Recovery of Subspace Structure from High-Rank Data with Missing Entries

João Carvalho, Manuel Marques, João P. Costeira

TÉCNICO LISBOA

Institute for Systems and Robotics Lisboa

joaopscarvalho

1. Problem

High-dimensional data lying in low-dimensional subspaces are very common in problems such as motion segmentation, structure from motion or face clustering. In practice, however, **observed data are incomplete** due to sensor failure, occlusion or data corruption.

We propose a method to **reconstruct and cluster incomplete high-dimensional data** lying in a union of K subspaces, $\{\mathcal{S}_k \subset \mathbb{R}^D\}_{k=1}^K$ with dimensions $\{d_k < D\}_{k=1}^K$.

We exploit the subspace structure of the data by using a *self-expressive* model:

$$\mathbf{X} = \mathbf{X}\mathbf{C}$$

Sparse coefficients, with $diag(\mathbf{C}) = 0$

Data: $[\mathbf{x}_1\, \mathbf{x}_2 \dots \mathbf{x}_N] \in \mathbb{R}^{D\times N}$

where

$$\mathbf{X} = \mathbf{X}_\Omega + \mathbf{X}_{\Omega^\complement}$$

Missing entries at the complementary positions of Ω and zeros otherwise

Known entries at $(i,j) \in \Omega$ and zeros otherwise

To recover the original subspaces, we estimate the missing entries, $\mathbf{X}_{\Omega^\complement}$, by imposing the subspace structure and then cluster the recovered data points. The missing data is estimated by solving the following **non-convex** optimization problem:

$$\min_{\mathbf{X}_{\Omega^\complement},\mathbf{C},\mathbf{E},\mathbf{Z}} \quad \|\mathbf{C}\|_1 + \lambda_e\|\mathbf{E}\|_1 + \frac{\lambda_z}{2}\|\mathbf{Z}\|_F^2 \qquad (1)$$

$$\text{s.t.} \quad \mathbf{X}_\Omega + \mathbf{X}_{\Omega^\complement} = \left(\mathbf{X}_\Omega + \mathbf{X}_{\Omega^\complement}\right)\mathbf{C} + \underbrace{\mathbf{E} + \mathbf{Z}}_{\text{Outlying entries and noise}}$$

$$diag(\mathbf{C}) = 0, \quad (\mathbf{X}_{\Omega^\complement})_\Omega = 0,$$

We propose a **tight convex relaxation** to solve this non-convex problem.

2. Accurate Subspace Segmentation by Successive Approximations

Consider we have a point $\mathbf{X}^{(i)}$ and $\mathbf{C}^{(i)}$. We define the following exact model

$$\mathbf{X}^{(i)} + \Delta\mathbf{X} = (\mathbf{X}^{(i)} + \Delta\mathbf{X})(\mathbf{C}^{(i)} + \Delta\mathbf{C}) + \mathbf{E} + \mathbf{Z}$$

where $\Delta\mathbf{X}_\Omega = 0$ and $diag(\Delta\mathbf{C}) = 0$.

To find $\Delta\mathbf{X}$ and $\Delta\mathbf{C}$, we consider the **linearization** of this model in a **trust region**. This approximate model leads to the following convex problem, similar (1) but with new constraints:

$$\min_{\Delta\mathbf{C},\Delta\mathbf{X},\mathbf{E},\mathbf{Z}} \quad \|\mathbf{C}^{(i)} + \Delta\mathbf{C}\|_1 + \lambda_e\|\mathbf{E}\|_1 + \frac{\lambda_z}{2}\|\mathbf{Z}\|_F^2$$

$$\text{s.t.} \quad \mathbf{X}^{(i)} + \Delta\mathbf{X} = (\mathbf{X}^{(i)} + \Delta\mathbf{X})\mathbf{C}^{(i)} + \mathbf{X}^{(i)}\Delta\mathbf{C} + \mathbf{E} + \mathbf{Z} \quad \text{Linearized model}$$

$$\|\Delta\mathbf{X}\|_\infty \leq \delta_X, \quad \|\Delta\mathbf{C}\|_\infty \leq \delta_C \quad \text{Trust region}$$

$$diag(\Delta\mathbf{C}) = 0$$

$$\Delta\mathbf{X}_\Omega = 0.$$

We compute a new solution to (1) by iteratively solving the previous problem and updating the current solution as

$$\mathbf{X}^{(i+1)} = \mathbf{X}^{(i)} + \Delta\mathbf{X}$$

$$\mathbf{C}^{(i+1)} = \mathbf{C}^{(i)} + \Delta\mathbf{C}$$

3. Robust Initialization for Matrix Completion [1]

Since (1) is biconvex, we use Alternate Convex Search [2]

1. Given $\mathbf{X}^{(i)}_{\Omega^\complement}$

$$\min_{\mathbf{C},\mathbf{E},\mathbf{Z}} \quad \|\mathbf{C}\|_1 + \lambda_e\|\mathbf{E}_\Omega\|_1 + \frac{\lambda_z}{2}\|\mathbf{Z}_\Omega\|_F^2$$

$$\text{s.t.} \quad \mathbf{X}_\Omega + \mathbf{X}^{(i)}_{\Omega^\complement} = (\mathbf{X}_\Omega + \mathbf{X}^{(i)}_{\Omega^\complement})\mathbf{C} + \mathbf{E} + \mathbf{Z}$$

