TV-DCT: METHOD TO IMPUTE GENE EXPRESSION DATA USING DCT BASED SPARSITY AND TOTAL VARIATION DENOISING

*Akanksha Farswan** *Anubha Gupta*

Signal Processing and Biomedical Imaging Lab (SBILab), Department of ECE, IIIT-Delhi, India.

ABSTRACT

Most of the bioinformatics tools used in the analysis of gene expression data require complete data matrices. Missing values in data can adversely influence the downstream analysis for diagnostics and treatment. Several methods to impute missing values in gene data have been developed. However, most of these work at high levels of observability. In this paper, we have proposed a novel 2-stage method, namely, TV-DCT for imputing incomplete gene expression matrices using Total Variation denoising and Discrete Cosine Transform Domain Sparsity (TV-DCT) that achieves smaller imputation errors, consistently, at all levels of observability. The proposed method has been compared with three state-of-the-art matrix completion methods on three different cancer datasets and is observed to perform better. The validation of imputed data has been demonstrated on the application of classification.

Index Terms— Gene expression data, matrix imputation, sparse recovery, machine learning, cancer treatment

1. INTRODUCTION

Microarray technology facilitates estimation of expression levels of thousands of genes simultaneously under different experimental conditions. Gene expression data generated from such experiments is subsequently analyzed using statistical or machine learning methods to extract relevant information for disease diagnostics and treatment, particularly, in cancer. However, gene expression data suffers from the problem of missing values that leads to inaccurate analysis. Missing values often occur due to various reasons, such as insufficient resolution, image corruption, dust or scratches on the slide [1]. A simple solution is to repeat the experiment. However, this is expensive as well as time consuming. Therefore, it is pertinent to recover missing values of microarray data by developing computational and statistical methods.

Some of the early methods that were implemented to deal with incomplete data include ZEROimpute (replacing missing entries with zeros), ROWimpute and COLimpute (replacing them with the averaged values of the observed entries of the corresponding rows or columns) [2]. However, their performance is sub-optimal because they do not account for the correlation among genes. There exists two main classes of methods depending on the manner in which correlation among genes is exploited, namely, local and global approaches [3]. k nearest-neighbor imputation (KNNimpute) [4], GMCimpute [5], SLSSimpute [6] are few examples of local approaches which utilize local correlation among the genes and perform optimally when the data is heterogeneous. Global approach based methods such as SVDimpute [4], Bayesian Principal Component Analysis (BPCA) [7] exploit the global covariance information resulting from the entire gene expression matrix and do not perform optimally when data is heterogenous. Hybrid methods such as HPM-MI [8], GA+SVR [9], MIGEC [10] also exist in literature that offer better performance irrespective of the type of correlation present in the data. However, most of these methods estimate gene expression values at high observability of data, e.g., when 70% or more data is available and 30% or less is missing. In recent times, researchers are predicting expression data values with very less amount of observed data that is as low as 10%. It is universally known that in any biological process, group of genes act together; thereby, contributing to the interdependence between the expression levels of genes. This interdependence leads to a highly correlated data matrix (of subjects versus genes). Therefore, gene expression matrix can be thought of as a low rank matrix that can be embedded into a lower dimensional subspace. Thus, the problem of imputing missing values of gene expression data can be considered as a matrix completion problem.

Today, matrix completion is an active research problem in various applications, say in recommender systems. Of the developed methods, LMaFit [11] (based on matrix factorization), LogDet [12] (implementing nuclear norm minimization), and Robust PCA (RPCA) [13] (robust to outliers and implements feature reduction) can be stated as three different types of state-of-the-art methods on matrix completion. These advanced methods are still not popular in genomics research, though one low rank constrained matrix completion method has been utilized recently in genomics [14].

In this paper, we propose a novel 2-stage method, namely, TV-DCT method for predicting missing values in gene expression data with the following salient features:

*Thanks to UGC, Govt. of India for UGC-Junior Research Fellowship. We thank Dr. Sriram K. IIIT Delhi for his valuable suggestions. We gratefully acknowledge the research funding support (Grant: EMR/2016/006183 and DST/ICPS/CPS-Individual/2018/279(G)) from the Department of Science and Technology, Govt. of India for this research work.

