completoin

Theorem

The iteration sequence generated by $\mathcal{X}^{t+1} \in \arg\min_{\mathcal{O}_\Omega = \mathcal{X}_\Omega} Q_\gamma(\mathcal{X}|\mathcal{X}^t)$ *is non-increasing, i.e.,* $\|\mathcal{X}^{t+1}\|_\gamma \leq \|\mathcal{X}^t\|_\gamma$ *and converges to some* Q^**. Besides, there exists a subsequence* $\{\mathcal{X}^{i_k}\}_{k=1}^{\infty}$ *converges to a minimal point* $\mathcal{X}_*$ *of* $\|\mathcal{X}\|_\gamma$ *on* $\{\mathcal{X}|\mathcal{O}_\Omega = \mathcal{X}_\Omega\}$*.*

Non-convex TRPCA

- Given a tensor $\mathcal{X}$, the goal of robust PCA is to decompose $\mathcal{X}$ into two parts: low-rank tensor $\mathcal{L}$ and sparse tensor $\mathcal{E}$. This problem can be formulated as

$$\min_{\mathcal{L},\mathcal{E}} \text{rank}(\mathcal{L}) + \|\mathcal{E}\|_0 \quad \text{s.t. } \mathcal{L} + \mathcal{E} = \mathcal{X}. \tag{14}$$

- Apply the proposed novel sparsity measure and tensor $\gamma-$norm, we obtain

$$\min_{\mathcal{L},\mathcal{E}} \|\mathcal{L}\|_{\gamma_1} + \Phi_{\lambda,\gamma_2}(\mathcal{E}) \quad \text{s.t. } \mathcal{L} + \mathcal{E} = \mathcal{X}. \tag{15}$$

- Majorization minimization:

$$\min_{\mathcal{L},\mathcal{E}} Q_{\gamma_1}(\mathcal{L}|\mathcal{L}^{\text{old}}) + Q_{\lambda,\gamma_2}(\mathcal{E}|\mathcal{E}^{\text{old}}) \quad \text{s.t. } \mathcal{L} + \mathcal{E} = \mathcal{X}. \tag{16}$$

Non-convex TRPCA

Theorem

The iteration sequence generated by

$$(\mathcal{L}^{t+1}, \mathcal{E}^{t+1}) \in \arg\min_{\mathcal{L}+\mathcal{E}=\mathcal{X}} Q_{\gamma_1}(\mathcal{L}|\mathcal{L}^t) + Q_{\lambda,\gamma_2}(\mathcal{E}|\mathcal{E}^t)$$

is non-increasing, i.e., $\|\mathcal{L}^{t+1}\|_{\gamma_1} + \Phi_{\lambda,\gamma_2}(\mathcal{E}^{t+1}) \leq \|\mathcal{L}^t\|_{\gamma_1} + \Phi_{\lambda,\gamma_2}(\mathcal{E}^t)$ *and converges to some* Q^**. Besides, there exists a subsequence* $\{(\mathcal{L}^{i_k}, \mathcal{E}^{i_k})\}_{k=1}^{\infty}$ *converges to a minimal point* $(\mathcal{L}_*, \mathcal{E}_*)$ *of* $\|\mathcal{L}\|_{\gamma_1} + \Phi_{\lambda,\gamma_2}(\mathcal{E})$ *on* $\{(\mathcal{L}, \mathcal{E})|\mathcal{L} + \mathcal{E} = \mathcal{X}\}$.

Experimental Results

TABLE I

TENSOR COMPLETION PERFORMANCES EVALUATION ON NATURAL IMAGES UNDER VARYING SAMPLING RATES.

