

RECOVERY OF SUBSPACE STRUCTURE FROM HIGH-RANK DATA WITH MISSING ENTRIES

João Carvalho[†], Manuel Marques^{†,*}, João P. Costeira^{*}

Institute for Systems and Robotics, LARSyS, Instituto Superior Técnico, Univ. Lisboa, Portugal

ABSTRACT

We propose a method to reconstruct and cluster incomplete high-dimensional data lying in a union of low-dimensional subspaces. Exploring the sparse representation model, we jointly estimate the missing data while imposing the intrinsic subspace structure. Although we have a non-convex problem, we propose an algorithm robust to initialization. Extensive experiments with synthetic and real data show that our approach leads to significant improvements in the reconstruction and segmentation, outperforming current state of the art for both low and high-rank data.

Index Terms— Motion Segmentation, Subspace Clustering, Missing Data, Matrix Completion, Sparse Representation

1. INTRODUCTION

Linear Subspaces are one of the most powerful mathematical tools to represent and model high-dimensional data. Signal Processing and especially Computer Vision communities use these tools in a wide variety of algorithms in domains such as classification, structure-from-motion, object recognition and image-segmentation. However, in all the above areas, observations in real scenarios are incomplete due to self-occlusion, sensor failure or data corruption, to name a few.

In this paper, we address the problem of subspace clustering with missing data by simultaneously completing the data and enforcing the subspace structure. So, given a set of incomplete high-dimensional points drawn from a union of linear subspaces, we aim to estimate the unknown entries and segment the reconstructed data points.

The problem stated above boils down to answering two fundamental questions: 1) how to reconstruct the data from the observed entries and; 2) how to aggregate data such that each cluster lies on a linear subspace. Several state of the art methods address the second question and consider only observed data, leading to biased and skewed estimates of the subspaces [1, 2, 3, 4, 5, 6]. An alternative, is to estimate the missing data with low-rank matrix completion methods [7, 8, 9] and then segment the reconstructed data with subspace clustering

methods [10, 11]. This approach, however, fails to exploit the subspace structure while reconstructing the data, leading to suboptimal performance. Some developments attempt to tackle the reconstruction and aggregation problems jointly. Mixture models are proposed in [12, 5, 13], where the goal is to find the data (low-rank) projections (onto each subspace) that best fit the known entries of the data. In [14], the author proposed an extension to [10] where the missing data and the model coefficients are jointly estimated, while minimizing the nuclear norm to satisfy a hard-rank constraint. Also taking advantage of self-representation based matrix completion, [15] imposes sparse representation and regularizes the data with the nuclear norm. Finally, [16] and [17] propose factorization models and sparse representations to recover the subspaces.

Following the approach of self-representation based matrix completion, we propose an algorithm that reconstructs the data through successive approximations and recovers its subspace structure. We name it Accurate Subspace Segmentation by Successive Approximations (ASSSA). Relying on a *sparse representation* [18, 19], we achieved greater accuracy on the estimates of the missing data and significantly improved the segmentation. We propose a robust initialization that enables convergence. Extensive experiments prove that this strategy outperforms current state of the art in low and high-rank data.

2. RECOVERY OF SUBSPACE STRUCTURE

In this section, we propose a method to estimate the missing entries of a partially prescribed matrix and to segment the reconstructed data into clusters corresponding to the underlying subspaces.

Consider K subspaces $\{\mathcal{S}_k \subset \mathbb{R}^D\}_{k=1}^K$, with dimensions $\{d_k < D\}_{k=1}^K$, and let $\{\mathbf{x}_j \in \mathbb{R}^D\}_{j=1}^N$ be the set of N data points lying in the union of the K subspaces¹. Denote the data matrix as $\mathbf{X} = [\mathbf{x}_1 \mathbf{x}_2 \dots \mathbf{x}_N] \in \mathbb{R}^{D \times N}$, where each \mathbf{x}_j is a data point. We build upon the work Sparse Subspace Clustering [10], which considers the model $\mathbf{X} = \mathbf{X}\mathbf{C}$, where \mathbf{C} is the coefficients matrix with $\text{diag}(\mathbf{C}) = 0$. This model translates

This work was funded by: [†] FCT project [CMU/ECE/0005/2017]; *H2020 project [AI4EU, GA825619]. J. Carvalho is also supported by FCT through grant [PD/BD/114429/2016].

