

Solving Low-Rank Representation Through Deep Newton Networks

Xianchao Xiu

Department of Automation



Operations Research Society of China, October 16-18, 2026

Joint work with [H. Chen](#) (SUEP), [X. Wang](#) (SUEP), [R. Yu](#) (SUEP) and [W. Liu](#) (SYSU)

Outline

Introduction

Proposed Methods

Numerical Experiments

Conclusions and Future Work

LRR

- Liu-Lin-Yan et al, Robust Recovery of Subspace Structures by Low-Rank Representation, *IEEE TPAMI*, 2013

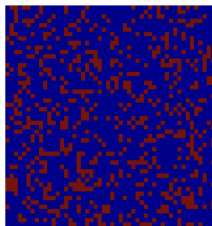
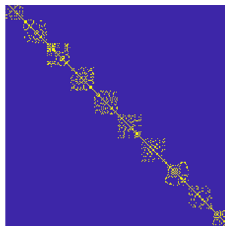
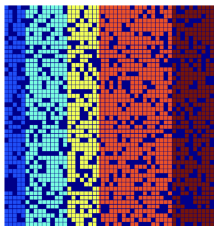
$$\min_{Z, E} \text{rank}(Z) + \lambda \|E\|_0$$

$$\text{s.t. } X = XZ + E$$

$\Rightarrow$

$$\min_{Z, E} \|Z\|_* + \lambda \|E\|_1$$

$$\text{s.t. } X = XZ + E$$



- Broad applications: clustering, classification, denoising, detection, fusion
- Related to low-rank regularization, low-rank approximation, low-rank decomposition

- Gregor-LeCun, Learning Fast Approximations of Sparse Coding, [ICML](#), 2010

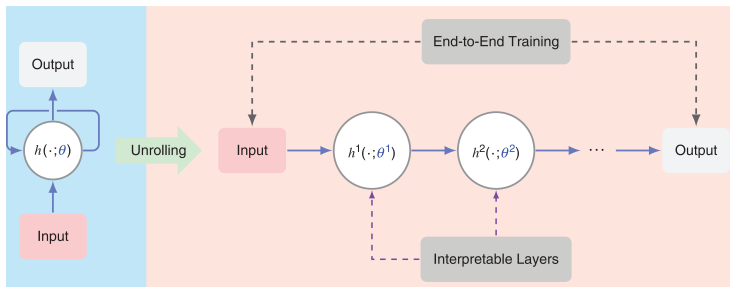
$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|Ax - b\|^2 + \lambda \|x\|_1$$

$$\Downarrow$$

$$\text{(ISTA)} \quad x^{k+1} = \text{Prox}_{\lambda \|\cdot\|_1}(x^k - \rho A^\top (Ax^k - b))$$

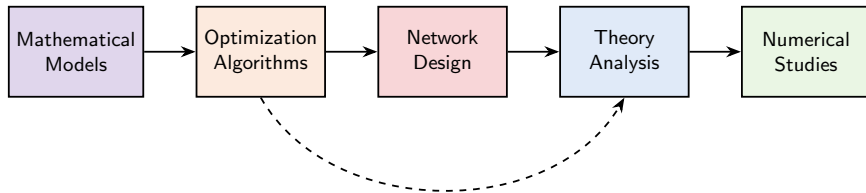
$$\Downarrow$$

$$\text{(LISTA)} \quad x^{k+1} = h_{\theta^k}(x^k - \rho A^\top (Ax^k - b))$$



Motivation

- ▶ Chen-Liu-Wang et al, Theoretical Linear Convergence of Unfolded ISTA and Its Practical Weights and Thresholds, [NeurIPS](#), 2018
- ▶ Yang-Sun-Li et al, ADMM-CSNet: A Deep Learning Approach for Image Compressive Sensing, [IEEE TPAMI](#), 2020
- ▶ Chen-Chen-Chen et al, Learning to Optimize: A Primer and A Benchmark, [JMLR](#), 2022
- ▶ Song-Zhang-Zhou et al, SOR-Unrolled Learning-to-Solve Method for Absolute Value Equations, [JORSC](#), 2026
- ▶ Xiu-He-Qu et al, A Survey of Deep Unfolding Methods for Image Compressive Sensing (in Chinese), [ORT](#), 2026



