BEYOND LOW RANK + SPARSE: MULTI-SCALE LOW RANK MATRIX DECOMPOSITION

Frank Ong, and Michael Lustig

Department of Electrical Engineering and Computer Sciences, University of California, Berkeley
{frankong, mlustig}@eecs.berkeley.edu

ABSTRACT

The recent low rank + sparse matrix decomposition [1, 2] enables us to decompose a matrix into sparse and globally low rank components. In this paper, we present a natural generalization and consider the decomposition of matrices into low rank components of multiple scales. The proposed multi-scale low rank decomposition is well motivated in practice, since natural data often exhibit multi-scale structure instead of globally or sparsely. Concretely, we propose a multi-scale low rank modeling to represent a data matrix as a sum of block-wise low rank matrices with increasing scales of block sizes. We then consider the inverse problem of decomposing the data matrix into its multi-scale low rank components, and approach the problem via a convex formulation. Theoretically, we show that under a deterministic incoherence condition, the convex program recovers the multi-scale low rank components exactly. Empirically, we show that the multi-scale low rank decomposition provides a more intuitive decomposition than existing low rank methods, and demonstrate its effectiveness in four applications, including illumination normalization for face images, motion separation for surveillance videos, multi-scale modeling of the dynamic contrast enhanced magnetic resonance imaging and collaborative filtering with age information.

***Index Terms*—** Multi-scale Modeling, Low Rank Matrix, Convex Relaxation, Structured Matrix, Signal Decomposition

1. INTRODUCTION

Signals and systems often exhibit different structures at different scales. Such multi-scale structure has inspired many multi-scale signal transforms, such as wavelets [3] and curvelets [4], that can represent natural signals compactly. By now, multi-scale modeling is associated with many success stories in sparse signal processing applications, including signal compression, denoising, compressed sensing [5, 6], and signal decomposition [7].

On the other hand, low rank methods are commonly used instead to exploit the data correlation and obtain a compact representation at the same time. Recent convex relaxation techniques [8] have further enabled low rank model to be easily adaptable to practical applications, including matrix completion [9], and system identification [10], making it ever more attractive.

In this paper, we present a multi-scale low rank matrix decomposition method that incorporates both multi-scale structure and low rank method. We argue that in practice data matrices are often correlated at different scales, so low rank methods should also exploit the multi-scale structure to obtain a more compact signal representation. Concretely, we propose a multi-scale modeling of matrices as a sum of block-wise low rank matrices with increasing scales of block sizes (more detail in Section 2), and consider the inverse problem of decomposing the matrix into its multi-scale components. Figure 1 illustrates the power of the proposed multi-scale low rank decomposition on a synthetic example.

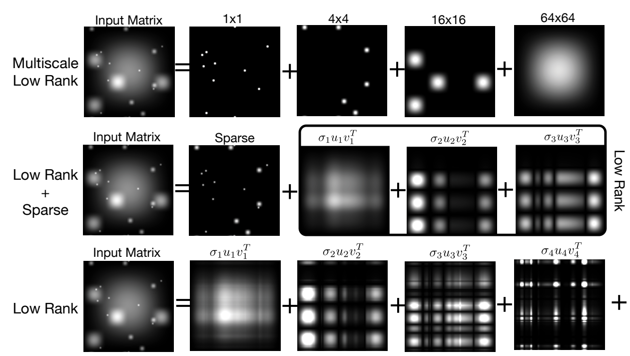

Fig. 1: An example of the proposed multi-scale low rank decomposition compared with other low rank methods. Each blob in the input matrix is a rank-1 outer product of hanning windows. Low rank + sparse fails to capture the blob locality fully. Only the multi-scale low rank decomposition exactly separates the blobs to their corresponding scales, thus representing them as compactly as possible.

Leveraging recent convex relaxation techniques, we propose a convex formulation to perform the multi-scale low rank matrix decomposition. We provide a theoretical analysis in Section 4 that extends the rank-sparsity incoherence results in Chandrasekaran et al. [1] and show that the proposed convex program can decompose the data matrix into its multi-scale components exactly under a deterministic incoherence condition. In addition, in Section 6, we applied the multi-scale low rank decomposition on real-world datasets and show that the proposed multi-scale low rank decomposition provides intuitive multi-scale decomposition and compact signal representation for a wide range of applications.

