ROBUST PCA VIA ALTERNATING ITERATIVELY REWEIGHTED LOW-RANK MATRIX FACTORIZATION

Paris V. Giampouras, Athanasios A. Rontogiannis, Konstantinos D. Koutroumbas

IAASARS, National Observatory of Athens, 15236, Penteli, Greece

ABSTRACT

Nowadays, many modern imaging applications generate large-scale and high-dimensional data. In order to efficiently handle these data, statistical tools amenable to exploiting their intrisic low-dimensional nature are needed. PCA is a ubiquitous method which has been widely applied in a variety of applications. However a major shortcoming of PCA is its sensitivity to gross errors - outliers. In light of this, robust PCA has been recently proposed. Robust PCA accounts for gross errors by assuming that the data matrix is the superposition of a low-rank matrix and a sparse matrix. In this work, a matrix factorization-based formulation of robust PCA which can efficiently handle large scale data is proposed. Low-rankness is imposed via a novel low-rank promoting term applied on the matrix factors, which can be viewed as a weighted version of the variational form of the nuclear norm. The newly formulated robust PCA problem is addressed via an alternating iteratively reweighted least squares-type algorithm. Simulated and real data experiments verify the effectiveness of the proposed algorithm as compared to other state-of-the-art robust PCA algorithms.

Index Terms— robust PCA, low-rank, matrix factorization

1. INTRODUCTION

Robust PCA has been at the heart of several image and video processing applications in recent years. Among others, it has been applied in background-foreground video substraction, image inpainting, face recognition etc. The main premise of robust PCA as compared to the traditional PCA is to account for data points that deviate from the adopted model and thus can be considered as samples that have been corrupted by gross errors, also known as *outliers*. Since these entries are not known beforehand, the robust PCA problem amounts to simultaneously detecting these entries and then estimate the "true" uncorrupted values, [1].

That said, robust PCA is an inherently ill-posed problem unless certain assumptions are made. Along these lines, robust PCA has been widely viewed as the task of decomposing a data matrix $\mathbf{X}$ into the sum of a *low-rank* matrix $\mathbf{L}$ and a sparse one $\mathbf{S}$. Convex formulations of the problem, [2], came into the scene for tranforming the problem to a well-posed one. In [2], it was shown that under certain conditions convex relaxations can efficiently recover the low-rank matrix provided that the matrix of the outliers is sparse enough.

Recently, large scale imaging applications necessitate the development of computational efficient methods for addressing the robust PCA problem. In light of this, nonconvex formulations of this problem with the aim to recude the comptational complexity of the resulting algorithms have been proposed. In that framework, matrix factorization based formulations of robust PCA have been recently introduced in the literature. The main idea of those approaches is to reduce the degrees of freedom of the problem by assuming that the low-rank component can be expressed via the product of two reduced-size matrices $\mathbf{U} \in \mathcal{R}^{m\times d}$, $\mathbf{V} \in \mathcal{R}^{n\times d}$ i.e., $\mathbf{L} = \mathbf{U}\mathbf{V}^T$. Since in most applications the rank r of $\mathbf{L}$ is unknown a priori, the inner dimension d of the matrix factors $\mathbf{U}$, $\mathbf{V}$ is taken as an overestimate of r i.e., $d \geq r$. Then, low-rank imposing terms are applied, that penalize the rank of $\mathbf{U}\mathbf{V}^T$. Among other approaches, the variational form of the nuclear norm has been adopted in robust PCA. Recently, matrix factorization approaches penalizing the Schatten-p norm of the factors $\mathbf{U}, \mathbf{V}$ were also introduced in [3].

