# Boosting Spectral Clustering on Incomplete Data via Kernel Correction and Affinity Learning

**Fangchen Yu**[1], Runze Zhao[1], Zhan Shi[1], Yiwen Lu[1]
Jicong Fan[1,2], Yicheng Zeng[2], Jianfeng Mao[1,2], Wenye Li[1,2](✉)

[1]The Chinese University of Hong Kong, Shenzhen, China
[2]Shenzhen Research Institute of Big Data, Shenzhen, China

37th Conference on Neural Information Processing Systems

December 10-16, 2023

# Outline

# 1. Introduction: Spectral Clustering on Incomplete Data

- **Spectral Clustering**'s performance highly relies on the quality of the similarity matrix, such as the kernel matrix and affinity matrix.

  **Step 1. Construct A Similarity Matrix: $S = K$ or $C$**
  - Calculate a (Gaussian) kernel $K \in \mathbb{R}^{n \times n}$ from $X \in \mathbb{R}^{d \times n}$.
  - Learn a self-expressive affinity matrix $C \in \mathbb{R}^{n \times n}$ from $K$.

  **Step 2. Perform Normalized Cut Algorithm [1]**
- **Incomplete Data** with missing values is commonly seen in real-life.

**Research Question** is how to estimate a high-quality kernel or affinity matrix for incomplete data, benefiting spectral clustering tasks.

# 2. Related Work

Traditional methods to deal with incomplete data:

- **Data Imputation**: to impute missing values with estimated values.
  - Statistical imputation: zero, mean, $k$NN, regression, ...
  - Matrix completion: e.g., min $\|UV^\top - X\|$
- **Distance Calibration**: to calibrate a non-metric distance to a metric.
- **Limitation**: no guarantee on the quality of the kernel matrix.

**Research Goal** is to propose new imputation-free algorithms based on properties of kernel matrices with a theoretical guarantee.

# 3.1 Kernel Correction Algorithm

**Motivation:**

- A valid kernel is a symmetric matrix being *positive semi-definite* [2].

**Formulation:**

- Estimation: estimate a naive kernel $K^0$ on incomplete data [3].
- Optimization: $\min_{K \in \mathbb{R}^{n \times n}} \|K - K^0\|_F^2 \ s.t.\ K \succeq 0,\ k_{ij} = k_{ji} \in [0,1],\ \forall\ i,j$
- Obtain the solution $\hat{K}$ by Dykstra's projection algorithm [4].

**Guarantee:**

- $\|K^* - \hat{K}\|_F \leq \|K^* - K^0\|_F$, where $K^*$ is the unknown ground-truth.

# 3.2 Kernel Self-expressive Learning Algorithms

- **K**ernel **S**elf-expressive **L**earning with **S**chatten **p**-norm (**KSL-Sp**)
  - Schatten $p$-norm: $\|C\|_{S_p} := (\sum_{i=1}^{n} \sigma_i^p(C))^{1/p}$ with flexible sparsity.

$$\min_{C \in \mathbb{R}^{n \times n}} \ \|\phi(X) - \phi(X)C\|_F^2 + \lambda \|C\|_{S_p}, \ s.t.\ 0 \leq c_{ij} \leq 1, \ \forall 1 \leq i, j \leq n.$$

- **A**daptive **K**ernel **L**east-**S**quares **R**epresentation (**AKLSR**)
  - Learn kernel and affinity matrices iteratively via KLSR and KC:

$$\min_{K \succeq 0, \ C} \ \|K - K^0\|_F^2 + \mathrm{Tr}(K - 2KC + C^\top KC) + \lambda \|C\|_F^2.$$

  - Solve it by ADMM [5] based on the augmented Lagrange function.

# 4.1 Performance on Kernel Estimation

**How to estimate a kernel matrix for incomplete data?**

- Data imputation: $X^0 \xrightarrow{\text{impute}} \hat{X} \to \hat{K}$
- Distance calibration: $X^0 \to D^0 \xrightarrow{\text{calibrate}} \hat{D} \to \hat{K}$
- Kernel correction: $X^0 \to K^0 \xrightarrow{\text{correct}} \hat{K}$

**Evaluation metrics:**

- Relative error (**RE** $= \frac{\|\hat{K}-K^*\|_F}{\|K^*\|_F}$): measures the accuracy of estimation.
- **Recall**: measures the accuracy of top-10 nearest neighbors.

Table 1: Gaussian kernel estimation on the Yale64 dataset under a missing ratio 80%.

| Metric | Naive | ZERO | MEAN | *k*NN | EM | SVT | GR | KFMC | DC | TRF | EE | KC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RE↓ | 0.113 | 0.382 | 0.195 | 0.381 | 0.195 | 0.380 | 0.376 | 0.335 | 0.180 | 0.112 | 0.097 | **0.089** |
| Recall↑ | 0.721 | 0.063 | 0.275 | 0.063 | 0.275 | 0.066 | 0.070 | 0.183 | 0.571 | 0.722 | 0.751 | **0.767** |

Our KC method achieves the smallest estimation error and the highest Recall with improved local relationships, which in turn benefits spectral clustering.

