Robust Matrix Decomposition with Sparse Corruptions

Daniel Hsu, Sham M. Kakade, and Tong Zhang

***Abstract*—Suppose a given observation matrix can be decomposed as the sum of a low-rank matrix and a sparse matrix, and the goal is to recover these individual components from the observed sum. Such additive decompositions have applications in a variety of numerical problems including system identification, latent variable graphical modeling, and principal components analysis. We study conditions under which recovering such a decomposition is possible via a combination of ℓ_1 norm and trace norm minimization. We are specifically interested in the question of how many sparse corruptions are allowed so that convex programming can still achieve accurate recovery, and we obtain stronger recovery guarantees than previous studies. Moreover, we do not assume that the spatial pattern of corruptions is random, which stands in contrast to related analyses under such assumptions via matrix completion.**

***Index Terms*—Matrix decompositions, sparsity, low-rank, outliers**

I. Introduction

THIS work studies additive decompositions of matrices into sparse and low-rank components. Such decompositions have found applications in a variety of numerical problems, including system identification [1], latent variable graphical modeling [2], and principal component analysis (PCA) [3]. In these settings, the user has an input matrix $Y \in \mathbb{R}^{m \times n}$ which is believed to be the sum of a sparse matrix X_S and a low-rank matrix X_L. For instance, in the application to PCA, X_L represents a matrix of m data points from a low-dimensional subspace of $\mathbb{R}^n$, and is corrupted by a sparse matrix X_S of errors before being observed as

$$Y \quad = \quad \underset{\text{(sparse)}}{X_S} \quad + \quad \underset{\text{(low-rank)}}{X_L}.$$

The goal is to recover the original data matrix X_L (and the error components X_S) from the corrupted observations Y. In the latent variable model application of Chandrasekaran *et al.* [2], Y represents the precision matrix over visible nodes of a Gaussian graphical model, and X_S represents the precision matrix over the visible nodes when conditioned on the hidden nodes. In general, Y may be dense as a result of dependencies between visible nodes through the hidden nodes. However, X_S will be sparse when the visible nodes are mostly independent after conditioning on the hidden nodes, and the difference $X_L = Y - X_S$ will be low-rank when the number of hidden nodes is small. The goal is then to infer the relevant dependency structure from just the visible nodes and measurements of their correlations.

Even if the matrix Y is exactly the sum of a sparse matrix X_S and a low-rank matrix X_L, it may be impossible to identify these components from the sum. For instance, the sparse matrix X_S may be low-rank, or the low-rank matrix X_L may be sparse. In such cases, these components may be confused for each other, and thus the desired decomposition of Y may not be identifiable. Therefore, one must impose conditions on the sparse and low-rank components in order to guarantee their identifiability from Y.

We present sufficient conditions under which X_S and X_L are identifiable from the sum Y. Essentially, we require that X_S not be too dense in any single row or column, and that the singular vectors of X_L not be too sparse. The level of denseness and sparseness are considered jointly in the conditions in order to obtain the weakest possible conditions. Under a mild strengthening of the condition, we also show that X_S and X_L can be recovered by solving certain convex programs, and that the solution is robust under small perturbations of Y. The first program we consider is

$$\min \ \lambda \|X_S\|_{\text{vec}(1)} + \|X_L\|_*$$

D. Hsu is with Microsoft Research New England, Cambridge, MA 02142 USA (e-mail: dahsu@microsoft.com).

S. M. Kakade is with Microsoft Research New England, Cambridge, MA 02142 USA, and also with the Department of Statistics, Wharton School, University of Pennsylvania, Philadelphia, PA 19104 (e-mail: skakade@wharton.upenn.edu). This author was partially supported by NSF grant IIS-1018651.

T. Zhang is with the Department of Statistics, Rutgers University (e-mail: tzhang@stat.rutgers.edu). This author was partially supported by the following grants: AFOSR FA9550-09-1-0425, NSA-AMS 081024, NSF DMS-1007527, and NSF IIS-1016061.

(subject to certain feasibility constraints such as $\|X_S + X_L - Y\| \leq \epsilon$), where $\|\cdot\|_{\text{vec}(1)}$ is the entry-wise 1-norm and $\|\cdot\|_*$ is the trace norm. These norms are natural convex surrogates for the sparsity of X_S and the rank of X_L [4], [5], which are generally intractable to optimize. We also consider a regularized formulation

$$\min \ \frac{1}{2\mu}\|X_S + X_L - Y\|_{\text{vec}(2)}^2 + \lambda\|X_S\|_{\text{vec}(1)} + \|X_L\|_*$$

where $\|\cdot\|_{\text{vec}(2)}$ is the Frobenius norm; this formulation may be more suitable in certain applications and enjoys different recovery guarantees.

A. Related work

Our work closely follows that of Chandrasekaran *et al*. [1], who initiated the study of rank-sparsity incoherence and its application to matrix decompositions. There, the authors identify parameters that characterize the incoherence of X_S and X_L sufficient to guarantee identifiability and recovery using convex programs. However, their analysis of this characterization yields conditions that are significantly stronger than those given in our present work. For instance, the allowed fraction of non-zero entries in X_S is quickly vanishing as a function of the matrix size, even under the most favorable conditions on X_L; our analysis does not have this restriction and allows X_S to have up to $\Omega(mn)$ non-zero entries when X_L is low-rank and has non-sparse singular vectors. In terms of the PCA application, our analysis allows for up to a constant fraction of the data matrix entries to be corrupted by noise of arbitrary magnitude, while the analysis of [1] requires that it decrease as a function of the matrix dimensions. Moreover, [1] only considers exact decompositions, which may be unrealistic in certain applications; we allow for approximate decompositions, and study the effect of perturbations on the accuracy of the recovered components.

The application to principal component analysis with gross sparse errors was studied by Candès *et al*. [3], building on previous results and analysis techniques for the related matrix completion problem (*e.g.*, [6], [7]). The sparse errors model of [3] requires that the support of the sparse matrix X_S be random, which can be unrealistic in some settings. However, the conditions are significantly weaker than those of [1]: for instance, they allow for $\Omega(mn)$ non-zero entries in X_S. Our work makes no probabilistic assumption on the sparsity pattern of X_S and instead studies purely deterministic structural conditions. The price we pay, however, is roughly a factor of $\text{rank}(X_L)$ in what is allowed for the support size of X_S (relative to the probabilistic analysis of [3]). Narrowing this gap with alternative deterministic conditions is an interesting open problem. Follow-up work to [3] studies the robustness of the recovery procedure [8], as well as quantitatively weaker conditions on X_S [9], but these works are only considered under the random support model. Our work is therefore largely complementary to these probabilistic analyses.

B. Outline

We describe our main results in Section II. In Section III, we review a number of technical tools such as matrix and operator norms that are used to characterize the rank-sparsity incoherence properties of the desired decomposition. Section IV analyzes these incoherence properties in detail, giving sufficient conditions for identifiability as well as for certifying the (approximate) optimality of a target decomposition for our optimization formulations. The main recovery guarantees are proved in Sections V and VI.

II. Main results

Fix an observation matrix $Y \in \mathbb{R}^{m\times n}$. Our goal is to (approximately) decompose the matrix Y into the sum of a sparse matrix X_S and a low-rank matrix X_L.

A. Optimization formulations

We consider two convex optimization problems over $(X_S, X_L) \in \mathbb{R}^{m\times n} \times \mathbb{R}^{m\times n}$. The first is the constrained formulation (parametrized by $\lambda > 0$, $\epsilon_{\text{vec}(1)} \geq 0$, and $\epsilon_* \geq 0$)

$$\begin{array}{ll} \min & \lambda\|X_S\|_{\text{vec}(1)} + \|X_L\|_* \\ \text{s.t.} & \|X_S + X_L - Y\|_{\text{vec}(1)} \leq \epsilon_{\text{vec}(1)} \\ & \|X_S + X_L - Y\|_* \leq \epsilon_* \end{array} \tag{1}$$

where $\|\cdot\|_{\text{vec}(1)}$ is the entry-wise 1-norm, and $\|\cdot\|_*$ is the trace norm (*i.e.*, sum of singular values). The second is the regularized formulation (with regularization parameter $\mu > 0$)

$$\min \frac{1}{2\mu}\|X_S + X_L - Y\|_{\text{vec}(2)}^2 + \lambda\|X_S\|_{\text{vec}(1)} + \|X_L\|_* \tag{2}$$

where $\|\cdot\|_{\text{vec}(2)}$ is the Frobenius norm (entry-wise 2-norm).

We also consider adding a constraint to control $\|X_L\|_{\text{vec}(\infty)}$, the entry-wise ∞-norm of X_L. To (1), we add the constraint

$$\|X_L\|_{\text{vec}(\infty)} \leq b$$

and to (2), we add

$$\|X_S - Y\|_{\text{vec}(\infty)} \leq b$$

The parameter b is intended as a natural bound for X_L and is typically known in applications. For example, in image processing, the values of interest may lie in the interval $[0, 255]$ (say), and hence, we might take $b = 500$ as a relaxation of the box constraint $[0, 255]$. The core of our analyses do not rely on these additional constraints; we only consider them to obtain improved robustness guarantees for recovering X_L, which may be important in some applications.

B. Identifiability conditions

Our first result is a refinement of the *rank-sparsity incoherence* notion developed by [1]. We characterize a target decomposition of Y into $Y = \bar{X}_S + \bar{X}_L$ by the projection operators to subspaces associated with $\bar{X}_S$ and $\bar{X}_L$. Let

$$\bar{\Omega} \;=\; \Omega(\bar{X}_S) \;:=\; \{X \in \mathbb{R}^{m\times n} : \mathrm{supp}(X) \subseteq \mathrm{supp}(\bar{X}_S)\}$$

be the space of matrices whose supports are subsets of the support of $\bar{X}_S$, and let $\mathcal{P}_{\bar{\Omega}}$ be the orthogonal projector to $\bar{\Omega}$ under the inner product $\langle A, B\rangle = \mathrm{tr}(A^\top B)$; this projection is given by

$$[\mathcal{P}_{\bar{\Omega}}(M)]_{i,j} = \begin{cases} M_{i,j} & \text{if } (i,j) \in \mathrm{supp}(\bar{X}_S) \\ 0 & \text{otherwise} \end{cases}$$

for all $i \in [m] := \{1, \ldots, m\}, j \in [n] := \{1, \ldots, n\}$. Furthermore, let

$$\begin{aligned} \bar{T} \;=\; T(\bar{X}_L) \;:=\; \\ \{X_1 + X_2 \in \mathbb{R}^{m\times n} : \mathrm{range}(X_1) \subseteq \mathrm{range}(\bar{X}_L), \\ \mathrm{range}(X_2^\top) \subseteq \mathrm{range}(\bar{X}_L^\top)\} \end{aligned}$$

be the span of matrices either with row-space contained in that of $\bar{X}_L$, or with column-space contained in that of $\bar{X}_L$. Let $\mathcal{P}_{\bar{T}}$ be the orthogonal projector to $\bar{T}$, again, under the inner product $\langle A, B\rangle = \mathrm{tr}(A^\top B)$; this projection is given by

$$\mathcal{P}_{\bar{T}}(M) = \bar{U}\bar{U}^\top M + M\bar{V}\bar{V}^\top - \bar{U}\bar{U}^\top M\bar{V}\bar{V}^\top$$

where $\bar{U} \in \mathbb{R}^{m\times\bar{r}}$ and $\bar{V} \in \mathbb{R}^{n\times\bar{r}}$ are, respectively, matrices of left and right orthonormal singular vectors corresponding to the non-zero singular values of $\bar{X}_L$, and $\bar{r}$ is the rank of $\bar{X}_L$. We will see that certain operator norms of $\mathcal{P}_{\bar{\Omega}}$ and $\mathcal{P}_{\bar{T}}$ can be bounded in terms of structural properties of $\bar{X}_S$ and $\bar{X}_L$.

The first property measures the *maximum number of non-zero entries in any row or column of* $\bar{X}_S$:

$$\alpha(\rho) \;:=\; \max\bigl\{\rho\|\mathrm{sign}(\bar{X}_S)\|_{1\to 1}, \; \rho^{-1}\|\mathrm{sign}(\bar{X}_S)\|_{\infty\to\infty}\bigr\}$$

where $\|M\|_{p\to q} := \max\{\|Mv\|_q : v \in \mathbb{R}^n, \|v\|_p \leq 1\}$,

$$\mathrm{sign}(M)_{i,j} = \begin{cases} -1 & \text{if } M_{i,j} < 0 \\ 0 & \text{if } M_{i,j} = 0 \\ +1 & \text{if } M_{i,j} > 0 \end{cases} \quad \forall i \in [m], j \in [n]$$

and $\rho > 0$ is a balancing parameter to accommodate disparity between the number of rows and columns; a natural choice for the balancing parameter is $\rho := \sqrt{n/m}$. We remark that ρ is only a parameter for the analysis; the optimization formulations do not directly involve ρ. Note that $\bar{X}_S$ may have $\Omega(mn)$ non-zero entries and $\alpha(\sqrt{n/m}) = O(\sqrt{mn})$ as long as the non-zero entries of $\bar{X}_S$ are spread out over the entire matrix. Conversely, a sparse matrix with just $O(m + n)$ could have $\alpha(\sqrt{n/m}) = \sqrt{mn}$ by having all of its non-zero entries in just a few rows and columns.

The second property measures the *sparseness of the singular vectors of* $\bar{X}_L$:

$$\begin{aligned} \beta(\rho) \;:=\; \rho^{-1}\|\bar{U}\bar{U}^\top\|_{\mathrm{vec}(\infty)} + \rho\|\bar{V}\bar{V}^\top\|_{\mathrm{vec}(\infty)} \\ + \|\bar{U}\|_{2\to\infty}\|\bar{V}\|_{2\to\infty}. \end{aligned}$$

For instance, if the singular vectors of $\bar{X}_L$ are perfectly aligned with the coordinate axes, then $\beta(\rho) = \Omega(1)$. On the other hand, if the left and right singular vectors have entries bounded by $\sqrt{c/m}$ and $\sqrt{c/n}$, respectively, for some $c \geq 1$, then $\beta(\sqrt{n/m}) \leq 3c\bar{r}/\sqrt{mn}$.

Our main identifiability result is the following.

Theorem 1. *If* $\inf_{\rho>0} \alpha(\rho)\beta(\rho) < 1$*, then* $\bar{\Omega} \cap \bar{T} = \{0\}$.

Theorem 1 is an immediate consequence of the following lemma (also given as Lemma 10).

Lemma 1. *For all* $M \in \mathbb{R}^{m\times n}$*,* $\|\mathcal{P}_{\bar{\Omega}}(\mathcal{P}_{\bar{T}}(M))\|_{\mathrm{vec}(1)} \leq \inf_{\rho>0} \alpha(\rho)\beta(\rho)\|M\|_{\mathrm{vec}(1)}$.

Proof of Theorem 1: Take any $M \in \bar{\Omega} \cap \bar{T}$. By Lemma 1, $\|\mathcal{P}_{\bar{\Omega}}(\mathcal{P}_{\bar{T}}(M))\|_{\mathrm{vec}(1)} \leq \alpha(\rho)\beta(\rho)\|M\|_{\mathrm{vec}(1)}$. On the other hand, $\mathcal{P}_{\bar{\Omega}}(\mathcal{P}_{\bar{T}}(M)) = M$, so $\alpha(\rho)\beta(\rho) < 1$ implies $\|M\|_{\mathrm{vec}(1)} = 0$, *i.e.*, $M = 0$. ■

Clearly, if $\bar{\Omega} \cap \bar{T}$ contains a matrix other than 0, then $\{(\bar{X}_S + M, \bar{X}_L - M) : M \in \bar{\Omega} \cap \bar{T}\}$ gives a family of sparse/low-rank decompositions of $Y = \bar{X}_S + \bar{X}_L$ with at least the same sparsity and rank as $(\bar{X}_S, \bar{X}_L)$. Conversely, if $\bar{\Omega} \cap \bar{T} = \{0\}$, then any matrix in the direct sum $\bar{\Omega} \oplus \bar{T}$ has exactly one decomposition into a matrix $A \in \bar{\Omega}$ plus a matrix $B \in \bar{T}$, and in this sense $(\bar{X}_S, \bar{X}_L)$ is identifiable.