¹Bold capital letters, \mathbf{A} , represent matrices. Bold lower-case letters, \mathbf{a} , represent column vectors. Bold lower-case letters with subscript, \mathbf{a}_i , represent the i^{th} column of matrix \mathbf{A} . Scalars are denoted by non-bold letters, a or A . The scalar element in row i and column j of matrix \mathbf{A} is denoted by a non-bold lower-case letter with two subscripts, a_{ij} .

the *self-expressiveness* property, which exploits the fact that each data point in a union of subspaces can be represented as a linear or affine combination of other points. Since only a subset of the entries of \mathbf{X} is observed, we define

$$\mathbf{X} = \mathbf{X}_\Omega + \mathbf{X}_{\Omega^c}, \quad (1)$$

where \mathbf{X}_Ω is a matrix with the known entries at $(i, j) \in \Omega$, and zero otherwise. \mathbf{X}_{Ω^c} is the matrix with the missing entries in the complementary positions, $(i, j) \in \Omega^c$, and zero otherwise. To recover the original subspaces, we first estimate the unknown \mathbf{X}_{Ω^c} by imposing the subspace model and then cluster the recovered data points. The missing data is estimated by solving the following optimization problem

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

$$\text{s.t. } \mathbf{X}_\Omega + \mathbf{X}_{\Omega^c} = (\mathbf{X}_\Omega + \mathbf{X}_{\Omega^c}) \mathbf{C} + \mathbf{E} + \mathbf{Z} \\ \text{diag}(\mathbf{C}) = 0, \quad (\mathbf{X}_{\Omega^c})_\Omega = 0,$$

where the first constraint is the data model and \mathbf{E} and \mathbf{Z} account for outlying entries and noise, respectively. By minimizing the ℓ_1 -norm of the coefficients matrix, \mathbf{C} , we favor a *sparse representation* of the data, in which, ideally, a point is represented by a linear combination of few points from its own subspace [18, 19]. For complete data, this model is guaranteed to recover the desired representation when the subspaces are sufficiently separated and data points are well distributed inside the subspaces [10, 20, 21].

Problem (2) is non-convex because of the product between \mathbf{X}_{Ω^c} and \mathbf{C} , so, we propose a tight convex relaxation and a robust initialization to solve it.

2.1. Accurate Subspace Segmentation by Successive Approximations

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

$$\mathbf{X}^{(i)} + \Delta \mathbf{X} = (\mathbf{X}^{(i)} + \Delta \mathbf{X})(\mathbf{C}^{(i)} + \Delta \mathbf{C}) + \mathbf{E} + \mathbf{Z}, \quad (3)$$

where $\Delta \mathbf{X}$ is zero in the entries corresponding to the known entries, *i.e.*, $\Delta \mathbf{X}_\Omega = 0$. To find $\Delta \mathbf{X}$ and $\Delta \mathbf{C}$, we consider a linearization of (3), where discard the quadratic term $\Delta \mathbf{X} \Delta \mathbf{C}$ and impose bounds on the norm of $\Delta \mathbf{X}$ and $\Delta \mathbf{C}$. This approximate model leads to the following convex problem, similar to (2) but with new constraints,

