# A Graph Regularized RPCA by Generalized Moreau Enhanced Model 

Ryota Komatsu, Masao Yamagishi, and Isao Yamada<br>Dept. of Information and Communications Engineering<br>Tokyo Institute of Technology, Tokyo, Japan<br>Email:\{komatsu,myamagi,isao\} @sp.ict.e.titech.ac.jp


#### Abstract

A robust principal component analysis (RPCA) on graphs [Shahid et al., 2016] shows that a quadratic function, say the graph regularizer, designed with two graph Laplacians, can serve as a computationally efficient low-rankness promoting regularizer. In this paper, we present a novel RPCA by combining the graph regularizer with a generalized-Moreau-nonconvexenhancement of L1 norm. The proposed graph regularized RPCA (GRPCA) model uses a nonconvex penalty while maintaining the overall convexity and can be solved with a proximal splitting type algorithm in [Abe, Yamagishi, and Yamada, 2020]. A numerical experiment in a scenario of foreground and background decomposition of a video demonstrates the efficacy of the proposed GRPCA model.


Index Terms-Robust PCA, Linearly involved generalized Moreau enhanced (LiGME) model, graph regularized PCA, proximal splitting algorithm

## I. Introduction

The classical principal component analysis (PCA), yet used most widely for data analysis [1], is known to have brittleness against grossly corrupted observations even at a small number of entries of data [2], [3]. The robust PCA (RPCA) [2]-[4] has been introduced as a decomposition of a matrix into the sum of a low-rank matrix and a sparse matrix. For a given $M \in \mathbb{R}^{p \times n}$, an ideal goal behind the RPCA is to solve

$$
\begin{equation*}
\operatorname{minimize}_{S \in \mathbb{R}^{p \times n}} \quad \operatorname{rank}(M-S)+\mu\|S\|_{0}, \quad \mu>0 \tag{1}
\end{equation*}
$$

where $\|\cdot\|_{0}$ stands for the number of nonzero entries in a matrix [3]. However, since this problem (1) is NP-hard, its convex relaxations and their algorithmic solutions have been investigated [2]-[4].

In [2], $\operatorname{rank}(\cdot)$ and $\|\cdot\|_{0}$ are replaced by their convex envelopes, i.e. the largest convex lower bounds, as

$$
\begin{equation*}
\operatorname{minimize}_{S \in \mathbb{R}^{p \times n}} \quad\|M-S\|_{\text {nuc }}+\mu\|S\|_{1}, \quad \mu>0 \tag{2}
\end{equation*}
$$

where $\|\cdot\|_{\text {nuc }}$ and $\|\cdot\|_{1}$ stand respectively for the sum of all singular values and the sum of absolute values of all entries. The problem (2) has been solved in [2] with the alternating direction methods [5], [6]. In [3], the DouglasRachford splitting [7] has been applied to the model (2), and this idea has also been extended to tensor completion [8] and hypercomplex generalization of RPCA [9]. To approximate the model (1) more than via (2), replacement of $\|\cdot\|_{\text {nuc }}$ and $\|\cdot\|_{1}$ in (2) with other nonconvex functions have been investigated. For

[^0]example, the $\ell_{p}$ norm $(0<p<1)$ and the minimax concave (MC) penalty [10] were used as more faithful nonconvex regularizers to $\operatorname{rank}(M-\cdot)$ and $\|\cdot\|_{0}$ [11], [12]. However, finding a global minimizer of such nonconvex models has not been guaranteed. Moreover, optimization algorithms in [2], [3], [11], [12] require the singular value decomposition (SVD) to cope with singular values at each iteration. Despite the great effort for computational reduction of the SVD (see, e.g. [13]), replaceable convex optimization models, of the model (2), not requiring the SVD have been desired.

Shahid et al. in [4] proposed to formulate the RPCA as

$$
\begin{equation*}
\underset{S \in \mathbb{R}^{p \times n}}{\operatorname{minimize}} \quad Q_{\omega}(M-S)+\mu\|S\|_{1}, \quad \mu>0 \tag{3}
\end{equation*}
$$

where $Q_{\omega}: \mathbb{R}^{p \times n} \rightarrow \mathbb{R}_{+}:$

$$
\begin{aligned}
X \mapsto & \frac{\omega}{2} \sum_{i=1}^{p} \sum_{j=1}^{n} W_{i, j}^{(1)}\left\|X_{\cdot, i}-X_{\cdot, j}\right\|^{2} \\
& +\frac{1-\omega}{2} \sum_{i=1}^{p} \sum_{j=1}^{n} W_{i, j}^{(2)}\left\|\left(X_{i, \cdot}-X_{j, \cdot}\right)^{\top}\right\|^{2}, \omega \in[0,1]
\end{aligned}
$$

and $W_{i, j}^{(1)}, W_{i, j}^{(2)} \geq 0$ are respectively designed by pairwise distances, e.g. the Euclidean norm $\|\cdot\|$, between column vectors and row vectors of $M$. In the model (3), we can expect that suppression of $Q_{\omega}(M-\cdot)$ enhances the closenesses of pair of column vectors $\left((M-S)_{\cdot, i},(M-S)_{\cdot, j}\right)$ for large $W_{i, j}^{(1)}$ and pair of row vectors $\left((M-S)_{i, .},(M-S)_{j, .}\right)$ for large $W_{i, j}^{(2)}$ and thus promotes a low-rankness of $M-S$ in the model (1). In the context of graph signal processing, $Q_{\omega}(M-\cdot)$ can be interpreted as an instance of the so-called graph regularizer (see (4) and [4], [14]). Indeed, Shahid et al. in [4] experimentally shows that $Q_{\omega}(M-\cdot)$ can serve as an approximation of $\operatorname{rank}(M-\cdot)$ in view of an eigenvalue analysis. We remark that the model (3) can be viewed as a kind of sparse regularized least square convex model, which motivates us to improve further the model (3) by employing a certain better nonconvex approximation of $\|\cdot\|_{0}$ while keeping the overall convexity for global optimization of such an improved model (Note: If $\|S\|_{1}$ is replaced by $\|S\|_{0}$ in the model (3), the problem becomes again NP-hard [15], [16].).