1. In the first stage, missing value recovery problem is formulated as compressive sensing based reconstruction with sparsity in the Discrete Cosine Transform (DCT). Such an approach has been recently used in [15] as a proof of concept to illustrate that DCT acts as approximate Karhunen-Loève transform for a large class of signals. Recently, this formulation has also been tried for data recovery in wireless sensor networks [16].
2. The second stage is formulated as the denoising problem assuming that the first stage recovery would need further improvement. This is carried out with total variation constraint applied on the data of each gene across patients, assuming that, in general, expression values of a particular gene across subjects will not vary much.
3. Missing value imputation is shown on three cancer dataset at low as well as high observability of data.
4. The performance of the proposed method is validated in the application of classification.

The performance of the proposed method is observed to be better compared to the existing state-of-the-art matrix completion methods (devised in other applications as well).

2. DATASET DESCRIPTION

Three publicly available microarray gene expression datasets: $ALLAML$ [17, 18], $lung$ [17, 19] and $Myeloma$ [20] have been used with details provided in Table-1. Dataset $ALLAML$ contains expression values of 7129 genes across 72 individuals. This data consists of two classes depending on whether an individual suffers from Acute Lymphocytic Leukemia ALL or Acute Myeloid Leukemia AML. Label '1' belongs to ALL and label '2' to AML. $lung$ dataset contains expression values of 3312 gene across 203 individuals and has five classes. Label '1' corresponds to lung adenocarcinomas, label '2' belongs to normal healthy individual, label '3' belongs to squamous-cell lung carcinoma, and label '4' represents pulmonary carcinoids, and label '5' belongs to small-cell lung carcinoma. $Myeloma$ dataset contains expression values of 33297 gene across 99 individuals and has four classes. Label '1' belongs to MGUS (precursor stage of Multiple Myeloma), label '2' belongs to Multiple Myeloma (MM), label '3' belongs to Smouldering Multiple Myeloma (SMM), and label '4' to healthy individuals.

Table 1: Dataset Description

Dataset	#Subjects	#Genes	#Classes
$ALLAML$	72	7129	2
$lung$	203	3312	5
$Myeloma$	99	33297	4

3. PROPOSED TV-DCT METHOD

The proposed TV-DCT method for completing the gene expression matrix is a 2-stage method. Stage-1 is the compressive sensing based framework used for matrix completion, while stage-2 is a denoising framework for the extraction of denoised data from the matrix recovered in stage-1.

Stage-1: Compressive Sensing based matrix Completion: First, we formulate the problem of matrix completion as CS-based reconstruction. To this end, we consider the incomplete matrix $\mathbf{Y}$ of size $m \times n$, where m denotes the number of subjects and n denotes the number of genes. Assuming that the expression of any gene for all subjects will be similar, data within a column would exhibit sparsity in some transform domain. Hence, we propose to recover missing data column-wise, i.e., by applying CS framework on each column of the matrix $\mathbf{Y}$. The sensing matrix Φ_i of size $r_i \times m$ is constructed for every i^{th} column, where r_i denotes the number of available data entries of that column. Corresponding to each observed entry (that is not missing) of the i^{th} column, there is a row in Φ_i with an entry '1' for the corresponding position and zeros in the rest of the positions. For example, assume $\mathbf{y} = \begin{bmatrix} x_1 & x_3 & x_6 \end{bmatrix}^T$ is the observed vector where only x_1, x_3 and x_6 are available and, x_2, x_4 and x_5 are missing. Then, the vector $\mathbf{y}$ can be written as $\mathbf{y} = \Phi\mathbf{x}$, where the sensing matrix is written as $\Phi = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$ and $\mathbf{x}$ is the desired vector to be recovered. Thus, missing values have been interpreted as if those values were not sensed. This formulation converts the missing data recovery problem to CS-based data reconstruction.

