



HAL
open science

Deep Unrolling for Nonconvex Robust Principal Component Analysis

Elizabeth Z. C. Tan, Caroline Chaux, Emmanuel Soubies, Vincent y F Tan

► **To cite this version:**

Elizabeth Z. C. Tan, Caroline Chaux, Emmanuel Soubies, Vincent y F Tan. Deep Unrolling for Nonconvex Robust Principal Component Analysis. IEEE MLSP 2023, Sep 2023, Rome, France. 10.1109/MLSP55844.2023.10285962 . hal-04160961

HAL Id: hal-04160961

<https://hal.science/hal-04160961>

Submitted on 13 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

DEEP UNROLLING FOR NONCONVEX ROBUST PRINCIPAL COMPONENT ANALYSIS

Elizabeth Z. C. Tan¹, Caroline Chaux², Emmanuel Soubies³, and Vincent Y. F. Tan^{1,4}¹ Department of Mathematics, National University of Singapore, Singapore² CNRS, IPAL, Singapore³ CNRS, IRIT, Université de Toulouse, Toulouse, France⁴ Department of Electrical and Computer Engineering, National University of Singapore, Singapore

ABSTRACT

We design algorithms for Robust Principal Component Analysis (RPCA) which consists in decomposing a matrix into the sum of a low rank matrix and a sparse matrix. We propose a deep unrolled algorithm based on an accelerated alternating projection algorithm which aims to solve RPCA in its non-convex form. The proposed procedure combines benefits of deep neural networks and the interpretability of the original algorithm and it automatically learns hyperparameters. We demonstrate the unrolled algorithm's effectiveness on synthetic datasets and also on a face modeling problem, where it leads to both better numerical and visual performances.

Index Terms— RPCA, Sparsity, low-rank, unrolled algorithm, hyperparameters.

1. INTRODUCTION

Robust Principal Component Analysis (RPCA) is the task of recovering a low rank matrix $\mathbf{L}^* \in \mathbb{R}^{d_1 \times d_2}$ and a sparse matrix $\mathbf{S}^* \in \mathbb{R}^{d_1 \times d_2}$ from their linear combination [1]

$$\mathbf{M}^* = \mathbf{L}^* + \mathbf{S}^*. \quad (1)$$

Finding an exact solution to the RPCA problem is challenging due to its combinatorial nature. Yet, RPCA has received considerable attention due to its importance in many fields. These include applications from *latent semantic indexing* [2] to *image processing* [3], to *learning graphical models with latent variables* [4], and to *collaborative filtering* [5].

The art of conventional RPCA: Some authors [6], [7] considered a convex relaxation of RPCA, where the low rank matrix is obtained throughout the minimization of the nuclear norm and the sparse matrix via an ℓ_1 -norm penalization. Such optimization problems can be solved by proximal gradient methods. However, such approaches are computationally expensive due to the proximal mapping of the nuclear norm, which involves a full singular value decomposition (SVD) of a $d_1 \times d_2$ matrix, amounting to at least $\mathcal{O}(d_1 d_2 \min(d_1, d_2))$

flops per iteration. In contrast, alternating algorithms have been proposed to solve the original nonconvex formulation of RPCA involving the ℓ_0 pseudo-norm and the rank function (Section 2.1). These include the alternating projections (AltProj) method [8], its accelerated version (AccAltProj) [9], and a block-based method based on the CUR decomposition [10]. Although faster and more closely related to model (1) compared to methods based on convex relaxations, their performance heavily rely on good initializations.

Learning-based strategies in RPCA: Deep neural networks (DNNs) have experienced a surge in popularity over the past decades, often attaining groundbreaking performance in various applications. In signal processing, incorporating deep learning approaches has become prominent because of their ability to automatically learn salient information from real world data. However, DNNs are known to suffer from two shortcomings. Firstly, their black-box nature (i.e., the lack of interpretability) hinders our understanding of why certain predictions are derived, which is crucial in detecting limitations. Secondly, they are susceptible to overfitting to the training data since they often have a large number of parameters compared to the amount of available training data.

To overcome these limitations, a technique known as *deep unrolling* (also known as *deep unfolding*) has been extensively explored [11] and has emerged as a promising approach in various signal processing problems. While the model parameters are fixed in the classical algorithms, the unrolled network replaces them with learnable parameters that can be optimised through end-to-end training using backpropagation. Therefore, a trained unrolled network can be viewed as a parameter-optimised algorithm, sharing both the benefits of conventional DNNs and interpretability of the original algorithm. Furthermore, as classical algorithms often have significantly fewer parameters than DNNs, unrolled networks can potentially mitigate the overfitting problem when there is insufficient data or when the training dataset is of low quality.

