A Proof of Lemma 2

Proof Suppose the SVD of X_0 is $U_0\Sigma_0V_0^T$, and the SVD of C_0 is $U_C\Sigma_CV_C^T$. Suppose $U_0^\perp$ and $U_C^\perp$ are the orthogonal complements of U_0 and U_C, respectively. By the independence between span (C_0) and span (X_0), $[U_0^\perp, U_C^\perp]$ spans the whole ambient space, and thus the following linear equation system has feasible solutions Y_0 and Y_C:

$$U_0^\perp(U_0^\perp)^TY_0 + U_C^\perp(U_C^\perp)^TY_C = \mathtt{I}.$$

Let $Y = \mathtt{I} - U_0^\perp(U_0^\perp)^TY_0$, then it can be computed that

$$X_0^TY = X_0^T \quad \text{and} \quad C_0^TY = 0,$$

i.e., $X_0 = YX_0$ and $YC_0 = 0$ are feasible. By $\mathcal{P}_{\mathcal{I}_0^c}(X) = X_0$, $\mathcal{P}_{\mathcal{I}_0}(X) = C_0$, $\mathcal{P}_{\mathcal{I}_0}(X_0) = X_0$ and $\mathcal{P}_{\mathcal{I}_0^c}(X_0) = 0$, the following linear equation system has feasible solutions Y:

$$X_0 = YX,$$

which simply leads to $V_0 \in \mathcal{P}_{V_X}^L$.

B Proof of Lemma 3

Proof Suppose $U_X\Sigma_XV_X^T$ is the SVD of X, $U_0\Sigma_0V_0^T$ is the SVD of X_0, U_C is the column space of C_0, and $U_C^\perp$ is the orthogonal complement of U_C. By $X = X_0 + C_0$, $(U_C^\perp)^TX = (U_C^\perp)^TX_0$ and thus

$$(U_C^\perp)^TU_X\Sigma_XV_X^T \quad = \quad (U_C^\perp)^TU_0\Sigma_0V_0^T,$$

from which it can be deduced that

$$(U_C^\perp)^TU_X = (U_C^\perp)^TU_0\Sigma_0(V_0^TV_X\Sigma_X^{-1}).$$

Since span (C_0) and span (X_0) are independent to each other, $(U_C^\perp)^TU_0$ is of full column rank. Let the SVD of $(U_C^\perp)^TU_0$ be $U_1\Sigma_1V_1^T$, then we have

$$V_0^TV_X\Sigma_X^{-1} = \Sigma_0^{-1}V_1\Sigma_1^{-1}U_1^T(U_C^\perp)^TU_X.$$

Hence,

$$\begin{aligned}\|V_0^TV_X\Sigma_X^{-1}\| &= \|\Sigma_0^{-1}V_1\Sigma_1^{-1}U_1^T(U_C^\perp)^TU_X\| \leq \|\Sigma_0^{-1}\|\|\Sigma_1^{-1}\| \\ &= \frac{1}{\sigma_{min}(X_0)\sin(\theta)},\end{aligned}$$

where $\|\Sigma_1^{-1}\| = 1/\sin(\theta)$ is concluded from (Knyazev et al., 2002). By $\|X\| \leq \|X_0\| + \|C_0\|$, we further have

$$\begin{aligned}\beta &= \frac{1}{\|\Sigma_X^{-1}V_X^TV_0\|\|X\|} \geq \frac{\sigma_{min}(X_0)\sin(\theta)}{\|X\|} \geq \frac{\sigma_{min}(X_0)\sin(\theta)}{\|X_0\| + \|C_0\|} \\ &= \frac{\sin(\theta)}{cond(X_0)(1 + \frac{\|C_0\|}{\|X_0\|})}.\end{aligned}$$

C Proof of Theorem 1

C.1 Roadmap of the Proof

In this section we provide an outline for the proof of Theorem 1. The proof follows three main steps.

1. **Equivalent Conditions:** Identify the necessary and sufficient conditions (called equivalent conditions), for any pair (Z', C') to produce the exact results (7).

For any feasible pair (Z', C') that satisfies $X = XZ' + C'$, let the SVD of Z' as $U'\Sigma'V'^T$ and the column support of C' as $\mathcal{I}'$. In order to produce the exact results (7), on the one hand, a necessary condition is that $\mathcal{P}_{V_0}^L(Z') = Z'$ and $\mathcal{P}_{\mathcal{I}_0}(C') = C'$, as this is nothing but U' is a subspace of V_0 and $\mathcal{I}'$ is a subset of $\mathcal{I}_0$. On the other hand, it can be proven that $\mathcal{P}_{V_0}^L(Z') = Z'$ and $\mathcal{P}_{\mathcal{I}_0}(C') = C'$ are sufficient to ensure $U'U'^T = V_0V_0^T$ and $\mathcal{I}' = \mathcal{I}_0$. So, the exactness described in (7) can be equally transformed into two constraints: $\mathcal{P}_{V_0}^L(Z') = Z'$ and $\mathcal{P}_{\mathcal{I}_0}(C') = C'$, which can be imposed as additional constraints in (2).

2. **Dual Conditions:** For a candidate pair (Z', C') that respectively has the desired row space and column support, identify the sufficient conditions for (Z', C') to be an optimal solution to the LRR problem (2). These conditions are call dual conditions.

   For the pair (Z', C') that satisfies $X = XZ' + C'$, $\mathcal{P}_{V_0}^L(Z') = Z'$ and $\mathcal{P}_{\mathcal{I}_0}(C') = C'$, let the SVD of Z' as $U'\Sigma'V'^T$ and the column-normalized version of C' as H'. That is, column $[H']_i = \frac{[C']_i}{\|[C']_i\|_2}$ for all $i \in \mathcal{I}_0$, and $[H']_i = 0$ for all $i \notin \mathcal{I}_0$ (note that the column support of C' is $\mathcal{I}_0$). Furthermore, define $\mathcal{P}_{T'}(\cdot) = \mathcal{P}_{U'}(\cdot) + \mathcal{P}_{V'}(\cdot) - \mathcal{P}_{U'}\mathcal{P}_{V'}(\cdot)$. With these notations, it can be proven that (Z', C') is an optimal solution to LRR if there exists a matrix Q that satisfies

$$\mathcal{P}_{T'}(X^TQ) = U'V'^T \qquad \|X^TQ - \mathcal{P}_{T'}(X^TQ)\| < 1$$
$$\mathcal{P}_{\mathcal{I}_0}(Q) = \lambda H' \qquad \|Q - \mathcal{P}_{\mathcal{I}_0}(Q)\|_{2,\infty} < \lambda.$$

   Although the LRR problem (2) may have multiple solutions, it can be further proven that any solution has the desired row space and column support, provided the above conditions have been satisfied. So, the left job is to prove the above dual conditions, i.e., construct the dual certificates.

3. **Dual Certificates:** Show that the dual conditions can be satisfied, i.e., construct the *dual certificates*.

   The construction of dual certificates mainly concerns a matrix Q that satisfies the dual conditions. However, since the dual conditions also depend on the pair (Z', C'), we actually need to obtain three matrices, Z', C' and Q. This is done by considering an alternate optimization problem, often called the "oracle problem". The oracle problem arises by imposing the success conditions as additional constraints in (2):

$$\textbf{Oracle Problem:} \quad \min_{Z,C} \|Z\|_* + \lambda\|C\|_{2,1}$$
$$X = XZ + C, \mathcal{P}_{V_0}^L(Z) = Z, \mathcal{P}_{\mathcal{I}_0}(C) = C.$$

   Note that the above problem is always feasible, as $(V_0V_0^T, C_0)$ is feasible. Thus, an optimal solution, denoted as $(\hat{Z}, \hat{C})$, exists. With this perspective, we would like to use $(\hat{Z}, \hat{C})$ to construct the dual certificates. Let the SVD of $\hat{Z}$ be $\hat{U}\hat{\Sigma}\hat{V}^T$, and the column-normalized version of $\hat{C}$ be $\hat{H}$. It is easy to see that there exists an orthonormal matrix $\bar{V}$ such that $\hat{U}\hat{V}^T = V_0\bar{V}^T$, where V_0 is the row space of X_0. Moreover, it is easy to show that $\mathcal{P}_{\hat{U}}(\cdot) = \mathcal{P}_{V_0}^L(\cdot)$, $\mathcal{P}_{\hat{V}}(\cdot) = \mathcal{P}_{\bar{V}}(\cdot)$, and hence the operator $\mathcal{P}_{\hat{T}}$ defined by $\hat{U}$ and $\hat{V}$, obeys $\mathcal{P}_{\hat{T}}(\cdot) = \mathcal{P}_{V_0}^L(\cdot) + \mathcal{P}_{\bar{V}}(\cdot) - \mathcal{P}_{V_0}^L\mathcal{P}_{\bar{V}}(\cdot)$. Finally, the dual certificates are finished by constructing Q as follows:

$$Q_1 \triangleq \lambda\mathcal{P}_{V_0}^L(X^T\hat{H}),$$
$$Q_2 \triangleq \lambda\mathcal{P}_{V_0^\perp}^L\mathcal{P}_{\mathcal{I}_0^c}\mathcal{P}_{\bar{V}}(\mathtt{I} + \sum_{i=1}^{\infty}(\mathcal{P}_{\bar{V}}\mathcal{P}_{\mathcal{I}_0}\mathcal{P}_{\bar{V}})^i)\mathcal{P}_{\bar{V}}(X^T\hat{H}),$$
$$Q \triangleq U_X\Sigma_X^{-1}V_X^T(V_0\bar{V}^T + \lambda X^T\hat{H} - Q_1 - Q_2),$$

   where $U_X\Sigma_XV_X^T$ is the SVD of the data matrix X.

C.2 Equivalent Conditions

Before starting the main proofs, we introduce the following lemmas, which are well-known and will be used multiple times in the proof.

