EXACT RECOVERY OF LOW-RANK PLUS COMPRESSED SPARSE MATRICES

Morteza Mardani, Gonzalo Mateos, and Georgios B. Giannakis

Dept. of ECE, University of Minnesota, Minneapolis, MN 55455, USA

ABSTRACT

Given the superposition of a low-rank matrix plus the product of a known fat compression matrix times a sparse matrix, the goal of this paper is to establish conditions under which exact recovery of the low-rank and sparse components becomes possible. This fundamental identifiability task subsumes compressed sensing and the timely low-rank plus sparse matrix recovery encountered in matrix decomposition problems. Leveraging the ability of ℓ_1- and nuclear norms to recover sparse and low-rank matrices, a convex program is formulated to estimate the unknowns. Analysis and simulations confirm that the said convex program can recover the unknowns for sufficiently low-rank and sparse enough components, along with a compression matrix possessing an isometry property.

1. INTRODUCTION

Let $\mathbf{X}_0 \in \mathbb{R}^{L\times T}$ be a low-rank matrix $[r := \text{rank}(\mathbf{X}_0) \ll \min(L,T)]$, and let $\mathbf{A}_0 \in \mathbb{R}^{F\times T}$ be sparse ($s := \|\mathbf{A}_0\|_0 \ll FT$, $\|\cdot\|_0$ counts the nonzero entries of its matrix argument). Given the compression matrix $\mathbf{R} \in \mathbb{R}^{L\times F}$ with $L \leq F$, and observations

$$\mathbf{Y} = \mathbf{X}_0 + \mathbf{R}\mathbf{A}_0 \tag{1}$$

this paper deals with the recovery of $\{\mathbf{X}_0, \mathbf{A}_0\}$. This is of interest e.g., to unveil anomalous flows in backbone networks [6,7]. In addition, this problem ties compressed sensing with low-rank plus sparse matrix decompositions. In the absence of the low-rank component ($\mathbf{X}_0 = \mathbf{0}_{L\times T}$), one is left with an under-determined sparse recovery problem [2]. When $\mathbf{Y} = \mathbf{X}_0 + \mathbf{A}_0$, the formulation boils down to principal components pursuit [3, 4]. Sufficient conditions for exact recovery are available for both of the aforementioned special cases. However, the superposition of a low-rank and a *compressed* sparse matrix further challenges identifiability of $\{\mathbf{X}_0, \mathbf{A}_0\}$.

The main contribution of this paper is to establish that given $\mathbf{Y}$ and $\mathbf{R}$ in (1), for small enough r and s one can *exactly* recover $\{\mathbf{X}_0, \mathbf{A}_0\}$ by solving the *convex* optimization problem

$$\text{(P1)} \quad \min_{\{\mathbf{X},\mathbf{A}\}} \ \|\mathbf{X}\|_* + \lambda\|\mathbf{A}\|_1 \quad \text{s.to} \ \ \mathbf{Y} = \mathbf{X} + \mathbf{R}\mathbf{A}.$$

where $\lambda \geq 0$; $\|\mathbf{X}\|_* := \sum_i \sigma_i(\mathbf{X})$ is the nuclear norm of $\mathbf{X}$ (σ_i stands for the i-th singular value); and, $\|\mathbf{X}\|_1 := \sum_{i,j} |x_{ij}|$ denotes the ℓ_1-norm – these are convex surrogates to the rank and ℓ_0-norm, respectively [2–4]. Recently, a greedy algorithm for recovering low-rank and sparse matrices from compressive measurements was put forth in [8]. However, convergence of the algorithm and its error performance are only assessed via numerical simulations.

A *deterministic* approach along the lines of [4] is adopted first to derive conditions under which (1) is locally identifiable (Section 2). Introducing a notion of incoherence between the additive components $\mathbf{X}_0$ and $\mathbf{R}\mathbf{A}_0$, and resorting to the restricted isometry constants of $\mathbf{R}$ [2], sufficient conditions are obtained in Section 3 to ensure that (P1) succeeds in exactly recovering the unknowns. Intuitively, the results here assert that if r and s are sufficiently small, the non-zero entries of $\mathbf{A}_0$ are sufficiently spread out, and subsets of columns of $\mathbf{R}$ behave as isometries, then (P1) exactly recovers $\{\mathbf{X}_0, \mathbf{A}_0\}$; see Section 4 for a sketch, and [7] for the detailed proof.

Notation: Operators $(\cdot)'$ and $\otimes$ will denote transposition and Kronecker product, respectively; $|\cdot|$, $\|\mathbf{x}\|$, $\|\mathbf{x}\|_\infty$ will denote the cardinality of a set, the ℓ_2- and ℓ_∞-norm of a vector, respectively. For matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$ $\langle \mathbf{A}, \mathbf{B}\rangle := \text{tr}(\mathbf{A}'\mathbf{B})$ denotes inner product, $\|\mathbf{A}\|_F$ the Frobenious norm, $\|\mathbf{A}\|$ the spectral norm, $\|\mathbf{A}\|_\infty := \max_{i,j} |a_{ij}|$ the ℓ_∞-norm, and $\|\mathbf{A}\|_{\infty,\infty} := \max_{\|\mathbf{x}\|_\infty=1} \|\mathbf{A}\mathbf{x}\|_\infty$ the induced ℓ_∞-norm. The $n \times n$ identity matrix will be represented by $\mathbf{I}_n$, and its i-th column by $\mathbf{e}_i$; likewise $\mathbf{0}_{n\times p} := \mathbf{0}_n\mathbf{0}'_p$. Define also the support set $\text{supp}(\mathbf{A}) := \{(i,j) : a_{ij} \neq 0\}$.