Existing unrolling strategies in the context of RPCA are currently limited to algorithms based on convex relaxations. These include CORONA [12], refRPCA [13], and other similar works [14], [15], [16], [17]. However, they inherit from

This work was supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) Programme.

the previously mentioned drawbacks of such convex relaxations. To the best of our knowledge, there does not exist unrolled versions of the alternating projections algorithm in RPCA, despite that being the state-of-the-art. Indeed, such an unrolled algorithm be beneficial in terms of having appealing computational properties and the closeness to model in (1) of nonconvex RPCA approaches, while mitigating existing shortcomings (sensitivity to the initialization and incognizance of hyperparameters).

Contributions: We propose an unrolled version of the Accelerated Alternating Projections algorithm [9]. The proposed procedure also incorporates the Minimax Concave Penalty (MCP), an alternative to hard thresholding owning numerous interesting properties and more suitable than the ℓ_1 -norm relaxation [18]. The overall proposed procedure performs excellently on benchmark synthetic datasets and real-world (face) datasets, exceeding the performances on the state-of-the-art (unrolled) approaches.

Outline: The paper is organised as follows. Preliminaries on RPCA algorithms and deep unrolling are presented in Section 2. The proposed method is then described in Section 3 and numerical experiments are conducted in Section 4. Finally, concluding remarks are presented in Section 5.

2. PRELIMINARIES

2.1. Algorithms for RPCA

RPCA may be formulated as the following non-convex optimization problem

$$\begin{aligned} \arg \min_{\mathbf{L}, \mathbf{S} \in \mathbb{R}^{d_1 \times d_2}} \|\mathbf{M}^* - \mathbf{L} - \mathbf{S}\|_{\text{F}}, \\ \text{subject to } \text{rank}(\mathbf{L}) \leq r \text{ and } \|\mathbf{S}\|_0 \leq |\Omega|, \end{aligned} \quad (2)$$

where $\|\cdot\|_{\text{F}}$ is the Frobenius norm, $r \geq \text{rank}(\mathbf{L}^*)$ upper bounds the rank of the low rank matrix \mathbf{L}^* , and $k \geq |\Omega|$ upper bounds the cardinality of the support of the sparse matrix \mathbf{S}^* .

Netrapalli *et al.* [8] proposed to solve (2) using the alternating projections (AltProj) method, which projects $\mathbf{M}^* - \mathbf{S}_k$ onto the space of low rank matrices and $\mathbf{M}^* - \mathbf{L}_k$ onto the set of sparse matrices in an alternating manner at each iteration k . It enjoys a computational complexity of $\mathcal{O}(d_1 d_2 r^2)$ per iteration. Building upon AltProj, Cai *et al.* [9] proposed an accelerated version known as AccAltProj with an improved complexity $\mathcal{O}(d_1 d_2 r)$. Later, Cai *et al.* [10] introduced the Iterated Robust CUR (IRCUR), which is a variant of AltProj with a per-iteration complexity of $\mathcal{O}(r^2 n \log n)$, where $n = \max(d_1, d_2)$. This is achieved by operating on submatrices, hence avoiding expensive computations on full matrices. However, it is widely acknowledged that CUR-based decompositions are less accurate than SVD-based ones.

We briefly describe AccAltProj in Alg. 1, before we move on to the proposed unrolled model. Here, \mathcal{M}_r denotes the set of rank- r matrices, and T_k denotes the tangent space of \mathcal{M}_r at \mathbf{L}_k . The i -th largest singular value of a matrix \mathbf{X} is denoted as

Algorithm 1: Accelerated Alternating Projections (AccAltProj)

Input: $\mathbf{M}^*, r, \epsilon, \beta_{\text{init}}, \beta, \gamma$

- 1 $\zeta_{-1} \leftarrow \beta_{\text{init}} \cdot \sigma_1^{(\mathbf{M}^*)}$;
- 2 $\mathbf{S}_{-1} \leftarrow \mathcal{T}_{\zeta_{-1}}(\mathbf{M}^*)$;
- 3 $\mathbf{L}_0 \leftarrow \mathcal{H}_r(\mathbf{M}^* - \mathbf{S}_{-1})$;
- 4 $\zeta_0 \leftarrow \beta \cdot \sigma_1^{(\mathbf{M}^* - \mathbf{S}_{-1})}$;
- 5 $\mathbf{S}_0 \leftarrow \mathcal{T}_{\zeta_0}(\mathbf{M}^* - \mathbf{L}_0)$;
- 6 $k \leftarrow 0$;
- 7 **while** $\|\mathbf{M}^* - \mathbf{L}_k - \mathbf{S}_k\|_{\text{F}} / \|\mathbf{M}^*\|_{\text{F}} \geq \epsilon$ **do**
- 8 $\mathbf{P}_{k+1} \leftarrow P_{T_k}(\mathbf{M}^* - \mathbf{S}_k)$;
- 9 $\mathbf{L}_{k+1} \leftarrow \mathcal{H}_r(\mathbf{P}_{k+1})$;
- 10 $\zeta_{k+1} \leftarrow \beta(\sigma_{r+1}^{(\mathbf{P}_{k+1})} + \gamma_{k+1} \sigma_1^{(\mathbf{P}_{k+1})})$;
- 11 $\mathbf{S}_{k+1} \leftarrow \mathcal{T}_{\zeta_{k+1}}(\mathbf{M}^* - \mathbf{L}_{k+1})$;
- 12 $k \leftarrow k + 1$;

end

Output: $\mathbf{L}_k, \mathbf{S}_k$

$\sigma_i^{(\mathbf{X})}$. The operator \mathcal{H}_r represents the truncated SVD operation at rank r and \mathcal{T}_{ζ} represents the hard-thresholding operator (i.e., the proximity operator of ℓ_0) with threshold ζ .