Note that, as we have argued above, the condition $\inf_{\rho>0} \alpha(\rho)\beta(\rho) < 1$ may be achieved even by matrices $\bar{X}_S$ with $\Omega(mn)$ non-zero entries, provided that the non-zero entries of $\bar{X}_S$ are sufficiently spread out, and that $\bar{X}_L$ is low-rank and has singular vectors far from the coordinate basis. This is in contrast with the conditions studied by [1]. Their analysis uses a different characterization of $\bar{X}_S$ and $\bar{X}_L$, which leads to a stronger identifiability condition in certain cases. Roughly, if $\bar{X}_S$ has an approximately symmetric sparsity pattern (so $\|\operatorname{sign}(\bar{X}_S)\|_{1\to 1} \approx \|\operatorname{sign}(\bar{X}_S)\|_{\infty\to\infty}$), then [1] requires $\alpha(1)\sqrt{\beta(1)} < 1$ for square $n \times n$ matrices.[1] Since $\beta(1) = \Omega(1/n)$ for any $\bar{X}_L \in \mathbb{R}^{n\times n}$, the condition implies $\alpha(1)^2 = O(n)$. Therefore $\bar{X}_S$ must have at most $O(n)$ non-zero entries (or else $\alpha(1)^2$ becomes super-linear). In other words, the fraction of non-zero entries allowed in $\bar{X}_S$ by the condition $\alpha(1)\sqrt{\beta(1)} < 1$ is quickly vanishing as a function of n.

C. Recovery guarantees

Our next results are guarantees for (approximately) recovering the sparse/low-rank decomposition $(\bar{X}_S, \bar{X}_L)$ from $Y = \bar{X}_S + \bar{X}_L$ via solving either convex optimization problems (1) or (2). We require a mild strengthening of the condition $\inf_{\rho>0} \alpha(\rho)\beta(\rho) < 1$, as well as appropriate settings of $\lambda > 0$ and $\mu > 0$ for our recovery guarantees. Before continuing, we first define another property of $\bar{X}_L$:

$$\gamma := \|\bar{U}\bar{V}^\top\|_{\mathrm{vec}(\infty)}$$

which is approximately the same as (in fact, bounded above by) the third term in the definition of $\beta(\rho)$. The quantities $\alpha(\rho)$, $\beta(\rho)$, and γ are central to our analysis. Therefore we state the following proposition for reference, which provides a more intuitive understanding of their behavior. We note that this is the only part in which any explicit dimensional dependencies comes into our analysis.

Proposition 1. *Let m_0 be the maximum number of non-zero entries of $\bar{X}_S$ per column and n_0 be the maximum number of non-zero entries of $\bar{X}_S$ per row. Let $\bar{r}$ be the rank of $\bar{U}$ and $\bar{V}$. Assume further that $m_0 \le c_1 m/\bar{r}$ and $n_0 \le c_1 n/\bar{r}$ for some $c_1 \in (0,1)$, and $\|\bar{U}\|_{\mathrm{vec}(\infty)} \le \sqrt{c_2/m}$ and $\|\bar{V}\|_{\mathrm{vec}(\infty)} \le \sqrt{c_2/n}$ for some $c_2 > 0$. Then with $\rho = \sqrt{n/m}$, we have*

$$\alpha(\rho) \le \frac{c_1}{\bar{r}}\sqrt{mn}, \qquad \beta(\rho) \le \frac{3c_2\bar{r}}{\sqrt{mn}}, \qquad \gamma \le \frac{c_2\bar{r}}{\sqrt{mn}}.$$

[1] [1] does not explicitly work out the non-square case, but claims that n can be replaced in their analysis by the larger matrix dimension $\max\{m,n\}$. However this does not seem possible, and the analysis there should only lead to the quite suboptimal dimensionality dependency $\min\{m,n\}$. This is because a rectangular matrix $\bar{X}_L$ will have left and right singular vectors of different dimensions and thus different allowable ranges of infinity norms.

We now proceed with conditions for the regularized formulation (2). Let $E := Y - (\bar{X}_S + \bar{X}_L)$ and

$$\begin{aligned} \epsilon_{2\to 2} &:= \|E\|_{2\to 2} \\ \epsilon_{\mathrm{vec}(\infty)} &:= \|E\|_{\mathrm{vec}(\infty)} + \|\mathcal{P}_{\bar{T}}(E)\|_{\mathrm{vec}(\infty)}. \end{aligned}$$

We require the following, for some $\rho > 0$ and $c > 1$:

$$\alpha(\rho)\beta(\rho) < 1 \tag{3}$$

$$\lambda \le \frac{(1-\alpha(\rho)\beta(\rho))(1-c\cdot\mu^{-1}\epsilon_{2\to 2})}{c\cdot\alpha(\rho)} - \frac{\alpha(\rho)\mu^{-1}\epsilon_{\mathrm{vec}(\infty)} + \alpha(\rho)\gamma}{\alpha(\rho)} \tag{4}$$

$$\lambda \ge c\cdot\frac{\gamma + \mu^{-1}(2-\alpha(\rho)\beta(\rho))\epsilon_{\mathrm{vec}(\infty)}}{1-\alpha(\rho)\beta(\rho) - c\cdot\alpha(\rho)\beta(\rho)} > 0. \tag{5}$$

For instance, if for some $\rho > 0$,

$$\alpha(\rho)\gamma \le \frac{1}{41} \quad\text{and}\quad \alpha(\rho)\beta(\rho) \le \frac{3}{41}, \tag{6}$$

then the conditions are satisfied for $c = 2$ provided that μ and λ are chosen to satisfy

$$\mu \ge \max\left\{4\cdot\epsilon_{2\to 2},\ \frac{2}{15}\cdot\frac{\epsilon_{\mathrm{vec}(\infty)}}{\lambda}\right\} \quad\text{and}\quad \frac{15}{2}\cdot\gamma \le \lambda \le \frac{15}{82}\cdot\frac{1}{\alpha(\rho)}. \tag{7}$$

Note that (6) can be satisfied when $c_1 \le c_2^{-1}/41$ in Proposition 1.

For the constrained formulation (1), our analysis requires the same conditions as above, except with E set to 0. Note that our analysis still allows for approximate decompositions; it is only the conditions that are formulated with $E = 0$. Specifically, we require for some $\rho > 0$ and $c > 1$:

$$\alpha(\rho)\beta(\rho) < 1 \tag{8}$$

$$\lambda \le \frac{1-\alpha(\rho)\beta(\rho) - c\cdot\alpha(\rho)\gamma}{c\cdot\alpha(\rho)} \tag{9}$$

$$\lambda \ge c\cdot\frac{\gamma}{1-\alpha(\rho)\beta(\rho) - c\cdot\alpha(\rho)\beta(\rho)} > 0. \tag{10}$$

For instance, if for some $\rho > 0$,

$$\alpha(\rho)\gamma \le \frac{1}{15} \quad\text{and}\quad \alpha(\rho)\beta(\rho) \le \frac{1}{5}, \tag{11}$$

then the conditions are satisfied for $c = 2$ provided that λ is chosen to satisfy

$$5\gamma \le \lambda \le \frac{1}{3\alpha(\rho)}. \tag{12}$$

Note that (11) can be satisfied when $c_1 \le c_2^{-1}/15$ in Proposition 1.

In summary, Proposition 1 shows that our results can be applied even with $m_0 = \Omega(m/\bar{r})$ and $n_0 = \Omega(n/\bar{r})$ corruptions. In contrast, the results of [1] only apply under the condition $\max(m_0, n_0) = O(\sqrt{\min(m,n)/\bar{r}})$, which is significantly stronger. Moreover, unlike the analysis of [3], we do not have to assume that $\mathrm{supp}(\bar{X}_S)$ is random.

The following theorem gives our recovery guarantee for the constrained formulation (1).

Theorem 2. *Fix a target pair* $(\bar{X}_S, \bar{X}_L) \in \mathbb{R}^{m\times n} \times \mathbb{R}^{m\times n}$ *satisfying* $\|Y - (\bar{X}_S + \bar{X}_L)\|_{\mathrm{vec}(1)} \le \epsilon_{\mathrm{vec}(1)}$ *and* $\|Y - (\bar{X}_S + \bar{X}_L)\|_* \le \epsilon_*$. *Assume the conditions* (8)*,* (9)*, and* (10) *hold for some* $\rho > 0$ *and* $c > 1$. *Let* $(\hat{X}_S, \hat{X}_L) \in \mathbb{R}^{m\times n}$ *be the solution to the convex optimization problem* (1)*. We have*

$$\begin{aligned}\max\left\{\|\hat{X}_S - \bar{X}_S\|_{\mathrm{vec}(1)},\ \|\hat{X}_L - \bar{X}_L\|_{\mathrm{vec}(1)}\right\} \\ \le \left(1 + (1-1/c)^{-1}\cdot\frac{2-\alpha(\rho)\beta(\rho)}{1-\alpha(\rho)\beta(\rho)}\right)\cdot\epsilon_{\mathrm{vec}(1)} \\ + (1-1/c)^{-1}\cdot\frac{2-\alpha(\rho)\beta(\rho)}{1-\alpha(\rho)\beta(\rho)}\cdot\epsilon_*/\lambda.\end{aligned}$$

If, in addition for some $b \ge \|\bar{X}_L\|_{\mathrm{vec}(\infty)}$*, either:*

- *the optimization problem* (1) *is augmented with the constraint* $\|X_L\|_{\mathrm{vec}(\infty)} \le b$ *(and letting* $\widetilde{X}_L := \hat{X}_L$*), or*
- $\hat{X}_L$ *is post-processed by replacing* $[\hat{X}_L]_{i,j}$ *with* $[\widetilde{X}_L]_{i,j} := \min\{\max\{[\hat{X}_L]_{i,j}, -b\}, b\}$ *for all* i, j*,*

then we also have

$$\begin{aligned}&\|\widetilde{X}_L - \bar{X}_L\|_{\mathrm{vec}(2)} \\ &\le \min\left\{\|\hat{X}_L - \bar{X}_L\|_{\mathrm{vec}(1)},\ \sqrt{2b\cdot\|\hat{X}_L - \bar{X}_L\|_{\mathrm{vec}(1)}}\right\}.\end{aligned}$$

The proof of Theorem 2 is in Section V. It is clear that if $Y = \bar{X}_S + \bar{X}_L$, then we can set $\epsilon_{\mathrm{vec}(1)} = \epsilon_* = 0$ and we obtain exact recovery: $\hat{X}_S = \bar{X}_S$ and $\hat{X}_L = \bar{X}_L$. Moreover, any perturbation $Y - (\bar{X}_S + \bar{X}_L)$ affects the accuracy of $(\hat{X}_S, \hat{X}_L)$ in entry-wise 1-norm by an amount $O(\epsilon_{\mathrm{vec}(1)} + \epsilon_*/\lambda)$. Note that here, the parameter λ serves to balance the entry-wise 1-norm and trace norm of the perturbation in the same way it is used in the objective function of (1). So, for instance, if we have the simplified conditions (11), then we may choose $\lambda = \sqrt{(5/3)\gamma/\alpha(\rho)}$ to satisfy (12), upon which the error bound becomes

$$\begin{aligned}\max\left\{\|\hat{X}_S - \bar{X}_S\|_{\mathrm{vec}(1)},\ \|\hat{X}_L - \bar{X}_L\|_{\mathrm{vec}(1)}\right\} \\ = O\left(\epsilon_{\mathrm{vec}(1)} + \sqrt{\frac{\alpha(\rho)}{\gamma}}\cdot\epsilon_*\right).\end{aligned}$$

It is possible to modify the constraints in (1) to use norms other than $\|\cdot\|_{\mathrm{vec}(1)}$ and $\|\cdot\|_*$; the analysis could at the very least be modified by simply using standard relationships to change between norms, although this may introduce new slack in the bounds. Finally, the second part of the theorem shows how the accuracy of $\hat{X}_L$ in Frobenius norm can be improved by adding an additional constraint or by post-processing the solution.

Now we state our recovery guarantees for the regularized formulation (2).

Theorem 3. *Fix a target pair* $(\bar{X}_S, \bar{X}_L) \in \mathbb{R}^{m\times n} \times \mathbb{R}^{m\times n}$*. Let* $E := Y - (\bar{X}_S + \bar{X}_L)$ *and*

$$\begin{aligned}\epsilon_{2\to2} &:= \|E\|_{2\to2} \\ \epsilon_{\mathrm{vec}(\infty)} &:= \|E\|_{\mathrm{vec}(\infty)} + \|\mathcal{P}_{\bar{T}}(E)\|_{\mathrm{vec}(\infty)} \\ \epsilon'_* &:= \|\mathcal{P}_{\bar{T}}(E)\|_*.\end{aligned}$$

Let $\bar{k} := |\mathrm{supp}(\bar{X}_S)|$ *and* $\bar{r} := \mathrm{rank}(\bar{X}_L)$*. Assume the conditions* (3)*,* (4)*, and* (5) *hold for some* $\rho > 0$ *and* $c > 1$*. Let* $(\hat{X}_S, \hat{X}_L) \in \mathbb{R}^{m\times n}$ *be the solution to the convex optimization problem* (2) *augmented with the constraint* $\|X_S - Y\|_{\mathrm{vec}(\infty)} \le b$ *for some* $b \ge \|\bar{X}_S - Y\|_{\mathrm{vec}(\infty)}$ *(*$b = \infty$ *is allowed). Let*

$$\begin{aligned}&\bar{r}' := \\ &\left(\lambda + \frac{\epsilon_{\mathrm{vec}(\infty)}}{\mu}\right)\cdot\frac{2\bar{k}}{1-\alpha(\rho)\beta(\rho)}\cdot\left(\lambda + \gamma + \frac{\epsilon_{\mathrm{vec}(\infty)}}{\mu}\right) \\ &+ \left(1 + 2\mu^{-1}\epsilon_{2\to2}\right)\cdot 2\bar{r} \\ &\cdot\left(\frac{2\alpha(\rho)}{1-\alpha(\rho)\beta(\rho)}\cdot\left(\lambda + \gamma + \frac{\epsilon_{\mathrm{vec}(\infty)}}{\mu}\right) + 1 + \frac{2\epsilon_{2\to2}}{\mu}\right).\end{aligned}$$

We have

$$\begin{aligned}&\|\hat{X}_S - \bar{X}_S\|_{\mathrm{vec}(1)} \\ &\le \frac{\frac{\bar{r}'\cdot\mu}{(1-1/c)\lambda} + \lambda\bar{k}\cdot\mu + 2\sqrt{\bar{k}\bar{r}}\cdot\mu + \bar{k}\cdot\epsilon_{\mathrm{vec}(\infty)}}{1-\alpha(\rho)\beta(\rho)} \\ &\|\hat{X}_S - \bar{X}_S\|_{\mathrm{vec}(2)} \\ &\le \min\left\{\|\hat{X}_S - \bar{X}_S\|_{\mathrm{vec}(1)},\ \sqrt{2b\cdot\|\hat{X}_S - \bar{X}_S\|_{\mathrm{vec}(1)}}\right\} \\ &\|\hat{X}_L - \bar{X}_L\|_* \le \sqrt{2\bar{r}}\cdot\|\hat{X}_S - \bar{X}_S\|_{\mathrm{vec}(2)} + \epsilon'_* \\ &\quad + \left(\frac{\bar{r}'\cdot(1-1/c)^{-1}}{2} + 2\bar{r}\right)\cdot\mu.\end{aligned}$$

The proof of Theorem 3 is in Section VI. As before, if $Y = \bar{X}_S + \bar{X}_L$ so $E = 0$, then we can set $\mu \to 0$ and obtain exact recovery with $\hat{X}_S = \bar{X}_S$ and $\hat{X}_L = \bar{X}_L$. When the perturbation E is non-zero, we control the accuracy of $\bar{X}_S$ in entry-wise 1-norm and 2-norm, and the accuracy of $\bar{X}_L$ in trace norm. Under the simplified conditions (6), we can choose $\lambda = (15/82)/\alpha(\rho)$ and

$\mu = \max\{4\epsilon_{2\to2}, 2\epsilon_{\mathrm{vec}(\infty)}/(15\lambda)\}$ to satisfy (7); this leads to the error bounds

$$\|\hat{X}_S - \bar{X}_S\|_{\mathrm{vec}(1)} = O\big(\bar{r}\alpha(\rho)\max\{\epsilon_{2\to2}, \alpha(\rho)\epsilon_{\mathrm{vec}(\infty)}\}\big)$$

$$\|\hat{X}_L - \bar{X}_L\|_* = O\bigg(\sqrt{\bar{r}}\min\bigg\{\sqrt{b\cdot\|\hat{X}_S - \bar{X}_S\|_{\mathrm{vec}(1)}}, \|\hat{X}_S - \bar{X}_S\|_{\mathrm{vec}(1)}\bigg\} + \epsilon_*' + \bar{r}\cdot\max\big\{\epsilon_{2\to2},\ \alpha(\rho)\epsilon_{\mathrm{vec}(\infty)}\big\}\bigg)$$

(here, we have used the facts $\bar{k} \leq \alpha(\rho)^2$, $\alpha(\rho)\lambda = \Theta(1)$, and $\bar{r}' = O(\bar{r})$, which also implies that $\bar{k}\cdot\epsilon_{\mathrm{vec}(\infty)} = O(\alpha(\rho)\cdot\alpha(\rho)\epsilon_{\mathrm{vec}(\infty)})$). Finally, note that if the constraint $\|X_S - Y\|_{\mathrm{vec}(\infty)} \leq b$ is added (*i.e.*, $b < \infty$), then the requirement $b \geq \|\bar{X}_S - Y\|_{\mathrm{vec}(\infty)}$ can be satisfied with $b := \|\bar{X}_S\|_{\mathrm{vec}(\infty)} + \epsilon_{\mathrm{vec}(\infty)}$. This allows for a possibly improved bound on $\|\hat{X}_L - \bar{X}_L\|_*$.