$$\begin{aligned} \min_{\Delta \mathbf{C}, \Delta \mathbf{X}, \mathbf{E}, \mathbf{Z}} & \|\mathbf{C}^{(i)} + \Delta \mathbf{C}\|_1 + \lambda_e \|\mathbf{E}\|_1 + \frac{\lambda_z}{2} \|\mathbf{Z}\|_F^2 \quad (4) \\ \text{s.t. } & \mathbf{X}^{(i)} + \Delta \mathbf{X} = (\mathbf{X}^{(i)} + \Delta \mathbf{X}) \mathbf{C}^{(i)} \\ & \quad + \mathbf{X}^{(i)} \Delta \mathbf{C} + \mathbf{E} + \mathbf{Z} \\ & \|\Delta \mathbf{X}\|_\infty \leq \delta_X \\ & \|\Delta \mathbf{C}\|_\infty \leq \delta_C \\ & \text{diag}(\Delta \mathbf{C}) = 0 \\ & \Delta \mathbf{X}_\Omega = 0. \end{aligned}$$

Thus, we compute a new solution closer to the minimum of (2) by updating the current solution as

$$\mathbf{X}^{(i+1)} = \mathbf{X}^{(i)} + \Delta \mathbf{X} \quad (5)$$

$$\mathbf{C}^{(i+1)} = \mathbf{C}^{(i)} + \Delta \mathbf{C}. \quad (6)$$

In summary, given an initial guess for \mathbf{X} and \mathbf{C} , we improve over that solution with updates (5) and (6), following the directions computed in (4). These updates are repeated until a chosen stopping criterion is met. In the remainder of this section we refer to this as Accurate Algorithm.

Robust Initialization for Matrix Completion. Problem (4), jointly with (5) and (6), is a tight convex relaxation of (2). However, it requires an initialization, so, we use the robust algorithm we proposed in [22], and present it here.

Although non-convex, we note that (2) is a biconvex problem, *i.e.*, it is convex when one of the variables, \mathbf{X}_{Ω^c} or \mathbf{C} , is fixed. A common approach to solve such problems is to use Alternate Convex Search [23], successively fixing a subset of variables and computing the others in an alternate manner. We take a similar approach, therefore, given the current estimate $\mathbf{X}_{\Omega^c}^{(i)}$, we compute \mathbf{C} by solving

$$\begin{aligned} \min_{\mathbf{C}, \mathbf{E}, \mathbf{Z}} & \|\mathbf{C}\|_1 + \lambda_e \|\mathbf{E}\|_1 + \frac{\lambda_z}{2} \|\mathbf{Z}\|_F^2 \quad (7) \\ \text{s.t. } & \mathbf{X}_\Omega + \mathbf{X}_{\Omega^c}^{(i)} = (\mathbf{X}_\Omega + \mathbf{X}_{\Omega^c}^{(i)}) \mathbf{C} + \mathbf{E} + \mathbf{Z} \\ & \text{diag}(\mathbf{C}) = 0, \end{aligned}$$

where \mathbf{E}_Ω and \mathbf{Z}_Ω are the error matrices indexed by Ω . Note that solving (2) for a fixed \mathbf{X}_{Ω^c} leads to biased \mathbf{C} if \mathbf{X}_{Ω^c} is too distant from the original data, as \mathbf{C} would be “shaped” to minimize the error not only in the known entries but also in the missing entries (whose values are distant from the real ones). So, to reduce the bias induced in \mathbf{C} , we account for the error only in the known entries, as in [2].

Given $\mathbf{C}^{(i+1)} = \mathbf{C}$, we compute the missing entries with the current data reconstruction, *i.e.*,

$$\mathbf{X}_{\Omega^c}^{(i+1)} = \left(\mathbf{X}^{(i)} \mathbf{C}^{(i+1)} \right)_{\Omega^c}, \quad (8)$$

where $\mathbf{X}^{(i)} = \mathbf{X}_\Omega + \mathbf{X}_{\Omega^c}^{(i)}$.