In this paper, to promote the sparsity than the model (3), we propose a new graph regularized RPCA (GRPCA) by using $Q_{\omega}(M-\cdot)$ together with a nonconvex sparsity promoting
function. The proposed sparsity promoting nonconvex function is designed, based on the generalized MC (GMC) penalty [17] and the Linearly involved generalized-Moreau-enhanced (LiGME) model [18] of which the latter was established recently, by subtracting Moreau-Yosida like regularization ${ }^{1}$ from $\|\cdot\|_{1}$. The proposed GRPCA model can maintain its overall convexity with a strategic parameter tuning found in [17]. Moreover, for the proposed GRPCA model, we can use proximal splitting type algorithms in [17], [18] which can produce a sequence converging to a global minimizer of the proposed model. A numerical experiment in a scenario of foreground and background decomposition of a video demonstrates that the proposed model recovers a low-rank component more clearly than the model (3).

## II. Preliminaries

## A. Notation

Let $\mathbb{N}, \mathbb{N}_{+}, \mathbb{R}, \mathbb{R}_{+}$, and $\mathbb{R}_{++}$be the sets of all nonnegative integers, positive integers, real numbers, nonnegative real numbers, and positive real numbers, respectively. $\|X\|_{2}$ is the maximum singular value of $X \in \mathbb{R}^{p \times n}$. The vectorization of a matrix is vec: $\mathbb{R}^{p \times n} \rightarrow \mathbb{R}^{p n}: X \mapsto\left[X_{\cdot, 1}{ }^{\top}, X_{\cdot, 2}{ }^{\top}, \cdots, X_{\cdot, n}{ }^{\top}\right]^{\top}$. Let $\left(\mathcal{H},\langle\cdot, \cdot\rangle_{\mathcal{H}},\|\cdot\|_{\mathcal{H}}\right)$ and $\left(\mathcal{K},\langle\cdot, \cdot\rangle_{\mathcal{K}},\|\cdot\|_{\mathcal{K}}\right)$ be finite dimensional real Hilbert spaces, i.e. finite dimensional inner product spaces over $\mathbb{R}$, and let $\mathcal{B}(\mathcal{H}, \mathcal{K})$ be the set of all bounded linear operators from $\left(\mathcal{H},\langle\cdot, \cdot\rangle_{\mathcal{H}},\|\cdot\|_{\mathcal{H}}\right)$ to $\left(\mathcal{K},\langle\cdot, \cdot\rangle_{\mathcal{K}},\|\cdot\|_{\mathcal{K}}\right)$. The positive definiteness and positive semidefiniteness of selfadjoint operator $X \in \mathcal{B}(\mathcal{H}, \mathcal{H})$ is denoted by $X \succ \mathrm{O}_{\mathcal{H}}$ and $X \succeq \mathrm{O}_{\mathcal{H}}$, respectively. Let $I_{n} \in \mathbb{R}^{n \times n}$ be the identity matrix, and let Id be the identity operator. Let $\Gamma_{0}(\mathcal{H})$ be the set of all proper lower semicontinuous convex functions $f: \mathcal{H} \rightarrow(-\infty, \infty]$. The proximity operator of $f \in \Gamma_{0}(\mathcal{H})$ is $\operatorname{Prox}_{f}: \mathcal{H} \rightarrow \mathcal{H}: x \mapsto \operatorname{argmin}_{y \in \mathcal{H}}\left(f(y)+(1 / 2)\|x-y\|_{\mathcal{H}}^{2}\right)$.

## B. Graph Construction for GRPCA

Let $G:=(V, E)$ be an undirected graph, where $V$ is the set of vertices and $E$ is the set of edges. The weighted adjacency matrix $W \in \mathbb{R}_{+}^{|V| \times|V|}$ is the symmetric matrix whose $(i, j)$ th entry is the weight of $\left(v_{i}, v_{j}\right)$ if $\left(v_{i}, v_{j}\right) \in E$ and 0 otherwise. The weighted degree matrix $D \in \mathbb{R}_{+}^{|V| \times|V|}$ is the diagonal matrix whose $(i, i)$ th entry is $D_{i, i}:=\sum_{j=1}^{|V|} W_{i, j}$. The graph Laplacian is $L:=D-W \in \mathbb{R}^{|V| \times|V|}$. Then, for any $n \in \mathbb{N}_{+}$ and $X \in \mathbb{R}^{n \times|V|}$, the following holds (See [20]):

$$
\begin{equation*}
\frac{1}{2} \sum_{i=1}^{|V|} \sum_{j=1}^{|V|} W_{i, j}\left\|X_{\cdot, i}-X_{\cdot, j}\right\|^{2}=\operatorname{tr}\left(X L X^{\top}\right) \tag{4}
\end{equation*}
$$

We show below how to construct $G$ from an $|V|$-tuple of $d$-dimensional real vectors $\boldsymbol{x}:=\left(x_{1}, x_{2}, \ldots, x_{|V|}\right) \in \mathbb{R}^{d} \times$ $\mathbb{R}^{d} \times \cdots \times \mathbb{R}^{d}$. In this paper, we employ the following graph construction found in [4].
(i) Let $K \in\{1,2, \ldots,|V|-1\}, V=\{1,2, \ldots,|V|\}$, and $E=\emptyset$. For all $i \in\{1,2, \ldots,|V|\}$, let $x_{i} \in \mathbb{R}^{d}$ be the graph signal on $i$.