Outline

Introduction

Proposed Methods

Numerical Experiments

Conclusions and Future Work

Model

- Consider the following LRR model

$$\begin{aligned} \min_{Z, E} \quad & \|Z\|_* + \lambda \|E\|_1 \\ \text{s.t.} \quad & X = XZ + E \end{aligned}$$

- Augmented Lagrangian method (ALM)

$$\begin{aligned} \min_{J, E, Z} \quad & \|J\|_* + \lambda \|E\|_1 \\ \text{s.t.} \quad & X = XZ + E, \quad Z = J \end{aligned}$$

which derives

$$\begin{aligned} \mathcal{L}_\beta(J, E, Z, Y_1, Y_2) = & \|J\|_* + \lambda \|E\|_1 + \langle Y_1, Z - J \rangle \\ & + \frac{\beta}{2} \|Z - J\|_F^2 + \langle Y_2, X - XZ - E \rangle + \frac{\beta}{2} \|X - XZ - E\|_F^2 \end{aligned}$$

Algorithm

- Update variables according to

$$\begin{cases} (J, E, Z) \leftarrow \arg \min_{J, E, Z} \mathcal{L}_\beta(J, E, Z, Y_1, Y_2) \\ Y_1 \leftarrow Y_1 + \beta(Z - J) \\ Y_2 \leftarrow Y_2 + \beta(X - XZ - E) \end{cases}$$

- First-order optimality conditions

$$\begin{cases} 0 \in \partial \|J\|_* - Y_1 + \beta(J - Z) \\ 0 \in \lambda \partial \|E\|_1 - Y_2 - \beta(X - XZ - E) \\ 0 = Y_1 + \beta(Z - J) - X^\top Y_2 - \beta X^\top (X - XZ - E) \end{cases}$$

which derives

$$J = \text{Prox}_{1/\beta, \|\cdot\|_*} (Z + Y_1/\beta), \quad E = \text{Prox}_{\lambda/\beta, \|\cdot\|_1} (X - XZ + Y_2/\beta)$$

SSN

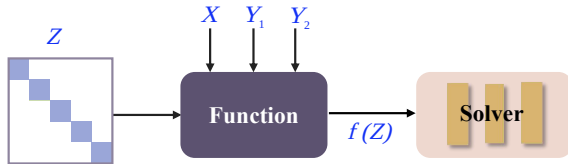
- Denote the nonlinear equation

$$f(Z) = Y_1 + \beta (Z - \text{Prox}_{1/\beta, \|\cdot\|_*}(Z + Y_1/\beta)) - X^\top Y_2 \\ - \beta X^\top (X - XZ - \text{Prox}_{\lambda/\beta, \|\cdot\|_1}(X - XZ + Y_2/\beta))$$

$\Downarrow$

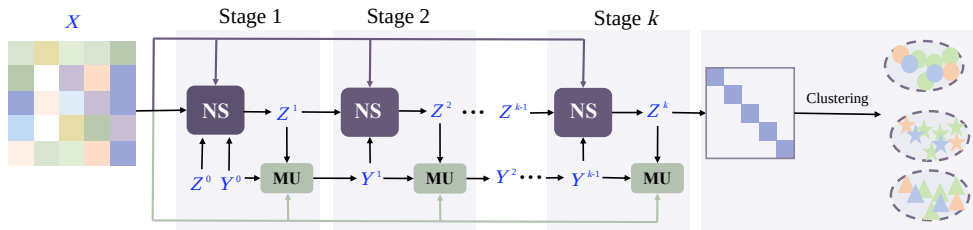
$$\begin{cases} g(Z)\Delta Z = -f(Z) \\ Z \leftarrow Z + \eta\Delta Z \end{cases}$$

- Newton solver (NS)