Lemma 4 *For any column space U, row space V and column support $\mathcal{I}$, the following holds.*

1. *Let the SVD of a matrix M be $U\Sigma V^T$, then $\partial\|M\|_* = \{UV^T + W|\mathcal{P}_T(W) = 0, \|W\| \le 1\}$.*

2. *Let the column support of a matrix M be $\mathcal{I}$, then $\partial\|M\|_{2,1} = \{H + L|\mathcal{P}_{\mathcal{I}}(H) = H, [H]_i = [M]_i/\|[M]_i\|_2, \forall i \in \mathcal{I}; \mathcal{P}_{\mathcal{I}}(L) = 0, \|L\|_{2,\infty} \leq 1\}$.*

3. *For any matrices M and N of consistent sizes, we have $\mathcal{P}_{\mathcal{I}}(MN) = M\mathcal{P}_{\mathcal{I}}(N)$.*

4. *For any matrices M and N of consistent sizes, we have $\mathcal{P}_U\mathcal{P}_{\mathcal{I}}(M) = \mathcal{P}_{\mathcal{I}}\mathcal{P}_U(M)$ and $\mathcal{P}_V^L\mathcal{P}_{\mathcal{I}}(N) = \mathcal{P}_{\mathcal{I}}\mathcal{P}_V^L(N)$.*

Lemma 5 *If a matrix H satisfies $\|H\|_{2,\infty} \leq 1$ and is support on $\mathcal{I}$, then $\|H\| \leq \sqrt{|\mathcal{I}|}$.*

Proof This lemma has been proven by (Xu et al., 2011). We present a proof here for the ease of reading.

$$\begin{aligned} \|H\| &= \|H^T\| = \max_{\|x\|_2\leq 1} \|H^T x\|_2 = \max_{\|x\|_2 \leq 1} \|x^T H\|_2 \\ &= \max_{\|x\|_2\leq 1} \sqrt{\sum_{i\in\mathcal{I}} (x^T[H]_i)^2} \leq \sqrt{\sum_{i\in\mathcal{I}} 1} = \sqrt{|\mathcal{I}|}. \end{aligned}$$

Lemma 6 *For any two column-orthonomal matrices U and V of consistent sizes, we have $\|UV^T\|_{2,\infty} = \max_i \|V^T\mathbf{e}_i\|_2$.*

Lemma 7 *For any matrices M and N of consistent sizes, we have*

$$\begin{aligned} &\|MN\|_{2,\infty} \leq \|M\|\|N\|_{2,\infty}, \\ &|\langle M, N\rangle| \leq \|M\|_{2,\infty}\|N\|_{2,1} \end{aligned}$$

Proof We have

$$\begin{aligned} \|MN\|_{2,\infty} &= \max_i \|MN\mathbf{e}_i\|_2 \\ &= \max_i \|M[N]_i\|_2 \leq \max_i \|M\|\|[N]_i\|_2 = \|M\| \max_i \|[N]_i\|_2 \\ &= \|M\|\|N\|_{2,\infty}. \end{aligned}$$

$$\begin{aligned} |\langle M, N\rangle| &= |\sum_i [M]_i^T[N]_i| \leq \sum_i |[M]_i^T[N]_i| \leq \sum_i \|[M]_i\|_2\|[N]_i\|_2 \\ &\leq \sum_i (\max_i \|[M]_i\|_2)\|[N]_i\|_2 = \|M\|_{2,\infty}\|N\|_{2,1}. \end{aligned}$$

The exactness described in (7) seems "mysterious". Actually, they can be "seamlessly" achieved by imposing two additional constraints in (2), as shown in the following theorem.

Theorem 2 *Let the pair (Z', C') satisfy $X = XZ' + C'$. Denote the SVD of Z' as $U'\Sigma'V'^T$, and the column support of C' as $\mathcal{I}'$. If $\mathcal{P}_{V_0}^L(Z') = Z'$ and $\mathcal{P}_{\mathcal{I}_0}(C') = C'$, then $U'U'^T = V_0V_0^T$ and $\mathcal{I}' = \mathcal{I}_0$.*

Remark 3 *The above theorem implies that the exactness described in (7) is equivalent to two linear constraints: $\mathcal{P}_{V_0}^L(Z^*) = Z^*$ and $\mathcal{P}_{\mathcal{I}_0}(C^*) = C^*$. As will be seen, this can largely facilitates the proof of Theorem 1.*

Proof To prove $U'U'^T = V_0V_0^T$, we only need to prove that $\operatorname{rank}(Z') \geq r_0$, as $\mathcal{P}_{V_0}^L(Z') = Z'$ implies that U' is a subspace of V_0. Notice that $\mathcal{P}_{\mathcal{I}_0^c}(X) = X_0$. Then we have

$$\begin{aligned} X_0 &= \mathcal{P}_{\mathcal{I}_0^c}(X) = \mathcal{P}_{\mathcal{I}_0^c}(XZ' + C') = \mathcal{P}_{\mathcal{I}_0^c}(XZ') \\ &= X\mathcal{P}_{\mathcal{I}_0^c}(Z'). \end{aligned}$$

So, $r_0 = \operatorname{rank}(X_0) = \operatorname{rank}\left(X\mathcal{P}_{\mathcal{I}_0^c}(Z')\right) \leq \operatorname{rank}\left(\mathcal{P}_{\mathcal{I}_0^c}(Z')\right) \leq \operatorname{rank}(Z')$.

To ensure $\mathcal{I}' = \mathcal{I}_0$, we only need to prove that $\mathcal{I}_0 \cap \mathcal{I}'^c = \emptyset$, since $\mathcal{P}_{\mathcal{I}_0}(C') = C'$ has produced $\mathcal{I}' \subseteq \mathcal{I}_0$. Via some computations, we have that

$$\begin{aligned} \mathcal{P}_{\mathcal{I}_0}(X_0) = 0 &\Rightarrow U_0 \Sigma_0 \mathcal{P}_{\mathcal{I}_0}(V_0^T) = 0 \\ &\Rightarrow \mathcal{P}_{\mathcal{I}_0}(V_0^T) = 0 \\ &\Rightarrow V_0 \mathcal{P}_{\mathcal{I}_0}(V_0^T) = 0. \end{aligned} \tag{8}$$

Also, we have

$$\begin{aligned} V_0 \in \mathcal{P}_{V_X}^L &\Rightarrow V_0^T = V_0^T V_X V_X^T \\ &\Rightarrow V_0 V_0^T = V_0 V_0^T V_X V_X^T, \end{aligned} \tag{9}$$

which simply leads to $V_0 V_0^T V_X \mathcal{P}_{\mathcal{I}_0}(V_X^T) = V_0 \mathcal{P}_{\mathcal{I}_0}(V_0^T)$. Recalling (8), we further have

$$\begin{aligned} V_0 \mathcal{P}_{\mathcal{I}_0}(V_0^T) = 0 &\Rightarrow V_0 V_0^T V_X \mathcal{P}_{\mathcal{I}_0}(V_X^T) = V_0 \mathcal{P}_{\mathcal{I}_0}(V_0^T) = 0 \\ &\Rightarrow V_0 V_0^T V_X \mathcal{P}_{\mathcal{I}_0 \cap \mathcal{I}'^c}(V_X^T) = 0, \end{aligned} \tag{10}$$

where the last equality holds because $\mathcal{I}_0 \cap \mathcal{I}'^c \subseteq \mathcal{I}_0$. Also, note that $\mathcal{I}_0 \cap \mathcal{I}'^c \subseteq \mathcal{I}'^c$. Then we have the following:

$$\begin{aligned} X = XZ' + C' \quad &\Rightarrow \quad \mathcal{P}_{\mathcal{I}_0 \cap \mathcal{I}'^c}(X) = X \mathcal{P}_{\mathcal{I}_0 \cap \mathcal{I}'^c}(Z') \\ &\Rightarrow \quad U_X \Sigma_X \mathcal{P}_{\mathcal{I}_0 \cap \mathcal{I}'^c}(V_X^T) = U_X \Sigma_X V_X^T \mathcal{P}_{\mathcal{I}_0 \cap \mathcal{I}'^c}(Z') \\ &\Rightarrow \quad \mathcal{P}_{\mathcal{I}_0 \cap \mathcal{I}'^c}(V_X^T) = V_X^T \mathcal{P}_{\mathcal{I}_0 \cap \mathcal{I}'^c}(Z') \\ &\Rightarrow \quad V_X \mathcal{P}_{\mathcal{I}_0 \cap \mathcal{I}'^c}(V_X^T) = V_X V_X^T \mathcal{P}_{\mathcal{I}_0 \cap \mathcal{I}'^c}(Z') \\ &\Rightarrow \quad V_0 V_0^T V_X \mathcal{P}_{\mathcal{I}_0 \cap \mathcal{I}'^c}(V_X^T) = V_0 V_0^T V_X V_X^T \mathcal{P}_{\mathcal{I}_0 \cap \mathcal{I}'^c}(Z') \end{aligned}$$

Recalling (9) and (10), then we have

$$\begin{aligned} V_0 V_0^T V_X \mathcal{P}_{\mathcal{I}_0 \cap \mathcal{I}'^c}(V_X^T) = 0 &\Rightarrow V_0 V_0^T V_X V_X^T \mathcal{P}_{\mathcal{I}_0 \cap \mathcal{I}'^c}(Z') = 0 \\ &\Rightarrow V_0 V_0^T \mathcal{P}_{\mathcal{I}_0 \cap \mathcal{I}'^c}(Z') = 0 \\ &\Rightarrow \mathcal{P}_{\mathcal{I}_0 \cap \mathcal{I}'^c}(Z') = 0, \end{aligned} \tag{11}$$

where the last equality is from the conclusion of $Z' = V_0 V_0^T Z'$. By $X = X_0 + C_0$,

$$\mathcal{P}_{\mathcal{I}_0 \cap \mathcal{I}'^c}(C_0) = \mathcal{P}_{\mathcal{I}_0 \cap \mathcal{I}'^c}(X - X_0) = \mathcal{P}_{\mathcal{I}_0 \cap \mathcal{I}'^c}(X).$$

Notice that $\mathcal{P}_{\mathcal{I}_0 \cap \mathcal{I}'^c}(X) = X \mathcal{P}_{\mathcal{I}_0 \cap \mathcal{I}'^c}(Z')$. Then by (11), we have

$$\mathcal{P}_{\mathcal{I}_0 \cap \mathcal{I}'^c}(C_0) = 0, \text{ and so } \mathcal{I}_0 \cap \mathcal{I}'^c = \emptyset.$$

C.3 Dual Conditions

To prove that LRR can exactly recover the row space and column support, Theorem 2 suggests us to prove that the pair (Z', C') is a solution to (2), and every solution to (2) also satisfies the two constraints in Theorem 2. To this end, we write down the optimal conditions of (2), resulting in the dual conditions for ensuring the exactness of LRR.

At first, we define two operators that are closely related to the subgradient of $\|C'\|_{2,1}$ and $\|Z'\|_*$.