[^1](ii) For all $i \in\{1,2, \ldots,|V|\}$, compare $\left\|x_{i}-x_{j}\right\|(j \in$ $\{1,2, \ldots,|V|\} \backslash\{i\})$ and register $(i, j)$ in $E$ if $x_{j}$ is in the $K$-nearest neighbors (KNN) from $x_{i}$.
(iii) Define the weighted adjacency matrix by
\[

W_{i, j}:= $$
\begin{cases}\exp \left(-\frac{\left\|x_{i}-x_{j}\right\|^{2}}{t^{2}}\right), & \text { if }(i, j) \in E \\ 0, & \text { otherwise }\end{cases}
$$
\]

where $t$ is the mean of distances of edges.
Let $M \in \mathbb{R}^{p \times n}$ be a matrix whose each column vector is a sample, let $W^{(1)} \in \mathbb{R}_{+}^{n \times n}$ and $L_{1} \in \mathbb{R}^{n \times n}$ be respectively the weighted adjacency matrix and the graph Laplacian of a graph constructed by samples, i.e. $\boldsymbol{x}:=\left(M_{\cdot, 1}, M_{\cdot, 2}, \ldots, M_{\cdot, n}\right)$, and let $W^{(2)} \in \mathbb{R}_{+}^{p \times p}$ and $L_{2} \in \mathbb{R}^{p \times p}$ be respectively the weighted adjacency matrix and the graph Laplacian of a graph constructed by features, i.e. $\boldsymbol{x}:=\left(M_{1, \cdot}{ }^{\top}, M_{2, \cdot}{ }^{\top}, \ldots, M_{p, \cdot}{ }^{\top}\right)$. In Sec. IV, we use GSPBOX [21] for implementation of the graph constructions.

## C. Linearly involved Generalized-Moreau-Enhanced Model

Let us introduce the linearly involved generalized-Moreauenhanced (LiGME) model [18] as a unified extension of [10], [17], [22], [23]. Inspired strongly by [17], the LiGME model [18] has been proposed with $\mu>0$ as

$$
\underset{x \in \mathcal{X}}{\operatorname{minimize}} \quad J_{\Psi_{B} \circ \mathfrak{L}}(x):=\frac{1}{2}\|y-A x\|_{\mathcal{Y}}^{2}+\mu \Psi_{B} \circ \mathfrak{L}(x),(5)
$$

where $\left(\mathcal{X},\langle\cdot, \cdot\rangle_{\mathcal{X}},\|\cdot\|_{\mathcal{X}}\right),\left(\mathcal{Y},\langle\cdot, \cdot\rangle_{\mathcal{Y}},\|\cdot\|_{\mathcal{Y}}\right),\left(\mathcal{Z},\langle\cdot, \cdot\rangle_{\mathcal{Z}},\|\cdot\|_{\mathcal{Z}}\right)$, and $\left(\tilde{\mathcal{Z}},\langle\cdot, \cdot\rangle_{\tilde{\mathcal{Z}}},\|\cdot\|_{\tilde{\mathcal{Z}}}\right)$ are finite dimensional real Hilbert spaces, $\operatorname{Prox}_{\gamma \Psi}\left(\forall \gamma \in \mathbb{R}_{++}\right)$is computable for coercive $\Psi \in \Gamma_{0}(\mathcal{Z}), \operatorname{dom} \Psi=\mathcal{Z}, \Psi \circ(-\mathrm{Id})=\Psi,(A, \mathfrak{L}, B, y, \mu) \in$ $\mathcal{B}(\mathcal{X}, \mathcal{Y}) \times \mathcal{B}(\mathcal{X}, \mathcal{Z}) \times \mathcal{B}(\mathcal{Z}, \tilde{\mathcal{Z}}) \times \mathcal{Y} \times \mathbb{R}_{++}$, and

$$
\Psi_{B}(\cdot):=\Psi(\cdot)-\min _{v \in \mathcal{Z}}\left[\Psi(v)+\frac{1}{2}\|B(\cdot-v)\|_{\tilde{\mathcal{Z}}}^{2}\right]
$$

is the generalized-Moreau-enhanced (GME) penalty function. A specialization of the model (5) for $(\mathcal{X}, \Psi)=\left(\mathbb{R}^{n},\|\cdot\|_{1}\right)$ and $B \in \mathbb{R}^{m \times n}$ reproduces the model [17]:

$$
\underset{x \in \mathbb{R}^{n}}{\operatorname{minimize}} \quad \frac{1}{2}\|y-A x\|^{2}+\mu\left(\|\cdot\|_{1}\right)_{B}(x), \quad \mu>0
$$