2. LOCAL IDENTIFIABILITY

Let $\mathbf{U}\mathbf{\Sigma}\mathbf{V}'$ denote the singular value decomposition (SVD) of $\mathbf{X}_0$, and consider the subspaces: s1) $\Phi(\mathbf{X}_0) := \{\mathbf{Z} \in \mathbb{R}^{L\times T} : \mathbf{Z} = \mathbf{U}\mathbf{W}'_1 + \mathbf{W}_2\mathbf{V}',\ \mathbf{W}_1 \in \mathbb{R}^{T\times r},\ \mathbf{W}_2 \in \mathbb{R}^{L\times r}\}$ of matrices in either the column or row space of $\mathbf{X}_0$; s2) $\Omega(\mathbf{A}_0) := \{\mathbf{H} \in \mathbb{R}^{F\times T} : \text{supp}(\mathbf{H}) \subseteq \text{supp}(\mathbf{A}_0)\}$ of matrices in $\mathbb{R}^{F\times T}$ with support contained in the support of $\mathbf{A}_0$; and s3) $\Omega_R(\mathbf{A}_0) := \{\mathbf{Z} \in \mathbb{R}^{L\times T} : \mathbf{Z} = \mathbf{R}\mathbf{H},\ \mathbf{H} \in \Omega(\mathbf{A}_0)\}$. For notational brevity, s1)-s3) will be henceforth denoted as $\{\Phi, \Omega_R, \Omega\}$. Noteworthy properties of these subspaces are: i) both Φ and $\Omega_R \subset \mathbb{R}^{L\times T}$, hence it is possible to directly compare elements from them; ii) $\mathbf{X}_0 \in \Phi$ and $\mathbf{R}\mathbf{A}_0 \in \Omega_R$; and iii) if $\mathbf{Z} \in \Phi^\perp$ is added to $\mathbf{X}_0$, then $\text{rank}(\mathbf{Z} + \mathbf{X}_0) > r$.

For now, assume that the subspaces Ω_R and Φ are also known. This extra information helps identifiability of (1), because potentially troublesome solutions $\{\mathbf{X}_0 + \mathbf{R}\mathbf{H}, \mathbf{A}_0 - \mathbf{H}\}$ are limited to a restricted class. If $\mathbf{X}_0 + \mathbf{R}\mathbf{H} \notin \Phi$ or $\mathbf{A}_0 - \mathbf{H} \notin \Omega$, that candidate solution is not admissible since it is known a priori that $\mathbf{A}_0 \in \Omega$ and $\mathbf{X}_0 \in \Phi$. Under these assumptions, the following lemma presents the necessary and sufficient conditions guaranteeing unique decomposability of $\mathbf{Y}$ according to (1) – a notion known as *local identifiability* [3]. Proofs are omitted here due to space limitations; see [7].

Lemma 1: *Matrix $\mathbf{Y}$ uniquely decomposes into $\mathbf{X}_0 + \mathbf{R}\mathbf{A}_0$ if and only if $\Phi \cap \Omega_R = \{\mathbf{0}_{L\times T}\}$, and $\mathbf{R}\mathbf{H} \neq \mathbf{0}_{L\times T}, \forall \mathbf{H} \in \Omega \backslash \{\mathbf{0}_{F\times T}\}$.*

In words, (1) is locally identifiable if and only if the subspaces Φ and Ω_R intersect transversally, and the sparse matrices in Ω are not annihilated by $\mathbf{R}$. This last condition is unique to the setting here, and is not present in [3, 4].

Remark 1 (Projection operators) $\mathcal{P}_\Omega(\mathbf{X})$ $(\mathcal{P}_{\Omega^\perp}(\mathbf{X}))$ denotes the projection of $\mathbf{X}$ onto the subspace Ω (orthogonal complement $\Omega^\perp$). It simply sets those elements of $\mathbf{X}$ not in $\text{supp}(\mathbf{A}_0)$ to zero. Likewise, $\mathcal{P}_\Phi(\mathbf{X})$ $(\mathcal{P}_{\Phi^\perp}(\mathbf{X}))$ denotes the projection

Work was supported by an AFOSR MURI grant no. FA 9550-10-1-0567.

of $\mathbf{X}$ onto the subspace Φ (orthogonal complement $\Phi^{\perp}$). Let $\mathbf{P}_U := \mathbf{U}\mathbf{U}'$ and $\mathbf{P}_V := \mathbf{V}\mathbf{V}'$ denote, respectively, projection onto the column and row spaces of $\mathbf{X}_0$. It can be shown that $\mathcal{P}_{\Phi}(\mathbf{X}) = \mathbf{P}_U\mathbf{X} + \mathbf{X}\mathbf{P}_V - \mathbf{P}_U\mathbf{X}\mathbf{P}_V$, while the projection onto the complement subspace is $\mathcal{P}_{\Phi^{\perp}}(\mathbf{X}) = (\mathbf{I} - \mathbf{P}_U)\mathbf{X}(\mathbf{I} - \mathbf{P}_V)$.