LRRDNN

► Overall architecture



► Training loss

$$\ell_{\text{total}} = \ell_{\text{self}} + \sigma_1 \ell_{\text{fun}} + \sigma_2 \ell_{\text{reg}}$$

$$= \left(\|X - XZ^K - E^K\|_F^2 + \gamma \|E^K\|_1 \right) + \sigma_1 \left(\frac{1}{N} \sum_{l=1}^N \|f(Z^K)\|_1 \right) + \sigma_2 \text{tr}(L_{Z^K})$$

LRRDNN+

- How to construct regularization terms?

$$\begin{array}{ll} \min_{Z,E} & \|Z\|_* + \lambda \|E\|_1 \\ \text{s.t.} & X = XZ + E \end{array} \quad \Rightarrow \quad \begin{array}{ll} \min_{Z,E} & \mathcal{R}(Z) + \lambda \mathcal{S}(E) \\ \text{s.t.} & X = XZ + E \end{array}$$

- Similar ALM and first-order optimality conditions

$$\begin{cases} 0 \in \partial \mathcal{R}(J) - Y_1 + \beta(J - Z) \\ 0 \in \lambda \partial \mathcal{S}(E) - Y_2 - \beta(X - XZ - E) \\ 0 = Y_1 + \beta(Z - J) - X^\top Y_2 - \beta X^\top (X - XZ - E) \end{cases}$$

which derives

$$J = \text{Prox}_{1/\beta, \mathcal{R}}(Z + Y_1/\beta), \quad E = \text{Prox}_{\lambda/\beta, \mathcal{S}}(X - XZ + Y_2/\beta)$$

LRRDNN+

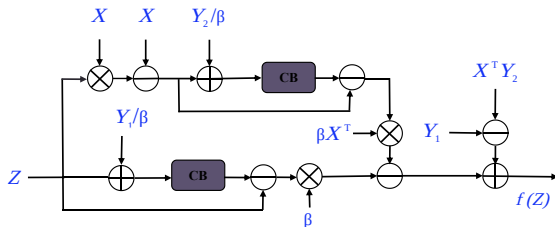
- Denote the nonlinear equation

$$f(Z) = Y_1 + \beta (Z - \text{Prox}_{1/\beta, \mathcal{R}}(Z + Y_1/\beta)) - X^\top Y_2 - \beta X^\top (X - XZ - \text{Prox}_{\lambda/\beta, \mathcal{S}}(X - XZ + Y_2/\beta))$$

$\Downarrow$

$$\begin{cases} g(Z)\Delta Z = -f(Z) \\ Z \leftarrow Z + \eta\Delta Z \end{cases}$$

- Convolutional block (CB)



Outline

Introduction

Proposed Methods

Numerical Experiments

Conclusions and Future Work

Experiments

- ▶ Compared methods
 - ▶ TLRR: Zhou-Lu-Feng et al, [IEEE TPAMI](#), 2021
 - ▶ ETLRR: Du-Liu-Shan et al, [KBS](#), 2022
 - ▶ IDR: Wei-Liu-Zhou et al, [IEEE TKDE](#), 2023
 - ▶ TS1-LLRR: Hu-Qu-Liu et al, [KBS](#), 2024
 - ▶ JSSC: Zhu-Xiu-Liu et al, [ESWA](#), 2025
 - ▶ TFLRR: Liu-Hu-Wang, [IEEE TSIPN](#), 2025
 - ▶ BTLRR: Ding-Zhao-Yang et al, [SIIMS](#), 2025
 - ▶ ISC: Zhang-Zhou-Yao et al, [IEEE TPAMI](#), 2026
- ▶ Implementation details
 - ▶ Hyperparameters: $\sigma_1 = 0.01$, $\sigma_2 = 1$, $K=6$
 - ▶ Learnable parameters: λ , β , $\mathcal{R}(J)$, $\mathcal{S}(E)$,
 - ▶ Clustering: k -means
 - ▶ Metrics: ACC ($\uparrow$)
- ▶ <https://xianchaoxiu.github.io>