Definition 4 *Let (Z', C') satisfy $X = XZ' + C'$, $\mathcal{P}_{V_0}^L(Z') = Z'$ and $\mathcal{P}_{\mathcal{I}_0}(C') = C'$. We define the following:*

$$\mathcal{B}(C') \quad \triangleq \quad \{H | \mathcal{P}_{\mathcal{I}_0^c}(H) = 0; \forall i \in \mathcal{I}_0 : [H]_i = \frac{[C']_i}{\|[C']_i\|_2}\}.$$

It is simple to see that $\mathcal{B}(C')$ is a column-normalized version of C'.

Let the SVD of Z' as $U' \Sigma' V'^T$, we further define the operator $\mathcal{P}_{T(Z')}$ as

$$\begin{aligned} \mathcal{P}_{T(Z')}(\cdot) \quad &\triangleq \quad \mathcal{P}_{U'}(\cdot) + \mathcal{P}_{V'}(\cdot) - \mathcal{P}_{U'} \mathcal{P}_{V'}(\cdot) \\ &= \quad \mathcal{P}_{V_0}^L(\cdot) + \mathcal{P}_{V'}(\cdot) - \mathcal{P}_{V_0}^L \mathcal{P}_{V'}(\cdot). \end{aligned}$$

Then, we present and prove the dual conditions for exactly recovering the row space and column support of X_0 and C_0, respectively.

Theorem 3 *Let (Z', C') satisfy $X = XZ' + C'$, $\mathcal{P}_{V_0}^{L}(Z') = Z'$ and $\mathcal{P}_{\mathcal{I}_0}(C') = C'$. Then (Z', C') is an optimal solution to (2) if there exists a matrix Q that satisfies*

$$\begin{array}{ll} (a) & \mathcal{P}_{T(Z')}(X^T Q) = U'V'^T, \\ (b) & \|\mathcal{P}_{T(Z')^{\perp}}(X^T Q)\| < 1, \\ (c) & \mathcal{P}_{\mathcal{I}_0}(Q) = \lambda \mathcal{B}(C'), \\ (d) & \|\mathcal{P}_{\mathcal{I}_0^c}(Q)\|_{2,\infty} < \lambda. \end{array}$$

Further, if $\mathcal{P}_{\mathcal{I}_0} \cap \mathcal{P}_{V'} = \{0\}$, then any optimal solution to (2) will have the exact row space and column support.

Proof By standard convexity arguments (Rockafellar, 1970), a feasible pair (Z', C') is an optimal solution to (2) if there exists Q' such that

$$Q' \in \partial \|Z'\|_* \quad \text{and} \quad Q' \in \lambda X^T \partial \|C'\|_{2,1}.$$

Note that (a) and (b) imply that $X^T Q \in \partial \|Z'\|_*$. Furthermore, letting $\mathcal{I}'$ be the column support of C', then by Theorem 2, we have $\mathcal{I}' = \mathcal{I}_0$. Therefore (c) and (d) imply that $Q \in \lambda \partial \|C'\|_{2,1}$, and so $X^T Q \in \lambda X^T \partial \|C'\|_{2,1}$. Thus, (Z', C') is an optimal solution to (2).

Notice that the LRR problem (2) may have multiple solutions. For any fixed $\Delta \neq 0$, assume that $(Z' + \Delta_1, C' - \Delta)$ is also optimal. Then by $X = X(Z' + \Delta_1) + (C' - \Delta) = XZ' + C'$, we have

$$\Delta = X\Delta_1.$$

By the well-known duality between operator norm and nuclear norm, there exists W_0 that satisfies $\|W_0\| = 1$ and $\langle W_0, \mathcal{P}_{T(Z')^{\perp}}(\Delta_1) \rangle = \|\mathcal{P}_{T(Z')^{\perp}}(\Delta_1)\|_*$. Let $W = \mathcal{P}_{T(Z')^{\perp}}(W_0)$, then we have that $\|W\| \leq 1$, $\langle W, \mathcal{P}_{T(Z')^{\perp}}(\Delta_1) \rangle = \|\mathcal{P}_{T(Z')^{\perp}}(\Delta_1)\|_*$ and $\mathcal{P}_{T(Z')}(W) = 0$. Let F be such that

$$[F]_i = \begin{cases} -\frac{[\Delta]_i}{\|[\Delta]_i\|_2}, & \text{if } i \notin \mathcal{I}_0 \text{ and } [\Delta]_i \neq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Then $\mathcal{P}_{T(Z')}(X^T Q) + W$ is a subgradient of $\|Z'\|_*$, and $\mathcal{P}_{\mathcal{I}_0}(Q)/\lambda + F$ is a subgradient of $\|C'\|_{2,1}$. By the convexity of nuclear norm and $\ell_{2,1}$ norm, we have

$$\begin{aligned} &\|Z' + \Delta_1\|_* + \lambda \|C' - \Delta\|_{2,1} \\ &\geq \|L'\|_* + \lambda \|C'\|_{2,1} + \langle \mathcal{P}_{T(Z')}(X^T Q) + W, \Delta_1 \rangle - \lambda \langle \mathcal{P}_{\mathcal{I}_0}(Q)/\lambda + F, \Delta \rangle \\ &= \|L'\|_* + \lambda \|C'\|_{2,1} + \|\mathcal{P}_{T(Z')^{\perp}}(\Delta_1)\|_* + \lambda \|\mathcal{P}_{\mathcal{I}_0^c}(\Delta)\|_{2,1} + \langle \mathcal{P}_{T(Z')}(X^T Q), \Delta_1 \rangle - \langle \mathcal{P}_{\mathcal{I}_0}(Q), \Delta \rangle. \end{aligned}$$

Notice that

$$\begin{aligned} &\langle \mathcal{P}_{T(Z')}(X^T Q), \Delta_1 \rangle - \langle \mathcal{P}_{\mathcal{I}_0}(Q), \Delta \rangle \\ &= \langle X^T Q - \mathcal{P}_{T(Z')^{\perp}}(X^T Q), \Delta_1 \rangle - \langle Q - \mathcal{P}_{\mathcal{I}_0^c}(Q), \Delta \rangle \\ &= \langle -\mathcal{P}_{T(Z')^{\perp}}(X^T Q), \Delta_1 \rangle + \langle \mathcal{P}_{\mathcal{I}_0^c}(Q), \Delta \rangle + \langle Q, X\Delta_1 - \Delta \rangle \\ &= \langle -\mathcal{P}_{T(Z')^{\perp}}(X^T Q), \Delta_1 \rangle + \langle \mathcal{P}_{\mathcal{I}_0^c}(Q), \Delta \rangle \\ &\geq -\|\mathcal{P}_{T(Z')^{\perp}}(X^T Q)\| \|\mathcal{P}_{T(Z')^{\perp}}(\Delta_1)\|_* - \|\mathcal{P}_{\mathcal{I}_0^c}(Q)\|_{2,\infty} \|\mathcal{P}_{\mathcal{I}_0^c}(\Delta)\|_{2,1}, \end{aligned}$$

where the last inequality is from Lemma 7, and the well-known conclusion that $|\langle MN \rangle| \leq \|M\| \|N\|_*$ holds for any matrices M and N.

The above deductions have proven that

$$\begin{aligned} \|Z' + \Delta_1\|_* + \lambda \|C' - \Delta\|_{2,1} \quad &\geq \quad \|L'\|_* + \lambda \|C'\|_{2,1} + (1 - \|\mathcal{P}_{T(Z')^{\perp}}(X^T Q)\|) \|\mathcal{P}_{T(Z')^{\perp}}(\Delta_1)\|_* \\ &+ \quad (\lambda - \|\mathcal{P}_{\mathcal{I}_0^c}(Q)\|_{2,\infty}) \|\mathcal{P}_{\mathcal{I}_0^c}(\Delta)\|_{2,1}. \end{aligned}$$

However, since both (Z', C') and $(Z' + \Delta_1, C' - \Delta)$ are optimal to (2), we must have

$$\|Z' + \Delta_1\|_* + \lambda\|C' - \Delta\|_{2,1} = \|L'\|_* + \lambda\|C'\|_{2,1},$$

and so

$$(1 - \|\mathcal{P}_{T(Z')^\perp}(X^T Q)\|)\|\mathcal{P}_{T(Z')^\perp}(\Delta_1)\|_* + (\lambda - \|\mathcal{P}_{\mathcal{I}_0^c}(Q)\|_{2,\infty})\|\mathcal{P}_{\mathcal{I}_0^c}(\Delta)\|_{2,1} \leq 0.$$

Recalling the conditions (b) and (d), then we have

$$\|\mathcal{P}_{T(Z')^\perp}(\Delta_1)\|_* = \|\mathcal{P}_{\mathcal{I}_0^c}(\Delta)\|_{2,1} = 0,$$

i.e., $\mathcal{P}_{T(Z')}(\Delta_1) = \Delta_1$ and $\mathcal{P}_{\mathcal{I}_0}(\Delta) = \Delta$. By Lemma 1,

$$Z' \in \mathcal{P}^L_{V_X}, Z' + \Delta_1 \in \mathcal{P}^L_{V_X} \quad \text{and so} \quad \Delta_1 \in \mathcal{P}^L_{V_X}.$$

Also, notice that $\Delta = X\Delta_1$. Thus, we have

$$\begin{aligned}
\mathcal{P}_{\mathcal{I}_0^c}(\Delta) = 0 \quad &\Rightarrow \quad X\mathcal{P}_{\mathcal{I}_0^c}(\Delta_1) = 0 \\
&\Rightarrow \quad V_X^T \mathcal{P}_{\mathcal{I}_0^c}(\Delta_1) = 0 \\
&\Rightarrow \quad \mathcal{P}^L_{V_X}\mathcal{P}_{\mathcal{I}_0^c}(\Delta_1) = 0 \\
&\Rightarrow \quad \mathcal{P}_{\mathcal{I}_0^c}(\mathcal{P}^L_{V_X}(\Delta_1)) = 0 \\
&\Rightarrow \quad \mathcal{P}_{\mathcal{I}_0^c}(\Delta_1) = 0,
\end{aligned}$$

which implies that $\mathcal{P}_{\mathcal{I}_0}(\Delta_1) = \Delta_1$. Furthermore, we have

$$\begin{aligned}
\mathcal{P}_{\mathcal{I}_0}(\Delta_1) &= \Delta_1 = \mathcal{P}_{T(Z')}(\Delta_1) = \mathcal{P}_{U'}(\Delta_1) + \mathcal{P}_{V'}\mathcal{P}_{U'^\perp}(\Delta_1) \\
&= \mathcal{P}_{U'}(\mathcal{P}_{\mathcal{I}_0}(\Delta_1)) + \mathcal{P}_{V'}\mathcal{P}_{U'^\perp}(\Delta_1) \\
&= \mathcal{P}_{\mathcal{I}_0}\mathcal{P}_{U'}(\Delta_1) + \mathcal{P}_{V'}\mathcal{P}_{U'^\perp}(\Delta_1) \\
&\Rightarrow \mathcal{P}_{\mathcal{I}_0}\mathcal{P}_{U'^\perp}(\Delta_1) = \mathcal{P}_{V'}\mathcal{P}_{U'^\perp}(\Delta_1).
\end{aligned}$$

