Robust Subspace Tracking with Missing Data and Outliers via ADMM

Le Trung Thanh[1], Nguyen Viet Dung[1,2], Nguyen Linh Trung[1,*], Karim Abed-Meraim[3]

[1] AVITECH Institute, University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam
[2] National Institute of Advanced Technologies of Brittany, Brest, France
[3] PRISME Laboratory, University of Orléans, Orléans, France

Abstract—**Robust subspace tracking is crucial when dealing with data in the presence of both outliers and missing observations. In this paper, we propose a new algorithm, namely PETRELS-ADMM, to improve performance of subspace tracking in such scenarios. Outliers residing in the observed data are first detected in an efficient way and removed by the alternating direction method of multipliers (ADMM) solver. The underlying subspace is then updated by the algorithm of parallel estimation and tracking by recursive least squares (PETRELS) in which each row of the subspace matrix was estimated in parallel. Based on PETRELS-ADMM, we also derive an efficient way for robust matrix completion. Performance studies show the superiority of PETRELS-ADMM as compared to the state-of-the-art algorithms. We also illustrate its effectiveness for the application of background-foreground separation.**

Index Terms—**Robust subspace tracking, robust PCA, robust matrix completion, missing data, outliers, alternating direction method of multipliers (ADMM).**

I. INTRODUCTION

Subspace estimation is a problem of finding a p-dimensional subspace $\mathbf{U}$ of $\mathbb{R}^n$, $p \ll n$, such that it represents the span of the observed signal (data) vectors, while assuming that these signals reside in a low dimensional subspace. It is generally referred to as principal component analysis (PCA), widely used for dimensionality reduction. Subspace estimation is typically obtained by batch approaches such as singular value decomposition (SVD) of the data matrix or eigenvalue decomposition (EVD) of its corresponding covariance matrix. These approaches are, however, not suitable for real-time applications because of their high computational complexity; $O(n^3)$, generally. To handle this problem, subspace tracking, also called streaming/dynamic PCA, is an excellent alternative solution and has much less complexity; see [1] for a review. However, in the presence of corruptions (e.g. noise, missing entries and outliers), the performance of these approaches may degrade.

Missing (incomplete) data are ubiquitous in many modern applications in general and subspace tracking in particular [2]. State-of-the-art algorithms for handling missing data aim to interpret RST via geometric lens (i.e., optimization), such as Grassmannian rank-one update subspace estimation (GROUSE) [3], parallel estimation and tracking by recursive least squares (PETRELS) [4] and online stochastic gradient descent (OSGD) [5]. Among state-of-the-arts, PETRELS provided competitive performance in terms of subspace estimation accuracy.

It is known that subspace tracking algorithms are sensitive to outliers (in a similar way as to PCA), thus demanding robust subspace tracking (RST), or robust streaming PCA. RST has recently brought much attention and been extensively studied

in [6]. Main approaches include: principal component pursuit (PCP) [7], alternating minimization (AltProj) [8], projected gradient decent (RPCA-GD) [9], recursive projected compressive sensing (ReProCS) [10], ℓ_p-norm robust online subspace tracking [11], [12], weighted least square-based RST (RO-BUSTA) [13] and their extensions. Among these approaches, only a few of them, for example GRASTA [11], ROSETA [12] and PETRELS-CFAR [13], are capable of dealing with RST in the presence of missing data.

In this paper, we consider the RST problem for streaming data in the presence of both outliers and missing entries. Often, we aim to reduce the effect of outliers and then apply a *robust* cost function, e.g. using GRASTA and ROSETA. However, in the presence of a large number of corrupted/missing data, the performance of these methods may not be adequate. On the other hand, we can also try to identify outliers and treat them as incomplete data first. Then a subspace tracking method for missing data (i.e., using non-robust cost function) is exploited to handle "outliers-removed data" as PETRELS-CFAR in [13]. This method can overcome the need to know the locations of corrupted entries in advance (as in MD-ISVD [14]), which is difficult to meet in practice. Moreover, beside its simplicity, we can also exploit advances in subspace tracking algorithms for missing data. The drawback, however, is that the performance of CFAR may degrade in the presence of missing data. Adopting the approach of PETRELS-CFAR but aiming to improve the tracking performance, we are interested in looking for a method that can remove outliers more correctly.

Our paper has two contributions. First, we propose an algorithm, namely PETRELS-ADMM, for RST with missing data and outliers. In particular, outliers residing in the observed data are first detected and removed by the alternating direction method of multipliers (ADMM) solver in an efficient way. The main idea is to eliminate the effect of outliers by augmenting on both sparse and weight vector instead of only weight one as in the existing methods (See section III-A for more details). The underlying subspace is then updated by PETRELS. Second, we also derive an efficient algorithm for robust matrix completion by exploiting advantage of PETRELS-ADMM. In particular, the data labelled as outliers by PETRELS-ADMM is treated as missing data. As a consequence, only "clean data" involves the completion process, thus improving overall performance.