In this paper, a novel approach for solving the robust PCA problem is proposed. Capitalizing on recent matrix factorization based schemes, we put forth a new low-rank imposing term applied on the matrix factors $\mathbf{U}, \mathbf{V}$. This term is in fact a weighted version of the variational form of the nuclear norm and encapsulates other approaches upon selecting appropriate weights. Going one step further we propose the utilization of a common reweighting matrix for both factors which induces jointly group-sparsity on the columns of $\mathbf{U}$, $\mathbf{V}$. In an effort to remain within the reweighting framework, a reweighted ℓ_1 norm is applied on the sparse matrix of outliers. Next, based on the block successive upper-bound minimization framework (BSUM), [4], an efficient alternating iteratively reweighted algorithm for solving the newly formulated

We acknowledge support of this work by the project PROTEAS II - Advanced Space Applications for Exploring the Universe of Space and Earth" (MIS 5002515) which is implemented under the Action Reinforcement f the Research and Innovation InfrastructurE", funded by the Operational Programme Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund).

robust PCA problem is derived. Simulated and real experimental results corroborate the effectiveness of the proposed algorithm as compared to other state-of-the-art robust PCA algorithms.

2. RELATED WORK

Robust PCA has been formulated in diverse ways in the literature. It first appeared as an NP-hard problem in [2],

$$\min_{\mathbf{L},\mathbf{S}} \ \mathrm{rank}(\mathbf{L}) + \|\mathbf{S}\|_0 \qquad \text{subject to} \quad \mathbf{X} = \mathbf{L} + \mathbf{S}, \quad (1)$$

where $\{\mathbf{X}, \mathbf{L}, \mathbf{S}\} \in \mathcal{R}^{m\times n}$, $\mathbf{X}$ denotes the data matrix, $\mathbf{L}$ its low-rank component and $\mathbf{S}$ the sparse one containing the outliers. The NP-hardness of this problem, coming up due to the combinatorial nature of $\mathrm{rank}(\cdot)$ and ℓ_0 norm, has sparked a wealth of computationally tractable altenatives for finding approximate solutions of (1). Among them, the combination of the convex envelopes of rank and ℓ_0 norm i.e., the nuclear norm $||\cdot||_*$ and the ℓ_1 norm $||\cdot||_1$ have been utilized as follows

$$\min_{\mathbf{L},\mathbf{S}} \ \|\mathbf{L}\|_* + \|\mathbf{S}\|_1 \qquad \text{subject to} \quad \mathbf{X} = \mathbf{L} + \mathbf{S}. \quad (2)$$

In an effort to overcome inherent shortcomings of the nuclear norm, i.e., the equal penalization of each singular value regardless of its magnitude, [5], a weighted version of the latter was introduced. The weighted nuclear norm is based on similar premises to those of the weighted version of the ℓ_1 norm, [6], and has been proven to provide significant merits in terms of data recovery performance. By incorporating the weighted nuclear norm, the robust PCA minimization problem is expressed as follows

$$\min_{\mathbf{L},\mathbf{S}} \ \|\mathbf{L}\|_{*,\mathbf{w}} + \|\mathbf{S}\|_1 \qquad \text{subject to} \quad \mathbf{X} = \mathbf{L} + \mathbf{S}, \quad (3)$$

where $\mathbf{w} = [w_1, w_2, \dots, w_n]^T, w_i \geq 0, i = 1, 2, \dots, n$ (assuming $m \geq n$) and $\|\mathbf{L}\|_{*,\mathbf{w}} = \sum_{i=1}^n w_i\sigma_i(\mathbf{L})$ is the resulting weighted version of the nuclear norm. As is analytically shown in [5], the weighted nuclear norm is convex when $w_{i+1} \leq w_i, i = 1, 2, \dots, n-1$. An interesting case arises when a reweighted version of this is adopted by defining the weights as follows

$$w_i = \frac{1}{\sigma_i(\mathbf{L}) + \epsilon} \quad (4)$$

where ϵ is a small constant that averts zero values on the denominator. It should be noted, that by setting the w_is as given above, the weighted nuclear norm becomes concave penalizing more heavily smaller values and less the larger ones. This scheme brings forth challenging theoretical issues, however experimental results have shown clear benefits arising by the use of the reweithted nuclear norm in different applications, [5]. Along similar research lines, the Schatten-p norm has been utilized in [7] for generalizing the nuclear norm to the case that the ℓ_p norm with $p \in [0, 1]$ is applied on the singular values. More specifically, in [7], the authors incorporated also the sparse component (which models the presence of outliers) into the formulated optimization problem by using the ℓ_p norm as the data fitting term. That said, robust PCA is now trasformed to the following optimization problem

$$\min_{\mathbf{L}} \ \|\mathbf{L}\|_{\mathcal{S}_p}^p + \|\mathbf{X} - \mathbf{L}\|_q^q \quad (5)$$

with $q \in [0, 1]$ and $\|\cdot\|_{\mathcal{S}_p}$ denoting the Schatten-p norm.