Since $\mathcal{P}_{\mathcal{I}_0}\mathcal{P}_{U'^\perp}(\Delta_1) = \mathcal{P}_{U'^\perp}(\Delta_1)$, the above result implies that

$$\mathcal{P}_{U'^\perp}(\Delta_1) \in \mathcal{P}_{\mathcal{I}_0} \cap \mathcal{P}_{V'}.$$

By the assumption of $\mathcal{P}_{\mathcal{I}_0} \cap \mathcal{P}_{V'} = \{0\}$, we have $\mathcal{P}_{U'^\perp}(\Delta_1) = 0$. Recalling Theorem 2, we have that $\mathcal{P}_{U'} = \mathcal{P}^L_{V_0}$, and so $\Delta_1 \in \mathcal{P}^L_{V_0}$. Thus, the solution $(Z'+\Delta_1, C'-\Delta)$ also satisfies $X = X(Z'+\Delta_1)+(C'-\Delta)$, $\mathcal{P}^L_{V_0}(Z'+\Delta_1) = Z'+\Delta_1$ and $\mathcal{P}_{\mathcal{I}_0}(C' - \Delta) = C' - \Delta$. Recalling Theorem 2 again, it can be concluded that the solution $(Z' + \Delta_1, C' - \Delta)$ also exactly recovers the row space and column support, i.e., all possible solutions to (2) equally produce the exact recovery.

C.4 Obtaining Dual Certificates

In this section, we complete the proof of Theorem 1 by constructing a matrix Q that satisfies the conditions in Theorem 3, and proving $\mathcal{P}_{\mathcal{I}_0} \cap \mathcal{P}_{V'} = \{0\}$ as well. This is done by considering an alternate optimization problem, often called the "oracle problem". The oracle problem arises by imposing the equivalent conditions as additional constraints in (2):

$$\textbf{Oracle Problem:} \quad \min_{Z,C} \|Z\|_* + \lambda\|C\|_{2,1} \quad X = XZ + C, \mathcal{P}^L_{V_0}(Z) = Z, \mathcal{P}_{\mathcal{I}_0}(C) = C. \tag{12}$$

Note that the above problem is always feasible, as $(V_0V_0^T, C_0)$ is feasible. Thus, an optimal solution, denoted as $(\hat{Z}, \hat{C})$, exists. With this perspective, we would like to show that $(\hat{Z}, \hat{C})$ is an optimal solution to (2), and obtain the dual certificates by the optimal conditions of (12).

Definition 5 *Let $(\hat{Z}, \hat{C})$ be an optimal solution to the oracle problem (12). Let $\hat{U}\hat{\Sigma}\hat{V}^T$ and $\hat{\mathcal{I}}$ be the SVD and column support of $\hat{Z}$ and $\hat{C}$, respectively. By Theorem 2,*

$$\hat{U}\hat{U}^T = V_0 V_0^T \quad \text{and} \quad \hat{\mathcal{I}} = \mathcal{I}_0.$$

Let

$$\bar{V} \triangleq \hat{V}\hat{U}^T V_0, \quad \text{then we have} \quad \hat{U}\hat{V}^T = V_0 \bar{V}^T.$$

Since $\hat{U}\hat{U}^T = V_0 V_0^T$ and $\bar{V}\bar{V}^T = \hat{V}\hat{V}^T$, we have

$$\begin{aligned} \mathcal{P}_{\hat{T}}(\cdot) &\triangleq \mathcal{P}_{\hat{U}}(\cdot) + \mathcal{P}_{\hat{V}}(\cdot) - \mathcal{P}_{\hat{U}}\mathcal{P}_{\hat{V}}(\cdot) \\ &= \mathcal{P}^L_{V_0}(\cdot) + \mathcal{P}_{\bar{V}}(\cdot) - \mathcal{P}^L_{V_0}\mathcal{P}_{\bar{V}}(\cdot). \end{aligned}$$

Lemma 8 *Let $\hat{H} = \mathcal{B}(\hat{C})$, then we have*

$$V_0 \mathcal{P}_{\mathcal{I}_0}(\bar{V}^T) = \lambda \mathcal{P}^L_{V_0}(X^T \hat{H}).$$

Proof Notice that the Lagrange dual function of the oracle problem (12) is

$$\begin{aligned} &\mathcal{L}(Z, C, Y, Y_1, Y_2) = \|Z\|_* + \lambda \|C\|_{2,1} + \langle Y, X - XZ - C\rangle \\ &+ \langle Y_1, \mathcal{P}^L_{V_0}(Z) - Z\rangle + \langle Y_2, \mathcal{P}_{\mathcal{I}_0}(C) - C\rangle, \end{aligned}$$

where Y, Y_1 and Y_2 are Lagrange multipliers. Since $(\hat{Z}, \hat{C})$ is a solution to problem (12), we have

$$0 \in \partial \mathcal{L}_Z(\hat{Z}, \hat{C}, Y, Y_1, Y_2) \quad \text{and} \quad 0 \in \partial \mathcal{L}_C(\hat{Z}, \hat{C}, Y, Y_1, Y_2).$$

Hence, there exists $\hat{W}$, $\hat{H}$ and $\hat{L}$ such that

$$\begin{aligned} &\mathcal{P}_{\hat{T}}(\hat{W}) = 0, \|\hat{W}\| \leq 1, V_0\bar{V}^T + \hat{W} \in \partial \|\hat{Z}\|_*, \\ &\hat{H} = \mathcal{B}(\hat{C}), \mathcal{P}_{\mathcal{I}_0}(\hat{L}) = 0, \|\hat{L}\|_{2,\infty} \leq 1, \hat{H} + \hat{L} \in \partial \|\hat{C}\|_{2,1}, \\ &V_0\bar{V}^T + \hat{W} - X^T Y - \mathcal{P}^L_{V_0^\perp}(Y_1) = 0, \\ &\lambda(\hat{H} + \hat{L}) - Y - \mathcal{P}_{\mathcal{I}_0^c}(Y_2) = 0. \end{aligned}$$

Let $A = \hat{W} - Y_1$ and $B = \lambda\hat{L} - Y_2$, then the last two equations above imply that

$$V_0\bar{V}^T + \mathcal{P}^L_{V_0^\perp}(A) = \lambda X^T \hat{H} + \mathcal{P}_{\mathcal{I}_0^c}(X^T B). \tag{13}$$

Furthermore, we have

$$\begin{aligned} \mathcal{P}^L_{V_0}\mathcal{P}_{\mathcal{I}_0}(V_0\bar{V}^T + \mathcal{P}^L_{V_0^\perp}(A)) &= \mathcal{P}^L_{V_0}\mathcal{P}_{\mathcal{I}_0}(V_0\bar{V}^T) + \mathcal{P}^L_{V_0}\mathcal{P}_{\mathcal{I}_0}\mathcal{P}^L_{V_0^\perp}(A) \\ &= V_0\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T) + \mathcal{P}^L_{V_0}\mathcal{P}^L_{V_0^\perp}\mathcal{P}_{\mathcal{I}_0}(A) \\ &= V_0\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T). \end{aligned} \tag{14}$$

Similarly, we have

$$\begin{aligned} \mathcal{P}^L_{V_0}\mathcal{P}_{\mathcal{I}_0}(\lambda X^T\hat{H} + \mathcal{P}_{\mathcal{I}_0^c}(X^T B)) &= \mathcal{P}^L_{V_0}\mathcal{P}_{\mathcal{I}_0}(\lambda X^T\hat{H}) + \mathcal{P}^L_{V_0}\mathcal{P}_{\mathcal{I}_0}\mathcal{P}_{\mathcal{I}_0^c}(X^T B) \\ &= \mathcal{P}^L_{V_0}\mathcal{P}_{\mathcal{I}_0}(\lambda X^T\hat{H}) = \lambda\mathcal{P}^L_{V_0}(X^T\mathcal{P}_{\mathcal{I}_0}(\hat{H})) \\ &= \lambda\mathcal{P}^L_{V_0}(X^T\hat{H}). \end{aligned} \tag{15}$$

Combing (13), (14) and (15) together, we have

$$V_0 \mathcal{P}_{\mathcal{I}_0}(\bar{V}^T) = \lambda \mathcal{P}^L_{V_0}(X^T \hat{H}).$$

Before constructing a matrix Q that satisfies the conditions in Theorem 3, we shall prove that $\mathcal{P}_{\mathcal{I}_0} \cap \mathcal{P}_{\hat{V}} = \{0\}$ can be satisfied by choosing appropriate parameter λ.

Definition 6 *Recalling the definition of $\bar{V}$, define matrix G as*

$$G \triangleq \mathcal{P}_{\mathcal{I}_0}(\bar{V}^T)(\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T))^T.$$

Then we have

$$G = \sum_{i\in\mathcal{I}_0}[\bar{V}^T]_i([\bar{V}^T]_i)^T \preccurlyeq \sum_i [\bar{V}^T]_i([\bar{V}^T]_i)^T = \bar{V}^T\bar{V} = \mathtt{I},$$

where $\preccurlyeq$ is the generalized inequality induced by the positive semi-definite cone. Hence, $\|G\| \leq 1$.

The following lemma states that $\|G\|$ can be far away from 1 by choosing appropriate λ.