Compared to GRASTA and ROSETA, the proposed PETRELS-ADMM algorithm has several advantages. First, our algorithm detects and removes outliers more efficiently. Second, the cost function in the subspace update step of the proposed method need not be robust. We note that, to have the "right" direction toward the true subspace, GRASTA and ROSETA require robust cost functions as well as additional adaptive parameter selection. Third, thanks to the use of

* Corresponding author: Nguyen Linh Trung, linhtrung@vnu.edu.vn.

PETRELS, our proposed algorithm has a good convergence rate and can converge to the global optimum given a full observation of the data or to a stationary point given a partial observation (See [4] and [13] for convergence analysis). In contrast, GRASTA uses the stochastic gradient descent on the Grassmannian manifold whose convergence rate is limited. In parallel, convergence of the heuristic subspace tracking algorithm in ROSETA has not been mathematically analyzed yet.

II. PROBLEM FORMULATION

At each time instance t, assume that we have a data vector $\mathbf{v}_t \in \mathbb{R}^{n \times 1}$ under the following signal model:

$$\mathbf{v}_t = \boldsymbol{\ell}_t + \mathbf{n}_t, \tag{1}$$

where $\mathbf{n}_t \in \mathbb{R}^n$ is additive white Gaussian noise, $\boldsymbol{\ell}_t \in \mathbb{R}^n$ is considered as the true signal that resides in a low dimensional subspace of $\mathbf{U}^{\text{true}} \in \mathbb{R}^{n \times p}$ ($p \ll n$), given by

$$\boldsymbol{\ell}_t = \mathbf{U}^{\text{true}} \mathbf{w}_t, \tag{2}$$

with $\mathbf{w}_t \in \mathbb{R}^p$ being a weight vector. Some entries of $\mathbf{v}_t$ may be missing and/or corrupted by outliers. So, the observed vector can be modeled as

$$\mathbf{v}_{t\Omega_t} = \mathcal{P}_{\Omega_t}(\mathbf{v}_t) + \mathbf{s}_t, \tag{3}$$

where $\mathcal{P}_{\Omega_t}$ is the projection under the observation mask Ω_t that indicates whether the k-th entry of $\mathbf{v}_t$ is observed (i.e., $\Omega_t(k) = 1$) or not (i.e., $\Omega_t(k) = 0$), $k = 1, \ldots, n$, and $\mathbf{s}_t \in \mathbb{R}^N$ is a sparse outlier vector.

RST problem for missing data and outliers: *Given a set of data vectors* $\{\mathbf{v}_{i\Omega_i}\}_{i=1}^t$ *at time instances* $1, \ldots, t$, *we wish to estimate a rank-p matrix* $\mathbf{U}_t \in \mathbb{R}^{n \times p}$ *that represents the span of the set of signal vectors* $\{\boldsymbol{\ell}_i\}_{i=1}^t$.

One type of optimization in RST is to minimize the total projection residual on the observed entries and account for outliers, as given by

$$\min \sum_{i=1}^t \frac{1}{2} \left\| \left(\mathbf{U}_{i\Omega_i} \mathbf{w}_i + \mathbf{s}_i - \mathbf{v}_{i\Omega_i} \right) \right\|_2^2 + \rho \left\| \mathbf{s}_i \right\|_0, \tag{4}$$

where the ℓ_0-norm applied to $\mathbf{s}_i$ is to control outlier density (sparsity) with the regularization weight ρ on outliers. However, the problem of (4) is NP-hard [15].

Since, the ℓ_1-norm $\|\mathbf{s}_i\|_1 = \sum_{k=1}^n |\mathbf{s}_i(k)|$ is a good convex approximation of $\|\mathbf{s}_i\|_0$ [15], we can relax (4) by

$$\min \sum_{i=1}^t \frac{1}{2} \left\| \left(\mathbf{U}_{i\Omega_i} \mathbf{w}_i + \mathbf{s}_i - \mathbf{v}_{i\Omega_i} \right) \right\|_2^2 + \rho \left\| \mathbf{s}_i \right\|_1, \tag{5}$$

which can be efficiently solved by convex optimization.

In particular, the solution of (5) can be obtained using alternating minimization, which can be decomposed into two steps. In the first step, we estimate the coefficients $\mathbf{w}_t$ and removes the outliers $\mathbf{s}_t$ by minimizing the following function:

$$f(\mathbf{U}, \mathbf{w}_i, \mathbf{s}_i) = \left\| \mathbf{U}_{\Omega_i} \mathbf{w}_i + \mathbf{s}_i - \mathbf{v}_{i\Omega_i} \right\|_2^2 + \rho \left\| \mathbf{s}_i \right\|_1, \tag{6}$$

for $i = 1, \ldots, t$. In the second step, we updates the subspace $\mathbf{U}_t$ by

$$\mathbf{U}_t = \underset{\mathbf{U}}{\arg\min}\, F_t(\mathbf{U}), \tag{7}$$

where

$$F_t(\mathbf{U}) = \frac{1}{t} \sum_{i=1}^t \lambda^{t-i} f(\mathbf{U}, \mathbf{w}_i, \mathbf{s}_i), \tag{8}$$

and λ with $0 \ll \lambda \leq 1$ is the forgetting factor aimed at discounting the effect of past observations.

