A Dictionary-Based Generalization of Robust PCA Part I: Study of Theoretical Properties

Sirisha Rambhatla, *Student Member, IEEE,* Xingguo Li, *Student Member, IEEE,* Jineng Ren, *Student Member, IEEE,* and Jarvis Haupt, *Senior Member, IEEE*

***Abstract*—We consider the decomposition of a data matrix assumed to be a superposition of a low-rank matrix and a component which is sparse in a known dictionary, using a convex demixing method. We consider two sparsity structures for the sparse factor of the dictionary sparse component, namely entry-wise and column-wise sparsity, and provide a unified analysis, encompassing both undercomplete and the overcomplete dictionary cases, to show that the constituent matrices can be successfully recovered under some relatively mild conditions on incoherence, sparsity, and rank. We corroborate our theoretical results by presenting empirical evaluations in terms of phase transitions in rank and sparsity, in comparison to related techniques. Investigation of a specific application in hyperspectral imaging is included in an accompanying paper.**

***Index Terms*—Low-rank, known dictionary, target localization, Robust PCA.**

I. Introduction

Leveraging structure of a given dataset is at the heart of all machine learning and data analysis tasks. *A priori* knowledge about the structure often makes the problem well-posed, leading to improvements in the solutions. Perhaps the most common of these, one that is often encountered in practice, is approximate low-rankness of the dataset, which is exploited by the popular principal component analysis (PCA) [1]. The low-rank structure encapsulates the model assumption that the data in fact spans a lower dimensional subspace than the ambient dimension of the data. However, in a number of applications, the data may not be inherently low-rank, but may be decomposed as a superposition of a low-rank component, and a component which has a sparse representation in a known *dictionary*. This scenario is particularly interesting in target identification applications [2], [3], where the *a priori* knowledge of the target *signatures* (dictionary), can be leveraged for localization.

In this work, we analyze a matrix demixing problem where a data matrix $\mathbf{M} \in \mathbb{R}^{n\times m}$ is formed via a superposition of a low-rank component $\mathbf{L} \in \mathbb{R}^{n\times m}$ of rank-r for $r < \min(n, m)$, and a dictionary sparse part $\mathbf{DS} \in \mathbb{R}^{n\times m}$. Here, the matrix $\mathbf{D} \in \mathbb{R}^{n\times d}$ is an *a priori* known dictionary, and $\mathbf{S} \in \mathbb{R}^{d\times m}$ is an unknown *sparse* coefficient matrix. Specifically, we will study the following model for $\mathbf{M}$:

$$\mathbf{M} = \mathbf{L} + \mathbf{DS}, \tag{1}$$

This work was supported by the DARPA YFA, Grant N66001-14-1-4047. Preliminary versions appeared in the proceedings of the 2016 IEEE Global Conference on Signal & Information Processing (GlobalSIP), 2017 Asilomar Conference on Signals, Systems, & Computers, and the 2018 IEEE International Conference on Acoustics, Speech & Signal Processing (ICASSP).

S. Rambhatla, J. Ren, and J. Haupt are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, 55455, USA e-mail: {rambh002, renxx282, jdhaupt}@umn.edu, respectively. X. Li is with the Computer Science Department, Princeton University, Princeton, NJ 08540, USA email: xingguol@cs.princeton.edu.

and identify the conditions under which the components $\mathbf{L}$ and $\mathbf{S}$ can be successfully recovered given $\mathbf{M}$ and $\mathbf{D}$ by solving appropriate convex formulations.

We consider the demixing problem described above for two different sparsity models on the matrix $\mathbf{S}$. First, we consider a case where $\mathbf{S}$ has at most s_e total non-zero entries (entry-wise sparse case), and second where $\mathbf{S}$ has s_c non-zero columns (column-wise sparse case). To this end, we develop the conditions under which solving

$$\min_{\mathbf{L},\mathbf{S}} \ \|\mathbf{L}\|_* + \lambda_e\|\mathbf{S}\|_1 \ \ \text{s.t.} \ \ \mathbf{M} = \mathbf{L} + \mathbf{DS}, \qquad \text{(D-RPCA(E))}$$

for the entry-wise sparsity case, and

$$\min_{\mathbf{L},\mathbf{S}} \ \|\mathbf{L}\|_* + \lambda_c\|\mathbf{S}\|_{1,2} \ \text{s.t.} \ \mathbf{M} = \mathbf{L} + \mathbf{DS}, \qquad \text{(D-RPCA(C))}$$

for the column-wise sparse case, will recover $\mathbf{L}$ and $\mathbf{S}$ for regularization parameters $\lambda_e \geq 0$ and $\lambda_c \geq 0$, respectively, given the data $\mathbf{M}$ and the dictionary $\mathbf{D}$. Here, the known dictionary $\mathbf{D}$ can be overcomplete (*fat*, i.e., $d > n$) or undercomplete (*thin*, i.e., $d \leq n$).

Here, "D-RPCA" refers to "Dictionary based Robust Principal Component Analysis", while the qualifiers "E" and "C" indicate the entry-wise and column-wise sparsity patterns, respectively. In addition, $\|.\|_*$, $\|.\|_1$, and $\|.\|_{1,2}$ refer to the nuclear norm, ℓ_1- norm of the vectorized matrix, and $\ell_{1,2}$ norm (sum of the ℓ_2 norm of the columns), respectively, which serve as convex relaxations of rank, sparsity, and column-wise sparsity inducing optimization, respectively.

These two types of sparsity patterns capture different structural properties of the dictionary sparse component. The entry-wise sparsity model allows individual data points to span low-dimensional subspaces, still allowing the dataset to span the entire space. In case of the column-wise sparse coefficient matrix $\mathbf{S}$, the component $\mathbf{DS}$ is also column-wise sparse. Therefore, this model effectively captures the structured (which depend upon the dictionary $\mathbf{D}$) corruptions in the otherwise low-rank structured columns of data matrix $\mathbf{M}$. Note that the non-zero columns of $\mathbf{S}$ are not restricted to be sparse in the column-wise sparsity model.

A. Background

A wide range of problems can be expressed in the form described in (1). Perhaps the most celebrated of these is principal component analysis (PCA) [1], which can be viewed as a special case of (1), with the matrix $\mathbf{D}$ set to zero. Next, in the absence of the component $\mathbf{L}$, the problem reduces to that of sparse recovery [4]–[6]; See [7] and references therein for an overview of related works. Further, the popular framework of Robust PCA tackles a case when the dictionary $\mathbf{D}$ is an identity matrix [8], [9]; variants include [10]–[13].

The model described in (1) is also closely related to the one considered in [14], which explores the overcomplete dictionary setting with applications to detection of network traffic anomalies. However, the analysis therein applies to a case where the dictionary $\mathbf{D}$ is overcomplete with orthogonal rows, and the coefficient matrix $\mathbf{S}$ has a small number of non-zero elements per row and column, which may be restrictive assumptions in some applications.

In particular, for the entry-wise case, the model shown in (1) is propitious in a number of applications. For example, it can be used for target identification in hyperspectral imaging [2], [3], and in topic modeling applications to identify documents with certain properties, on similar lines as [15]. We analyze and demonstrate the application of this model for a hyperspectral demixing task in an application extension of this work [16]. Further, in source separation tasks, a variant of this model was used in singing voice separation in [17], [18]. In addition, we can also envision source separation tasks where $\mathbf{L}$ is not low-rank, but can in turn be modeled as being sparse in a known [19] or unknown [20] dictionary.

For the column-wise setting, model (1) is also closely related to outlier identification [21]–[24], which is motivated by a number of contemporary "big data" applications. Here, the sparse matrix $\mathbf{S}$, also called outliers in this regime, may be of interest and can be used in identifying malicious responses in collaborative filtering applications [25], finding anomalous patterns in network traffic [26] or estimating visually salient regions of images [27]–[29].

B. Our Contributions

As described above, we propose and analyze a dictionary based generalization of robust PCA as shown in (1). Here, we consider two distinct sparsity patterns of $\mathbf{S}$, i.e., entry-wise and column-wise sparse $\mathbf{S}$, arising from different structural assumptions on the dictionary sparse component. Our specific contributions for each sparsity pattern are summarized below.

Entry-wise case: We make the following contributions towards guaranteeing the recovery of $\mathbf{L}$ and $\mathbf{S}$ via the convex optimization problem in D-RPCA(E). First, we analyze the *thin* case (i.e. $d \leq n$), where we assume that the matrix $\mathbf{S}$ has at most $s_e = \mathcal{O}(\frac{m}{r})$ non-zero elements *globally*, i.e., $\|\mathbf{S}\|_0 \leq s_e$ Next, for the *fat* case, we first extend the analysis presented in [14] to eliminate the orthogonality constraint on the rows of the dictionary $\mathbf{D}$. Further, we relax the sparsity constraints required by [14] on rows and columns of the sparse coefficient matrix $\mathbf{S}$, to study the case when $\|\mathbf{S}\|_0 \leq s_e$ with at most $k = \mathcal{O}(d/\log(n))$ non-zero elements per column [30]. Hence, we provide a unified analysis for both the *thin* and the *fat* case, making the model (1) amenable to a wide range of applications.

Column-wise case: We propose and analyze a dictionary based generalization of robust PCA, specifically *Outlier Pursuit* (OP) [21], wherein the coefficient matrix $\mathbf{S}$ admits a column sparse structure which can be viewed as "outliers"; see also [3].

Note that, in this case there is an inherent ambiguity regarding the recovery of the true component pair $(\mathbf{L}, \mathbf{S})$ corresponding to the low-rank part and the dictionary sparse component, respectively. Specifically, any pair $(\mathbf{L}_0, \mathbf{S}_0)$ satisfying $\mathbf{M} = \mathbf{L}_0 + \mathbf{D}\mathbf{S}_0 = \mathbf{L} + \mathbf{D}\mathbf{S}$, where $\mathbf{L}_0$ and $\mathbf{L}$ have the same column space, and $\mathbf{S}_0$ and $\mathbf{S}$ have the identical column support, is a solution of D-RPCA(C). To this end, we develop the sufficient conditions under which solving the convex optimization in D-RPCA(C) recovers the column space of the low-rank component $\mathbf{L}$, while identifying the outlier columns of $\mathbf{S}$. Here, the difference between D-RPCA(C) and OP being the inclusion of the known dictionary [21].

Next, we demonstrate how the *a priori* knowledge of the dictionary $\mathbf{D}$ helps us identify the corrupted columns via phase transitions in rank and sparsity for recovery of the outlier columns. Specifically, we show that in comparison to OP, D-RPCA(C) works for potentially higher ranks of $\mathbf{L}$, when s_c is a fixed proportion of m.

The *thin* dictionary case – an interesting result: As suggested by [14], when the dictionary is *thin*, i.e., $d < n$, one can envision a pseudo-inversed based technique, wherein we pre-multiply both sides in (1) with the Moore-Penrose pseudo-inverse $\mathbf{D}^\dagger \in \mathbb{R}^{d\times n}$, i.e., $\mathbf{D}^\dagger\mathbf{D} = \mathbf{I}$ (this is not applicable for the *fat* case due to the non-trivial null space of the pseudo-inverse). This operation leads to a formulation which resembles the robust PCA (RPCA) [8], [9] model for the entry-wise case and Outlier Pursuit (OP) [21] for the column-wise case, i.e.,

$$\mathbf{D}^\dagger\mathbf{M} = \mathbf{D}^\dagger\mathbf{L} + \mathbf{S}, \ (\text{RPCA}^\dagger) \qquad \mathbf{D}^\dagger\mathbf{M} = \mathbf{D}^\dagger\mathbf{L} + \mathbf{S}. \ (\text{OP}^\dagger)$$

An interesting finding of our work is that although this transformation algebraically reduces the entry-wise and column-wise sparsity cases to Robust PCA and OP settings, respectively, the specific model assumptions of Robust PCA and OP may not hold for all choices of dictionary size d and rank r. Specifically, we find that in cases where $d < r$, this pre-multiplication may not lead to a "low-rank" $\mathbf{D}^\dagger\mathbf{L}$. This suggests that the notion of "low" or "high" rank is relative to the maximum possible rank of $\mathbf{D}^\dagger\mathbf{L}$, which in this case is $\min(d, r)$. Therefore, if $d < r$, $\mathbf{D}^\dagger\mathbf{L}$ can be full-rank, and the low-rank assumptions of RPCA and OP may no longer hold. As a result, these two models (the pseudo inversed case and the current work) cannot be used interchangeably for the thin dictionary case. We corroborate these via experimental evaluations presented in Section V[1]

The rest of the paper is organized as follows. We formalize the problem, introduce the notation, and describe various considerations on the structure of the component matrices in Section II. In Section III, we present our main theorems for the entry-wise and column-wise cases along with discussion on the implication of the results, followed by an outline of the analysis in Section IV. Numerical evaluations are provided in Section V. Finally, we summarize our contributions and conclude this discussion in Section VI with insights on future work.

Notation: Given a matrix $\mathbf{X}$, we use $\|\mathbf{X}\| := \sigma_{\max}(\mathbf{X})$ for the spectral norm, where $\sigma_{\max}(\mathbf{X})$ denotes the maximum singular value of the matrix, $\|\mathbf{X}\|_\infty := \max\limits_{i,\ j}|\mathbf{X}_{ij}|$, $\|\mathbf{X}\|_{\infty,\infty} := \max\limits_{i}\|\mathbf{e}_i^\top\mathbf{X}\|_1$, and $\|\mathbf{X}\|_{\infty,2} := \max\limits_{i}\|\mathbf{X}\mathbf{e}_i\|$. Here, $\mathbf{X}_{i,j}$ denotes the (i, j) element of $\mathbf{X}$ and $\mathbf{e}_i$ denotes the cannonical basis vector with 1 at the i-th location.

[1]The code is made available at `github.com/srambhatla/Dictionary-based-Robust-PCA`, and the results are therefore reproducible.

II. Preliminaries

We start formalizing the problem set-up and introduce model parameters pertinent to our analysis. We begin our discussion with our notion of optimality for the two sparsity modalities; we also summarize the notation in Table I in the appendix.

A. Optimality of the Solution Pair

For the entry-wise case, we recover the low-rank component $\mathbf{L}$, and the sparse coefficient matrix $\mathbf{S}$, given the dictionary $\mathbf{D}$, and data $\mathbf{M}$ generated according to the model described in (1). Recall that s_e is the global sparsity, k denotes the number of non-zero entries in a column of $\mathbf{S}$ when the dictionary is *fat*.

In the the column-wise sparsity setting, due to the inherent ambiguity in the model (1), as discussed in Section I-B, we can only hope to recover the column-space for the low-rank matrix and the identities of the non-zero columns for the sparse matrix. Therefore, in this case any solution in the *Oracle Model* (defined below) is deemed to be optimal.

Definition D.1 (Oracle Model for Column-wise Sparsity Case)**.** Let the pair $(\mathbf{L}, \mathbf{S})$ be the matrices forming the data $\mathbf{M}$ as per (1), define the oracle model $\{\mathbf{M}, \mathcal{U}, \mathcal{I}_{\mathcal{S}_c}\}$. Then, any pair $(\mathbf{L}_0, \mathbf{S}_0)$ is in the *Oracle Model* $\{\mathbf{M}, \mathcal{U}, \mathcal{I}_{\mathcal{S}_c}\}$, if $\mathcal{P}_{\mathcal{U}}(\mathbf{L}_0) = \mathbf{L}$, $\mathcal{P}_{\mathcal{S}_c}(\mathbf{D}\mathbf{S}_0) = \mathbf{D}\mathbf{S}$ and $\mathbf{L}_0 + \mathbf{D}\mathbf{S}_0 = \mathbf{L} + \mathbf{D}\mathbf{S} = \mathbf{M}$ hold simultaneously, where $\mathcal{P}_{\mathcal{U}}$ and $\mathcal{P}_{\mathcal{S}_c}$ are projections onto the column space $\mathcal{U}$ of $\mathbf{L}$ and column support $\mathcal{I}_{\mathcal{S}_c}$ of $\mathbf{S}$, respectively.

B. Conditions on the Dictionary

We require that the dictionary $\mathbf{D}$ follows the *generalized frame property* (GFP) as defined as follows.

Definition D.2. A matrix $\mathbf{D}$ satisfies the *generalized frame property* (GFP), on vectors $\mathbf{v} \in \mathcal{R}$, if for any fixed vector $\mathbf{v} \in \mathcal{R}$ where $\mathbf{v} \neq \mathbf{0}$, we have

$$\alpha_\ell \|\mathbf{v}\|_2^2 \leq \|\mathbf{D}\mathbf{v}\|_2^2 \leq \alpha_u \|\mathbf{v}\|_2^2,$$

where α_ℓ and α_u are the lower and upper *generalized frame bounds* with $0 < \alpha_\ell \leq \alpha_u < \infty$.

The GFP shown above is met as long as the vectors $\mathbf{v}$ are not in the null-space of the matrix $\mathbf{D}$, and $\mathbf{D}$ has a finite $\|\mathbf{D}\|$. Therefore, for the *thin* dictionary setting $d \leq n$ for both entry-wise and column-wise cases $\mathcal{R}$ can be the entire space, and GFP is satisfied as long as $\mathbf{D}$ has full column rank. For example, $\mathbf{D}$ being a *frame* [31] suffices; see [32] for a brief overview of frames.

On the other hand, for the *fat* dictionary setting, we need the space $\mathcal{R}$ to be structured such that the GFP is met for both the entry-wise and column-wise sparsity cases. Specifically, for the entry-wise sparsity case, we also require that the frame bounds α_u and α_ℓ be close to each other. To this end, we assume that $\mathbf{D}$ satisfies the *restricted isomtery property* (RIP) of order $k = \mathcal{O}(d/\log(n))$ with a *restricted isometric constant* (RIC) of δ in this case, and that $\alpha_u = (1+\delta)$ and $\alpha_\ell = (1-\delta)$.

C. Relevant Subspaces

We now define the subspaces relevant for our discussion. For the following discussion, let the pair $(\mathbf{L_0}, \mathbf{S_0})$ denote the solution to D-RPCA(E) in the entry-wise sparse case. Further, for the column-wise sparse setting, let $(\mathbf{L_0}, \mathbf{S_0})$ denote a solution pair in the oracle model $\{\mathbf{M}, \mathcal{U}, \mathcal{I}_{\mathcal{S}_c}\}$ as defined in **D.1**, obtained by solving D-RPCA(C).