We solve (7) and enforce (8) iteratively until a stopping criterion is reached. This approach does not provide convergence guarantees, however, it has succeeded in several estimation problems with missing data, such as [22, 24, 25]. In summary, we first run this Initialization Algorithm and use its solution \mathbf{X}_{Ω^c} , \mathbf{C} as starting point for the Accurate Algorithm, which computes the solution for (2)². After solving the completion problem, we build an affinity matrix from \mathbf{C} and apply spectral clustering to segment the data in K clusters.

²Both problems (4) and (7) can be efficiently solved with the ADMM algorithm [26]. For the first iteration of (7) we set $\mathbf{X}_{\Omega^c}^{(0)} = 0$.

Implementation Details. To make our algorithm independent of the λ 's, we propose the following. First, we run the Initialization Algorithm with a set of increasing λ 's and stop when the model error in the known entries stops decreasing, *i.e.*, $\|(\mathbf{X}^{(i)} - \mathbf{X}^{(i)}\mathbf{C}^{(i)})_{\Omega}\|_F$ stops decreasing. Then, using the λ 's that minimize this self-expressiveness error (residual), we compute a solution for (2) by running the Accurate Algorithm until the stopping condition is met. The intuition behind this procedure of starting with small λ 's and iteratively increasing them is the following: if in the first iterations the error in Ω is highly penalized (high λ 's), \mathbf{C} will get skewed in order to minimize the error given the current (distant) estimates of the missing data, significantly deteriorating subsequent updates of \mathbf{X}_{Ω^c} , (8). When the “plateau” is reached, increasing λ will not result in lower residual but only in less sparse \mathbf{C} (losing the subspace structure).

3. EXPERIMENTS

In this section, we assess the performance of our method and compare it to state of the art methods. We consider two metrics: clustering error and reconstruction error, defined as $e_c = \frac{\#\text{missclassified points}}{\#\text{points}}$ and $e_r = \frac{\|\hat{\mathbf{X}} - \mathbf{X}\|_F}{\|\mathbf{X}\|_F}$, where $\hat{\mathbf{X}}$ and \mathbf{X} are the estimated and true data matrices, respectively, and $\|\cdot\|_F$ the Frobenius norm. We compare with SSC-EWZF [2], EMSC [12], SRME-MC [15], and SVT [7] and LMaFit [8] followed by SSC (SVT-SSC and LMaFit-SSC)³.

Synthetic Data. We draw N_k points per subspace from a union of K random subspaces of dimension $d \ll D$ in \mathbb{R}^D . For simplicity, we draw the same number of points N_k per subspace $k \in \{1, \dots, K\}$ and assume all subspaces have dimension $d \leq N_k$. The missing data is generated by selecting uniformly at random, with probability ρ , the set of entries corresponding to missing values, Ω^c . We generate the data points in the range $[0, 1]$. For each experiment and percentage of missing data, we run 20 trials.

Tables 1 and 2 show the reconstruction and clustering error as a function of the percentage of missing data, ρ for low-rank data. Our method outperforms other methods recovering the original data, achieving the optimal solution for low ρ values. Note that only SRME-MC achieves similar performance. On high-rank data, Table 3 and 4, the results are similar but our method significantly outperforms SRME-MC on clustering with highly incomplete data.

Figure 1 shows the error versus the ambient space dimension D , going from high to low-rank, with $\rho = 0.70$. This experiment stresses the improvement achieved by our method over state of the art methods in high-rank cases and high missing rate, namely in clustering performance.

³SSC-EWZF does not provide completion, however, we evaluate its reconstruction with $\hat{\mathbf{X}} = \mathbf{X}_{\Omega^c}\mathbf{C}$, corresponding to the starting point (iteration 0) of our Initialization Algorithm. Also, note that for EMSC the clusters result from the mixture weights.

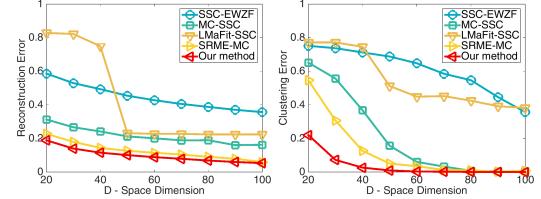


Fig. 1. Reconstruction error (left) and clustering error (right) as a function of the ambient space dimension, D , with $\rho = 0.70$, $N_k = 50$, $K = 7$ and $d = 5$.