2.1. Incoherence measures

Building on Lemma 1, simpler sufficient conditions are derived here to ensure local identifiability. To quantify the overlap between Φ and Ω_R, introduce the *incoherence* parameter

$$\mu(\Omega_R, \Phi) := \max_{\mathbf{Z}\in\Omega_R\setminus\{\mathbf{0}_{L\times T}\}} \|\mathcal{P}_{\Phi}(\mathbf{Z})\|_F/\|\mathbf{Z}\|_F. \quad (2)$$

Observe that $\mu(\Omega_R, \Phi) \in [0, 1]$. The lower bound is achieved when Φ and Ω_R are orthogonal, while the upper bound is attained when $\Phi\cap\Omega_R$ contains a non-zero element. Assuming $\Phi\cap\Omega_R = \{\mathbf{0}_{L\times T}\}$, then $\mu(\Omega_R, \Phi) < 1$ represents the cosine of the angle between Φ and Ω_R [5]. From Lemma 1, it appears that $\mu(\Omega_R, \Phi) < 1$ guarantees $\Phi \cap \Omega_R = \{\mathbf{0}_{L\times T}\}$. As it will become clear later on, tighter conditions on $\mu(\Omega_R, \Phi)$ will prove instrumental to guarantee exact recovery of $\{\mathbf{X}_0, \mathbf{A}_0\}$ by solving (P1).

To measure the incoherence among subsets of columns of $\mathbf{R}$, which is tightly related to the second condition in Lemma 1, the restricted isometry constants (RICs) come handy [2]. The constant $\delta_k(\mathbf{R})$ measures the extent to which a k-subset of columns of $\mathbf{R}$ behaves like an isometry. It is defined as the smallest value satisfying

$$c(1 - \delta_k(\mathbf{R})) \leq \|\mathbf{R}\mathbf{u}\|^2/\|\mathbf{u}\|^2 \leq c(1 + \delta_k(\mathbf{R})) \quad (3)$$

for every $\mathbf{u} \in \mathbb{R}^F$ with $\|\mathbf{u}\|_0 \leq k$ and for some positive normalization constant $c < 1$ [2]. For later use, introduce $\theta_{s_1,s_2}(\mathbf{R})$ which measures 'how orthogonal' are two disjoint column subsets of $\mathbf{R}$ with cardinality s_1 and s_2. It is the smallest value that satisfies

$$|\langle\mathbf{R}\mathbf{u}_1, \mathbf{R}\mathbf{u}_2\rangle| \leq c\,\theta_{s_1,s_2}(\mathbf{R})\|\mathbf{u}_1\|\|\mathbf{u}_2\| \quad (4)$$

for every $\mathbf{u}_1, \mathbf{u}_2 \in \mathbb{R}^F$, where $\text{supp}(\mathbf{u}_1) \cap \text{supp}(\mathbf{u}_2) = \emptyset$ and $\|\mathbf{u}_1\|_0 \leq s_1, \|\mathbf{u}_2\|_0 \leq s_2$. The normalization constant c plays the same role as in $\delta_k(\mathbf{R})$. A wide family of matrices with small RICs have been introduced in e.g., [2].

All the elements are in place to state this section's main result.

Proposition 1: *Assume that each column of $\mathbf{A}_0$ contains at most k nonzero elements. If $\mu(\Omega_R, \Phi) < 1$ and $\delta_k(\mathbf{R}) < 1$, then $\Omega_R\cap\Phi = \{\mathbf{0}_{L\times T}\}$ and $\mathbf{R}\mathbf{H} \neq \mathbf{0}_{L\times T}, \forall\mathbf{H} \in \Omega\setminus\{\mathbf{0}_{F\times T}\}$.*
For small k, the extra assumption in Proposition 1 precludes the entries of $\mathbf{A}_0$ from clustering in a few columns of $\mathbf{A}_0$.

3. EXACT RECOVERY VIA CONVEX OPTIMIZATION

In addition to $\mu(\Omega_R, \Phi)$, there are other incoherence measures which play an important role in the conditions for exact recovery. Consider a feasible solution $\{\mathbf{X}_0 + a_{ij}\mathbf{R}\mathbf{e}_i\mathbf{e}_j', \mathbf{A}_0 - a_{ij}\mathbf{e}_i\mathbf{e}_j'\}$, where $(i, j) \notin \text{supp}(\mathbf{A}_0)$ and thus $a_{ij}\mathbf{e}_i\mathbf{e}_j' \notin \Omega$. It may then happen that $a_{ij}\mathbf{R}\mathbf{e}_i\mathbf{e}_j' \in \Phi$ and $\text{rank}(\mathbf{X}_0 + a_{ij}\mathbf{R}\mathbf{e}_i\mathbf{e}_j') = \text{rank}(\mathbf{X}_0) - 1$, while $\|\mathbf{A}_0 - a_{ij}\mathbf{e}_i\mathbf{e}_j'\|_0 = \|\mathbf{A}_0\|_0 + 1$, challenging identifiability when Φ and Ω_R are unknown. Similar complications will arise if $\mathbf{X}_0$ has a sparse row space that could be confused with the row space of $\mathbf{A}_0$. These issues motivate defining

$$\gamma_R(\mathbf{U}) := \max_{i,j} \frac{\|\mathbf{P}_U\mathbf{R}\mathbf{e}_i\mathbf{e}_j'\|_F}{\|\mathbf{R}\mathbf{e}_i\mathbf{e}_j'\|_F},\ \gamma(\mathbf{V}) := \max_i \|\mathbf{P}_V\mathbf{e}_i\|_F$$

where $\gamma_R(\mathbf{U}), \gamma(\mathbf{V}) \leq 1$. The maximum of $\gamma_R(\mathbf{U})(\gamma(\mathbf{V}))$ is attained when $\mathbf{R}\mathbf{e}_i\mathbf{e}_j'(\mathbf{e}_i)$ is in the column (row) space of $\mathbf{X}_0$ for some (i, j). Small values of $\gamma_R(\mathbf{U})$ and $\gamma(\mathbf{V})$ imply that the column and row spaces of $\mathbf{X}_0$ do not contain the columns of $\mathbf{R}$ and sparse vectors, respectively.