In order to choose the sparsifying transform, we studied the columns of the gene expression matrix in the DCT domain and observe that these are highly sparse. This appears intuitively correct because, as stated earlier, every column represents the expression values of a particular gene across subjects. Biologically, these expression values would be similar and hence, data within any column would be slowly varying. Since DCT acts as a KL-type basis for slow-varying signals [15], data exhibits sparsity in the DCT domain. With this information, we recover each column of matrix $\mathbf{Y}$ using the CS-based reconstruction with the sparsity constraint on the columns in the DCT domain. The following optimization problem is solved to recover the i^{th} column of matrix $\mathbf{Y}$

$$\underset{\tilde{\mathbf{x}}_i}{min}(||\mathbf{y}_i - \Phi_i\tilde{\mathbf{x}}_i||_2^2 + \lambda_1||\mathbf{D}\tilde{\mathbf{x}_i}||_1), \tag{1}$$

where $\mathbf{y}_i$ contains the observed entries of the i^{th} column of matrix $\mathbf{Y}$, Φ_i is the corresponding sensing matrix for the i^{th} column, and $\tilde{\mathbf{x}}_i$ is the corresponding recovered column. The above problem is also called as analysis-prior formulation and is non-separable because of the presence of DCT matrix $\mathbf{D}$ with $\tilde{\mathbf{x}}_i$. Since DCT is an orthogonal transform, we can easily transform it to synthesis prior formulation as

$$\underset{\tilde{\mathbf{z}}_i}{min}(||\mathbf{y}_i - \Phi_i\mathbf{D}^T\mathbf{z}_i||_2^2 + \lambda_1||\mathbf{z}_i||_1), \tag{2}$$

where $\mathbf{D}\tilde{\mathbf{x}}_i = \mathbf{z}_i$. The above problem is now separable and can be easily solved using the iterative soft thresholding algorithm (ISTA) [21]. The update rule for $\mathbf{z}_i$ is

$$\mathbf{z}_i^{k+1} = soft\left\{\mathbf{z}_i^k + \frac{1}{\alpha}(\mathbf{D}\Phi_i^T)(\mathbf{y}_i - \Phi_i\mathbf{D}^T\mathbf{z}_i^k), \frac{\lambda_1}{2\alpha}\right\}, \tag{3}$$

$$\tilde{\mathbf{x}}_i = \mathbf{D}^T\mathbf{z}_i, \tag{4}$$

$$\widetilde{\mathbf{X}} = [\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, ...\tilde{\mathbf{x}}_n \tag{5}$$

where $\tilde{\mathbf{X}}$ is the recovered complete matrix and '$soft$' denotes the soft thresholding operator. The above optimization problem is solved using 'spgl' solver that optimally chooses the regularization parameter λ_1 [22] [23].

Stage-2: TV Denoising: Matrix $\tilde{\mathbf{X}}$ recovered from stage-1 is assumed to be noisy and hence, total variation (TV) based denoising is used in the second stage of the proposed algorithm. It is often used in image processing applications to reduce noise in the image and simultaneously preserve its edges [24]. We use TV for noise removal on individual gene's data across subjects assuming that the expression values of a particular gene will vary slowly over different subjects. Total variation filtering algorithm presented in [25] is used in TV-DCT which is formulated as

$$\min_{\mathbf{x}_i}(||\mathbf{x}_i - \tilde{\mathbf{x}}_i||_2^2 + \lambda_2||\mathbf{A}\mathbf{x}_i||_1), \tag{6}$$

where i ranges from 1 to n (number of columns/ genes). $\mathbf{A}$ is a difference operator defined as $\mathbf{A} = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & & \ddots & \\ & & & -1 & 1 \end{bmatrix}$ and it maps a vector $\mathbf{x}_i$ to

$$(\mathbf{A}\mathbf{x}_i^k) = \mathbf{x}_i^k - \mathbf{x}_i^{k+1}. \tag{7}$$

Dual formulation of above is used to solve the optimization framework (because of non-differentiability of l_1-norm) as