Lemma 9 *Let $\psi = \|G\|$, then $\psi \leq \lambda^2\|X\|^2\gamma n$.*

Proof Notice that

$$\begin{aligned}\psi &= \|\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T)(\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T))^T\| = \|V_0\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T)(\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T))^T V_0^T\|\\ &= \|(V_0\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T))(V_0\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T))^T\|.\end{aligned}$$

By Lemma 8, we have

$$\begin{aligned}\psi &= \|\lambda\mathcal{P}^L_{V_0}(X^T\hat{H})(\lambda\mathcal{P}^L_{V_0}(X^T\hat{H}))^T\|\\ &= \lambda^2\|\mathcal{P}^L_{V_0}(X^T\hat{H})(\mathcal{P}^L_{V_0}(X^T\hat{H}))^T\|\\ &\leq \lambda^2\|\mathcal{P}^L_{V_0}(X^T\hat{H})\|\|(\mathcal{P}^L_{V_0}(X^T\hat{H}))^T\|\\ &\leq \lambda^2\|X^T\hat{H}\|^2 \leq \lambda^2\|X\|^2\|\hat{H}\|^2\\ &\leq \lambda^2\|X\|^2|\mathcal{I}_0| = \lambda^2\|X\|^2\gamma n,\end{aligned}$$

where $\|H\|^2 \leq |\mathcal{I}_0| = \gamma n$ is due to Lemma 5.

The above lemma bounds ψ far way from 1. In particular, for $\lambda \leq \frac{3}{7\|X\|\sqrt{\gamma n}}$, we have $\psi \leq \frac{1}{4}$. So we can assume that $\psi < 1$ in sequel.

Lemma 10 *If $\psi < 1$, then $\mathcal{P}_{\hat{V}} \cap \mathcal{P}_{\mathcal{I}_0} = \mathcal{P}_{\bar{V}} \cap \mathcal{P}_{\mathcal{I}_0} = \{0\}$.*

Proof Let $M \in \mathcal{P}_{\bar{V}} \cap \mathcal{P}_{\mathcal{I}_0}$, then we have

$$\begin{aligned}\|M\|^2 &= \|MM^T\| = \|\mathcal{P}_{\mathcal{I}_0}(M)(\mathcal{P}_{\mathcal{I}_0}(M))^T\| = \|\mathcal{P}_{\mathcal{I}_0}(M\bar{V}\bar{V}^T)(\mathcal{P}_{\mathcal{I}_0}(M\bar{V}\bar{V}^T))^T\|\\ &= \|M\bar{V}\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T)(\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T))^T\bar{V}^T M^T\|\\ &\leq \|M\|^2\|\bar{V}\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T)(\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T))^T\bar{V}^T\| = \|M\|^2\|\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T)(\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T))^T\| = \|M\|^2\psi\\ &\leq \|M\|^2.\end{aligned}$$

Since $\psi < 1$, the last equality can hold only if $\|M\| = 0$, and hence $M = 0$. Also, note that $\mathcal{P}_{\hat{V}} = \mathcal{P}_{\bar{V}}$, which completes the proof.

The following lemma plays a key role in constructing Q that satisfies the conditions in Theorem 3.

Lemma 11 *If $\psi < 1$, then the operator $\mathcal{P}_{\bar{V}}\mathcal{P}_{\mathcal{I}_0^c}\mathcal{P}_{\bar{V}}$ is an injection from $\mathcal{P}_{\bar{V}}$ to $\mathcal{P}_{\bar{V}}$, and its inverse operator is $\mathtt{I} + \sum_{i=1}^{\infty}(\mathcal{P}_{\bar{V}}\mathcal{P}_{\mathcal{I}_0}\mathcal{P}_{\bar{V}})^i$.*

Proof For any matrix M such that $\|M\| = 1$, we have

$$\begin{aligned}\mathcal{P}_{\bar{V}}\mathcal{P}_{\mathcal{I}_0}\mathcal{P}_{\bar{V}}(M) &= \mathcal{P}_{\bar{V}}\mathcal{P}_{\mathcal{I}_0}(M\bar{V}\bar{V}^T)\\ &= \mathcal{P}_{\bar{V}}(M\bar{V}\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T))\\ &= M\bar{V}\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T)\bar{V}\bar{V}^T\\ &= M\bar{V}(\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T)\bar{V})\bar{V}^T\\ &= M\bar{V}(\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T)(\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T))^T)\bar{V}^T\\ &= M\bar{V}G\bar{V}^T,\end{aligned}$$

which leads to $\|\mathcal{P}_{\bar{V}}\mathcal{P}_{\mathcal{I}_0}\mathcal{P}_{\bar{V}}\| \leq \|G\| = \psi$. Since $\psi < 1$, $\mathtt{I} + \sum_{i=1}^{\infty}(\mathcal{P}_{\bar{V}}\mathcal{P}_{\mathcal{I}_0}\mathcal{P}_{\bar{V}})^i$ is well defined, and has a spectral norm not larger than $1/(1-\psi)$.

Note that

$$\mathcal{P}_{\bar{V}}\mathcal{P}_{\mathcal{I}_0^c}\mathcal{P}_{\bar{V}} = \mathcal{P}_{\bar{V}}(I - \mathcal{P}_{\mathcal{I}_0})\mathcal{P}_{\bar{V}} = \mathcal{P}_{\bar{V}}(\mathtt{I} - \mathcal{P}_{\bar{V}}\mathcal{P}_{\mathcal{I}_0}\mathcal{P}_{\bar{V}}),$$

thus for any $M \in \mathcal{P}_{\bar{V}}$ the following holds

$$\begin{aligned}\mathcal{P}_{\bar{V}}\mathcal{P}_{\mathcal{I}_0^c}\mathcal{P}_{\bar{V}}(\mathtt{I} + \sum_{i=1}^{\infty}(\mathcal{P}_{\bar{V}}\mathcal{P}_{\mathcal{I}_0}\mathcal{P}_{\bar{V}})^i)(M) &= \mathcal{P}_{\bar{V}}(\mathtt{I} - \mathcal{P}_{\bar{V}}\mathcal{P}_{\mathcal{I}_0}\mathcal{P}_{\bar{V}})(\mathtt{I} + \sum_{i=1}^{\infty}(\mathcal{P}_{\bar{V}}\mathcal{P}_{\mathcal{I}_0}\mathcal{P}_{\bar{V}})^i)(M) \\ &= \mathcal{P}_{\bar{V}}(M) = M.\end{aligned}$$

Lemma 12 *We have*

$$\|\mathcal{P}_{\mathcal{I}_0^c}(\bar{V}^T)\|_{2,\infty} \leq \sqrt{\frac{\mu r_0}{(1-\gamma)n}}.$$

Proof Notice that $X = X\hat{Z} + \hat{C}$ and $\mathcal{P}_{\mathcal{I}_0^c}(X) = X_0 = \mathcal{P}_{\mathcal{I}_0^c}(X_0)$. Then we have

$$\begin{aligned}X = X\hat{Z} + \hat{C} &\Rightarrow \mathcal{P}_{\mathcal{I}_0^c}(X_0) = X\mathcal{P}_{\mathcal{I}_0^c}(\hat{Z}) \\ &\Rightarrow V_0^T = \mathcal{P}_{\mathcal{I}_0^c}(V_0^T) = \Sigma_0^{-1}U_0^T X\hat{U}\hat{\Sigma}\mathcal{P}_{\mathcal{I}_0^c}(\hat{V}^T),\end{aligned}$$

which implies that the rows of $\mathcal{P}_{\mathcal{I}_0^c}(\hat{V}^T)$ span the rows of V_0^T. However, the rank of $\mathcal{P}_{\mathcal{I}_0^c}(\hat{V}^T)$ is at most r_0 (this is because the rank of both $\hat{U}$ and $\hat{V}$ is r_0). Thus, it can be concluded that $\mathcal{P}_{\mathcal{I}_0^c}(\hat{V}^T)$ is of full row rank. At the same time, we have

$$0 \preccurlyeq \mathcal{P}_{\mathcal{I}_0^c}(\hat{V}^T)(\mathcal{P}_{\mathcal{I}_0^c}(\hat{V}^T))^T \preccurlyeq \mathtt{I}.$$

So, there exists a symmetric, invertible matrix $Y \in \mathbb{R}^{r_0 \times r_0}$ such that

$$\|Y\| \leq 1 \quad \text{and} \quad Y^2 = \mathcal{P}_{\mathcal{I}_0^c}(\hat{V}^T)(\mathcal{P}_{\mathcal{I}_0^c}(\hat{V}^T))^T.$$

This in turn implies that $Y^{-1}\mathcal{P}_{\mathcal{I}_0^c}(\hat{V}^T)$ has orthonomal rows. Since $\mathcal{P}_{\mathcal{I}_0^c}(V_0^T) = V_0^T$ is also row orthonomal, it can be concluded that there exists a row orthonomal matrix R such that

$$Y^{-1}\mathcal{P}_{\mathcal{I}_0^c}(\hat{V}^T) = R\mathcal{P}_{\mathcal{I}_0^c}(V_0^T).$$

Then we have

$$\begin{aligned}\|\mathcal{P}_{\mathcal{I}_0^c}(\hat{V}^T)\|_{2,\infty} &= \|YR\mathcal{P}_{\mathcal{I}_0^c}(V_0^T)\|_{2,\infty} \\ &\leq \|Y\|\|R\mathcal{P}_{\mathcal{I}_0^c}(V_0^T)\|_{2,\infty} \leq \|R\mathcal{P}_{\mathcal{I}_0^c}(V_0^T)\|_{2,\infty} \\ &\leq \|\mathcal{P}_{\mathcal{I}_0^c}(V_0^T)\|_{2,\infty} \\ &\leq \sqrt{\frac{\mu r_0}{(1-\gamma)n}},\end{aligned}$$

where the last inequality is from the definition of μ.

By the definition of $\bar{V}$, we further have

$$\begin{aligned}\|\mathcal{P}_{\mathcal{I}_0^c}(\bar{V}^T)\|_{2,\infty} &= \|\mathcal{P}_{\mathcal{I}_0^c}(V_0^T\hat{U}\hat{V}^T)\|_{2,\infty} = \|V_0^T\hat{U}\mathcal{P}_{\mathcal{I}_0^c}(\hat{V}^T)\|_{2,\infty} \leq \|\mathcal{P}_{\mathcal{I}_0^c}(\hat{V}^T)\|_{2,\infty} \\ &\leq \sqrt{\frac{\mu r_0}{(1-\gamma)n}}.\end{aligned}$$

Now we define Q_1 and Q_2 used to construct the matrix Q that satisfies the conditions in Theorem 3.