Our proposed multi-scale low rank matrix decomposition draws many inspirations from recent developments in rank minimization [8–14]. In particular, the multi-scale low rank matrix decomposition is a generalization of the low rank + sparse decomposition proposed by Chandrasekaran et al. [1] and Candès et al. [2]. Our multi-scale low rank convex formulation also fits into the convex demixing framework proposed by McCoy et al. [15, 16] and can be viewed as a concrete and practical example of the convex demixing framework. Bakshi et al. [17] proposed a multi-scale principal component analysis by applying principal component analysis on wavelet transformed signals, but such method implicitly constrains the signal to lie on a predefined wavelet subspace. Various multi-resolution matrix factorization techniques [18, 19] were proposed to greedily peel off components of each scale by recursively applying matrix factorization, but these factorization methods are not straightforward to incorporate with other reconstruction problems as models.

2. MULTI-SCALE LOW RANK MATRIX MODELING

Low rank matrix modeling are used frequently in many applications including biomedical imaging [20], face recognition [21] and collaborative filtering [22]. While low rank modeling captures the notion of data similarity, it ignores any locality information that may be present in the data matrix. Since natural data often exhibits multi-scale structure, a multi-scale low rank modeling is intuitively a more appropriate modeling for many signal processing applications.

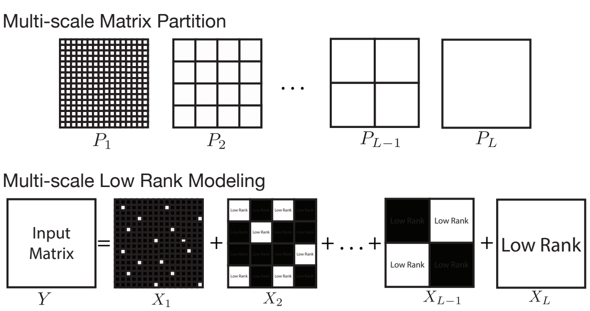

Fig. 2: Illustration of a multi-scale matrix partition and its associated multi-scale low rank modeling. Our multi-scale modeling naturally extends to sparse matrices as 1×1 low rank matrices.

To concretely formulate the multi-scale low rank model, we first assume that we can partition the matrix of interest Y into different scales. Specifically, we assume that we are given a multi-scale partition $\{P_i\}_{i=1}^L$ of the indices of an $M \times N$ matrix, where each block in P_i is an order magnitude larger than the blocks in the previous scale P_{i-1}. To easily transform between the data matrix and the block matrices, we define a block reshape operator $R_b(X)$ to extract a block b from the matrix X and reshapes it into an $m_i \times n_i$ matrix. Its adjoint operator $R_b^\top$ takes the block matrix and embeds it to its original position in a full-size zero matrix.

Given an $M \times N$ input matrix Y and its corresponding multi-scale partition and block reshape operators, we propose the following multi-scale low rank modeling:

$$Y = \sum_{i=1}^{L} X_i, \quad X_i = \sum_{b \in P_i} R_b^\top (U_b S_b V_b^\top) \tag{1}$$

where U_b, S_b, and V_b are matrices with sizes $m_i \times r_b$, $r_b \times r_b$ and $n_i \times r_b$ respectively and form the rank-r_b reduced singular value decomposition (SVD) of $R_b(X_i)$. Figure 2 illustrates one example of the modeling with its associated partition.

3. PROBLEM FORMULATION

Given a data matrix Y that fits the multi-scale low rank model, our goal is to recover $\{X_i\}_{i=1}^L$ from Y. While solving for the decomposition in general seems hopeless, recent development in convex relaxations suggests that rank and sparsity minimization problems can often be relaxed to a convex program via nuclear norm [8, 11] and $l1$-norm minimization [5, 6]. Hence, there is hope that, along the same line, we can perform the multi-scale low rank decomposition exactly via a convex formulation.