For the low-rank matrix $\mathbf{L}$, let the compact singular value decomposition (SVD) be defined as

$$\mathbf{L} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top,$$

where $\mathbf{U} \in \mathbb{R}^{n\times r}$ and $\mathbf{V} \in \mathbb{R}^{m\times r}$ are the left and right singular vectors of $\mathbf{L}$, respectively, and $\boldsymbol{\Sigma}$ is the diagonal matrix with singular values on the diagonal. Here, matrices $\mathbf{U}$ and $\mathbf{V}$ each have orthogonal columns, and the non-negative entries $\boldsymbol{\Sigma}_{ii} = \sigma_i$ are arranged in descending order. We define $\mathcal{L}$ as the linear subspace consisting of matrices spanning the same row or column space as $\mathbf{L}$, i.e.,

$$\mathcal{L} := \{\mathbf{U}\mathbf{W}_1^\top + \mathbf{W}_2\mathbf{V}^\top, \mathbf{W}_1 \in \mathbb{R}^{m\times r}, \mathbf{W}_2 \in \mathbb{R}^{n\times r}\}.$$

Next, let $\mathcal{S}_e$ ($\mathcal{S}_c$ for the column-wise sparsity setting) be the space spanned by $d \times m$ matrices with the same non-zero support (column support, denoted as $\text{csupp}(\cdot)$) as $\mathbf{S}$, and let the space $\mathcal{D}$ denote the space spanned by the dictionary sparse component under our model be defined as

$$\mathcal{D} := \{\mathbf{D}\mathbf{H}\}, \text{where} \begin{cases} \mathbf{H} \in \mathcal{S}_e \text{ for entry-wise case,} \\ \text{csupp}(\mathbf{H}) \subseteq \mathcal{I}_{\mathcal{S}_c} \text{ for column-wise case.} \end{cases}$$

Here, $\mathcal{I}_{\mathcal{S}_c}$ denotes the index set containing the non-zero column index set of $\mathbf{S}$ for the column-wise case.

Also, we denote the corresponding complements of the spaces described above by appending '$\perp$'. In addition, we use calligraphic '$\mathcal{P}_{\mathcal{G}}(\cdot)$' to denote the projection operator onto a subspace $\mathcal{G}$, and '$\mathbf{P}_{\mathbf{G}}$' to denote the corresponding projection matrix. For instance, we define $\mathcal{P}_{\mathcal{U}}(\cdot)$ and $\mathcal{P}_{\mathcal{V}}(\cdot)$ as the projection operators corresponding to the column space $\mathcal{U}$ and row space $\mathcal{V}$ of the low-rank component $\mathbf{L}$. Therefore, for a given matrix $\mathbf{X} \in \mathbb{R}^{n\times m}$,

$$\mathcal{P}_{\mathcal{U}}(\mathbf{X}) = \mathbf{P}_{\mathbf{U}}\mathbf{X} \text{ and } \mathcal{P}_{\mathcal{V}}(\mathbf{X}) = \mathbf{X}\mathbf{P}_{\mathbf{V}},$$

where $\mathbf{P}_{\mathbf{U}} = \mathbf{U}\mathbf{U}^\top$ and $\mathbf{P}_{\mathbf{V}} = \mathbf{V}\mathbf{V}^\top$. With this, the projection operators onto, and orthogonal to the subspace $\mathcal{L}$ are respectively defined as

$$\mathcal{P}_{\mathcal{L}}(\mathbf{X}) = \mathbf{P}_{\mathbf{U}}\mathbf{X} + \mathbf{X}\mathbf{P}_{\mathbf{V}} - \mathbf{P}_{\mathbf{U}}\mathbf{X}\mathbf{P}_{\mathbf{V}}, \quad \text{and}$$
$$\mathcal{P}_{\mathcal{L}^\perp}(\mathbf{X}) = (\mathbf{I} - \mathbf{P}_{\mathbf{U}})\mathbf{X}(\mathbf{I} - \mathbf{P}_{\mathbf{V}}).$$

D. Incoherence Measures and Other Parameters

We employ various notions of incoherence to identify the conditions under which our procedures succeed. To this end, we first define the incoherence parameter μ, that characterizes the relationship between the low-rank part, $\mathbf{L}$, and the dictionary sparse part $\mathbf{D}\mathbf{S}$ as,

$$\mu := \max_{\mathbf{Z} \in \mathcal{D}\backslash\{\mathbf{0}\}} \frac{\|\mathcal{P}_{\mathcal{L}}(\mathbf{Z})\|_{\text{F}}}{\|\mathbf{Z}\|_{\text{F}}}. \quad (2)$$

The parameter $\mu \in [0, 1]$ is the measure of degree of similarity between the low-rank part and the dictionary sparse component. Here, a larger μ implies that the dictionary sparse component is close to the low-rank part, while a small μ indicates otherwise. In addition, we also define the parameter β_U as

$$\beta_{\mathbf{U}} := \max_{\|\mathbf{u}\|=1} \frac{\|(\mathbf{I} - \mathbf{P}_{\mathbf{U}})\mathbf{D}\mathbf{u}\|^2}{\|\mathbf{D}\mathbf{u}\|^2}, \quad (3)$$

which measures the similarity between the orthogonal complement of the column-space $\mathcal{U}$ and the dictionary $\mathbf{D}$.

The next two measures of incoherence can be interpreted as a way to identify the cases where for $\mathbf{L}$ with SVD as $\mathbf{L}=\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$: (a) $\mathbf{U}$ resembles the dictionary $\mathbf{D}$, and/or (b) $\mathbf{V}$ resembles the sparse coefficient matrix $\mathbf{S}$. In these cases, the low-rank part may mimic the dictionary sparse component. To this end, similar to [14], we define the following to measure these properties respectively as

$$\text{(a) } \gamma_{\mathbf{U}} := \max_i \frac{\|\mathbf{P}_{\mathbf{U}}\mathbf{D}\mathbf{e}_i\|^2}{\|\mathbf{D}\mathbf{e}_i\|^2} \text{ and (b) } \gamma_{\mathbf{V}} := \max_i \|\mathbf{P}_{\mathbf{V}}\mathbf{e}_i\|^2. \quad (4)$$

Here, $\gamma_{\mathbf{U}} \leq 1$, and achieves the upper bound when a dictionary element is exactly aligned with the column space $\mathcal{U}$ of $\mathbf{L}$. Moreover, $\gamma_{\mathbf{V}} \in [r/nm, 1]$ achieves the upper bound when the row-space of $\mathbf{L}$ is "spiky", i.e., a certain row of $\mathbf{V}$ is 1-sparse, meaning that a column of $\mathbf{L}$ is supported by (can be expressed as a linear combination of) a column of $\mathbf{U}$. The lower bound here is attained when it is "spread-out", i.e., each column of $\mathbf{L}$ is a linear combination of all columns of $\mathbf{U}$. In general, our recovery of the two components is easier when the incoherence parameters $\gamma_{\mathbf{U}}$ and $\gamma_{\mathbf{V}}$ are closer to their lower bounds. Further, for notational convenience, we define constants

$$\xi_e := \|\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top\|_\infty \text{ and } \xi_c := \|\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top\|_{\infty,2}. \quad (5)$$

Here, ξ_e is the maximum absolute entry of $\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top$, which measures how close columns of $\mathbf{D}$ are to the singular vectors of $\mathbf{L}$. Similarly, for the column-wise case, ξ_c measures the closeness of columns of $\mathbf{D}$ to the singular vectors of $\mathbf{L}$ under a different metric (column-wise maximum ℓ_2-norm).

III. Main Results

We present the main results corresponding to each sparsity structure of $\mathbf{S}$ in this section.

A. Exact Recovery for Entry-wise Sparsity Case

Our main result establishes the existence of a regularization parameter λ_e, for which solving the optimization problem D-RPCA(E) will recover the components $\mathbf{L}$ and $\mathbf{S}$ exactly. To this end, we will show that such a λ_e belongs to a non-empty interval $[\lambda_e^{\min}, \lambda_e^{\max}]$ with $\lambda_e^{\min}$ and $\lambda_e^{\max}$ defined as

$$\lambda_e^{\min} := \tfrac{1+C_e}{1-C_e}\,\xi_e \text{ and } \lambda_e^{\max} := \tfrac{\sqrt{\alpha_\ell}(1-\mu)-\sqrt{r\alpha_u}\mu}{\sqrt{s_e}}, \quad (6)$$

where $0 \leq C_e < 1$ is a constant that captures the relationship between different model parameters, and is defined as

$$C_e := \tfrac{c}{\alpha_\ell(1-\mu)^2 - c},$$

and c is defined as

$$c := \begin{cases} c_t = \frac{\alpha_u((1+2\gamma_{\mathbf{U}})(\min(s_e,d)+s_e\gamma_{\mathbf{V}})+2\gamma_{\mathbf{V}}\min(s_e,m))}{2} \\ \qquad - \frac{\alpha_\ell(\min(s_e,d)+s_e\gamma_{\mathbf{V}})}{2}, & \text{for } d \leq n, \\ c_f = \frac{\alpha_u((1+2\gamma_{\mathbf{U}})(k+s_e\gamma_{\mathbf{V}})+2\gamma_{\mathbf{V}}\min(s_e,m))}{2} \\ \qquad - \frac{\alpha_\ell(k+s_e\gamma_{\mathbf{V}})}{2}, & \text{for } d > n. \end{cases}$$

Given these definitions, we formalize the theorem for the entry-wise case as following, and its corresponding analysis is provided in Section IV-A.

Theorem 1. *Suppose $\mathbf{M}=\mathbf{L}+\mathbf{D}\mathbf{S}$, where $\operatorname{rank}(\mathbf{L})=r$ and $\mathbf{S}$ has at most s_e non-zeros, i.e., $\|\mathbf{S}\|_0 \leq s_e \leq s_e^{\max} := \frac{(1-\mu)^2}{2}\frac{m}{r}$. Given $\mu \in [0,1]$, $\gamma_{\mathbf{U}}$, $\gamma_{\mathbf{V}} \in [r/m, 1]$, ξ_e defined in (2), (4), (5), and any $\lambda_e \in [\lambda_e^{\min}, \lambda_e^{\max}]$ with $\lambda_e^{\max} > \lambda_e^{\min} \geq 0$ defined in (6), the dictionary $\mathbf{D} \in \mathbb{R}^{n\times d}$ obeys the generalized frame property **D.2** with frame bounds $[\alpha_\ell, \alpha_u]$, solving D-RPCA(E) will recover matrices $\mathbf{L}$ and $\mathbf{S}$ if the following conditions hold:*

- *For $d \leq n$, $\mathcal{R}$ may contain the entire space and $\gamma_{\mathbf{U}}$ follows*

$$\gamma_{\mathbf{U}} \leq \begin{cases} \frac{(1-\mu)^2-2s_e\gamma_{\mathbf{V}}}{2s_e(1+\gamma_{\mathbf{V}})}, & \text{for } s_e \leq \min\ (d, s_e^{\max}) \\ \frac{(1-\mu)^2-2s_e\gamma_{\mathbf{V}}}{2(d+s_e\gamma_{\mathbf{V}})}, & \text{for } d < s_e \leq s_e^{\max} \end{cases}; \quad (7)$$

- *For $d > n > C_1\ k\log(n)$ for a constant C_1, $\mathcal{R}$ consists of all k sparse vectors, and $\gamma_{\mathbf{U}}$ follows*

$$\gamma_{\mathbf{U}} \leq \tfrac{(1-\mu)^2-2s_e\gamma_{\mathbf{V}}}{2(k+s_e\gamma_{\mathbf{V}})}. \quad (8)$$

Theorem 1 establishes the sufficient conditions for the existence of λ_e to guarantee recovery of $(\mathbf{L}, \mathbf{S})$ for both the *thin* and the *fat* cases. The conditions on $\gamma_{\mathbf{U}}$ dictated by (7) and (8), for the thin and fat case, respectively, arise from ensuring that $\lambda_e^{\min} \geq 0$. Further, the condition $\lambda_e^{\min} < \lambda_e^{\max}$, translates to the following sufficient condition on rank r in terms of the sparsity s_e,

$$r < \left(\sqrt{\tfrac{\alpha_\ell}{\alpha_u}}\tfrac{1-\mu}{\mu} - \tfrac{\xi_e}{\sqrt{\alpha_u}\mu}\tfrac{1+C_e}{1-C_e}\sqrt{s_e}\right)^2, \quad (9)$$

for the recovery of $(\mathbf{L}, \mathbf{S})$. This relationship matches with our empirical evaluations and will be revisited in Section V-A.

We note that for both, *thin* and *fat* dictionary case, the conditions are closely related to the incoherence measures (μ, $\gamma_{\mathbf{V}}$, and $\gamma_{\mathbf{U}}$) between the low-rank part, $\mathbf{L}$, the dictionary, $\mathbf{D}$, and the sparse component $\mathbf{S}$. In general, smaller sparsity, rank, and incoherence parameters are sufficient for ensuring the recovery of the components for a particular problem. This is in line with our intuition – the more distinct the two components, the easier it should be to tease them apart. Moreover, we observe that the theorem imposes an an upper-bound on the global sparsity, i.e., $s_e \leq s_e^{\max} = \mathcal{O}(\frac{m}{r})$. This bound is similar to the result in [21], and is due to the deterministic nature of our analysis w.r.t. the locations of the non-zero elements of coefficients $\mathbf{S}$.

B. Exact Recovery for Column-wise Sparsity Case

Recall that we consider the oracle model in this case as described in **D.1** owing to the intrinsic ambiguity in recovery of $(\mathbf{L}, \mathbf{S})$; see our discussion in Section I-B. To demonstrate its recoverability, the following lemma establishes the sufficient conditions for the existence of an optimal pair $(\mathbf{L}_0, \mathbf{S}_0)$. The proof is provided in Appendix B.

Lemma 2. *Given $\mathbf{M}$, $\mathbf{D}$, and $(\mathcal{L}, \mathcal{S}_c, \mathcal{D})$, any pair $(\mathbf{L}_0, \mathbf{S}_0) \in \{\mathbf{M}, \mathcal{U}, \mathcal{I}_{\mathcal{S}_c}\}$ satisfies $\operatorname{span}\{col(\mathbf{L}_0)\} = \mathcal{U}$ and $csupp(\mathbf{S}_0) = \mathcal{I}_{\mathcal{S}_c}$ if $\mu < 1$.*

Analogous to the entry-wise case, we will show the existence of a non-empty interval $[\lambda_c^{\min}, \lambda_c^{\max}]$ for the regularization parameter λ_c, for which solving D-RPCA(C) recovers an optimal pair as per Lemma 2. Here, for a constant $C_c := \frac{\alpha_u}{\alpha_\ell}\frac{1}{(1-\mu)^2}\gamma_{\mathbf{V}}\beta_{\mathbf{U}}$, $\lambda_c^{\min}$ and $\lambda_c^{\max}$ are defined as

$$\lambda_c^{\min} := \tfrac{\xi_c+\sqrt{rs_c\alpha_u}\mu C_c}{1-s_cC_c} \text{ and } \lambda_c^{\max} := \tfrac{\sqrt{\alpha_\ell}(1-\mu)-\sqrt{r\alpha_u}\mu}{\sqrt{s_c}}. \quad (10)$$

Then, our main result for the column-wise case is as follows, and its analysis is provided in Section IV-B.

Theorem 3. *Suppose* $\mathbf{M} = \mathbf{L} + \mathbf{DS}$ *with* $(\mathbf{L}, \mathbf{S})$ *defining the oracle model* $\{\mathbf{M}, \mathcal{U}, \mathcal{I}_{\mathcal{S}_c}\}$*, where* $\text{rank}(\mathbf{L}) = r$*,* $|\mathcal{I}_{\mathcal{S}_c}| = s_c$ *for* $s_c \leq s_c^{\max} := \frac{\alpha_\ell}{\alpha_u \gamma_{\mathbf{V}}} \cdot \frac{(1-\mu)^2}{\beta_{\mathbf{U}}}$*. Given* $\mu \in [0, 1)$*,* $\beta_{\mathbf{U}}$*,* $\gamma_{\mathbf{V}} \in [r/m, 1]$*,* ξ_c *defined in* (2)*,* (3)*,* (4)*,* (5)*, and any* $\lambda_c \in [\lambda_c^{\min}, \lambda_c^{\max}]$*, for* $\lambda_c^{\max} > \lambda_c^{\min} \geq 0$ *defined in* (10)*, solving D-RPCA(C) will recover a pair of components* $(\mathbf{L}_0, \mathbf{S}_0) \in \{\mathbf{M}, \mathcal{U}, \mathcal{I}_{\mathcal{S}_c}\}$*, if the space* $\mathcal{R}$ *is structured such that the dictionary* $\mathbf{D} \in \mathbb{R}^{n \times d}$ *obeys the generalized frame property* ***D.2*** *with frame bounds* $[\alpha_\ell, \alpha_u]$*, for* $\alpha_\ell > 0$.

Theorem 3 states the conditions under which the solution to the optimization problem D-RPCA(C) will be in the oracle model defined in **D.1**. The condition on the column sparisty $s_c \leq s_c^{\max}$ is a result of the constraint that $\lambda_c^{\min} \geq 0$. Similar to (9), requiring $\lambda_c^{\max} > \lambda_c^{\min}$ leads to the following sufficient condition on the rank r in terms of the sparsity s_c,

$$r < \left(\sqrt{\frac{\alpha_\ell}{\alpha_u}} \frac{1-\mu}{\mu} - \frac{\xi_c}{\sqrt{\alpha_u}\mu} \sqrt{s_c} \right)^2. \quad (11)$$

Moreover, suppose that $1 \lesssim \alpha_l \leq \alpha_u \lesssim 1$, which can be easily met by a tight frame when $d < n$, or a RIP type condition when $d > n$. Further, if $\frac{(1-\mu)^2}{\beta_U}$ is a constant, then since $\gamma_{\mathbf{V}} = \Theta(\frac{r}{m})$, we have that $s_c^{\max} = \mathcal{O}(\frac{m}{r})$. This is of the same order with the upper bound of s_c in the Outlier Pursuit (OP) [21].

Our numerical results in Section V further show that D-RPCA(C) can be much more robust than OP, and may recover $\{\mathcal{U}, \mathcal{I}_{\mathcal{S}_c}\}$ even when the rank of $\mathbf{L}$ is high and the number of outliers s_c is a constant proportion of m.

IV. Proof of Main Results

A. *Proof of Theorem 1*

We use dual certificate construction procedure to prove the main result in Theorem. 1; the proofs of all lemmata used here are listed in Appendix A. To this end, we start by constructing a dual certificate for the convex problem shown in D-RPCA(E). Here, we first show the conditions the dual certificate needs to satisfy via the following lemma.

