

APPENDIX

A. Proof of theorem 1

We start with the sampling of the rows. Theorem 5 in [28] shows that for any $\delta_r, \epsilon_r \in (0, 1)$, with probability at least $1 - \epsilon_r$,

$$(1 - \delta_r) \|z\|_2^2 \leq \frac{p}{\rho_r} \|M_r z\|_2^2 \leq (1 + \delta_r) \|z\|_2^2$$

for all $z \in \text{span}(P_{k_r})$ provided that

$$\rho_r \geq \frac{3}{\delta_r^2} \nu_{k_r}^2 \log \left(\frac{2k_r}{\epsilon_r} \right). \quad (26)$$

Notice that Theorem 5 in [28] is a uniform result. As a consequence, with probability at least $1 - \epsilon_r$,

$$(1 - \delta_r) \|y_i\|_2^2 \leq \frac{p}{\rho_r} \|M_r y_i\|_2^2 \leq (1 + \delta_r) \|y_i\|_2^2, \quad i = 1, \dots, n, \quad (27)$$

for all $y_1, \dots, y_n \in \text{span}(P_{k_r})$ provided that (26) holds. Summing the previous inequalities over all i shows that, with probability at least $1 - \epsilon_r$,

$$(1 - \delta_r) \|Y\|_F^2 \leq \frac{p}{\rho_r} \|M_r Y\|_F^2 \leq (1 + \delta_r) \|Y\|_F^2, \quad (28)$$

for all $Y \in \mathbb{R}^{p \times n}$ with column-vectors in $\text{span}(P_{k_r})$.

Let us continue with the sampling of the columns. Again, Theorem 5 in [28] shows that for any $\delta_c, \epsilon_c \in (0, 1)$, with probability at least $1 - \epsilon_c$,

$$(1 - \delta_c) \|w\|_2^2 \leq \frac{n}{\rho_c} \|w^\top M_c\|_2^2 \leq (1 + \delta_c) \|w\|_2^2$$

for all $w \in \text{span}(Q_{k_c})$ provided that

$$\rho_c \geq \frac{3}{\delta_c^2} \nu_{k_c}^2 \log \left(\frac{2k_c}{\epsilon_c} \right). \quad (29)$$

As a consequence, with probability at least $1 - \epsilon_c$,

$$(1 - \delta_c) \|z_i\|_2^2 \leq \frac{n}{\rho_c} \|z_i^\top M_c\|_2^2 \leq (1 + \delta_c) \|z_i\|_2^2, \quad i = 1, \dots, \rho_r, \quad (30)$$

for all $z_1, \dots, z_{\rho_r} \in \text{span}(Q_{k_c})$ provided that (29) holds. Summing the previous inequalities over all i shows that, with probability at least $1 - \epsilon_c$,

$$(1 - \delta_c) \|Z\|_F^2 \leq \frac{n}{\rho_c} \|Z M_c\|_F^2 \leq (1 + \delta_c) \|Z\|_F^2 \quad (31)$$

for all $Z \in \mathbb{R}^{\rho_r \times n}$ with row-vectors in $\text{span}(Q_{k_c})$. In particular, this property holds, with at least the same probability, for all matrices Z of the form $M_r Y$ where $Y \in \mathbb{R}^{p \times n}$ is a matrix with row-vectors in $\text{span}(Q_{k_c})$.

We now continue by combining (28) and (31). We obtain that

$$(1 - \delta_c)(1 - \delta_r) \|Y\|_F^2 \leq \frac{np}{\rho_c \rho_r} \|M_r Y M_c\|_F^2 \leq (1 + \delta_c)(1 + \delta_r) \|Y\|_F^2 \quad (32)$$

for all $Y \in \mathbb{R}^{p \times n}$ with column-vectors in $\text{span}(P_{k_r})$ and row-vectors in $\text{span}(Q_{k_c})$, provided that (26) and (29) hold. It remains to compute the probability that (32) holds. Property (32) does not hold if (28) or (31) do not hold. Using the union bound, (32) does not hold with probability at most $\epsilon_r + \epsilon_c$. To finish the proof, one just need to choose $\epsilon_r = \epsilon_c = \epsilon/2$ and $\delta_r = \delta_c = \delta/3$, and notice that $(1 + \delta/3)^2 \leq 1 + \delta$ and $(1 - \delta/3)^2 \geq 1 - \delta$ for $\delta \in (0, 1)$.