Another identifiability issue arises when $\mathbf{X}_0 = \mathbf{R}\mathbf{H}$ for some sparse matrix $\mathbf{H} \in \Omega$. In this case, each column of $\mathbf{X}_0$ is spanned by a few columns of $\mathbf{R}$. Consider the parameter

$$\xi_R(\mathbf{U}, \mathbf{V}) := \|\mathbf{R}'\mathbf{U}\mathbf{V}'\|_\infty = \max_{i,j} |\mathbf{e}_i{}'\mathbf{R}'\mathbf{U}\mathbf{V}\mathbf{e}_j|.$$

A small value of $\xi_R(\mathbf{U}, \mathbf{V})$ implies that each column of $\mathbf{X}_0$ is spanned by sufficiently many columns of $\mathbf{R}$. To understand this property, suppose for simplicity that all nonzero singular values of $\mathbf{X}_0$ are identical and equal to σ, say. The k-th column of $\mathbf{X}_0$ is then $\sum_{i=1}^r \sigma\mathbf{u}_i v_{i,k}$, and its projection onto the l-th column of $\mathbf{R}$ is

$$\Big|\langle\mathbf{R}\mathbf{e}_l, \sum_{i=1}^r \sigma\mathbf{u}_i v_{i,k}\rangle\Big| = \sigma\Big|\sum_{i=1}^r \langle\mathbf{R}\mathbf{e}_l, \mathbf{u}_i\rangle v_{i,k}\Big| \leq \sigma\xi_R(\mathbf{U}, \mathbf{V}).$$

Since the energy of $\sum_{i=1}^r \sigma\mathbf{u}_i v_{i,k}$ is somehow allocated along the directions $\mathbf{R}\mathbf{e}_l$, if all the aforementioned projections can be made arbitrarily small, then sufficiently many nonzero terms in the expansion are needed to account for all this energy.

3.1. Main Result

Theorem 1: *Consider given matrices $\mathbf{Y} \in \mathbb{R}^{L\times T}$ and $\mathbf{R} \in \mathbb{R}^{L\times F}$, obeying $\mathbf{Y} = \mathbf{X}_0 + \mathbf{R}\mathbf{A}_0 = \mathbf{U}\mathbf{\Sigma}\mathbf{V}' + \mathbf{R}\mathbf{A}_0$, where $r := \text{rank}(\mathbf{X}_0)$ and $s := \|\mathbf{A}_0\|_0$. Assume that every row and column of $\mathbf{A}_0$ has at most k nonzero elements, and that $\mathbf{R}$ has orthonormal rows. If $\mu(\Omega_R, \Phi)$, $\delta_k(\mathbf{R})$, $\theta_{1,1}(\mathbf{R})$, $\gamma_R(\mathbf{U})$, $\gamma(\mathbf{V})$ and $\xi_R(\mathbf{U}, \mathbf{V})$ are sufficiently small such that*

I) $\mu(\Omega_R, \Phi) < 1$ and $\delta_k(\mathbf{R}) < 1$

II) $(1 - \mu(\Omega_R, \Phi))^2(1 - \delta_k(\mathbf{R})) > \omega_{max}$

III) $$(1 + \alpha_{max})\left(\frac{1 + \beta_{max}}{1 - \beta_{max}}\right)\xi_R(\mathbf{U}, \mathbf{V})\sqrt{s} + \mu(\Omega_R, \Phi)(1 + \delta_k(\mathbf{R}))^{1/2}(1 + \alpha_{max})\sqrt{r} < 1$$

hold, where

$$\omega_{max} := k\theta_{1,1}(\mathbf{R}) + (1 + \delta_1(\mathbf{R}))\,[k\gamma_R(\mathbf{U}) + s\gamma(\mathbf{V}) + s\gamma_R(\mathbf{U})\gamma(\mathbf{V})]$$

$$\alpha_{max} := \left\{\frac{1}{c(1 - \delta_k(\mathbf{R}))(1 - \mu(\Omega_R, \Phi))^2} - 1\right\}^{1/2}$$

$$\beta_{max} := \frac{1}{(1 - \mu(\Omega_R, \Phi))^2(1 - \delta_k(\mathbf{R}))\omega_{max}^{-1} - 1}$$

then there exists $\lambda > 0$ for which (P1) exactly recovers $\{\mathbf{X}_0, \mathbf{A}_0\}$.

Satisfaction of the conditions in Theorem 1 hinges upon the values of the incoherence parameters $\mu(\Omega_R, \Phi), \gamma_R(\mathbf{U}), \gamma_R(\mathbf{V})$, $\xi_R(\mathbf{U}, \mathbf{V})$, and the RICs $\delta_k(\mathbf{R})$ and $\theta_{1,1}(\mathbf{R})$. In particular, $\{\omega_{\max}, \alpha_{\max}, \beta_{\max}\}$ are increasing functions of these parameters. Therefore, it is readily observed from I)-III) that the smaller these parameters, the easier the conditions are met. Furthermore, the incoherence parameters are increasing functions of the rank r and sparsity level s. The RIC $\delta_k(\mathbf{R})$ is also an increasing function of k, the maximum number of nonzero elements per row/column of $\mathbf{A}_0$.

Therefore, for sufficiently small values of $\{r, s, k\}$, the sufficient conditions of Theorem 1 can be indeed satisfied.