Our analysis centers around the construction of a dual certificate using a least-squares method similar to that in related works [1], [3]. The construction requires the invertibility of $\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}}$ (a composition of projection operators), which is established in our analysis by studying certain operator norms of $\mathcal{P}_{\bar{\Omega}}$ and $\mathcal{P}_{\bar{T}}$ (in previous works, invertibility is established only under probabilistic assumptions [3] or stricter sparsity conditions [1]). The rest of the analysis then relates the accuracy of the solutions to (1) and (2) to properties of the constructed dual certificate.

D. Examples

We illustrate our main results with some simple examples.

1) Random models: We first consider a random model for the matrices $\bar{X}_S$ and $\bar{X}_L$ [1]. Let the support of $\bar{X}_S$ be chosen uniformly at random $\tilde{k}$ times over the $[m]\times[n]$ matrix entries (so that one entry can be selected multiple times). The value of the entries in the chosen support can be arbitrary. With high probability, we have

$$\|\operatorname{sign}(\bar{X}_S)\|_{1\to1} = O\left(\frac{\tilde{k}\log n}{n}\right) \quad \text{and}$$

$$\|\operatorname{sign}(\bar{X}_S)\|_{\infty\to\infty} = O\left(\frac{\tilde{k}\log m}{m}\right)$$

so for $\rho := \sqrt{(n\log m)/(m\log n)}$, we have

$$\alpha(\rho) = O\left(\tilde{k}\sqrt{\frac{(\log m)(\log n)}{mn}}\right).$$

The logarithmic factors are due to collisions in the random process. Now let $\bar{U}$ and $\bar{V}$ be chosen uniformly at random over all families of $\bar{r}$ orthonormal vectors in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively. Using arguments similar to those in [6], one can show that with high probability,

$$\|\bar{U}\bar{U}^\top\|_{\mathrm{vec}(\infty)} = O\left(\frac{\bar{r}\log m}{m}\right)$$

$$\|\bar{V}\bar{V}^\top\|_{\mathrm{vec}(\infty)} = O\left(\frac{\bar{r}\log n}{n}\right)$$

$$\|\bar{U}\|_{2\to\infty} = O\left(\sqrt{\frac{\bar{r}\log m}{m}}\right)$$

$$\|\bar{V}\|_{2\to\infty} = O\left(\sqrt{\frac{\bar{r}\log n}{n}}\right),$$

so for the previously chosen ρ, we have

$$\beta(\rho) = O\left(\bar{r}\sqrt{\frac{(\log m)(\log n)}{mn}}\right) \quad \text{and}$$

$$\gamma = O\left(\bar{r}\sqrt{\frac{(\log m)(\log n)}{mn}}\right).$$

Therefore

$$\alpha(\rho)\beta(\rho) = O\left(\frac{\tilde{k}\bar{r}(\log m)(\log n)}{mn}\right) \quad \text{and}$$

$$\alpha(\rho)\gamma = O\left(\frac{\tilde{k}\bar{r}(\log m)(\log n)}{mn}\right),$$

both of which are $\ll 1$ provided that

$$\tilde{k} \leq \delta\cdot\frac{mn}{\bar{r}(\log m)(\log n)}$$

for a small enough constant $\delta \in (0,1)$. In other words, when $\bar{X}_L$ is low-rank, the matrix $\bar{X}_S$ can have nearly a constant fraction of its entries be non-zero while still allowing for exact decomposition of $Y = \bar{X}_S + \bar{X}_L$. Our guarantee improves over that of [1] by roughly a factor of $\Omega((mn)^{1/4})$, but is worse by a factor of $\bar{r}(\log m)(\log n)$ relative to the guarantees of [3] for the random model. Therefore there is a gap between our generic deterministic analysis and a direct probabilistic analysis of this random model, and this gap seems unavoidable with sparsity conditions based on $\alpha(\rho)$. This is because $\bar{X}_L$ could be an $n\times n$ (for simplicity) block diagonal matrix with r blocks of $n/r \times n/r$ rank-1 matrices; such a matrix guarantees $\beta(1) = O(r/n)$ but has just n^2/r non-zero entries. It is an interesting open problem to find alternative characterizations of $\operatorname{supp}(\bar{X}_S)$ that can narrow or close this gap.

2) Principal component analysis with sparse corruptions: Suppose $\bar{X}_L$ is matrix of m data points lying in a low-dimensional subspace of $\mathbb{R}^n$, and Z is a random matrix with independent Gaussian noise entries with variance σ^2. Then $Y' = \bar{X}_L + Z$ is the standard model for principal component analysis. We augment the model with a sparse noise component $\bar{X}_S$ to obtain $Y = \bar{X}_S + \bar{X}_L + Z$; here, we allow the non-zero entries of $\bar{X}_S$ to possibly approach infinity.

According to Theorem 3, we need to estimate $\|Z\|_{2\to 2}$, $\|Z\|_{\text{vec}(\infty)}$, $\|\mathcal{P}_{\bar{T}}(Z)\|_{\text{vec}(\infty)}$, and $\|\mathcal{P}_{\bar{T}}(Z)\|_*$. We have the following with high probability [10],

$$\|Z\|_{2\to 2} \le \sigma\sqrt{m} + \sigma\sqrt{n} + O(\sigma).$$

Using standard arguments with the rotational invariance of the Gaussian distribution, we also have

$$\|Z\|_{\text{vec}(\infty)} \le O(\sigma \log(mn)) \quad \text{and} \quad \|\mathcal{P}_{\bar{T}}(Z)\|_{\text{vec}(\infty)} \le O(\sigma \log(mn))$$

with high probability. Finally, by Lemma 5, we have

$$\|\mathcal{P}_{\bar{T}}(Z)\|_* \le 2\bar{r}\|Z\|_{2\to 2} \le 2\bar{r}\sigma\sqrt{m} + 2\bar{r}\sigma\sqrt{n} + O(\bar{r}\sigma).$$

Suppose $(\bar{X}_S, \bar{X}_L)$ has $\alpha(\rho) \le c_1(\sqrt{mn}/\bar{r})$, $\beta(\rho) = \Theta(\bar{r}/\sqrt{mn})$, and $\gamma = \Theta(\bar{r}/\sqrt{mn})$ and satisfies the simplified condition (6). This can be achieved with $c_1 c_2 \le 1/41$ in Proposition 1. Also assume λ and μ are chosen to satisfy (7), and that $b \ge \|\bar{X}_L\|_{\text{vec}(\infty)} + \epsilon_{\text{vec}(\infty)}$. Then we note that $\bar{k} = O(c_1^2 mn/\bar{r}^2)$, and thus have from Theorem 3 (see the discussion thereafter):

$$\begin{aligned}
&\|\hat{X}_S - \bar{X}_S\|_{\text{vec}(1)} \\
&= O\big(c_1\sqrt{mn}\max\{\sigma\sqrt{m} + \sigma\sqrt{n}, \sigma\sqrt{mn}\log(mn)/\bar{r}\}\big) \\
&= O\left(\sigma c_1 mn \log(mn)/\bar{r}\right) \\
&\|\hat{X}_L - \bar{X}_L\|_* = O\big(\sqrt{b\sigma c_1 mn \log(mn)/\bar{r}} \\
&\quad + \bar{r}\sigma(\sqrt{m} + \sqrt{n})) + c_1\sqrt{mn}\big),
\end{aligned}$$

where we may take $b = O(\sigma \log(mn) + \|X_L\|_{\text{vec}(\infty)}))$.

Now consider the situation where both $m, n \to \infty$, and assume that $\|\bar{X}_L\|_{\text{vec}(\infty)}$ remains bounded. If $c_1(\log(mn))^2 = o(1)$—which means that the number of corruptions per column is $o(m/(\log(mn))^2)$ and the number of deterministic corruptions per row is $o(n/(\log(mn))^2)$—then

$$\|\hat{X}_L - \bar{X}_L\|_* = O(\bar{r}\sigma(\sqrt{m} + \sqrt{n}))$$

so the normalized trace norm error of $\hat{X}_L$ tends to zero

$$\frac{1}{\sqrt{mn}}\|\hat{X}_L - \bar{X}_L\|_* \to 0.$$

This means that we can correctly recover the principal components of $\bar{X}_L$ with both deterministic corruptions and random noise, when both m and n are large and $c_1(\log(mn))^2 = o(1)$ in Proposition 1.

III. Technical preliminaries

A. Norms, inner products, and projections

Our analysis involves a variety of norms of vectors, matrices (viewed as elements of a vector space as well as linear operators of vectors), and linear operators of matrices; we define these and related notions in this section.

1) Entry-wise norms: For any $p \in [1, \infty]$, define $\|v\|_p := (\sum_i |v_i|^p)^{1/p}$ be the p-norm of a vector v (with $\|v\|_\infty := \max_i |v_i|$). Also, define $\|M\|_{\text{vec}(p)} := (\sum_{i,j} |M_{i,j}|^p)^{1/p}$ to be the entry-wise p-norm of a matrix M (again, with $\|M\|_{\text{vec}(\infty)} := \max_{i,j} |M_{i,j}|$). Note that $\|\cdot\|_{\text{vec}(2)}$ corresponds to the Frobenius norm.

2) Inner products, linear operators, and orthogonal projections: We endow $\mathbb{R}^{m\times n}$ with the inner product $\langle\cdot,\cdot\rangle$ between matrices that induces the Frobenius norm $\|\cdot\|_{\text{vec}(2)}$; this is given by $\langle M, N\rangle = \text{tr}(M^\top N)$.

For a linear operator $\mathcal{T} : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$, we denote its adjoint by $\mathcal{T}^*$; this is the unique linear operator that satisfies $\langle \mathcal{T}^*(M), N\rangle = \langle M, \mathcal{T}(N)\rangle$ for all $M \in \mathbb{R}^{m\times n}$ and $N \in \mathbb{R}^{m\times n}$ (in this work, we only consider bounded linear operators). For any two linear operators $\mathcal{T}_1$ and $\mathcal{T}_2$, we let $\mathcal{T}_1 \circ \mathcal{T}_2$ denote their composition as defined by $(\mathcal{T}_1 \circ \mathcal{T}_2)(M) := \mathcal{T}_1(\mathcal{T}_2(M))$.

Given a subspace $W \subseteq \mathbb{R}^{m\times n}$, we let $W^\perp$ denote its orthogonal complement, and let $\mathcal{P}_W : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ denote the orthogonal projector to W with respect to $\langle\cdot,\cdot\rangle$, *i.e.*, the unique linear operator with range W and satisfying ${\mathcal{P}_W}^* = \mathcal{P}_W$ and $\mathcal{P}_W \circ \mathcal{P}_W = \mathcal{P}_W$.

3) Induced norms: For any two vector norms $\|\cdot\|_p$ and $\|\cdot\|_q$, define $\|M\|_{p\to q} := \max_{x\neq 0} \|Mx\|_q/\|x\|_p$ to be the corresponding induced operator norm of a matrix M. Our analysis uses the following special cases which have alternative definitions:

- $\|M\|_{1\to 1} = \max_j \|Me_j\|_1$,
- $\|M\|_{1\to 2} = \max_j \|Me_j\|_2$,
- $\|M\|_{2\to 2} =$ spectral norm of M (*i.e.*, largest singular value of M),
- $\|M\|_{2\to\infty} = \max_i \|M^\top e_i\|_2$, and
- $\|M\|_{\infty\to\infty} = \max_i \|M^\top e_i\|_1$.

Here, e_i is the ith coordinate vector which has a 1 in the ith position and 0 elsewhere.

Finally, we also consider induced operator norms of linear matrix operators $\mathcal{T} : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ (in particular, projection operators with respect to $\langle\cdot,\cdot\rangle$). For any two matrix norms $\|\cdot\|_\diamondsuit$ and $\|\cdot\|_\heartsuit$, define $\|\mathcal{T}\|_{\diamondsuit\to\heartsuit} := \max_{M\neq 0} \|\mathcal{T}(M)\|_\heartsuit/\|M\|_\diamondsuit$.

4) Other norms: The trace norm (or nuclear norm) $\|M\|_*$ of a matrix M is the sum of the singular values of M. We will also make use of a hybrid matrix norm

$\|\cdot\|_{\sharp(\rho)}$, parametrized by $\rho > 0$, which we define by

$$\|M\|_{\sharp(\rho)} := \max\{\rho\|M\|_{1\to1},\ \rho^{-1}\|M\|_{\infty\to\infty}\}.$$

Also define $\|M\|_{\flat(\rho)} := \sup_{\|N\|_{\sharp(\rho)}\le 1}\langle M,N\rangle$, *i.e.*, the dual of $\|\cdot\|_{\sharp(\rho)}$ (see below).

5) Dual pairs: The matrix norm $\|\cdot\|_{\heartsuit}$ is said to be dual to $\|\cdot\|_{\spadesuit}$ if, for all $M \in \mathbb{R}^{m\times n}$, $\|M\|_{\heartsuit} = \sup_{\|N\|_{\spadesuit}\le 1}\langle M,N\rangle$.

Proposition 2. *Fix any matrix norm $\|\cdot\|_{\spadesuit}$, and let $\|\cdot\|_{\heartsuit}$ be its dual. For all $M \in \mathbb{R}^{m\times n}$ and $N \in \mathbb{R}^{m\times n}$, we have*

$$\langle M,N\rangle \le \|M\|_{\spadesuit}\|N\|_{\heartsuit}.$$

Proposition 3. *Fix any any linear matrix operator $\mathcal{T} : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ and any pair of matrix norms $\|\cdot\|_{\spadesuit}$ and $\|\cdot\|_{\clubsuit}$. We have*

$$\|\mathcal{T}\|_{\spadesuit\to\clubsuit} = \|\mathcal{T}^*\|_{\diamondsuit\to\heartsuit},$$

where $\|\cdot\|_{\heartsuit}$ is dual to $\|\cdot\|_{\spadesuit}$, and $\|\cdot\|_{\diamondsuit}$ is dual to $\|\cdot\|_{\clubsuit}$.

The following pairs of matrix norms are dual to each other:

1) $\|\cdot\|_{\mathrm{vec}(p)}$ and $\|\cdot\|_{\mathrm{vec}(q)}$ where $1/p + 1/q = 1$;
2) $\|\cdot\|_*$ and $\|\cdot\|_{2\to2}$;
3) $\|\cdot\|_{\sharp(\rho)}$ and $\|\cdot\|_{\flat(\rho)}$ (by definition).

6) Some lemmas: First we show that the $\|\cdot\|_{\sharp(\rho)}$ norm (for any $\rho > 0$) bounds the spectral norm $\|\cdot\|_{2\to2}$.

Lemma 2. *For any $M \in \mathbb{R}^{m\times n}$, we have for all $\rho > 0$,*

$$\|M\|_{2\to2} \le \|M\|_{\sharp(\rho)}.$$

Proof: Let σ be the largest singular value of M, and let $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$ be, respectively, associated left and right singular vectors. Then

$$\begin{aligned}&\left\|\begin{bmatrix} 0 & \rho M \\ \rho^{-1}M^\top & 0\end{bmatrix}\begin{bmatrix}\rho^{1/2}u \\ \rho^{-1/2}v\end{bmatrix}\right\|_1 \\ &\quad= \left\|\begin{bmatrix}\rho^{1/2}Mv \\ \rho^{-1/2}M^\top u\end{bmatrix}\right\|_1 = \sigma\left\|\begin{bmatrix}\rho^{1/2}u \\ \rho^{-1/2}v\end{bmatrix}\right\|_1.\end{aligned}$$

Moreover, by definition of $\|\cdot\|_{1\to1}$,

$$\begin{aligned}&\left\|\begin{bmatrix} 0 & \rho M \\ \rho^{-1/2}M^\top & 0\end{bmatrix}\begin{bmatrix}\rho^{1/2}u \\ \rho^{-1/2}v\end{bmatrix}\right\|_1 \\ &\quad\le \left\|\begin{bmatrix} 0 & \rho M \\ \rho^{-1}M^\top & 0\end{bmatrix}\right\|_{1\to1}\left\|\begin{bmatrix}\rho^{1/2}u \\ \rho^{-1/2}v\end{bmatrix}\right\|_1.\end{aligned}$$

Therefore

$$\begin{aligned}\|M\|_{2\to2} = \sigma &\le \left\|\begin{bmatrix} 0 & \rho M \\ \rho^{-1}M^\top & 0\end{bmatrix}\right\|_{1\to1} \\ &= \max\{\|\rho^{-1}M^\top\|_{1\to1}, \|\rho M\|_{1\to1}\} \\ &= \max\{\rho^{-1}\|M\|_{\infty\to\infty}, \rho\|M\|_{1\to1}\} \\ &= \|M\|_{\sharp(\rho)}.\end{aligned}$$

■

The following lemma is the dual of Lemma 2.