Recently, low-rank constraints have been incorporated in the optimization problems via *matrix factorization* based approaches. The main premise of this class of methods is the reduced computational complexity they offer which makes them amenable to handling large scale data. Assuming a rank-r matrix $\mathbf{L}$, it is known that there exist matrices $\mathbf{U}, \mathbf{V}$ so that $\mathbf{L} = \mathbf{U}\mathbf{V}^T$. In that respect, robust PCA can now be stated as a nonconvex optimization problem with respect to matrices $\mathbf{U}, \mathbf{V}$ which can be solved via alternating minimization strategies see e.g. [8]. However, in many real applications the rank is not known beforehand. This fact has given rise to schemes that first assume an overestimate $d \geq r$ for the number of columns of $\mathbf{U}, \mathbf{V}$. Then, both matrix factors are suitably penalized so that -ideally- a matrix $\mathbf{L}$ of rank-r is produced from their product. In that framework, the variational form of the nuclear norm, defined as

$$\|\mathbf{L}\|_* = \min_{\mathbf{L}=\mathbf{U}\mathbf{V}^T, \mathbf{U}\in\mathcal{R}^{m\times d}, \mathbf{V}\in\mathcal{R}^{n\times d}} \ \frac{1}{2}\left(\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2\right) \quad (6)$$

has been applied. In [9], (6) was incorporated in the robust PCA problem, as follows,

$$\min_{\mathbf{U}\in\mathcal{R}^{m\times d}, \mathbf{V}\in\mathcal{R}^{n\times d}} \ \|\mathbf{X} - \mathbf{U}\mathbf{V}^T\|_1 + \frac{\lambda}{2}\left(\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2\right), \quad (7)$$

where λ is a low-rank regularization parameter. In (7) the data fitting term is penalized via the ℓ_1 norm in order to take into account the fact that data have been corrupted by outliers.

More recently two novel matrix factorization based robust PCA methods were proposed in [3]. The crux of those methods was the generalization of the variational form of the nuclear norm to the case of Schatten-p norms with $p \in [0, 1]$ as follows,

$$\|\mathbf{L}\|_{S_p}^p = \min_{\mathbf{L}=\mathbf{U}\mathbf{V}^T, \mathbf{U}\in\mathcal{R}^{m\times d}, \mathbf{V}\in\mathcal{R}^{n\times d}} \ \frac{p_2\|\mathbf{U}\|_{S_{p_1}}^{p_1} + p_1\|\mathbf{V}\|_{S_{p_2}}^{p_2}}{p_1 + p_2} \quad (8)$$

where $p = \frac{p_1 p_2}{p_1 + p_2}$. Two special instances arising for $\{p_1 = 1, p_2 = 1, p = \frac{1}{2}\}$ and $\{p_1 = 2, p_2 = 1, p = \frac{2}{3}\}$ were then utilized for the robust PCA problem. With the aim to remain to the ℓ_p quasi-norm minimization setting, the authors utilized the relevant ℓ_p norm for penalizing the sparse component, giving thus rise to the two different optimization problems arising by minimizing two different sums of norms a) $\|\mathbf{L}\|_{S_{\frac{1}{2}}}^{\frac{1}{2}} + \|\mathbf{S}\|_{\frac{1}{2}}^{\frac{1}{2}}$ and b) $\|\mathbf{L}\|_{S_{\frac{2}{3}}}^{\frac{2}{3}} + \|\mathbf{S}\|_{\frac{2}{3}}^{\frac{2}{3}}$.