Concretely, let us define $\|\cdot\|_{\text{nuc}}$ to be the nuclear norm, and $\|\cdot\|_{\text{msv}}$ be the maximum singular value norm. We define the block-wise nuclear norm for the ith scale as, $\|\cdot\|_{(i)} = \sum_{b \in P_i} \|R_b(\cdot)\|_{\text{nuc}}$ and its associated dual norm as, $\|\cdot\|_{(i)}^* = \max_{b \in P_i} \|R_b(\cdot)\|_{\text{msv}}$. Then, we consider the following convex relaxation for the multi-scale low rank decomposition:

$$\begin{aligned} \underset{X_1,\ldots,X_L}{\text{minimize}} \quad & \sum_{i=1}^{L} \lambda_i \|X_i\|_{(i)} \\ \text{subject to} \quad & Y = \sum_{i=1}^{L} X_i \end{aligned} \tag{2}$$

where $\{\lambda_i\}_{i=1}^L$ are the regularization parameters and their selection will be described in detail in section 5. The convex formulation can be solved by many convex algorithms efficiently. In particular, first-order update steps can be obtained using ADMM [23], and block-wise SVD's. For more detail about the algorithm, please see our extended arXiv paper [24].

We note that the proposed multi-scale low rank convex formulation is a natural generalization of the low rank + sparse convex formulation [1, 2]. In particular, with the two sided matrix partition (Fig. 2), the nuclear norm applied to the 1×1 blocks becomes the element-wise $l1$-norm and the norm for the largest scale is the nuclear norm. In addition, in our extended paper [24] and briefly described in the next section, we show that the core theoretical guarantees in Chandrasekaren et al. [1] for low rank + sparse can be generalized to the multi-scale setting, thereby providing theoretical justification for the proposed convex formualtion.

4. THEORY

In this section, we provide a brief summary of the theoretical aspect behind the proposed convex formulation. At a high level, we show that as long as we can choose the regularization parameters $\{\lambda_i\}_{i=1}^L$ to "balance" the coherence between the matrix components $\{X_i\}_{i=1}^L$, then the proposed convex formulation (2) recovers $\{X_i\}_{i=1}^L$ from Y exactly. For more detail, please refer to our extended arXiv paper [24].

Following Chandrasekaren et al. [1], we consider a deterministic measure of incoherence through the block-wise column and row spaces of $\{X_i\}_{i=1}^L$. Formally, let us first define the block-wise column and row spaces of X_i as:

$$T_i = \{ \sum_{b \in P_i} R_b^\top (U_b X_b^\top + Y_b V_b^\top) : X_b \in \mathbb{C}^{n_i \times r_i}, Y_b \in \mathbb{C}^{m_i \times r_i} \} \quad (3)$$

Then, intuitively, we can say the matrix component X_i is incoherent with the other scales if its block-wise column and row space T_i is not "spiky" with respect to the other scale norms. Formally, we can capture this by defining the coherence parameter for the jth scale signal component X_j with respect to the ith scale to be the following:

$$\mu_{ij} = \max_{N \in T_j, \ \|N\|_{(j)}^* \leq 1} \|N\|_{(i)}^* \quad (4)$$

With the above incoherence definition, the following theorem states our main theoretical result:

Theorem 4.1. If we can choose regularization parameters $\{\lambda_i\}_{i=1}^L$ such that,

$$\sum_{j \neq i} \mu_{ij} \frac{\lambda_j}{\lambda_i} < \frac{1}{2}, \quad \text{for } i = 1, \ldots, L \quad (5)$$

then the proposed convex problem (2) recovers $\{X_i\}_{i=1}^L$ from Y as the unique optimizer.

In particular when the number of scales $L = 2$, the condition on $\{\mu_{12}, \mu_{21}\}$ reduces to a similar form as the rank-sparsity incoherence result in Chandrasekaren et al. [1]. For further detail, please refer to our extended arXiv paper [24].

5. REGULARIZATION PARAMETERS

While theoretically we can establish a criterion on selecting the regularization parameters (see Section 4), such parameters are not straightforward to calculate in practice. In this section, we provide guidance on selecting the regularization parameters $\{\lambda_i\}_{i=1}^L$.