It is worth noting that not only s, but also the position of the nonzero entries in $\mathbf{A}_0$ plays an important role in satisfying I)-III). This is manifested through k, for which a small value indicates that $\mathbf{A}_0$ entries are sufficiently spread out. Moreover, no restriction is placed on the magnitude of these entries, since as seen later on it is only the positions that affect optimal recovery via (P1).

Remark 2 (Row orthonormality of $\mathbf{R}$) The assumption $\mathbf{R}\mathbf{R}' = \mathbf{I}_L$ is equivalent to full-rank $\mathbf{R}$. This is because for full row-rank $\mathbf{R} = \mathbf{U}_R\mathbf{\Sigma}_R\mathbf{V}'_R$, one can pre-multiply both sides of (1) with $\mathbf{\Sigma}_R^{-1}\mathbf{U}'_R$ to obtain $\tilde{\mathbf{R}} := \mathbf{V}'_R$ with orthonormal rows.

Remark 3 (Satisfiability) Albeit I)-III) are not verifiable since e.g., $\delta_k(\mathbf{R})$ is NP-hard to compute [2], finding a class of (possibly random) matrices $\{\mathbf{X}_0, \mathbf{A}_0, \mathbf{R}\}$ satisfying I)-III) is detailed in [7].

4. PROOF OF THE MAIN RESULT

In what follows, conditions are first derived under which $\{\mathbf{X}_0, \mathbf{A}_0\}$ is the *unique* optimal solution of (P1). In essence, these conditions are expressed in terms of certain dual certificates. Then, Section 4.2 deals with the construction of a valid dual certificate.

4.1. Unique Optimality Conditions

Recall the *nonsmooth* optimization problem (P1), and its Lagrangian

$$\mathcal{L}(\mathbf{X}, \mathbf{A}, \mathbf{M}) = \|\mathbf{X}\|_* + \lambda\|\mathbf{A}\|_1 + \langle \mathbf{M}, \mathbf{Y} - \mathbf{X} - \mathbf{R}\mathbf{A}\rangle \qquad (5)$$

where $\mathbf{M} \in \mathbb{R}^{L\times T}$ is the matrix of dual variables(multipliers) associated with the constraint in (P1). From the characterization of the subdifferential for nuclear and ℓ_1-norm (see e.g., [1]), the subdifferential of the Lagrangian at $\{\mathbf{X}_0, \mathbf{A}_0\}$ is given by

$$\partial_{\mathbf{X}}\mathcal{L}(\mathbf{X}_0, \mathbf{A}_0, \mathbf{M}) = \left\{\mathbf{U}\mathbf{V}' + \mathbf{W} - \mathbf{M} : \|\mathbf{W}\| \le 1, \mathcal{P}_\Phi(\mathbf{W}) = \mathbf{0}_{L\times T}\right\}$$
$$\partial_{\mathbf{A}}\mathcal{L}(\mathbf{X}_0, \mathbf{A}_0, \mathbf{M}) = \left\{\lambda \text{sign}(\mathbf{A}_0) + \lambda\mathbf{F} - \mathbf{R}'\mathbf{M} : \|\mathbf{F}\|_\infty \le 1, \mathcal{P}_\Omega(\mathbf{F}) = \mathbf{0}_{F\times T}\right\}.$$

The optimality conditions for (P1) assert that $\{\mathbf{X}_0, \mathbf{A}_0\}$ is an optimal (not necessarily unique) solution if and only if

$$\mathbf{0}_{F\times T} \in \partial_{\mathbf{A}}\mathcal{L}(\mathbf{X}_0, \mathbf{A}_0, \mathbf{M}) \text{ and } \mathbf{0}_{L\times T} \in \partial_{\mathbf{X}}\mathcal{L}(\mathbf{X}_0, \mathbf{A}_0, \mathbf{M}).$$

This can be shown equivalent to finding the pair $\{\mathbf{W}, \mathbf{F}\}$ that satisfies: i) $\|\mathbf{W}\| \le 1$, $\mathcal{P}_\Phi(\mathbf{W}) = \mathbf{0}_{L\times T}$; ii) $\|\mathbf{F}\|_\infty \le 1$, $\mathcal{P}_\Omega(\mathbf{F}) = \mathbf{0}_{F\times T}$; and iii) $\lambda\text{sign}(\mathbf{A}_0) + \lambda\mathbf{F} = \mathbf{R}'(\mathbf{U}\mathbf{V}' + \mathbf{W})$. In general, i)-iii) may hold for multiple solution pairs. However, the next lemma asserts that a slight tightening of the optimality conditions i)-iii) leads to a *unique* optimal solution for (P1).

Lemma 2: *Assume that I) holds. If there exists a dual certificate $\mathbf{\Gamma} \in \mathbb{R}^{L\times T}$ satisfying*
C1) $\mathcal{P}_\Phi(\mathbf{\Gamma}) = \mathbf{U}\mathbf{V}'$
C2) $\mathcal{P}_\Omega(\mathbf{R}'\mathbf{\Gamma}) = \lambda\text{sign}(\mathbf{A}_0)$
C3) $\|\mathcal{P}_{\Phi^\perp}(\mathbf{\Gamma})\| < 1$
C4) $\|\mathcal{P}_{\Omega^\perp}(\mathbf{R}'\mathbf{\Gamma})\|_\infty < \lambda$
then $\{\mathbf{X}_0, \mathbf{A}_0\}$ is the unique optimal solution of (P1).

The remainder of the proof deals with the construction of a dual certificate $\mathbf{\Gamma}$ that meets C1)-C4). To this end, extra conditions [II) and III) in Theorem 1] for the existence of $\mathbf{\Gamma}$ are derived in terms of the incoherence parameters and the RICs. For the special case $\mathbf{R} = \mathbf{I}_L$, the conditions in Lemma 2 boil down to those in [4, Prop. 2] for principal components pursuit. However, the dual certificate construction techniques used in [4] do not carry over to the setting considered here, where a compression matrix $\mathbf{R}$ is present.

