A Appendix

A.1 Proof of Theorem 2

Proof. First note that $\boldsymbol{W}_{encoder} \in \mathbb{R}^{r\times d}$ since we limit the dimension of the code layer of the AutoEncoder to be $r = \text{rank}(\boldsymbol{L}^*)$. The problem we consider is:

$$\min_{\boldsymbol{L},\boldsymbol{E},\boldsymbol{W}_{encoder}\in\mathbb{R}^{r\times d}} \|\boldsymbol{W}_{encoder}\boldsymbol{L}\|_{2,1} + \lambda\|\boldsymbol{E}^T\|_{2,1}, s.t.\ \boldsymbol{H} = \boldsymbol{L} + \boldsymbol{E}, \boldsymbol{L} = \mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{L})) \tag{11}$$

Since we have the constraint $\boldsymbol{L} = \mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{L}))$, the rank of any feasible $\boldsymbol{L}$ should be no more than r, and $\boldsymbol{L}$ must lie within the row-space of $\boldsymbol{W}_{encoder}$. So we can write $\boldsymbol{W}_{encoder} = \boldsymbol{R}\boldsymbol{U}^T$ where $\boldsymbol{R} \in \mathbb{R}^{r\times r}$, and $\boldsymbol{U}$ are the top-r left singular vectors of $\boldsymbol{L}$. As the rows of $\boldsymbol{W}_{encoder}$ are orthornormal, *i.e.*, $\boldsymbol{W}_{encoder}\boldsymbol{W}_{encoder}^T = \boldsymbol{I}$, so $\boldsymbol{R}\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{R}^T = \boldsymbol{I}$ and therefore $\boldsymbol{R}\boldsymbol{R}^T = \boldsymbol{I}$ (so this square matrix $\boldsymbol{R}$ is unitary). Further, since setting $\boldsymbol{W}_{decoder} = \boldsymbol{W}_{encoder}^T$ will always meet the constraint $\boldsymbol{L} = \mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{L}))$, the main problem we need to consider becomes:

$$\min_{\substack{\{\boldsymbol{L}|\text{rank}(\boldsymbol{L})\le r\},\boldsymbol{E},\\ \boldsymbol{W}_{encoder}=\boldsymbol{R}\boldsymbol{U}^T,\boldsymbol{R}\text{ is unitary}}} \|\boldsymbol{W}_{encoder}\boldsymbol{L}\|_{2,1} + \lambda\|\boldsymbol{E}^T\|_{2,1}, s.t.\ \boldsymbol{H} = \boldsymbol{L} + \boldsymbol{E}, \tag{12}$$

Since $\|\boldsymbol{L}\|_* = \|\boldsymbol{R}\boldsymbol{U}^T\boldsymbol{L}\|_* = \|\boldsymbol{W}_{encoder}\boldsymbol{L}\|_* = \|\boldsymbol{I}(\boldsymbol{W}_{encoder}\boldsymbol{L})\|_* = \|\sum_{i=1}^r \boldsymbol{e}_i(\boldsymbol{W}_{encoder}\boldsymbol{L})_{i,:}\|_* \le \sum_{i=1}^r \|\boldsymbol{e}_i(\boldsymbol{W}_{encoder}\boldsymbol{L})_{i,:}\| = \sum_{i=1}^r \|(\boldsymbol{W}_{encoder}\boldsymbol{L})_{i,:}\|_2 \triangleq \|\boldsymbol{W}_{encoder}\boldsymbol{L}\|_{2,1}$. The equality is achieved when $\boldsymbol{R} = \boldsymbol{I}$, *i.e.*, $\boldsymbol{W}_{encoder} = \boldsymbol{U}^T$, because $\|\boldsymbol{W}_{encoder}\boldsymbol{L}\|_{2,1} = \|\boldsymbol{U}^T\boldsymbol{L}\|_{2,1} = \|\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\|_{2,1} = \|\boldsymbol{\Sigma}\boldsymbol{V}^T\|_{2,1} = \sum_{i=1}^r \sigma_i = \|\boldsymbol{L}\|_*$.

So

$$\min_{\substack{\{\boldsymbol{L}|\text{rank}(\boldsymbol{L})\le r\},\boldsymbol{E},\\ \boldsymbol{W}_{encoder}=\boldsymbol{R}\boldsymbol{U}^T,\boldsymbol{R}\text{ is unitary}}} \|\boldsymbol{W}_{encoder}\boldsymbol{L}\|_{2,1} + \lambda\|\boldsymbol{E}^T\|_{2,1} \ge \min_{\{\boldsymbol{L}|\text{rank}(\boldsymbol{L})\le r\},\boldsymbol{E}} \|\boldsymbol{L}\|_* + \lambda\|\boldsymbol{E}^T\|_{2,1}, \tag{13}$$

with equality achieved when $\boldsymbol{W}_{encoder}$ equals top-r left singular vectors of $\boldsymbol{L}$.