$$\min_{\mathbf{x}_i} \max_{|\mathbf{w}_i|\leq 1} (||\mathbf{x}_i - \tilde{\mathbf{x}}_i||_2^2 + \lambda_2 \mathbf{w}_i^T \mathbf{A}\mathbf{x}_i), \tag{8}$$

where $\mathbf{w}_i$ is an auxiliary vector such that

$$||\mathbf{x}_i||_1 = \max_{|\mathbf{w}_i|\leq 1} (\mathbf{w}_i^T \mathbf{x}_i). \tag{9}$$

TV denoising problem is minimized using iterative clipping algorithm with update equations as

$$\mathbf{x}_i^{k+1} = \tilde{\mathbf{x}}_i - \mathbf{A}^T \mathbf{w}_i^k, \tag{10}$$

$$\mathbf{w}_i^{k+1} = clip\left\{\mathbf{w}_i^k + (\frac{1}{\alpha})\mathbf{A}\mathbf{x}_i^{k+1}, \frac{\lambda_2}{2}\right\}, \tag{11}$$

for $k \geq 0$ with $\mathbf{w}^{(0)} = \mathbf{0}$ and $\alpha \geq \text{maxeig}(\mathbf{A}\mathbf{A}^T)$.

The regularization parameter λ_2 controls how much smoothing is performed and is determined empirically using grid search. It was set to value 1.2 for $ALLAML$ dataset, 0.3 for $lung$ dataset and 0.3 for $Myeloma$ dataset in our experiments. After denoising, the recovered matrix is organized as

$$\mathbf{x}_{j,i} = \hat{\mathbf{x}}_{j,i}, \;\; if \;\; \Omega_{j,i} = 1, \tag{12}$$

where $\Omega_{j,i} = 1$ if the entry at $(j,i)^{\text{th}}$ position is observed in the incomplete matrix and $\Omega_{j,i} = 0$ if the entry is missing. $\hat{\mathbf{x}}_{j,i}$ are the entries in the original matrix $\hat{\mathbf{X}}$.

4. RESULTS

4.1. Experiments and Evaluation

Simulations are carried out by randomly introducing missing values in the complete gene expression data matrix $\mathbf{X}$ with missing rates starting from 20% to 90%. Missing values are estimated using the proposed imputation method and some existing conventional methods. Imputed values are then compared to the original values for evaluating the performance of the method. Normalized root mean squared error (NRMSE) is used as the evaluation metric and is defined as

$$\text{NRMSE} = \frac{||\hat{\mathbf{X}}(original) - \mathbf{X}(recovered)||_F}{||\hat{\mathbf{X}}(original)||_F} \tag{13}$$

Algorithm 1: Proposed TV-DCT Method

Stage 1 - Matrix Recovery
Input: $\mathbf{Y}$ (Input incomplete matrix), DCT matrix $\mathbf{D}$
for loop from $i = 1.....n$
Calculate Φ_i for all i using $\mathbf{y}_i$
Update $\mathbf{z}_i$ using (3)
Calculate $\tilde{\mathbf{x}}_i = \mathbf{D}^T \mathbf{z}_i$
$end\ for$
Obtain $\tilde{\mathbf{X}}$ from $\tilde{\mathbf{x}}_i$
Output: $\tilde{\mathbf{X}}$ (Recovered Matrix from Stage-1)
Stage 2 - Denoising
Input: $\tilde{\mathbf{X}}$(Noisy matrix), $\mathbf{A}$ (Difference Operator)
for loop from $i = 1.....n$
$while$ converge:
Update $\mathbf{x}_i$ using (10)
Update $\mathbf{w}_i$ using (11)
$end\ while$
$end\ for$
Obtain $\mathbf{X}$ from $\mathbf{x}_i$
Replace the already observed entries in $\mathbf{X}$ using (12)
Output: $\mathbf{X}$ (Recovered Matrix)