Definition 7 *Define Q_1 and Q_2 as follows:*

$$\begin{aligned}Q_1 &\triangleq \lambda\mathcal{P}_{V_0}^L(X^T\hat{H}) = V_0\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T), \\ Q_2 &\triangleq \lambda\mathcal{P}_{V_0^\perp}^L\mathcal{P}_{\mathcal{I}_0^c}\mathcal{P}_{\bar{V}}(\mathtt{I} + \sum_{i=1}^{\infty}(\mathcal{P}_{\bar{V}}\mathcal{P}_{\mathcal{I}_0}\mathcal{P}_{\bar{V}})^i)\mathcal{P}_{\bar{V}}(X^T\hat{H}) \\ &= \lambda\mathcal{P}_{\mathcal{I}_0^c}\mathcal{P}_{\bar{V}}(\mathtt{I} + \sum_{i=1}^{\infty}(\mathcal{P}_{\bar{V}}\mathcal{P}_{\mathcal{I}_0}\mathcal{P}_{\bar{V}})^i)\mathcal{P}_{\bar{V}}\mathcal{P}_{V_0^\perp}^L(X^T\hat{H}),\end{aligned}$$

where the equalities are due to Lemma 8 and Lemma 4.

The following Theorem almost finishes the proof of Theorem 1.

Theorem 4 *Let the SVD of the dictionary matrix X as $U_X \Sigma_X V_X^T$. Assume $\psi < 1$. Let*

$$Q \triangleq U_X \Sigma_X^{-1} V_X^T (V_0 \bar{V}^T + \lambda X^T \hat{H} - Q_1 - Q_2).$$

If

$$\frac{\gamma}{1-\gamma} < \frac{\beta^2 (1-\psi)^2}{(3-\psi+\beta)^2 \mu r_0},$$

and

$$\frac{(1-\psi)\sqrt{\frac{\mu r_0}{1-\gamma}}}{\|X\|\sqrt{n}(\beta(1-\psi) - (1+\beta)\sqrt{\frac{\gamma}{1-\gamma}\mu r_0})} < \lambda < \frac{1-\psi}{\|X\|\sqrt{\gamma n}(2-\psi)},$$

then Q satisfies the conditions in Theorem 3, i.e., it is the dual certificate.

Proof By Lemma 10, it is concluded that $\psi < 1$ can ensure that $\mathcal{P}_{\hat{V}} \cap \mathcal{P}_{\mathcal{I}_0} = \{0\}$. Hence it is sufficient to show that Q simultaneously satisfies

$$\begin{array}{ll}
(\mathbf{S1}) & \mathcal{P}_{\hat{U}}(X^T Q) = \hat{U}\hat{V}^T, \\
(\mathbf{S2}) & \mathcal{P}_{\hat{V}}(X^T Q) = \hat{U}\hat{V}^T, \\
(\mathbf{S3}) & \mathcal{P}_{\mathcal{I}_0}(Q) = \lambda \hat{H}, \\
(\mathbf{S4}) & \|\mathcal{P}_{\hat{T}^\perp}(X^T Q)\| < 1, \\
(\mathbf{S5}) & \|\mathcal{P}_{\mathcal{I}_0^c}(Q)\|_{2,\infty} < \lambda.
\end{array}$$

We prove that each of these five conditions holds, in **S1-S5**. Then in **S6**, we show that the condition on λ is not vacuous, i.e., the lower bound is strictly less than the upper bound.

First of all, we shall simplify the formula of $X^T Q$ that will be used several times in the following process. Recalling the setting (3) that assumes $\mathcal{P}_{V_X}^L(V_0) = V_0$, we have that $\mathcal{P}_{V_X}^L(Q_1) = Q_1$ and

$$\begin{aligned}
\mathcal{P}_{V_X}^L(Q_2) &= \lambda \mathcal{P}_{\mathcal{I}_0^c} \mathcal{P}_{\bar{V}} (\mathtt{I} + \sum_{i=1}^{\infty} (\mathcal{P}_{\bar{V}} \mathcal{P}_{\mathcal{I}_0} \mathcal{P}_{\bar{V}})^i) \mathcal{P}_{\bar{V}} \mathcal{P}_{V_X}^L \mathcal{P}_{V_0^\perp}^L (X^T \hat{H}) \\
&= \lambda \mathcal{P}_{\mathcal{I}_0^c} \mathcal{P}_{\bar{V}} (\mathtt{I} + \sum_{i=1}^{\infty} (\mathcal{P}_{\bar{V}} \mathcal{P}_{\mathcal{I}_0} \mathcal{P}_{\bar{V}})^i) \mathcal{P}_{\bar{V}} \mathcal{P}_{V_X}^L (\mathtt{I} - V_0 V_0^T) X^T \hat{H}) \\
&= \lambda \mathcal{P}_{\mathcal{I}_0^c} \mathcal{P}_{\bar{V}} (\mathtt{I} + \sum_{i=1}^{\infty} (\mathcal{P}_{\bar{V}} \mathcal{P}_{\mathcal{I}_0} \mathcal{P}_{\bar{V}})^i) \mathcal{P}_{\bar{V}} (\mathtt{I} - V_0 V_0^T) X^T \hat{H}) \\
&= \lambda \mathcal{P}_{\mathcal{I}_0^c} \mathcal{P}_{\bar{V}} (\mathtt{I} + \sum_{i=1}^{\infty} (\mathcal{P}_{\bar{V}} \mathcal{P}_{\mathcal{I}_0} \mathcal{P}_{\bar{V}})^i) \mathcal{P}_{\bar{V}} \mathcal{P}_{V_0^\perp}^L (X^T \hat{H}) \\
&= Q_2.
\end{aligned}$$

Further, we have

$$\begin{aligned}
X^T Q &= V_X V_X^T (V_0 \bar{V}^T + \lambda X^T \hat{H} - Q_1 - Q_2) = \mathcal{P}_{V_X}^L (V_0 \bar{V}^T + \lambda X^T \hat{H} - Q_1 - Q_2) \\
&= \mathcal{P}_{V_X}^L (V_0 \bar{V}^T) + \lambda \mathcal{P}_{V_X}^L (X^T \hat{H}) - \mathcal{P}_{V_X}^L (Q_1) - \mathcal{P}_{V_X}^L (Q_2) \\
&= V_0 \bar{V}^T + \lambda X^T \hat{H} - \mathcal{P}_{V_X}^L (Q_1) - \mathcal{P}_{V_X}^L (Q_2) \\
&= V_0 \bar{V}^T + \lambda X^T \hat{H} - Q_1 - \mathcal{P}_{V_X}^L (Q_2) \\
&= V_0 \bar{V}^T + \lambda X^T \hat{H} - Q_1 - Q_2.
\end{aligned}$$

S1: Note that $\mathcal{P}^L_{V_0}(Q_1) = \lambda\mathcal{P}^L_{V_0}(X^T\hat{H})$ and $\mathcal{P}^L_{V_0}(Q_2) = 0$. Thus we have

$$\begin{aligned}
\mathcal{P}_{\hat{U}}(X^T Q) &= \mathcal{P}_{\hat{U}}(V_0\bar{V}^T + \lambda X^T\hat{H} - Q_1 - Q_2) \\
&= \mathcal{P}^L_{V_0}(V_0\bar{V}^T + \lambda X^T\hat{H} - Q_1 - Q_2) \\
&= V_0\bar{V}^T + \lambda\mathcal{P}^L_{V_0}(X^T\hat{H}) - \mathcal{P}^L_{V_0}(Q_1) - \mathcal{P}^L_{V_0}(Q_2) \\
&= V_0\bar{V}^T - \mathcal{P}^L_{V_0}(Q_2) \\
&= V_0\bar{V}^T = \hat{U}\hat{V}^T.
\end{aligned}$$

S2: First note that

$$\begin{aligned}
\mathcal{P}_{\bar{V}}(Q_2) &= \lambda\mathcal{P}_{\bar{V}}\mathcal{P}_{\mathcal{I}_0^c}\mathcal{P}_{\bar{V}}(\mathtt{I} + \sum_{i=1}^{\infty}(\mathcal{P}_{\bar{V}}\mathcal{P}_{\mathcal{I}_0}\mathcal{P}_{\bar{V}})^i)\mathcal{P}_{\bar{V}}\mathcal{P}^L_{V_0^\perp}(X^T\hat{H}) \\
&= \lambda\mathcal{P}_{\bar{V}}\mathcal{P}^L_{V_0^\perp}(X^T\hat{H}),
\end{aligned}$$

which is from that the operator $\mathcal{P}_{\bar{V}}\mathcal{P}_{\mathcal{I}_0^c}\mathcal{P}_{\bar{V}}$ is an injection from $\mathcal{P}_{\bar{V}}$ to $\mathcal{P}_{\bar{V}}$, and its inverse is given by $\mathtt{I} + \sum_{i=1}^{\infty}(\mathcal{P}_{\bar{V}}\mathcal{P}_{\mathcal{I}_0}\mathcal{P}_{\bar{V}})^i$.

Thus we have

$$\begin{aligned}
\mathcal{P}_{\hat{V}}(X^T Q) &= \mathcal{P}_{\hat{V}}(V_0\bar{V}^T + \lambda X^T\hat{H} - Q_1 - Q_2) \\
&= \mathcal{P}_{\bar{V}}(V_0\bar{V}^T + \lambda X^T\hat{H} - Q_1 - Q_2) \\
&= V_0\bar{V}^T + \lambda\mathcal{P}_{\bar{V}}(X^T\hat{H}) - \lambda\mathcal{P}_{\bar{V}}\mathcal{P}^L_{V_0}(X^T\hat{H}) - \mathcal{P}_{\bar{V}}(Q_2) \\
&= V_0\bar{V}^T + \lambda\mathcal{P}_{\bar{V}}\mathcal{P}^L_{V_0^\perp}(X^T\hat{H}) - \mathcal{P}_{\bar{V}}(Q_2) \\
&= V_0\bar{V}^T = \hat{U}\hat{V}^T.
\end{aligned}$$

S3: We have

$$\begin{aligned}
\mathcal{P}_{\mathcal{I}_0}(Q) &= \mathcal{P}_{\mathcal{I}_0}(U_X\Sigma_X^{-1}V_X^T(V_0\bar{V}^T + \lambda X^T\hat{H} - Q_1 - Q_2)) \\
&= U_X\Sigma_X^{-1}V_X^T V_0\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T) + \lambda U_X U_X^T\mathcal{P}_{\mathcal{I}_0}(\hat{H}) - U_X\Sigma_X^{-1}V_X^T\mathcal{P}_{\mathcal{I}_0}(Q_1) \\
&= U_X\Sigma_X^{-1}V_X^T V_0\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T) + \lambda U_X U_X^T\hat{H} - U_X\Sigma_X^{-1}V_X^T\mathcal{P}_{\mathcal{I}_0}(Q_1) \\
&= U_X\Sigma_X^{-1}V_X^T V_0\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T) + \lambda U_X U_X^T\hat{H} - U_X\Sigma_X^{-1}V_X^T V_0\mathcal{P}_{\mathcal{I}_0}(\bar{V}^T) \\
&= \lambda U_X U_X^T\hat{H} = \lambda\mathcal{P}^L_{V_X}(\hat{H}).
\end{aligned}$$

By $\hat{C} = X(\mathtt{I} - \hat{Z})$, we have that $\hat{C} \in \mathcal{P}_{U_X}$ and so

$$\hat{H} = \mathcal{B}(\hat{C}) \in \mathcal{P}_{U_X},$$

which finishes the proof of $\mathcal{P}_{\mathcal{I}_0}(Q) = \lambda\hat{H}$.