Results

► Clustering comparisons

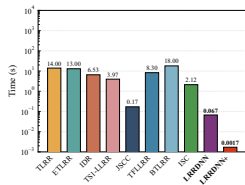
Datasets	TLRR	ETLRR	IDR	TS1-LLRR	JSCC	TFLRR	BTLRR	ISC	LRRDNN	LRRDNN+
ORL	59.03%	58.24%	73.32%	73.46%	74.53%	68.70%	53.53%	74.75%	75.25%	90.25%
PalmData25	77.94%	81.89%	95.27%	91.25%	93.80%	93.55%	92.52%	93.55%	97.45%	95.95%
FRDUE	74.23%	84.18%	95.31%	95.31%	91.35%	96.71%	94.66%	96.65%	97.24%	96.02%
USPS	40.97%	62.69%	79.80%	79.80%	69.62%	74.20%	55.59%	65.30%	80.50%	99.40%
LUNG	64.33%	42.91%	86.21%	75.81%	85.52%	63.04%	66.01%	74.88%	87.69%	79.31%
EYaleB	80.68%	76.99%	81.28%	73.86%	85.96%	71.27%	79.94%	75.59%	83.21%	99.09%
MNIST	33.44%	30.37%	59.59%	58.61%	46.60%	48.27%	37.65%	32.09%	54.93%	93.43%
AR	36.51%	12.79%	72.64%	76.46%	78.52%	71.67%	61.40%	54.42%	73.14%	89.95%
Average	58.39%	56.26%	80.43%	78.07%	78.24%	73.43%	67.66%	70.90%	81.18%	92.93%

► Ablation studies

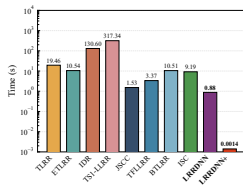
Models	SSN	ORL	PalmData25	FRDUE	USPS	LUNG	EYaleB	MNIST	AR	Average
LRRDNN	×	73.50%	91.75%	92.96%	76.10%	86.70%	79.79%	52.17%	72.61%	78.20%
	✓	75.25%	97.45%	97.24%	80.50%	87.69%	83.21%	54.93%	73.14%	81.18%
LRRDNN+	×	90.50%	94.00%	92.60%	97.10%	91.63%	98.81%	89.65%	82.69%	92.12%
	✓	98.75%	99.35%	98.72%	99.40%	94.09%	99.09%	93.43%	89.95%	96.60%

Efficiency

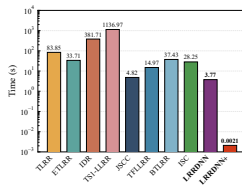
► Runtime in a logarithmic scale



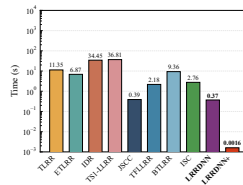
(a) ORL



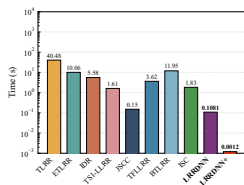
(b) PalmData25



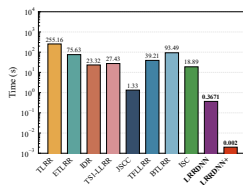
(c) FRDUE



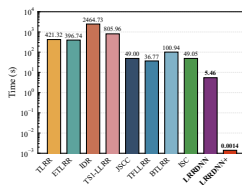
(d) USPS



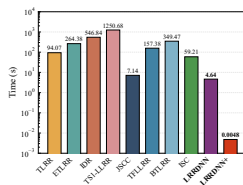
(e) LUNG



(f) EYaleB



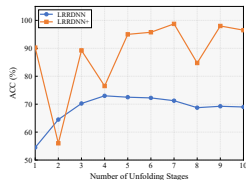
(g) MNIST



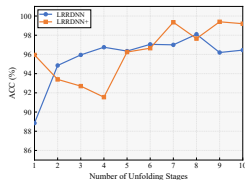
(h) AR

Analysis

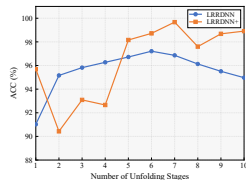
► Effect of the number of unrolling stages K