As shown in Fact 1 (ii), we can bridge parametrically the gap between $\|\cdot\|_{0}$ and $\|\cdot\|_{1}$ by introducing a parameter $B \in \mathbb{R}^{m \times n}$ in Moreau-Yosida like regularization of $\|\cdot\|_{1}$ and subtracting this regularization from $\|\cdot\|_{1}$ [18]. In [18], a proximal splitting algorithm to solve the general model (5) is also presented.
Fact 1 (See [18, Proposition 1, Example 2]). The LiGME penalty $\Psi_{B} \circ \mathfrak{L}$ has the following properties:
(i) Suppose that

$$
\begin{equation*}
A^{*} A-\mu \mathfrak{L}^{*} B^{*} B \mathfrak{L} \succeq \mathrm{O}_{\mathcal{X}} \tag{6}
\end{equation*}
$$

where $(\cdot)^{*}$ denotes adjoint. Then the LiGME model maintains the overall convexity, i.e. $J_{\Psi_{B} \circ \mathfrak{L}} \in \Gamma_{0}(\mathcal{X})$.
(ii) The LiGME penalty with $\mathfrak{L}=\mathrm{Id}$, i.e. the GMC penalty in [17], bridges the gap between $\|\cdot\|_{0}$ and $\|\cdot\|_{1}$ as

$$
\begin{aligned}
\left(\|\cdot\|_{1}\right)_{\mathrm{O}_{m, n}} \circ \mathrm{Id} & =\|\cdot\|_{1}, \\
\lim _{\gamma \downarrow 0} \frac{2}{\gamma}\left(\|\cdot\|_{1}\right)_{\frac{1}{\sqrt{\gamma}} \mathrm{Id}} \circ \mathrm{Id} & =\|\cdot\|_{0} .
\end{aligned}
$$

## III. A Moreau Enhancement for GrPCA

## A. GRPCA by Generalized-Moreau-Enhanced (GME) Model

To achieve the goal of RPCA, we can employ by [4] the graph regularizer $Q_{\omega}(M-\cdot)$ as a low-rankness promoting function in place of $\|\cdot\|_{\text {nuc }}$. Since this graph regularizer is a quadratic convex function, we can promote the sparsity of $S$ with the aid of a more faithful promoting function (than $\|\cdot\|_{1}$ ) to $\|\cdot\|_{0}$. Along this strategy, the model (3) can be enhanced naturally as

$$
\operatorname{minimize}_{S \in \mathbb{R}^{p \times n}} Q_{\omega}(M-S)+\mu\left(\|\cdot\|_{1}\right)_{B} \circ \operatorname{Id}(\operatorname{vec}(S)), \mu>0,(7)
$$

where $B \in \mathbb{R}^{p n \times p n}$. Clearly, the model (7) can be reformulated as an instance of the model (5).

Let $s:=\operatorname{vec}(S)$, let $\boldsymbol{m}:=\operatorname{vec}(M)$, and let $A \in \mathbb{R}^{p n \times p n}$ such that $A^{\top} A=\omega\left(L_{1} \otimes I_{p}\right)+(1-\omega)\left(I_{n} \otimes L_{2}\right)$, where $X \otimes Y \in \mathbb{R}^{p n \times p n}$ stands for the Kronecker product of $X \in$ $\mathbb{R}^{n \times n}$ and $Y \in \mathbb{R}^{p \times p}$. Then, by letting $\boldsymbol{y}:=A \boldsymbol{m}$, the proposed model (7) can be expressed as

$$
\begin{equation*}
\operatorname{minimize}_{\boldsymbol{s} \in \mathbb{R}^{p n}} J_{\left(\|\cdot\|_{1}\right)_{B}}(\boldsymbol{s}):=\frac{1}{2}\|\boldsymbol{y}-A \boldsymbol{s}\|^{2}+\mu\left(\|\cdot\|_{1}\right)_{B}(\boldsymbol{s}) \tag{8}
\end{equation*}
$$

where $J_{\left(\|\cdot\|_{1}\right)_{B}} \in \Gamma_{0}\left(\mathbb{R}^{p n}\right)$ is achieved (See [17]) if

$$
\begin{equation*}
A^{\top} A-\mu B^{\top} B \succeq \mathrm{O}_{\mathbb{R}^{p n}} \tag{9}
\end{equation*}
$$

The model (7) reproduces (3) for $B=\mathrm{O}_{\mathbb{R}^{p n}}$ (See [17]).

## B. Optimization Algorithm

Under general overall convexity condition (6) for the model (5), a proximal splitting type algorithm of guaranteed convergence to a global minimizer was established in [18]. We apply this algorithm with $(\Psi, \mathfrak{L}, B)=\left(\|\cdot\|_{1}, \operatorname{Id}, \sqrt{\theta / \mu} A\right)(\theta \in$ $[0,1])$ to the model (8). Note that Selesnick [17] showed that the overall convexity condition (9) is achieved by

$$
\begin{equation*}
B:=\sqrt{\theta / \mu} A, \quad \theta \in[0,1] . \tag{10}
\end{equation*}
$$