Lemma 3. *For any $M \in \mathbb{R}^{m\times n}$, we have for all $\rho > 0$,*

$$\|M\|_{\flat(\rho)} \le \|M\|_*.$$

Proof: We know that $\|M\|_{\flat(\rho)} = \langle M,N\rangle$ for some matrix N such that $\|N\|_{\sharp(\rho)} = 1$. Therefore $\|N\|_{2\to2} \le 1$ from Lemma 2, and thus using Proposition 2,

$$\|M\|_{\flat(\rho)} = \langle M,N\rangle \le \|M\|_*\|N\|_{2\to2} \le \|M\|_*.$$

■

Finally we state a lemma concerning the invertibility of a certain block-form operator used in our analysis.

Lemma 4. *Fix any matrix norm $\|\cdot\|_{\spadesuit}$ on $\mathbb{R}^{m\times n}$ and linear operators $\mathcal{T}_1 : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ and $\mathcal{T}_2 : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$. Let $\mathcal{I} : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ be the identity operator, and suppose $\|\mathcal{T}_1\circ\mathcal{T}_2\|_{\spadesuit\to\spadesuit} < 1$.*

1) *$\mathcal{I} - \mathcal{T}_1\circ\mathcal{T}_2$ is invertible and satisfies*

$$\|(\mathcal{I} - \mathcal{T}_1\circ\mathcal{T}_2)^{-1}\|_{\spadesuit\to\spadesuit} \le \frac{1}{1 - \|\mathcal{T}_1\circ\mathcal{T}_2\|_{\spadesuit\to\spadesuit}}.$$

2) *The linear operator on $\mathbb{R}^{m\times n}\times\mathbb{R}^{m\times n}$*

$$\begin{bmatrix}\mathcal{I} & \mathcal{T}_1 \\ \mathcal{T}_2 & \mathcal{I}\end{bmatrix}$$

is invertible, and its inverse is given by

$$\begin{aligned}&\begin{bmatrix}\mathcal{I} & \mathcal{T}_1 \\ \mathcal{T}_2 & \mathcal{I}\end{bmatrix}^{-1} \\ &= \begin{bmatrix}(\mathcal{I} - \mathcal{T}_1\circ\mathcal{T}_2)^{-1} & -(\mathcal{I} - \mathcal{T}_1\circ\mathcal{T}_2)^{-1}\circ\mathcal{T}_1 \\ -(\mathcal{I} - \mathcal{T}_2\circ\mathcal{T}_1)^{-1}\circ\mathcal{T}_2 & (\mathcal{I} - \mathcal{T}_2\circ\mathcal{T}_1)^{-1}\end{bmatrix}.\end{aligned}$$

Proof: The first claim is a standard application of Taylor expansions. The second claim then follows from formulae of block matrix inverses using Schur complements. ■

B. Projection operators and subdifferential sets

Recall the definitions of the following subspaces

$$\Omega(X_S) := \{X \in \mathbb{R}^{m\times n} : \mathrm{supp}(X) \subseteq \mathrm{supp}(X_S)\}$$

and

$$\begin{aligned}T(X_L) := \{X_1 + X_2 \in \mathbb{R}^{m\times n} :&\\ &\mathrm{range}(X_1) \subseteq \mathrm{range}(X_L),\\ &\mathrm{range}(X_2^\top) \subseteq \mathrm{range}(X_L^\top)\}.\end{aligned}$$

The orthogonal projectors to these spaces are given in the following proposition.

Proposition 4. *Fix any $X_S \in \mathbb{R}^{m\times n}$ and $X_L \in \mathbb{R}^{m\times n}$. For any matrix $M \in \mathbb{R}^{m\times n}$,*

$$[\mathcal{P}_{\Omega(X_S)}(M)]_{i,j} = \begin{cases} M_{i,j} & \text{if } (i,j) \in \operatorname{supp}(X_S) \\ 0 & \text{otherwise} \end{cases}$$

for all $1 \le i \le m$ and $1 \le j \le n$, and

$$\mathcal{P}_{T(X_L)}(M) = UU^\top M + MVV^\top - UU^\top MVV^\top$$

where U and V are the matrices of left and right singular vectors of X_L.

Lemma 5. *Under the setting of Proposition 4, with $k := |\operatorname{supp}(X_S)|$,*

$$\begin{aligned}
\|\mathcal{P}_{\Omega(X_S)}(M)\|_{\text{vec}(1)} &\le \sqrt{k}\|\mathcal{P}_{\Omega(X_S)}(M)\|_{\text{vec}(2)} \\
&\le \sqrt{k}\|M\|_{\text{vec}(2)} \\
\|\mathcal{P}_{\Omega(X_S)}(M)\|_{\text{vec}(1)} &\le k\|\mathcal{P}_{\Omega(X_S)}(M)\|_{\text{vec}(\infty)} \\
&\le k\|M\|_{\text{vec}(\infty)} \\
\|\mathcal{P}_{T(X_L)}(M)\|_{2\to 2} &\le 2\|M\|_{2\to 2} \\
\|\mathcal{P}_{T(X_L)}(M)\|_{*} &\le 2\operatorname{rank}(X_L)\|M\|_{2\to 2} \\
\|\mathcal{P}_{T(X_L)}(M)\|_{\text{vec}(2)} &\le 2\sqrt{\operatorname{rank}(X_L)}\|M\|_{2\to 2}.
\end{aligned}$$

Proof: The first and second claims rely on the fact that $|\operatorname{supp}(\mathcal{P}_{\Omega(X_S)}(M))| \le |\operatorname{supp}(X_S)|$, as well as the fact that $\mathcal{P}_{\Omega(X_S)}$ is an orthonormal projector with respect to the inner product that induces the $\|\cdot\|_{\text{vec}(2)}$ norm. For the third claim, note that

$$\begin{aligned}
&\|\mathcal{P}_{T(X_L)}(M)\|_{2\to 2} \\
&\le \|UU^\top M\|_{2\to 2} + \|(I - UU^\top)MVV^\top\|_{2\to 2} \\
&\le 2\|M\|_{2\to 2}.
\end{aligned}$$

The remaining claims use a similar decomposition as the third claim as well as the fact that $\operatorname{rank}(UU^\top M) \le \operatorname{rank}(X_L)$ and $\operatorname{rank}((I-UU^\top)MVV^\top)\} \le \operatorname{rank}(X_L)$. ∎

Define

$$\operatorname{sign}(X_S) \in \{-1, 0, +1\}^{m\times n}$$

to be the matrix whose (i,j)th entry is $\operatorname{sign}([X_S]_{i,j})$, and define

$$\operatorname{orth}(X_L) := UV^\top,$$

where U and V, respectively, are matrices of the left and right orthonormal singular vectors of X_L corresponding to non-zero singular values. The following proposition characterizes the subdifferential sets for the non-smooth norms $\|\cdot\|_{\text{vec}(1)}$ and $\|\cdot\|_*$ [11].

Proposition 5. *The subdifferential set of $X_S \mapsto \|X_S\|_{\text{vec}(1)}$ is*

$$\begin{aligned}
&\partial_{X_S}(\|X_S\|_{\text{vec}(1)}) \\
&= \{G \in \mathbb{R}^{m\times n} : \|G\|_{\text{vec}(\infty)} \le 1, \mathcal{P}_{\Omega(X_S)}(G) \\
&= \operatorname{sign}(X_S)\};
\end{aligned}$$

the subdifferential set of $X_L \mapsto \|X_L\|_*$ is

$$\begin{aligned}
\partial_{X_L}(\|X_L\|_*) = \{G \in \mathbb{R}^{m\times n} :& \\
\|G\|_{2\to 2} \le 1, \mathcal{P}_{T(X_L)}(G) &= \operatorname{orth}(X_L)\}.
\end{aligned}$$

The following lemma is a simple consequence of subgradient properties.

Lemma 6. *Fix $\lambda > 0$ and define the function $g(X_S, X_L) := \lambda\|X_S\|_{\text{vec}(1)} + \|X_L\|_*$. Consider any $(\bar{X}_S, \bar{X}_L)$ in $\mathbb{R}^{m\times n} \times \mathbb{R}^{m\times n}$. If there exists $Q \in \mathbb{R}^{m\times n}$ such that: Q is a subgradient of $\lambda\|X_S\|_{\text{vec}(1)}$ at $X_S = \bar{X}_S$, Q is a subgradient of $\|X_L\|_*$ at $X_L = \bar{X}_L$, and $\|\mathcal{P}_{\Omega(\bar{X}_S)^\perp}(Q)\|_{\text{vec}(\infty)} \le \lambda/c$ and $\|\mathcal{P}_{T(\bar{X}_L)^\perp}(Q)\|_{2\to 2} \le 1/c$ for some $c > 1$, then*

$$\begin{aligned}
&g(X_S, X_L) - g(\bar{X}_S, \bar{X}_L) \\
&\ge \langle Q, X_S + X_L - \bar{X}_S - \bar{X}_L\rangle \\
&\quad + (1 - 1/c)\lambda\|\mathcal{P}_{\bar{\Omega}^\perp}(X_S - \bar{X}_S)\|_{\text{vec}(1)} \\
&\quad + (1 - 1/c)\|\mathcal{P}_{\bar{T}^\perp}(X_L - \bar{X}_L)\|_*
\end{aligned}$$

for all $(X_S, X_L) \in \mathbb{R}^{m\times n} \times \mathbb{R}^{m\times n}$.

Proof: Let $\bar{\Omega} := \Omega(\bar{X}_S)$, $\bar{T} := T(\bar{X}_L)$, $\Delta_S := X_S - \bar{X}_S$, and $\Delta_L : X_L - \bar{X}_L$. For any subgradient $G \in \partial_{X_S}(\lambda\|\bar{X}_S\|_{\text{vec}(1)})$, we have $G - Q = \mathcal{P}_{\bar{\Omega}}(G) + \mathcal{P}_{\bar{\Omega}^\perp}(G) - \mathcal{P}_{\bar{\Omega}}(Q) - \mathcal{P}_{\bar{\Omega}^\perp}(Q) = \mathcal{P}_{\bar{\Omega}^\perp}(G) - \mathcal{P}_{\bar{\Omega}^\perp}(Q)$. Therefore

$$\begin{aligned}
&\lambda\|\bar{X}_S + \Delta_S\|_{\text{vec}(1)} - \lambda\|\bar{X}_S\|_{\text{vec}(1)} - \langle Q, \Delta_S\rangle \\
&\ge \sup\{\langle G, \Delta_S\rangle - \langle Q, \Delta_S\rangle : G \in \partial_{X_S}(\lambda\|\bar{X}_S\|_{\text{vec}(1)})\} \\
&\ge \sup\{\langle G - Q, \Delta_S\rangle : G \in \partial_{X_S}(\lambda\|\bar{X}_S\|_{\text{vec}(1)})\} \\
&= \sup\{\langle \mathcal{P}_{\bar{\Omega}^\perp}(G) - \mathcal{P}_{\bar{\Omega}^\perp}(Q), \Delta_S\rangle : \\
&\qquad G \in \partial_{X_S}(\lambda\|\bar{X}_S\|_{\text{vec}(1)})\} \\
&= \sup\{\langle \mathcal{P}_{\bar{\Omega}^\perp}(G) - \mathcal{P}_{\bar{\Omega}^\perp}(Q), \mathcal{P}_{\bar{\Omega}^\perp}(\Delta_S)\rangle : \\
&\qquad G \in \partial_{X_S}(\lambda\|\bar{X}_S\|_{\text{vec}(1)})\} \\
&= \sup\{\langle \mathcal{P}_{\bar{\Omega}^\perp}(G), \mathcal{P}_{\bar{\Omega}^\perp}(\Delta_S)\rangle \\
&\qquad - \langle \mathcal{P}_{\bar{\Omega}^\perp}(Q), \mathcal{P}_{\bar{\Omega}^\perp}(\Delta_S)\rangle : G \in \partial_{X_S}(\lambda\|\bar{X}_S\|_{\text{vec}(1)})\} \\
&= \lambda\|\mathcal{P}_{\bar{\Omega}^\perp}(\Delta_S)\|_{\text{vec}(1)} - \langle \mathcal{P}_{\bar{\Omega}^\perp}(Q), \mathcal{P}_{\bar{\Omega}^\perp}(\Delta_S)\rangle \\
&\ge \lambda\|\mathcal{P}_{\bar{\Omega}^\perp}(\Delta_S)\|_{\text{vec}(1)} \\
&\qquad - \|\mathcal{P}_{\bar{\Omega}^\perp}(Q)\|_{\text{vec}(\infty)}\|\mathcal{P}_{\bar{\Omega}^\perp}(\Delta_S)\|_{\text{vec}(1)} \\
&\ge \lambda(1 - 1/c)\|\mathcal{P}_{\bar{\Omega}^\perp}(\Delta_S)\|_{\text{vec}(1)}
\end{aligned}$$

where the second-to-last inequality uses the duality of $\|\cdot\|_{\mathrm{vec}(1)}$ and $\|\cdot\|_{\mathrm{vec}(\infty)}$ and Proposition 3. Similarly,

$$\|\bar{X}_L - \Delta_L\|_* - \|\bar{X}_L\|_* - \langle Q, \Delta_L\rangle \geq (1 - 1/c)\|\mathcal{P}_{\bar{T}^\perp}(\Delta_L)\|_*$$

by noting the duality of $\|\cdot\|_*$ and $\|\cdot\|_{2\to 2}$. Combining these gives the desired inequality. ■

IV. Rank-sparsity incoherence

Throughout this section, we fix a target $(\bar{X}_S, \bar{X}_L) \in \mathbb{R}^{m\times n} \times \mathbb{R}^{m\times n}$, and let $\bar{\Omega} := \Omega(\bar{X}_S)$ and $\bar{T} := T(\bar{X}_L)$. Also let $\bar{U}$ and $\bar{V}$ be, respectively, matrices of the left and right singular vectors of $\bar{X}_L$ corresponding to non-zero singular values. Recall the following structural properties of $\bar{X}_S$ and $\bar{X}_L$:

$$\begin{aligned}
\alpha(\rho) &:= \|\operatorname{sign}(\bar{X}_S)\|_{\sharp(\rho)} \\
&= \max\{\rho\|\operatorname{sign}(\bar{X}_S)\|_{1\to 1},\ \rho^{-1}\|\operatorname{sign}(\bar{X}_S)\|_{\infty\to\infty}\}; \\
\beta(\rho) &:= \rho^{-1}\|\bar{U}\bar{U}^\top\|_{\mathrm{vec}(\infty)} + \rho\|\bar{V}\bar{V}^\top\|_{\mathrm{vec}(\infty)} \\
&\quad + \|\bar{U}\|_{2\to\infty}\|\bar{V}\|_{2\to\infty}; \\
\gamma &:= \|\operatorname{orth}(\bar{X}_L)\|_{\mathrm{vec}(\infty)} = \|\bar{U}\bar{V}^\top\|_{\mathrm{vec}(\infty)}.
\end{aligned}$$

The parameter ρ is a balancing parameter to handle disparity between row and column dimensions. The quantity $\alpha(\rho)$ is the maximum number of non-zero entries in any single row or column. The quantities $\beta(\rho)$ and γ measure the coherence of the singular vectors of $\bar{X}_L$, that is, the alignment of the singular vectors with the coordinate basis. For instance, under the conditions of Proposition 1, we have (with $\rho = \sqrt{n/m}$)

$$\alpha(\rho) \leq c_1\sqrt{mn},$$

$$\beta(\rho) \leq \frac{3c_2\operatorname{rank}(\bar{X}_L)}{\sqrt{mn}} \quad\text{and}\quad \gamma \leq \frac{c_2\operatorname{rank}(\bar{X}_L)}{\sqrt{mn}}$$

for some constants c_1 and c_2.

A. Operator norms of projection operators

We show that under the condition $\inf_{\rho>0}\alpha(\rho)\beta(\rho) < 1$, the pair $(\bar{X}_S, \bar{X}_L)$ is identifiable from its sum $\bar{X}_S + \bar{X}_L$ (Theorem 1). This is achieved by proving that the composition of projection operators $\mathcal{P}_{\bar{\Omega}}$ and $\mathcal{P}_{\bar{T}}$ is a contraction as per Lemma 1, which in turn implies that $\bar{\Omega} \cap \bar{T} = \{0\}$.

The following two lemmas bound the projection operators $\mathcal{P}_{\bar{\Omega}}$ and $\mathcal{P}_{\bar{T}}$ in complementary norms.