Thanks to the law of large numbers, the observation mean, $F_t(\mathbf{U})$, without discounting (i.e., $\lambda = 1$), will converge to the true mean, $F(\mathbf{U})$, when t approaches infinity. Therefore, the true signal subspace can be asymptotically obtained by

$$\mathbf{U}^{\text{true}} = \underset{\mathbf{U} \in \mathbb{R}^{n \times p}}{\arg\min}\, F(\mathbf{U}) \tag{9}$$

In the next section, we will propose an efficient algorithm to minimize $f(\mathbf{U}, \mathbf{w}_i, \mathbf{s}_i)$ and $F_t(\mathbf{U})$, and then indicate that its solution, $\mathbf{U}_t$, will converge almost surely to the local optimum of $F(\mathbf{U})$.

III. PROPOSED PETRELS-ADMM ALGORITHMS

Now, we propose the PETRELS-ADMM algorithm for RST with missing data and outliers. The algorithm first applies the ADMM framework in [16], which has been widely used in previous works for solving (6) GRASTA [11] and ROSETA [12], and then uses PETRELS to tackle (7). However, the main difference in our method is that we propose to augment on both sparse and weight vectors to further reduce the effect of outliers. The first part of this section deals with RST. In the second part, we apply the proposed robust algorithm to matrix completion.

A. Robust Subspace Tracking

We show here how to solve (6) step-by-step:

1) Update $\mathbf{s}_t$ *and* $\mathbf{w}_t$: Under the assumption that the underlying subspace $\mathbf{U}_t$ changes slowly, we have the following approximation, $\mathbf{U}_t \cong \mathbf{U}_{t-1}$. Therefore, at each time instance t, the weights in $\mathbf{w}_t$ and the outliers in $\mathbf{s}_t$ can be estimated from the data vector $\mathbf{v}_{\Omega_t}$ and $\mathbf{U}_{t-1}$ by rewriting (6) as

$$f(\mathbf{w}, \mathbf{s}) = \left\| \mathbf{U}_{t-1\Omega_t} \mathbf{w} + \mathbf{s} - \mathbf{v}_{t\Omega_t} \right\|_2^2 + \rho \left\| \mathbf{s} \right\|_1. \tag{10}$$

Update $\mathbf{s}_t$: To estimate $\mathbf{s}_t$ given $\mathbf{w}$, we exploit that the fact that (10) can be cast into the ADMM form as follows:

$$\min_{\mathbf{u}, \mathbf{s}} h(\mathbf{u}) + g(\mathbf{s}), \quad \text{subject to } \mathbf{u} - \mathbf{s} = \mathbf{0}, \tag{11}$$

where $\mathbf{u}$ is an additional decision variable, $h(\mathbf{u}) = \frac{1}{2}\|\mathbf{U}_{t-1\Omega_t}\mathbf{w} + \mathbf{u} - \mathbf{v}_{t\Omega_t}\|_2^2$ and $g(\mathbf{s}) = \rho\|\mathbf{s}\|_1$. The corresponding augmented Lagrangian with the dual variable vector $\boldsymbol{\beta}$ is thus given by

$$\mathcal{L}(\mathbf{s}, \mathbf{u}, \mathbf{v}) = g(\mathbf{s}) + h(\mathbf{u}) + \boldsymbol{\beta}^T(\mathbf{u} - \mathbf{s}) + \frac{\rho_1}{2}\|\mathbf{u} - \mathbf{s}\|_2^2.$$

We emphasize that we propose to focus on augmenting $\mathbf{s}$, unlike GRASTA and ROSETA, on augmenting $\mathbf{w}$.

Let $\mathbf{r} = \boldsymbol{\beta}/\rho_1$ be a scaled version of the dual variable. We obtain the following rule for updating $\mathbf{s}_t$:

$$\mathbf{u}^{k+1} = \arg\min_{\mathbf{u}} \left\{ h(\mathbf{u}) + \frac{\rho_1}{2}\|\mathbf{u} - (\mathbf{r}^k - \mathbf{s}^k)\|_2^2 \right\}$$

$$= \frac{1}{\rho_1}(\mathbf{v}_{\Omega_t} - \mathbf{U}_{t-1\Omega_t}\mathbf{w}) - (\mathbf{r}^k - \mathbf{s}^k),$$

$$\mathbf{s}^{k+1} = \arg\min_{\mathbf{s}} \left\{ g(\mathbf{s}) + \frac{\rho_1}{2}\|\mathbf{u}^{k+1} - (\mathbf{r}^k - \mathbf{s})\|_2^2 \right\}$$

$$= S_{1/\rho_1}(\mathbf{u}^{k+1} + \mathbf{r}^k),$$

$$\mathbf{r}^{k+1} = \mathbf{r}^k + \mathbf{u}^{k+1} - \mathbf{s}^{k+1},$$

where $S_\alpha(x)$ is the soft thresholding, defined as

$$S_\alpha(x) = \begin{cases} 0, & \text{if } |x| \le \alpha, \\ x - \alpha, & \text{if } x > \alpha, \\ x + \alpha, & \text{if } x < -\alpha, \end{cases}$$

which is a proximity operator of the ℓ_1-norm [16].