Lemma 4. *If there exists a dual certificate* $\mathbf{\Gamma} \in \mathbb{R}^{n \times m}$ *satisfying*

(C1) $\mathcal{P}_{\mathcal{L}}(\mathbf{\Gamma}) = \mathbf{U}\mathbf{V}^\top$, **(C2)** $\mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top \mathbf{\Gamma}) = \lambda_e \ \text{sign}(\mathbf{S}_0)$,
(C3) $\|\mathcal{P}_{\mathcal{L}^\perp}(\mathbf{\Gamma})\| < 1$, *and* **(C4)** $\|\mathcal{P}_{\mathcal{S}_e^\perp}(\mathbf{D}^\top \mathbf{\Gamma})\|_\infty < \lambda_e$.

then the pair $(\mathbf{L}_0, \ \mathbf{S}_0)$ *is the unique solution of D-RPCA(E).*

We will now proceed with the construction of the dual certificate which satisfies the conditions outlined by **(C1)-(C4)** by Lemma 4. Using the analysis similar to [14] (Section V. B.), we construct the dual certificate as

$$\mathbf{\Gamma} = \mathbf{U}\mathbf{V}^\top + (\mathbf{I} - \mathbf{P_U})\mathbf{X}(\mathbf{I} - \mathbf{P_V}),$$

for arbitrary $\mathbf{X} \in \mathbb{R}^{n \times m}$. The condition **(C1)** is readily satisfied by our choice of $\mathbf{\Gamma}$. For **(C2)**, we substitute the expression for $\mathbf{\Gamma}$ to arrive at

$$\mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top \mathbf{U}\mathbf{V}^\top) + \mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top (\mathbf{I} - \mathbf{P_U})\mathbf{X}(\mathbf{I} - \mathbf{P_V})) = \lambda_e \ \text{sign}(\mathbf{S}_0). \quad (12)$$

Letting $\mathbf{Z} := \mathbf{D}^\top (\mathbf{I} - \mathbf{P_U})\mathbf{X}(\mathbf{I} - \mathbf{P_V})$ and

$$\mathbf{B}_{\mathcal{S}_\mathbf{e}} := \lambda_e \ \text{sign}(\mathbf{S}_0) - \mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top \mathbf{U}\mathbf{V}^\top),$$

we can write (12) as $\mathcal{P}_{\mathcal{S}_e}(\mathbf{Z}) = \mathbf{B}_{\mathcal{S}_\mathbf{e}}$. Further, we can vectorize the equation above as $\mathcal{P}_{\mathcal{S}_e}(\text{vec}(\mathbf{Z})) = \text{vec}(\mathbf{B}_{\mathcal{S}_\mathbf{e}})$. Let $\mathbf{b}_{\mathcal{S}_e}$ be a length s_e vector containing elements of $\mathbf{B}_{\mathcal{S}_e}$ corresponding to the support of $\mathbf{S}_0$. Now, note that $\text{vec}(\mathbf{Z})$ can be represented in terms of a Kronecker product as follows,

$$\text{vec}(\mathbf{Z}) = [(\mathbf{I} - \mathbf{P_V}) \otimes \mathbf{D}^\top (\mathbf{I} - \mathbf{P_U})]\text{vec}(\mathbf{X}).$$

On defining $\mathbf{A} := (\mathbf{I} - \mathbf{P_V}) \otimes \mathbf{D}^\top (\mathbf{I} - \mathbf{P_U}) \in \mathbb{R}^{md \times mn}$, we have $\text{vec}(\mathbf{Z}) = \mathbf{A}\text{vec}(\mathbf{X})$. Further, let $\mathbf{A}_{\mathcal{S}_\mathbf{e}} \in \mathbb{R}^{s \times nm}$ denote the rows of $\mathbf{A}$ that correspond to support of $\mathbf{S}_0$, and let $\mathbf{A}_{\mathcal{S}_\mathbf{e}^\perp}$ correspond to the remaining rows of $\mathbf{A}$. Using these definitions and results, we have $\mathbf{A}_{\mathcal{S}_e}\text{vec}(\mathbf{X}) = \mathbf{b}_{\mathcal{S}_e}$. Thus, for conditions **(C1)** and **(C2)** to be satisfied, we need

$$\text{vec}(\mathbf{X}) = \mathbf{A}_{\mathcal{S}_e}^\top (\mathbf{A}_{\mathcal{S}_e} \mathbf{A}_{\mathcal{S}_e}^\top)^{-1} \mathbf{b}_{\mathcal{S}_e}. \quad (13)$$

Here, the following result ensures the existence of the inverse.

Lemma 5. *If* $\mu < 1$ *and* $\alpha_\ell > 0$*,* $\sigma_{\min}(\mathbf{A}_{\mathcal{S}_e})$ *satisfies the bound* $\sigma_{\min}(\mathbf{A}_{\mathcal{S}_e}) \geq \sqrt{\alpha_\ell}(1 - \mu)$.

Now, we look at the condition **(C3)** $\|\mathcal{P}_{\mathcal{L}^\perp}(\mathbf{\Gamma})\| < 1$. This is where our analysis departs from [14]; we write

$$\begin{aligned} \|\mathcal{P}_{\mathcal{L}^\perp}(\mathbf{\Gamma})\| &= \|(\mathbf{I} - \mathbf{P_U})\mathbf{X}(\mathbf{I} - \mathbf{P_V})\| \\ &\leq \|\mathbf{X}\| \leq \|\mathbf{X}\|_\text{F} \leq \|\mathbf{A}_{\mathcal{S}_e}^\top (\mathbf{A}_{\mathcal{S}_e} \mathbf{A}_{\mathcal{S}_e}^\top)^{-1}\| \|\mathbf{b}_{\mathcal{S}_e}\|_2, \end{aligned}$$

where we have used the fact that $\|(\mathbf{I} - \mathbf{P_U})\| \leq 1$ and $\|(\mathbf{I} - \mathbf{P_V})\| \leq 1$. Now, as $\mathbf{A}_{\mathcal{S}_e}^\top (\mathbf{A}_{\mathcal{S}_e} \mathbf{A}_{\mathcal{S}_e}^\top)^{-1}$ is the pseudo-inverse of $\mathbf{A}_{\mathcal{S}_e}$, i.e., $\mathbf{A}_{\mathcal{S}_e} \mathbf{A}_{\mathcal{S}_e}^\top (\mathbf{A}_{\mathcal{S}_e} \mathbf{A}_{\mathcal{S}_e}^\top)^{-1} = \mathbf{I}$, we have that $\|\mathbf{A}_{\mathcal{S}_e}^\top (\mathbf{A}_{\mathcal{S}_e} \mathbf{A}_{\mathcal{S}_e}^\top)^{-1}\| = 1/\sigma_{\min}(\mathbf{A}_{\mathcal{S}_e})$, where $\sigma_{\min}(\mathbf{A}_{\mathcal{S}_e})$ is the smallest singular value of $\mathbf{A}_{\mathcal{S}_e}$. Therefore, we have

$$\|\mathcal{P}_{\mathcal{L}^\perp}(\mathbf{\Gamma})\| \leq \frac{\|\mathbf{b}_{\mathcal{S}_e}\|_2}{\sigma_{\min}(\mathbf{A}_{\mathcal{S}_e})}. \quad (14)$$

The following lemma establishes an upper bound on $\|\mathbf{b}_{\mathcal{S}_e}\|_2$.

Lemma 6. $\|\mathbf{b}_{\mathcal{S}_e}\|_2$ *satisfies the bound* $\|\mathbf{b}_{\mathcal{S}_e}\|_2 \leq \lambda_e \sqrt{s_e} + \sqrt{r \alpha_u} \mu$.

Combining (14), Lemma 5, and Lemma 6, we have

$$\|\mathcal{P}_{\mathcal{L}^\perp}(\mathbf{\Gamma})\| \leq \frac{\lambda_e \sqrt{s_e} + \sqrt{r \alpha_u} \mu.}{\sqrt{\alpha_\ell}(1-\mu)}. \quad (15)$$

Now, combining (15) and the upper bound on λ_e defined in (6), we have that **(C3)** holds.

Now, we move onto finding conditions under which **(C4)** is satisfied by our dual certificate. For this we will bound $\|\mathcal{P}_{\mathcal{S}_e^\perp}(\mathbf{D}^\top \mathbf{\Gamma})\|_\infty$. Our analysis follows the similar procedure as employed in deriving (16) in [14], reproduced here for completeness. First, by the definition of $\mathbf{\Gamma}$ and properties of the $\|.\|_\infty$ norm, we have

$$\|\mathcal{P}_{\mathcal{S}_e^\perp}(\mathbf{D}^\top \mathbf{\Gamma})\|_\infty \leq \|\mathcal{P}_{\mathcal{S}_e^\perp}(\mathbf{Z})\|_\infty + \|\mathcal{P}_{\mathcal{S}_e^\perp}(\mathbf{D}^\top \mathbf{U}\mathbf{V})\|_\infty. \quad (16)$$

We now focus on simplifying the term $\|\mathcal{P}_{\mathcal{S}_e^\perp}(\mathbf{Z})\|_\infty$. By definition of $\mathbf{A}$, and using the fact that $\text{vec}(\mathbf{Z}) = \mathbf{A}\text{vec}(\mathbf{X})$, we have $\mathcal{P}_{\mathcal{S}_e^\perp}(\mathbf{Z}) = \mathbf{A}_{\mathcal{S}_e^\perp}\text{vec}(\mathbf{X})$, which implies

$$\begin{aligned} \|\mathcal{P}_{\mathcal{S}_e^\perp}(\mathbf{Z})\|_\infty &= \|\mathbf{A}_{\mathcal{S}_e^\perp}\text{vec}(\mathbf{X})\|_\infty \\ &= \|\mathbf{A}_{\mathcal{S}_e^\perp} \mathbf{A}_{\mathcal{S}_e}^\top (\mathbf{A}_{\mathcal{S}_e} \mathbf{A}_{\mathcal{S}_e}^\top)^{-1} \mathbf{b}_{\mathcal{S}_e}\|_\infty, \end{aligned}$$

where we have used the result on $\text{vec}(\mathbf{X})$ shown in (13).

Now, since $\|\mathbf{b}_{\mathcal{S}_e}\|_\infty$ can be written as

$$\|\mathbf{b}_{\mathcal{S}_e}\|_\infty = \|\mathbf{B}_{\mathcal{S}_e}\|_\infty = \|\lambda_e \text{sign}(\mathbf{A}_0) - \mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top \mathbf{U}\mathbf{V}^\top)\|_\infty.$$

Now, using the following upper bound on $\|\mathbf{b}_{\mathcal{S}_e}\|_\infty$,

Lemma 7. $\|\mathbf{b}_{\mathcal{S}_e}\|_\infty$ *satisfies the bound* $\|\mathbf{b}_{\mathcal{S}_e}\|_\infty \leq \lambda_e + \|\mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top \mathbf{U}\mathbf{V}^\top)\|_\infty$.

and on defining

$$\mathbf{Q} := \mathbf{A}_{\mathcal{S}_e^\perp}\mathbf{A}_{\mathcal{S}_e}^\top(\mathbf{A}_{\mathcal{S}_e}\mathbf{A}_{\mathcal{S}_e}^\top)^{-1},$$

we have

$$\begin{aligned}\|\mathcal{P}_{\mathcal{S}_e^\perp}(\mathbf{Z})\|_\infty &= \|\mathbf{Q}\mathbf{b}_{\mathcal{S}_e}\|_\infty \leq \|\mathbf{Q}\|_{\infty,\infty}\|\mathbf{b}_{\mathcal{S}_e}\|_\infty \\ &= \|\mathbf{Q}\|_{\infty,\infty}\|\lambda_e \text{sign}(\mathbf{A}_0) - \mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)\|_\infty, \\ &\leq \|\mathbf{Q}\|_{\infty,\infty}(\lambda_e + \|\mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)\|_\infty),\end{aligned}$$

where we have the following bound for $\|\mathbf{Q}\|_{\infty,\infty}$.

Lemma 8. $\|\mathbf{Q}\|_{\infty,\infty}$ *satisfies the bound* $\|\mathbf{Q}\|_{\infty,\infty} \leq C_e(\alpha_u, \alpha_\ell, \gamma_\mathbf{U}, \gamma_\mathbf{V}, s_e, d, k, \mu)$*, where*

$$C_e := \tfrac{c}{\alpha_\ell(1-\mu)^2 - c}$$

where $0 \leq C_e < 1$ *and* c *is defined as*

$$c := \begin{cases} c_t = \frac{\alpha_u((1+2\gamma_\mathbf{U})(\min(s_e,d)+s_e\gamma_\mathbf{V})+2\gamma_\mathbf{V}\min(s_e,m))}{2} \\ \qquad - \frac{\alpha_\ell(\min(s_e,d)+s_e\gamma_\mathbf{V})}{2}, \; \textit{for } d \leq n, \\ c_f = \frac{\alpha_u((1+2\gamma_\mathbf{U})(k+s_e\gamma_\mathbf{V})+2\gamma_\mathbf{V}\min(s_e,m))}{2} \\ \qquad - \frac{\alpha_\ell(k+s_e\gamma_\mathbf{V})}{2}, \qquad \textit{for } d > n. \end{cases} \tag{17}$$

Combining this with (16) and Lemma 8, we have

$$\begin{aligned}\|\mathcal{P}_{\mathcal{S}_e^\perp}(\mathbf{D}^\top\boldsymbol{\Gamma})\|_\infty \leq C_e\Big(\lambda_e + \|\mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)\|_\infty\Big) \\ + \|\mathcal{P}_{\mathcal{S}_e^\perp}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)\|_\infty. \end{aligned} \tag{18}$$

By simplifying (18), we arrive at the lower bound $\lambda_e^{\min}$ for λ_e as in (6), from which **(C4)** holds. Gleaning from the expressions for $\lambda_e^{\max}$ and $\lambda_e^{\min}$, we observe that $\lambda_e^{\max} > \lambda_e^{\min} \geq 0$ for the existence of λ_e that can recover the desired matrices. This completes the proof. ■

Characterizing $\lambda_e^{\min}$: In the previous section, we characterized the $\lambda_e^{\min}$ and $\lambda_e^{\max}$ based on the dual certificate construction procedure. For the recovery of the true pair $(\mathbf{L}, \mathbf{S})$, we require $\lambda_e^{\max} > \lambda_e^{\min} \geq 0$. Since $\xi_e \geq 0$ and $c \geq 0$ by definition, we need $0 \leq C_e < 1$ for $\lambda_e^{\min} > 0$, i.e.,

$$c < \tfrac{1}{2}\alpha_\ell(1-\mu)^2 \geq \tfrac{\alpha_\ell}{2}. \tag{19}$$

***Conditions for thin* D**: To simplify the analysis we assume, without loss of generality, that $d < m$. Specifically, we will assume that $d \leq \frac{m}{\alpha r}$, where $\alpha > 1$ is a constant. With this assumption in mind, we will analyze the following cases for the global sparsity, when $s_e \leq d$ and $d < s_e \leq m$.
Case 1: $s_e \leq d$. For this case c_t is given by

$$c_t = \tfrac{s_e\alpha_u}{2}[(1+2\gamma_\mathbf{U})(1+\gamma_\mathbf{V}) + 2\gamma_\mathbf{V}] - \tfrac{s_e\alpha_\ell}{2}(1+\gamma_\mathbf{V})$$

From (19), we have $\alpha_\ell(1-\mu)^2 - 2c_t > 0$, which leads to

$$\tfrac{\alpha_u}{\alpha_\ell} < \tfrac{(1-\mu)^2 + s_e(1+\gamma_\mathbf{V})}{s_e(1+2\gamma_\mathbf{U})(1+\gamma_\mathbf{V}) + 2s_e\gamma_\mathbf{V}}.$$

As per the GFP of **D.2**, we also require that $\alpha_u/\alpha_\ell \geq 1$. Therefore we arrive at

$$\gamma_\mathbf{U} < \tfrac{(1-\mu)^2 - 2s_e\gamma_\mathbf{V}}{2s_e(1+\gamma_\mathbf{V})}.$$

Further, since $\gamma_\mathbf{U} \geq 0$, we require the numerator to be positive, and since the lower bound on $\gamma_\mathbf{V} \geq \frac{r}{m}$, we have

$$s_e \leq \tfrac{(1-\mu)^2}{2}\tfrac{m}{r} := s_e^{\max},$$

which also implies $s_e \leq m$. Now, the condition $c_t \geq 0$ implies

$$\tfrac{\alpha_u}{\alpha_\ell} \geq \tfrac{1+\gamma_\mathbf{V}}{(1+2\gamma_\mathbf{U})(1+\gamma_\mathbf{V})+2\gamma_\mathbf{V}}.$$

Since, the R.H.S. of this inequality is upper bounded by 1 (achieved when $\gamma_\mathbf{U}$ and $\gamma_\mathbf{V}$ are zero). This condition on c_t is satisfied by our assumption that $\alpha_u/\alpha_\ell \geq 1$.
Case 2: $d < s_e \leq m$. For this case, we have

$$c_t = \tfrac{\alpha_u}{2}((1+2\gamma_\mathbf{U})(d+s_e\gamma_\mathbf{V}) + 2s_e\gamma_\mathbf{V}) - \tfrac{\alpha_\ell}{2}(d+s_e\gamma_\mathbf{V}).$$

From (19), we have

$$\tfrac{\alpha_u}{\alpha_\ell} < \tfrac{(1-\mu)^2+(d+s_e\gamma_\mathbf{V})}{(1+2\gamma_\mathbf{U})(d+s_e\gamma_\mathbf{V})+2s_e\gamma_\mathbf{V}}.$$

Again, due to the requirement that $\alpha_u/\alpha_\ell \geq 1$, following a similar argument as in the previous case we conclude that

$$\gamma_\mathbf{U} \leq \tfrac{(1-\mu)^2 - 2s_e\gamma_\mathbf{V}}{2(d+s_e\gamma_\mathbf{V})} \text{ and } s_e \leq \tfrac{(1-\mu)^2}{2}\tfrac{m}{r}.$$

As in the previous case the $c_t \geq 0$ is met by our due to our assumption on the frame bounds.

***Conditions for fat* D**: To simplify the analysis, we suppose that $k < m$. Note that in this case, we require that the coefficient matrix $\mathbf{S}$ has k-sparse columns. Now, $c = c_f$ is given by

$$c_f := \tfrac{\alpha_u}{2}((1+2\gamma_\mathbf{U})(k+s_e\gamma_\mathbf{V}) + 2\gamma_\mathbf{V}s_e) - \tfrac{\alpha_l}{2}(k+s_e\gamma_\mathbf{V})$$

As for the *thin* case, we substitute the expression for c_f in (19) as follows

$$\tfrac{\alpha_u}{\alpha_\ell} < \tfrac{(1-\mu)^2+(k+s_e\gamma_\mathbf{V})}{(1+2\gamma_\mathbf{U})(k+s_e\gamma_\mathbf{V})+2s_e\gamma_\mathbf{V}}$$

Again, by GFP we require that $\alpha_u/\alpha_\ell \geq 1$, therefore we have

$$\gamma_\mathbf{U} < \tfrac{(1-\mu)^2 - 2s_e\gamma_\mathbf{V}}{2(k+s_e\gamma_\mathbf{V})} \text{ and } s_e \leq \tfrac{(1-\mu)^2}{2}\tfrac{m}{r},$$

Also, the condition that $c_f \geq 0$ is met by the assumption that $\mathbf{D}$ obeys GFP.

Characterizing $\lambda_e^{\max}$: Further, the condition $\lambda_e^{\min} < \lambda_e^{\max}$, translates to a relationship between rank r, and the sparsity s_e, as shown in (9) for $s_e \leq s_e^{\max}$.