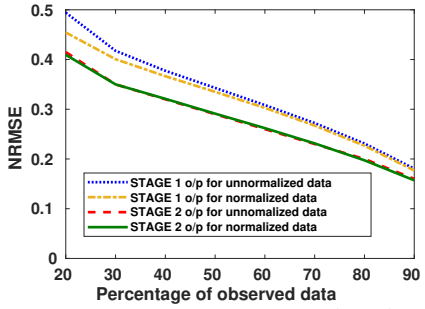
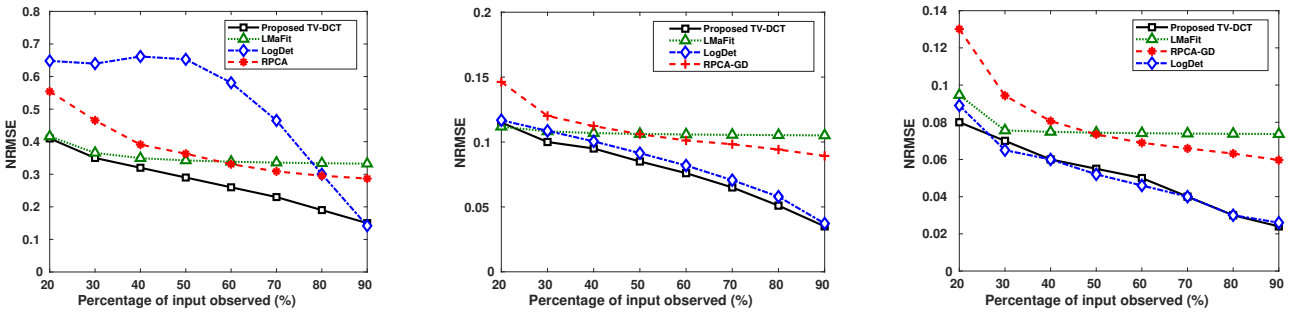


Fig. 1: NRMSE on imputed matrices of $ALLAML$ at varying sampling ratios.

We evaluated our proposed TV-DCT method for log normalized as well as unnormalized data. It is clear from Fig. 1 that Stage-1 performs better with normalized data than with unnormalized data. After Stage-2 denoising, results are almost consistent with each other. Rest of the results are shown in Fig. 2. At every sampling percentage, NRMSE is averaged over 10 separate runs with random sampling in every run. We also compared the performance of the proposed TV-DCT method with three existing state-of-the-art methods for matrix completion namely, LogDet [12], RPCA-GD [13] and LMaFit [11], and used LRSlibrary [26] for computing results with these methods (refer to Fig. 2 and 3).

From Figure 2, TV-DCT method is observed to outperform other methods. Least NRMSE is 0.15, 0.035 and 0.024



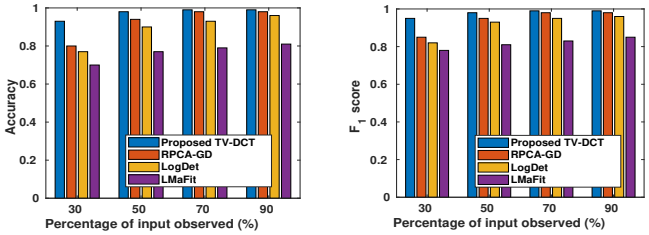
(a) $ALLAML$ dataset (b) $lung$ dataset (c) $Myeloma$ dataset

Fig. 2: Comparison of the proposed TV-DCT method with existing methods at different percentages of observed input data.

Table 2: Classification accuracy and F_1 scores on different sampling percentage of incomplete matrix and the recovered/imputed matrix using proposed TV-DCT method for $ALLAML$ dataset.