3. PROPOSED PROBLEM FORMULATION

In this paper, capitalizing on the advances that the above-mentioned matrix factorization based approaches have offered to robust PCA, we propose a novel and generic low-rank imposing scheme which applies on both matrix factors

$\mathbf{U}, \mathbf{V}$. This scheme can be considered as a weighted version of the variational form of the nuclear norm and is defined as

$$h(\mathbf{U}, \mathbf{V}) = \|\mathbf{U}\mathbf{W}_{\mathbf{U}}^{\frac{1}{2}}\|_F^2 + \|\mathbf{V}\mathbf{W}_{\mathbf{V}}^{\frac{1}{2}}\|_F^2. \quad (9)$$

Due to the relation of the weighted Frobenius norm to the Schatten-p norm, [10], it can be easily shown that $h(\mathbf{U}, \mathbf{V})$ can be seen as a generalization of (6). Interestingly, the proposed low-rank imposing term besides (6), it also encompasses the recently proposed regularization (8), for specific matrices $\mathbf{W}_{\mathbf{U}}, \mathbf{W}_{\mathbf{V}}$, [11]. Herein, we focus on a special case of $h(\mathbf{U}, \mathbf{V})$ which arises for a common weighting matrix $\mathbf{W} = \mathbf{W}_{\mathbf{U}} = \mathbf{W}_{\mathbf{V}}$ defined as

$$\mathbf{W} = \mathrm{diag}\Big(\left(\|\boldsymbol{u}_1\|_2^2 + \|\boldsymbol{v}_1\|_2^2 + \epsilon\right)^{p-1}, \left(\|\boldsymbol{u}_2\|_2^2 + \|\boldsymbol{v}_2\|_2^2 + \epsilon\right)^{p-1}, \ldots, \left(\|\boldsymbol{u}_d\|_2^2 + \|\boldsymbol{v}_d\|_2^2 + \epsilon\right)^{p-1} \Big), \quad (10)$$

where $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ are the ith columns of $\mathbf{U}$ and $\mathbf{V}$, respectively and the small positive scalar ϵ has been included as in (4). Note that the diagonal matrix $\mathbf{W}$ weighs equally the corresponding columns of the factors $\mathbf{U}$ and $\mathbf{V}$. Moreover, this particular selection of $\mathbf{W}$ gives rise to a reweighting-type scheme since each of the columns $\boldsymbol{u}_i, \boldsymbol{v}_i$ in (9) is weighted by a term $\left(\|\boldsymbol{u}_i\|_2^2 + \|\boldsymbol{v}_i\|_2^2 + \epsilon\right)^{\frac{p-1}{2}}$ which contains its squared ℓ_2 norm and resembles the matrix factorization analogue of the reweighted nuclear norm minimization problem described above. To make it more clear, it can be considered that each column $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ ($i = 1, 2, \ldots, d$) of matrices $\mathbf{U}, \mathbf{V}$ is weighted by a w_i defined as

$$w_i = \frac{1}{\left(\|\boldsymbol{u}_i\|_2^2 + \|\boldsymbol{v}_i\|_2^2 + \epsilon\right)^{\frac{1-p}{2}}}. \quad (11)$$

Using (11), we get

$$h(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^{d} \left(\|\boldsymbol{u}_i\|_2^2 + \|\boldsymbol{v}_i\|_2^2 + \epsilon\right)^p. \quad (12)$$

Remark 1: The regularization term of eq. (12) can be actually viewed as a group sparsity imposing term on the columns $\begin{bmatrix} \boldsymbol{u}_i \\ \boldsymbol{v}_i \end{bmatrix}$ *of the concatenated matrix* $\begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix}$. *Specifically, for* $p = \frac{1}{2}$ *it is the* ℓ_1/ℓ_2 *norm of this matrix.*

In the framework of robust PCA and in an effort to remain into the same reweighting philosophy, we propose to account for the sparse matrix of outliers $\mathbf{S}$ by utilizing the reweighted ℓ_1 norm i.e., $\|\mathbf{S}\|_{1,\mathbf{P}} = \sum_{i=1,j=1}^{m,n} p_{ij}|s_{ij}|$ where $p_{ij} = \frac{1}{|s_{ij}|+\epsilon}$.