To select the regularization parameters $\{\lambda_i\}_{i=1}^L$, we follow the suggestions from Wright et al. [25] and Fogel et al. [26], and set each regularization parameter λ_i to be the Gaussian complexity of each norm $\| \cdot \|_{(i)}$, which can be found as [27]:

$$\lambda_i \sim \sqrt{m_i} + \sqrt{n_i} + \sqrt{\log(MN / \max\{m_i, n_i\})} \quad (6)$$

This regularization parameter selection is consistent with the ones recommended for low rank + sparse decomposition by Candès et al. [2]. In practice, we found that the suggested regularization parameter selection allows exact multi-scale decomposition when the signal model is matched (for example Figure 1) and provides visually intuitive decomposition for real-world datasets.

6. APPLICATIONS

To test for practical performance, we applied the multi-scale low rank decomposition on four different real-world applications that often use low rank methods. Regularization parameters λ_i were chosen exactly as $\sqrt{m_i} + \sqrt{n_i} + \sqrt{\log(MN/\max\{m_i, n_i\})}$ for multi-scale low rank and $\max(m_i, n_i)$ for low rank + sparse decomposition. Partial cycle spinning was used for multi-scale low rank decomposition to reduce blocking artifacts. For more detail about the experiments, please refer to our extended version on arXiv [24]. In the spirit of reproducible research, we provide a software package (in C and partially in MATLAB) to reproduce most of the results described in this paper: `https://github.com/frankong/multi_scale_low_rank`

6.1. Multi-scale Illumination Normalization for Face Recognition Pre-processing

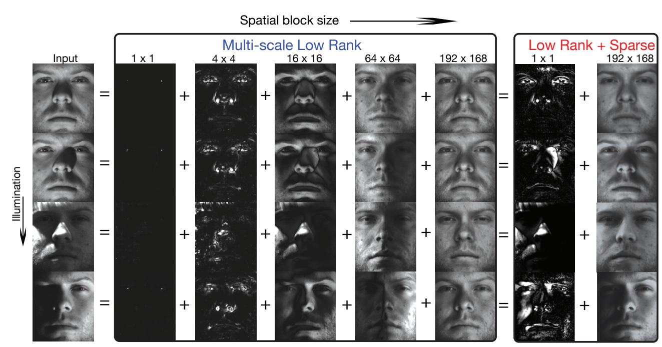

Fig. 3: Multi-scale low rank versus low rank + sparse on faces with uneven illumination.

Low rank + sparse decomposition [2] was recently proposed to capture and remove uneven illumination as sparse errors, yet most shadows are not sparse and contain structure. Hence, we propose using multi-scale low rank decomposition to capture the local spatial correlation of illumination changes and recover face images as the globally low rank component.

Figure 3 shows one of the comparison results on the Yale B face database [28]. Multi-scale low rank decomposition recovered almost shadow-free faces. In particular, the sparkles in the eyes were represented in the 1×1 block size and the larger illumination changes were represented in bigger blocks, thus capturing most of the uneven illumination changes. In contrast, low rank + sparse decomposition could only recover from small illumination changes.

6.2. Multi-scale Motion Separation for Surveillance Videos

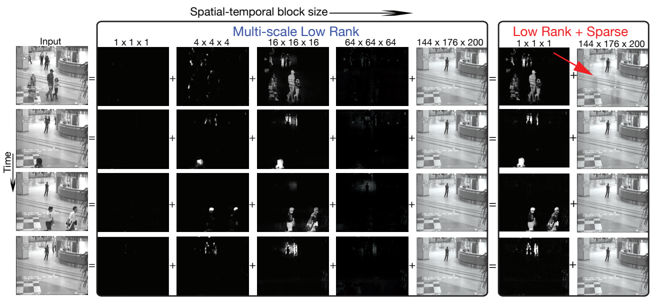

Fig. 4: Multi-scale low rank versus low rank + sparse decomposition on a surveillance video.

Low rank + sparse decomposition [2] was proposed to extract foreground objects as sparse components from the low rank surveillance video background, yet sparsity alone often cannot capture motion compactly, as pointed out by the red arrow in Figure 4. Since video dynamics are correlated at multiple scales in space and time, we propose using the multi-scale low rank modeling to capture the multi-scale motion.