B. Proof of Theorem 3

In this section we prove Theorem 3; the proofs of all lemmata are listed in Appendix B. The Lagrangian of the nonsmooth optimization problem D-RPCA(C) is

$$\mathcal{F}(\mathbf{L}, \mathbf{S}, \boldsymbol{\Lambda}) = \|\mathbf{L}\|_* + \lambda_c\|\mathbf{S}\|_{1,2} + \langle \boldsymbol{\Lambda}, \mathbf{M} - \mathbf{L} - \mathbf{D}\mathbf{S}\rangle, \tag{20}$$

where $\boldsymbol{\Lambda} \in \mathbb{R}^{n\times m}$ is a dual variable. The subdifferentials of (20) with respect to $(\mathbf{L}, \mathbf{S})$ are

$$\begin{aligned}\partial_\mathbf{L}\mathcal{F}(\mathbf{L},\mathbf{S},\boldsymbol{\Lambda}) &= \{\mathbf{U}\mathbf{V}^\top + \mathbf{W} - \boldsymbol{\Lambda}, \|\mathbf{W}\|_2 \leq 1, \mathcal{P}_\mathcal{L}(\mathbf{W}) = \mathbf{0}\}, \\ \partial_\mathbf{S}\mathcal{F}(\mathbf{L},\mathbf{S},\boldsymbol{\Lambda}) &= \Big\{\lambda_c\mathbf{H} + \lambda_c\mathbf{F} - \mathbf{D}^\top\boldsymbol{\Lambda}, \mathcal{P}_{\mathcal{S}_c}(\mathbf{H}) = \mathbf{H}, \\ &\qquad \mathbf{H}_{:,j} = \tfrac{\mathbf{S}_{:,j}}{\|\mathbf{S}_{:,j}\|_2}, \mathcal{P}_{\mathcal{S}_c}(\mathbf{F}) = \mathbf{0}, \|\mathbf{F}\|_{\infty,2} \leq 1\Big\}. \end{aligned} \tag{21}$$

Then we claim that a pair $(\mathbf{L}, \mathbf{S})$ is an optimal point of D-RPCA(C) if and only if the following hold by the optimality conditions:

$$\mathbf{0}_{n\times m} \in \partial_\mathbf{L}\mathcal{F}(\mathbf{L},\mathbf{S},\boldsymbol{\Lambda}) \text{ and} \tag{22}$$

$$\mathbf{0}_{d\times m} \in \partial_\mathbf{S}\mathcal{F}(\mathbf{L},\mathbf{S},\boldsymbol{\Lambda}). \tag{23}$$

The following lemma states the optimality conditions for the optimal solution pair $(\mathbf{L}, \mathbf{S})$.

Lemma 9. *Given* $\mathbf{M}$ *and* $\mathbf{D}$*, let* $(\mathbf{L}, \mathbf{S})$ *define the oracle model* $\{\mathbf{M}, \mathcal{U}, \mathcal{I}_{\mathcal{S}_c}\}$*. Then any solution* $(\mathbf{L}_0, \mathbf{S}_0) \in \{\mathbf{M}, \mathcal{U}, \mathcal{I}_{\mathcal{S}_c}\}$ *is the an optimal solution pair of D-RPCA(C), if there exists a dual certificate* $\mathbf{\Gamma} \in \mathbb{R}^{n\times m}$ *that satisfies*
$(\mathbf{C1})$ $\mathcal{P}_{\mathcal{L}}(\mathbf{\Gamma}) = \mathbf{U}\mathbf{V}^\top$*,* $(\mathbf{C2})$ $\mathcal{P}_{\mathcal{S}_c}(\mathbf{D}^\top\mathbf{\Gamma}) = \lambda_c\mathbf{H}$*, where* $\mathbf{H}_{:,j} = \mathbf{S}_{:,j}/\|\mathbf{S}_{:,j}\|_2$ *for all* $j \in \mathcal{I}_{\mathcal{S}_c}$*;* $\mathbf{0}$*, otherwise,*
$(\mathbf{C3})$ $\|\mathcal{P}_{\mathcal{L}^\perp}(\mathbf{\Gamma})\|_2 < 1$*, and* $(\mathbf{C4})$ $\|\mathcal{P}_{\mathcal{S}_c{}^\perp}(\mathbf{D}^\top\mathbf{\Gamma})\|_{\infty,2} < \lambda_c$*.*

We first propose $\mathbf{\Gamma}$ as the dual certificate, where

$$\mathbf{\Gamma} = \mathbf{U}\mathbf{V}^\top + (\mathbf{I} - \mathbf{P_U})\,\mathbf{X}\,(\mathbf{I} - \mathbf{P_V})\,, \text{ for any } \mathbf{X} \in \mathbb{R}^{n\times m}.$$

Hence, the condition **(C1)** is readily satisfied by our choice of $\mathbf{\Gamma}$. Now, the condition **(C2)**, defined as $\mathcal{P}_{\mathcal{S}_c}(\mathbf{D}^\top\mathbf{\Gamma}) = \lambda_c\tilde{\mathbf{S}}$, where $\tilde{\mathbf{S}}_{:,j} = \frac{\mathbf{S}_{:,j}}{\|\mathbf{S}_{:,j}\|_2}$ for all $j \in \mathcal{I}_{\mathcal{S}_c}$; $\mathbf{0}$, otherwise. Substituting the expression for $\mathbf{\Gamma}$, we need the following condition to hold

$$\mathcal{P}_{\mathcal{S}_c}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top) + \mathcal{P}_{\mathcal{S}_c}(\mathbf{D}^\top(\mathbf{I}-\mathbf{P_U})\mathbf{X}\,(\mathbf{I}-\mathbf{P_V})) = \lambda_c\tilde{\mathbf{S}}. \quad (24)$$

Letting $\mathbf{Z} := \mathbf{D}^\top\,(\mathbf{I}-\mathbf{P_U})\,\mathbf{X}\,(\mathbf{I}-\mathbf{P_V})$ and $\mathbf{B}_{\mathcal{S}_c} := \lambda_c\tilde{\mathbf{S}} - \mathcal{P}_{\mathcal{S}_c}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)$, we have $\mathcal{P}_{\mathcal{S}_c}(\mathbf{Z}) = \mathbf{B}_{\mathcal{S}_c}$. Further, vectorizing the equation above, we have

$$\mathcal{P}_{\mathcal{S}_c}(\mathrm{vec}(\mathbf{Z})) = \mathbf{b}_{\mathcal{S}_c}, \quad (25)$$

where $\mathbf{b}_{\mathcal{S}_c} := \mathrm{vec}(\mathbf{B}_{\mathcal{S}_c})$. Next, by letting $\mathbf{A} := (\mathbf{I}-\mathbf{P_V}) \otimes \mathbf{D}^\top(\mathbf{I}-\mathbf{P_U})$, using the definition of $\mathbf{Z}$ and the properties of the Kronecker product we have $\mathrm{vec}(\mathbf{Z}) = \mathbf{A}\mathrm{vec}(\mathbf{X})$. Now, let $\mathbf{A}_{\mathcal{S}_c}$ denote the rows of $\mathbf{A}$ corresponding to the non-zero rows of $\mathrm{vec}(\mathbf{S})$ and $\mathbf{A}_{\mathcal{S}_c^\perp}$ denote the remaining rows, then

$$\mathcal{P}_{\mathcal{S}_c}(\mathrm{vec}(\mathbf{Z})) = \mathbf{A}_{\mathcal{S}_c}\mathrm{vec}(\mathbf{X}). \quad (26)$$

From (25) and (26), we have $\mathbf{A}_{\mathcal{S}_c}\mathrm{vec}(\mathbf{X}) = \mathbf{b}_{\mathcal{S}_c}$. Therefore, we need the following

$$\mathrm{vec}(\mathbf{X}) = \mathbf{A}_{\mathcal{S}_c}^\top(\mathbf{A}_{\mathcal{S}_c}\mathbf{A}_{\mathcal{S}_c}^\top)^{-1}\mathbf{b}_{\mathcal{S}_c}, \quad (27)$$

which corresponds to the least norm solution i.e., $\mathbf{X} = \mathrm{argmin}_{\mathbf{X}}\ \|\mathbf{X}\|_{\mathrm{F}}$, s.t. $\mathbf{A}_{\mathcal{S}_c}\mathrm{vec}(\mathbf{X}) = \mathbf{b}_{\mathcal{S}_c}$). For this choice of $\mathbf{X}$ (24) is satisfied and consequently the condition **(C2)**. Here, the existence of the inverse is ensured by the following lemma.

Lemma 10. *If* $\mu < 1$ *and* $\alpha_\ell > 0$*, the minimum singular values of* $\mathbf{A}_{\mathcal{S}_c}$ *is bounded away from* 0 *and is given by* $\sqrt{\alpha_\ell}(1-\mu)$

Upon the existence of such $\mathbf{X}$ as defined in (27), **(C3)** is satisfied if the following condition holds

$$\begin{aligned}\|\mathcal{P}_{\mathcal{L}^\perp}(\mathbf{\Gamma})\|_2 &= \|\,(\mathbf{I}-\mathbf{P_V})\,\mathbf{X}\,(\mathbf{I}-\mathbf{P_U})\,\|_2\\ &\le \|\mathbf{I}-\mathbf{P_V}\|_2\|\mathbf{X}\|_2\|\mathbf{I}-\mathbf{P_U}\|_2 = \|\mathbf{X}\|_2 \le \|\mathbf{X}\|_{\mathrm{F}} < 1.\end{aligned}$$

From (27), this condition translates to

$$\|\mathbf{A}_{\mathcal{S}_c}^\top(\mathbf{A}_{\mathcal{S}_c}\mathbf{A}_{\mathcal{S}_c}^\top)^{-1}\|\|\mathbf{b}_{\mathcal{S}_c}\|_2 < 1.$$

Now, since $\|\mathbf{A}_{\mathcal{S}_c}^\top(\mathbf{A}_{\mathcal{S}_c}\mathbf{A}_{\mathcal{S}_c}^\top)^{-1}\|\| = 1/\sigma_{\min}(\mathbf{A}_{\mathcal{S}_c})$ (see the analogous analysis for the entry-wise case), we need

$$\|\mathcal{P}_{\mathcal{L}^\perp}(\mathbf{\Gamma})\|_2 \le \tfrac{\|\mathbf{b}_{\mathcal{S}_c}\|_2}{\sigma_{\min}(\mathbf{A}_{\mathcal{S}_c})} < 1.$$

Now, using Lemma 10 and the following bound on $\|\mathbf{b}_{\mathcal{S}_c}\|_2$,

Lemma 11. $\|\mathbf{b}_{\mathcal{S}_c}\|_2$ *is upper bounded by* $\lambda_c\sqrt{s_c} + \sqrt{r\alpha_u}\mu$*.*

we have that the condition **(C3)** holds if

$$\|\mathcal{P}_{\mathcal{L}^\perp}(\mathbf{\Gamma})\|_2 \le \tfrac{\lambda_c\sqrt{s_c}+\sqrt{r\alpha_u}\mu}{\sqrt{\alpha_\ell}(1-\mu)} < 1,$$

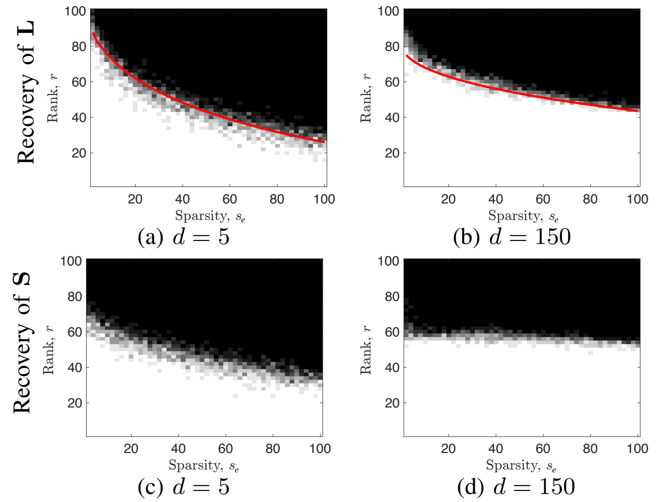


Fig. 1: Recovery for varying rank of $\mathbf{L}$, sparsity of $\mathbf{S}$ and number of dictionary elements in $\mathbf{D}$ as per Theorem 1. Each plot shows average recovery across 10 trials for varying ranks and sparsity up to $s_e^{\max} = m$, where $n = m = 100$ and the white region represents correct recovery. We decide success if $\|\mathbf{L} - \hat{\mathbf{L}}\|_{\mathrm{F}}/\|\mathbf{L}\|_{\mathrm{F}} \le 0.02$ and $\|\mathbf{S} - \hat{\mathbf{S}}\|_{\mathrm{F}}/\|\mathbf{S}\|_{\mathrm{F}} \le 0.02$, where $\hat{\mathbf{L}}$ and $\hat{\mathbf{S}}$ are the recovered $\mathbf{L}$ and $\mathbf{S}$, respectively. Panels (a)-(b) show the recovery of the low-rank part $\mathbf{L}$ and (c)-(d) show the recovery of the sparse part with varying dictionary sizes $d = 5$ and 150, respectively. Also, the predicted trend between rank r and sparsity s_e as per Theorem 1, eq.(9) is shown in red in panels (a-b).

which is satisfied by our choice of $\lambda_c^{\max}$ (10). Now, for the condition **(C4)** we need the following condition to hold true:

$$\begin{aligned}&\|\mathcal{P}_{\mathcal{S}_c^\perp}(\mathbf{D}^\top\mathbf{\Gamma})\|_{\infty,2}\\ &\le \|\mathcal{P}_{\mathcal{S}_c^\perp}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)\|_{\infty,2} + \|\mathcal{P}_{\mathcal{S}_c^\perp}(\mathbf{D}^\top(\mathbf{I}-\mathbf{P_U})\mathbf{X}(\mathbf{I}-\mathbf{P_V}))\|_{\infty,2}\\ &= \|\mathcal{P}_{\mathcal{S}_c^\perp}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)\|_{\infty,2} + \|\mathcal{P}_{\mathcal{S}_c^\perp}(\mathbf{Z})\|_{\infty,2} < \lambda_c.\end{aligned}$$

Note that, here $\|\mathcal{P}_{\mathcal{S}_c}(\mathbf{D}^T\mathbf{U}\mathbf{V}^T)\|_{\infty,2} \le \xi_c$. Therefore, using the following result,

Lemma 12. *An upper bound on* $\|\mathcal{P}_{\mathcal{S}_c^\perp}(\mathbf{Z})\|_{\infty,2}$ *is given by* $(\lambda_c s_c + \sqrt{r\alpha_u s_c}\mu)C_c$*.*

the condition **(C4)** implies that,

$$\xi_c + \tfrac{\alpha_u}{\alpha_\ell(1-\mu)^2}\sqrt{s_c}\gamma_{\mathbf{V}}\beta_{\mathbf{U}}(\lambda_c\sqrt{s_c} + \sqrt{r\alpha_u}\mu) < \lambda_c.$$

To this end, if we let $C_c := \frac{\alpha_u}{\alpha_\ell(1-\mu)^2}\gamma_{\mathbf{V}}\beta_{\mathbf{U}}$, **(C4)** is satisfied by $\lambda_c^{\min}$ defined in (10). This completes the proof. ■

Characterizing $\lambda_c^{\min}$**:** From (10), we need $\lambda_c^{\min} := \frac{\xi_c + \sqrt{r s_c \alpha_u}\mu C_c}{1 - s_c C_c} \ge 0$, where $C_c := \frac{\alpha_u}{\alpha_\ell(1-\mu)^2}\gamma_{\mathbf{V}}\beta_{\mathbf{U}} \ge 0$. Then from $s_c C_c < 1$, we require $s_c < s_c^{\max} := \frac{\alpha_\ell(1-\mu)^2}{\alpha_u\gamma_{\mathbf{V}}\beta_{\mathbf{U}}}$.

Characterizing $\lambda_c^{\max}$**:** Since we need $\lambda_c^{\min} < \lambda_c^{\max}$, substituting the expressions for $\lambda_c^{\min}$ and $\lambda_c^{\max}$, and using the fact that $s_c C_c < 1$, we arrive at (11).

V. Numerical Simulations

In this section, we empirically evaluate the properties of D-RPCA(E) and D-RPCA(C) via phase transition in rank and sparsity, and compare its performance to related techniques, and to the behavior predicted by Theorem 1 and Theorem 3 in (9) and (11), respectively.

A. Entry-Wise Sparsity Case

Experimental Set-up: We employ the accelerated proximal gradient (APG) algorithm outlined in Algorithm 1 of [14]

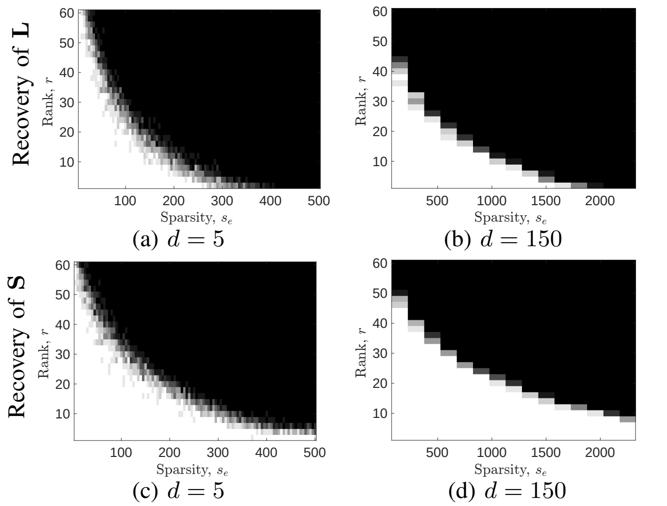


Fig. 2: Recovery for varying rank of $\mathbf{L}$, sparsity of $\mathbf{S}$ and number of dictionary elements in $\mathbf{R}$. Panels (a)-(b) show the recovery of the low-rank part $\mathbf{L}$ and (c)-(d) show the recovery of the sparse part with varying dictionary sizes $d = 5$ and 150, respectively. The experimental set-up and the success metric remains the same as in Fig. 1.

and Algorithm 1 of [16] to solve the optimization problem D-RPCA(E). For these evaluations, we fix $n = m = 100$, and generate the low-rank part $\mathbf{L}$ by outer product of two column normalized random matrices of sizes $n \times r$ and $m \times r$, with entries drawn from the standard normal distribution. In addition, we draw s_e non-zero entries of the sparse component $\mathbf{S}$ from the Rademacher distribution, and the dictionary $\mathbf{D}$ from the standard normal distribution with normalized columns. We then run 10 Monte-Carlo trials for each pair of rank and sparsity, and for each of these, we scan across 100 values of λ_es in the range of $[\lambda_e^{\min}, \lambda_e^{\max}]$ to find the best pair of $(\mathbf{L}_0, \mathbf{S}_0)$ to compile the results. [2]

Discussion: Phase transition in rank and sparsity averaged over 10 trials for dictionaries of sizes $d = 5$ (thin) and $d = 150$ (fat), are shown in Fig. 1 and Fig. 2, respectively. We note from Fig. 1 that indeed the empirical relationship between rank and sparsity for the recovery of $(\mathbf{L_0}, \mathbf{S_0})$ has the same trend as predicted by

$$r < \left(\sqrt{\tfrac{\alpha_\ell}{\alpha_u} \tfrac{1-\mu}{\mu}} - \tfrac{\xi_e}{\sqrt{\alpha_u}\mu} \tfrac{1+C_e}{1-C_e} \sqrt{s_e} \right)^2,$$

as shown in (9) in Section III for $s_e \leq s_e^{\max}$. Here, the parameters corresponding to the predicted trend (shown in red) have been hand-tuned for best fit.