Lemma 7. *For any $M \in \mathbb{R}^{m\times n}$ and $p \in \{1, \infty\}$, we have*

$$\|\mathcal{P}_{\bar{\Omega}}(M)\|_{p\to p} \leq \|\operatorname{sign}(\bar{X}_S)\|_{p\to p}\|M\|_{\mathrm{vec}(\infty)}.$$

This implies, for all $\rho > 0$,

$$\|\mathcal{P}_{\bar{\Omega}}\|_{\mathrm{vec}(\infty)\to\sharp(\rho)} \leq \alpha(\rho).$$

Proof: Define $s(X_S) \in \{0,1\}^{m\times n}$ to be the entry-wise absolute value of $\operatorname{sign}(X_S)$. We have

$$\begin{aligned}
\|\mathcal{P}_{\bar{\Omega}}(M)\|_{p\to p} &= \max\{\|\mathcal{P}_{\bar{\Omega}}(M)v\|_p : \|v\|_p \leq 1\} \\
&\leq \|\mathcal{P}_{\bar{\Omega}}(M)\|_{\mathrm{vec}(\infty)} \\
&\qquad \max\{\|s(\mathcal{P}_{\bar{\Omega}}(M))v\|_p : \|v\|_p \leq 1\} \\
&\leq \|M\|_{\mathrm{vec}(\infty)} \\
&\qquad \max\{\|s(\bar{X}_S)v\|_p : \|v\|_p \leq 1\} \\
&= \|M\|_{\mathrm{vec}(\infty)}\|\operatorname{sign}(\bar{X}_S)\|_{p\to p}.
\end{aligned}$$

The second part follows from the definitions of $\|\cdot\|_{\sharp(\rho)}$ and $\alpha(\rho)$. ■

Lemma 8. *For any $M \in \mathbb{R}^{m\times n}$, we have*

$$\begin{aligned}
&\|\mathcal{P}_{\bar{T}}(M)\|_{\mathrm{vec}(\infty)} \\
&\leq \|\bar{U}\bar{U}^\top\|_{\mathrm{vec}(\infty)}\|M\|_{1\to 1} + \|\bar{V}\bar{V}^\top\|_{\mathrm{vec}(\infty)}\|M\|_{\infty\to\infty} \\
&\quad + \|\bar{U}\|_{2\to\infty}\|\bar{V}\|_{2\to\infty}\|M\|_{2\to 2}.
\end{aligned}$$

This implies, for all $\rho > 0$,

$$\|\mathcal{P}_{\bar{T}}\|_{\sharp(\rho)\to\mathrm{vec}(\infty)} \leq \beta(\rho).$$

Proof: We have $\|\mathcal{P}_{\bar{T}}(M)\|_{\mathrm{vec}(\infty)} = \|\bar{U}\bar{U}^\top M + M\bar{V}\bar{V}^\top - \bar{U}\bar{U}^\top M\bar{V}\bar{V}^\top\|_{\mathrm{vec}(\infty)} \leq \|\bar{U}\bar{U}^\top M\|_{\mathrm{vec}(\infty)} + \|M\bar{V}\bar{V}^\top\|_{\mathrm{vec}(\infty)} + \|\bar{U}\bar{U}^\top M\bar{V}\bar{V}^\top\|_{\mathrm{vec}(\infty)}$ by the triangle inequality. The bounds for each term now follow from the definitions:

$$\begin{aligned}
\|\bar{U}\bar{U}^\top M\|_{\mathrm{vec}(\infty)} &= \max_i \|M^\top\bar{U}\bar{U}^\top e_i\|_\infty \\
&\leq \|M^\top\|_{\infty\to\infty}\max_i\|\bar{U}\bar{U}^\top e_i\|_\infty \\
&= \|M\|_{1\to 1}\|\bar{U}\bar{U}^\top\|_{\mathrm{vec}(\infty)};
\end{aligned}$$

$$\begin{aligned}
\|M\bar{V}\bar{V}^\top\|_{\mathrm{vec}(\infty)} &= \max_j \|M\bar{V}\bar{V}^\top e_j\|_\infty \\
&\leq \|M\|_{\infty\to\infty}\max_j\|\bar{V}\bar{V}^\top e_j\|_\infty \\
&= \|M\|_{\infty\to\infty}\|\bar{V}\bar{V}\|_{\mathrm{vec}(\infty)};
\end{aligned}$$

and

$$\begin{aligned}
&\|\bar{U}\bar{U}^\top M\bar{V}\bar{V}^\top\|_{\mathrm{vec}(\infty)} \\
&= \max_{i,j}|e_i^\top\bar{U}(\bar{U}^\top M\bar{V})\bar{V}^\top e_j| \\
&\leq \max_{i,j}\|\bar{U}^\top e_i\|_2\|\bar{U}^\top M\bar{V}\|_{2\to 2}\|\bar{V}^\top e_j\|_2 \\
&\leq \|M\|_{2\to 2}\|\bar{U}\|_{2\to\infty}\|\bar{V}\|_{2\to\infty} \\
&\leq \|M\|_{\sharp(\rho)}\|\bar{U}\|_{2\to\infty}\|\bar{V}\|_{2\to\infty}
\end{aligned}$$

where the second step follows by Cauchy-Schwarz, and the fourth step follows from Lemma 2. The second part now follows the definition of $\beta(\rho)$. ■

Now we show that the composition of $\mathcal{P}_{\bar{\Omega}}$ and $\mathcal{P}_{\bar{T}}$ gives a contraction under the certain norms and their duals.

Lemma 9. *For all $\rho > 0$,*

1) $\|\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}}\|_{\sharp(\rho)\to\sharp(\rho)} \le \alpha(\rho)\beta(\rho)$;
2) $\|\mathcal{P}_{\bar{T}} \circ \mathcal{P}_{\bar{\Omega}}\|_{\mathrm{vec}(\infty)\to\mathrm{vec}(\infty)} \le \alpha(\rho)\beta(\rho)$;

Proof: Immediate from Lemma 7 and Lemma 8. ∎

Lemma 10. *For all $\rho > 0$,*

1) $\|\mathcal{P}_{\bar{T}} \circ \mathcal{P}_{\bar{\Omega}}\|_{\flat(\rho)\to\flat(\rho)} \le \alpha(\rho)\beta(\rho)$;
2) $\|\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}}\|_{\mathrm{vec}(1)\to\mathrm{vec}(1)} \le \alpha(\rho)\beta(\rho)$.

Proof: First note that $(\mathcal{P}_{\bar{T}} \circ \mathcal{P}_{\bar{\Omega}})^* = \mathcal{P}_{\bar{\Omega}}^* \circ \mathcal{P}_{\bar{T}}^* = \mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}}$ because $\mathcal{P}_{\bar{\Omega}}$ and $\mathcal{P}_{\bar{T}}$ are self-adjoint, and similarly $(\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}})^* = \mathcal{P}_{\bar{T}} \circ \mathcal{P}_{\bar{\Omega}}$. Now the claim follows by Proposition 3 and Lemma 9, using the facts that $\|\cdot\|_{\flat(\rho)}$ is dual to $\|\cdot\|_{\sharp(\rho)}$ and that $\|\cdot\|_{\mathrm{vec}(1)}$ is dual to $\|\cdot\|_{\mathrm{vec}(\infty)}$. ∎

Note that Lemma 1 is encompassed by Lemma 10. Another consequence of these contraction properties is the following uncertainty principle, analogous to one stated by [1], which effectively states that a matrix X cannot have both $\|\operatorname{sign}(X)\|_{\sharp(\rho)}$ and $\|\operatorname{orth}(X)\|_{\mathrm{vec}(\infty)}$ simultaneously small.

Theorem 4. *If $X = \bar{X}_S = \bar{X}_L \neq 0$, then* $\inf_{\rho>0} \alpha(\rho)\beta(\rho) \ge 1$.

Proof: Note that the non-zero element X lives in $\bar{\Omega} \cap \bar{T}$, so we get the conclusion by the contrapositive of Theorem 1. ∎

B. Dual certificate

The incoherence properties allow us to construct an approximate dual certificate $(Q_{\bar{\Omega}}, Q_{\bar{T}}) \in \bar{\Omega} \times \bar{T}$ that is central to the analysis of the optimization problems (1) and (2).

The certificate is constructed as the solution to the linear system

$$\begin{cases} \mathcal{P}_{\bar{\Omega}}(Q_{\bar{\Omega}} + Q_{\bar{T}} + \mu^{-1}E) = \lambda \operatorname{sign}(\bar{X}_S) \\ \mathcal{P}_{\bar{T}}(Q_{\bar{\Omega}} + Q_{\bar{T}} + \mu^{-1}E) = \operatorname{orth}(\bar{X}_L) \end{cases}$$

for some matrix $E \in \mathbb{R}^{m\times n}$; this can be equivalently written as

$$\begin{bmatrix} \mathcal{I} & \mathcal{P}_{\bar{\Omega}} \\ \mathcal{P}_{\bar{T}} & \mathcal{I} \end{bmatrix} \begin{bmatrix} Q_{\bar{\Omega}} \\ Q_{\bar{T}} \end{bmatrix} = \begin{bmatrix} \lambda \operatorname{sign}(\bar{X}_S) - \mu^{-1}\mathcal{P}_{\bar{\Omega}}(E) \\ \operatorname{orth}(\bar{X}_L) - \mu^{-1}\mathcal{P}_{\bar{T}}(E) \end{bmatrix}.$$

We show the existence of the dual certificate $(Q_{\bar{\Omega}}, Q_{\bar{T}})$ under the conditions (3), (4), and (5) relative to an arbitrary matrix E. Recall that the recovery guarantees for the constrained formulation requires the conditions with $E = 0$, while the guarantees for the regularized formulation takes $E = Y - (\bar{X}_S + \bar{X}_L)$.

Theorem 5. *Pick any $c > 1$, $\rho > 0$, and $E \in \mathbb{R}^{m\times n}$. Let $\bar{k} := |\operatorname{supp}(\bar{X}_S)|$ and $\bar{r} := \operatorname{rank}(\bar{X}_L)$. Let*

$$\epsilon_{2\to2} := \|E\|_{2\to2}$$
$$\epsilon_{\mathrm{vec}(\infty)} := \|E\|_{\mathrm{vec}(\infty)} + \|\mathcal{P}_{\bar{T}}(E)\|_{\mathrm{vec}(\infty)}.$$

If the following conditions hold:

$$\alpha(\rho)\beta(\rho) < 1 \tag{13}$$

$$\lambda \le \frac{(1-\alpha(\rho)\beta(\rho))(1 - c\cdot\mu^{-1}\epsilon_{2\to2})}{c\cdot\alpha(\rho)} - \frac{\alpha(\rho)\mu^{-1}\epsilon_{\mathrm{vec}(\infty)} + \alpha(\rho)\gamma}{\alpha(\rho)} \tag{14}$$

$$\lambda \ge c\cdot\frac{\gamma + \mu^{-1}(2-\alpha(\rho)\beta(\rho))\epsilon_{\mathrm{vec}(\infty)}}{1-\alpha(\rho)\beta(\rho) - c\cdot\alpha(\rho)\beta(\rho)} > 0 \tag{15}$$

(these are a restatement of (3), (4), *and* (5)*), then*

$$Q_{\bar{\Omega}} := (\mathcal{I} - \mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}})^{-1}\big(\lambda \operatorname{sign}(\bar{X}_S) - \mathcal{P}_{\bar{\Omega}}(\operatorname{orth}(\bar{X}_L)) - \mu^{-1}(\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}^\perp})(E)\big) \in \bar{\Omega} \quad \textit{and}$$
$$Q_{\bar{T}} := (\mathcal{I} - \mathcal{P}_{\bar{T}} \circ \mathcal{P}_{\bar{\Omega}})^{-1}\big(\operatorname{orth}(\bar{X}_L) - \lambda\mathcal{P}_{\bar{T}}(\operatorname{sign}(\bar{X}_S)) - \mu^{-1}(\mathcal{P}_{\bar{T}} \circ \mathcal{P}_{\bar{\Omega}^\perp})(E)\big) \in \bar{T}$$

are well-defined and satisfy

$$\mathcal{P}_{\bar{\Omega}}(Q_{\bar{\Omega}} + Q_{\bar{T}} + \mu^{-1}E) = \lambda \operatorname{sign}(\bar{X}_S)$$
$$\mathcal{P}_{\bar{T}}(Q_{\bar{\Omega}} + Q_{\bar{T}} + \mu^{-1}E) = \operatorname{orth}(\bar{X}_L)$$

and

$$\|\mathcal{P}_{\bar{\Omega}^\perp}(Q_{\bar{\Omega}} + Q_{\bar{T}} + \mu^{-1}E)\|_{\mathrm{vec}(\infty)} \le \lambda/c$$
$$\|\mathcal{P}_{\bar{T}^\perp}(Q_{\bar{\Omega}} + Q_{\bar{T}} + \mu^{-1}E)\|_{2\to2} \le 1/c.$$

Moreover,

$$\|Q_{\bar{\Omega}}\|_{2\to2} \le \frac{\alpha(\rho)}{1-\alpha(\rho)\beta(\rho)}\cdot\big(\lambda + \gamma + \mu^{-1}\epsilon_{\mathrm{vec}(\infty)}\big)$$

$$\|Q_{\bar{T}}\|_{2\to2} \le \frac{2\alpha(\rho)}{1-\alpha(\rho)\beta(\rho)}\cdot\big(\lambda + \gamma + \mu^{-1}\epsilon_{\mathrm{vec}(\infty)}\big) + 1 + 2\mu^{-1}\epsilon_{2\to2}$$

$$\|Q_{\bar{T}}\|_* \le 2\bar{r}\|Q_{\bar{T}}\|_{2\to2}$$

$$\|Q_{\bar{T}}\|_{\mathrm{vec}(\infty)} \le \frac{1}{1-\alpha(\rho)\beta(\rho)}\cdot\big(\lambda + \gamma + \mu^{-1}\epsilon_{\mathrm{vec}(\infty)}\big)$$

$$\|Q_{\bar{\Omega}}\|_{\mathrm{vec}(\infty)} \le \frac{2}{1-\alpha(\rho)\beta(\rho)}\cdot\big(\lambda + \gamma + \mu^{-1}\epsilon_{\mathrm{vec}(\infty)}\big)$$

$$\|Q_{\bar{\Omega}}\|_{\mathrm{vec}(1)} \le \bar{k}\|Q_{\bar{\Omega}}\|_{\mathrm{vec}(\infty)}$$

$$\|Q_{\bar{\Omega}} + Q_{\bar{T}}\|_{\mathrm{vec}(2)}^2 \le \lambda\|Q_{\bar{\Omega}}\|_{\mathrm{vec}(1)}\big(1 + \mu^{-1}\lambda^{-1}\epsilon_{\mathrm{vec}(\infty)}\big) + \|Q_{\bar{T}}\|_*\big(1 + 2\mu^{-1}\epsilon_{2\to2}\big).$$

Remark 1. The dual certificate constitutes an approximate subgradient in the sense that $Q_{\bar{\Omega}} + Q_{\bar{T}} + \mu^{-1}E$

is a subgradient of both $\lambda\|X_S\|_{\text{vec}(1)}$ at $X_S = \bar{X}_S$, and $\|X_L\|_*$ at $X_L = \bar{X}_L$.