The algorithm in [18] solves primal problem and dual problem simultaneously by introducing dual variables $\boldsymbol{v}$ and $\boldsymbol{w}$.
Fact 2 (See [18, Theorem 1 with $\mathfrak{L}=\mathrm{Id}$, Example 2]). For the model (8) under the overall convexity condition (9), let $\mathcal{H}:=$ $\mathbb{R}^{p n} \times \mathbb{R}^{p n} \times \mathbb{R}^{p n}$ and define $T_{\text {LiGME }}: \mathcal{H} \rightarrow \mathcal{H}:(s, \boldsymbol{v}, \boldsymbol{w}) \mapsto$ $(\boldsymbol{\xi}, \boldsymbol{\zeta}, \boldsymbol{\eta})$, with $(\sigma, \tau) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$, by
$\boldsymbol{\xi}:=\left[I_{p n}-\frac{1}{\sigma}\left(A^{\top} A-\mu B^{\top} B\right)\right] s-\frac{\mu}{\sigma} B^{\top} B \boldsymbol{v}-\frac{\mu}{\sigma} \boldsymbol{w}$ $+\frac{1}{\sigma} A^{\top} A \boldsymbol{m}$,
$\boldsymbol{\zeta}:=\operatorname{Prox} \frac{\mu}{\tau}\|\cdot\|_{1}\left[\frac{2 \mu}{\tau} B^{\top} B \boldsymbol{\xi}-\frac{\mu}{\tau} B^{\top} B \boldsymbol{s}+\left(I_{p n}-\frac{\mu}{\tau} B^{\top} B\right) \boldsymbol{v}\right]$
$\boldsymbol{\eta}:=2 \boldsymbol{\xi}-\boldsymbol{s}+\boldsymbol{w}-\operatorname{Prox}_{\|\cdot\|_{1}}(2 \boldsymbol{\xi}-\boldsymbol{s}+\boldsymbol{w})$.
Then the following hold:
(i) $\operatorname{argmin}_{\boldsymbol{s} \in \mathbb{R}^{p n}} J_{\left(\|\cdot\|_{1}\right)_{B}}(\boldsymbol{s})=\left\{\boldsymbol{s}^{\star} \in \mathbb{R}^{p n} \mid\left(\boldsymbol{s}^{\star}, \boldsymbol{v}^{\star}, \boldsymbol{w}^{\star}\right) \in\right.$ $\left.\operatorname{Fix}\left(T_{\text {LiGME }}\right)\right\}$, where $\operatorname{Fix}\left(T_{\text {LiGME }}\right):=\{x \in \mathcal{H} \mid x=$ $\left.T_{\text {LiGME }}(x)\right\}$ is the set of all fixed points of $T_{\text {LiGME }}$.

```
Algorithm 1 Algorithm for the models (7) and (3)
Input: \((\mu, \omega, \theta, \kappa, I, \varepsilon) \in \mathbb{R}_{++} \times[0,1] \times[0,1] \times(1, \infty) \times \mathbb{N}_{+} \times\)
    \(\mathbb{R}_{+}\)
Output: \(\left(M-S_{k+1}, S_{k+1}\right) \in \mathbb{R}^{p \times n} \times \mathbb{R}^{p \times n}\)
    \(\sigma:=(\kappa / 2)\left[\omega\left\|L_{1}\right\|_{2}+(1-\omega)\left\|L_{2}\right\|_{2}\right]+\mu+\kappa-1\)
    \(\tau:=(\kappa / 2+2 / \kappa) \theta\left[\omega\left\|L_{1}\right\|_{2}+(1-\omega)\left\|L_{2}\right\|_{2}\right]+\kappa-1\)
    \(\left(S_{0}, V_{0}, W_{0}\right) \in \mathbb{R}^{p \times n} \times \mathbb{R}^{p \times n} \times \mathbb{R}^{p \times n}\)
    for \(k=0,1, \ldots, I-1\) do
        \(S_{k+1} \leftarrow S_{k}-\frac{\mu}{\sigma} W_{k}-\frac{1}{\sigma}\left\{\omega\left[(1-\theta) S_{k}+\theta V_{k}-M\right] L_{1}\right.\)
                                \(\left.+(1-\omega) L_{2}\left[(1-\theta) S_{k}+\theta V_{k}-M\right]\right\}\)
        \(V_{k+1} \leftarrow \operatorname{Prox} \frac{\mu}{\tau}\|\cdot\|_{1}\left\{V_{k}+\frac{\theta}{\tau}\left[\omega\left(2 S_{k+1}-S_{k}-V_{k}\right) L_{1}\right.\right.\)
                                \(\left.\left.+(1-\omega) L_{2}\left(2 S_{k+1}-S_{k}-V_{k}\right)\right]\right\}\)
        \(W_{k+1} \leftarrow 2 S_{k+1}-S_{k}+W_{k}-\operatorname{Prox}_{\|\cdot\|_{1}}\left(2 S_{k+1}-S_{k}+W_{k}\right)\)
        if \(\|\left(\operatorname{vec}\left(S_{k+1}\right), \operatorname{vec}\left(V_{k+1}\right), \operatorname{vec}\left(W_{k+1}\right)\right)\)
                        \(-\left(\operatorname{vec}\left(S_{k}\right), \operatorname{vec}\left(V_{k}\right), \operatorname{vec}\left(W_{k}\right)\right) \|_{\mathfrak{P}}<\varepsilon\) then
            break
        end if
    end for
```

(ii) Choose $(\sigma, \tau, \kappa) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times(1, \infty)$ such that

$$
\begin{align*}
& \sigma I_{p n}-(\kappa / 2) A^{\top} A-\mu I_{p n} \succ \mathrm{O}_{\mathbb{R}^{p n}} \\
& \tau \geq(\kappa / 2+2 / \kappa) \mu\|B\|_{2}^{2} \tag{11}
\end{align*}
$$

Then, for

$$
\mathfrak{P}:=\left[\begin{array}{ccc}
\sigma I_{p n} & -\mu B^{\top} B & -\mu I_{p n} \\
-\mu B^{\top} B & \tau I_{p n} & \mathrm{O}_{\mathbb{R}^{p n}} \\
-\mu I_{p n} & \mathrm{O}_{\mathbb{R}^{p n}} & \mu I_{p n}
\end{array}\right] \succ \mathrm{O}_{\mathcal{H}}
$$