B. Proof of Theorem 2

Using the optimality condition we have, for any $Z \in \mathbb{R}^{p \times n}$,

$$\|M_r X^* M_c - \tilde{X}\|_F \leq \|M_r Z M_c - \tilde{X}\|_F.$$

For $Z = \bar{X}$, we have

$$\|M_r X^* M_c - \tilde{X}\|_F \leq \|M_r \bar{X} M_c - \tilde{X}\|_F,$$

which gives

$$\|M_r X^* M_c - M_r \bar{X} M_c - \tilde{E}\|_F \leq \|\tilde{E}\|_F.$$

As (2) holds, we have

$$\begin{aligned}\|M_r X^* M_c - M_r \bar{X} M_c - \tilde{E}\|_F &\geq \|M_r(X^* - \bar{X})M_c\|_F - \|\tilde{E}\|_F \\ &\geq \sqrt{\frac{\rho_r \rho_c (1 - \delta)}{np}} \|X^* - \bar{X}\|_F - \|\tilde{E}\|_F.\end{aligned}$$

Therefore, by combining the above equations we get

$$\|X^* - \bar{X}\|_F \leq 2\sqrt{\frac{np}{\rho_r \rho_c (1 - \delta)}} \|\tilde{E}\|_F.$$

C. Proof of Theorem 3

Using the optimality condition we have for any $Z \in \mathbb{R}^{p \times n}$ and optimal solution $X^* = \bar{X}^* + E^*$:

$$\|M_r X^* M_c - \tilde{X}\|_F^2 + \bar{\gamma}_c \text{tr}(X^* \mathcal{L}_c X^{*\top}) + \bar{\gamma}_r \text{tr}(X^{*\top} \mathcal{L}_r X^*) \leq \|M_r Z M_c - \tilde{X}\|_F^2 + \bar{\gamma}_c \text{tr}(Z \mathcal{L}_c Z^\top) + \bar{\gamma}_r \text{tr}(Z^\top \mathcal{L}_r Z) \quad (33)$$

using $Z = \bar{X} = P_{k_r} Y_b Q_{k_c}^\top$ as in the proof of theorem 2 in [36], where $Y_b \in \mathbb{R}^{k_r \times k_c}$ and it is not necessarily diagonal. Note that $\|Y_b\|_F = \|\bar{X}\|_F$, $Q_{k_c}^\top Q_{k_c} = I_{k_c}$, $P_{k_r}^\top P_{k_r} = I_{k_r}$, $\bar{Q}_{k_c}^\top Q_{k_c} = 0$, $\bar{P}_{k_r}^\top P_{k_r} = 0$. From the proof of theorem 2 in [36] we also know that:

$$\begin{aligned}\text{tr}(\bar{X} \mathcal{L}_c (\bar{X})^\top) &\leq \lambda_{k_c} \|\bar{X}\|_F^2 \\ \text{tr}(\bar{X}^\top \mathcal{L}_r (\bar{X})) &\leq \lambda_{k_r} \|\bar{X}\|_F^2 \\ \text{tr}(X^* \mathcal{L}_c (X^*)^\top) &\geq \lambda_{k_c+1} \|X^* \bar{Q}_{k_c}\|_F^2 \\ \text{tr}(X^* \mathcal{L}_r (X^*)^\top) &\geq \lambda_{k_r+1} \|\bar{P}_{k_r}^\top X^*\|_F^2\end{aligned}$$

Now using all this information in (33) we get

$$\|M_r X^* M_c - \tilde{X}\|_F^2 + \bar{\gamma}_c \lambda_{k_c+1} \|X^* \bar{Q}_{k_c}\|_F^2 + \bar{\gamma}_r \lambda_{k_r+1} \|\bar{P}_{k_r}^\top X^*\|_F^2 \leq \|\tilde{E}\|_F^2 + (\bar{\gamma}_c \lambda_{k_c} + \bar{\gamma}_r \lambda_{k_r}) \|\bar{X}\|_F^2$$