AccAltProj differs from AltProj by performing a tangent space projection on T_k rather than directly projecting $\mathbf{M}^* - \mathbf{S}_k$ onto \mathcal{M}_k . This is followed by projecting the intermediate matrix onto \mathcal{M}_r to obtain \mathbf{L}_{k+1} before projecting $\mathbf{M}^* - \mathbf{L}_{k+1}$ back onto the set of sparse matrices. Cai *et al.* [9] derived the projection operator onto T_k as:

$$P_{T_k}(\mathbf{A}) = [\mathbf{U}_k \quad \mathbf{Q}_1] \begin{bmatrix} \mathbf{U}_k^{\top} \mathbf{A} \mathbf{V}_k & \mathbf{R}_2^{\top} \\ \mathbf{R}_1 & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_k^{\top} \\ \mathbf{Q}_2^{\top} \end{bmatrix}, \quad (3)$$

where $\mathbf{U}_k, \mathbf{V}_k$ contain the singular vectors from the truncated SVD of $\mathbf{L}_k = \mathbf{U}_k \Sigma_k \mathbf{V}_k^{\top}$, and $(\mathbf{Q}_1, \mathbf{R}_1)$ and $(\mathbf{Q}_2, \mathbf{R}_2)$ are the factors from the QR decompositions of $(\mathbf{I} - \mathbf{V}_k \mathbf{V}_k^{\top})(\mathbf{M}^* - \mathbf{S}_k)^{\top} \mathbf{U}_k$ and $(\mathbf{I} - \mathbf{U}_k \mathbf{U}_k^{\top})(\mathbf{M}^* - \mathbf{S}_k) \mathbf{V}_k$ respectively.

2.2. Deep Unrolling

A tedious task in implementing iterative optimization algorithms is to tune their hyperparameters (e.g., stepsize, regularisation parameters). To circumvent this problem, *unrolled* versions of standard algorithms [11] have recently been developed. In essence, algorithm unrolling or unfolding consists in converting an iterative algorithm into a neural network. One iteration of the iterative algorithm is being transformed to one layer of the neural network. The benefits of this approach include neural network interpretability and automatic parameter learning. Following this line of thought, we propose to unroll the Accelerated Alternating Projections algorithm (Alg. 1).

3. PROPOSED UNROLLED AccAltProj

We adopt AccAltProj as our baseline model to unroll as it is fast compared to most existing algorithms and is more robust

compared to IRCUR. We follow the idea from the Learned Iterative Soft Thresholding Algorithm [19] to design a non-linear feed-forward architecture with a fixed number of layers.

As β and γ are fixed heuristically in AccAltProj, we chose to learn them in the unrolled network. The parameter β controls the variance of matrix elements of recovered $\hat{\mathbf{L}}$ while γ controls the rate of convergence [8]. They also play a key role for the theoretical guarantee of AccAltProj. More precisely, if properly chosen, the initial guesses \mathbf{S}_{-1} and \mathbf{L}_0 generated at Lines 1 to 5 of Alg. 1 fulfill the required condition for local convergence of AccAltProj [9, Theorem 1]. Learning β and γ automatically allows our model to be customisable to use cases where datasets share similar properties for the underlying low-rank and sparse components.

3.1. Using the Minimax Concave Penalty (MCP) instead of the ℓ_0 or ℓ_1 norms

One challenge when developing an unrolled version of AccAltProj is that we are unable to directly use hard-thresholding for the non-linear activation. This is because it is not sub-differentiable, a property needed to deploy gradient-based optimizers to learn the parameters β and γ [19].

LRPCA [18] tackled this problem by replacing the hard-thresholding operator with the soft-thresholding operator in their unrolled model. However, the soft-thresholding operator is the proximal mapping of the ℓ_1 norm while the hard-thresholding operator is the proximal mapping of the ℓ_0 pseudo-norm. As such, vanilla soft-thresholding is not suitable for our objective in (2).

Taking the best of both worlds, we consider in this work the *Minimax Concave Penalty* (MCP) [20], defined as

$$\text{MCP}(x; \zeta, v) = \begin{cases} \frac{v\zeta^2}{2}, & \text{if } |x| > v\zeta \\ \zeta|x| - \frac{x^2}{2v}, & \text{if } |x| \leq v\zeta, \end{cases} \quad (4)$$

where ζ is the threshold, and $v > 1$ is a parameter controlling the concavity of the penalty. It has a close relationship to the ℓ_0 pseudo-norm, from both the statistical [20] and optimization viewpoints [21], while being subdifferentiable.