$\left(\mathcal{H},\langle\cdot, \cdot\rangle_{\mathfrak{P}},\|\cdot\|_{\mathfrak{P}}\right)$ is a real Hilbert space, where $\langle\cdot, \cdot\rangle_{\mathfrak{P}}: \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}:(x, y) \mapsto\langle x, \mathfrak{P} y\rangle_{\mathcal{H}}$ and $\|\cdot\|_{\mathfrak{P}}:=$ $\sqrt{\langle\cdot, \cdot\rangle_{\mathfrak{P}}} . T_{\mathrm{LiGME}}$ is $\kappa /(2 \kappa-1)$-averaged nonexpansive in $\left(\mathcal{H},\langle\cdot, \cdot\rangle_{\mathfrak{B}},\|\cdot\|_{\mathfrak{F}}\right)$.
(iii) Suppose that $(\sigma, \tau, \kappa) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times(1, \infty)$ satisfies (11). Then, for any $\left(s_{0}, \boldsymbol{v}_{0}, \boldsymbol{w}_{0}\right) \in \mathcal{H}$, the sequence $\left(\boldsymbol{s}_{k}, \boldsymbol{v}_{k}, \boldsymbol{w}_{k}\right)_{k \in \mathbb{N}} \subset \mathcal{H}$ generated by

$$
\begin{equation*}
\left(\boldsymbol{s}_{k+1}, \boldsymbol{v}_{k+1}, \boldsymbol{w}_{k+1}\right):=T_{\mathrm{LiGME}}\left(\boldsymbol{s}_{k}, \boldsymbol{v}_{k}, \boldsymbol{w}_{k}\right) \tag{12}
\end{equation*}
$$

converges to $\left(s^{\star}, \boldsymbol{v}^{\star}, \boldsymbol{w}^{\star}\right) \in \operatorname{Fix}\left(T_{\text {LiGME }}\right)$ such that

$$
s^{\star} \in \underset{s \in \mathbb{R}^{p n}}{\operatorname{argmin}} J_{\left(\|\cdot\|_{1}\right)_{B}}(s) .
$$

For the parameters $(\sigma, \tau, \kappa)$ satisfying (11), we use

$$
\begin{aligned}
\sigma & :=(\kappa / 2)\left[\omega\left\|L_{1}\right\|_{2}+(1-\omega)\left\|L_{2}\right\|_{2}\right]+\mu+\kappa-1 \\
\tau & :=(\kappa / 2+2 / \kappa) \theta\left[\omega\left\|L_{1}\right\|_{2}+(1-\omega)\left\|L_{2}\right\|_{2}\right]+\kappa-1
\end{aligned}
$$

which is a variant of the example in [18, Theorem 1]. Since we vectorized $S \in \mathbb{R}^{p \times n}$ into $s \in \mathbb{R}^{p n}$ in the model (8), we have to handle the large size matrices $A^{\top} A, B^{\top} B \in \mathbb{R}^{p n \times p n}$ in the algorithm (12). Fortunately, since $\operatorname{Prox}_{\gamma\|\cdot\|_{1}}: \mathbb{R}^{p \times n} \rightarrow$ $\mathbb{R}^{p \times n}:\left[X_{i, j}\right] \mapsto\left[\operatorname{sgn}\left(X_{i, j}\right) \max \left\{0,\left|X_{i, j}\right|-\gamma\right\}\right]\left(\gamma \in \mathbb{R}_{++}\right)$, which is the so-called soft-thresholding, is entrywise, we can matricize $\left(s_{k}, \boldsymbol{v}_{k}, \boldsymbol{w}_{k}\right) \in \mathcal{H}$ in (12) into $\left(S_{k}, V_{k}, W_{k}\right) \in$ $\mathbb{R}^{p \times n} \times \mathbb{R}^{p \times n} \times \mathbb{R}^{p \times n}$ in Algorithm 1. We remark that Algorithm 1 is scalable because it does not require any SVD and matrix inversion at each iteration.


Fig. 1: Foreground and background decomposition of a video. Full size and cropped frames are respectively shown in the first and second rows. The first and fourth columns show the original frames, and the second and third columns show the estimated low-rank components. We used the original frame 1683, just after the bus passed by, as the (estimated) ground truth of the low-rank component of the frame 1669.

Note that the model (3) can also be solved by Algorithm 1 with $\theta=0$, i.e. $B=\mathrm{O}_{\mathbb{R}^{p n}}$. Moreover, since the proposed model (8) with $\mathfrak{L}=\mathrm{Id}$ is rather simple compared with the model (5), the model (8) can also be solved by [17, Proposition $15]$ : for any $\left(\theta, \boldsymbol{s}_{0}, \boldsymbol{v}_{0}\right) \in[0,1) \times \mathbb{R}^{p n} \times \mathbb{R}^{p n}$,

$$
\begin{aligned}
& \boldsymbol{s}_{k+1}:=\operatorname{Prox}_{\nu \mu\|\cdot\|_{1}}\left(\boldsymbol{s}_{k}-\nu A^{\top} A\left(\boldsymbol{s}_{k}+\theta\left(\boldsymbol{v}_{k}-\boldsymbol{s}_{k}\right)-\boldsymbol{m}\right)(13)\right. \\
& \boldsymbol{v}_{k+1}:=\operatorname{Prox}_{\nu \mu\|\cdot\|_{1}}\left(\boldsymbol{v}_{k}-\nu \theta A^{\top} A\left(\boldsymbol{v}_{k}-\boldsymbol{s}_{k}\right)\right),
\end{aligned}
$$

where $\nu:=\nu^{\prime} /\left(\max \{1, \theta /(1-\theta)\}\left\|A^{\top} A\right\|_{2}\right)\left(\nu^{\prime} \in(0,2)\right)$ is the available stepsize to guarantee $\lim _{k \rightarrow \infty} \boldsymbol{s}_{k}=\boldsymbol{s}^{\star} \in$ $\operatorname{argmin}_{s \in \mathbb{R}^{p n}} J_{\left(\|\cdot\|_{1}\right){ }_{\sqrt{\theta / \mu} A}}(s)$. We remark that, in the model (8) with $B:=\sqrt{\theta / \mu} A$ in (10), the larger $\theta \in[0,1]$ enhances more the nonconvexity of $\left(\|\cdot\|_{1}\right)_{B}$ (Note: the algorithm (13) is not applicable to $\theta=1$ achieving (9) while Algorithm 1 is applicable to any $\theta \in[0,1]$ ).