Proof: Under the condition (13), we have $\alpha(\rho)\beta(\rho) < 1$, and therefore Lemma 9 and Lemma 4 imply that the operators $\mathcal{I} - \mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}}$ and $\mathcal{I} - \mathcal{P}_{\bar{T}} \circ \mathcal{P}_{\bar{\Omega}}$ are invertible and satisfy

$$\|(\mathcal{I} - \mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}})^{-1}\|_{\sharp(\rho)\to\sharp(\rho)} \leq \frac{1}{1-\alpha(\rho)\beta(\rho)},$$
$$\|(\mathcal{I} - \mathcal{P}_{\bar{T}} \circ \mathcal{P}_{\bar{\Omega}})^{-1}\|_{\text{vec}(\infty)\to\text{vec}(\infty)} \leq \frac{1}{1-\alpha(\rho)\beta(\rho)}.$$

Thus $Q_{\bar{\Omega}}$ and $Q_{\bar{T}}$ are well-defined. We can bound $\|Q_{\bar{\Omega}}\|_{2\to2}$ as

$$\begin{aligned}
&\|Q_{\bar{\Omega}}\|_{2\to2} \leq \|Q_{\bar{\Omega}}\|_{\sharp(\rho)} \quad \text{(Lemma 2)}\\
&= \big\|(I - \mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}})^{-1}\big(\lambda \operatorname{sign}(\bar{X}_S) - \mathcal{P}_{\bar{\Omega}}(\operatorname{orth}(\bar{X}_L))\\
&\quad - \mu^{-1}(\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}^\perp})(E)\big)\big\|_{\sharp(\rho)}\\
&\leq \frac{1}{1-\alpha(\rho)\beta(\rho)} \cdot \big\|\lambda \operatorname{sign}(\bar{X}_S) - \mathcal{P}_{\bar{\Omega}}(\operatorname{orth}(\bar{X}_L))\\
&\quad - \mu^{-1}(\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}^\perp})(E)\big\|_{\sharp(\rho)}\\
&\leq \frac{1}{1-\alpha(\rho)\beta(\rho)} \cdot \big(\lambda\|\operatorname{sign}(\bar{X}_S)\|_{\sharp(\rho)}\\
&\quad + \|\mathcal{P}_{\bar{\Omega}}(\operatorname{orth}(\bar{X}_L))\|_{\sharp(\rho)} + \mu^{-1}\|(\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}^\perp})(E)\|_{\sharp(\rho)}\big)\\
&\leq \frac{\alpha(\rho)}{1-\alpha(\rho)\beta(\rho)} \cdot \big(\lambda + \gamma + \mu^{-1}\|\mathcal{P}_{\bar{T}^\perp}(E)\|_{\text{vec}(\infty)}\big)\\
&\quad\quad \text{(Lemma 7)}\\
&\leq \frac{\alpha(\rho)}{1-\alpha(\rho)\beta(\rho)} \cdot \big(\lambda + \gamma + \mu^{-1}\epsilon_{\text{vec}(\infty)}\big).
\end{aligned}$$

Above, we have used the bound $\|\mathcal{P}_{\bar{T}^\perp}(E)\|_{\text{vec}(\infty)} = \|E - \mathcal{P}_{\bar{T}}(E)\|_{\text{vec}(\infty)} \leq \epsilon_{\text{vec}(\infty)}$. Therefore,

$$\begin{aligned}
&\|\mathcal{P}_{\bar{T}^\perp}(Q_{\bar{\Omega}} + \mu^{-1}E)\|_{2\to2}\\
&\leq \|(I - \bar{U}\bar{U}^\top)Q_{\bar{\Omega}}(I - \bar{V}\bar{V}^\top)\|_{2\to2} + \frac{\|P_{\bar{T}^\perp}(E)\|_{2\to2}}{\mu}\\
&\leq \|Q_{\bar{\Omega}}\|_{2\to2} + \mu^{-1}\epsilon_{2\to2}\\
&\leq \frac{\alpha(\rho)}{1-\alpha(\rho)\beta(\rho)} \cdot (\lambda + \gamma + \mu^{-1}\epsilon_{\text{vec}(\infty)}) + \mu^{-1}\epsilon_{2\to2}.
\end{aligned}$$

The condition (14) now implies that this quantity is at most $1/c$.

Now we bound $\|Q_{\bar{T}}\|_{\text{vec}(\infty)}$ as

$$\begin{aligned}
&\|Q_{\bar{T}}\|_{\text{vec}(\infty)}\\
&= \big\|(\mathcal{I} - \mathcal{P}_{\bar{T}} \circ \mathcal{P}_{\bar{\Omega}})^{-1}\big(\operatorname{orth}(\bar{X}_L) - \lambda\mathcal{P}_{\bar{T}}(\operatorname{sign}(\bar{X}_S)\\
&\quad - \mu^{-1}(\mathcal{P}_{\bar{T}} \circ \mathcal{P}_{\bar{\Omega}^\perp})(E)\big)\big\|_{\text{vec}(\infty)}\\
&\leq \frac{1}{1-\alpha(\rho)\beta(\rho)} \cdot \big\|\operatorname{orth}(\bar{X}_L) - \lambda\mathcal{P}_{\bar{T}}(\operatorname{sign}(\bar{X}_S)\\
&\quad - \mu^{-1}(\mathcal{P}_{\bar{T}} \circ \mathcal{P}_{\bar{\Omega}^\perp})(E)\big\|_{\text{vec}(\infty)}\\
&\leq \frac{1}{1-\alpha(\rho)\beta(\rho)} \cdot \big(\|\operatorname{orth}(\bar{X}_L)\|_{\text{vec}(\infty)}\\
&\quad + \lambda\|\mathcal{P}_{\bar{T}}(\operatorname{sign}(\bar{X}_S))\|_{\text{vec}(\infty)}\\
&\quad + \mu^{-1}\|(\mathcal{P}_{\bar{T}} \circ \mathcal{P}_{\bar{\Omega}^\perp})(E)\|_{\text{vec}(\infty)}\big)\\
&\leq \frac{1}{1-\alpha(\rho)\beta(\rho)} \cdot \big(\gamma + \lambda\alpha(\rho)\beta(\rho) + \mu^{-1}\epsilon_{\text{vec}(\infty)}\big)\\
&\quad\quad \text{(Lemma 9)}.
\end{aligned}$$

Above, we have used the bound $\|(\mathcal{P}_{\bar{T}} \circ \mathcal{P}_{\bar{\Omega}^\perp})(E)\|_{\text{vec}(\infty)} = \|\mathcal{P}_{\bar{T}}(E) - (\mathcal{P}_{\bar{T}} \circ \mathcal{P}_{\bar{\Omega}})(E)\|_{\text{vec}(\infty)} \leq \|\mathcal{P}_{\bar{T}}(E)\|_{\text{vec}(\infty)} + \alpha(\rho)\beta(\rho)\|E\|_{\text{vec}(\infty)} \leq \epsilon_{\text{vec}(\infty)}$. Therefore,

$$\begin{aligned}
&\|\mathcal{P}_{\bar{\Omega}^\perp}(Q_{\bar{T}} + \mu^{-1}E)\|_{\text{vec}(\infty)}\\
&\leq \|Q_{\bar{T}}\|_{\text{vec}(\infty)} + \mu^{-1}\|\mathcal{P}_{\bar{\Omega}^\perp}(E)\|_{\text{vec}(\infty)}\\
&\leq \frac{1}{1-\alpha(\rho)\beta(\rho)} \cdot \big(\gamma + \lambda\alpha(\rho)\beta(\rho) + \mu^{-1}\epsilon_{\text{vec}(\infty)}\big)\\
&\quad + \mu^{-1}\epsilon_{\text{vec}(\infty)}.
\end{aligned}$$

The condition (15) now implies that this quantity is at most λ/c.

We also have

$$\begin{aligned}
&\|Q_{\bar{T}}\|_{2\to2} = \|\mathcal{P}_{\bar{T}}(Q_{\bar{\Omega}} + \mu^{-1}E) - \operatorname{orth}(\bar{X}_L)\|_{2\to2}\\
&\leq \frac{2\alpha(\rho)}{1-\alpha(\rho)\beta(\rho)} \cdot \big(\lambda + \gamma + \mu^{-1}\epsilon_{\text{vec}(\infty)}\big)\\
&\quad + 1 + 2\mu^{-1}\epsilon_{2\to2}
\end{aligned}$$

since $\|\mathcal{P}_{\bar{T}}(Q_{\bar{\Omega}})\|_{2\to2} \leq 2\|Q_{\bar{\Omega}}\|_{2\to2}$ and $\|\mathcal{P}_{\bar{T}}(E)\|_{2\to2} \leq 2\epsilon_{2\to2}$ by Lemma 5, and

$$\begin{aligned}
&\|Q_{\bar{\Omega}}\|_{\text{vec}(\infty)} = \|P_{\bar{\Omega}}(Q_{\bar{T}} + \mu^{-1}E) - \lambda \operatorname{sign}(\bar{X}_S)\|_{\text{vec}(\infty)}\\
&\leq \frac{1}{1-\alpha(\rho)\beta(\rho)} \cdot \big(\lambda + \gamma + \mu^{-1}\epsilon_{\text{vec}(\infty)}\big)\\
&\quad + \lambda + \mu^{-1}\epsilon_{\text{vec}(\infty)}.
\end{aligned}$$

The bounds on $\|Q_{\bar{T}}\|_*$ and $\|Q_{\bar{\Omega}}\|_{\text{vec}(1)}$ follow from the facts that $\operatorname{rank}(Q_{\bar{T}}) \leq 2\bar{r}$ and $\|\operatorname{supp}(Q_{\bar{\Omega}})\| \leq \bar{k}$.

Finally,

$$\begin{aligned}
&\|Q_{\bar\Omega} + Q_{\bar T}\|_{\mathrm{vec}(2)}^2 \\
&= \langle Q_{\bar\Omega}, \mathcal{P}_{\bar\Omega}(Q_{\bar\Omega} + Q_{\bar T})\rangle + \langle Q_{\bar T}, \mathcal{P}_{\bar T}(Q_{\bar\Omega} + Q_{\bar T})\rangle \\
&= \langle Q_{\bar\Omega}, \lambda\mathcal{P}_{\bar\Omega}(\mathrm{sign}(\bar X_S)) - \mu^{-1}\mathcal{P}_{\bar\Omega}(E)\rangle \\
&\quad + \langle Q_{\bar T}, \mathcal{P}_{\bar T}(\mathrm{orth}(\bar X_L)) - \mu^{-1}\mathcal{P}_{\bar T}(E)\rangle \\
&\le \lambda\|Q_{\bar\Omega}\|_{\mathrm{vec}(1)}\big(1 + \mu^{-1}\lambda^{-1}\|\mathcal{P}_{\bar\Omega}(E)\|_{\mathrm{vec}(\infty)}\big) \\
&\quad + \|Q_{\bar T}\|_*\big(1 + \mu^{-1}\|\mathcal{P}_{\bar T}(E)\|_{2\to2}\big) \\
&\le \lambda\|Q_{\bar\Omega}\|_{\mathrm{vec}(1)}\big(1 + \mu^{-1}\lambda^{-1}\epsilon_{\mathrm{vec}(\infty)}\big) \\
&\quad + \|Q_{\bar T}\|_*\big(1 + 2\mu^{-1}\epsilon_{2\to2}\big).
\end{aligned}$$

■

V. Analysis of constrained formulation

Throughout this section, we fix a target decomposition $(\bar X_S, \bar X_L)$ that satisfies the constraints of (1), and let $(\hat X_S, \hat X_L)$ be the optimal solution to (1). Let $\Delta_S := \hat X_S - \bar X_S$ and $\Delta_L := \hat X_L - \bar X_L$. We show that under the conditions of Theorem 5 with $E = 0$ (recall that this does not mean we assume $Y - \bar X_S - \bar X_L = 0$) and appropriately chosen λ, solving (1) accurately recovers the target decomposition $(\bar X_S, \bar X_L)$.

We decompose the errors into symmetric and anti-symmetric parts $\Delta_{\mathrm{avg}} := (\Delta_S + \Delta_L)/2$ and $\Delta_{\mathrm{mid}} := (\Delta_S - \Delta_L)/2$. The constraints allow us to easily bound Δ_{avg}, so most of the analysis involves bounding Δ_{mid} in terms of Δ_{avg}.

Lemma 11. $\|\Delta_{\mathrm{avg}}\|_{\mathrm{vec}(1)} \le \epsilon_{\mathrm{vec}(1)}$ *and* $\|\Delta_{\mathrm{avg}}\|_* \le \epsilon_*$.

Proof: Since both $(\hat X_S, \hat X_L)$ and $(\bar X_S, \bar X_L)$ are feasible solutions to (1), we have for $\diamondsuit \in \{\mathrm{vec}(1), *\}$,

$$\begin{aligned}
&\|\Delta_{\mathrm{avg}}\|_\diamondsuit = {}^1\!/_2\|\Delta_S + \Delta_L\|_\diamondsuit \\
&= {}^1\!/_2\|(\hat X_S + \hat X_L - Y) - (\bar X_S + \bar X_L - Y)\|_\diamondsuit \\
&\le {}^1\!/_2\left(\|\hat X_S + \hat X_L - Y\|_\diamondsuit + \|\bar X_S + \bar X_L - Y\|_\diamondsuit\right) \\
&\le \epsilon_\diamondsuit.
\end{aligned}$$

■

Lemma 12. *Assume the conditions of Theorem 5 hold with $E = 0$. We have*

$$\begin{aligned}
&\lambda\|\mathcal{P}_{\bar\Omega^\perp}(\Delta_{\mathrm{mid}})\|_{\mathrm{vec}(1)} + \|\mathcal{P}_{\bar T^\perp}(\Delta_{\mathrm{mid}})\|_* \\
&\qquad \le (1 - 1/c)^{-1}\left(\lambda\|\Delta_{\mathrm{avg}}\|_{\mathrm{vec}(1)} + \|\Delta_{\mathrm{avg}}\|_*\right).
\end{aligned}$$

Proof: Let $Q := Q_{\bar\Omega} + Q_{\bar T}$ be the dual certificate guaranteed by Theorem 5. Note that Q satisfies the conditions of Lemma 6, so we have

$$\begin{aligned}
&\lambda\|\bar X_S + \Delta_{\mathrm{mid}}\|_{\mathrm{vec}(1)} + \|\bar X_L - \Delta_{\mathrm{mid}}\|_* \\
&\quad - \lambda\|\bar X_S\|_{\mathrm{vec}(1)} - \|\bar X_L\|_* \\
&\ge (1 - 1/c)\left(\lambda\|\mathcal{P}_{\bar\Omega^\perp}(\Delta_{\mathrm{mid}})\|_{\mathrm{vec}(1)} + \|\mathcal{P}_{\bar T^\perp}(\Delta_{\mathrm{mid}})\|_*\right).
\end{aligned}$$

Using the triangle inequality, we have

$$\begin{aligned}
&\lambda\|\hat X_S\|_{\mathrm{vec}(1)} + \|\hat X_L\|_* \\
&= \lambda\|\bar X_S + \Delta_S\|_{\mathrm{vec}(1)} + \|\bar X_L + \Delta_L\|_* \\
&= \lambda\|\bar X_S + \Delta_{\mathrm{mid}} + \Delta_{\mathrm{avg}}\|_{\mathrm{vec}(1)} \\
&\quad + \|\bar X_L - \Delta_{\mathrm{mid}} + \Delta_{\mathrm{avg}}\|_* \\
&\ge \lambda\|\bar X_S + \Delta_{\mathrm{mid}}\|_{\mathrm{vec}(1)} - \lambda\|\Delta_{\mathrm{avg}}\|_{\mathrm{vec}(1)} \\
&\quad + \|\bar X_L - \Delta_{\mathrm{mid}}\|_* - \|\Delta_{\mathrm{avg}}\|_*.
\end{aligned}$$

Now using the fact that $\lambda\|\hat X_S\|_{\mathrm{vec}(1)} + \|\hat X_L\|_* \le \lambda\|\bar X_S\|_{\mathrm{vec}(1)} + \|\bar X_L\|_*$ gives the claim. ■

Lemma 13. *Let $\bar k := |\operatorname{supp}(\bar X_S)|$. Assume the conditions of Theorem 5 hold with $E = 0$. We have*

$$\begin{aligned}
&\|\mathcal{P}_{\bar\Omega}(\Delta_{\mathrm{mid}})\|_{\mathrm{vec}(1)} \\
&\quad \le \frac{(1 - 1/c)^{-1}}{1 - \alpha(\rho)\beta(\rho)} \cdot (\|\Delta_{\mathrm{avg}}\|_{\mathrm{vec}(1)} + \|\Delta_{\mathrm{avg}}\|_*/\lambda).
\end{aligned}$$

Proof: Because $\Delta_{\mathrm{mid}} = \mathcal{P}_{\bar\Omega}(\Delta_{\mathrm{mid}}) + \mathcal{P}_{\bar\Omega^\perp}(\Delta_{\mathrm{mid}}) = \mathcal{P}_{\bar T}(\Delta_{\mathrm{mid}}) + \mathcal{P}_{\bar T^\perp}(\Delta_{\mathrm{mid}})$, we have the equation

$$\mathcal{P}_{\bar\Omega}(\Delta_{\mathrm{mid}}) - \mathcal{P}_{\bar T}(\Delta_{\mathrm{mid}}) = -\mathcal{P}_{\bar\Omega^\perp}(\Delta_{\mathrm{mid}}) + \mathcal{P}_{\bar T^\perp}(\Delta_{\mathrm{mid}}).$$

Separately applying $\mathcal{P}_{\bar\Omega}$ and $\mathcal{P}_{\bar T}$ to both sides gives

$$\begin{bmatrix} \mathcal{I} & \mathcal{P}_{\bar\Omega} \\ \mathcal{P}_{\bar T} & \mathcal{I} \end{bmatrix} \begin{bmatrix} \mathcal{P}_{\bar\Omega}(\Delta_{\mathrm{mid}}) \\ -\mathcal{P}_{\bar T}(\Delta_{\mathrm{mid}}) \end{bmatrix} = \begin{bmatrix} (\mathcal{P}_{\bar\Omega} \circ \mathcal{P}_{\bar T^\perp})(\Delta_{\mathrm{mid}}) \\ -(\mathcal{P}_{\bar T} \circ \mathcal{P}_{\bar\Omega^\perp})(\Delta_{\mathrm{mid}}) \end{bmatrix}.$$