4.2. Dual Certificate Construction

Condition C1) in Lemma 2 implies that $\mathbf{\Gamma} = \mathbf{U}\mathbf{V}' + (\mathbf{I} - \mathbf{P}_U)\mathbf{X}(\mathbf{I} - \mathbf{P}_V)$, for arbitrary $\mathbf{X} \in \mathbb{R}^{L\times T}$ (cf. Remark 1). Upon defining $\mathbf{Z} := \mathbf{R}'(\mathbf{I} - \mathbf{P}_U)\mathbf{X}(\mathbf{I} - \mathbf{P}_U)$ and $\mathbf{B}_\Omega := \lambda\text{sign}(\mathbf{A}_0) - \mathcal{P}_\Omega(\mathbf{R}'\mathbf{U}\mathbf{V}')$, C1)-C2) are equivalent to $\mathcal{P}_\Omega(\mathbf{Z}) = \mathbf{B}_\Omega$.

To express this last condition in terms of $\mathbf{X}$, first vectorize $\mathbf{Z}$ to obtain $\text{vec}(\mathbf{Z}) = [(\mathbf{I} - \mathbf{P}_V) \otimes \mathbf{R}'(\mathbf{I} - \mathbf{P}_U)]\,\text{vec}(\mathbf{X})$. Define $\mathbf{A} := (\mathbf{I} - \mathbf{P}_V) \otimes \mathbf{R}'(\mathbf{I} - \mathbf{P}_U)$ and an $s \times LT$ matrix $\mathbf{A}_\Omega$ formed with those s rows of $\mathbf{A}$ associated with those elements in supp$(\mathbf{A}_0)$. Likewise, define $\mathbf{A}_{\Omega^\perp}$ which collects the remaining rows from $\mathbf{A}$ such that $\mathbf{A} = \mathbf{\Pi}[\mathbf{A}'_\Omega, \mathbf{A}'_{\Omega^\perp}]'$ for a suitable row permutation matrix $\mathbf{\Pi}$. Finally, let $\mathbf{b}_\Omega$ be the vector of length s containing those elements of $\mathbf{B}_\Omega$ with indices in supp$(\mathbf{A}_0)$. With these definitions, C1)-C2) can be expressed as $\mathbf{A}_\Omega\text{vec}(\mathbf{X}) = \mathbf{b}_\Omega$.

To upper-bound the left-hand side of C3) in terms of $\mathbf{X}$, use the assumption $\mathbf{R}\mathbf{R}' = \mathbf{I}_L$ to arrive at

$$\|\mathcal{P}_{\Phi^\perp}(\mathbf{\Gamma})\| = \|\mathbf{R}'(\mathbf{I} - \mathbf{P}_U)\mathbf{X}(\mathbf{I} - \mathbf{P}_V)\|$$
$$\le \|\mathbf{R}'(\mathbf{I} - \mathbf{P}_U)\mathbf{X}(\mathbf{I} - \mathbf{P}_V)\|_F = \|\mathbf{A}\text{vec}(\mathbf{X})\|.$$

Similarly, the left-hand side of C4) can be bounded as

$$\|\mathcal{P}_{\Omega^\perp}(\mathbf{R}'\mathbf{\Gamma})\|_\infty = \|\mathcal{P}_{\Omega^\perp}(\mathbf{Z}) + \mathcal{P}_{\Omega^\perp}(\mathbf{R}'\mathbf{U}\mathbf{V}')\|_\infty$$
$$\le \|\mathcal{P}_{\Omega^\perp}(\mathbf{Z})\|_\infty + \|\mathcal{P}_{\Omega^\perp}(\mathbf{R}'\mathbf{U}\mathbf{V}')\|_\infty$$
$$= \|\mathbf{A}_{\Omega^\perp}\text{vec}(\mathbf{X})\|_\infty + \|\mathcal{P}_{\Omega^\perp}(\mathbf{R}'\mathbf{U}\mathbf{V}')\|_\infty.$$

In a nutshell, if one could find $\mathbf{X} \in \mathbb{R}^{L\times T}$ such that
c1) $\mathbf{A}_\Omega\text{vec}(\mathbf{X}) = \mathbf{b}_\Omega$
c2) $\|\mathbf{A}\text{vec}(\mathbf{X})\| < 1$
c3) $\|\mathbf{A}_{\Omega^\perp}\text{vec}(\mathbf{X})\|_\infty + \|\mathcal{P}_{\Omega^\perp}(\mathbf{R}'\mathbf{U}\mathbf{V}')\|_\infty < \lambda$
hold for some positive λ, then C1)-C4) are satisfied as well.

The final steps of the proof entail: i) finding an appropriate candidate solution $\hat{\mathbf{X}}$ such that c1) holds; and ii) deriving conditions in terms of the incoherence parameters and RICs that guarantee $\hat{\mathbf{X}}$ meets the required bounds in c2) and c3) for a range of λ values. The following lemma is instrumental to accomplishing i).