Update $\mathbf{w}_t$: To estimate $\mathbf{w}_t$ given $\mathbf{s}$, we minimize the augmented Lagrangian of (10), which is

$$\mathcal{L}(\mathbf{w}, \mathbf{p}, \mathbf{q}) = \frac{1}{2}||(\mathbf{v}_{t\Omega_t} - \mathbf{p}) - \mathbf{U}_{t-1\Omega_t}\mathbf{w}||_2^2 + \frac{\rho_2}{2}||\mathbf{w} - \mathbf{q}||_2^2, \quad (12)$$

where $\mathbf{p}$ and $\mathbf{q}$ are the additional decision variable vectors for $\mathbf{s}$ and $\mathbf{w}$, respectively. However, (12) is still affected by outliers because $\mathbf{s}$ and its decision variable $\mathbf{p}$ may not be completely rejected in each iteration. Therefore, $\mathcal{L}(\mathbf{w}, \mathbf{p}, \mathbf{q})$ can be cast further into the ADMM form such that it can lie between least-squares and least-absolute deviations to reduce the effect of outliers. The Huber fitting can provide a transition between the quadratic and absolute terms of $\mathcal{L}(\mathbf{w}, \mathbf{p}, \mathbf{q})$ [16], as

$$f^{\text{Hub}}(x) = \begin{cases} x^2/2, & |x| \le 1, \\ |x| - 1/2, & |x| > 1. \end{cases}$$

It means that we apply the Huber fitting to two terms of (12). As a result, $\mathbf{q}$-updates for estimating $\mathbf{w}$ involve the proximity operator of the Huber function, as in

$$\mathbf{q}^{k+1} = \frac{\rho_2}{1+\rho}\mathbf{z}^{k+1} + \frac{1}{1+\rho_2}S_{1+\frac{1}{\rho_2}}(\mathbf{z}^{k+1}),$$

where $\mathbf{z}$ is a dummy variable with $\mathbf{z}^{k+1} = \mathbf{U}_{t-1\Omega_t}\mathbf{w}^{k+1} + \mathbf{p}^k - \mathbf{v}_{\Omega_t}$ at the k-th iteration. Hence, at the $(k+1)$-th iteration, $\mathbf{w}^{k+1}$ can be updated using the following closed-form solution of the convex quadratic form:

$$\mathbf{w}^{k+1} = (\mathbf{U}_{t-1\Omega_t}^T\mathbf{U}_{t-1\Omega_t} + \rho_2\mathbf{I})^{-1}\mathbf{U}_{t-1\Omega_t}^T(\mathbf{v}_{t\Omega_t} - \mathbf{p}^k + \mathbf{q}^k),$$

where parameter $\rho_2 > 0$ is to ensure that the matrix $\mathbf{U}_{t-1\Omega_t}^T\mathbf{U}_{t-1\Omega_t} + \rho_2\mathbf{I}$ is invertible.

To sum up, the rule for updating $\mathbf{w}_t$ is be given by

$$\mathbf{w}^{k+1} = (\mathbf{U}_{t-1\Omega_t}^T\mathbf{U}_{t-1\Omega_t} + \rho_2\mathbf{I})^{-1}\mathbf{U}_{t-1\Omega_t}^T(\mathbf{v}_{t\Omega_t} - \mathbf{p}^k + \mathbf{q}^k),$$
$$\mathbf{z}^{k+1} = \mathbf{U}_{t-1\Omega_t}\mathbf{w}^{k+1} + \mathbf{p}^k - \mathbf{v}_{t\Omega_t},$$
$$\mathbf{q}^{k+1} = \frac{\rho_2}{1+\rho}\mathbf{z}^{k+1} + \frac{1}{1+\rho_2}S_{1+\frac{1}{\rho_2}}(\mathbf{z}^{k+1}),$$
$$\mathbf{p}^{k+1} = \mathbf{p}^k + (\mathbf{U}_{t-1\Omega_t}\mathbf{w}^{k+1} - \mathbf{q}^{k+1} - \mathbf{v}_{t\Omega_t}).$$

We note that, by using the Huber fitting operator, our algorithm is better than GRASTA and ROSETA, which use ℓ_2 regularization, in reducing the effect of outliers.

2) Update $\mathbf{U}_t$: Having estimated $\mathbf{s}_t$, we can rewrite (7) as

$$\mathbf{U}_t = \arg\min_{\mathbf{U}\in\mathbb{R}^{n\times p}}\sum_{i=1}^t \lambda^{t-i}||\mathbf{v}_{i\Omega_i}^{\text{re}} - \mathbf{U}_{\Omega_i}\mathbf{w}_i||_2^2, \quad (13)$$

where the recovered signal $\mathbf{v}_{\Omega_i}^{\text{re}}$ is determined by

$$\mathbf{v}_{i\Omega_i}^{\text{re}}(k) = \begin{cases} \frac{||\mathbf{s}_i||_0}{n}\mathbf{v}_{i\Omega_i}(k), & \text{if } \mathbf{s}_i(k) = 0, \\ 0, & \text{otherwise,} \end{cases}$$

and the ℓ_1-norm term of outliers $\mathbf{s}_t$ can be eliminated. The problem of (13) can be solved by using PETRELS [4], which can be decomposed into subproblems for each row of $\mathbf{U}$. Note that subspace tracking in this way is efficient since we can

ignore the m-th row if the m-th entry of $\mathbf{v}_{t\Omega_t}^{\text{re}}$ is labeled as corrupted. More details can be found in [4].