Under the condition $\alpha(\rho)\beta(\rho) < 1$, Lemma 10 and Lemma 4 imply that

$$\|(\mathcal{I} - \mathcal{P}_{\bar\Omega} \circ \mathcal{P}_{\bar T})^{-1}\|_{\mathrm{vec}(1)\to\mathrm{vec}(1)} \le \frac{1}{1 - \alpha(\rho)\beta(\rho)}$$

and that

$$\begin{aligned}
\mathcal{P}_{\bar\Omega}(\Delta_{\mathrm{mid}}) = (\mathcal{I} - \mathcal{P}_{\bar\Omega} \circ \mathcal{P}_{\bar T})^{-1}\big(&(\mathcal{P}_{\bar\Omega} \circ \mathcal{P}_{\bar T^\perp})(\Delta_{\mathrm{mid}}) \\
&+ (\mathcal{P}_{\bar\Omega} \circ \mathcal{P}_{\bar T} \circ \mathcal{P}_{\bar\Omega^\perp})(\Delta_{\mathrm{mid}})\big).
\end{aligned}$$

Therefore

$$
\begin{aligned}
&\|\mathcal{P}_{\bar{\Omega}}(\Delta_{\mathrm{mid}})\|_{\mathrm{vec}(1)} \\
&\leq \frac{1}{1-\alpha(\rho)\beta(\rho)} \cdot \big(\|(\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}^\perp})(\Delta_{\mathrm{mid}})\|_{\mathrm{vec}(1)} \\
&\quad + \|(\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}} \circ \mathcal{P}_{\bar{\Omega}^\perp})(\Delta_{\mathrm{mid}})\|_{\mathrm{vec}(1)}\big) \\
&\leq \frac{1}{1-\alpha(\rho)\beta(\rho)} \cdot \big(\sqrt{\bar{k}} \cdot \|\mathcal{P}_{\bar{T}^\perp}(\Delta_{\mathrm{mid}})\|_{\mathrm{vec}(2)} \\
&\quad + \alpha(\rho)\beta(\rho) \cdot \|\mathcal{P}_{\bar{\Omega}^\perp}(\Delta_{\mathrm{mid}})\|_{\mathrm{vec}(1)}\big) \quad \text{(Lemma 10)} \\
&\leq \frac{1}{1-\alpha(\rho)\beta(\rho)} \cdot \big(\sqrt{\bar{k}} \cdot \|\mathcal{P}_{\bar{T}^\perp}(\Delta_{\mathrm{mid}})\|_{*} \\
&\quad + \alpha(\rho)\beta(\rho) \cdot \|\mathcal{P}_{\bar{\Omega}^\perp}(\Delta_{\mathrm{mid}})\|_{\mathrm{vec}(1)}\big) \\
&\leq \frac{(1-1/c)^{-1}}{1-\alpha(\rho)\beta(\rho)} \cdot \max\{\sqrt{\bar{k}},\ \alpha(\rho)\beta(\rho)/\lambda\} \\
&\quad \cdot (\lambda\|\Delta_{\mathrm{avg}}\|_{\mathrm{vec}(1)} + \|\Delta_{\mathrm{avg}}\|_{*}) \quad \text{(Lemma 12)} \\
&\leq \frac{(1-1/c)^{-1}}{1-\alpha(\rho)\beta(\rho)} \cdot (\|\Delta_{\mathrm{avg}}\|_{\mathrm{vec}(1)} + \|\Delta_{\mathrm{avg}}\|_{*}/\lambda)
\end{aligned}
$$

where the last inequality uses the facts $\bar{k} \leq \alpha(\rho)^2$, $\alpha(\rho)\beta(\rho) < 1$, and $\lambda\alpha(\rho) \leq 1$ (this last inequality uses the condition in (14)). ■

We now prove Theorem 2, which we restate here for convenience.

Theorem 6 (Theorem 2 restated)**.** *Assume the conditions of Theorem 5 hold with $E = 0$. We have*

$$
\begin{aligned}
&\max\{\|\Delta_S\|_{\mathrm{vec}(1)},\ \|\Delta_L\|_{\mathrm{vec}(1)}\} \\
&\leq \left(1 + (1-1/c)^{-1} \cdot \frac{2-\alpha(\rho)\beta(\rho)}{1-\alpha(\rho)\beta(\rho)}\right) \cdot \epsilon_{\mathrm{vec}(1)} \\
&\qquad + (1-1/c)^{-1} \cdot \frac{2-\alpha(\rho)\beta(\rho)}{1-\alpha(\rho)\beta(\rho)} \cdot \epsilon_{*}/\lambda.
\end{aligned}
$$

If, in addition for some $b \geq \|\bar{X}_L\|_{\mathrm{vec}(\infty)}$, either:

- *the optimization problem* (1) *is augmented with the constraint $\|X_L\|_{\mathrm{vec}(\infty)} \leq b$ (and letting $\widetilde{X}_L := \hat{X}_L$), or*
- *$\hat{X}_L$ is post-processed by replacing $[\hat{X}_L]_{i,j}$ with $[\widetilde{X}_L]_{i,j} := \min\{\max\{[\hat{X}_L]_{i,j}, -b\}, b\}$ for all i, j,*

then we also have

$$
\|\widetilde{X}_L - \bar{X}_L\|_{\mathrm{vec}(2)} \leq \min\left\{\|\Delta_L\|_{\mathrm{vec}(1)},\ \sqrt{2b\|\Delta_L\|_{\mathrm{vec}(1)}}\right\}.
$$

Proof: First note that since $\Delta_S = \Delta_{\mathrm{avg}} + \Delta_{\mathrm{mid}}$ and $\Delta_L = \Delta_{\mathrm{avg}} - \Delta_{\mathrm{mid}}$, we have $\max\{\|\Delta_S\|_{\mathrm{vec}(1)}, \|\Delta_L\|_{\mathrm{vec}(1)}\} \leq \|\Delta_{\mathrm{avg}}\|_{\mathrm{vec}(1)} + \|\Delta_{\mathrm{mid}}\|_{\mathrm{vec}(1)}$. We can bound $\|\Delta_{\mathrm{mid}}\|_{\mathrm{vec}(1)}$ as

$$
\begin{aligned}
\|\Delta_{\mathrm{mid}}\|_{\mathrm{vec}(1)} &\leq \|\mathcal{P}_{\bar{\Omega}^\perp}(\Delta_{\mathrm{mid}})\|_{\mathrm{vec}(1)} + \|\mathcal{P}_{\bar{\Omega}}(\Delta_{\mathrm{mid}})\|_{\mathrm{vec}(1)} \\
&\leq (1-1/c)^{-1} \cdot \left(1 + \frac{1}{1-\alpha(\rho)\beta(\rho)}\right) \\
&\quad \cdot (\|\Delta_{\mathrm{avg}}\|_{\mathrm{vec}(1)} + \|\Delta_{\mathrm{avg}}\|_{*}/\lambda)
\end{aligned}
$$

by Lemma 12 and Lemma 13. The bounds on $\|\Delta_S\|_{\mathrm{vec}(1)}$ and $\|\Delta_L\|_{\mathrm{vec}(1)}$ follow from the bounds on $\|\Delta_{\mathrm{mid}}\|_{\mathrm{vec}(1)}$, $\|\Delta_{\mathrm{avg}}\|_{\mathrm{vec}(1)}$, and $\|\Delta_{\mathrm{avg}}\|_{*}$ (from Lemma 11).

If the constraint $\|X_L\|_{\mathrm{vec}(\infty)} \leq b$ is added, then we can use the facts

$$
\begin{aligned}
\|\Delta_L\|_{\mathrm{vec}(\infty)} &\leq \|\hat{X}_L\|_{\mathrm{vec}(\infty)} + \|\bar{X}_L\|_{\mathrm{vec}(\infty)} \\
&\leq 2b
\end{aligned}
$$

and

$$
\begin{aligned}
\|\Delta_L\|_{\mathrm{vec}(2)} &\leq \sqrt{\|\Delta_L\|_{\mathrm{vec}(\infty)}\|\Delta_L\|_{\mathrm{vec}(1)}} \\
&\leq \sqrt{2b\|\Delta_L\|_{\mathrm{vec}(1)}}.
\end{aligned}
$$

If $\hat{X}_L$ is post-processed, then (letting $\mathrm{clip}(\hat{X}_L)$ be the result of the post-processing)

$$
|[\widetilde{X}_L]_{i,j} - [\bar{X}_L]_{i,j}| \leq |[\hat{X}_L]_{i,j} - [\bar{X}_L]_{i,j}|
$$

for all i, j, so

$$
\|\widetilde{X}_L - \bar{X}_L\|_{\mathrm{vec}(1)} \leq \|\Delta_L\|_{\mathrm{vec}(1)}
$$

and

$$
\begin{aligned}
\|\widetilde{X}_L - \bar{X}_L\|_{\mathrm{vec}(2)} &\leq \sqrt{2b\|\widetilde{X}_L - \bar{X}_L\|_{\mathrm{vec}(1)}} \\
&\leq \sqrt{2b\|\Delta_L\|_{\mathrm{vec}(1)}}.
\end{aligned}
$$

■

VI. Analysis of regularized formulation

Throughout this section, we fix a target decomposition $(\bar{X}_S, \bar{X}_L)$ that satisfies $\|\bar{X}_S - Y\|_{\mathrm{vec}(\infty)} \leq b$, and let $(\hat{X}_S, \hat{X}_L)$ be the optimal solution to (2) augmented with the constraint $\|X_S - Y\|_{\mathrm{vec}(\infty)} \leq b$ for some $b \geq \|\bar{X}_S - Y\|_{\mathrm{vec}(\infty)}$ ($b = \infty$ is allowed). Let $\Delta_S := \hat{X}_S - \bar{X}_S$ and $\Delta_L := \hat{X}_L - \bar{X}_L$. We show that under the conditions of Theorem 5 with $E = Y - (\bar{X}_S + \bar{X}_L)$ and appropriately chosen λ and μ, solving (2) accurately recovers the target decomposition $(\bar{X}_S, \bar{X}_L)$.

Lemma 14. *There exists $G_S, G_L, H_S \in \mathbb{R}^{m \times n}$ such that*

1) $\mu^{-1}(\hat{X}_S + \hat{X}_L - Y) + \lambda G_S + H_S = 0$*;* $\|G_S\|_{\mathrm{vec}(\infty)} \leq 1$*;*
2) $\mu^{-1}(\hat{X}_S + \hat{X}_L - Y) + G_L = 0$*;* $\|G_L\|_{2\to 2} \leq 1$*;*
3) $[H_S]_{i,j}[\Delta_S]_{i,j} \geq 0\ \forall i, j$.

Proof: We express the constraint $\|X_S - Y\|_{\mathrm{vec}(\infty)} \leq b$ in (2) as $2mn$ constraints $[X_S]_{i,j} - Y_{i,j} - b \leq 0$ and $-[X_S]_{i,j} + Y_{i,j} - b \leq 0$ for all i, j. Now the corresponding Lagrangian is

$$
\begin{aligned}
&\frac{1}{2\mu}\|X_S + X_L - Y\|_{\mathrm{vec}(2)}^2 + \lambda\|X_S\|_{\mathrm{vec}(1)} + \|X_L\|_{*} \\
&+ \langle \Lambda^+, X_S - Y - b1_{m,n}\rangle + \langle \Lambda^-, -X_S + Y - b1_{m,n}\rangle
\end{aligned}
$$

where $\Lambda^+, \Lambda^- \geq 0$ and $1_{m,n}$ is the all-ones $m \times n$ matrix. First-order optimality conditions imply that there exists a subgradient G_S of $\|X_S\|_{\text{vec}(1)}$ at $X_S = \hat{X}_S$ and a subgradient G_L of $\|X_L\|_*$ at $X_L = \hat{X}_L$ such that

$$\mu^{-1}(\hat{X}_S + \hat{X}_L - Y) + \lambda G_S + (\Lambda^+ - \Lambda^-) = 0$$

and

$$\mu^{-1}(\hat{X}_S + \hat{X}_L - Y) + G_L = 0.$$

Now since $\|\bar{X}_S - Y\|_{\text{vec}(\infty)} \leq b$, we have $[\bar{X}_S]_{i,j} \leq Y_{i,j} + b$ and $-[\bar{X}_S]_{i,j} \leq -Y_{i,j} + b$. By complementary slackness, if $\Lambda^+_{i,j} > 0$, then $[\hat{X}_S]_{i,j} - Y_{i,j} - b = 0$, which means $[\hat{X}_S]_{i,j} - [\bar{X}_S]_{i,j} \geq [\hat{X}_S]_{i,j} - (Y_{i,j} + b) = 0$. So $\Lambda^+_{i,j}[\Delta_S]_{i,j} \geq 0$. Similarly, if $\Lambda^-_{i,j} > 0$, then $[\hat{X}_S]_{i,j} - [\bar{X}_S]_{i,j} \leq 0$. So $\Lambda^-_{i,j}[\Delta_S]_{i,j} \leq 0$. Therefore $H := \Lambda^+ - \Lambda^-$ satisfies $H_{i,j}[\Delta_S]_{i,j} \geq 0$. ■

Lemma 15. *Assume the conditions of Theorem 5 hold with $E = Y - (\bar{X}_S + \bar{X}_L)$, and let $(Q_{\bar{\Omega}}, Q_{\bar{T}})$ be the dual certificate from the conclusion. We have*

$$\lambda\|\mathcal{P}_{\bar{\Omega}^\perp}(\Delta_S)\|_{\text{vec}(1)} + \|\mathcal{P}_{\bar{T}^\perp}(\Delta_L)\|_* \leq (1 - 1/c)^{-1}\|Q_{\bar{\Omega}} + Q_{\bar{T}}\|^2_{\text{vec}(2)}\mu/2.$$

Proof: Let $Q := Q_{\bar{\Omega}} + Q_{\bar{T}}$ and $\Delta := \Delta_S + \Delta_L$. Since $Q + \mu^{-1}E$ satisfies the conditions of Lemma 6,

$$\begin{aligned}&(1 - 1/c)\left(\lambda\|\mathcal{P}_{\bar{\Omega}^\perp}(\Delta_S)\|_{\text{vec}(1)} + \|\mathcal{P}_{\bar{T}^\perp}(\Delta_L)\|_*\right)\\ &\leq (\lambda\|\hat{X}_S\|_{\text{vec}(1)} + \|\hat{X}_L\|_*) - (\lambda\|\bar{X}_S\|_{\text{vec}(1)} + \|\bar{X}_L\|_*)\\ &\quad - \langle Q + \mu^{-1}E, \Delta_S + \Delta_L\rangle.\end{aligned}$$

Furthermore, by the optimality of $(\hat{X}_S, \hat{X}_L)$,

$$\begin{aligned}&(\lambda\|\hat{X}_S\|_{\text{vec}(1)} + \|\hat{X}_L\|_*) - (\lambda\|\bar{X}_S\|_{\text{vec}(1)} + \|\bar{X}_L\|_*)\\ &\leq \frac{1}{2\mu}\|\bar{X}_S + \bar{X}_L - Y\|^2_{\text{vec}(2)} - \frac{1}{2\mu}\|\hat{X}_S + \hat{X}_L - Y\|^2_{\text{vec}(2)}\\ &= \frac{1}{2\mu}\|E\|^2_{\text{vec}(2)} - \frac{1}{2\mu}\|\Delta_S + \Delta_L - E\|^2_{\text{vec}(2)}\\ &= \frac{1}{2\mu}(2\langle E, \Delta\rangle - \langle \Delta, \Delta\rangle).\end{aligned}$$

Combining the inequalities gives

$$\begin{aligned}&(1 - 1/c)\left(\lambda\|\mathcal{P}_{\bar{\Omega}^\perp}(\Delta_S)\|_{\text{vec}(1)} + \|\mathcal{P}_{\bar{T}^\perp}(\Delta_L)\|_*\right)\\ &\leq -\langle Q, \Delta\rangle - \frac{1}{2\mu}\langle \Delta, \Delta\rangle \leq \|Q\|^2_{\text{vec}(2)}\mu/2\end{aligned}$$

where the last inequality follows by taking the maximum value over Δ at $\Delta = -\mu Q$. ■

Now we prove Theorem 3, restated below (with an additional result for $\|\Delta_L\|_{\flat(\rho)}$).