Lemma 3: *If I) holds, matrix $\mathbf{A}_\Omega$ has full row rank, and its minimum singular value is bounded below by*

$$\sigma_{min}(\mathbf{A}_\Omega) \ge c^{1/2}(1 - \delta_k(\mathbf{R}))^{1/2}(1 - \mu(\Omega_R, \Phi)).$$

According to Lemma 3, the least-norm (LN) solution $\hat{\mathbf{X}}_{\text{LN}} := \arg\min_{\mathbf{X}}\left\{\|\mathbf{X}\|_F^2 : \mathbf{A}_\Omega\text{vec}(\mathbf{X}) = \mathbf{b}_\Omega\right\}$ exists, and is given by

$$\text{vec}(\hat{\mathbf{X}}_{\text{LN}}) = \mathbf{A}'_\Omega\left(\mathbf{A}_\Omega\mathbf{A}'_\Omega\right)^{-1}\mathbf{b}_\Omega. \qquad (6)$$

Remark 4 (Candidate dual certificate) From the arguments at the beginning of this section, the candidate dual certificate is $\hat{\mathbf{\Gamma}} = \mathbf{U}\mathbf{V}' + (\mathbf{I} - \mathbf{P}_U)\hat{\mathbf{X}}_{\text{LN}}(\mathbf{I} - \mathbf{P}_V)$.

The LN solution is an attractive choice, since it facilitates satisfying c2) and c3) which require norms of $\text{vec}(\mathbf{X})$ to be small. Plugging the LN solution (6) in the left-hand side of c2) yields (define $\mathbf{Q} :=$

$\mathbf{A}_{\Omega^\perp}\mathbf{A}'_\Omega\left(\mathbf{A}_\Omega\mathbf{A}'_\Omega\right)^{-1}$ for notational brevity)

$$\begin{aligned}\|\mathbf{A}\text{vec}(\hat{\mathbf{X}}_{\text{LN}})\| &= \left\|\begin{pmatrix}\mathbf{A}_\Omega\\ \mathbf{A}_{\Omega^\perp}\end{pmatrix}\mathbf{A}'_\Omega\left(\mathbf{A}_\Omega\mathbf{A}'_\Omega\right)^{-1}\mathbf{b}_\Omega\right\|\\ &= \left\|\begin{pmatrix}\mathbf{I}\\ \mathbf{Q}\end{pmatrix}\mathbf{b}_\Omega\right\| \leq (1+\|\mathbf{Q}\|)\,\|\mathbf{b}_\Omega\|. \quad (7)\end{aligned}$$

Moreover, substituting (6) in the left-hand side of c3) results in

$$\begin{aligned}&\|\mathbf{Q}\mathbf{b}_\Omega\|_\infty + \|\mathcal{P}_{\Omega^\perp}(\mathbf{R}'\mathbf{U}\mathbf{V}')\|_\infty\\ &\quad\leq \|\mathbf{Q}\|_{\infty,\infty}\|\mathbf{b}_\Omega\|_\infty + \|\mathcal{P}_{\Omega^\perp}(\mathbf{R}'\mathbf{U}\mathbf{V}')\|_\infty. \quad (8)\end{aligned}$$

Next, upper-bounds are obtained for $\|\mathbf{Q}\|$ and $\|\mathbf{Q}\|_{\infty,\infty}$.

Lemma 4: *If I) holds, then*

$$\|\mathbf{Q}\| \leq \alpha_{\max} := \left\{\frac{1}{c(1-\delta_k(\mathbf{R}))(1-\mu(\Omega_R,\Phi))^2} - 1\right\}^{1/2}.$$

If II) holds as well, then

$$\|\mathbf{Q}\|_{\infty,\infty} \leq \beta_{\max} := \frac{\omega_{max}}{(1-\mu(\Omega_R,\Phi))^2(1-\delta_k(\mathbf{R}))-\omega_{max}}.$$

Going back to (7)-(8), note that $\|\mathbf{B}_\Omega\|_\infty = \|\mathbf{b}_\Omega\|_\infty$ and $\|\mathbf{B}_\Omega\|_F = \|\mathbf{b}_\Omega\|$, which can be respectively upper-bounded as

$$\begin{aligned}\|\mathbf{B}_\Omega\|_\infty &= \|\lambda\text{sign}(\mathbf{A}_0) - \mathcal{P}_\Omega(\mathbf{R}'\mathbf{U}\mathbf{V}')\|_\infty\\ &\leq \lambda + \|\mathcal{P}_\Omega(\mathbf{R}'\mathbf{U}\mathbf{V}')\|_\infty \quad (9)\\ \|\mathbf{B}_\Omega\|_F &= \|\lambda\text{sign}(\mathbf{A}_0) - \mathcal{P}_\Omega(\mathbf{R}'\mathbf{U}\mathbf{V}')\|_F\\ &\leq \lambda\sqrt{s} + \|\mathcal{P}_\Omega(\mathbf{R}'\mathbf{U}\mathbf{V}')\|_F. \quad (10)\end{aligned}$$

It is shown that $\|\mathcal{P}_\Omega(\mathbf{R}'\mathbf{U}\mathbf{V}')\|_F \leq \sqrt{r}\mu(\Omega_R,\Phi)c^{1/2}(1+\delta_k(\mathbf{R}))^{1/2}$ [7], and plugging this bound back into (10) one arrives at

$$\|\mathbf{B}_\Omega\|_F \leq \lambda\sqrt{s} + \sqrt{r}\mu(\Omega_R,\Phi)c^{1/2}(1+\delta_k(\mathbf{R}))^{1/2}. \quad (11)$$