All in all, robust PCA is now formulated as follows

$$\min_{\mathbf{U}\in\mathcal{R}^{m\times d}, \mathbf{V}\in\mathcal{R}^{n\times d}, \mathbf{S}\in\mathcal{R}^{m\times n}} \|\mathbf{X} - (\mathbf{U}\mathbf{V}^T + \mathbf{S})\|_F^2 + \lambda \sum_{i=1}^{d} \left(\|\boldsymbol{u}_i\|_2^2 + \|\boldsymbol{v}_i\|_2^2 + \epsilon\right)^p + \mu\|\mathbf{S}\|_{1,\mathbf{P}} \quad (13)$$

where μ is the regularization parameter of the reweighted ℓ_1 norm.

4. OPTIMIZATION ALGORITHM

In this section, we aim at deriving an efficient algorithm for solving (13). Let us focus on the the cost function $f(\mathbf{U}, \mathbf{V}, \mathbf{S})$ for $p = \frac{1}{2}$ i.e.,

$$f(\mathbf{U}, \mathbf{V}, \mathbf{S}) = \frac{1}{2}\|\mathbf{X} - (\mathbf{U}\mathbf{V}^T + \mathbf{S})\|_F^2 + \lambda \sum_{i=1}^{d} \sqrt{\|\boldsymbol{u}_i\|_2^2 + \|\boldsymbol{v}_i\|_2^2 + \epsilon} + \|\mathbf{S}\|_{1,\mathbf{P}}. \quad (14)$$

It should be noted that this cost function is nonconvex, and hence any derived minimization algorithm must be carefully initialized so as to avoid "bad" local minima. Moreover, the problem is non-separable (actually this happens for any $p \in [0, 1]$) with respect to $\mathbf{U}, \mathbf{V}$. That said, closed form solutions for $\mathbf{U}, \mathbf{V}$ are not available rendering a plain alternating minimization scheme not applicable.

To circumvent this difficulty we come up with a block successive upper bound minimization approach (BSUM), [4]. The main idea of this approach is to alternatingly update matrices $\mathbf{U}, \mathbf{V}$ and $\mathbf{S}$ by iteratively minimizing local tight upper-bounds of the cost function. In this framework, at iteration $k+1$, matrices $\mathbf{U}$ and $\mathbf{V}$ are updated by $\mathbf{U}^{k+1} = \arg\min_{\mathbf{U}} l(\mathbf{U}|\mathbf{U}^k, \mathbf{V}^k, \mathbf{S}^k)$ and $\mathbf{V}^{k+1} = \arg\min_{\mathbf{V}} g(\mathbf{V}|\mathbf{U}^{k+1}, \mathbf{V}^k, \mathbf{S}^k)$ where

$$l(\mathbf{U}|\mathbf{U}^k, \mathbf{V}^k, \mathbf{S}^k) = \|\mathbf{X} - (\mathbf{U}\mathbf{V}^{k,T} + \mathbf{S}^k)\|_F^2 + \lambda \sum_{i=1}^{d} \frac{\|\boldsymbol{u}_i\|_2^2 + \|\boldsymbol{v}_i^k\|_2^2 + \epsilon}{\sqrt{\|\boldsymbol{u}_i^k\|_2^2 + \|\boldsymbol{v}_i^k\|_2^2 + \epsilon}} + \mu\|\mathbf{S}^k\|_{1,\mathbf{P}} \quad (15)$$

$$g(\mathbf{U}|\mathbf{U}^{k+1}, \mathbf{V}^k, \mathbf{S}^k) = \|\mathbf{X} - (\mathbf{U}^{k+1}\mathbf{V}^T + \mathbf{S}^k)\|_F^2 + \lambda \sum_{i=1}^{d} \frac{\|\boldsymbol{u}_i^{k+1}\|_2^2 + \|\boldsymbol{v}_i\|_2^2 + \epsilon}{\sqrt{\|\boldsymbol{u}_i^{k+1}\|_2^2 + \|\boldsymbol{v}_i^k\|_2^2 + \epsilon}} + \mu\|\mathbf{S}^k\|_{1,\mathbf{P}}. \quad (16)$$