The following theorem, whose proof is omitted here due to the space limitation but can be found in our technical report [17], indicates the convergence of PETRELS-ADMM.

Theorem 1 (Convergence of PETRELS-ADMM): Let $\{\mathbf{U}_t\}_{t=1}^\infty$ be a sequence of solutions generated by PETRELS-ADMM, then the sequence converges to a stationary point of the expected loss function $F(\mathbf{U})$ when $t \to \infty$.

B. Robust Matrix Completion

Motivated by the advantages of the proposed PETRELS-ADMM algorithm, we apply it to the problem of robust matrix completion (RMC), that is, to recover corrupted entries affected by missing data and outliers. The main idea is to treat outliers as missing data and only "clean" data is used to compute the weight vector. Particularly, $\mathbf{v}$ can be divided into two components: "clean"entries (without outliers and missing) and corrupted entries, denoted by $\mathbf{v}_{\text{clean}}$ and $\mathbf{v}_{\text{cor}}$ respectively. These components can be obtained by from the projection $\mathcal{P}$ under a mask Ω_{clean} and under the mask for the remaining (corrupted) entries Ω_{cor} respectively as

$$\mathbf{v}_{\text{clean}} = \mathcal{P}_{\Omega_{\text{clean}}}(\mathbf{v}_t) \text{ and } \mathbf{v}_{\text{cor}} = \mathcal{P}_{\Omega_{\text{cor}}}(\mathbf{v}_t).$$

Then, the problem of matrix completion is formulated as

$$(\mathbf{w}^*, \mathbf{v}_{\text{cor}}^{\text{re}}) = \arg\min_{\mathbf{w}, \mathbf{v}_{\text{cor}}}||(\mathbf{U}_{\Omega_{\text{clean}}}\mathbf{w} - \mathbf{v}_{\text{clean}}) + (\mathbf{U}_{\Omega\text{cor}}\mathbf{w} - \mathbf{v}_{\text{cor}})||_2^2.$$

Since PETRELS-ADMM is effective, as latter shown in the experiments, in correctly locating the missing data and outliers (i.e., $\mathbf{v}_{\text{cor}}$) we can reduce their effects by setting them to zero. As a result, the matrix completion problem can be reformulated as

$$\mathbf{w}^* = \arg\min_{\mathbf{w}}||\mathbf{U}_{\Omega_{\text{clean}}}\mathbf{w} - \mathbf{v}_{\text{clean}}||_2^2,$$
$$\mathbf{v}_{\text{cor}}^{\text{re}} = \arg\min_{\mathbf{v}_{\text{cor}}}||\mathbf{U}_{\Omega_{\text{cor}}}\mathbf{w}^* - \mathbf{v}_{\text{cor}}||_2^2.$$

Thus, the closed-form solutions are given by

$$\mathbf{w}^* = (\mathbf{U}_{\Omega_{\text{clean}}}^T\mathbf{U}_{\Omega_{\text{clean}}})^{-1}\mathbf{U}_{\Omega_{\text{clean}}}^T\mathbf{v}_{\text{clean}}, \quad (14)$$
$$\mathbf{v}_{\text{cor}}^{\text{re}} = \mathbf{U}_{\Omega_{\text{cor}}}\mathbf{w}^*. \quad (15)$$

IV. EXPERIMENTS

In this section, we assess performance of the proposed PETRELS-ADMM algorithm by comparing to state-of-the-arts in three scenarios: robust subspace tracking, robust matrix completion and video background-foreground separation[1].

A. Robust Subspace Tracking

State-of-the-art algorithms for comparison are: GRASTA [11], ROSETA [12] and PETRELS-CFAR [13]. To have a fair comparison, the parameters of these algorithms are set default. In the following experiments, the data vectors $\{\mathbf{v}_t\}_{t\ge1}$ were randomly generated using the standard signal model

$$\mathbf{v}_t = \mathbf{A}\mathbf{x}_t + \mathbf{n}_t, \quad \mathbf{v}_{\Omega_t} = \mathcal{P}_{\Omega_t}(\mathbf{v}_t) + \mathbf{s}_t.$$

where $\mathbf{A}$ denotes a mixing matrix $\mathbb{R}^{n\times p}$, $\mathbf{x}_t$ is a random vector living on the $\mathbb{R}^p$ space and both of them are i.i.d. Gaussian of $\mathcal{N}(0, 1)$, $\mathbf{n}_t$ is white Gaussian noise of $\mathcal{N}(0, \sigma^2)$,