S4: Since $\mathcal{P}_{\hat{T}^\perp}(V_0\bar{V}^T) = \mathcal{P}_{\hat{T}^\perp}(Q_1) = 0$, we have

$$\begin{aligned}
\mathcal{P}_{\hat{T}^\perp}(X^T Q) &= \mathcal{P}_{\hat{T}^\perp}(V_0\bar{V}^T + \lambda X^T\hat{H} - Q_1 - Q_2) \\
&= \lambda\mathcal{P}_{\bar{V}^\perp}\mathcal{P}^L_{V_0^\perp}(X^T\hat{H}) - \lambda\mathcal{P}^L_{V_0^\perp}\mathcal{P}_{\bar{V}^\perp}\mathcal{P}_{\mathcal{I}_0^c}\mathcal{P}_{\bar{V}}(\mathtt{I} + \sum_{i=1}^{\infty}(\mathcal{P}_{\bar{V}}\mathcal{P}_{\mathcal{I}_0}\mathcal{P}_{\bar{V}})^i)\mathcal{P}_{\bar{V}}(X^T\hat{H}).
\end{aligned}$$

First, it can be calculated that

$$\|\mathcal{P}_{\bar{V}^\perp}\mathcal{P}^L_{V_0^\perp}(X^T\hat{H})\| \leq \|X^T\hat{H}\| \leq \|X\|\|\hat{H}\| \leq \|X\|\sqrt{\gamma n},$$

where $\|\hat{H}\| \leq \sqrt{\gamma n}$ is due to Lemma 5.

Second, we have the following

$$\begin{aligned}
&\|\mathcal{P}^{L}_{V_0^{\perp}}\mathcal{P}_{\bar{V}^{\perp}}\mathcal{P}_{\mathcal{I}_0^c}\mathcal{P}_{\bar{V}}(\mathtt{I}+\sum_{i=1}^{\infty}(\mathcal{P}_{\bar{V}}\mathcal{P}_{\mathcal{I}_0}\mathcal{P}_{\bar{V}})^i)\mathcal{P}_{\bar{V}}(X^T\hat{H})\| \\
&\leq \|\mathcal{P}_{\mathcal{I}_0^c}\mathcal{P}_{\bar{V}}(\mathtt{I}+\sum_{i=1}^{\infty}(\mathcal{P}_{\bar{V}}\mathcal{P}_{\mathcal{I}_0}\mathcal{P}_{\bar{V}})^i)\mathcal{P}_{\bar{V}}(X^T\hat{H})\| \\
&\leq \|(\mathtt{I}+\sum_{i=1}^{\infty}(\mathcal{P}_{\bar{V}}\mathcal{P}_{\mathcal{I}_0}\mathcal{P}_{\bar{V}})^i)\mathcal{P}_{\bar{V}}(X^T\hat{H})\| \\
&\leq \frac{1}{1-\psi}\|\mathcal{P}_{\bar{V}}(X^T\hat{H})\| \\
&\leq \frac{\|X\|\sqrt{\gamma n}}{1-\psi}.
\end{aligned}$$

Thus we have that

$$\begin{aligned}
\|\mathcal{P}_{\hat{T}^{\perp}}(X^TQ)\| < 1 \quad &\Leftarrow \quad \lambda(\|X\|\sqrt{\gamma n}+\frac{\|X\|\sqrt{\gamma n}}{1-\psi}) < 1 \\
&\Leftarrow \quad \lambda < \frac{1-\psi}{\|X\|\sqrt{\gamma n}(2-\psi)}.
\end{aligned}$$

S5: Note that $\mathcal{P}_{\mathcal{I}_0^c}(X^T\hat{H}) = \mathcal{P}_{\mathcal{I}_0^c}(Q_1) = 0$. So we only need to bound the rest two parts.

By Lemma 7, we have

$$\begin{aligned}
\|\mathcal{P}_{\mathcal{I}_0^c}(U_X\Sigma_X^{-1}V_X^TV_0\bar{V}^T)\|_{2,\infty} &= \|U_X\Sigma_X^{-1}V_X^TV_0\mathcal{P}_{\mathcal{I}_0^c}(\bar{V}^T)\|_{2,\infty} \\
&\leq \|U_X\Sigma_X^{-1}V_X^TV_0\|\|\mathcal{P}_{\mathcal{I}_0^c}(\bar{V}^T)\|_{2,\infty} \\
&= \|\Sigma_X^{-1}V_X^TV_0\|\|\mathcal{P}_{\mathcal{I}_0^c}(\bar{V}^T)\|_{2,\infty} \\
&\leq \frac{1}{\beta\|X\|}\|\mathcal{P}_{\mathcal{I}_0^c}(\bar{V}^T)\|_{2,\infty} \\
&\leq \frac{1}{\beta\|X\|}\sqrt{\frac{\mu r_0}{(1-\gamma)n}},
\end{aligned} \tag{16}$$

where $\|\Sigma_X^{-1}V_X^TV_0\| \leq \frac{1}{\beta\|X\|}$ is due the the definition of β, and the last inequality is due to Lemma 12.

We expand Q_2 for convenience:

$$\begin{aligned}
Q_2 &= \lambda\mathcal{P}_{\mathcal{I}_0^c}\mathcal{P}_{\bar{V}}(\mathtt{I}+\sum_{i=1}^{\infty}(\mathcal{P}_{\bar{V}}\mathcal{P}_{\mathcal{I}_0}\mathcal{P}_{\bar{V}})^i)\mathcal{P}_{\bar{V}}\mathcal{P}^{L}_{V_0^{\perp}}(X^T\hat{H}) \\
&= \lambda(\mathtt{I}-V_0V_0^T)(X^T\hat{H})\bar{V}\bar{V}^T(\mathtt{I}+\sum_{i=1}^{\infty}\bar{V}G^i\bar{V}^T)\bar{V}\mathcal{P}_{\mathcal{I}_0^c}(\bar{V}^T).
\end{aligned}$$

Write $Q_2 = \lambda(\bar{Q}_2 - \tilde{Q}_2)$, with

$$\begin{aligned}
\bar{Q}_2 &\triangleq X^T\hat{H}\bar{V}\bar{V}^T(\mathtt{I}+\sum_{i=1}^{\infty}\bar{V}G^i\bar{V}^T)\bar{V}\mathcal{P}_{\mathcal{I}_0^c}(\bar{V}^T), \\
\tilde{Q}_2 &\triangleq V_0V_0^TX^T\hat{H}\bar{V}\bar{V}^T(\mathtt{I}+\sum_{i=1}^{\infty}\bar{V}G^i\bar{V}^T)\bar{V}\mathcal{P}_{\mathcal{I}_0^c}(\bar{V}^T).
\end{aligned}$$

Then we have

$$\begin{aligned}
\|\mathcal{P}_{\mathcal{I}_0^c}(U_X\Sigma_X^{-1}V_X^T\bar{Q}_2)\|_{2,\infty} &= \|U_X\Sigma_X^{-1}V_X^T\mathcal{P}_{\mathcal{I}_0^c}(\bar{Q}_2)\|_{2,\infty} \\
&= \|U_XU_X^T\hat{H}\bar{V}\bar{V}^T(\mathtt{I}+\sum_{i=1}^{\infty}\bar{V}G^i\bar{V}^T)\bar{V}\mathcal{P}_{\mathcal{I}_0^c}(\bar{V}^T)\|_{2,\infty} \\
&\le \|\hat{H}\bar{V}\bar{V}^T(\mathtt{I}+\sum_{i=1}^{\infty}\bar{V}G^i\bar{V}^T)\bar{V}\mathcal{P}_{\mathcal{I}_0^c}(\bar{V}^T)\|_{2,\infty} \\
&\le \|\hat{H}\|\|\bar{V}\bar{V}^T(\mathtt{I}+\sum_{i=1}^{\infty}\bar{V}G^i\bar{V}^T)\bar{V}\mathcal{P}_{\mathcal{I}_0^c}(\bar{V}^T)\|_{2,\infty} \\
&\le \|\hat{H}\|\|\bar{V}\bar{V}^T\|\|(\mathtt{I}+\sum_{i=1}^{\infty}\bar{V}G^i\bar{V}^T)\|\|\bar{V}\|\|\mathcal{P}_{\mathcal{I}_0^c}(\bar{V}^T)\|_{2,\infty} \\
&\le \sqrt{\gamma n}\frac{1}{1-\psi}\sqrt{\frac{\mu r_0}{(1-\gamma)n}} \\
&= \frac{1}{1-\psi}\sqrt{\frac{\gamma}{1-\gamma}\mu r_0}, \qquad (17)
\end{aligned}$$

and

$$\begin{aligned}
&\|\mathcal{P}_{\mathcal{I}_0^c}(U_X\Sigma_X^{-1}V_X^T\tilde{Q}_2)\|_{2,\infty} = \|U_X\Sigma_X^{-1}V_X^T\mathcal{P}_{\mathcal{I}_0^c}(\tilde{Q}_2)\|_{2,\infty} \\
&= \|U_X\Sigma_X^{-1}V_X^TV_0V_0^TX^T\hat{H}\bar{V}\bar{V}^T(\mathtt{I}+\sum_{i=1}^{\infty}\bar{V}G^i\bar{V}^T)\bar{V}\mathcal{P}_{\mathcal{I}_0^c}(\bar{V}^T)\|_{2,\infty} \\
&\le \|\Sigma_X^{-1}V_X^TV_0\|\|V_0^TX^T\|\|\hat{H}\|\|\bar{V}\bar{V}^T\|\|(\mathtt{I}+\sum_{i=1}^{\infty}\bar{V}G^i\bar{V}^T)\|\|\bar{V}\|\|\mathcal{P}_{\mathcal{I}_0^c}(\bar{V}^T)\|_{2,\infty} \\
&\le \frac{1}{\beta\|X\|}\|X\|\sqrt{\gamma n}\frac{1}{1-\psi}\sqrt{\frac{\mu r_0}{(1-\gamma)n}} \\
&= \frac{1}{\beta(1-\psi)}\sqrt{\frac{\gamma}{1-\gamma}\mu r_0}. \qquad (18)
\end{aligned}$$