It can be easily shown that $l(\mathbf{U}|\mathbf{U}^k, \mathbf{V}^k, \mathbf{S}^k) \geq f(\mathbf{U}, \mathbf{V}^k, \mathbf{S}^k)$ and $g(\mathbf{U}|\mathbf{U}^{k+1}, \mathbf{V}^k, \mathbf{S}^k) \geq f(\mathbf{U}^{k+1}, \mathbf{V}, \mathbf{S}^k)$ by using the arithmetic-geometric mean inequality, [4]. Lastly, matrix $\mathbf{S}$ is updated by solving a minimization problem of the form

$$\mathbf{S}^{k+1} = \underset{\mathbf{S}}{\mathrm{argmin}} \|(\mathbf{X} - \mathbf{U}^{k+1}\mathbf{V}^{T,k+1,}) - \mathbf{S}\|_F^2 + \mu\|\mathbf{S}\|_{1,\mathbf{P}}. \quad (17)$$

This problem has a closed form solution, given as

$$[\mathbf{S}^{k+1}]_{ij} = \mathrm{ST}([\mathbf{X} - \mathbf{U}^{k+1}\mathbf{V}^{T,k+1}]_{ij}, \frac{\mu}{|x_{ij} - [\mathbf{U}^{k+1}\mathbf{V}^{T,k+1}]_{ij}| + \epsilon}) \quad (18)$$

where $\mathrm{ST}(x, \alpha) = \mathrm{sign}(x) \times (|x| - a)$. The resulting algorithm is given in Algorithm 1 and its convergence properties are summarized in the following Proposition, whose proof is not provided due to space limitations.

Proposition 1: The sequence of $\{\mathbf{U}^k, \mathbf{V}^k, \mathbf{S}^k\}$ *generated by Algorithm 1 converges to a stationary point of the cost function* $f(\mathbf{U}, \mathbf{V}, \mathbf{S})$.

Proof: Convergence to stationary points can be established using arguments similar to those provided in [11].

5. EXPERIMENTAL RESULTS

In this section, we present simulated and real data experiments that corroborate the efficiency of the proposed Algorithm 1 in addressing the robust PCA problem. In order to

Algorithm 1: Robust PCA Alternating iteratively reweighted least squares (RAIRLS) algorithm

Input: $\mathbf{X}, \lambda > 0, \mu > 0, d$
Initialize: $k = 0, \mathbf{V}_0, \mathbf{U}_0$
repeat
$\mathbf{D} = \text{diag}(\mathbf{d}), \text{d}_i = \frac{1}{\sqrt{\left(\|u_i^k\|_2^2 + \|v_i^k\|_2^2 + \epsilon\right)}}, i = 1, 2, \ldots, d$
$\mathbf{U}_{k+1} = \left(\mathbf{X} - \mathbf{S}^k\right)^T \mathbf{V}_k \left(\mathbf{V}_k^T \mathbf{V}_k + \lambda \mathbf{D}\right)^{-1}$
$\mathbf{V}_{k+1} = \left(\mathbf{X} - \mathbf{S}^k\right) \mathbf{U}_{k+1}^T \left(\mathbf{U}_{k+1}^T \mathbf{U}_{k+1} + \lambda \mathbf{D}\right)^{-1}$
$\mathbf{S}_{k+1}$ is computed from (18)
$k = k + 1$
until *convergence*
Output: $\hat{\mathbf{U}} = \mathbf{U}_{k+1}, \hat{\mathbf{V}} = \mathbf{V}_{k+1}, \hat{\mathbf{S}} = \mathbf{S}^{k+1}$

m=n	RAIRLS		$(S+L)_{\frac{1}{2}}$		$(S+L)_{\frac{2}{3}}$		WNNM	
	NRE	F-M	NRE	F-M	NRE	F-M	NRE	F-M
500	0.052	0.845	0.072	0.840	0.077	0.839	0.058	0.842
1000	0.037	0.848	0.044	0.847	0.047	0.846	0.045	0.846

Table 1. Simulated data - NRE and F-M results.