---

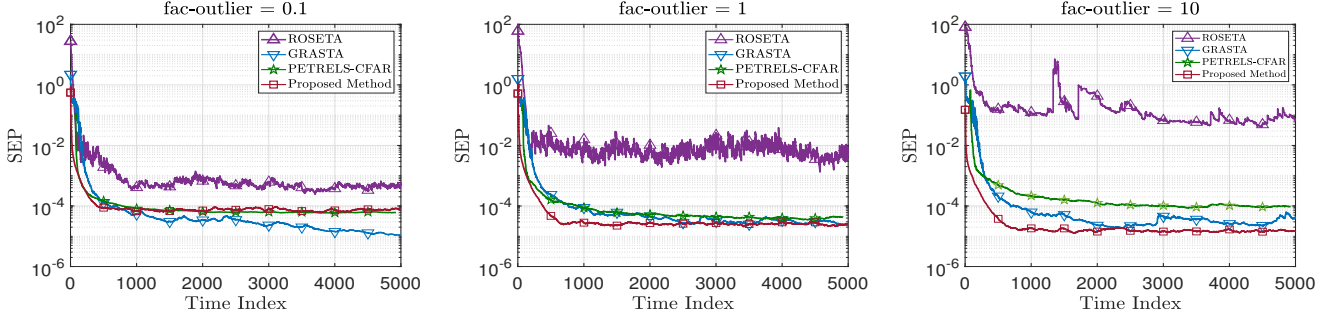[1]MATLAB codes are available at https://github.com/thanhtbt/RST.

Fig. 1. Impact of outlier intensity on algorithm performance: $n = 100$, $p = 2$, 90% entries observed, outlier density of 5% and SNR = 20 dB.
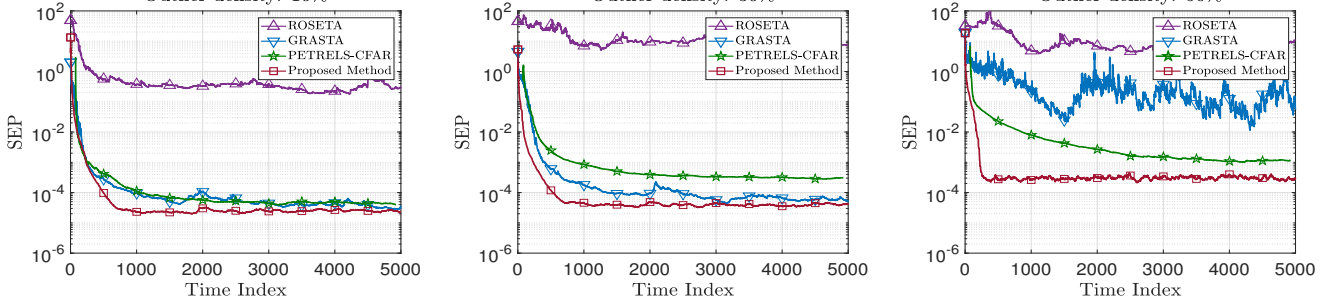

Fig. 2. Impact of outlier density on algorithm performance: $n = 100$, $p = 2$, 70% entries observed, outlier intensity fac-outlier = 5 and SNR = 20 dB.

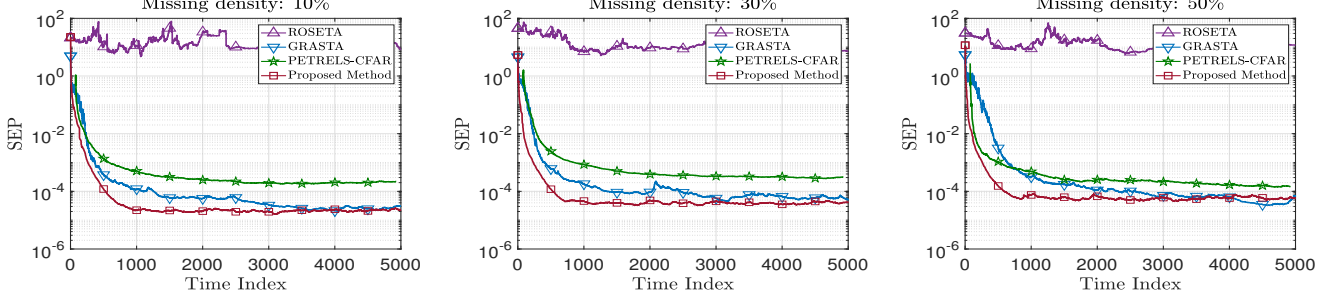Fig. 3. Impact of missing density on algorithm performance: $n = 100, p = 2$, outlier density of 30%, outlier intensity fac-outlier = 5 and SNR = 20 dB.

with SNR $= -10 \log_{10}(\sigma^2)$ is the signal-to-noise ratio to control the effect of noise to the algorithms. performance, $\mathcal{P}_{\Omega_t}$ is the projection under the observation mask Ω_t with a given percentage of missing data $k\%$, $\mathbf{s}_t$ is i.i.d. uniform over $[0, \text{fac-outlier}]$, where fac-outlier determines the maximum magnitude of outliers. We use random initialization in all experiments.