Theorem 7 (Theorem 3 restated)**.** *Let $\bar{k} := |\operatorname{supp}(\bar{X}_S)|$ and $\bar{r} := \operatorname{rank}(\bar{X}_L)$. Assume the conditions of Theorem 5 hold with $E = Y - (\bar{X}_S + \bar{X}_L)$, and let $(Q_{\bar{\Omega}}, Q_{\bar{T}})$ be the dual certificate from the conclusion. We have*

$$\begin{aligned}&(1 - \alpha(\rho)\beta(\rho)) \cdot \|\Delta_S\|_{\text{vec}(1)}\\ &\leq \lambda^{-1}(1 - 1/c)^{-1}\|Q_{\bar{\Omega}} + Q_{\bar{T}}\|^2_{\text{vec}(2)}\mu\\ &\quad + \lambda\bar{k}\mu + 2\sqrt{\bar{k}\bar{r}}\mu + \bar{k}\|(\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}^\perp})(E)\|_{\text{vec}(\infty)},\end{aligned}$$

$$\|\Delta_S\|_{\text{vec}(2)} \leq \min\left\{\|\Delta_S\|_{\text{vec}(1)},\ \sqrt{2b\|\Delta_S\|_{\text{vec}(1)}}\right\},$$

$$\begin{aligned}\|\Delta_L\|_{\flat(\rho)} &\leq (1 - 1/c)^{-1}\|Q_{\bar{\Omega}} + Q_{\bar{T}}\|^2_{\text{vec}(2)}\mu/2\\ &\quad + \min\left\{\beta(\rho)\|\Delta_S\|_{\text{vec}(1)},\ \sqrt{2\bar{r}}\|\Delta_S\|_{\text{vec}(2)}\right\}\\ &\quad + \|\mathcal{P}_{\bar{T}}(E)\|_* + 2\bar{r}\mu,\end{aligned}$$

and

$$\begin{aligned}\|\Delta_L\|_* &\leq (1 - 1/c)^{-1}\|Q_{\bar{\Omega}} + Q_{\bar{T}}\|^2_{\text{vec}(2)}\mu/2\\ &\quad + \sqrt{2\bar{r}}\|\Delta_S\|_{\text{vec}(2)} + \|\mathcal{P}_{\bar{T}}(E)\|_* + 2\bar{r}\mu.\end{aligned}$$

Proof: From Lemma 14, we obtain $G_S, G_L, H_S \in \mathbb{R}^{m\times n}$ and the following equations:

$$\begin{aligned}&\lambda\mathcal{P}_{\bar{\Omega}}(G_S) = -\mu^{-1}(\mathcal{P}_{\bar{\Omega}}(\Delta_S) + \mathcal{P}_{\bar{\Omega}}(\Delta_L) - \mathcal{P}_{\bar{\Omega}}(E))\\ &\quad - \mathcal{P}_{\bar{\Omega}}(H_S) \qquad (16)\\ &\mathcal{P}_{\bar{T}}(G_L) = -\mu^{-1}(\mathcal{P}_{\bar{T}}(\Delta_S) + \mathcal{P}_{\bar{T}}(\Delta_L) - \mathcal{P}_{\bar{T}}(E)) \qquad (17)\\ &(\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}})(G_L) = -\mu^{-1}\big((\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}})(\Delta_S)\\ &\quad + (\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}})(\Delta_L) - (\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}})(E)\big). \qquad (18)\end{aligned}$$

Subtracting (18) from (16) gives

$$\begin{aligned}&\mu^{-1}\big(\mathcal{P}_{\bar{\Omega}}(\Delta_S) - (\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}} \circ \mathcal{P}_{\bar{\Omega}})(\Delta_S)\\ &\quad - (\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}} \circ \mathcal{P}_{\bar{\Omega}^\perp})(\Delta_S) + (\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}^\perp})(\Delta_L)\big)\\ &\quad + \mathcal{P}_{\bar{\Omega}}(H_S)\\ &= -\lambda\mathcal{P}_{\bar{\Omega}}(G_S) + (\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}})(G_L)\\ &\quad + \mu^{-1}(\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}^\perp})(E).\end{aligned}$$

Moreover, we have $\langle \operatorname{sign}(\Delta_S), \mathcal{P}_{\bar{\Omega}}(\Delta_S)\rangle = \|\mathcal{P}_{\bar{\Omega}}(\Delta_S)\|_{\text{vec}(1)}$ and $\langle \operatorname{sign}(\Delta_S), \mathcal{P}_{\bar{\Omega}}(H_S)\rangle = \|\mathcal{P}_{\bar{\Omega}}(H_S)\|_{\text{vec}(1)}$, so taking inner products with $\operatorname{sign}(\Delta_S)$ on both sides of the equation gives the following chain of inequalities:

$$\begin{aligned}
&\mu^{-1}\|\mathcal{P}_{\bar{\Omega}}(\Delta_S)\|_{\mathrm{vec}(1)} + \|\mathcal{P}_{\bar{\Omega}}(H_S)\|_{\mathrm{vec}(1)} \\
&\leq \mu^{-1}\|(\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}} \circ \mathcal{P}_{\bar{\Omega}})(\Delta_S)\|_{\mathrm{vec}(1)} \\
&\quad + \mu^{-1}\|(\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}} \circ \mathcal{P}_{\bar{\Omega}^\perp})(\Delta_S)\|_{\mathrm{vec}(1)} \\
&\quad + \mu^{-1}\|(\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}^\perp})(\Delta_L)\|_{\mathrm{vec}(1)} \\
&\quad + \lambda\|\mathcal{P}_{\bar{\Omega}}(G_S)\|_{\mathrm{vec}(1)} + \|(\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}})(G_L)\|_{\mathrm{vec}(1)} \\
&\quad + \mu^{-1}\|(\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}^\perp})(E)\|_{\mathrm{vec}(1)} \\
&\leq \mu^{-1}\alpha(\rho)\beta(\rho)\|\mathcal{P}_{\bar{\Omega}}(\Delta_S)\|_{\mathrm{vec}(1)} \\
&\quad + \mu^{-1}\alpha(\rho)\beta(\rho)\|\mathcal{P}_{\bar{\Omega}^\perp}(\Delta_S)\|_{\mathrm{vec}(1)} + \lambda\bar{k} \\
&\quad + \mu^{-1}\sqrt{\bar{k}}\|\mathcal{P}_{\bar{T}^\perp}(\Delta_L)\|_{\mathrm{vec}(2)} + \sqrt{\bar{k}}\|\mathcal{P}_{\bar{T}}(G_L)\|_{\mathrm{vec}(2)} \\
&\quad + \mu^{-1}\bar{k}\|(\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}^\perp})(E)\|_{\mathrm{vec}(\infty)} \\
&\leq \mu^{-1}\alpha(\rho)\beta(\rho)\|\mathcal{P}_{\bar{\Omega}}(\Delta_S)\|_{\mathrm{vec}(1)} \\
&\quad + \mu^{-1}\alpha(\rho)\beta(\rho)\|\mathcal{P}_{\bar{\Omega}^\perp}(\Delta_S)\|_{\mathrm{vec}(1)} \\
&\quad + \mu^{-1}\sqrt{\bar{k}}\|\mathcal{P}_{\bar{T}^\perp}(\Delta_L)\|_{\mathrm{vec}(2)} \\
&\quad + \lambda\bar{k} + 2\sqrt{\bar{k}\bar{r}}\|G_L\|_{2\to 2} \\
&\quad + \mu^{-1}\bar{k}\|(\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}^\perp})(E)\|_{\mathrm{vec}(\infty)} \\
&\leq \mu^{-1}\alpha(\rho)\beta(\rho)\|\mathcal{P}_{\bar{\Omega}}(\Delta_S)\|_{\mathrm{vec}(1)} \\
&\quad + \mu^{-1}\alpha(\rho)\beta(\rho)\|\mathcal{P}_{\bar{\Omega}^\perp}(\Delta_S)\|_{\mathrm{vec}(1)} \\
&\quad + \mu^{-1}\sqrt{\bar{k}}\|\mathcal{P}_{\bar{T}^\perp}(\Delta_L)\|_{\mathrm{vec}(2)} \\
&\quad + \lambda\bar{k} + 2\sqrt{\bar{k}\bar{r}} + \mu^{-1}\bar{k}\|(\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}^\perp})(E)\|_{\mathrm{vec}(\infty)}.
\end{aligned}$$

The second and third inequalities above follow from Lemma 5 and Lemma 10, and the fourth inequality uses the fact that $\|G_L\|_{2\to 2} \leq 1$. Rearranging the inequality and applying Lemma 15 gives

$$\begin{aligned}
&(1 - \alpha(\rho)\beta(\rho)) \cdot \|\mathcal{P}_{\bar{\Omega}}(\Delta_S)\|_{\mathrm{vec}(1)} \\
&\leq \alpha(\rho)\beta(\rho)\|\mathcal{P}_{\bar{\Omega}^\perp}(\Delta_S)\|_{\mathrm{vec}(1)} + \sqrt{\bar{k}}\|\mathcal{P}_{\bar{T}^\perp}(\Delta_L)\|_{\mathrm{vec}(2)} \\
&\quad + \lambda\bar{k}\mu + 2\sqrt{\bar{k}\bar{r}}\mu + \bar{k}\|(\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}^\perp})(E)\|_{\mathrm{vec}(\infty)} \\
&\leq \max\{\alpha(\rho)\beta(\rho)/\lambda,\ \sqrt{\bar{k}}\} \\
&\quad \cdot (1 - 1/c)^{-1}\|Q_{\bar{\Omega}} + Q_{\bar{T}}\|^2_{\mathrm{vec}(2)}\mu/2 \\
&\quad + \lambda\bar{k}\mu + 2\sqrt{\bar{k}\bar{r}}\mu + \bar{k}\|(\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}^\perp})(E)\|_{\mathrm{vec}(\infty)} \\
&\leq \lambda^{-1}(1 - 1/c)^{-1}\|Q_{\bar{\Omega}} + Q_{\bar{T}}\|^2_{\mathrm{vec}(2)}\mu/2 \\
&\quad + \lambda\bar{k}\mu + 2\sqrt{\bar{k}\bar{r}}\mu + \bar{k}\|(\mathcal{P}_{\bar{\Omega}} \circ \mathcal{P}_{\bar{T}^\perp})(E)\|_{\mathrm{vec}(\infty)}
\end{aligned}$$

since $\bar{k} \leq \alpha(\rho)^2$, $\alpha(\rho)\beta(\rho) < 1$, and $\lambda\alpha(\rho) \leq 1$. Now we combine this with $\|\Delta_S\|_{\mathrm{vec}(1)} \leq \|\mathcal{P}_{\bar{\Omega}^\perp}(\Delta_S)\|_{\mathrm{vec}(1)} + \|\mathcal{P}_{\bar{\Omega}}(\Delta_S)\|_{\mathrm{vec}(1)}$ and Lemma 15 to get the first bound.

For the second bound, we use the facts $\|\Delta_S\|_{\mathrm{vec}(\infty)} \leq \|\hat{X}_S - Y\|_{\mathrm{vec}(\infty)} + \|\bar{X}_S - Y\|_{\mathrm{vec}(\infty)} \leq 2b$ and $\|\Delta_S\|_{\mathrm{vec}(2)} \leq \sqrt{\|\Delta_S\|_{\mathrm{vec}(1)}\|\Delta_S\|_{\mathrm{vec}(\infty)}} \leq \sqrt{2b\|\Delta_S\|_{\mathrm{vec}(1)}}$.

For the third and fourth bounds, we obtain from (17)

$$\begin{aligned}
&\|\mathcal{P}_{\bar{T}}(\Delta_L)\|_{\flat(\rho)} \\
&\leq \|\mathcal{P}_{\bar{T}}(\Delta_S)\|_{\flat(\rho)} + \|\mathcal{P}_{\bar{T}}(E)\|_{\flat(\rho)} + \mu\|\mathcal{P}_{\bar{T}}(G_L)\|_{\flat(\rho)} \\
&\leq \|\mathcal{P}_{\bar{T}}\|_{\mathrm{vec}(1)\to\flat(\rho)}\|\Delta_S\|_{\mathrm{vec}(1)} + \|\mathcal{P}_{\bar{T}}(E)\|_* \\
&\quad + \mu\|\mathcal{P}_{\bar{T}}(G_L)\|_* \quad \text{(Lemma 3)} \\
&= \|\mathcal{P}^*_{\bar{T}}\|_{\sharp(\rho)\to\mathrm{vec}(\infty)}\|\Delta_S\|_{\mathrm{vec}(1)} + \|\mathcal{P}_{\bar{T}}(E)\|_* \\
&\quad + \mu\|\mathcal{P}_{\bar{T}}(G_L)\|_* \quad \text{(Proposition 3)} \\
&\leq \beta(\rho)\|\Delta_S\|_{\mathrm{vec}(1)} + \|\mathcal{P}_{\bar{T}}(E)\|_* + \mu\|\mathcal{P}_{\bar{T}}(G_L)\|_* \\
&\quad \text{(Lemma 8)} \\
&\leq \beta(\rho)\|\Delta_S\|_{\mathrm{vec}(1)} + \|\mathcal{P}_{\bar{T}}(E)\|_* + 2\bar{r}\mu \\
&\quad \text{(Lemma 5 and } \|G_L\|_{2\to 2} \leq 1)
\end{aligned}$$

and

$$\begin{aligned}
&\|\mathcal{P}_{\bar{T}}(\Delta_L)\|_* \\
&\leq \|\mathcal{P}_{\bar{T}}(\Delta_S)\|_* + \|\mathcal{P}_{\bar{T}}(E)\|_* + \mu\|\mathcal{P}_{\bar{T}}(G_L)\|_* \\
&\leq \sqrt{2\bar{r}}\|\Delta_S\|_{\mathrm{vec}(2)} + \|\mathcal{P}_{\bar{T}}(E)\|_* + 2\bar{r}\mu \\
&\quad \text{(Lemma 5 and } \|G_L\|_{2\to 2} \leq 1).
\end{aligned}$$

Now we combine these with

$$\begin{aligned}
\|\Delta_L\|_{\flat(\rho)} &\leq \|\mathcal{P}_{\bar{T}^\perp}(\Delta_L)\|_{\flat(\rho)} + \|\mathcal{P}_{\bar{T}}(\Delta_L)\|_{\flat(\rho)} \\
&\leq \|\mathcal{P}_{\bar{T}^\perp}(\Delta_L)\|_* \\
&\quad + \min\{\|\mathcal{P}_{\bar{T}}(\Delta_L)\|_*, \|\mathcal{P}_{\bar{T}}(\Delta_L)\|_{\flat(\rho)}\} \\
&\quad \text{(Lemma 3)} \\
\|\Delta_L\|_* &\leq \|\mathcal{P}_{\bar{T}^\perp}(\Delta_L)\|_* + \|\mathcal{P}_{\bar{T}}(\Delta_L)\|_*
\end{aligned}$$

and Lemma 15. ■

Note that we have an error bound for Δ_L in $\|\cdot\|_{\flat(\rho)}$ norm, which can be significantly smaller than the bound for the trace norm of Δ_L.

Acknowledgment

We thank Emmanuel Candès for clarifications about the results in [3].

References

[1] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, "Rank-Sparsity Incoherence for Matrix Decomposition," *ArXiv e-prints*, Jun. 2009.

[2] V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky, "Latent Variable Graphical Model Selection via Convex Optimization," *ArXiv e-prints*, Aug. 2010.

[3] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust Principal Component Analysis?" *ArXiv e-prints*, Dec. 2009.

[4] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal. Statist. Soc B.*, vol. 58, no. 1, pp. 267–288, 1996.

[5] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, Department of Electrical Engineering, Stanford University, 2002.

[6] E. Candès and R. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, pp. 717–772, 2009.

[7] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," *ArXiv e-prints*, Oct. 2009.
[8] Z. Zhou, X. Li, J. Wright, E. J. Candès, and Y. Ma, "Stable principal component pursuit," in *Proceedings of International Symposium on Information Theory*, 2010.
[9] A. Ganesh, J. Wright, X. Li, E. J. Candès, and Y. Ma., "Dense error correction for low-rank matrices via principal component pursuit," in *Proceedings of International Symposium on Information Theory*, 2010.
[10] K. R. Davidson and S. J. Szarek, "Local operator theory, random matrices and banach spaces," in *Handbook of the geometry of Banach spaces, Vol. I*, 2001, pp. 317–366.
[11] G. A. Watson, "Characterization of the subdifferential of some matrix norms," *Linear Algebra and Applications*, vol. 170, pp. 1039–1053, 1992.

Daniel Hsu received a B.S. in Computer Science and Engineering from the University of California, Berkeley in 2004 and a Ph.D. in Computer Science from the University of California, San Diego in 2010. From 2010 to 2011, he was a postdoctoral scholar at Rutgers University and a visiting scholar at the Wharton School of the University of Pennsylvania. He is currently a postdoctoral researcher at Microsoft Research New England. His research interests are in algorithmic statistics and machine learning.

Sham M. Kakade is currently an associate professor of statistics at the Wharton School at the University of Pennsylvania, and a Senior Researcher at Microsoft Research New England. He received his B.A. in Physics from the California Institute of Technology and his Ph.D. from the Gatsby Computational Neuroscience Unit affiliated with University College London. He spent the following two years as a Postdoctoral Researcher at the Department of Computer and Information Science at the University of Pennsylvania. Subsequently, he joined the Toyota Technological Institute, where he was an assistant professor for four years. His research focuses on artificial intelligence and machine learning, and their connections to other areas such as game theory and economics.

Tong Zhang received a B.A. in Mathematics and Computer Science from Cornell University in 1994 and a Ph.D. in Computer Science from Stanford University in 1998. After graduation, he worked at IBM T.J. Watson Research Center in Yorktown Heights, New York, and Yahoo Research in New York city. He is currently a professor of statistics at Rutgers University. His research interests include machine learning, algorithms for statistical computation, their mathematical analysis and applications.