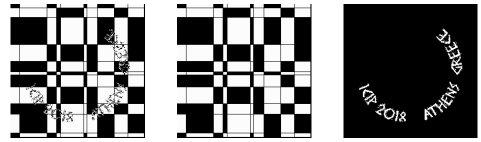
better highlight the efficiency of the proposed approach we compare it to three state-of-art algorithms i.e., Weighted Nuclear Norm Minimization (WNNM), [12], and the two matrix factorization based algorithms of [3], namely the $(S+L)_{\frac{1}{2}}$ and the $(S+L)_{\frac{2}{3}}$. As performance metrics we utilized the normalized reconstruction error defined as $\text{NRE} = \frac{\|\mathbf{L}-\mathbf{L}_0\|_F^2}{\|\mathbf{L}_0\|_F^2}$ for evaluating the recovery of the low-rank component, and the F-measure, [3], in order to evaluate the recovery of the sparse component. The proposed algorithm stops when either a maximum number of 500 iterations limit is reached or the following criterion is satisfied: $\frac{\|\hat{\mathbf{U}}^{k+1}\hat{\mathbf{V}}^{T,k+1} - \hat{\mathbf{U}}^k\hat{\mathbf{V}}^{T,k}\|_F^2}{\|\hat{\mathbf{U}}^k\hat{\mathbf{V}}^{T,k}\|_F^2} \leq 10^{-4}$. The same stopping rule is followed for the rest algorithms with their stopping criteria adopted as proposed in the relevant papers, [12], [3]. Finally, the regularization parameters of the proposed algorithm and its rivals have been fine tuned (WNNM, $(S+L)_{\frac{2}{3}}$ and $(S+L)_{\frac{1}{2}}$ where optimized as in [3]). It should be also noted, that a *column pruning mechanism* for the columns that have been zeroed by the proposed RAIRLS algorithm was adopted, as further explained in [11].

5.1. Simulated data experiment

In this experiment we generate a low-rank matix $\mathbf{L}_0 \in \mathcal{R}^{m\times n}$ of rank r as the product $\mathbf{U}_0\mathbf{V}_0^T$ of matrices $\mathbf{U}_0 \in \mathcal{R}^{m\times r}$ and $\mathbf{V}_0 \in \mathcal{R}^{n\times r}$ whose elements are i.i.d. random samples from a zero mean and unit variance Gaussian distribution. A sparse matrix $\mathbf{S}_0$ is also generated with support set chosen uniformly at random and its non-zero entries being uniformly i.i.d. in the interval $[-5, 5]$. Data matrix $\mathbf{X}$ is also corrupted by addtional

m=n	Average time (sec)			
	RAIRLS	$(S+L)_{\frac{1}{2}}$	$(S+L)_{\frac{2}{3}}$	WNNM
500	5.66	8.42	4.82	14.92
1000	27.24	44.50	43.50	98.05

Table 2. Simulated data - average time of execution.



(a) (b) (c)

Fig. 1. a) input image b) $\mathbf{L}_0$ and c) $\mathbf{S}_0$

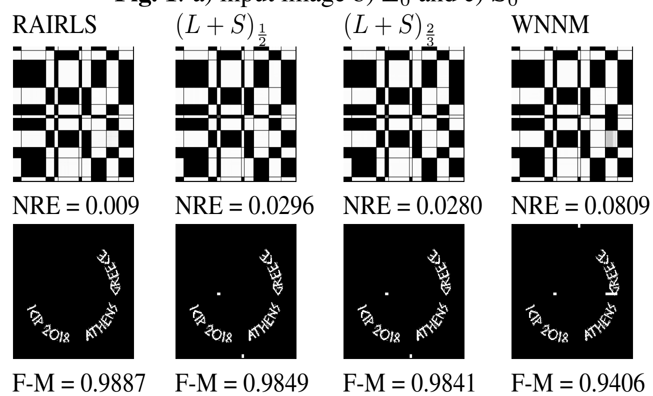


Fig. 2. Recoverd low-rank and text images.