The subspace estimation performance (SEP) metric [13], defined below, is used to assess the subspace estimation accuracy:

$$\text{SEP} = \frac{\text{tr}\{\mathbf{U}_{es}^T(\mathbf{I} - \mathbf{U}_{ex}\mathbf{U}_{ex}^T)\mathbf{U}_{es}\}}{\text{tr}\{\mathbf{U}_{es}^T(\mathbf{U}_{ex}\mathbf{U}_{ex}^T)\mathbf{U}_{es}\}},$$

where $\mathbf{U}_{ex}$ and $\mathbf{U}_{es}$ are the true and the estimated subspaces correspondingly. The lower SEP is, the better the performance of algorithm.

Fig. 1 indicates the effect of outlier intensity on algorithm performance. As we can see, at low intensity, all algorithms yielded good accuracy with fast convergence, though ROSETA provided higher SEP as compared to that by the three remaining algorithms. Meanwhile, at high intensity (e.g. fac-outlier = 0.1, 1 or 10), PETRELS-ADMM provided the best performance in terms of both convergence speed and accuracy.

Fig. 2 shows the performance in terms of outlier density. We can see that PETRELS-ADMM outperformed GRASTA, ROSETA and PETRELS-CFAR. In the presence of a high

percentage of outliers, e.g. 50% as in Fig. 2, PETRELS-ADMM yielded reasonable performance in terms of accuracy, SEP $\approx 10e\text{-}4$, while the other algorithms failed. When the measurement data were corrupted by a smaller number of outliers, PETRELS-ADMM still provided better performance than the others, as shown in Fig. 2.

The effect of missing data density is presented in the Fig. 3. Similarly, PETRELS-ADMM yielded good performance in three cases of missing data: 10%, 30% and 50%. PETRELS-CFAR provided similar performance but with slower convergence, while ROSETA and GRASTA were only good for the cases of low percentage of missing data (e.g. $\leq 50\%$).

B. Robust Matrix Completion

We compare the performance RMC using PETRELS-ADMM, GRASTA [11] and RPCA-GD [9]. The measurement data $\mathbf{X} = \mathbf{AS}$ used for this task were the rank-2 matrices with size of 400×400. We generated the mixing matrix $\mathbf{A} \in \mathbb{R}^{400 \times 2}$ and the signal matrix $\mathbf{S} \in \mathbb{R}^{2 \times 400}$ at random. The entries were i.i.d. Gaussian of $\mathcal{N}(0, 1)$. The measurement data $\mathbf{X}$ was added with white Gaussian noise $\mathbf{N} \in \mathbb{R}^{400 \times 400}$ whose SNR is set to 40 dB. The measurement data matrices were corrupted by different percentages of missing and outliers from $0\% - 90\%$. The location and value of corrupted entries (including missing and outliers) were uniformly distributed.

Fig. 4 shows that the proposed algorithm of PETRELS-ADMM-based RMC outperformed GRASTA-based and
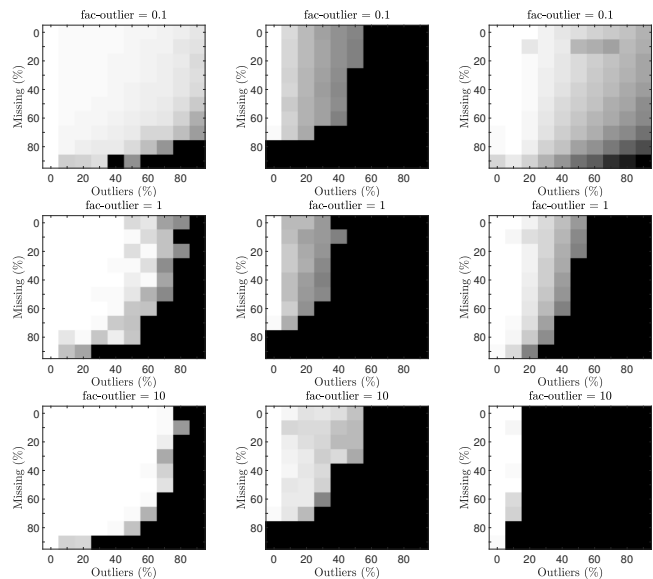
Fig. 4. Effect of outlier intensity on robust matrix completion performance. White colour denotes perfect recovery, black colour denotes failure and gray colour is in between. From left to right column: PETRELS-ADMM, GRASTA, RCPA-GD.
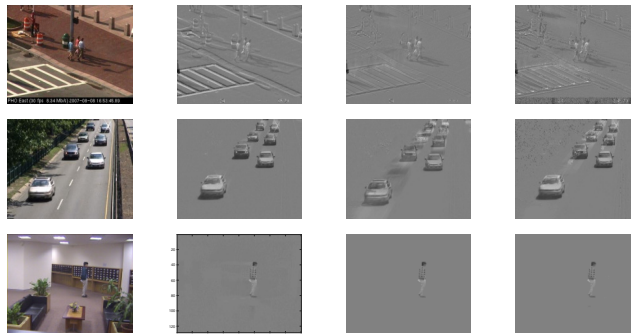

Fig. 5. Video Background-Foreground Separation. From left to right column: original data, PETRELS-ADMM, GRASTA, PETRELS-CFAR.