Method	20%			40%			60%			80%			time (s)
	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	
SiLRTC	23.59	0.798	0.822	27.987	0.899	0.915	32.24	0.951	0.964	37.47	**0.977**	0.988	19.95
HaLRTC	23.82	0.797	0.828	28.39	0.902	0.920	33.038	0.953	0.968	39.27	**0.978**	0.991	31.32
FBCP	24.08	0.668	0.794	26.40	0.753	0.837	27.35	0.799	0.857	27.71	0.82	0.865	103.68
t-SVD	24.13	0.764	0.835	29.703	0.893	0.931	36.03	0.950	0.977	45.04	0.969	0.992	33.47
t-TNN	25.30	0.841	0.864	30.50	0.923	0.943	36.27	0.952	0.978	44.14	0.967	0.991	**3.03**
LRTC_{mcp}	**25.70**	**0.845**	**0.869**	**31.06**	**0.927**	**0.946**	**36.87**	**0.959**	**0.980**	**45.46**	0.973	**0.993**	3.79
LRTC_{scad}	**25.70**	**0.844**	**0.869**	**31.04**	**0.926**	**0.946**	**36.84**	**0.959**	**0.980**	**45.45**	0.973	**0.993**	3.83

TABLE II

TENSOR COMPLETION PERFORMANCES EVALUATION ON HYPERSPECTRAL IMAGES UNDER VARYING SAMPLING RATES. THE UNIT IS 10^{-4} FOR MSE.

Method	20%			40%			60%			80%			time (s)
	PSNR	MSE	ERGAS	PSNR	MSE	ERGAS	PSNR	MSE	ERGAS	PSNR	MSE	ERGAS	
SiLRTC	41.71	4.70	30.912	45.46	1.95	21.524	49.14	0.827	14.412	52.86	0.354	10.341	42.21
HaLRTC	42.11	4.53	29.626	45.95	1.90	20.556	49.79	0.801	13.514	53.67	0.342	9.660	53.11
FBCP	37.09	14.47	52.931	43.25	3.85	29.318	46.00	2.225	22.344	46.67	2.011	20.688	210.18
t-SVD	41.64	5.10	31.835	45.52	2.12	22.142	49.42	0.886	14.685	53.49	0.365	10.157	224.21
t-TNN	42.46	3.81	28.702	46.07	1.60	20.135	49.82	0.667	**13.272**	53.61	**0.290**	**9.515**	**47.29**
LRTC_{mcp}	**42.91**	**3.58**	**27.799**	**46.75**	**1.51**	**19.684**	**50.47**	**0.665**	**13.293**	**54.11**	**0.316**	**9.642**	70.95
$\text{LRTC}_{\text{scad}}$	**42.91**	**3.58**	**27.804**	**46.76**	**1.51**	**19.684**	**50.46**	**0.666**	13.298	**54.11**	**0.316**	**9.642**	71.14