From above we have:

$$\|M_r X^* M_c - \tilde{X}\|_F \leq \|\tilde{E}\|_F + \sqrt{(\bar{\gamma}_c \lambda_{k_c} + \bar{\gamma}_r \lambda_{k_r})} \|\bar{X}\|_F \quad (34)$$

and

$$\sqrt{(\bar{\gamma}_c \lambda_{k_c+1} \|X^* \bar{Q}_{k_c}\|_F^2 + \bar{\gamma}_r \lambda_{k_r+1} \|\bar{P}_{k_r}^\top X^*\|_F^2)} \leq \|\tilde{E}\|_F + \sqrt{(\bar{\gamma}_c \lambda_{k_c} + \bar{\gamma}_r \lambda_{k_r})} \|\bar{X}\|_F \quad (35)$$

using

$$\bar{\gamma}_c = \gamma \frac{1}{\lambda_{k_c+1}} \quad \text{and} \quad \bar{\gamma}_r = \gamma \frac{1}{\lambda_{k_r+1}},$$

and

$$\|E^*\|_F^2 = \|X^* \bar{Q}_{k_c}\|_F^2 = \|\bar{P}_{k_r}^\top X^*\|_F^2$$

we get:

$$\|M_r X^* M_c - \tilde{X}\|_F \leq \|\tilde{E}\|_F + \sqrt{\gamma \left(\frac{\lambda_{k_c}}{\lambda_{k_c+1}} + \frac{\lambda_{k_r}}{\lambda_{k_r+1}} \right)} \|\bar{X}\|_F \quad (36)$$

and

$$\sqrt{2\gamma} \|E^*\|_F \leq \|\tilde{E}\|_F + \sqrt{\gamma \left(\frac{\lambda_{k_c}}{\lambda_{k_c+1}} + \frac{\lambda_{k_r}}{\lambda_{k_r+1}} \right)} \|\bar{X}\|_F \quad (37)$$

which implies

$$\|E^*\|_F \leq \frac{\|\tilde{E}\|_F}{\sqrt{2\gamma}} + \frac{1}{\sqrt{2}} \sqrt{\gamma \left(\frac{\lambda_{k_c}}{\lambda_{k_c+1}} + \frac{\lambda_{k_r}}{\lambda_{k_r+1}} \right)} \|\bar{X}\|_F \quad (38)$$

Focus on $\|M_r X^* M_c - \tilde{X}\|_F^2$ now. As M_r, M_c are constructed with a sampling without replacement, we have $\|M_r E^* M_c\|_F \leq \|E^*\|_F$. Now using the above facts and the RIP we get:

$$\begin{aligned}\|M_r X^* M_c - \tilde{X}\|_F &= \|M_r(\bar{X}^* + E^*)M_c - M_r \bar{X} M_c - \tilde{E}\|_F \\ &\geq \sqrt{\frac{\rho_r \rho_c (1 - \delta)}{np}} \|\bar{X}^* - \bar{X}\|_F - \|\tilde{E}\|_F - \|E^*\|_F\end{aligned}$$

this implies

$$\|\bar{X}^* - \bar{X}\|_F \leq \sqrt{\frac{np}{\rho_c \rho_r (1-\delta)}} \left[\left(2 + \frac{1}{\sqrt{2\gamma}}\right) \|\tilde{E}\|_F + \left(\frac{1}{\sqrt{2}} + \sqrt{\gamma}\right) \sqrt{\left(\frac{\lambda_{k_c}}{\lambda_{k_c+1}} + \frac{\lambda_{k_r}}{\lambda_{k_r+1}}\right)} \|\bar{X}\|_F \right]$$

Discussion Let $A_1, A_2 \in \mathbb{R}^{p \times n}$ and $A_1 = U_1 S_1 V_1^T$, $A_2 = U_2 S_2 V_2^T$ then if $\|A_1 - A_2\|_F^2 \rightarrow 0$, then $S_1 \rightarrow S_2$.