Gaussian i.i.d noise of variance $\sigma^2 = 0.25$. Experiments are conducted for two different cases $m = n = \{500, 1000\}$. In both cases, the matrix factorization based algorithms are provided an overestimate $d = 20$ of the true rank $r = 10$. Both experiments are executed for 10 different realizations and the mean values of the NRE and the F-M for each of the algorithms is given in Table 1. As it can be seen from Table 1 the proposed RAIRLS presents promising results outperforming its counterparts in terms of both the performance metrics. Moreover, as shown in Table 2, RAIRLS requires much less time than the other algorithms in the large scale setting i.e., $m = n = 1000$. This favorable property of RAIRLS results from the pruning of columns of the matrix factors that have been zeroed due to the group-sparsity inducing nature of the proposed low-rank regularization term.

5.2. Real data experiment

Herein, we test the performance of the proposed algorithm to a low-level vision problem i.e, that of text removal. To this end, we apply RAIRLS and the other competing algorithms on the 256×256 pixels image of rank $r = 10$, used in [3], which is corrupted by outlier noise, which corresponds to text. The corrupted input image, the low-rank ground truth image and the text image are given in Fig. 1. All algorithms taking part in the experiment are initialized to an overestimate of $d = 20$ of the true rank r. In Fig. 2, the recovered (by all the algorithms) low-rank components of the image as well as the text images are shown. Moreover, as quantitative metrics of the performance we provide the NRE of the low-rank component and the F-M of the text image. The obtained by RAIRLS results are comparable with those of the state-of-the-art relevants and slightly better when it comes to the NRE, thus showing the competence of RAIRLS in handling real data.

6. REFERENCES

[1] Fernando De La Torre and Michael J Black, "A framework for robust subspace learning," *International Journal of Computer Vision*, vol. 54, no. 1-3, pp. 117–142, 2003.

[2] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright, "Robust principal component analysis?," *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 11, 2011.

[3] Fanhua Shang, James Cheng, Yuanyuan Liu, Zhi-Quan Luo, and Zhouchen Lin, "Bilinear factor matrix norm minimization for robust pca: Algorithms and applications," *IEEE transactions on Pattern Analysis and Machine Intelligence*, 2017.

[4] Mingyi Hong, Meisam Razaviyayn, Zhi-Quan Luo, and Jong-Shi Pang, "A unified algorithmic framework for block-structured optimization involving big data: With applications in machine learning and signal processing," *IEEE Signal Processing Magazine*, vol. 33, no. 1, pp. 57–77, 2016.

[5] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2862–2869.

[6] Emmanuel J Candes, Michael B Wakin, and Stephen P Boyd, "Enhancing sparsity by reweighted 1 minimization," *Journal of Fourier analysis and applications*, vol. 14, no. 5-6, pp. 877–905, 2008.

[7] Feiping Nie, Heng Huang, and Chris HQ Ding, "Low-rank matrix recovery via efficient schatten p-norm minimization.," in *AAAI*, 2012.

[8] Prateek Jain, Praneeth Netrapalli, and Sujay Sanghavi, "Low-rank matrix completion using alternating minimization," in *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*. ACM, 2013, pp. 665–674.

[9] Ricardo Cabral, Fernando De la Torre, João P Costeira, and Alexandre Bernardino, "Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition," in *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE, 2013, pp. 2488–2495.

[10] Karthik Mohan and Maryam Fazel, "Iterative reweighted algorithms for matrix rank minimization," *Journal of Machine Learning Research*, vol. 13, no. Nov, pp. 3441–3473, 2012.

[11] P. V Giampouras, A. A Rontogiannis, and K. D Koutroumbas, "Alternating iteratively reweighted minimization algorithms for low-rank matrix factorization," *arXiv preprint arXiv:1710.02004*, 2017.

[12] Shuhang Gu, Qi Xie, Deyu Meng, Wangmeng Zuo, Xiangchu Feng, and Lei Zhang, "Weighted nuclear norm minimization and its applications to low level vision," *International Journal of Computer Vision*, vol. 121, no. 2, pp. 183–208, 2017.