Its proximal mapping $\mathcal{P}(x; \zeta, v) := \text{prox}_{\text{MCP}}(x; \zeta, v)$ is

$$\mathcal{P}(x; \zeta, v) = \text{sign}(x) \min \left\{ \frac{v \max(|x| - \zeta, 0)}{v - 1}, |x| \right\} \quad (5)$$

which corresponds to the ‘‘firm thresholding operator’’ [22], a compromise between soft- and hard- thresholding. In the unrolled version of Alg. 1, we use $\mathcal{P}(\cdot; \zeta, v)$ in place of the hard thresholding operator \mathcal{T}_ζ . The penalty functions and their corresponding proximal mappings are shown in Fig. 1.

3.2. Unrolled AccAltProj

We consider an unrolled version of the Modified Accelerated Alternating Projections and refer to it as the unrolled RPCA algorithm. Each iteration is thus transformed in one layer as

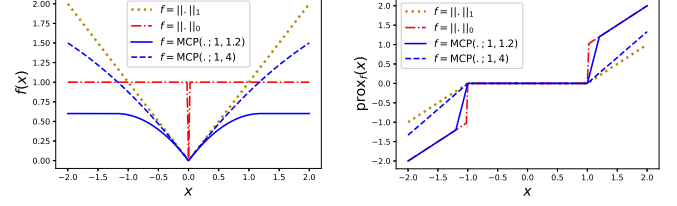


Fig. 1. Penalty functions and their proximal mappings.

shown in Fig. 2. We use this neural network to learn β and γ while keeping v fixed ($v = 1.05$).

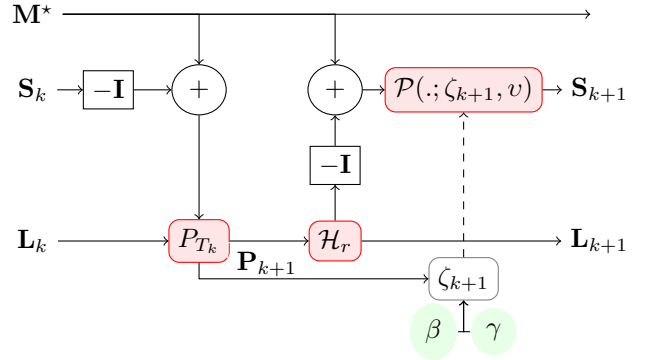


Fig. 2. One layer of the unrolled RPCA algorithm. Note that ζ_{k+1} is a function of (β, γ) , defined in Line 10 of Alg. 1.

3.3. Training Criteria

While it is possible to use adaptive parameters (i.e., separate β_k, γ_k for each layer k), we choose to learn only a single (β, γ) that is *shared* across the layers. In the unrolled model, we initialise $\beta = \frac{1}{2 \cdot \sqrt{d_1 \times d_2}}$ and $\gamma = 0.7$ since they are the default values used in AccAltProj [9].

Consider a set of input data $\{\mathbf{M}_{\text{train}}^q\}_{q=1}^Q$ and the associated sparse and low-rank decomposition that we denote by $\{(\mathbf{L}_{\text{train}}^q, \mathbf{S}_{\text{train}}^q)\}_{q=1}^Q$. These can be obtained either via simulations or through the application of classical RPCA iterative algorithms on $\mathbf{M}_{\text{train}}^q$. Then, following [12], we learn the two parameters γ and β via

$$(\hat{\gamma}, \hat{\beta}) \in \arg \min_{(\gamma, \beta) \in \mathbb{R}^2} \sum_{q=1}^Q \mathcal{L}(\mathbf{L}_{\text{train}}^q, \mathbf{L}^q) + \mathcal{L}(\mathbf{S}_{\text{train}}^q, \mathbf{S}^q) \quad (6)$$

$$\text{subject to } (\mathbf{L}^q, \mathbf{S}^q) = \mathcal{N}(\gamma, \beta; \mathbf{M}_{\text{train}}^q)$$

where \mathcal{N} is defined by cascading layers as in Fig. 2 (unrolled network). Finally, we set the loss \mathcal{L} to be the relative error $\mathcal{L}(\hat{\mathbf{X}}, \mathbf{X}) = \|\hat{\mathbf{X}} - \mathbf{X}\|_{\mathbb{F}}^2 / \|\mathbf{X}\|_{\mathbb{F}}^2$.

4. NUMERICAL EXPERIMENTS

To illustrate the effectiveness of the proposed approach, we performed experiments on two settings: a fully controlled one

through synthetic simulations and a realistic one in the context of face modelling. The code to reproduce the simulations will be released.