Then we only need to consider the right-hand-side of Eq. 13, *i.e.*,

$$\min_{\{\boldsymbol{L}|\text{rank}(\boldsymbol{L})\le r\},\boldsymbol{E}} \|\boldsymbol{L}\|_* + \lambda\|\boldsymbol{E}^T\|_{2,1} \quad s.t.\ \boldsymbol{H} = \boldsymbol{L} + \boldsymbol{E} \tag{14}$$

Recall that under the conditions of Theorem 1, the solution $\hat{\boldsymbol{L}}$ of Eq. 2 has exactly the same column space as $\boldsymbol{L}^*$ ($\hat{\boldsymbol{L}}$ may not and not necessary to be equal to $\boldsymbol{L}^*$), so $\text{rank}(\hat{\boldsymbol{L}}) = r$. Then the solution $\hat{\boldsymbol{L}}$ of Eq. 2 must also be the global optimal solution of Eq. 14. Finally, as the row-space of $\boldsymbol{W}_{encoder}$ equals the column-space of $\hat{\boldsymbol{L}}$, it recovers the underlying subspace of $\boldsymbol{L}^*$ exactly. □

A.2 Proof of Theorem 3

Proof. A) Suppose $\mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}'))$ is the global optimal solution of Eq. 10 that is different from $\mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}^*))$. Let $\mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}')) = \mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}^*)) - \boldsymbol{v}$. So we have $\boldsymbol{v} \in Range(\boldsymbol{W}_{end})\backslash 0$.

$$\sum_{i=1}^{M} \|[\boldsymbol{h} - \mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}'))]_i\|_2 \tag{15}$$

$$= \sum_{i=1}^{M} \|[\boldsymbol{h} - \mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}^*)) + \boldsymbol{v}]_i\|_2 \tag{16}$$

$$= \sum_{i=1}^{M} \|[\boldsymbol{h} - \mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}^*))]_i + \boldsymbol{v}_i\|_2 \tag{17}$$

$$= \sum_{i\in S} \|[\boldsymbol{h} - \mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}^*))]_i + \boldsymbol{v}_i\|_2 + \sum_{i\in \bar{S}} \|[\boldsymbol{h} - \mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}^*))]_i + \boldsymbol{v}_i\|_2 \tag{18}$$

$$= \sum_{i\in S} \|\boldsymbol{v}_i\|_2 + \sum_{i\in \bar{S}} \|[\boldsymbol{h} - \mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}^*))]_i + \boldsymbol{v}_i\|_2 \tag{19}$$

$$\geq \sum_{i\in S} \|\boldsymbol{v}_i\|_2 + \sum_{i\in \bar{S}} \|[\boldsymbol{h} - \mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}^*))]_i\|_2 - \sum_{i\in \bar{S}} \|\boldsymbol{v}_i\|_2 \tag{20}$$

$$> \sum_{i\in \bar{S}} \|[\boldsymbol{h} - \mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}^*))]_i\|_2 \tag{21}$$

$$= \sum_{i=1}^{M} \|[\boldsymbol{h} - \mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}^*))]_i\|_2 \tag{22}$$

where S is the index set of size g such that $\boldsymbol{e}_i^* = \boldsymbol{0}, \forall i \in S$. And the last inequality follows from the assumed range space property since $\boldsymbol{v} \in Range(\boldsymbol{W}_{end})\backslash 0$.

The above contradicts the assumption that $\mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}'))$ is the global optimal solution that is different from $\mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}^*))$.

B) First, note that $\boldsymbol{h} = \mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}^*)) + \boldsymbol{e}^*$ and $\sum_{i=1}^{M} \mathbb{1}\{\boldsymbol{e}_i^* \neq \boldsymbol{0}\} \leq M - g$. The following proof strategy is motivated from the robust linear regression with block-sparse corruptions [12]. Suppose $\mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}'))$ is the global optimal solution of Eq. 9 that is different from $\mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}^*))$. Let $\boldsymbol{h} = \mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}')) + \boldsymbol{e}'$, then we have

$$\sum_{i=1}^{M} \mathbb{1}\{\boldsymbol{e}_i' \neq \boldsymbol{0}\} \leq \sum_{i=1}^{M} \mathbb{1}\{\boldsymbol{e}_i^* \neq \boldsymbol{0}\} \leq M - g \tag{23}$$

and $\mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}')) + \boldsymbol{e}' = \mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}^*)) + \boldsymbol{e}^*$.

From Eq. 23 we know that

$$\sum_{i=1}^{M} \mathbb{1}\{[\mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}^*)) - \mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}'))]_i = \boldsymbol{0}\} \geq M - \sum_{i=1}^{M} \mathbb{1}\{\boldsymbol{e}_i^* \neq \boldsymbol{0}\} - \sum_{i=1}^{M} \mathbb{1}\{\boldsymbol{e}_i' \neq \boldsymbol{0}\} \tag{24}$$

$$\geq M - (M-g) - (M-g) = 2g - M \tag{25}$$

Let $\mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}')) = \mathcal{D}_\phi(\mathcal{E}_\psi(\boldsymbol{l}^*)) - \boldsymbol{v}$, so we have $\boldsymbol{v} \in Range(\boldsymbol{W}_{end})\backslash 0$ and $\sum_{i=1}^{M} \mathbb{1}\{\boldsymbol{v}_i = \boldsymbol{0}\} \geq 2g - M$. Now split the M blocks into 3 disjoint sets $\{\Omega_0, \Omega_1, \Omega_2\}$, where Ω_0 is any subset with size $2g - M$ such that $\boldsymbol{v}_{\Omega_0} = \boldsymbol{0}$, and $|\Omega_1| = |\Omega_2| = M - g$. Since $|\Omega_0 \cup \Omega_1| = g$, by our assumption, we have $\sum_{i\in\Omega_0\cup\Omega_1} \|\boldsymbol{v}_i\|_2 > \sum_{i\in\Omega_2} \|\boldsymbol{v}_i\|_2$. Since $|\Omega_0 \cup \Omega_2| = g$, by our assumption, we have $\sum_{i\in\Omega_0\cup\Omega_2} \|\boldsymbol{v}_i\|_2 > \sum_{i\in\Omega_1} \|\boldsymbol{v}_i\|_2$. However, this leads to a contradiction since $\sum_{i\in\Omega_0} \|\boldsymbol{v}_i\|_2 = 0$. □