In fact, as shown in Fig. 2, this trend continues for sparsity levels much greater than $s_e^{\max}$. This can be potentially attributed to the limitations of the deterministic analysis (on the locations of the non-zero elements of $\mathbf{S}$) presented here.

Further, Fig. 3 shows the results of RPCA† (in green, shows the area where at least one of the 10 Monte-Carlo simulations succeeds) in comparision to the results obtained by D-RPCA(E) for $d = 5$ and $d = 50$. We observe that D-RPCA(E) outperforms RPCA† across the board. In fact, we notice that the RPCA† technique only succeeds when $r < d$. We believe that this is because when $d < r$ the component $\mathbf{D}^\dagger\mathbf{L}$ is not low-rank (full-rank in this case) w.r.t. the maximum potential rank of $\mathbf{D}^\dagger\mathbf{L}$. As a result, the model

[2]For ease of computation we run on modest values of n and m.

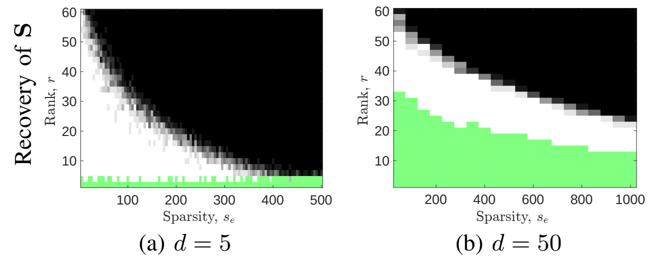


Fig. 3: Comparision of phase transitions in rank and sparsity between D-RPCA(E) and RPCA† for recovery of $\mathbf{S}$ for different dictionary sizes. Panels (a) and (b) correspond to $d = 5$ and $d = 50$, respectively. Experimental set-up and the success metric remains same as Fig. 2. The area in green corresponds to recovery by RPCA† where at least 1 out of the 10 Monte-Carlo trials succeeds.

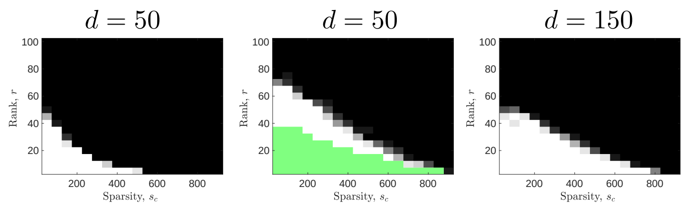
assumptions of the robust PCA problem do not apply; see Section I-B. In contrast, the proposed framework of D-RPCA(E) can handle these cases effectively (see Fig. 3) since $\mathbf{L}$ is low-rank irrespective of the dictionary size. This phenomenon highlights the applicability of the proposed approach to cases where $d < r$, and simultaneous recovery of the low-rank component in one-shot.

B. Column-wise Sparsity Case

We now present phase transition in rank r and number of outlierss_c to evaluate the performance of D-RPCA(C). In particular, we compare with Outlier Pursuit (OP) [21] that solves D-RPCA(C) with $\mathbf{D} = \mathbf{I}$, and OP† to demonstrate that the *a priori* knowledge of the dictionary provides superior recovery properties.

Experimental Set-up: Again, we employ a variant of the APG algorithm outlined in Algorithm 1 of [14] specialized for the column-wise sparsity case to solve the optimization problem D-RPCA(C); see Algorithm 1 of [16]. We set $n = 100$, $m = 1000$, and for each pair of r and s_c we run 10 Monte-Carlo trials for $r \in \{5, 10, 15 \ldots, 100\}$ and $s_c \in \{50, 100, 150, \ldots, 900\}$. For our experiments, we form $\mathbf{L} = [\mathbf{U}\mathbf{V}^\top \mathbf{0}_{n \times s_c}] \in \mathbb{R}^{n \times m}$, where $\mathbf{U} \in \mathbb{R}^{n \times r}$, $\mathbf{V} \in \mathbb{R}^{(m-s_c) \times r}$ have i.i.d. $\mathcal{N}(0, 1)$ entries, which are then column normalized. Next, we generate $\mathbf{S} = [\mathbf{0}_{d \times (m-s_c)} \mathbf{W}] \in \mathbb{R}^{d \times m}$ where each entry of $\mathbf{W} \in \mathbb{R}^{d \times s_c}$ is i.i.d. $\mathcal{N}(0, 1)$. Also, the known dictionary $\mathbf{D} \in \mathbb{R}^{n \times d}$ is formed by normalizing the columns of a random matrix with i.i.d. $\mathcal{N}(0, 1)$ entries. For each method, we scan through 100 values of the regularization parameter $\lambda_c \in [\lambda_c^{\min}, \lambda_c^{\max}]$ to find a solution pair $(\mathbf{L}_0, \mathbf{S}_0)$ with the best precision, i.e., best detection of the outliers and rejection of false positives. We declare an experiment successful if it acheives a precision i.e. $(\text{True Positives}/(\text{True Positives} + \text{False Positives}))$ of 0.99 or higher. Here, we threshold the column norms at 2×10^{-3} before we evaluate the precision.

Discussion: Fig. 4 (a)-(c) shows the phase transition in rank r and column-sparsity s_c for the outlier identification performance (in terms of precison) of OP for $d = 50$, D-RPCA(C) for $d = 50$ (and OP† in green, marking the region where precision is greater than 0), and D-RPCA(C) for $d = 150$, respectively. We observe that the *a priori* knowledge of the dictionary $\mathbf{D}$ significantly boosts the performance of D-RPCA(C) as compared to OP. This showcases the superior outlier identification properties of the proposed technique D-RPCA(C). Further, similar to the entry-wise case, we note



(a) OP (b) D-RPCA(C); Green denotes OP† (c) D-RPCA(C)

Fig. 4: Phase transitions in rank r and column sparsity s_c across 10 Monte-Carlo simulations. Panels (a), (b), (c) show the precision i.e., (True Positives/(True Positives + False Positives)) in identifying the outlier columns of $\mathbf{S}$ for $d = 50$ using (a) OP and (b) OP†, and D-RPCA(C) for $d = 150$, respectively. In addition, panel (b) also shows the performance by OP† for $d = 50$ in green, marking the region where precision is greater than 0, super imposed over D-RPCA(C). Here, we threshold the column norms of the recovered matrix $\mathbf{S}$ at 2×10^{-3} before computing the precision, and a trial is declared successful if it achieves a precision of 0.99 or higher.

that the pseudo-inversed based technique OP† (in green) fails when $r > d$. For the $d = 150$ case the proposed technique D-RPCA(C) is able to identify the outlier columns with high precision. This is interesting since our technique succeeds even when the outlier columns are not themselves sparse (we draw the entries of the outlier columns from $\mathcal{N}(0, 1)$). This corroborates our theoretical assumption that $\mathcal{R}$ needs to be structured such that GFP is met.

Our empirical evaluations paves way to potential improvements via a probabilistic analysis of the model instead of the case considered here. Additionally, the recent results on non-convex low-rank matrix estimation formulations [33], [34] may potentially lead to computationally efficient algorithms by replacing the expensive SVD step in every iteration. Exploration of these extensions are left for future work.

VI. Conclusions and Future Work

We analyze a dictionary based generalization of robust PCA. Here, we model the acquired data as a superposition of a low-rank component and a dictionary sparse component, considering two distinct sparsity patterns – entry-wise sparsity and column-wise sparsity, respectively. Specifically, for the entry-wise sparsity case, we extend the theoretical guarantees presented in [14] to a setting wherein the dictionary $\mathbf{D}$ may have arbitrary number of columns, and the coefficient matrix $\mathbf{S}$ has *global* sparsity of s_e, i.e. $\|\mathbf{S}\|_0 = s_e \leq s_e^{\max}$, rendering the results useful for a potentially wide range of applications. Further, we propose a column-wise sparsity model, which can be viewed as a dictionary based generalization of Outlier Pursuit [21]. For this case, we analyze and develop the conditions under which solving a convex program will recover the correct column-space of the low-rank part while identifying the outlier columns in the dictionary sparse part. To corroborate our theoretical results, we provide empirical evaluations via phase transition plots in rank and sparsity.

Appendix

In the following appendices, we provide the proofs of the lemmata employed to establish our main results. We also summarize the notation in Table I.

TABLE I: Summary of important notation and parameters

Matrices	
$\mathbf{M} \in \mathbb{R}^{n \times m}$	The data matrix
$\mathbf{L} \in \mathbb{R}^{n \times m}$	The low-rank matrix with rank-r and singular value decomposition $\mathbf{L} = \mathbf{U\Sigma V}^\top$
$\mathbf{D} \in \mathbb{R}^{n \times d}$	The known dictionary either *thin* ($d \leq n$) or *fat* ($d > n$)
$\mathbf{S} \in \mathbb{R}^{d \times m}$	The sparse component with the following properties –(1) in case of entry-wise sparsity: s_e non-zero entries and when $d > n$ has at most k non-zeros per column, and (2) in case of column-wise sparsity: s_c non-zero columns
Regularization Parameters	
$\lambda_e \in \mathbb{R}$	The regularization parameter for the entry-wise sparsity case
$\lambda_c \in \mathbb{R}$	The regularization parameter for the column sparsity case
Subspaces	
$\mathcal{L}$	The set of matrices which span the same column or row space as $\mathbf{L}$, i.e., $\mathcal{L} := \{\mathbf{U}\mathbf{W}_1^\top + \mathbf{W}_2\mathbf{V}^\top, \mathbf{W}_1 \in \mathbb{R}^{m \times r}, \mathbf{W}_2 \in \mathbb{R}^{n \times r}\}$.
$\mathcal{S}_e$	The set of matrices with the same support as $\mathbf{S}$ (for the entry-wise sparse case).
$\mathcal{S}_c$	The set of matrices with the same column support as $\mathbf{S}$ (for the column-wise sparse case).
$\mathcal{D}$	The set of matrices whose columns span the subspace spanned by columns of $\mathbf{D}$, i.e. $\mathcal{D} := \{\mathbf{Z} = \mathbf{RH}, \mathbf{H} \in \mathcal{S}_e \text{ or } \mathbf{H} \in \mathcal{S}_c\}$
$\mathcal{U}$	The column space of $\mathbf{L}$
$\mathcal{V}$	The row space of $\mathbf{L}$
Index Sets	
$\mathcal{I}_{\mathcal{S}_e}$	Support of matrix $\mathbf{S}$ (entry-wise case)
$\mathcal{I}_{\mathcal{S}_c}$	Column support of matrix $\mathbf{S}$ (the outliers)
$\mathcal{I}_{\mathbf{L}}$	Index set of the inliers (column-wise case)
Projection	
$\mathcal{P}_{\mathcal{G}}(\cdot)$	Projection operator corresponding to any subspace $\mathcal{G}$
$\mathbf{P}_{\mathbf{G}}$	Projection matrix corresponding to the operator $\mathcal{P}_{\mathcal{G}}(\cdot)$
Parameters for analysis	
μ	The incoherence parameter between the low-rank component and the dictionary, defined as $\mu := \max\limits_{\mathbf{Z} \in \mathcal{D} \setminus \{\mathbf{0}_{d \times m}\}} \frac{\|\mathcal{P}_{\mathcal{L}}(\mathbf{Z})\|_{\mathrm{F}}}{\|\mathbf{Z}\|_{\mathrm{F}}}$
$\gamma_{\mathbf{V}}$	Defined as $\gamma_{\mathbf{V}} := \max\limits_i \|\mathbf{P_V}\mathbf{e}_i\|^2$
$\gamma_{\mathbf{U}}$	Defined as $\gamma_{\mathbf{U}} := \max\limits_i \frac{\|\mathbf{P_U}\mathbf{De}_i\|^2}{\|\mathbf{De}_i\|^2}$
$\beta_{\mathbf{U}}$	Defined as $\beta_{\mathbf{U}} := \max\limits_{\|\mathbf{u}\|=1} \frac{\|(\mathbf{I} - \mathbf{P_U})\mathbf{Du}\|^2}{\|\mathbf{Du}\|^2}$
ξ_e	Defined as $\xi_e := \|\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top\|_\infty$
ξ_c	Defined as $\xi_c := \|\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top\|_{\infty,2}$
α_ℓ	Lower generalized frame bound
α_u	Upper generalized frame bound

A. Proofs for Entry-wise Case

We present the details of the proofs in this section for the entry-wise case. We first start by deriving the optimality conditions.

Proof of Lemma 4. Let $\{\mathbf{L}_0, \mathbf{S}_0\}$ be a solution of the problem posed above. Notice that this pair is not necessarily unique. For example, as shown in proof of Lemma 2 in [14], $\{\mathbf{L}_0 + \mathbf{DH}, \mathbf{S}_0 - \mathbf{H}\}$, with arbitrary $\mathbf{H}$, is another feasible solution of the problem satisfying the optimality conditions (derived in this section).

We begin by writing the Lagrangian, $\mathcal{F}(\mathbf{L}, \mathbf{S}, \mathbf{\Lambda})$, for the given problem as follows.

$$\mathcal{F}(\mathbf{L}, \mathbf{S}, \mathbf{\Lambda}) = \|\mathbf{L}\|_* + \lambda_e\|\mathbf{S}\|_1 \ + \langle\mathbf{\Lambda},\ \mathbf{M} - \mathbf{L} - \mathbf{DS}\rangle,$$

where $\mathbf{\Lambda} \in \mathbb{R}^{n \times m}$ are the Lagrange multipliers.

Let the singular value decomposition (SVD) of $\mathbf{L}_0$ be represented as $\mathbf{U\Sigma V}^\top$. Then the sub-differential set of $\|\mathbf{L}\|_*$ can be represented as

$$\partial_{\mathbf{L}}\|\mathbf{L}\|_*\Big|_{\mathbf{L}=\mathbf{L}_0} = \{\mathbf{U}\mathbf{V}^\top + \mathbf{W} : \|\mathbf{W}\| \leq 1, \mathcal{P}_{\mathcal{L}}(\mathbf{W}) = \mathbf{0}\},$$

as shown in [35]. Also, the subdifferential set corresponding to $\|\mathbf{S}\|_1$ is given by

$$\partial_{\mathbf{S}}\|\mathbf{S}\|_1\Big|_{\mathbf{S}=\mathbf{S}_0} = \{\text{sign}(\mathbf{S}_0) + \mathbf{F} : \|\mathbf{F}\|_\infty \leq 1, \mathcal{P}_{\mathcal{S}_e}(\mathbf{F}) = \mathbf{0}\},$$

Using these results, we write the sub-differential of the Lagrangian with respect to $\mathbf{L}$ and $\mathbf{S}$ at $\{\mathbf{L}_0, \mathbf{S}_0\}$ as

$$\partial_{\mathbf{L}}\mathcal{F}(\mathbf{L}_0,\mathbf{S}_0,\mathbf{\Lambda}) = \{\mathbf{U}\mathbf{V}^\top + \mathbf{W} - \mathbf{\Lambda} : \|\mathbf{W}\| \leq 1, \mathcal{P}_{\mathcal{L}}(\mathbf{W}) = \mathbf{0}\},$$
$$\partial_{\mathbf{S}}\mathcal{F}(\mathbf{L}_0,\mathbf{S}_0,\mathbf{\Lambda}) = \{\lambda_{\mathbf{e}}\text{sign}(\mathbf{S}_0) + \lambda_{\mathbf{e}}\mathbf{F} - \mathbf{D}^\top\mathbf{\Lambda}, \|\mathbf{F}\|_\infty \leq 1, \mathcal{P}_{\mathcal{S}_e}(\mathbf{F}) = \mathbf{0}\}.$$

Then optimality conditions are

$$\mathbf{0}_{n\times m} \in \partial_{\mathbf{L}}\mathcal{F}(\mathbf{L}_0,\mathbf{S}_0,\mathbf{\Lambda}) \text{ and } \mathbf{0}_{d\times m} \in \partial_{\mathbf{S}}\mathcal{F}(\mathbf{L}_0,\mathbf{S}_0,\mathbf{\Lambda}),$$

which implies that the dual solution $\mathbf{\Lambda}$ must obey the following,

$$\mathbf{\Lambda} \in \mathbf{U}\mathbf{V}^\top + \mathbf{W},\ \|\mathbf{W}\| \leq 1,\ \mathcal{P}_{\mathcal{L}}(\mathbf{W}) = \mathbf{0}_{n\times m} \text{ and}$$
$$\mathbf{D}^\top\mathbf{\Lambda} \in \lambda_{\mathbf{e}}\text{sign}(\mathbf{S}_0) + \lambda_{\mathbf{e}}\mathbf{F},\ \|\mathbf{F}\|_\infty \leq 1,\ \mathcal{P}_{\mathcal{S}_e}(\mathbf{F}) = \mathbf{0}_{d\times m}.$$

Our aim here is to find the conditions on $\mathbf{W}$ and $\mathbf{F}$ such that the pair $\{\mathbf{L}_0,\ \mathbf{S}_0\}$ is a unique solution to the problem at hand.