## IV. Numerical Experiments

## A. Foreground and Background Decomposition of a Video

We applied Algorithm 1 to GRPCA-GME (7), i.e. the proposed GRPCA with a GME penalty, and GRPCA-L1 (3) as its special case with $\theta=0$, in a scenario of foreground and background decomposition of a video. In this scenario, a static background and a dynamic foreground can be seen as a low-rank component and a sparse component, respectively. Note that this scenario corresponds to a challenge for estimation of the hidden image information covered by an obstacle, i.e. the bus. We used the scene just after the bus passed by as the (estimated) ground truth of the low-rank component. Following [4], we used the first 1000 frames of a video in [24], which was found at http://vis-www.cs. umass.edu/ $\sim$ narayana/castanza/I2Rdataset/, converted them to grayscale, and rescaled a frame to $h \times w=64 \times 80$.

We defined the matrix $M \in \mathbb{R}^{h w \times 1000}$, whose $i$ th column $M_{\cdot, i} \in \mathbb{R}^{h w}$ is the vectorization of the $i$ th frame of a video. As the tunable parameters in Sec. II-B and Algorithm 1, we used $K=30$ for $L_{1}, K=10$ for $L_{2},(\mu, \omega, \kappa, I, \varepsilon)=$ $\left(0.1,0.9,1.1,10^{4}, 0\right),\left(S_{0}, V_{0}, W_{0}\right)=\left(\mathrm{O}_{p, n}, \mathrm{O}_{p, n}, \mathrm{O}_{p, n}\right)$ for both models, $\theta=0$ for GRPCA-L1, and $\theta=0.3$ for GRPCAGME. Among these parameters, $(\mu, \omega)=(0.1,0.9)$ is first chosen from $\{0.01,0.05,0.1,0.5\} \times\{0.7,0.8,0.9\}$ to achieve the least squared error between "the estimated background (by GRPCA-L1) from frame 1669 " and "the original frame 1683 (just after the bus passed by)". Then the parameter $\theta=0.3$ for GRPCA-GME is chosen from $\{0,0.1, \ldots, 0.9,1\}$ to achieve the least squared error. As seen in the second row in Fig. 1, GRPCA-GME seems to estimate the hidden image information (by the bus) more clearly than GRPCA-L1.

## B. Comparison of Estimation Accuracy

By recalling that the original goal of RPCA in [2] is estimation of a low-rank matrix $\mathcal{L}_{\star}$ and a sparse matrix $\mathcal{S}_{\star}$ from a given matrix $M\left(=\mathcal{L}_{\star}+\mathcal{S}_{\star}\right)$. In this subsection, we compare quantitatively the models (3) and (7) from the view point of estimation of $\mathcal{L}_{\star}$. Following [3, Sec. 5], we set $\mathcal{L}_{\star}:=X_{1} X_{2}^{\top} \in \mathbb{R}^{100 \times 100}$ and $\mathcal{S}_{\star} \in \mathbb{R}^{100 \times 100}$, where all entries of $X_{1}, X_{2} \in \mathbb{R}^{100 \times 5}$ are drawn from $\mathcal{N}(0,1)$, and randomly chosen 1000 nonzero entries of $\mathcal{S}_{\star}$ are drawn from the uniform distribution on $(-500,500)$. As the tunable parameters in Sec. II-B, Algorithm 1, and (13), we used $\left(K, \omega, \kappa, I, \varepsilon, \nu^{\prime}\right)=\left(10,0.5,1.1,10^{3}, 0,1.9\right),(\mu, \theta)=(3,0)$ for GRPCA-L1, $(\mu, \theta) \in\{13\} \times\{0.2,0.8,1.0\}$ for GRPCAGME, $\left(S_{0}, V_{0}, W_{0}\right)=\left(\mathrm{O}_{p, n}, \mathrm{O}_{p, n}, \mathrm{O}_{p, n}\right)$ for Algorithm 1, and $\left(s_{0}, \boldsymbol{v}_{0}\right)=(\mathbf{0}, \mathbf{0})$ for the algorithm (13). Among these parameters, $(\mu, \theta)=(3,0)$ for GRPCA-L1 is chosen from $\{1,2, \ldots, 19,20\} \times\{0\}$ to achieve the least squared error from


Fig. 2: The relative error $\left\|M-S_{k}-\mathcal{L}_{\star}\right\|_{F} /\left\|\mathcal{L}_{\star}\right\|_{F}$, where $\|\cdot\|_{F}$ is the Frobenius norm, and the value $J_{\left(\|\cdot\|_{1}\right)_{\sqrt{\theta / \mu} A}}\left(s_{k}\right)$.
$\mathcal{L}_{\star}$ just after 5000 iterations by Algorithm 1. For GRPCAGME, $(\mu, \theta)=(13,0.2)$ is chosen from $\{1,2, \ldots, 19,20\} \times$ $\{0.2,0.4,0.6,0.8,1\}$ to achieve the least squared error. As seen in Fig. 2, GRPCA-GME outperforms GRPCA-L1 in terms of the relative error. However, in view of [3, Table 1 (DR-PCP)], both models (3) and (7) do not seem to achieve comparable accuracies as estimations of $\mathcal{L}_{\star}$ in terms of the relative error in Fig. 2. We also compared convergence performances of Algorithm 1 and the algorithm (13). As seen in Fig. 2, the algorithm (13) converges faster than Algorithm 1 for small $\theta=0.2$, and Algorithm 1 converges faster than the algorithm (13) for large $\theta=0.8$. Fig. 2(c) shows that Algorithm 1 is applicable to $\theta=1$ (See the end of Sec. III).