Upon substituting (9), (11) and the bounds in Lemma 4 into (7) and (8), one finds that c2) and c3) hold if there exists $\lambda > 0$ such that

$$(1+\alpha_{\max})\left\{\lambda\sqrt{s} + \sqrt{r}\mu(\Omega_R,\Phi)c^{1/2}(1+\delta_k(\mathbf{R}))^{1/2}\right\} < 1 \quad (12a)$$

$$\beta_{\max}\left(\lambda + \|\mathcal{P}_\Omega(\mathbf{R}'\mathbf{U}\mathbf{V}')\|_\infty\right) + \|\mathcal{P}_{\Omega^\perp}(\mathbf{R}'\mathbf{U}\mathbf{V}')\|_\infty < \lambda \quad (12b)$$

hold. Recognizing that $\xi_R(\mathbf{U},\mathbf{V}) = \max\{\|\mathcal{P}_\Omega(\mathbf{R}'\mathbf{U}\mathbf{V}')\|_\infty, \|\mathcal{P}_{\Omega^\perp}(\mathbf{R}'\mathbf{U}\mathbf{V}')\|_\infty\}$, the left-hand side of (12b) can be further bounded. After straightforward manipulations, one deduces that conditions (12a) and (12b) are satisfied for $\lambda \in (\lambda_{\min}, \lambda_{\max})$, where

$$\begin{aligned}\lambda_{\min} &:= \frac{1+\beta_{\max}}{1-\beta_{\max}}\xi_R(\mathbf{U},\mathbf{V})\\ \lambda_{\max} &:= \frac{1}{\sqrt{s}}\left\{(1+\alpha_{\max})^{-1} - \sqrt{r}\mu(\Omega_R,\Phi)c^{1/2}(1+\delta_k(\mathbf{R}))^{1/2}\right\}\end{aligned}$$

Clearly, one still needs to ensure $\lambda_{\max} > \lambda_{\min}$ so that the LN solution (6) meets the requirements c1)-c3) [equivalently, $\hat{\mathbf{\Gamma}}$ in Remark 4 satisfies C1)-C4) from Lemma 2]. Condition $\lambda_{\max} > \lambda_{\min}$ is equivalent to III) in Theorem 1, and the proof is now complete.

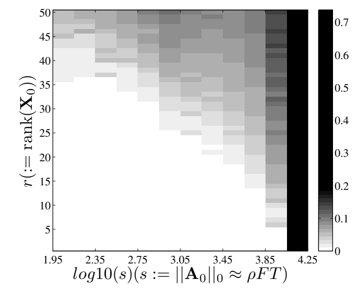

Fig. 1. Relative error $e_r := \|\mathbf{A}_0 - \hat{\mathbf{A}}\|_F/\|\mathbf{A}_0\|_F$ for various values of r and s where $L = 105$, $F = 210$, and $T = 420$. White represents exact recovery ($e_r \approx 0$), while black represents $e_r \approx 1$.

5. NUMERICAL TESTS

To test the exact recovery claims, a data matrix was simulated according to $\mathbf{Y} = \mathbf{X}_0 + \mathbf{V}'_R\mathbf{A}_0$, where $\mathbf{V}_R \in \mathbb{R}^{F\times L}$ contain the right singular vectors of the random matrix $\mathbf{R} = \mathbf{U}_R\mathbf{\Sigma}_R\mathbf{V}'_R$ with entries being Bernoulli distributed with parameter $1/2$. Low-rank matrices are generated as $\mathbf{X}_0 = \mathbf{V}'_R\mathbf{W}_1\mathbf{W}'_2$, where $\mathbf{W}_1 \in \mathbb{R}^{F\times r}$ and $\mathbf{W}_2 \in \mathbb{R}^{T\times r}$ contain i.i.d. zero-mean Gaussian entries of variance $1/\sqrt{FT}$. Every entry of $\mathbf{A}_0$ is drawn independently at random from the set $\{-1,0,1\}$, with $\Pr(a_{ij}=-1) = \Pr(a_{ij}=1) = \rho/2$. Also, set $L = 105$, $F = 210$, and $T = 420$.

To demonstrate that (P1) is capable of recovering the exact values of $\{\mathbf{X}_0, \mathbf{A}_0\}$, the optimization problem is solved for a wide range of values of r and s. Let $\hat{\mathbf{A}}$ represent the solution obtained by (P1) for a suitable value of λ. Fig. 1 shows the relative error in recovering $\mathbf{A}_0$, namely $\|\hat{\mathbf{A}} - \mathbf{A}_0\|_F/\|\mathbf{A}_0\|_F$ for various values of r and s. It is apparent that (P1) succeeds in recovering $\mathbf{A}_0$ for sufficiently sparse $\mathbf{A}_0$ and low-rank $\mathbf{X}_0$ from the observed data $\mathbf{Y}$. A similar trend is observed for the recovery of $\mathbf{X}_0$. A corroborating plot is omitted here due to space limitations.

6. REFERENCES

[1] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[2] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Info. Theory*, vol. 51, pp. 4203–4215, 2005.

[3] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. of the ACM*, vol. 58, pp. 1–37, 2011.

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

[5] F. Deutsch, *Best Approximation in Inner Product Spaces*, Springer-Verlag, second edition, 2001.

[6] A. Lakhina, M. Crovella, and C. Diot, "Diagnosing network-wide traffic anomalies," in *Proc. of ACM SIGCOMM*, Portland, OR, Aug. 2004.

[7] M. Mardani, G. Mateos, and G. B. Giannakis, "Recovery of low-Rank plus compressed sparse matrices with application to unveiling traffic anomalies," *IEEE Trans. Info. Theory*, 2012, see also arXiv:1204.6537v1 [cs.IT].

[8] A. C. Sankaranarayanan, A. E. Waters, and R. G. Baraniuk, "SpaRCS: Recovering low-rank and sparse matrices from compressive measurements," in *Proc. of Neural Info. Proc. Systems*, Granada, Spain, Dec. 2011.