Using these conditions, we see that $\mathcal{P}_{\mathcal{L}}(\mathbf{\Lambda}) = \mathbf{U}\mathbf{V}^\top$ and $\mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top\mathbf{\Lambda}) = \lambda_e\text{sign}(\mathbf{S}_0)$; these correspond to conditions **(C1)** and **(C2)**, respectively. Now consider a feasible solution $\{\mathbf{L_0}+\mathbf{D}\mathbf{H}, \mathbf{S_0}-\mathbf{H}\}$ for a non-zero $\mathbf{H} \in \mathbb{R}^{d\times m}$. Let $\mathbf{W}$, with $\|\mathbf{W}\| = 1$ and $\mathcal{P}_{\mathcal{L}}(\mathbf{W}) = \mathbf{0}$, then by duality of norms,

$$\langle\mathbf{W},\ \mathbf{D}\mathbf{H}\rangle = \langle\mathbf{W},\ \mathcal{P}_{\mathcal{L}^\perp}(\mathbf{D}\mathbf{H})\rangle = \|\mathcal{P}_{\mathcal{L}^\perp}(\mathbf{D}\mathbf{H})\|_*.$$

Further, let $\mathbf{F}$, with $\|\mathbf{F}\|_\infty = 1$ and $\mathcal{P}_{\mathcal{S}_e}(\mathbf{F}) = \mathbf{0}$, be such that

$$\mathbf{F}_{ij} = \begin{cases} -\text{sign}(\mathbf{H}_{ij}) & , \text{ if } \{i,\ j\} \notin \mathcal{S}_e \text{ and } \mathbf{H}_{ij} \neq 0 \\ 0 & , \text{ otherwise} \end{cases},$$

where $\mathbf{F}_{ij}$ denotes the $(i,j)^{\text{th}}$ element of $\mathbf{F}$. Then, we arrive at the following simplification for $\langle\mathbf{F},\ \mathbf{H}\rangle$ by duality of norms,

$$\langle\mathbf{F},\ \mathbf{H}\rangle = \langle\mathbf{F},\ \mathcal{P}_{\mathcal{S}_e^\perp}(\mathbf{H})\rangle = -\|\mathcal{P}_{\mathcal{S}_e^\perp}(\mathbf{H})\|_1.$$

We first write the sub-gradient optimality condition,

$$\|\mathbf{L}_0 + \mathbf{D}\mathbf{H}\|_* + \lambda_e\ \|\mathbf{S_0} - \mathbf{H}\|_1 \geq\ \|\mathbf{L}_0\|_* + \lambda_e\ \|\mathbf{S}_0\|_1 + \langle\mathbf{U}\mathbf{V}^\top + \mathbf{W},\ \mathbf{D}\mathbf{H}\rangle - \langle\lambda_e\text{sign}(\mathbf{S}_0) + \lambda_e\mathbf{F},\ \mathbf{H}\rangle. \quad (28)$$

Next, we use the relationships derived above to simplify the following term:

$$\begin{aligned}&\langle\mathbf{U}\mathbf{V}^\top + \mathbf{W},\ \mathbf{D}\mathbf{H}\rangle - \langle\lambda_e\text{sign}(\mathbf{S}_0) + \lambda_e\mathbf{F},\ \mathbf{H}\rangle,\\ &= \langle\mathbf{W},\mathbf{D}\mathbf{H}\rangle - \lambda_e\langle\mathbf{F},\mathbf{H}\rangle + \langle\mathcal{P}_{\mathcal{L}}(\mathbf{\Lambda}),\mathbf{D}\mathbf{H}\rangle - \langle\mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top\mathbf{\Lambda}),\mathbf{H}\rangle\\ &= \|\mathcal{P}_{\mathcal{L}^\perp}(\mathbf{D}\mathbf{H})\|_* + \lambda_e\|\mathcal{P}_{\mathcal{S}_e^\perp}(\mathbf{H})\|_1 + \langle\mathcal{P}_{\mathcal{L}}(\mathbf{\Lambda}),\mathbf{D}\mathbf{H}\rangle - \langle\mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top\mathbf{\Lambda}),\mathbf{H}\rangle\end{aligned}$$

We now simplify $\langle\mathcal{P}_{\mathcal{L}}(\mathbf{\Lambda}),\ \mathbf{D}\mathbf{H}\rangle - \langle\mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top\mathbf{\Lambda}),\ \mathbf{H}\rangle$ using Holder's inequality.

$$\begin{aligned}&\langle\mathcal{P}_{\mathcal{L}}(\mathbf{\Lambda}),\mathbf{D}\mathbf{H}\rangle - \langle\mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top\mathbf{\Lambda}),\mathbf{H}\rangle\\ &= \langle\mathbf{\Lambda} - \mathcal{P}_{\mathcal{L}^\perp}(\mathbf{\Lambda}),\ \mathbf{D}\mathbf{H}\rangle - \langle\mathbf{D}^\top\mathbf{\Lambda} - \mathcal{P}_{\mathcal{S}_e^\perp}(\mathbf{D}^\top\mathbf{\Lambda}),\mathbf{H}\rangle\\ &\geq -\|\mathcal{P}_{\mathcal{L}^\perp}(\mathbf{D}\mathbf{H})\|_*\|\mathcal{P}_{\mathcal{L}^\perp}(\mathbf{\Lambda})\| - \|\mathcal{P}_{\mathcal{S}_e^\perp}(\mathbf{D}^\top\mathbf{\Lambda})\|_\infty\|\mathcal{P}_{\mathcal{S}_e^\perp}(\mathbf{H})\|_1\end{aligned}$$

Finally, we simplify the optimality condition in shown in (28),

$$\begin{aligned}&\|\mathbf{L}_0 + \mathbf{D}\mathbf{H}\|_* + \lambda_e\ \|\mathbf{S_0} - \mathbf{H}\|_1\\ &\geq\ \|\mathbf{L}_0\|_* + \lambda_e\ \|\mathbf{S}_0\|_1 + (1 - \|\mathcal{P}_{\mathcal{L}^\perp}(\mathbf{\Lambda})\|)\|\mathcal{P}_{\mathcal{L}^\perp}(\mathbf{D}\mathbf{H})\|_*\\ &\quad + (\lambda_e - \|\mathcal{P}_{\mathcal{S}_e^\perp}(\mathbf{D}^\top\mathbf{\Lambda})\|_\infty)\|\mathcal{P}_{\mathcal{S}_e^\perp}(\mathbf{H})\|_1.\end{aligned}$$

Here, we note that if $\|\mathcal{P}_{\mathcal{L}^\perp}(\mathbf{\Lambda})\| < 1$ and $\|\mathcal{P}_{\mathcal{S}_e^\perp}(\mathbf{D}^\top\mathbf{\Lambda})\|_\infty < \lambda_e$, then the pair $\{\mathbf{L}_0, \mathbf{S}_0\}$ is the unique solution of the problem. Consequently, these are the required necessary conditions **(C3)** and **(C4)**, respectively. ■

Proof of Lemma 5. First, note that we need $\mathbf{A}_{\mathcal{S}_e}$ to have full row rank, i.e, its smallest singular value should be greater than zero. To this end, we first derive a lower bound on the smallest singular value, $\sigma_{\min}(\mathbf{A}_{\mathcal{S}_e})$ of $\mathbf{A}_{\mathcal{S}_e}$ as follows:

$$\sigma_{\min}(\mathbf{A}_{\mathcal{S}_e}) = \min_{\mathbf{H}\in\mathcal{S}_e\setminus\{\mathbf{0}\}} \tfrac{\|\mathbf{A}^\top\text{vec}(\mathbf{H})\|}{\|\text{vec}(\mathbf{H})\|}.$$

Now, using the definition of $\mathbf{A}^\top$ and properties of Kronecker products namely, transpose and vectorization of product of three matrices, we have

$$\sigma_{\min}(\mathbf{A}_{\mathcal{S}_e}) = \min_{\mathbf{H}\in\mathcal{S}_e\setminus\{\mathbf{0}\}} \tfrac{\|(\mathbf{I}-\mathbf{P_U})\mathbf{D}\mathbf{H}(\mathbf{I}-\mathbf{P_V})\|_{\text{F}}}{\|\mathbf{H}\|_{\text{F}}}.$$

Now, since $(\mathbf{I} - \mathbf{P_U})\mathbf{D}\mathbf{H}(\mathbf{I} - \mathbf{P_V}) = \mathcal{P}_{\mathcal{L}^\perp}(\mathbf{D}\mathbf{H})$,

$$\sigma_{\min}(\mathbf{A}_{\mathcal{S}_e}) = \min_{\mathbf{H}\in\mathcal{S}_e\setminus\{\mathbf{0}\}} \tfrac{\|\mathcal{P}_{\mathcal{L}^\perp}(\mathbf{D}\mathbf{H})\|_{\text{F}}}{\|\mathbf{D}\mathbf{H}\|_{\text{F}}}\tfrac{\|\mathbf{D}\mathbf{H}\|_{\text{F}}}{\|\mathbf{H}\|_{\text{F}}}.$$

Using the GFP, we have the following lower bound:

$$\sigma_{\min}(\mathbf{A}_{\mathcal{S}_e}) \geq \sqrt{\alpha_\ell}\ \min_{\mathbf{Z}\in\mathcal{D}\setminus\{\mathbf{0}\}} \tfrac{\|\mathcal{P}_{\mathcal{L}^\perp}(\mathbf{Z})\|_{\text{F}}}{\|\mathbf{Z}\|_{\text{F}}}.$$

Further, simplifying using properties of the projection operator, the reverse triangle inequality and the definition of μ,

$$\begin{aligned}\sigma_{\min}(\mathbf{A}_{\mathcal{S}_e}) &= \sqrt{\alpha_\ell}\ \min_{\mathbf{Z}\in\mathcal{D}\setminus\{\mathbf{0}\}} \tfrac{\|\mathbf{Z}-\mathcal{P}_{\mathcal{L}}(\mathbf{Z})\|_{\text{F}}}{\|\mathbf{Z}\|_{\text{F}}}\\ &\geq \sqrt{\alpha_\ell}\ \big(1 - \max_{\mathbf{Z}\in\mathcal{D}\setminus\{\mathbf{0}\}} \tfrac{\|\mathcal{P}_{\mathcal{L}}(\mathbf{Z})\|_{\text{F}}}{\|\mathbf{Z}\|_{\text{F}}}\big) = \sqrt{\alpha_\ell}(1-\mu).\end{aligned}$$

Therefore, we note that if $\mu < 1$ and $\alpha_\ell > 0$, $\mathbf{A}_{\mathcal{S}_e}$ has full row rank, and the lower bound on the smallest singular value is given by $\sqrt{\alpha_\ell}(1-\mu)$. ■

Proof of Lemma 6. We begin with the definition of $\mathbf{b}_{\mathcal{S}_e}$. Since $\|\mathbf{b}_{\mathcal{S}_e}\|_2 = \|\mathbf{B}_{\mathcal{S}_e}\|_{\text{F}}$ and $\mathbf{B}_{\mathcal{S}_\mathbf{e}} := \lambda_e\text{sign}(\mathbf{S}_0) - \mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)$,

$$\begin{aligned}\|\mathbf{b}_{\mathcal{S}_e}\|_2 &= \|\lambda_e\text{sign}(\mathbf{S_0}) - \mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)\|_{\text{F}},\\ &\leq \lambda_e\sqrt{s_e} + \|\mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)\|_{\text{F}}.\end{aligned}$$

Now for an upper bound on $\|\mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)\|_{\text{F}}$ we start by analyzing $\|\mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)\|_{\text{F}}^2$,

$$\|\mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)\|_{\text{F}}^2 = |\langle\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top,\ \mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)\rangle|.$$

Using properties of the inner products and using the fact that $\mathcal{P}_{\mathcal{L}}(\mathbf{U}\mathbf{V}^\top) = \mathbf{U}\mathbf{V}^\top$,

$$\|\mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)\|_{\text{F}}^2 = |\langle\mathcal{P}_{\mathcal{L}}(\mathbf{U}\mathbf{V}^\top),\ \mathbf{D}\mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)\rangle|.$$

Further simplifying using Cauchy Schwarz inequality and the definition of μ we have

$$\begin{aligned}&\|\mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)\|_{\text{F}}^2 \leq \|\mathcal{P}_{\mathcal{L}}(\mathbf{U}\mathbf{V}^\top)\|_{\text{F}}\|\mathcal{P}_{\mathcal{L}}(\mathbf{D}\mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top))\|_{\text{F}}\\ &\leq \mu\|\mathbf{U}\mathbf{V}^\top\|_{\text{F}}\|\mathbf{D}\mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)\|_{\text{F}}\end{aligned}$$

Now, since $\|\mathbf{U}\mathbf{V}^\top\|_{\text{F}} = \sqrt{r}$ and using the GFP we have $\|\mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)\|_{\text{F}} \leq \mu\sqrt{r\alpha_u}$. Therefore, an upper bound for $\|\mathbf{b}_{\mathcal{S}_e}\|_2$ is given by $\|\mathbf{b}_{\mathcal{S}_e}\|_2 \leq \lambda_e\sqrt{s_e} + \sqrt{r\alpha_u}\mu$. ■

Proof of Lemma 7. Since $\|\mathbf{b}_{\mathcal{S}_e}\|_\infty = \|\mathbf{B}_{\mathcal{S}_e}\|_\infty$ and $\mathbf{B}_{\mathcal{S}_\mathbf{e}} := \lambda_e\text{sign}(\mathbf{S}_0) - \mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)$, we have the upper bound $\|\mathbf{b}_{\mathcal{S}_e}\|_\infty \leq \lambda_e + \|\mathcal{P}_{\mathcal{S}_e}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)\|_\infty$. ■

Proof of Lemma 8. We begin by simplifying the quantity of interest as follows:

$$\begin{aligned}\|\mathbf{Q}\|_{\infty,\infty} &= \|\mathbf{A}_{\mathcal{S}_e{}^\perp}\mathbf{A}_{\mathcal{S}_e}^\top(\mathbf{A}_{\mathcal{S}_e}\mathbf{A}_{\mathcal{S}_e}^\top)^{-1}\|_{\infty,\infty}\\ &\leq \|\mathbf{A}_{\mathcal{S}_e{}^\perp}\mathbf{A}_{\mathcal{S}_e}^\top\|_{\infty,\infty}\|(\mathbf{I} - (\mathbf{I} - \mathbf{A}_{\mathcal{S}_e}\mathbf{A}_{\mathcal{S}_e}^\top))^{-1}\|_{\infty,\infty}\\ &\leq \tfrac{\|\mathbf{A}_{\mathcal{S}_e{}^\perp}\mathbf{A}_{\mathcal{S}_e}^\top\|_{\infty,\infty}}{1-\|\mathbf{I}-\mathbf{A}_{\mathcal{S}_e}\mathbf{A}_{\mathcal{S}_e}^\top\|_{\infty,\infty}}. \quad (29)\end{aligned}$$

Now, we derive appropriate bounds on the numerator and the denominator of (29) separately. Consider the numerator $\|\mathbf{A}_{\mathcal{S}_e^\perp}\mathbf{A}_{\mathcal{S}_e}^\top\|_{\infty,\infty}$. Here, we are interested in the maximum ℓ_1-norm of the rows of $\mathbf{A}_{\mathcal{S}_e^\perp}\mathbf{A}_{\mathcal{S}_e}^\top$, i.e.,

$$\|\mathbf{A}_{\mathcal{S}_e^\perp}\mathbf{A}_{\mathcal{S}_e}^\top\|_{\infty,\infty} = \max_i \|\mathbf{e}_i^\top \mathbf{A}_{\mathcal{S}_e^\perp}\mathbf{A}_{\mathcal{S}_e}^\top\|_1.$$

Let $\mathcal{I}_{\mathcal{S}_e}$ refer to the support of $\mathbf{S}_0$, and $\bar{\mathcal{I}}_{\mathcal{S}_e}$ to its complement. Then, the expression can be written in terms of $\mathcal{I}_{\mathcal{S}_e}$ and $\bar{\mathcal{I}}_{\mathcal{S}_e}$:

$$\|\mathbf{A}_{\mathcal{S}_e^\perp}\mathbf{A}_{\mathcal{S}_e}^\top\|_{\infty,\infty} = \max_{j\in\bar{\mathcal{I}}_{\mathcal{S}_e}} \sum_{\ell\in\mathcal{I}_{\mathcal{S}_e}} |\mathbf{e}_l^\top \mathbf{A}\mathbf{A}^\top \mathbf{e}_j|.$$

Now, $\mathbf{A}$ is defined as $(\mathbf{I}-\mathbf{P_V})\otimes\mathbf{D}^\top(\mathbf{I}-\mathbf{P_U})$, therefore using the property of the product of two Kronecker products and product of projection matrices, $\mathbf{A}\mathbf{A}^\top$ can be written as

$$\mathbf{A}\mathbf{A}^\top = (\mathbf{I}-\mathbf{P_V})\otimes\mathbf{D}^\top(\mathbf{I}-\mathbf{P_U})\mathbf{D}.$$

We are interested in the $\{\ell, j\}$ entry of $\mathbf{A}\mathbf{A}^\top$. Since, $\mathbf{A}\mathbf{A}^\top$ has a Kronecker product structure, an entry of $\mathbf{A}\mathbf{A}^\top$ is given by the product of elements of the matrices in the Kronecker product, therefore

$$\max_{j\in\bar{\mathcal{I}}_{\mathcal{S}_e}} \sum_{\ell\in\mathcal{I}_{\mathcal{S}_e}} |\mathbf{e}_l^\top \mathbf{A}\mathbf{A}^\top \mathbf{e}_j| = \max_{j_1,j_2\in\bar{\mathcal{I}}_{\mathcal{S}_e}} \sum_{\ell_1,\ell_2\in\mathcal{I}_{\mathcal{S}_e}} g(j_1,j_2,\ell_1,\ell_2), \quad (30)$$

where $g(j_1,j_2,\ell_1,\ell_2)$ is given by

$$g(j_1,j_2,\ell_1,\ell_2) = |\mathrm{Tr}(\mathbf{e}_{\ell_2}\mathbf{e}_{\ell_1}^\top\mathbf{D}^\top(\mathbf{I}-\mathbf{P_U})\mathbf{D}\mathbf{e}_{j_1}\mathbf{e}_{j_2}^\top(\mathbf{I}-\mathbf{P_V}))|.$$

Now, consider $g(j_1,j_2,\ell_1,\ell_2)$, which can be simplified as

$$\begin{aligned} g(j_1,j_2,\ell_1,\ell_2) = |&\mathrm{Tr}(\mathbf{e}_{\ell_2}\mathbf{e}_{\ell_1}^\top\mathbf{D}^\top(\mathbf{I}-\mathbf{P_U})\mathbf{D}\mathbf{e}_{j_1}\mathbf{e}_{j_2}^\top) \\ &- \mathrm{Tr}(\mathbf{e}_{\ell_2}\mathbf{e}_{\ell_1}^\top\mathbf{D}^\top(\mathbf{I}-\mathbf{P_U})\mathbf{D}\mathbf{e}_{j_1}\mathbf{e}_{j_2}^\top\mathbf{P_V})|. \end{aligned}$$

Since trace is invariant under cyclic permutations, we have

$$\begin{aligned} g(j_1,j_2,\ell_1,\ell_2) = |&\mathbf{e}_{\ell_1}^\top\mathbf{D}^\top(\mathbf{I}-\mathbf{P_U})\mathbf{D}\mathbf{e}_{j_1}\mathbb{1}_{\{j_2=\ell_2\}} \\ &- \mathbf{e}_{\ell_1}^\top\mathbf{D}^\top(\mathbf{I}-\mathbf{P_U})\mathbf{D}\mathbf{e}_{j_1}\mathbf{e}_{j_2}^\top\mathbf{P_V}\mathbf{e}_{\ell_2}|. \end{aligned}$$

Denote $x := \mathbf{e}_{\ell_1}^\top\mathbf{D}^\top(\mathbf{I}-\mathbf{P_U})\mathbf{D}\mathbf{e}_{j_1}$ and $y := \mathbf{e}_{j_2}^\top\mathbf{P_V}\mathbf{e}_{\ell_2}$, then we have

$$g(j_1,j_2,\ell_1,\ell_2) = |x\mathbb{1}_{\{j_2=\ell_2\}} - xy|.$$

Now, the following upper bound on $g(j_1,j_2,\ell_1,\ell_2)$ can be evaluated by squaring both sides and simplifying

$$g(j_1,j_2,\ell_1,\ell_2) \leq x\sqrt{\mathbb{1}_{\{j_2=\ell_2\}} + y^2}. \quad (31)$$

First consider x, which can be written as $x = x\mathbb{1}_{\{j_1=\ell_1\}} + x\mathbb{1}_{\{j_1\neq\ell_1\}}$. Here, $x\mathbb{1}_{\{j_1=\ell_1\}}$ can be upper bounded as shown below using the GFP

$$x = (\mathbf{e}_{\ell_1}^\top\mathbf{D}^\top(\mathbf{I}-\mathbf{P_U})\mathbf{D}\mathbf{e}_{\ell_1}) \leq \mathbf{e}_{\ell_1}^\top\mathbf{D}^\top\mathbf{D}\mathbf{e}_{\ell_1} \leq \alpha_u.$$