We observe that

$$\|A_1 - A_2\|_F^2 = \|U_1 S_1 V_1^T - U_2 S_2 V_2^T\|_F^2 = \|U_2^T U_1 S_1 V_1^T V_2 - S_2\|_F^2$$

which implies that $U_2^T U_1 S_1 V_1^T V_2 \approx S_2$. This is equivalent to saying that for the significant values of S_2 , the orthonormal matrices $U_2^T U_1$ and $V_1^T V_2$ have to be almost diagonal. As a result, for the significant values of S_2 , U_2 and V_2 have to be aligned with U_1 and V_1 . The same reason also implies that $S_1 \approx S_2$.

D. Solution of eq. (11)

Let us examine how to solve (11). The problem can be reformulated as:

$$\min_U \text{tr}(U^T \mathcal{L}_r U) \quad \text{s.t.:} \quad U^T U = I_k, \quad \|M_r U - \tilde{U}\|_F^2 < \epsilon$$

Let U' is the zero appended matrix of \tilde{U} , then we can re-write it as:

$$\min_U \text{tr}(U^T \mathcal{L}_r U) \quad \text{s.t.:} \quad U^T U = I_k, \quad \|M_r(U - U')\|_F^2 < \epsilon$$

The above problem is equivalent to (11), as the term $\|M_r(U - U')\|_F^2$ has been removed from the objective and introduced as a constraint. Note that the constant γ_r is not needed anymore. The new model parameter ϵ controls the radius of the L_2 ball $\|M_r(U - U')\|_F^2$. In simple words it controls how much noise is tolerated by the projection of U on the ball that is centered at U' . To solve the above problem one needs to split it down into two sub-problems and solve iteratively between:

- 1) The optimization $\min_U \text{tr}(U^T \mathcal{L}_r U) \quad \text{s.t.:} \quad U^T U = I_k$. The solution to this problem is given by the lowest k eigenvectors of \mathcal{L}_r . Thus it requires a complexity of $\mathcal{O}((n+p)k^2)$ for solving both problems (11).
- 2) The projection on the L_2 ball $\|M_r(U - U')\|_F^2$ whose complexity is $\mathcal{O}(\rho_c + \rho_r)$.

Thus the solution requires a double iteration with a complexity of $\mathcal{O}(Ink^2)$ and is almost as expensive as FRPCAG.

E. Proof of Theorem 4

We can write (12) and (13) as following:

$$\min_{u_1 \dots u_p} \sum_{i=1}^p [\|M_r u_i - \tilde{u}_i\|_2^2 + \gamma'_r u_i^T \mathcal{L}_r u_i] \quad (39)$$

$$\min_{v_1 \dots v_n} \sum_{i=1}^n [\|M_c^T v_i - \tilde{v}_i\|_2^2 + \gamma'_c v_i^T \mathcal{L}_c v_i] \quad (40)$$

In this proof, we only treat Problem (39) and the recovery of \bar{U} . The proof for Problem (13) and the recovery of \bar{V} is identical. The above two problems can be solved independently for every i . From theorem 3.2 of [28] we obtain:

$$\|\bar{u}_i^* - \bar{u}_i\|_2 \leq \sqrt{\frac{p}{\rho_r(1-\delta)}} \left[\left(2 + \frac{1}{\sqrt{\gamma'_r \lambda_{k_r+1}}}\right) \|\tilde{e}_i^u\|_2 + \left(\sqrt{\frac{\lambda_{k_r}}{\lambda_{k_r+1}}} + \sqrt{\gamma'_r \lambda_{k_r}}\right) \|\bar{u}_i\|_2 \right], \quad (41)$$

and

$$\|e_i^*\|_2 \leq \frac{1}{\sqrt{\gamma'_r \lambda_{k_r+1}}} \|\tilde{e}_i^u\|_2 + \sqrt{\frac{\lambda_{k_r}}{\lambda_{k_r+1}}} \|\bar{u}_i\|_2,$$