4.1. Simulated/Sythetic Data

Problem setup: The synthetic data are generated as in [9], i.e., let $\mathbf{L}^* = \mathbf{U}\mathbf{V}^\top$, where $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{d \times r}$ contain elements generated i.i.d. from the standard normal distribution. Similarly, the components of \mathbf{S}^* are sampled i.i.d. and uniformly from the interval $[-c \cdot \mathbb{E}(|[\mathbf{L}^*]_{ij}|), c \cdot \mathbb{E}(|[\mathbf{L}^*]_{ij}|)]$ where $c > 0$. The positions of the non-zero elements are randomly sampled without replacement. In the following, the matrix \mathbf{S}^* is said to be α -sparse if each of its rows and columns contain at most αd non-zero elements. Finally, given a generated pair $(\mathbf{L}^*, \mathbf{S}^*)$, we generate an input-target training data matrix as $\mathbf{M}_{\text{train}} = \mathbf{M}^* = \mathbf{L}^* + \mathbf{S}^*$ and $(\mathbf{L}_{\text{train}}, \mathbf{S}_{\text{train}})$ is obtained via IRCUR applied on $\mathbf{M}_{\text{train}}$.

For this experiment, we fix the dimensions to $d_1 = d_2 = d = 250$ and the rank to $r = 2$. We consider several simulated data sets generated by varying the sparsity level (controlled by α) and the amplitude (controlled by c) of the sparse component \mathbf{S}^* . More precisely, we consider the following cases to assess the performance of our unrolled network:

| | Case 1 | Case 2 | Case 3 | Case 4 |
|---------------|----------|----------|-----------|-----------|
| (α, c) | (0.1, 1) | (0.3, 1) | (0.01, 1) | (0.1, 10) |

Table 1. Table showing different experimental settings.

For each case, we generate a total of 300 samples, and split them into 180 training samples and 120 test samples. The unrolled network is trained for a total of 8 epochs.

The metrics that we use to quantify the performances of the unrolled model and its competitors are as follows:

$$\epsilon_L(\mathbf{L}_{\text{out}}) := \|\mathbf{L}^* - \mathbf{L}_{\text{out}}\|_F, \quad (7)$$

$$\epsilon_S(\mathbf{S}_{\text{out}}) := \|\mathbf{S}^* - \mathbf{S}_{\text{out}}\|_F, \quad (8)$$

$$\epsilon_M(\mathbf{L}_{\text{out}}, \mathbf{S}_{\text{out}}) := \|\mathbf{M}^* - \mathbf{L}_{\text{out}} - \mathbf{S}_{\text{out}}\|_F / \|\mathbf{M}^*\|_F, \quad (9)$$

$$\epsilon_{\text{supp}}(\mathbf{S}_{\text{out}}) := \frac{1}{d^2} (\mathbb{1}_{\{[\mathbf{S}^*]_{ij}=0, [\mathbf{S}_{\text{out}}]_{ij} \neq 0\}} + \mathbb{1}_{\{[\mathbf{S}^*]_{ij} \neq 0, [\mathbf{S}_{\text{out}}]_{ij}=0\}}), \quad (10)$$

where $\mathbf{L}_{\text{out}}, \mathbf{S}_{\text{out}}$ are placeholders for the outputs that could be computed from IRCUR, AccAltProj, or the unrolled model (after training). These four errors respectively quantify the accuracies on 1) the estimation of \mathbf{L}^* , 2) the estimation of \mathbf{S}^* , 3) the overall matrix \mathbf{M}^* and 4) support recovery of \mathbf{S}^* .

Results: In Fig. 3, we report the four errors described in the four cases. We compare the performance of the proposed approach with IRCUR [10] and AccAltProj [9] (which are not unrolled algorithms).

We observe from Fig. 3 that the proposed unrolled algorithm improves over its classical counterpart, which means

that the hyperparameters are learned well. The lowest error in \mathbf{M} is always achieved by the IRCUR method and remains small (order 10^{-8} versus 10^{-7}) for the other methods. This can be explained by the fact that the hyperparameters are learnt so as to minimize the error on \mathbf{L} and \mathbf{S} . This is confirmed by the results obtained individually on matrices \mathbf{L} and \mathbf{S} for which the smallest error is always obtained by the proposed unrolled method. Finally, as expected, the unrolled algorithm using the ℓ_1 -norm instead of the MCP does not perform well.

| | Case 1 | Case 2 | Case 3 | Case 4 |
|----------|-----------------------|-----------------------|-----------------------|-----------------------|
| γ | 7.74×10^{-1} | 7.71×10^{-1} | 7.88×10^{-1} | 7.40×10^{-1} |
| β | 7.03×10^{-2} | 7.71×10^{-2} | 5.00×10^{-2} | 4.58×10^{-2} |

Table 2. Parameters learnt from the trained unrolled network in the different experimental settings.