Table 1. Reconstruction error as a function of missing rate, ρ , for the low-rank case, $K = 3$, $D = 50$, $d_k = 5$, $N_k = 50$.

ρ	0.15	0.30	0.50	0.70	0.80	0.90
EMSC	0.017	0.032	0.120	0.347	0.635	0.970
SSC-EWZF	0.098	0.135	0.232	0.430	0.597	0.816
LMaFit-SSC	0.088	0.082	0.172	0.586	0.744	0.881
SVT-SSC	0.088	0.079	0.079	0.161	0.244	0.422
SRME-MC	0.006	0.007	0.015	0.089	0.148	0.262
Our method	3.8×10^{-5}	1.5×10^{-4}	0.005	0.082	0.138	0.259

Table 2. Clustering error as a function of missing rate, ρ , for the low-rank case, $K = 3$, $D = 50$, $d_k = 5$, $N_k = 50$.

ρ	0.15	0.30	0.50	0.70	0.80	0.90
EMSC	0.062	0.132	0.263	0.488	0.573	0.598
SSC-EWZF	0.001	0.007	0.053	0.424	0.580	0.616
LMaFit-SSC	0	5×10^{-4}	0.040	0.555	0.603	0.606
SVT-SSC	0	0	0	0.011	0.200	0.570
SRME-MC	0	0	0	0.007	0.047	0.324
Our method	0	0	0	0.003	0.024	0.363

Table 3. Reconstruction error as a function of missing rate, ρ , for the high-rank, $K = 10$, $D = 80$, $d_k = 10$, $N_k = 50$.

ρ	0.15	0.30	0.50	0.70	0.80	0.90
EMSC	0.064	0.113	0.223	0.479	0.688	0.919
SSC-EWZF	0.093	0.121	0.205	0.384	0.550	0.781
SRME-MC	0.009	0.024	0.060	0.101	0.126	0.207
Our method	1×10^{-4}	0.001	0.062	0.087	0.111	0.191

Table 4. Clustering error as a function of missing rate, ρ , for the high-rank, $K = 10$, $D = 80$, $d_k = 10$, $N_k = 50$.

ρ	0.15	0.30	0.50	0.70	0.80	0.90
EMSC	0.289	0.402	0.639	0.812	0.827	0.832
SSC-EWZF	0.002	0.021	0.141	0.769	0.821	0.833
SRME-MC	0	0	7×10^{-4}	0.021	0.078	0.776
Our method	0	0	0.001	0.002	0.020	0.497

Motion Segmentation. We consider the problem of motion segmentation with Hopkins 155 dataset, where we aim to cluster trajectories of feature points belonging to multiple objects in a video sequence. Since this dataset contains only complete trajectories, we add missing data by selecting uniformly at random with probability ρ the set of entries corresponding to missing values.

Tables 5 and 6 show that for all levels of ρ , our method achieves the lowest clustering errors. These experiments also show that having lower reconstruction error does not necessarily mean having better clustering performance.

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

Motion Capture. We consider the motion capture scenario, where several sensors capture the motion of a human performing various activities. Here, we consider the completion problem for randomly selected sequences from the CMU Mocap dataset, where each point in the data matrix corresponds to the location of all the keypoints on that frame. To ensure high-rank data, we only use 1/3 of the keypoints, which are randomly selected. We generate missing data uniformly at random with probability ρ .

Table 7 shows the completion error for this experiment. As before, the improvement achieved by our method is specially significant for higher values of ρ .