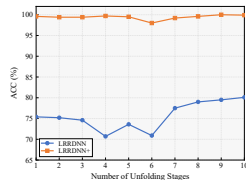
(a) ORL



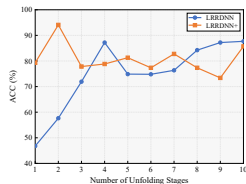
(b) PalmData25



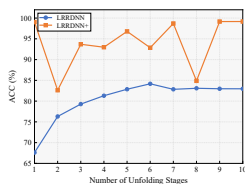
(c) FRDUE



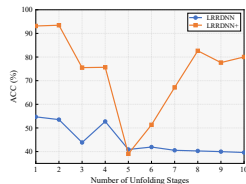
(d) USPS



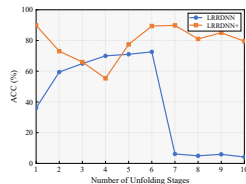
(e) LUNG



(f) EYaleB



(g) MNIST



(h) AR

Outline

Introduction

Proposed Methods

Numerical Experiments

Conclusions and Future Work

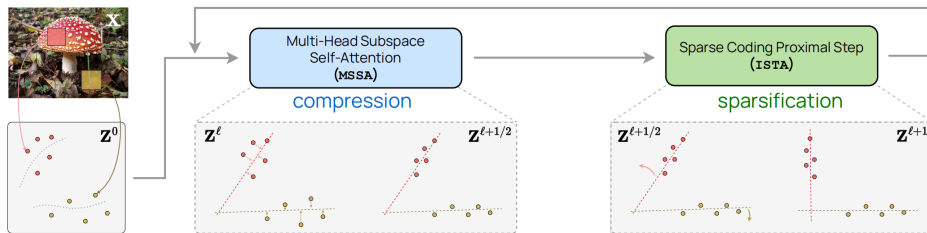
Conclusions

- ▶ Chen-Wang-Xiu-Liu, Rethinking Low-Rank Representation with Deep Semi-Smooth Newton, [PRCV](#), 2026
- ▶ Chen-Yu-Wang-Xiu-Liu, Solving Low-Rank Representation Through Deep Newton Networks, [working in process](#)
- ▶ Chen-Wang-Xiu-Liu, LLM-IMVC: Large Language Model-Guided Deep Unrolling for Incomplete Multi-View Clustering, [working in process](#)

Methods	HandWrite	BDGP	MSRC	COIL20	Caltech5V	USPS	Average
LMVSC (AAAI'20)	67.1%	55.6%	35.2%	69.2%	46.1%	56.0%	54.9%
SFMC (TPAMI'22)	75.7%	37.8%	60.0%	74.8%	46.8%	98.9%	65.7%
FPMVS-CAG (TIP'22)	82.3%	55.6%	80.5%	60.2%	80.9%	78.3%	73.0%
FDAGF (AAAI'23)	82.3%	52.5%	73.9%	74.0%	81.3%	56.1%	70.0%
RCAGL (TKDE'24)	79.4%	51.8%	77.6%	65.1%	83.6%	60.9%	69.7%
ALPC (AAAI'25)	63.2%	92.2%	64.3%	76.4%	46.1%	50.2%	65.4%
TLRLF4 (TPAMI'25)	76.4%	92.7%	—	76.3%	75.5%	99.2%	84.0%
LLM-IMVC	99.9%	100.0%	88.1%	77.2%	100.0%	99.9%	94.2%

Future Work

- ▶ Yang-Huang-Wipf, Transformers from An Optimization Perspective, [NeurIPS](#), 2022
- ▶ Yu-Buchanan-Pai et al, White-box Transformers via Sparse Rate Reduction, [NeurIPS](#), 2023
- ▶ Porras-Hadou-Ribeiro, A Constrained Optimization Perspective of Unrolled Transformers, [ICML](#), 2026
- ▶ Tai-Liu-Li et al, A Mathematical Explanation of Transformers, [SIIMS](#), 2026



Thank you for your attention!

xcxiu@shu.edu.cn