From Table 2, we observe that the learned γ 's are similar across different settings, where they are all slightly larger than their initialised value of 0.7. This suggests that the default value of $\gamma = 0.7$ suggested in [9] is a fairly good estimate. The slight increase may be because AccAltProj implements an early stopping criterion, where stops once the error $\frac{\|\mathbf{M}^* - \mathbf{L}_k - \mathbf{S}_k\|_F}{\|\mathbf{M}^*\|_F}$ at iteration $k < 50$ is below the tolerance of 10^{-6} . As such, for a larger fixed number of layers, a larger γ would be needed so that the network converges more slowly to the same point. Conversely, the learned γ would be smaller if we reduced the number of layers for the unrolled network. This means that learned γ is optimised for the given fixed number of layers in the network. The learned β exhibits much more variation across the different cases. In particular, they increased by about 2 times from before training in Cases 1 and 2, and about 1.5 times in Cases 3 and 4. This observation is in line with the interpretation of β in [8], which stated that a higher value of β results in $\hat{\mathbf{L}}$ that is more ‘‘spiky’’ and $\hat{\mathbf{S}}$ that is more heavily diffused. We can take Case 1 as the baseline for the other cases to compare against. In Case 2 where α is greater than in Case 1, there would be more non-zero values in \mathbf{S}^* , making it more diffused. In contrast, with smaller α in Case 3, the few non-zero elements of \mathbf{S}^* become more prominent against the backdrop of the other zero-valued elements, making it less diffused. In Case 4 where the magnitude of non-zero elements in \mathbf{S}^* is 10 times of that in Case 1, the non-zero values are more pronounced and hence \mathbf{S}^* is less diffused. As the learned β matches our expectation from theory for each case, this demonstrates that our unrolled model is indeed able to automatically fine-tune the parameter β to the different settings, which is an advantage over the classical AccAltProj.

4.2. Face Dataset

Problem setup: We now test the proposed unrolled model on the Yale Face Database [23] for the application of face model-

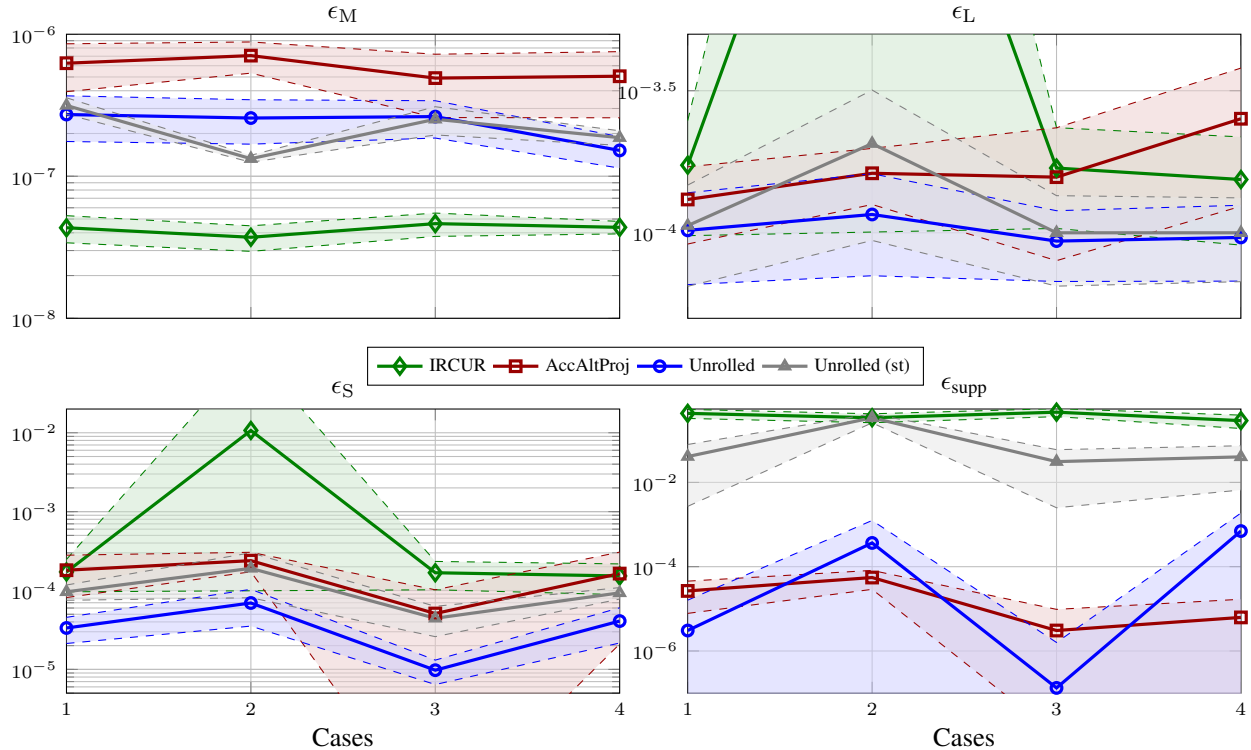


Fig. 3. Means and standard deviations of errors. The “Unrolled (st)” curve stands for unrolled algorithm with *soft-thresholding*.

ing. The Yale database consists of 11 grayscale facial images each for a total of 15 subjects. The 11 images, each having a dimension of 243×320 , show the same individual with different facial expressions, lighting conditions or accessories such as spectacles on the face. The task is of face modeling is to recover the occlusion-free image for facial recognition [3].