RPCA-GD-based algorithms. At low outlier intensity (i.e., fac-outlier = 0.1), PETRELS-ADMM-based RMC and RCPA-GD-based RMC provided excellent performance even when the data were corrupted by a very high fraction of outliers and the missing data were recovered perfectly. At high outlier intensity (i.e., fac-outlier $\geq$ 1), PETRELS-ADMM-based RMC provided the best performance in terms of matrix reconstruction error, GRASTA-based RMC still retained good performance, while RPCA-GD-based RMC failed to recover corrupted entries.

C. Video Background/Foreground Separation

We further illustrate the effectiveness of the proposed PETRELS-ADMM algorithm in the application of RST for video background/foreground separation, and compare it with GRASTA and PETRELS-CFAR. Datasets "Highway" including 1700 frames of size 240×320 pixels and "Sidewalk" including 1200 frames of size 240×352 pixels were obtained from CD.net2012[2]. The "Lobby" has 1546 frames of size 144×176 pixels from GRASTA. We can see from Fig. 5 that PETRELS-ADMM was capable of detecting objects in video and provided competitive performance to GRASTA and PETRELS-CFAR.

V. Conclusions

In this work, we have studied the problem of robust subspace tracking to deal with corrupted data in the presence of both outliers and missing observations. A new efficient algorithm, namely PETRELS-ADMM, was proposed for robust subspace tracking and for robust matrix completion. Experiments were conducted to illustrate the effectiveness of the proposed algorithms in terms of both quantity and quality.

VI. Acknowledgment

[2]CD.net2012: http://jacarini.dinf.usherbrooke.ca/dataset2012.

References

[1] J. P. Delmas, "Subspace tracking for signal processing," *Adaptive Signal Processing: Next Generation Solutions*, pp. 211–270, 2010.

[2] L. Balzano, Y. Chi, and Y. M. Lu, "Streaming pca and subspace tracking: The missing data case," *Proceedings of the IEEE*, vol. 106, no. 8, pp. 1293–1310, 2018.

[3] L. Balzano, R. Nowak, and B. Recht, "Online identification and tracking of subspaces from highly incomplete information," in *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on.* IEEE, 2010, pp. 704–711.

[4] Y. Chi, Y. C. Eldar, and R. Calderbank, "Petrels: Parallel subspace estimation and tracking by recursive least squares from partial observations," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5947–5959, 2013.

[5] M. Mardani, G. Mateos, and G. B. Giannakis, "Subspace Learning and Imputation for Streaming Big Data Matrices and Tensors," *IEEE Transactions on Signal Processing*, vol. 63, no. 10, pp. 2663–2677, 2015.

[6] N. Vaswani, T. Bouwmans, S. Javed, and P. Narayanamurthy, "Robust Subspace Learning: Robust PCA, Robust Subspace Tracking, and Robust Subspace Recovery," *IEEE Signal Processing Magazine*, vol. 35, no. 4, pp. 32–55, 2018.

[7] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM (JACM)*, vol. 58, no. 3, p. 11, 2011.

[8] P. Netrapalli, U. Niranjan, S. Sanghavi, A. Anandkumar, and P. Jain, "Non-convex robust PCA," in *Advances in Neural Information Processing Systems*, 2014, pp. 1107–1115.

[9] X. Yi, D. Park, Y. Chen, and C. Caramanis, "Fast algorithms for robust PCA via gradient descent," in *Advances in Neural Information Processing Systems*, 2016, pp. 4152–4160.

[10] C. Qiu, N. Vaswani, B. Lois, and L. Hogben, "Recursive Robust PCA or Recursive Sparse Recovery in Large but Structured Noise," *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 5007–5039, 2014.

[11] J. He, L. Balzano, and A. Szlam, "Incremental gradient on the grassmannian for online foreground and background separation in subsampled video," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on.* IEEE, 2012, pp. 1568–1575.

[12] H. Mansour and X. Jiang, "A robust online subspace estimation and tracking algorithm," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on.* IEEE, 2015, pp. 4065–4069.

[13] N. Linh-Trung, V. Nguyen, M. Thameri, T. Minh-Chinh, and K. Abed-Meraim, "Low-complexity adaptive algorithms for robust subspace tracking," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 6, pp. 1197–1212, 2018.

[14] M. Brand, "Incremental singular value decomposition of uncertain data with missing values," in *European Conference on Computer Vision.* Springer, 2002, pp. 707–720.

[15] J. A. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1030–1051, 2006.

[16] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[17] T. T. Le, V.-D. Nguyen, N. Linh-Trung, and K. Abed-Meraim, "Robust subspace tracking with missing data and outliers: Novel algorithm and performance guarantee," *VNU University of Engineering and Technology, Vietnam, Tech. Rep. UET-AVITECH-2019003, May 2019.*