Figure 4 shows one of the results on a surveillance video from Li et al. [29]. Multi-scale low rank decomposition recovered a mostly artifact free background video in the globally low rank component whereas low rank + sparse decomposition exhibited ghosting artifact in certain segments of the video. For the multi-scale low rank decomposition, body motion was mostly captured in the $16 \times 16 \times 16$ scale while fine-scale motion was captured in $4 \times 4 \times 4$ scale.

6.3. Multi-scale Low Rank Modeling for Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI)

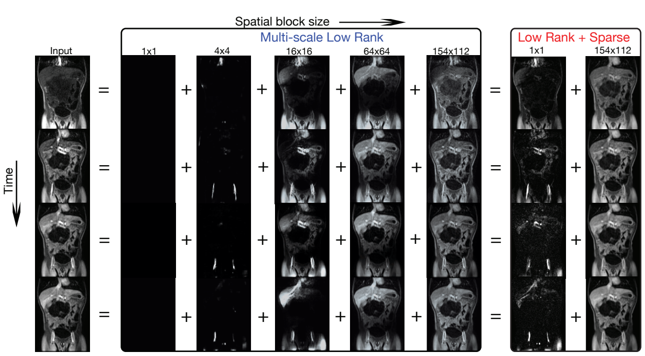

Fig. 5: Multi-scale low rank versus low rank + sparse decomposition on a dynamic contrast enhanced magnetic resonance image series.

Low rank + sparse modeling [30] was proposed to model the static background and dynamics in DCE-MRI as low rank and sparse matrices respectively, yet dynamics in DCE-MRI are almost never sparse. Hence, we propose using a multi-scale low rank modeling to capture contrast dynamics over multiple scales.

Figure 5 shows one of the results on a fully sampled dynamic contrast enhanced image data, acquired in a pediatric patient. In the multi-scale low rank decomposition result, small contrast dynamics in vessels were captured in 4×4 blocks while contrast dynamics in the liver were captured in 16×16 blocks. Hence, different types of contrast dynamics were captured compactly in their suitable scales. In contrast, the low rank + sparse modeling could only provide a coarse separation of dynamics and static tissue, which resulted in neither truly sparse nor truly low rank components.

6.4. Multi-scale Age Grouping for Collaborative Filtering

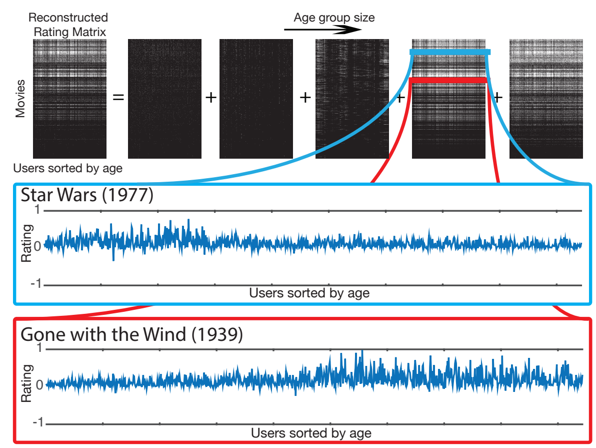

Fig. 6: Multi-scale low rank reconstructed matrix of the 100K MovieLens dataset.

Low rank matrix completion is commonly used in collaborative filtering [9, 11, 14], yet does not exploit the fact that users with similar demographic backgrounds have similar taste. In particular, users of similar age should have similar taste. Hence, we incorporated the proposed multi-scale low rank modeling with matrix completion by partitioning users according to their age and compared it with low rank matrix completion on the 100K MovieLens dataset.

Figure 6 shows a multi-scale low rank reconstructed user rating matrix. Using multiple scales of block-wise low rank matrices, correlations in different age groups were captured. For example, one of the scales shown in Figure 6 captures the tendency that younger users rated Star Wars higher whereas the more senior users rated Gone with the Wind higher. The multi-scale low rank reconstructed matrix achieved a root mean-squared-error of 0.9385 compared to 0.9552 for the low rank reconstructed matrix.

7. REFERENCES

[1] V Chandrasekaran, S Sanghavi, and P A Parrilo, "Rank-sparsity incoherence for matrix decomposition," *SIAM Journal on Optimization*, vol. 21, no. 2, pp. 572–596, 2011.