## V. Conclusion

In this paper, we have proposed a sparsity-enhanced model with a GME penalty for GRPCA. The proposed GRPCA model uses a nonconvex GME penalty while maintaining the overall convexity. We also show that a proximal splitting type algorithm in [18] can be used to obtain a global minimizer of the proposed model. A numerical experiment in a scenario of foreground and background decomposition of a video demonstrates the efficacy of the proposed GRPCA model.

## REFERENCES

[1] I. T. Jolliffe, Principal Component Analysis. Springer, 1986.
[2] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" J. ACM, vol. 58, no. 3, Jun. 2011.
[3] S. Gandy and I. Yamada, "Convex optimization techniques for the efficient recovery of a sparsely corrupted low-rank matrix," Journal of Math-for-Industry, vol. 2, no. 5, pp. 147-156, 2010.
[4] N. Shahid, N. Perraudin, V. Kalofolias, G. Puy, and P. Vandergheynst, "Fast robust PCA on graphs," IEEE Journal of Selected Topics in Signal Processing, vol. 10, no. 4, pp. 740-756, 2016.
[5] Z. Lin, M. Chen, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," arXiv preprint arXiv:1009.5055, 2010.
[6] X. Yuan and J. Yang, "Sparse and low-rank matrix decomposition via alternating direction method," Pacific Journal of Optimization, vol. 9, no. 1, pp. 167-180, 2013.
[7] P. L. Combettes and J. C. Pesquet, "A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery," IEEE Journal of Selected Topics in Signal Processing, vol. 1, no. 4, pp. 564-574, 2007.
[8] S. Gandy, B. Recht, and I. Yamada, "Tensor completion and low-n-rank tensor recovery via convex optimization," Inverse Problems, vol. 27, no. 2, p. 025010, jan 2011.
[9] T. Mizoguchi and I. Yamada, "Hypercomplex principal component pursuit via convex optimization," in Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Nov. 2018, pp. 579-586
[10] C.-H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," The Annals of Statistics, vol. 38, no. 2, pp. $894-942$, 2010.
[11] J. Wang, M. Wang, X. Hu, and S. Yan, "Visual data denoising with a unified schatten-p norm and $\ell_{q}$ norm regularized principal component pursuit," Pattern Recognition, vol. 48, no. 10, pp. 3135-3144, 2015.
[12] S. Wang, D. Liu, and Z. Zhang, "Nonconvex relaxation approaches to robust matrix recovery." in IJCAI. Citeseer, 2013, pp. 1764-1770
[13] J. Dongarra, M. Gates, A. Haidar, J. Kurzak, P. Luszczek, S. Tomov, and I. Yamazaki, "The singular value decomposition: Anatomy of optimizing an algorithm for extreme scale," SIAM Review, vol. 60, no. 4, pp. 808865, 2018.
[14] Z. Zhang and K. Zhao, "Low-rank matrix approximation with manifold regularization," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 7, pp. 1717-1729, 2013
[15] B. K. Natarajan, "Sparse approximate solutions to linear systems," SIAM Journal on Computing, vol. 24, no. 2, pp. 227-234, 1995.
[16] T. T. Nguyen, C. Soussen, J. Idier, and E. Djermoune, "Np-hardness of $\ell_{0}$ minimization problems: revision and extension to the non-negative setting," in International conference on Sampling Theory and Applications (SampTA), 2019, pp. 1-4.
[17] I. Selesnick, "Sparse regularization via convex analysis," IEEE Transactions on Signal Processing, vol. 65, no. 17, pp. 4481-4494, 2017.
[18] J. Abe, M. Yamagishi, and I. Yamada, "Linearly involved generalized Moreau enhanced models and their proximal splitting algorithm under overall convexity condition," Inverse Problems, vol. 36, no. 3, p. 035012, Feb. 2020.
[19] K. Yosida, "On the differentiability and the representation of oneparameter semi-group of linear operators," Journal of the Mathematical Society of Japan, vol. 1, no. 1, pp. 15-21, 1948.
[20] M. Belkin and P. Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation," Neural Computation, vol. 15, no. 6, pp. 1373-1396, 062003.
[21] N. Perraudin, J. Paratte, D. Shuman, L. Martin, V. Kalofolias, P. Vandergheynst, and D. K. Hammond, "GSPBOX: A toolbox for signal processing on graphs," 2016.
[22] L. Yin, A. Parekh, and I. Selesnick, "Stable principal component pursuit via convex analysis," IEEE Transactions on Signal Processing, vol. 67, no. 10, pp. 2595-2607, 2019.
[23] J. Abe, M. Yamagishi, and I. Yamada, "Convexity-edge-preserving signal recovery with linearly involved generalized minimax concave penalty function," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2019, pp. 4918-4922.
[24] L. Li, W. Huang, I. Y.-H. Gu, and Q. Tian, "Statistical modeling of complex backgrounds for foreground object detection," IEEE Transactions on Image Processing, vol. 13, no. 11, pp. 1459-1472, 2004.


[^0]:    This work was supported in part by JSPS Grants-in-Aid (18K19804).

[^1]:    ${ }^{1}$ See [19] for the origin of Yosida regularization for maximally monotone linear operator.