Combing (16), (17) and (18) together, we have

$$\begin{aligned}
\|\mathcal{P}_{\mathcal{I}_0^c}(Q)\|_{2,\infty} \quad &\le \quad \|\mathcal{P}_{\mathcal{I}_0^c}(U_X\Sigma_X^{-1}V_X^TV_0\bar{V}^T)\|_{2,\infty} + \lambda\|\mathcal{P}_{\mathcal{I}_0^c}(U_X\Sigma_X^{-1}V_X^T\bar{Q}_2)\|_{2,\infty} \\
&+ \quad \lambda\|\mathcal{P}_{\mathcal{I}_0^c}(U_X\Sigma_X^{-1}V_X^T\tilde{Q}_2)\|_{2,\infty} \\
&\le \quad \frac{1}{\beta\|X\|}\sqrt{\frac{\mu r_0}{(1-\gamma)n}} + \frac{\lambda}{(1-\psi)}\sqrt{\frac{\gamma}{1-\gamma}\mu r_0} + \frac{\lambda}{\beta(1-\psi)}\sqrt{\frac{\gamma}{1-\gamma}\mu r_0} \\
&= \quad \frac{1}{\beta\|X\|}\sqrt{\frac{\mu r_0}{(1-\gamma)n}} + \frac{\lambda(1+\beta)}{\beta(1-\psi)}\sqrt{\frac{\gamma}{1-\gamma}\mu r_0}.
\end{aligned}$$

Hence,

$$\begin{aligned}
&\|\mathcal{P}_{\mathcal{I}_0^c}(Q)\|_{2,\infty} < \lambda \\
&\Leftarrow \frac{1}{\beta\|X\|}\sqrt{\frac{\mu r_0}{(1-\gamma)n}} + \frac{\lambda(1+\beta)}{\beta(1-\psi)}\sqrt{\frac{\gamma}{1-\gamma}\mu r_0} < \lambda \\
&\Leftarrow \frac{1}{\beta\|X\|}\sqrt{\frac{\mu r_0}{(1-\gamma)n}} < \lambda(1-\frac{1+\beta}{\beta(1-\psi)}\sqrt{\frac{\gamma}{1-\gamma}\mu r_0}) \\
&\Leftarrow \frac{1-\psi}{\|X\|}\sqrt{\frac{\mu r_0}{(1-\gamma)n}} < \lambda(\beta(1-\psi)-(1+\beta)\sqrt{\frac{\gamma}{1-\gamma}\mu r_0}) \\
&\Leftarrow \lambda > \frac{(1-\psi)\sqrt{\frac{\mu r_0}{(1-\gamma)}}}{\|X\|\sqrt{n}(\beta(1-\psi)-(1+\beta)\sqrt{\frac{\gamma}{1-\gamma}\mu r_0})},
\end{aligned}$$

as long as $\beta(1-\psi) - (1+\beta)\sqrt{\frac{\gamma}{1-\gamma}\mu r_0} > 0$, which is proven in the following step.

S6: We have shown that each of the 5 conditions hold. Finally, we show that the bounds on λ can be satisfied. But this amounts to a condition on the outlier fraction γ. Indeed, we have

$$\frac{(1-\psi)\sqrt{\frac{\mu r_0}{(1-\gamma)}}}{\|X\|\sqrt{n}(\beta(1-\psi)-(1+\beta)\sqrt{\frac{\gamma}{1-\gamma}\mu r_0})} < \frac{1-\psi}{\|X\|\sqrt{n}(2-\psi)\sqrt{\gamma}}$$
$$\Leftarrow (2-\psi)\sqrt{\frac{\gamma}{(1-\gamma)}\mu r_0} < \beta(1-\psi) - (1+\beta)\sqrt{\frac{\gamma}{1-\gamma}\mu r_0}$$
$$\Leftarrow \frac{\gamma}{1-\gamma} < \frac{\beta^2(1-\psi)^2}{(3-\psi+\beta)^2\mu r_0},$$

which can be satisfied, since the right hand side does not depends on γ. Moreover, this condition also ensures $\beta(1-\psi) - (1+\beta)\sqrt{\frac{\gamma}{1-\gamma}\mu r_0} > 0$.

We have thus shown that if $\psi < 1$ and λ is within the given bounds, we can construct a dual certificate. From here, the following lemma immediately establishes our main result, Theorem 1.

Lemma 13 *Let γ^* be such that*

$$\frac{\gamma^*}{1-\gamma^*} = \frac{324\beta^2}{49(11+4\beta)^2\mu r_0},$$

then LRR, with $\lambda = \frac{3}{7\|X\|\sqrt{\gamma * n}}$, strictly succeeds as long as $\gamma \leq \gamma^*$.

Proof First note that

$$\frac{324\beta^2}{49(11+4\beta)^2\mu r_0} = \frac{36}{49}\frac{\beta^2(1-\frac{1}{4})^2}{(3-\frac{1}{4}+\beta)^2\mu r_0}.$$

Lemma 9 implies that as long as $\gamma \leq \gamma^*$ we have the following:

$$\psi \leq \lambda^2\|X\|^2\gamma n = \frac{9\gamma}{49\gamma *} \leq \frac{9}{49} < \frac{1}{4}.$$

Hence, we have

$$\frac{\beta^2(1-\psi)^2}{(3-\psi+\beta)^2\mu r_0} > \frac{\beta^2(1-\frac{1}{4})^2}{(3-\frac{1}{4}+\beta)^2\mu r_0}$$
$$\Rightarrow \frac{\gamma^*}{1-\gamma^*} < \frac{36}{49}\frac{\beta^2(1-\psi)^2}{(3-\psi+\beta)^2\mu r_0}$$
$$\Rightarrow \mu r_0 < \frac{36}{49}\frac{\beta^2(1-\psi)^2(1-\gamma^*)}{(3-\psi+\beta)^2\gamma^*}.$$

Note that $\frac{(1-\psi)\sqrt{\frac{\mu r_0}{(1-\gamma)}}}{\|X\|\sqrt{n}(\beta(1-\psi)-(1+\beta)\sqrt{\frac{\gamma}{1-\gamma}\mu r_0})}$ as a function of $\sqrt{\frac{\gamma}{1-\gamma}\mu r_0}$ is strictly increasing. Moreover, $\sqrt{\frac{\gamma}{1-\gamma}\mu r_0} <$

$\frac{\beta(1-\psi)}{3-\psi+\beta}$, and thus

$$\begin{aligned}
\frac{(1-\psi)\sqrt{\frac{\mu r_0}{(1-\gamma)}}}{\|X\|\sqrt{n}(\beta(1-\psi)-(1+\beta)\sqrt{\frac{\gamma}{1-\gamma}\mu r_0})} &< \frac{(1-\psi)\sqrt{\frac{\mu r_0}{(1-\gamma)}}(3-\psi+\beta)}{\|X\|\sqrt{n}\beta(1-\psi)(2-\psi)} \\
&< \frac{\frac{6}{7}\frac{\beta(1-\psi)^2}{3-\psi+\beta}\sqrt{\frac{1-\gamma^*}{1-\gamma}}(3-\psi+\beta)}{\|X\|\sqrt{\gamma^* n}\beta(1-\psi)(2-\psi)} \\
&= \frac{\frac{6}{7}(1-\psi)\sqrt{\frac{1-\gamma^*}{1-\gamma}}}{\|X\|\sqrt{\gamma^* n}(2-\psi)} \\
&\leq \frac{\frac{6}{7}(1-\psi)}{\|X\|\sqrt{\gamma^* n}(2-\psi)} \\
&\leq \frac{3}{7\|X\|\sqrt{\gamma^* n}},
\end{aligned}$$

where the last inequality holds because $\psi \geq 0$.

By $\psi < 1/4$, we also have

$$\frac{1-\psi}{\|X\|\sqrt{\gamma n}(2-\psi)} \geq \frac{1-\psi}{\|X\|\sqrt{\gamma^* n}(2-\psi)} > \frac{1-\frac{1}{4}}{\|X\|\sqrt{\gamma^* n}(2-\frac{1}{4})} = \frac{3}{7\|X\|\sqrt{\gamma^* n}}.$$

Hence, $\lambda = \frac{3}{7\|X\|\sqrt{\gamma^* n}}$ always satisfies the given bounds, as long as the outlier fraction γ is not higher than γ^*.

D List of Notations

X	The observed data matrix.
X_0	The ground truth of the data matrix.
C_0	The ground truth of the outliers.
$cond(\cdot)$	The condition number of a matrix.
d	The ambient data dimension, i.e., number of rows of X.
n	The number of data points, i.e., number of columns of X.
$\mathcal{I}_0$	The indices of outliers, i.e., non-zero columns of C_0.
γ	Fraction of outliers, which equals $\|\mathcal{I}_0\|/n$.
U_0, V_0	The left and right singular vectors of X_0.
μ	Incoherence parameter of V_0.
β	RWD parameter of the dictionary X.
$\hat{Z}, \hat{C}$	The optimal solution of the Oracle Problem.
$\hat{U}, \hat{V}$	The left and right singular vectors of $\hat{Z}$.
$\bar{V}$	An auxiliary matrix defined in Definition 5.
$\mathcal{B}(\cdot)$	An operator defined in Definition 4.
$\hat{H}$	An auxiliary matrix defined in Lemma 8, as $\hat{H} = \mathcal{B}(\hat{C})$.
G	An auxiliary matrix defined in Definition 6.
ϕ	Defined in Lemma 9 as $\psi = \\|G\\|$.