We vectorise the images of each subject, which are then stacked to form a 77760×11 matrix M^* . The static occlusion-free image of the subject forms the low-rank component of the matrix while the varied facial expressions, shadows, and objects covering the face form the low-rank component. Since all 11 images share one common underlying occlusion-free facial image, we assume that $\text{rank}(M^*) = 1$.

Subjects 1 to 7 and Subjects 8 to 15 are used for training and testing respectively. Similar to the experiment on synthetic datasets, we use IRCUR to obtain initial estimates and train the unrolled network for 8 epochs.

Results: Visual results are displayed in Figs. 4 and 5 for the low rank and sparse parts respectively. The methods enable the separation of the original images into expressionless faces and expression details. While the images learned by IRCUR are poor, the proposed unrolled strategy adaptively learned hyperparameters (γ, β) that result in sharper edges. Tests were performed on an Intel(R) Core(TM) i7-1185G7 @3.00GHz, with 32Go RAM. The total (7 subjects) training time is 300s. Testing times (per subject on average) are as follows: IRCUR: 8s, AccAltProj: 0.75s and unrolled procedure: 4.375s, showing that the unrolled procedure has a good

accuracy-computation time tradeoff.

5. CONCLUSION

We proposed an unrolled algorithm to solve the RPCA problem in its nonconvex form. This results in an unrolled version of the AccAltProj algorithm but incorporates the Minimax Concave Penalty. The underlying learning strategy, which has the advantage of learning hyperparameters γ and β automatically, allows us to improve the state-of-the-art performances on benchmark synthetic datasets used in existing works as well as on real-world face datasets. In future work, we plan to improve on the training criterion in Section 3.3 as well as the automatic learning of more parameters such as the ones that parametrize the MCP, i.e., ζ and v in (4).

6. REFERENCES

- [1] E. J. Candès, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?,” *J. ACM*, vol. 58, no. 3, pp. 1–37, 2011.
- [2] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, “Indexing by latent semantic analysis,” *J. Am. Soc. Inf. Sci.*, vol. 41, no. 6, pp. 391–407, 1990.
- [3] T. Bouwmans, S. Javed, H. Zhang, Z. Lin, and R. Otazo, “On the applications of robust PCA in image and video

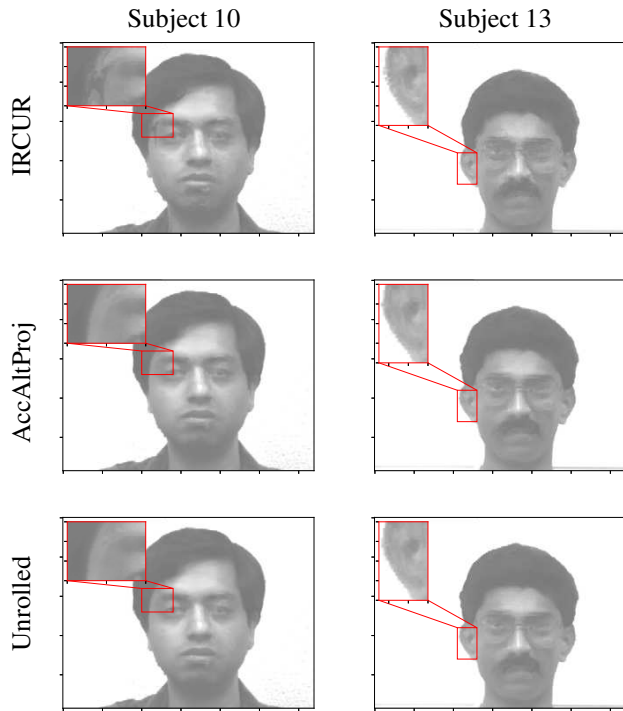


Fig. 4. Low-rank components of subjects 10 and 13 recovered from IRCUR, AccAltProj, and unrolled network. AccAltProj and its unrolled version result in images with sharper edges.

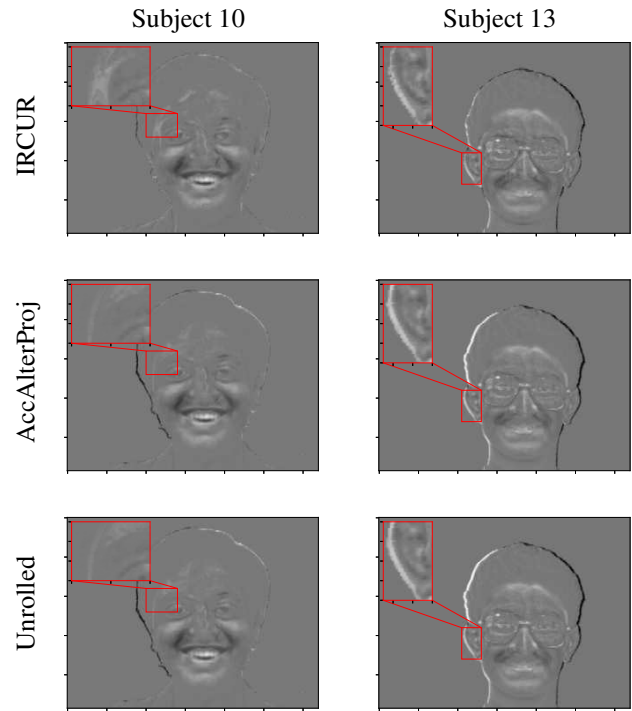


Fig. 5. Sparse components (happy) of subjects 10 and 13 recovered from IRCUR, AccAltProj and unrolled network. Observe that AccAltProj and its unrolled version result in smoother estimates while preserving details.