which implies

$$\|\bar{u}_i^* - \bar{u}_i\|_2^2 \leq 2 \frac{p}{\rho_r(1-\delta)} \left[\left(2 + \frac{1}{\sqrt{\gamma'_r \lambda_{k_r+1}}}\right)^2 \|\tilde{e}_i^u\|_2^2 + \left(\sqrt{\frac{\lambda_{k_r}}{\lambda_{k_r+1}}} + \sqrt{\gamma'_r \lambda_{k_r}}\right)^2 \|\bar{u}_i\|_2^2 \right],$$

and

$$\|e_i^*\|_2^2 \leq \frac{2}{\gamma'_r \lambda_{k_r+1}} \|\tilde{e}_i^u\|_2^2 + 2 \frac{\lambda_{k_r}}{\lambda_{k_r+1}} \|\bar{u}_i\|_2^2.$$

Summing the previous inequalities over all i 's yields

$$\|\bar{U}^* - \bar{U}\|_F^2 \leq 2 \frac{p}{\rho_r(1-\delta)} \left[\left(2 + \frac{1}{\sqrt{\gamma'_r \lambda_{k_r+1}}} \right)^2 \|\tilde{E}^u\|_F^2 + \left(\sqrt{\frac{\lambda_{k_r}}{\lambda_{k_r+1}}} + \sqrt{\gamma'_r \lambda_{k_r}} \right)^2 \|\bar{U}\|_F^2 \right],$$

and

$$\|E^*\|_F^2 \leq \frac{2}{\gamma'_r \lambda_{k_r+1}} \|\tilde{E}^u\|_F^2 + 2 \frac{\lambda_{k_r}}{\lambda_{k_r+1}} \|\bar{U}\|_F^2.$$

Taking the square root of both inequalities terminates the proof. Similarly, the expressions for \bar{V} can be derived:

$$\|\bar{V}^* - \bar{V}\|_F \leq \sqrt{\frac{2n}{\rho_c(1-\delta)}} \left[\left(2 + \frac{1}{\sqrt{\gamma'_c \lambda_{k_c+1}}} \right) \|\tilde{E}^v\|_F + \left(\sqrt{\frac{\lambda_{k_c}}{\lambda_{k_c+1}}} + \sqrt{\gamma'_c \lambda_{k_c}} \right) \|\bar{V}\|_F \right]$$

and

$$\|E^*\|_F \leq \sqrt{\frac{2}{\gamma'_c \lambda_{k_c+1}}} \|\tilde{E}^v\|_F + \sqrt{2 \frac{\lambda_{k_c}}{\lambda_{k_c+1}}} \|\bar{V}\|_F.$$

F. Proof of Lemma 1

Let $S = [S_a^\top | S_b^\top]^\top$. Further we split \mathcal{L} into submatrices as follows:

$$\mathcal{L} = \begin{bmatrix} \mathcal{L}_{aa} & \mathcal{L}_{ab} \\ \mathcal{L}_{ba} & \mathcal{L}_{bb} \end{bmatrix}$$

Now (17) can be written as:

$$\begin{aligned} \min_{S_a} & \begin{bmatrix} S_a \\ S_b \end{bmatrix}^\top \begin{bmatrix} \mathcal{L}_{aa} & \mathcal{L}_{ab} \\ \mathcal{L}_{ba} & \mathcal{L}_{bb} \end{bmatrix} \begin{bmatrix} S_a \\ S_b \end{bmatrix} \\ \text{s.t.} & S_b = R \end{aligned}$$

further expanding we get:

$$\min_{S_a} S_a^\top \mathcal{L}_{aa} S_a + S_a^\top \mathcal{L}_{ab} R + R^\top \mathcal{L}_{ba} S_a + R \mathcal{L}_{bb} R$$

using $\nabla S_a = 0$, we get:

$$2\mathcal{L}_{ab}R + 2\mathcal{L}_{aa}S_a = 0$$

$$S_a = -\mathcal{L}_{aa}^{-1} \mathcal{L}_{ab} R$$

G. Other Approximate Decoders

Alternatively, if the complete data matrix Y is available then we can reduce the complexity further by performing a graph-upsampling for only one of the two subspaces U or V .