Further, we can derive an upper bound on $x\mathbb{1}_{\{j_1\neq\ell_1\}}$ using the parallelogram law for inner-products as follows.

$$\begin{aligned} x &\leq |\mathbf{e}_{j_1}^\top\mathbf{D}^\top\mathbf{D}\mathbf{e}_{\ell_1}| + |\mathbf{e}_{j_1}^\top\mathbf{D}^\top\mathbf{P_U}\mathbf{D}\mathbf{e}_{\ell_1}| \\ &\leq \tfrac{\alpha_u-\alpha_\ell}{2} + \alpha_u\gamma_{\mathbf{U}} = \tfrac{\alpha_u(1+2\gamma_{\mathbf{U}})}{2} - \tfrac{\alpha_\ell}{2}. \end{aligned}$$

Therefore, we have

$$x \leq \alpha_u\mathbb{1}_{\{j_1=\ell_1\}} + (\tfrac{\alpha_u(1+2\gamma_{\mathbf{U}})}{2} - \tfrac{\alpha_\ell}{2})\mathbb{1}_{\{j_1\neq\ell_1\}}.$$

Now, consider $\sqrt{\mathbb{1}_{\{j_2=\ell_2\}} + y^2}$, since $y = \mathbf{e}_{j_2}^\top\mathbf{P_V}\mathbf{P_V}\mathbf{e}_{\ell_2}$, and further, since $\sqrt{a^2+b^2} < (a+b)$ for $a>0$ and $b>0$, we have $\sqrt{\mathbb{1}_{\{j_2=\ell_2\}} + y^2} \leq \mathbb{1}_{\{j_2=\ell_2\}} + \gamma_{\mathbf{V}}$. Now, substituting in (31), i.e., the expression for $g(j_1,j_2,\ell_1,\ell_2)$, we have,

$$\begin{aligned} &g(j_1,j_2,\ell_1,\ell_2) \leq \\ &(\alpha_u\mathbb{1}_{\{j_1=\ell_1\}} + (\tfrac{\alpha_u(1+2\gamma_{\mathbf{U}})}{2} - \tfrac{\alpha_\ell}{2})\mathbb{1}_{\{j_1\neq\ell_1\}})(\mathbb{1}_{\{j_2=\ell_2\}} + \gamma_{\mathbf{V}}), \end{aligned}$$

and finally substituting in (30) and noting that since $j_1, j_2 \in \bar{\mathcal{I}}_{\mathcal{S}_e}$ and $\ell_1,\ell_2 \in \mathcal{I}_{\mathcal{S}_e}$, $\mathbb{1}_{\{j_1=\ell_1\}}\mathbb{1}_{\{j_2=\ell_2\}} = 0$,

$$\begin{aligned} \|\mathbf{A}_{\mathcal{S}_e^\perp}\mathbf{A}_{\mathcal{S}_e}^\top\|_{\infty,\infty} \leq &\max_{j_1,j_2\in\bar{\mathcal{I}}_{\mathcal{S}_e}} \sum_{\ell_1,\ell_2\in\mathcal{I}_{\mathcal{S}_e}} (\tfrac{\alpha_u(1+2\gamma_{\mathbf{U}})}{2} - \tfrac{\alpha_\ell}{2})\mathbb{1}_{\substack{\{j_1\neq\ell_1\},\\ \{j_2=\ell_2\}}} \\ &+ \alpha_u\gamma_{\mathbf{V}}\mathbb{1}_{\{j_1=\ell_1\}} + (\tfrac{\alpha_u(1+2\gamma_{\mathbf{U}})\gamma_{\mathbf{V}}}{2} - \tfrac{\alpha_\ell\gamma_{\mathbf{V}}}{2})\mathbb{1}_{\{j_1\neq\ell_1\}}. \end{aligned} \quad (32)$$

Now, for $\mathbf{A}_0 \in \mathbb{R}^{d\times m}$, the maximum number of non-zeros per row is $\min(s_e, m)$, while those in a column are $\min(s_e, d)$ for the *thin* case and $\min(s_e, k)$ for the *fat* case. Then we have

$$\|\mathbf{A}_{\mathcal{S}_e^\perp}\mathbf{A}_{\mathcal{S}_e}^\top\|_{\infty,\infty} \leq c. \quad (33)$$

Here, the constant c is as defined in (17). Now, to bound the denominator of (29), we have

$$\begin{aligned} \|\mathbf{I}-\mathbf{A}_{\mathcal{S}_e}\mathbf{A}_{\mathcal{S}_e}^\top\|_{\infty,\infty} &= \max_i \|\mathbf{e}_i^\top(\mathbf{I}-\mathbf{A}_{\mathcal{S}_e}\mathbf{A}_{\mathcal{S}_e}^\top)\|_1 \\ &= \max_{j,\ell\in\mathcal{S}} |1-\|\mathbf{e}_j^\top\mathbf{A}\|^2| + \sum_{j\neq\ell} |\langle\mathbf{e}_j^\top\mathbf{A}, \mathbf{e}_l^\top\mathbf{A}\rangle| \end{aligned} \quad (34)$$

We proceed to bound $|1-\|\mathbf{e}_j^\top\mathbf{A}\|^2|$. For this, we derive a lower bound on $\|\mathbf{e}_j^\top\mathbf{A}\|^2$. Note that $\mathbf{e}_j^\top\mathbf{A}$ selects the j-th row of $\mathbf{A}$, which has a Kronecker product structure. Therefore,

$$\begin{aligned} \|\mathbf{e}_j^\top\mathbf{A}\| &= \|(\mathbf{I}-\mathbf{P_U})\mathbf{D}\mathbf{e}_{j_1}\mathbf{e}_{j_2}^\top(\mathbf{I}-\mathbf{P_V})\|_{\mathrm{F}} = \|\mathcal{P}_{\mathcal{L}^\perp}(\mathbf{D}\mathbf{e}_{j_1}\mathbf{e}_{j_2}^\top)\|_{\mathrm{F}} \\ &\geq \|\mathbf{D}\mathbf{e}_{j_1}\mathbf{e}_{j_2}^\top\| - \|\mathcal{P}_{\mathcal{L}}(\mathbf{D}\mathbf{e}_{j_1}\mathbf{e}_{j_2}^\top)\|_{\mathrm{F}} \geq \sqrt{\alpha_\ell}(1-\mu). \end{aligned}$$

Therefore, since $\mu < 1$ and $\alpha_\ell > 0$, then if $\alpha_\ell \leq \frac{1}{(1-\mu)^2}$, we have $|1-\|\mathbf{e}_j^\top\mathbf{A}\|^2| \leq 1-\alpha_\ell(1-\mu)^2$. The analysis for deriving an upper bound for the second term in (34) closely follows that used in (33), as shown below.

$$\sum_{j\neq\ell} |\langle\mathbf{e}_j^\top\mathbf{A}_{\mathcal{S}_e}, \mathbf{e}_l^\top\mathbf{A}_{\mathcal{S}_e}\rangle| = \sum_{(\ell_1,\ell_2)\in\mathcal{S}\setminus\{(j_1,j_2)\}} g(j_1,j_2,\ell_1,\ell_2) \leq c.$$

Combining these results, we have the following bound for

$$\|\mathbf{I}-\mathbf{A}_{\mathcal{S}_e}\mathbf{A}_{\mathcal{S}_e}^\top\|_{\infty,\infty} \leq 1-\alpha_\ell(1-\mu)^2 + c.$$

Finally, substituting these results in (29) we have $\|\mathbf{Q}\|_{\infty,\infty} \leq C_e := \frac{c}{\alpha_\ell(1-\mu)^2 - c}$, where c is given by (**??**). ■

B. Proofs for Column-wise Case

Proof of Lemma 2. We show that for any $(\mathbf{L}_0, \mathbf{S}_0) \in \{\mathbf{M}, \mathcal{U}, \mathcal{I}_{\mathcal{S}_c}\}$, if $\mathrm{span}\{\mathrm{col}(\mathbf{L}_0)\} = \mathcal{U}$ and $\mathrm{csupp}(\mathbf{D}\mathbf{S}_0) = \mathcal{I}_{\mathcal{S}_c}$ do not hold simultaneously, then $\mu = 1$.

Let $\mathbf{L} + \mathbf{D}\mathbf{S} = \mathbf{M}$, as per our model shown in (1). Now, let $(\mathbf{L}_0, \mathbf{S}_0)$ be any other pair in our Oracle Model $\{\mathbf{M}, \mathcal{U}, \mathcal{I}_{\mathcal{S}_c}\}$,

$$\mathbf{L}_0 = \mathbf{L} + \mathbf{\Delta}_1 \in \mathcal{U} \ \text{ and } \ \mathbf{D}\mathbf{S}_0 = \mathbf{D}\mathbf{S} + \mathbf{\Delta}_2 \in \mathcal{S}_c,$$

for some $\mathbf{\Delta}_1$ and $\mathbf{\Delta}_2$, then we have that $\mathbf{\Delta}_1 + \mathbf{\Delta}_2 = \mathbf{0}$. This implies that $\mathrm{csupp}(\mathbf{\Delta}_1) \in \mathcal{S}_c$. Further, this implies that $\mathbf{L}$ and $\mathbf{L}_0$ at least match in the columns indexed by the inliers, i.e., $\mathcal{P}_{\mathcal{I}_\mathbf{L}}(\mathbf{L}) = \mathcal{P}_{\mathcal{I}_\mathbf{L}}(\mathbf{L}_0)$, and we have

$$\mathcal{U} = \mathrm{span}\{\mathrm{col}(\mathbf{L}_0)\} = \mathrm{span}\{\mathrm{col}(\mathcal{P}_{\mathcal{I}_\mathbf{L}}(\mathbf{L}_0))\} = \mathrm{span}\{\mathrm{col}(\mathcal{P}_{\mathcal{I}_\mathbf{L}}(\mathbf{L}))\}.$$

Therefore, $\mathrm{csupp}(\mathbf{D}\mathbf{S}_0) \subseteq \mathcal{I}_{\mathcal{S}_c}$. Specifically, this implies that there may exist a $j \in \mathcal{I}_{\mathcal{S}_c}$ for which $\mathbf{D}\mathbf{S}_{:,j} - (\mathbf{\Delta}_1)_{:,j} = 0$, which will imply that $\mathcal{P}_{\mathcal{U}^\perp}(\mathbf{D}\mathbf{S}_{:,j}) = 0$. This condition implies that $\mu = 1$. Therefore, we require $\mathrm{span}\{\mathrm{col}(\mathbf{L}_0)\} = \mathcal{U}$ and $\mathrm{csupp}(\mathbf{D}\mathbf{S}_0) = \mathcal{I}_{\mathcal{S}_c}$ to hold simultaneously for $\mu < 1$. ■

Proof of Lemma 9. Let $(\mathbf{L}_0, \mathbf{S}_0)$ be an optimal solution pair of (D-RPCA(C)). From the optimality conditions (22) and (23), we seek $\boldsymbol{\Lambda}$ such that

$$\boldsymbol{\Lambda} \in \mathbf{U}\mathbf{V}^\top + \mathbf{W} \quad \text{and} \quad \mathbf{D}^\top \boldsymbol{\Lambda} \in \lambda_c \mathbf{H} + \lambda_c \mathbf{F}. \tag{35}$$

Now consider a feasible solution $\{\mathbf{L_0} + \mathbf{D}\boldsymbol{\Delta}, \mathbf{S_0} - \boldsymbol{\Delta}\}$ for a non-zero $\boldsymbol{\Delta} \in \mathbb{R}^{d\times m}$. Then by the optimality of $(\mathbf{L}_0, \mathbf{S}_0)$ using the subgradient inequality, we have

$$\begin{aligned}\|\mathbf{L}_0 + \mathbf{D}\boldsymbol{\Delta}\|_* + \lambda_c\|\mathbf{S}_0 - \boldsymbol{\Delta}\|_{1,2} &\geq \|\mathbf{L}_0\|_* + \lambda_c\|\mathbf{S}_0\|_{1,2}\\ &+ \langle \mathbf{U}\mathbf{V}^\top + \mathbf{W}, \mathbf{D}\boldsymbol{\Delta}\rangle - \lambda_c\langle \mathbf{H} + \mathbf{F}, \boldsymbol{\Delta}\rangle.\end{aligned}$$

Let $G(\boldsymbol{\Delta}) = \langle \mathbf{U}\mathbf{V}^\top + \mathbf{W}, \mathbf{D}\boldsymbol{\Delta}\rangle - \lambda_c\langle \mathbf{H} + \mathbf{F}, \boldsymbol{\Delta}\rangle$. We will show that if **(q1)-(q4)** hold, then $G(\boldsymbol{\Delta}) > 0$, which proves the optimality of $(\mathbf{L}_0, \mathbf{S}_0)$. Rewrite $G(\boldsymbol{\Delta})$ as

$$G(\boldsymbol{\Delta}) = \langle \mathbf{W}, \mathbf{D}\boldsymbol{\Delta}\rangle - \lambda_c\langle \mathbf{F}, \boldsymbol{\Delta}\rangle + \langle \mathbf{D}^\top\mathbf{U}\mathbf{V}^\top - \lambda_c\mathbf{H}, \boldsymbol{\Delta}\rangle. \tag{36}$$

Let $\mathbf{W}$, with $\|\mathbf{W}\| = 1$ and $\mathcal{P}_{\mathcal{L}}(\mathbf{W}) = \mathbf{0}$, then by duality of norms,

$$\langle \mathbf{W}, \mathbf{D}\boldsymbol{\Delta}\rangle = \langle \mathbf{W}, \mathcal{P}_{\mathcal{L}^\perp}(\mathbf{D}\boldsymbol{\Delta})\rangle = \|\mathcal{P}_{\mathcal{L}^\perp}(\mathbf{D}\boldsymbol{\Delta})\|_*. \tag{37}$$

Further, let $\mathbf{F}$, with $\|\mathbf{F}\|_{\infty,2} = 1$ and $\mathcal{P}_{\mathcal{S}_c}(\mathbf{F}) = \mathbf{0}$, be such that

$$\mathbf{F}_{:,j} = \begin{cases} -\frac{\boldsymbol{\Delta}_{:,j}}{\|\boldsymbol{\Delta}_{:,j}\|}, & \text{if } j \notin \mathcal{I}_{\mathcal{S}_c} \text{ and } \boldsymbol{\Delta}_{:,j} \neq 0\\ 0, & \text{otherwise}\end{cases},$$

where $\mathbf{F}_{:,j}$ denotes the j^{th} column of $\mathbf{F}$. Then, we arrive at the following simplification for $\langle \mathbf{F}, \ \boldsymbol{\Delta}\rangle$ by duality of norms,

$$\langle \mathbf{F}, \ \boldsymbol{\Delta}\rangle = \langle \mathbf{F}, \ \mathcal{P}_{\mathcal{S}_c^\perp}(\boldsymbol{\Delta})\rangle = -\|\mathcal{P}_{\mathcal{S}_c^\perp}(\boldsymbol{\Delta})\|_{1,2}. \tag{38}$$

Since $\mathcal{P}_{\mathcal{L}}(\boldsymbol{\Lambda}) = \mathbf{U}\mathbf{V}^\top$ and $\mathcal{P}_{\mathcal{S}_c}(\mathbf{D}^\top\boldsymbol{\Lambda}) = \lambda_c\mathbf{H}$ by optimality conditions of (35),

$$\begin{aligned}&\langle \mathbf{D}^\top\mathbf{U}\mathbf{V}^\top - \lambda_c\mathbf{H}, \boldsymbol{\Delta}\rangle = -\langle \mathcal{P}_{\mathcal{L}^\perp}(\boldsymbol{\Lambda}), \mathbf{D}\boldsymbol{\Delta}\rangle + \langle \mathcal{P}_{\mathcal{S}_c^\perp}(\mathbf{D}^\top\boldsymbol{\Lambda}), \boldsymbol{\Delta}\rangle\\ &\geq -\|\mathcal{P}_{\mathcal{L}^\perp}(\mathbf{D}\boldsymbol{\Delta})\|_*\|\mathcal{P}_{\mathcal{L}^\perp}(\boldsymbol{\Lambda})\|\\ &\quad - \|\mathcal{P}_{\mathcal{S}_c^\perp}(\boldsymbol{\Delta})\|_{1,2}\|\mathcal{P}_{\mathcal{S}_c^\perp}(\mathbf{D}^\top\boldsymbol{\Lambda})\|_{\infty,2},\end{aligned} \tag{39}$$

where we use Holder's inequality in the last step.