processing.” *Proc. IEEE*, vol. 106, no. 8, pp. 1427–1457, 2018.

- [4] V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky, “Latent variable graphical modeling via convex optimization,” *The Annals of Statistics*, vol. 40, no. 4, pp. 1935–1967, 2012.
- [5] Y. Koren, S. Rendle, and R. Bell, “Advances in collaborative filtering,” *Recommender systems handbook*, pp. 91–142, 2021.
- [6] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, “Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization,” in *Adv. Neural Inf. Process. Syst.*, Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, Eds. 2009, vol. 22, Curran Associates, Inc.
- [7] Z. Lin, A. Ganesh, J. Wright, L. Wu, M. Chen, and Y. Ma, “Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix,” *Proc. IEEE Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, vol. 61, 2009.
- [8] P. Netrapalli, U. N. Niranjan, S. Sanghavi, A. Anandkumar, and P. Jain, “Non-convex robust PCA,” 2014.

- [9] H. Cai, J.-F. Cai, and K. Wei, “Accelerated alternating projections for robust principal component analysis,” *J. Mach. Learn. Res.*, vol. 20, no. 1, pp. 685–717, Jan 2019.
- [10] H. Cai, K. Hamm, L. Huang, J. Li, and T. Wang, “Rapid robust principal component analysis: CUR accelerated inexact low rank estimation,” *IEEE Signal Process. Lett.*, vol. 28, pp. 116–120, Feb 2021.
- [11] V. Monga, Y. Li, and Y. Eldar, “Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing,” *IEEE Signal Process. Mag.*, vol. 38, no. 2, pp. 18–44, 3 2021.
- [12] O. Solomon, R. Cohen, Y. Zhang, Y. Yang, Q. He, J. Luo, R. J. G. van Sloun, and Y. C. Eldar, “Deep unfolded robust PCA with application to clutter suppression in ultrasound,” *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 1051–1063, 2020.
- [13] H. Van Luong, B. Joukovsky, Y. C. Eldar, and N. Deligiannis, “A deep-unfolded reference-based RPCA network for video foreground-background separation,” *Proc. Eur. Sig. Image Proc. Conf.*, pp. 1432–1436, 2021.

- [14] R. Cohen, Y. Zhang, O. Solomon, D. Toberman, L. Taieb, R. J. G. van Sloun, and Y. C. Eldar, "Deep convolutional robust PCA with application to ultrasound imaging," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, 2019, pp. 3212–3216.
- [15] R. J. G. van Sloun, R. Cohen, and Y. C. Eldar, "Deep learning in ultrasound imaging," *Proc. IEEE*, vol. 108, no. 1, pp. 11–29, 2020.
- [16] B. Qin, H. Mao, Y. Liu, J. Zhao, Y. Lv, Y. Zhu, S. Ding, and X. Chen, "Robust PCA unrolling network for super-resolution vessel extraction in x-ray coronary angiography," *IEEE Trans. Med. Imag.*, vol. 41, no. 11, pp. 3087–3098, 2022.
- [17] S. Markowitz, C. Snyder, Y. C. Eldar, and M. N. Do, "Multimodal unrolled robust PCA for background foreground separation," *IEEE Trans. Image Process.*, vol. 31, pp. 3553–3564, 2022.
- [18] H. Cai, J. Liu, and W. Yin, "Learned robust PCA: A scalable deep unfolding approach for high-dimensional outlier detection," in *Adv. Neural Inf. Process. Syst.*, A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, Eds., 2021, pp. 16977–16989.
- [19] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proc. Int. Conf. Mach. Learn.*, Madison, WI, USA, 2010, ICML'10, p. 399–406, Omnipress.
- [20] C.-H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *Ann. Statist.*, vol. 38, no. 2, pp. 894 – 942, 2010.
- [21] E. Soubies, L. Blanc-Féraud, and G. Aubert, "A unified view of exact continuous penalties for l2-l0 minimization," *SIAM J. Optim.*, vol. 27, no. 3, pp. 2034–2060, 2017.
- [22] J. Woodworth and R. Chartrand, "Compressed sensing recovery via nonconvex shrinkage penalties," *Inverse Problems*, vol. 32, no. 7, pp. 075004, may 2016.
- [23] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, 1997.