	Accuracy				F_1 score			
Classifier	Random Forest		Linear SVM		Random Forest		Linear SVM	
SR	Observed	Recovered	Observed	Recovered	Observed	Recovered	Observed	Recovered
20	0.65	**0.96**	0.67	**0.90**	0.77	**0.96**	.79	**.90**
30	0.65	**0.96**	0.67	**0.93**	0.77	**0.96**	.80	**.95**
40	0.69	**0.97**	0.72	**0.94**	0.79	**0.97**	.82	**.95**
50	0.71	**0.97**	0.74	**0.98**	0.80	**0.97**	.83	**.98**
60	0.75	**0.97**	0.81	**0.99**	0.83	**0.97**	.87	**.99**
70	0.77	**0.96**	0.86	**0.99**	0.84	**0.97**	.90	**.99**
80	0.80	**0.95**	0.91	**0.99**	0.86	**0.96**	.93	**.99**
90	0.85	**0.95**	0.94	**0.99**	0.88	**0.96**	.96	**.99**

at 90% observed data for $ALLAML$, $lung$ and $Myeloma$ datasets respectively. TV-DCT method outperforms other methods even at low observability of 20% of the data on both *ALLAML* and *lung* dataset and is as good as LogDet on $Myeloma$ dataset. However, LogDet is computationally expensive as compared to our method for large matrices of size 99×33297. Moreover, all these methods perform optimally when their parameters are tuned properly. It should be noted that we created missing data little *conservatively* by dropping data at matrix level because this may lead to missing of all subjects' data for any gene at low observability. While the three state-of-the-art methods exploiting the properties of 2D data matrix could still function, it could pose challenges to the proposed method that works on each gene data individually. Despite that, we observe that TV-DCT performs largely better to the existing methods.



(a) Accuracy (b) F_1 score

Fig. 3: Classification accuracy and F_1 scores obtained on imputed matrices of $ALLAML$ at varying sampling ratios.

4.2. Validation

We validated simulation results by performing classification on incomplete matrices and imputed matrices of the $ALLAML$ dataset. We performed features reduction using mutual informarion criterion, where the number of optimal features were obtained by grid search. 5-fold cross validation was performed 20 times and averaged accuracy has been reported. These experiments were performed with Python 2.7 and Sklearn 0.19.1 library. We calculated classification accuracy and F_1 score at each sampling ratio from 20% to 90%. The accuracy and F_1 score are defined as: Accuracy $= \frac{1}{N}\sum_{i=1}^{N} 1(x_i = \tilde{x}_i)$ and $F_1 = \frac{2\times\text{precision}\times\text{recall}}{\text{precision}+\text{recall}}$, where N is the total number of samples in the dataset, x_i is the class label of the i^{th} sample, and $\tilde{x}_i$ is the class label determined by the classifier. We used Random Forest and Linear SVM classifiers. Linear SVM provided best classification results. It is evident from Table-2 that the classification accuracy and F_1 scores are low on incomplete matrices compared to those obtained on imputed matrices. We also compared the classification accuracy on imputed matrices obtained via existing methods as shown in Fig. 3. Classification accuracy is highest when gene expression data is imputed by the proposed TV-DCT method. Owing to space constraints, classification results are not shown on the other two datasets, although similar performance is noted.

5. CONCLUSION

Missing value imputation in gene expression data is important for appropriate analysis in cancer research. In this study, we have presented a novel 2-stage TV-DCT matrix imputation method. TV-DCT is tested on three different cancer datasets at low as well as high observability of data. The comparative performance of the TV-DCT method is observed to be superior to the state-of-the-art matrix completion methods in terms of NRMSE and classification accuracy.

6. REFERENCES

[1] Qian Xiang, Xianhua Dai, Yangyang Deng, Caisheng He, Jiang Wang, Jihua Feng, and Zhiming Dai, "Missing value imputation for microarray gene expression data using histone acetylation information," *BMC bioinformatics*, vol. 9, no. 1, pp. 252, 2008.

[2] A.A. Alizadeh et al., "Distinct types of diffuse large b-cell lymphoma identified by gene expression profiling," *Nature*, vol. 403, no. 6769, pp. 503, 2000.

[3] A. Wee-Chung Liew et al., "Missing value imputation for gene expression data: computational techniques to recover missing data from available information," *Briefings in Bioinformatics*, vol. 12, no. 5, pp. 498–513, 2010.