Combining (36), (37), (38), and (39), we have

$$\begin{aligned}G(\boldsymbol{\Delta}) &\geq (1 - \|\mathcal{P}_{\mathcal{L}^\perp}(\boldsymbol{\Lambda})\|)\|\mathcal{P}_{\mathcal{L}^\perp}(\mathbf{D}\boldsymbol{\Delta})\|_*\\ &+ (\lambda_c - \|\mathcal{P}_{\mathcal{S}_c^\perp}(\mathbf{D}^\top\boldsymbol{\Lambda})\|_{\infty,2})\|\mathcal{P}_{\mathcal{S}_c^\perp}(\boldsymbol{\Delta})\|_{1,2}\end{aligned}$$

Since we have an arbitrary $\boldsymbol{\Delta}$ with $\boldsymbol{\Delta} \neq \mathbf{0}$ and $(\mathbf{L}_0 + \mathbf{D}\boldsymbol{\Delta}, \mathbf{S}_0 - \boldsymbol{\Delta}) \notin \{\mathcal{U}, \mathcal{I}_{\mathcal{S}_c}\}$, $\|\mathcal{P}_{\mathcal{L}^\perp}(\mathbf{D}\boldsymbol{\Delta})\|_* = \|\mathcal{P}_{\mathcal{S}_c^\perp}(\boldsymbol{\Delta})\|_{1,2} = 0$ does not hold. Therefore, to ensure the uniqueness of the solution $(\mathbf{L}_0, \mathbf{S}_0)$, we need $\|\mathcal{P}_{\mathcal{L}^\perp}(\boldsymbol{\Lambda})\| < 1$ and $\|\mathcal{P}_{\mathcal{S}_c^\perp}(\mathbf{D}^\top\boldsymbol{\Lambda})\|_{\infty,2} < \lambda_c$. Hence, any dual certificate which obeys the conditions **(C1)-(C4)** guarantees optimality of the solution. ■

Proof of Lemma 10. We begin by writing the definition of $\sigma_{\min}(\mathbf{A}_{\mathcal{S}_c}^\top)$ as

$$\sigma_{\min}(\mathbf{A}_{\mathcal{S}_c}^\top) = \min_{\mathbf{H}\in\mathcal{S}_c/\{\mathbf{0}_{d\times m}\}} \frac{\|\mathbf{A}^\top\text{vec}(\mathbf{H})\|_2}{\|\text{vec}(\mathbf{H})\|_2}.$$

By the definition of $\mathbf{A}$ and using the property of Kronecker product for multiplication by a vector we have

$$\sigma_{\min}(\mathbf{A}_{\mathcal{S}_c}^\top) = \min_{\mathbf{H}\in\mathcal{S}_c/\{\mathbf{0}_{d\times m}\}} \frac{\|(\mathbf{I}-\mathbf{P_U})\mathbf{D}\mathbf{H}(\mathbf{I}-\mathbf{P_V})\|_{\text{F}}}{\|\mathbf{H}\|_{\text{F}}}.$$

Further $(\mathbf{I} - \mathbf{P_U})\,\mathbf{D}\mathbf{H}\,(\mathbf{I} - \mathbf{P_V}) = \mathcal{P}_{\mathcal{L}^\perp}(\mathbf{D}\mathbf{H})$, and we can write that expression above as follows

$$\begin{aligned}\sigma_{\min}(\mathbf{A}_{\mathcal{S}_c}^\top) &= \min_{\mathbf{H}\in\mathcal{S}_c/\{\mathbf{0}_{d\times m}\}} \frac{\|\mathbf{D}\mathbf{H}\|_{\text{F}}}{\|\mathbf{H}\|_{\text{F}}}\cdot\frac{\|(\mathbf{I}-\mathcal{P}_{\mathcal{L}})(\mathbf{D}\mathbf{H})\|_{\text{F}}}{\|\mathbf{D}\mathbf{H}\|_{\text{F}}}\\ &\overset{(i)}{\geq} \sqrt{\alpha_\ell}\Big(1 - \max_{\mathbf{Z}\in\mathcal{D}/\{\mathbf{0}_{n\times m}\}} \frac{\|\mathcal{P}_{\mathcal{L}}(\mathbf{Z})\|_{\text{F}}}{\|\mathbf{Z}\|_{\text{F}}}\Big) \overset{(ii)}{\geq} \sqrt{\alpha_\ell}(1-\mu).\end{aligned}$$

Here (i) is due to the GFP condition D.2 and the reverse triangle inequality, and (ii) from the incoherence property in (2). ■

Proof of Lemma 11. We start by using the correspondence between the vector $\mathbf{b}_{\mathcal{S}_c}$ and the matrix $\mathbf{B}_{\mathcal{S}_c}$, i.e.,

$$\|\mathbf{b}_{\mathcal{S}_c}\|_2 = \|\mathbf{B}_{\mathcal{S}_c}\|_{\text{F}} = \|\lambda_c\tilde{\mathbf{S}} - \mathcal{P}_{\mathcal{S}_c}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)\|_{\text{F}}.$$

Now, since $\tilde{\mathbf{S}}_{:,j} = \mathbf{S}_{:,j}/\|\mathbf{S}_{:,j}\|_2$ for all $j \in \mathcal{I}_{\mathcal{S}_c}$; and is $\mathbf{0}$ otherwise (i.e., when $j \notin \mathcal{I}_{\mathcal{S}_c}$), using triangle inequality, we have

$$\|\mathbf{b}_{\mathcal{S}_c}\|_2 \leq \lambda_c\sqrt{s_c} + \|\mathcal{P}_{\mathcal{S}_c}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)\|_{\text{F}}. \tag{40}$$

Since we have

$$\begin{aligned}&\|\mathcal{P}_{\mathcal{S}_c}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)\|_{\text{F}}^2 \leq \|\mathcal{P}_{\mathcal{L}}(\mathbf{U}\mathbf{V}^\top)\|_{\text{F}}\|\mathcal{P}_{\mathcal{L}}(\mathbf{D}\mathcal{P}_{\mathcal{S}_c}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top))\|_{\text{F}}\\ &\overset{(i)}{\leq}\mu\|\mathbf{U}\mathbf{V}^\top\|_{\text{F}}\|\mathbf{D}\mathcal{P}_{\mathcal{S}_c}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)\|_{\text{F}} \overset{(ii)}{\leq} \sqrt{r\alpha_u}\mu\|\mathcal{P}_{\mathcal{S}_c}(\mathbf{D}^\top\mathbf{U}\mathbf{V}^\top)\|_{\text{F}},\end{aligned} \tag{41}$$

where (i) is from subspace incoherence property and (ii) is from the GFP D.2. Combining (40) and (41), we have

$$\|\text{vec}(\mathbf{B}_{\mathcal{S}_c})\|_2 \leq \lambda_c\sqrt{s_c} + \sqrt{r\alpha_u}\mu. \quad ■$$

Proof of Lemma 12. We begin by analyzing the quantity of interest – $\|\mathcal{P}_{\mathcal{S}_c^\perp}(\mathbf{Z})\|_{\infty,2}$, i.e., we are interested in the maximum column norm of the matrix $\mathcal{P}_{\mathcal{S}_c^\perp}(\mathbf{Z})$. Note that $\mathbf{Z}$ is defined as

$$\mathbf{Z} = \mathbf{D}^\top\,(\mathbf{I} - \mathbf{P_U})\,\mathbf{X}\,(\mathbf{I} - \mathbf{P_V})\,,$$

and we have $\text{vec}(\mathbf{Z}) = \mathbf{A}\text{vec}(\mathbf{X})$. Further, we have that

$$\mathcal{P}_{\mathcal{S}_c^\perp}\,(\text{vec}(\mathbf{Z})) = \mathbf{A}_{\mathcal{S}_c^\perp}\text{vec}(\mathbf{X}).$$

Now, observe that the columns of matrix $\mathcal{P}_{\mathcal{S}_c^\perp}(\mathbf{Z})$ appear as blocks of size $n\times 1$ in the vector $\mathcal{P}_{\mathcal{S}_c^\perp}\,(\text{vec}(\mathbf{Z}))$. Moreover, the elements of vector $\mathcal{P}_{\mathcal{S}_c^\perp}\,(\text{vec}(\mathbf{Z}))$ are formed due to the inner product between the rows of Kronecker product structured matrix $\mathbf{A}_{\mathcal{S}_c^\perp}$ and $\text{vec}(X)$. Therefore, to identify a column of $\mathcal{P}_{\mathcal{S}_c^\perp}\,(\mathbf{Z})$ we need to focus on the interaction between correponding rows of $\mathbf{A}_{\mathcal{S}_c^\perp}$ and $\text{vec}(\mathbf{X})$.

Consider the Kronecker product structured matrix $\mathbf{A}_{\mathcal{S}_c^\perp}$. Since the rows in $\mathbf{A}_{\mathcal{S}_c^\perp}$ correspond to all rows outside the column support $\mathcal{S}_c$, this corresponds to selecting those rows of $m\times m$ matrix $(\mathbf{I} - \mathbf{P_V})$ which correspond to $\mathcal{S}_c^\perp$, which we denote by $(\mathbf{I} - \mathbf{P_V})_{\mathcal{S}_c^\perp}$ i.e.,

$$\mathbf{A}_{\mathcal{S}_c^\perp} = (\mathbf{I} - \mathbf{P_V})_{\mathcal{S}_c^\perp} \otimes \mathbf{D}^\top(\mathbf{I} - \mathbf{P_U}).$$

For simplicity of the upcoming analysis, we denote the matrix $(\mathbf{I} - \mathbf{P_V})$ as

$$(\mathbf{I} - \mathbf{P_V}) = \begin{bmatrix} v_{11} & \cdots & v_{1m}\\ \vdots & \ddots & \vdots\\ v_{m1} & \cdots & v_{mm}\end{bmatrix}.$$

Using this notation, the j-th block of vector $\mathcal{P}_{\mathcal{S}_c^\perp}\,(\text{vec}(\mathbf{Z}))$ (which is also the j-th column of $\mathcal{P}_{\mathcal{S}_c^\perp}\,(\mathbf{Z})$), can be written as

$$\mathbf{Z}_{:,j} = (v_{j,:} \otimes \mathbf{D}^\top(\mathbf{I} - \mathbf{P_U}))\text{vec}(\mathbf{X})$$

for some $j \in \mathcal{I}_{\mathcal{S}_c^\perp}$. Now, further since $\text{vec}(\mathbf{X}) := \mathbf{A}_{\mathcal{S}_c}^\top(\mathbf{A}_{\mathcal{S}_c}\mathbf{A}_{\mathcal{S}_c}^\top)^{-1}\text{vec}(\mathbf{B}_{\mathcal{S}_c})$, therefore we are interested in maximum 2-norm of

$$\mathbf{Z}_{:,j} = (v_{j,:} \otimes \mathbf{D}^\top(\mathbf{I} - \mathbf{P_U}))\mathbf{A}_{\mathcal{S}_c}^\top(\mathbf{A}_{\mathcal{S}_c}\mathbf{A}_{\mathcal{S}_c}^\top)^{-1}\text{vec}(\mathbf{B}_{\mathcal{S}_c}),$$

for some $j \in \mathcal{I}_{\mathcal{S}_c^\perp}$. Note that $\mathbf{A}_{\mathcal{S}_c}^\top$ itself is a Kronecker product structured matrix given by

$$\mathbf{A}_{\mathcal{S}_c} = (\mathbf{I} - \mathbf{P_V})_{\mathcal{S}_c}^\top \otimes (\mathbf{I} - \mathbf{P_U})\mathbf{D}.$$

Using the mixed product rule for Kronecker products we have $\mathbf{Z}_{:,j} = (v_{j,:}(\mathbf{I}-\mathbf{P_V})^\top_{\mathcal{S}_c} \otimes \mathbf{D}^\top(\mathbf{I}-\mathbf{P_U})\mathbf{D})(\mathbf{A}_{\mathcal{S}_c}\mathbf{A}_{\mathcal{S}_c}^\top)^{-1}\mathbf{b}_{\mathcal{S}_c}$, for some $j \in \mathcal{I}_{\mathcal{S}_c^\perp}$. Further, since for two matrices $\mathbf{A}$ and $\mathbf{B}$, $\|\mathbf{A}\otimes\mathbf{B}\| = \|\mathbf{A}\|\|\mathbf{B}\|$, we have

$$\begin{aligned}\|\mathbf{Z}_{:,j}\| \leq\ & \|\mathbf{e}_j^\top(\mathbf{I}-\mathbf{P_V})_{\mathcal{S}_c^\perp}(\mathbf{I}-\mathbf{P_V})^\top_{\mathcal{S}_c}\| \\ & \times \|\mathbf{D}^\top(\mathbf{I}-\mathbf{P_U})\mathbf{D}\|\|(\mathbf{A}_{\mathcal{S}_c}\mathbf{A}_{\mathcal{S}_c}^\top)^{-1}\|\|\mathbf{b}_{\mathcal{S}_c}\|, \quad (42)\end{aligned}$$

where we also use the fact that $v_{j,:} = \mathbf{e}_j^\top(\mathbf{I}-\mathbf{P_V})_{\mathcal{S}_c^\perp}$. We will now proceed to bound the first term in (42). Note that

$$\begin{aligned}&\max_{j\in\mathcal{S}_c^\perp}\|\mathbf{e}_j^\top(\mathbf{I}-\mathbf{P_V})_{\mathcal{S}_c^\perp}(\mathbf{I}-\mathbf{P_V})^\top_{\mathcal{S}_c}\|^2 \\ &= \max_{j\in\mathcal{S}_c^\perp}\sum_{i\in\mathcal{S}_c}\langle(\mathbf{I}-\mathbf{P_V})^\top\mathbf{e}_j, (\mathbf{I}-\mathbf{P_V})^\top\mathbf{e}_i\rangle^2.\end{aligned}$$

Now, each term in the summation can be bounded as

$$\begin{aligned}&\max_{j\in\mathcal{S}_c^\perp, i\in\mathcal{S}_c}|\langle(\mathbf{I}-\mathbf{P_V})^\top\mathbf{e}_j, (\mathbf{I}-\mathbf{P_V})^\top\mathbf{e}_i\rangle| \\ &= \max_{j\in\mathcal{S}_c^\perp, i\in\mathcal{S}_c}|-\langle\mathbf{P_V}\mathbf{e}_j, \mathbf{P_V}\mathbf{e}_i\rangle| \leq \|\mathbf{P_V}\mathbf{e}_j\|\|\mathbf{P_V}\mathbf{e}_i\| \leq \gamma_\mathbf{V}.\end{aligned}$$

This implies $\|\mathbf{e}_j^\top(\mathbf{I}-\mathbf{P_V})_{\mathcal{S}_c^\perp}(\mathbf{I}-\mathbf{P_V})^\top_{\mathcal{S}_c}\| \leq \sqrt{s_c}\gamma_\mathbf{V}$. Further, note that $\|(\mathbf{A}_{\mathcal{S}_c}\mathbf{A}_{\mathcal{S}_c}^\top)^{-1}\| \leq \|\mathbf{A}_{\mathcal{S}_c}^{-1}\|^2 = \frac{1}{\sigma_{\min}(\mathbf{A}_{\mathcal{S}_c})^2}$. Substituting this into (42), for a $j \in \mathcal{S}_c^\perp$, we have

$$\|\mathbf{Z}_{:,j}\| \leq \tfrac{\sqrt{s_c}\gamma_\mathbf{V}}{\sigma_{\min}(\mathbf{A}_{\mathcal{S}_c})^2}\|\mathbf{D}^\top(\mathbf{I}-\mathbf{P_U})\mathbf{D}\|\|\mathbf{b}_{\mathcal{S}_c}\|. \quad (43)$$

We can further write $\|\mathbf{D}^\top(\mathbf{I}-\mathbf{P_U})\mathbf{D}\|$ as follows

$$\|\mathbf{D}^\top(\mathbf{I}-\mathbf{P_U})\mathbf{D}\| = \max_{\|u\|=1}\tfrac{\|(\mathbf{I}-\mathbf{P_U})\mathbf{D}u\|^2}{\|\mathbf{D}u\|^2}\|\mathbf{D}u\|^2 \leq \beta_\mathbf{U}\alpha_u.$$

Substituting this result in (43), using Lemma 10 and Lemma 11,

$$\|\mathcal{P}_{\mathcal{S}_c^\perp}(\mathbf{Z})\|_{\infty,2} \leq \sqrt{s_c}C_c(\lambda_c\sqrt{s_c} + \sqrt{r\alpha_u}\mu).$$

■

References

[1] I. Jolliffe, *Principal component analysis*, Wiley Online Library, 2002.

[2] S. Rambhatla, X. Li, and J. Haupt, "Target-based hyperspectral demixing via generalized robust PCA," in *51st Asilomar Conference on Signals, Systems, and Computers, ACSSC 2017, Pacific Grove, CA, USA, October 29 - November 1, 2017*, 2017, pp. 420–424.

[3] X. Li, J. Ren, S. Rambhatla, Y. Xu, and J. Haupt, "Robust pca via dictionary based outlier pursuit," in *2018 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP),*. IEEE, 2018.

[4] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227–234, 1995.

[5] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.

[6] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[7] H. Rauhut, "Compressive sensing and structured random matrices," *Theoretical foundations and numerical methods for sparse recovery*, vol. 9, pp. 1–92, 2010.

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

[9] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, "Rank-sparsity incoherence for matrix decomposition," *SIAM Journal on Optimization*, vol. 21, no. 2, pp. 572–596, 2011.

[10] Z. Zhou, X. Li, J. Wright, E. J. Candès, and Y. Ma, "Stable principal component pursuit," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*. IEEE, 2010, pp. 1518–1522.

[11] X. Ding, L. He, and L. Carin, "Bayesian robust principal component analysis," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3419–3430, 2011.

[12] J. Wright, A. Ganesh, K. Min, and Y. Ma, "Compressive principal component pursuit," *Information and Inference*, vol. 2, no. 1, pp. 32–68, 2013.

[13] Y. Chen, A. Jalali, S. Sanghavi, and C. Caramanis, "Low-rank matrix recovery from errors and erasures," *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4324–4337, 2013.

[14] M. Mardani, G. Mateos, and G. B. Giannakis, "Recovery of low-rank plus compressed sparse matrices with application to unveiling traffic anomalies," *IEEE Transactions on Information Theory*, vol. 59, no. 8, pp. 5186–5205, 2013.

[15] K. Min, Z. Zhang, J. Wright, and Y. Ma, "Decomposing background topics from keywords by principal component pursuit," in *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, New York, NY, USA, 2010, CIKM '10, pp. 269–278, ACM.

[16] S. Rambhatla, X. Li, J. Ren, and J. Haupt, "A Dictionary-Based Generalization of Robust PCA Part II: Applications to Target Localization in Hyperspectral Imaging," *Journal In Review*, 2018.

[17] P. S. Huang, S. D. Chen, P. Smaragdis, and M. J. Hasegawa, "Singing-voice separation from monaural recordings using robust principal component analysis," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),*. IEEE, 2012, pp. 57–60.

[18] P. Sprechmann, A. M. Bronstein, and G. Sapiro, "Real-time online singing voice separation from monaural recordings using robust low-rank modeling.," in *ISMIR*, 2012, pp. 67–72.

[19] J. L. Starck, Y. Moudden, J. Bobin, M. Elad, and D. L. Donoho, "Morphological component analysis," in *Optics & Photonics 2005*. International Society for Optics and Photonics, 2005, pp. 59140Q–59140Q.

[20] S. Rambhatla and J. Haupt, "Semi-blind source separation via sparse representations and online dictionary learning," in *Signals, Systems and Computers, 2013 Asilomar Conference on*. IEEE, 2013, pp. 1687–1691.

[21] H. Xu, C. Caramanis, and S. Sanghavi, "Robust pca via outlier pursuit," in *Neural Information Processing Systems*, 2010, pp. 2496–2504.

[22] X. Li and J. Haupt, "Identifying outliers in large matrices via randomized adaptive compressive sampling," *Trans. Signal Processing*, vol. 63, no. 7, pp. 1792–1807, 2015.

[23] X. Li and J. Haupt, "Locating salient group-structured image features via adaptive compressive sensing," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2015.

[24] M. Rahmani and G. Atia, "Randomized robust subspace recovery for high dimensional data matrices," *arXiv preprint arXiv:1505.05901*, 2015.

[25] B. Mehta and W. Nejdl, "Attack resistant collaborative filtering," in *Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval*, 2008, pp. 75–82.

[26] A. Lakhina, M. Crovella, and C. Diot, "Diagnosing network-wide traffic anomalies," in *ACM SIGCOMM Computer Communication Review*. ACM, 2004, vol. 34, pp. 219–230.

[27] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[28] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency.," in *Advances in Neural Information Processing Systems*, 2006, pp. 545–552.

[29] T. Liu, J. Sun, N. Zheng, X. Tang, and H. Shum, "Learning to detect a salient object," in *Proc. CVPR*, 2007.

[30] S. Rambhatla, X. Li, and J. Haupt, "A dictionary based generalization of robust PCA," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2016.

[31] R. J. Duffin and A. C. Schaeffer, "A class of nonharmonic fourier series," *Transactions of the American Mathematical Society*, vol. 72, no. 2, pp. 341–366, 1952.

[32] C. Heil, "What is ... a frame?," *Notices of the American Mathematical Society*, vol. 60, no. 6, June/July 2013.

[33] S. Tu, R. Boczar, M. Soltanolkotabi, and B. Recht, "Low-rank solutions of linear matrix equations via procrustes flow," *arXiv preprint arXiv:1507.03566*, 2015.

[34] Y. Chen and M. Wainwright, "Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees," *arXiv preprint arXiv:1509.03025*, 2015.

[35] G.A. Watson, "Characterization of the subdifferential of some matrix norms," *Linear Algebra and its Applications*, vol. 170, pp. 33 – 45, 1992.