1) *Approximate decoder 2* : Suppose we do the upsampling only for U , then the approximate decoder 2 can be written as:

$$\min_U \text{tr}(U^\top \mathcal{L}_r U) \quad \text{s.t.} \quad M_r U = \tilde{U}.$$

The solution for U is given by eq. 18. Then, we can write V as:

$$V = Y^\top U \tilde{\Sigma}^{-1} \sqrt{\frac{\rho_c \rho_r (1-\delta)}{np}}$$

However, we do not need to explicitly determine V here. Instead the low-rank X can be determined directly from U with the projection given below:

$$X = U \tilde{\Sigma} \sqrt{\frac{np}{\rho_c \rho_r (1-\delta)}} V^\top = U U^\top Y.$$

2) *Approximate decoder 3* : Similar to the approximate decoder 2, we can propose another approximate decoder 3 which performs a graph upsampling on V and then determines U via matrix multiplication operation.

$$\min_V \text{tr}(V^\top \mathcal{L}_c V) \quad \text{s.t:} \quad M_c^\top V = \tilde{V}$$

The solution for V is given by eq. 18. Using the similar trick as for the approximate decoder 2, we can compute X without computing U . Therefore, $X = YVV^\top$.

For the proposed approximate decoders, we would need to do one SVD to determine the singular values $\tilde{\Sigma}$. However, note that this SVD is on the compressed matrix $\tilde{X} \in \mathbb{R}^{\rho_r \times \rho_c}$. Thus, it is inexpensive $\mathcal{O}(\rho_r^2 \rho_c)$ assuming that $\rho_r < \rho_c$.

H. Computational Complexities & Additional Results

We present the computational complexity of all the models considered in this work. For a matrix $X \in \mathbb{R}^{p \times n}$, let I denote the number of iterations for the algorithms to converge, p is the number of features, n is the number of samples, ρ_r, ρ_c are the number of features and samples for the compressed data \tilde{Y} and satisfy eq. (1) and theorem 1, k is the rank of the low-dimensional space or the number of clusters, \mathcal{K} is the number of nearest neighbors for graph construction, O_l, O_c correspond to the number of iterations in the Lancos and Chebyshev approximation methods. All the models which use the graph G_c are marked by '+'. The construction of graph G_r is included only in FRPCAG and CPCA. Furthermore,

- 1) We assume that $\mathcal{K}, k, \rho_r, \rho_c, p \ll n$ and $n + p + k + \mathcal{K} + \rho_r + \rho_c \approx n$.
- 2) The complexity of $\|Y - X\|_1$ is $\mathcal{O}(np)$ per iteration and that of $\|\tilde{Y} - \tilde{X}\|_1$ is $\mathcal{O}(\rho_c \rho_r)$.
- 3) The complexity of the computations corresponding to the graph regularization $\text{tr}(X \mathcal{L}_c X^\top) + \text{tr}(X^\top \mathcal{L}_r X) = \mathcal{O}(p|\mathcal{E}_c| + n|\mathcal{E}_r|) = \mathcal{O}(pn\mathcal{K} + np\mathcal{K})$, where $\mathcal{E}_r, \mathcal{E}_c$ denote the number of non-zeros in $\mathcal{L}_r, \mathcal{L}_c$ respectively. Note that we use the \mathcal{K} -nearest neighbors graphs so $\mathcal{E}_r \approx \mathcal{K}p$ and $\mathcal{E}_c \approx \mathcal{K}n$.
- 4) The complexity for the construction of $\tilde{\mathcal{L}}_c$ and $\tilde{\mathcal{L}}_r$ for compressed data \tilde{Y} is negligible if FLANN is used, i.e, $\mathcal{O}(\rho_c \rho_r \log(\rho_c))$ and $\mathcal{O}(\rho_c \rho_r \log(\rho_r))$. However, if the kron reduction strategy of Section IV-A is used then the cost is $\mathcal{O}(\mathcal{K}O_l(n + p)) \approx \mathcal{O}(\mathcal{K}O_l n)$.
- 5) We use the complexity $\mathcal{O}(np^2)$ for all the SVD computations on the matrix $X \in \mathbb{R}^{p \times n}$ and $\mathcal{O}(\rho_c \rho_r^2)$ for $\tilde{X} \in \mathbb{R}^{\rho_c \times \rho_r}$.
- 6) The complexity of $\|M_r X M_c - \tilde{X}\|_F^2$ is negligible as compared to the graph regularization terms $\text{tr}(X \mathcal{L}_c X^\top) + \text{tr}(X^\top \mathcal{L}_r X)$.
- 7) We use the approximate decoders for low-rank recovery in the complexity calculations (eq. (18) in Section V-C). All the decoders for low-rank recovery are summarized in Table X.
- 8) The complexity of k-means [10] is $\mathcal{O}(Inkp)$ for a matrix $X \in \mathbb{R}^{p \times n}$ and $\mathcal{O}(I\rho_r \rho_c k)$ for a matrix $\tilde{X} \in \mathbb{R}^{\rho_r \times \rho_c}$.