$$diag(\mathbf{C}) = 0,$$

2. Given $\mathbf{C}^{(i+1)} = \mathbf{C}$

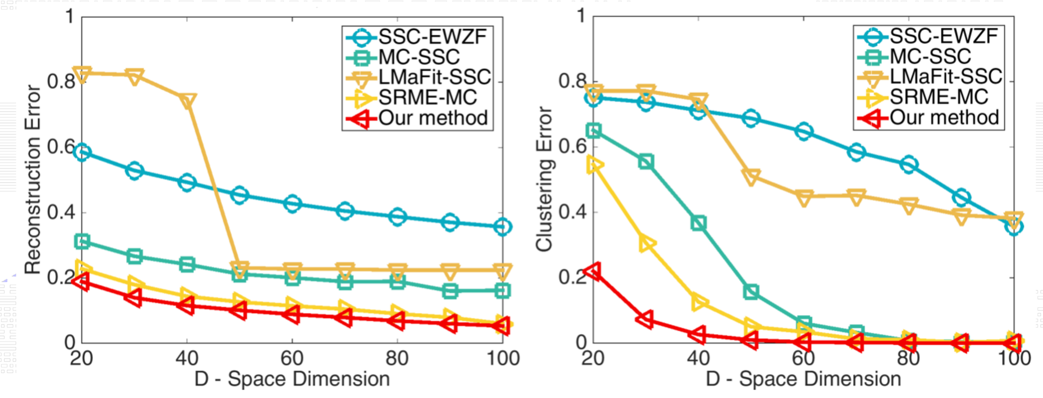
$$\mathbf{X}^{(i+1)}_{\Omega^\complement} = (\mathbf{X}^{(i)}\mathbf{C}^{(i+1)})_{\Omega^\complement} \quad \text{where} \quad \mathbf{X}^{(i)} = \mathbf{X}_\Omega + \mathbf{X}^{(i)}_{\Omega^\complement}$$

[1] J. Carvalho, M. Marques, and JP Costeira, "Subspace Segmentation by Successive Approximations: A Method for Low-Rank and High-Rank Data with Missing Entries," arXiv preprint arXiv:1709.01467, 2017.

[2] J. Gorski, F. Pfeuffer, and K. Klamroth, "Biconvex sets and optimization with biconvex functions: a survey and extensions," Mathematical Methods of Operations Research, vol. 66, no. 3, pp. 373–407, 2007.

4. Experiments

Synthetic Data



$\rho = 0.70$, $N_k = 50$, $K = 7$ and $d = 5$

ρ - Missing rate
D - Ambient space dimension
d - Subspace dimension
K - Number subspaces
N_k - Number points per subspace

Reconstruction Error

$$e_r = \frac{\|\hat{\mathbf{X}} - \mathbf{X}\|_F}{\|\mathbf{X}\|_F}$$

Clustering Error

$$e_c = \frac{\#\text{missclassified points}}{\#\text{points}}$$

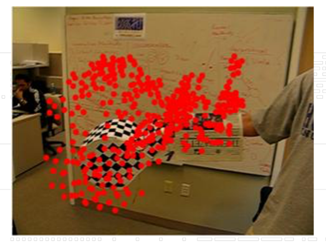
Hopkins 155

Table 5. Reconstruction error for all sequences in Hopkins 155, with 8 trials per sequence (one per missing rate).

ρ	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80
SSC-EWZF	0.070	0.101	0.133	0.183	0.253	0.351	0.481	0.654
LMaFit-SSC	0.072	0.077	0.088	0.101	0.106	0.121	0.125	0.179
SRME-MC	0.005	0.005	0.005	0.005	**0.006**	**0.010**	**0.022**	**0.077**
Our method	**0.001**	**0.002**	**0.003**	**0.004**	0.009	0.018	0.035	0.113

Table 6. Clustering error for all sequences in Hopkins 155.

ρ	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80
SSC-EWZF	0.180	0.204	0.226	0.245	0.257	0.275	0.296	0.318
LMaFit-SSC	0.228	0.216	0.207	0.304	0.315	0.318	0.325	0.332
SRME-MC	0.112	0.109	0.108	0.121	0.129	0.143	0.176	0.224
Our method	**0.073**	**0.070**	**0.071**	**0.068**	**0.071**	**0.080**	**0.095**	**0.129**



Skeleton Completion

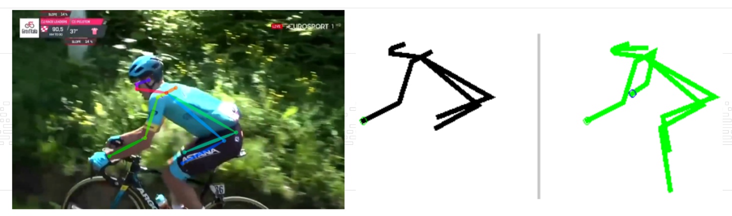


Table 8. Reconstruction error for the skeleton completion experiment (20 trials each).

ρ	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80
SSC-EWZF	0.088	0.127	0.194	0.270	0.359	0.467	0.610	0.751
SRME-MC	0.044	0.059	**0.085**	0.103	0.132	0.166	0.248	0.364
Our method	**0.040**	**0.055**	0.086	**0.100**	**0.116**	**0.133**	**0.175**	**0.240**

CMU Motion Capture

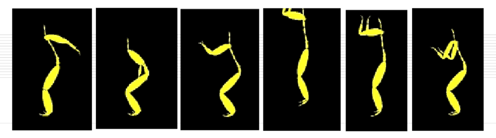


Table 7. Reconstruction error for the experiments (20 trials each) with the CMU Mocap dataset.

ρ	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80
SSC-EWZF	0.075	0.143	0.211	0.295	0.382	0.493	0.615	0.763
SRME-MC	0.019	0.031	0.047	0.070	0.100	0.160	0.232	0.384
Our method	**0.012**	**0.021**	**0.034**	**0.051**	**0.071**	**0.128**	**0.160**	**0.245**

Group Image Inpainting

COIL-20 (ρ = 0.70) Extended Yale B (ρ = 0.80)