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

Skeleton Completion. Pose estimation with skeleton detection is a relevant problem in several contexts [27], however, due to misdetections and occlusions, state of the art [27] fails to detect all keypoints (joints). Here, we consider the completion of skeletons detected on video sequences from cycling races, where cyclists have similar poses during the race development. Specifically, we consider a summary video from the 2018 Giro D’Italia, using [27] to detect skeletons in the sequence. Figure 2 shows a frame from the video and the corresponding incomplete and complete skeletons. We perform quantitative experiments by randomly sampling 600 skeletons from the video sequence, resulting in high-rank data, and artificially generate missing data.

Table 8 shows the results, where for high missing rate we significantly outperform SRME-MC for high missing rate.

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

Image Inpainting. We test our method in the group-image inpainting problem, with the COIL-20 object dataset [28] and



Fig. 2. Cycling race video frame with detected skeletons (left). Example of incomplete and recovered skeleton (right).

with the Extended Yale Face Database B [29]. We resize the images in the COIL dataset to 32×32 pixels and randomly select a subset of 500 from the total of 1440 images. For the Yale dataset, we selected the first 20 images from 5 randomly selected subjects. As before, we generate the missing data uniformly at random.

Figure 3 shows the results for LMaFit [8] and our method for the COIL dataset with 70% of missing data. Our method (fourth column) captures the details of the objects (e.g., second and third rows), while LMaFit (third column) fails to correctly capture the object shape. Similarly, with 80% of missing on the Yale dataset, LMaFit fails to capture the shape of the mouth (e.g., first and last examples), the details in the eyes (second face) or even the face expression (third and fourth faces).

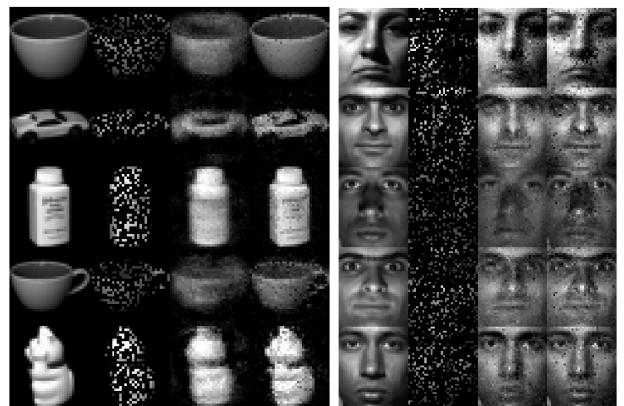


Fig. 3. Group-image inpainting problem. Columns: original image, image with missing data, LMaFit completion, our method completion. COIL-20 with 70% missing (left). Extended Yale B with 80% missing (right).

4. CONCLUSIONS

We proposed Accurate Subspace Segmentation by Successive Approximations, a method for subspace segmentation with incomplete data lying in a union of affine subspaces. It recovers missing entries by exploiting the sparse representation of the data (for both observed and unobserved entries), constraining it to lie in the respective subspace. Using a two stage approach, we significantly improve the reconstruction and clustering performance. Our algorithm is robust to initialization and does not require parameter tuning. Extensive synthetic and real data experiments showed that our method outperforms current state of the art methods for both low and high-rank problems.