[4] O. Troyanskaya et al., "Missing value estimation methods for dna microarrays," *Bioinformatics*, vol. 17, no. 6, pp. 520–525, 2001.

[5] Ming Ouyang, William J Welsh, and Panos Georgopoulos, "Gaussian mixture clustering and imputation of microarray data," *Bioinformatics*, vol. 20, no. 6, pp. 917–923, 2004.

[6] Xiaobai Zhang, Xiaofeng Song, Huinan Wang, and Huanping Zhang, "Sequential local least squares imputation estimating missing value of microarray data," *Computers in biology and medicine*, vol. 38, no. 10, pp. 1112–1120, 2008.

[7] S. Oba et al., "A Bayesian missing value estimation method for gene expression profile data," *Bioinformatics*, vol. 19, no. 16, pp. 2088–2096, 2003.

[8] Archana Purwar and Sandeep Kumar Singh, "Hybrid prediction model with missing value imputation for medical data," *Expert Systems with Applications*, vol. 42, no. 13, pp. 5621–5631, 2015.

[9] Ibrahim Berkan Aydilek and Ahmet Arslan, "A hybrid method for imputation of missing values using optimized fuzzy c-means with support vector regression and a genetic algorithm," *Information Sciences*, vol. 233, pp. 25–35, 2013.

[10] Jing Tian, Bing Yu, Dan Yu, and Shilong Ma, "Missing data analyses: a hybrid multiple imputation algorithm using gray system theory and entropy based on clustering," *Applied intelligence*, vol. 40, no. 2, pp. 376–388, 2014.

[11] 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*, vol. 4, no. 4, pp. 333–361, 2012.

[12] Z. Kang, C. Peng, and Q. Cheng, "Top-N recommender system via matrix completion.," in *AAAI*, 2016, pp. 179–185.

[13] X. Yi et al., "Fast algorithms for robust pca via gradient descent," in *Advances in NIPS*, 2016, pp. 4152–4160.

[14] A. Kapur et al., "Gene expression prediction using low-rank matrix completion," *BMC Bioinformatics*, vol. 17, no. 1, pp. 243, 2016.

[15] A. Gupta, S.D. Joshi, and P. Singh, "On the approximate discrete KLT of fractional Brownian motion and applications," *Journal of the Franklin Institute*, 2018.

[16] N. Jain, A. Gupta, and V. Ashok Bohara, "PCI-MDR: Missing Data Recovery in Wireless Sensor Networks using Partial Canonical Identity Matrix," *IEEE Wireless Communications Letters*, 2018.

[17] J. Li et al., "Feature selection: A data perspective," *ACM Computing Surveys (CSUR)*, vol. 50, no. 6, pp. 94, 2017.

[18] T.R. Golub et al., "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring," *science*, vol. 286, no. 5439, pp. 531–537, 1999.

[19] A. Bhattacharjee et al., "Classification of human lung carcinomas by mrna expression profiling reveals distinct adenocarcinoma subclasses," *Proceedings of the National Academy of Sciences*, vol. 98, no. 24, pp. 13790–13795, 2001.

[20] L. López-Corral et al., "Transcriptome analysis reveals molecular profiles associated with evolving steps of monoclonal gammopathies," *Haematologica*, pp. haematol–2013, 2014.

[21] Amir Beck and Marc Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[22] E. van den Berg and M. P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.

[23] E. van den Berg and M. P. Friedlander, "SPGL1: A solver for large-scale sparse reconstruction," June 2007, http://www.cs.ubc.ca/labs/scl/spgl1.

[24] Leonid I Rudin, Stanley Osher, and Emad Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[25] Ivan W Selesnick and Ilker Bayram, "Total variation filtering," *White paper*, 2010.

[26] A. Sobral et al., "LRSlibrary: Low-rank and sparse tools for background modeling and subtraction in videos," in *Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing*. CRC Press, Taylor and Francis Group., 2015.