Table X: A summary and computational complexities of all the decoders proposed in this work. The Lancos method used here is presented in [39].

Type	Low-rank			
	model	complexity	Algo	parallel
ideal	$\min_X \ M_r X M_c - \tilde{X}\ _F^2$ s.t: $X^\top \in \text{span}(Q_{k_c})$ $X \in \text{span}(P_{k_r})$	$\mathcal{O}(n^3)$	–	–
alter-nate	$\min_X \ M_r X M_c - \tilde{X}\ _F^2$ $+\gamma_c \text{tr}(X \mathcal{L}_c X^\top)$ $+\gamma_r \text{tr}(X^\top \mathcal{L}_r X)$	$\mathcal{O}(Inp\mathcal{K})$	gradient descent	no
approx-imate	$\min_{U'} \ M_r U - \tilde{U}\ _F^2$ $+\gamma_r \text{tr}(U^\top \mathcal{L}_r U)$	$\mathcal{O}(In\mathcal{K})$	gradient descent	yes
	$\min_{V'} \ M_c^\top V - \tilde{V}\ _F^2$ $+\gamma_c \text{tr}(V^\top \mathcal{L}_c V)$	$\mathcal{O}(Ip\mathcal{K})$	gradient descent	yes
	$X = U \tilde{\Sigma} V^\top$	$\mathcal{O}(\rho_r^2 \rho_c)$	SVD	
Subspace-Upsampling	$\min_U \text{tr}(U^\top \mathcal{L}_r U)$ s.t: $M_r U = \tilde{U}$	$\mathcal{O}(pkO_l\mathcal{K})$	PCG	
	$\min_V \text{tr}(V^\top \mathcal{L}_c V)$ s.t: $M_c^\top V = \tilde{V}$	$\mathcal{O}(nkO_l\mathcal{K})$	PCG	yes
	$X = U \tilde{\Sigma} V^\top$	$\mathcal{O}(\rho_r^2 \rho_c)$	SVD	
approx-imate 2	$\min_U \text{tr}(U^\top \mathcal{L}_r U)$ s.t: $M_r U = \tilde{U}$ $X = U U^\top Y$	$\mathcal{O}(pkO_l\mathcal{K})$	PCG	yes
approx-imate 3	$\min_V \text{tr}(V^\top \mathcal{L}_c V)$ s.t: $M_c^\top V = \tilde{V}$ $X = Y V V^\top$	$\mathcal{O}(nkO_l\mathcal{K})$	PCG	yes