5. REFERENCES

- [1] R. Heckel and H. Bölcskei, “Robust subspace clustering via thresholding,” *IEEE Transactions on Information Theory*, vol. 61, no. 11, pp. 6320–6342, 2015.
- [2] C. Yang, D. Robinson, and R. Vidal, “Sparse subspace clustering with missing entries,” in *Proceedings ICML*, 2015, pp. 2463–2472.
- [3] L. Balzano, A. Szelam, B. Recht, and R. Nowak, “K-subspaces with missing data,” in *SSP Workshop*. IEEE, 2012, pp. 612–615.
- [4] B. Eriksson, L. Balzano, and R. D. Nowak, “High-rank matrix completion..,” in *AISTATS*, 2012, pp. 373–381.
- [5] D Pimentel-Alarcón, L Balzano, R Marcia, R Nowak, and R Willett, “Group-sparse subspace clustering with missing data,” in *SSP Workshop*. IEEE, 2016, pp. 1–5.
- [6] Renli Liang, Yanqin Bai, and Hai Xiang Lin, “An inexact splitting method for the subspace segmentation from incomplete and noisy observations,” *Journal of Global Optimization*, pp. 1–19, 2018.
- [7] J. Cai, E J Candès, and Z Shen, “A singular value thresholding algorithm for matrix completion,” *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [8] Z. Wen, W. Yin, and Y. Zhang, “Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm,” *Mathematical Programming Computation*, pp. 1–29, 2012.
- [9] Z. Lin, R. Liu, and Z. Su, “Linearized alternating direction method with adaptive penalty for low-rank representation,” in *NIPS*, 2011, pp. 612–620.
- [10] E. Elhamifar and R. Vidal, “Sparse subspace clustering: Algorithm, theory, and applications,” *TPAMI*, vol. 35, no. 11, pp. 2765–2781, 2013.
- [11] R. Vidal and P. Favaro, “Low rank subspace clustering (lrsc),” *Patt. Rec. Letters*, vol. 43, pp. 47–61, 2014.
- [12] D. Pimentel, R Nowak, and L. Balzano, “On the sample complexity of subspace clustering with missing data,” in *SSP, Workshop on*. IEEE, 2014, pp. 280–283.
- [13] D Pimentel-Alarcón, “Mixture matrix completion,” in *NIPS*, 2018, pp. 2197–2207.
- [14] E. Elhamifar, “High-rank matrix completion and clustering under self-expressive models,” in *NIPS*, 2016, pp. 73–81.
- [15] J Fan and TWS Chow, “Sparse subspace clustering for data with missing entries and high-rank matrix completion,” *Neural Networks*, vol. 93, pp. 36–44, 2017.
- [16] J Fan, M Zhao, and TWS Chow, “Matrix completion via sparse factorization solved by accelerated proximal alternating linearized minimization,” *IEEE Transactions on Big Data*, 2018.
- [17] M Kwon, H Kim, and H Choi, “Improving low-rank matrix completion with self-expressiveness,” in *ACM ICIKM*. ACM, 2018, pp. 1651–1654.
- [18] EJ Candes, JK Romberg, and T Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Communications on pure and applied mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [19] EJ Candes, “The restricted isometry property and its implications for compressed sensing,” *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589–592, 2008.
- [20] M Soltanolkotabi, EJ Candes, et al., “A geometric analysis of subspace clustering with outliers,” *The Annals of Statistics*, vol. 40, no. 4, pp. 2195–2238, 2012.
- [21] M. Soltanolkotabi, E. Elhamifar, E. J. Candes, et al., “Robust subspace clustering,” *The Annals of Statistics*, vol. 42, no. 2, pp. 669–699, 2014.
- [22] 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.
- [23] 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.
- [24] R Guerreiro and P Aguiar, “Estimation of rank deficient matrices from partial observations: two-step iterative algorithms,” in *International Workshop on Energy Minimization Methods in CVPR*. Springer, 2003, pp. 450–466.
- [25] Manuel Marques and João Costeira, “Estimating 3d shape from degenerate sequences with missing data,” *CVIU*, vol. 113, no. 2, pp. 261–272, 2009.
- [26] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends in ML*, vol. 3, no. 1, pp. 1–122, 2011.
- [27] Z. Cao, T. Simon, S. Wei, and Y. Sheikh, “Realtime multi-person 2d pose estimation using part affinity fields,” in *IEEE CVPR*, 2017, pp. 7291–7299.
- [28] SA Nene, SK Nayar, H Murase, et al., “Columbia object image library (coil-20),” 1996.
- [29] K.C. Lee, J. Ho, and D. Kriegman, “Acquiring linear subspaces for face recognition under variable lighting,” *IEEE TPAMI*, vol. 27, no. 5, pp. 684–698, 2005.