Experimental Results

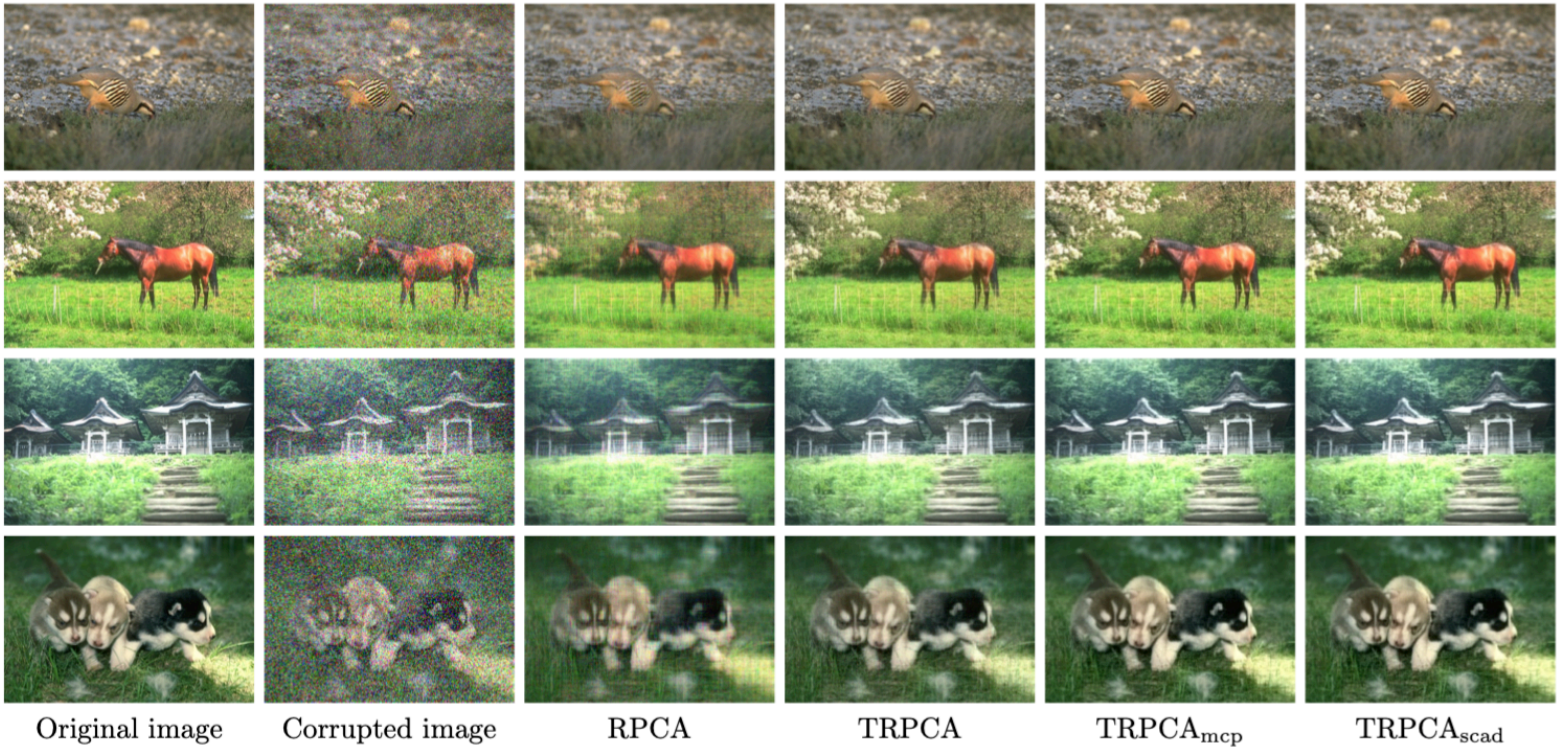


Fig. 4. Tensor RPCA performance comparison on example images. From top to bottom: $p_n = 0.1, 0.2, 0.3, 0.4$.

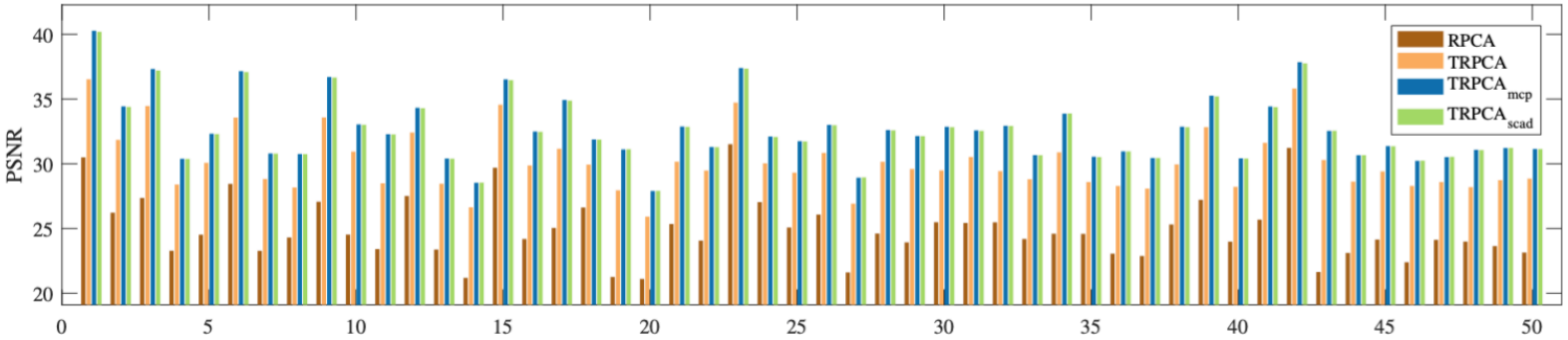


Fig. 5. Comparison of PSNR values obtained by RPCA, TRPCA, TRPCA$_{mcp}$,TRPCA$_{scad}$ on randomly selected 50 images.