# 4.2 Performance on Standard Spectral Clustering

- Values of $X^0$ are missing completely at random.
- Use the estimated Gaussian kernel as the input of normalized cut algorithm.
- Accuracy (ACC), Normalized Mutual Information (NMI), Adjusted Rand Index (ARI)

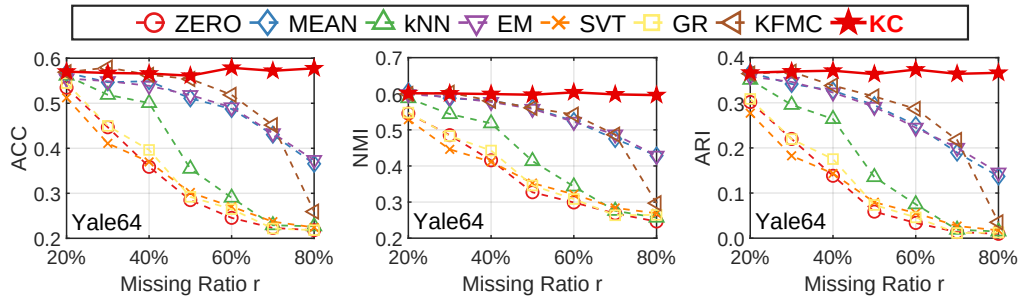


Figure 1: Robustness analysis of standard spectral clustering on the Yale64 dataset.

Our KC method maintains consistent and superior performance even with large missingness.

# 4.3 Performance on Self-expressive Affinity Learning

Use the learned affinity matrix as the input of normalized cut algorithm.

- KSSC [6]: $\min_C \|\phi(X) - \phi(X)C\|_F^2 + \lambda\|C\|_1$
- KLSR [7]: $\min_C \|\phi(X) - \phi(X)C\|_F^2 + \lambda\|C\|_F^2$
- **KSL-Sp**: $\min_C \|\phi(X) - \phi(X)C\|_F^2 + \lambda\|C\|_{S_p}$
- **AKLSR**: $\min_{K\succeq 0, C} \|K - K^0\|_F^2 + \mathrm{Tr}(K - 2KC + C^\top KC) + \lambda\|C\|_F^2$

Table 2: NMI performance of self-expressive affinity on Yale64 dataset under a missing ratio 80%.

| Method | Naive | ZERO | MEAN | $k$NN | EM | SVT | GR | KFMC | DC | TRF | EE | KC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| KSSC | 0.219 | 0.215 | 0.167 | 0.173 | 0.177 | 0.218 | 0.208 | 0.259 | 0.588 | 0.210 | 0.209 | **0.616** |
| KLSR | 0.606 | 0.311 | 0.604 | 0.320 | 0.609 | 0.321 | 0.327 | 0.318 | 0.597 | 0.603 | 0.604 | **0.616** |
| KSL-Sp | 0.370 | 0.315 | 0.581 | 0.303 | 0.579 | 0.305 | 0.304 | 0.295 | 0.555 | 0.364 | 0.599 | **0.619** |
| AKLSR | 0.452 | 0.327 | 0.606 | 0.338 | 0.605 | 0.308 | 0.338 | 0.312 | 0.570 | 0.464 | 0.575 | **0.614** |

Both KSL-Sp and AKLSR algorithms employ corrected kernels to yield dependable affinity matrices, thereby elevating spectral clustering performance.

# 5. Conclusion

**How to boost spectral clustering on incomplete data?**

- Proposed an imputation-free framework:
- Learned a high-quality *kernel* matrix by the kernel correction algorithm.
- Learned a high-quality *affinity* matrix by the kernel self-expressive affinity learning algorithms.
- Experiments show the effectiveness of kernel correction method, compared to existing data imputation and distance calibration approaches.

# More Information

- Codes are available at `https://github.com/SciYu/Spectral-Clustering-on-Incomplete-Data`.
- Welcome to visit the homepage and explore potential collaborations in the future.

Paper Github Video Poster Homepage

# References


[1] Andrew Ng, Michael Jordan, and Yair Weiss. "On spectral clustering: Analysis and an algorithm". In: *Advances in Neural Information Processing Systems* 14 (2001).

[2] Rafic Nader et al. "On the positive semi-definite property of similarity matrices". In: *Theoretical Computer Science* 755 (2019), pp. 13–28.

[3] Fangchen Yu et al. "Online estimation of similarity matrices with incomplete data". In: *Uncertainty in Artificial Intelligence*. PMLR. 2023, pp. 2454–2464.

[4] Heinz H Bauschke and Jonathan M Borwein. "Dykstra's alternating projection algorithm for two sets". In: *Journal of Approximation Theory* 79.3 (1994), pp. 418–443.

[5] Stephen Boyd et al. "Distributed optimization and statistical learning via the alternating direction method of multipliers". In: *Foundations and Trends® in Machine learning* 3.1 (2011), pp. 1–122.

[6] Vishal M Patel and René Vidal. "Kernel sparse subspace clustering". In: *2014 IEEE International Conference on Image Processing*. IEEE. 2014, pp. 2849–2853.

[7] Jicong Fan et al. "A simple approach to automated spectral clustering". In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 9907–9921.