Robust PCA and Tensor RPCA

Robust PCA by convex optimization
\[ \min_{||L||_1} \lambda ||S||_1, \text{ s.t. } X = L + S. \]


Tensor RPCA: given \( \mathbf{X} \in \mathbb{R}^{m \times n \times k} \), how to recover the underlying low-rank tensor \( \mathbf{L} \) and sparse tensor \( \mathbf{S} \) perfectly?
\[ \min_{||\mathbf{L}||_1, ||\mathbf{S}||_1} \lambda ||\mathbf{S}||_1, \text{ s.t. } \mathbf{X} = \mathbf{L} + \mathbf{S}. \]  

(1)

Notations and Preliminaries

- For \( \mathbf{A} \in \mathbb{R}^{m \times n} \), denote \( \mathbf{A}^{(i)} \) as the \( i \)-th frontal slice of \( \mathbf{A} \).
- Block circulant matrix \( \text{circ}(\mathbf{A}) = \begin{bmatrix} \mathbf{A} & \mathbf{A}^{(2)} & \cdots & \mathbf{A}^{(n)} \end{bmatrix} \in \mathbb{R}^{n \times nm} \).
- A key property: block circulant matrix can be block diagonalized in the Fourier domain, i.e., \( \text{circ}(\mathbf{A}) \) = \( \mathbf{P} \text{diag}(\hat{\mathbf{A}}) \mathbf{P}^T \), where \( \mathbf{P} \) is the Discrete Fourier Transform (DFT) matrix with property \( \mathbf{F}^H \mathbf{F} = n \mathbf{I} = \mathbf{F} \mathbf{F}^H \).

Tensor Nuclear Norm

Tensor nuclear norm: \( \| \mathbf{A} \|_{nuc} = \max_{\mathbf{U}, \mathbf{V}} \langle \mathbf{U}, \mathbf{A}, \mathbf{V} \rangle \), subject to \( \mathbf{A} = \mathbf{U} \mathbf{V}^H \) is the unique minimizer to (1) with \( \mathbf{A} = \mathbf{L} + \mathbf{S} \), and that \( \mathbf{L} \) has low rank and \( \mathbf{S} \) is sparse. How to perfectly recover both \( \mathbf{L} \) and \( \mathbf{S} \) from \( \mathbf{X} \)?

Tensor Incoherence Conditions. For \( \mathbf{L} \in \mathbb{R}^{m \times n} \), assume that \( \| \mathbf{L} \|_F = r \) and it has the skinny SVD \( \mathbf{L} = \mathbf{U} \mathbf{S} \mathbf{V}^H \), where \( \mathbf{U} \in \mathbb{R}^{m \times m} \) and \( \mathbf{V} \in \mathbb{R}^{n \times n} \) satisfy \( \mathbf{U}^H \mathbf{U} = \mathbf{I} \) and \( \mathbf{V}^H \mathbf{V} = \mathbf{I} \), and \( \mathbf{S} \in \mathbb{R}^{m \times n} \) is f-diagonal. Then \( \mathbf{L} \) is said to satisfy the tensor incoherence conditions with parameter \( \rho \).

\[ \begin{aligned} & \max_{i \neq j} \langle \mathbf{U}(:,i), \mathbf{V}(:,j) \rangle \leq \frac{\sqrt{rn}}{\sqrt{m+n}}, \\ & \max_{i \neq j} \| \mathbf{U}(:,i) \|_2 \cdot \| \mathbf{V}(:,j) \|_2 \leq \sqrt{mn} \rho, \\ & \| \mathbf{U}^H \mathbf{V} \|_F \leq \sqrt{mn} \rho, \\ & \| \mathbf{U} \|_F \leq \sqrt{m} \rho, \\ & \| \mathbf{V} \|_F \leq \sqrt{n} \rho, \\ & \| \mathbf{S} \|_F \leq \sqrt{mn} \rho. \end{aligned} \]

Define \( \beta = \max(n, r) \) and \( \gamma = \min(n, r) \). Then \( \beta \leq \mu \leq \sqrt{mn} \).

Theorem: Suppose \( \mathbf{L} \in \mathbb{R}^{m \times n} \) obeys (2)-(4). Fix any \( m \times n \) tensor \( \mathbf{A} \) of signs. Suppose that the support set \( \Omega \) of \( \mathbf{S} \) is uniformly distributed among all sets of cardinality \( m \) and that \( \| \mathbf{S}(\bar{x}) \|_F = \| \mathbf{A}(\bar{x}) \|_F \) for all \( (i, j, k) \in \Omega \). Then, there exist universal constants \( c_1, c_2 > 0 \) such that with probability at least \( 1 - 4^{-n/2} \) (over the choice of support of \( \mathbf{S} \), \( \mathbf{L} \), \( \mathbf{S} \) is the unique minimizer to (1) with \( \lambda = 1/\sqrt{m} \), provided that
\[ \frac{\mu \sqrt{n}}{\sqrt{m} \rho}, \text{ and } m \leq \rho \beta^2, \]
where \( \rho \) and \( \beta \) are positive constants. If \( \mathbf{L} \in \mathbb{R}^{m \times n} \) has rectangular frontal slices, TRPCA with \( \lambda = 1/\sqrt{mn} \) succeeds with probability at least \( 1 - 4^{-n/2} \), provided that \( \| \mathbf{L} \|_F \leq \sqrt{mn} \rho / \sqrt{m+n} \), and \( m \leq \rho \beta^2 \).

Remarks: (1) The perfect recovery guarantee is identical to RPCA. (2) When \( \gamma = 1, \) TRPCA and its recovery guarantee reduce to RPCA.

Optimization by ADMM

- The main per-iteration cost lies in the update of \( \mathbf{A} \), which requires computing FFT and \( n \) SVDs of \( n \times n \) matrices.
- The per-iteration complexity of ADMM for TRPCA is \( O \left( mn \beta \log(n) + \beta \gamma \right) \).