Table XI: Computational complexity of all the models considered in this work

Model	Complexity G_c $\mathcal{O}(np \log(n))$	Complexity G_r $\mathcal{O}(np \log(p))$	Complexity Algorithm for $p \ll n$	Overall Complexity (low-rank)		Overall Complexity (clustering)	
				Fast SVD for $p \ll n$	Decoder Total	k-means [10]	Decoder Total
LE [4]	+	–	$\mathcal{O}(n^3)$	–	–	$\mathcal{O}(Inkp)$	–
LLE [32]	–	–	$\mathcal{O}((p+k)n^2)$	–	–	$\mathcal{O}(Inkp)$	–
PCA	–	–	$\mathcal{O}(p^2 n)$	–	$\mathcal{O}(n(p^2 \log(n) + p^2))$	$\mathcal{O}(Inkp)$	$\mathcal{O}(np^2 \log(n) + np^2)$
GLPCA [14]	+	–	$\mathcal{O}(n^3)$	–	$\mathcal{O}(n(p \log(n) + n^2))$	$\mathcal{O}(Inkp)$	$\mathcal{O}(n^3)$
NMF [17]	–	–	$\mathcal{O}(Inpk)$	–	$\mathcal{O}(Inpk)$	$\mathcal{O}(Inkp)$	$\mathcal{O}(Inp(k+K))$
GNMF [6]	+	–	$\mathcal{O}(Inpk)$	–	$\mathcal{O}(np \log(n))$	$\mathcal{O}(Inkp)$	$\mathcal{O}(np \log(n))$
MMF [46]	+	–	$\mathcal{O}(((p+k)^2 + pk)I)$	–	$\mathcal{O}(np \log(n))$	$\mathcal{O}(Inkp)$	$\mathcal{O}(np \log(n))$
RPCA [7]	–	–	$\mathcal{O}(Imp^2)$	–	$\mathcal{O}(np^2 I + n \log(n))$	$\mathcal{O}(Inkp)$	$\mathcal{O}(np^2 I + n \log(n))$
RPCAG [35]	–	–	$\mathcal{O}(Imp^2)$	–	$\mathcal{O}(np^2 I + n \log(n))$	$\mathcal{O}(Inkp)$	$\mathcal{O}(np^2 I + n \log(n))$
FRPCAG [36]	+	+	$\mathcal{O}(Imp\mathcal{K})$	–	$\mathcal{O}(np \log(n))$	$\mathcal{O}(Inkp)$	$\mathcal{O}(np \log(n))$
CPCA	+	+	$\mathcal{O}(I\rho_c\rho_r\mathcal{K})$	$\mathcal{O}(\rho_c\rho_r^2)$	$\mathcal{O}(nkO_l\mathcal{K})$	$\mathcal{O}(I\rho_c\rho_r)$	$\mathcal{O}(nkO_l)$

Table XII: Details of the datasets used for clustering experiments in this work.

Dataset	Samples	Dimension	Classes
ORL large	400	56×46	40
ORL small	300	28×23	30
CMU PIE	1200	32×32	30
YALE	165	32×32	11
MNIST	70000	28×28	10
MNIST small	1000	28×28	10
USPS large	10000	16×16	10
USPS small	3500	16×16	10

Table XIII: Range of parameter values for each of the models considered in this work. k is the rank or dimension of subspace or the number of clusters, λ is the weight associated with the sparse term for Robust PCA framework [7] and γ, α are the parameters associated with the graph regularization term.

Model	Parameters	Parameter Range
LLE [32], PCA LE [4]	k	$k \in \{2^1, 2^2, \dots, \min(n, p)\}$
GLPCA [14]	k, γ	$k \in \{2^1, 2^2, \dots, \min(n, p)\}$ $\gamma \implies \beta$ using [14] $\beta \in \{0.1, 0.2, \dots, 0.9\}$
MMF [46]	k, γ	$k \in \{2^1, 2^2, \dots, \min(n, p)\}$ $\gamma \in \{2^{-3}, 2^{-2}, \dots, 2^{10}\}$
NMF [17]	k	
GNMF [6]	k, γ	
RPCA [7]	λ	$\lambda \in \left\{ \frac{2^{-3}}{\sqrt{\max(n, p)}} : 0.1 : \frac{2^3}{\sqrt{\max(n, p)}} \right\}$
RPCAG [35]	λ, γ	$\gamma \in \{2^{-3}, 2^{-2}, \dots, 2^{10}\}$
FRPCAG [36]	γ_r, γ_c	$\gamma_r, \gamma_c \in (0, 30)$
CPCA	γ_r, γ_c	$\gamma_r, \gamma_c \in (0, 30)$
	k (approximate decoder)	$\Sigma_{k,k} / \Sigma_{1,1} < 0.1$