[2] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright, "Robust principal component analysis?," *Journal of the ACM*, vol. 58, no. 3, pp. 1–37, May 2011.

[3] S G Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *Pattern Analysis and Machine Intelligence*, 1989.

[4] E Candes, L Demanet, D Donoho, and L Ying, "Fast discrete curvelet transforms," *Multiscale Modeling & Simulation*, vol. 5, no. 3, pp. 861–899, 2006.

[5] D L Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[6] E J Candes, J Romberg, and T Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.

[7] S S Chen, D L Donoho, and M A Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[8] M Fazel, *Matrix rank minimization with applications*, Ph.D. thesis, Stanford University, Mar. 2002.

[9] Emmanuel J Candès and Benjamin Recht, "Exact Matrix Completion via Convex Optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, Dec. 2009.

[10] Maryam Fazel, Ting Kei Pong, Defeng Sun, and Paul Tseng, "Hankel Matrix Rank Minimization with Applications to System Identification and Realization," *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 3, pp. 946–977, July 2013.

[11] Benjamin Recht, Maryam Fazel, and Pablo A Parrilo, "Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.

[12] Alekh Agarwal, Sahand Negahban, and Martin J Wainwright, "Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions," *The Annals of Statistics*, vol. 40, no. 2, pp. 1171–1197, Apr. 2012.

[13] Huan Xu, Constantine Caramanis, and Sujay Sanghavi, "Robust PCA via Outlier Pursuit," *Information Theory, IEEE Transactions on*, vol. 58, no. 5, pp. 3047–3064, May 2012.

[14] E J Candes and T Tao, "The Power of Convex Relaxation: Near-Optimal Matrix Completion," *Information Theory, IEEE Transactions on*, vol. 56, no. 5, pp. 2053–2080, May 2010.

[15] Michael B McCoy and Joel A Tropp, "The achievable performance of convex demixing," *arXiv Preprint, arXiv:1309.7478*, Sept. 2013, Available online: `http://arxiv.org/abs/1309.7478v1`.

[16] Michael B McCoy and Joel A Tropp, "Sharp recovery bounds for convex demixing, with applications," *Foundations of Computational Mathematics*, vol. 14, no. 3, pp. 1–51, Jan. 2014.

[17] Bhavik R Bakshi, "Multiscale PCA with application to multivariate statistical process monitoring," *AIChE Journal*, vol. 44, no. 7, pp. 1596–1610, July 1998.

[18] Risi Kondor, Nedelina Teneva, and Vikas Garg, "Multiresolution Matrix Factorization," in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, 2014, pp. 1620–1628.

[19] R Kakarala and P O Ogunbona, "Signal analysis using a multiresolution form of the singular value decomposition," *Image Processing*, vol. 10, no. 5, pp. 724–735, 2001.

[20] Z P Liang, "Spatiotemporal imagingwith partially separable functions," in *Biomedical Imaging: From Nano to Macro*. 2007, pp. 988–991, IEEE.

[21] Ronen Basri and David W Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 218–233, Feb. 2003.

[22] Ken Goldberg, Theresa Roeder, Dhruv Gupta, and Chris Perkins, "Eigentaste: A Constant Time Collaborative Filtering Algorithm," *Information Retrieval*, vol. 4, no. 2, pp. 133–151, 2001.

[23] Stephen Boyd, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[24] Frank Ong and Michael Lustig, "Beyond Low Rank + Sparse: Multi-scale Low Rank Matrix Decomposition," *arXiv Preprint, arXiv:1507.0875*, July 2015, Available online: `http://arxiv.org/abs/1507.08751`.

[25] J Wright, A Ganesh, K Min, and Y Ma, "Compressive principal component pursuit," *Information and Inference: A Journal of the IMA*, 2013.

[26] R Foygel and L Mackey, "Corrupted sensing: Novel guarantees for separating structured signals," *Information Theory*, vol. 60, no. 2, pp. 1223–1247, 2014.

[27] Afonso S Bandeira and Ramon van Handel, "Sharp nonasymptotic bounds on the norm of random matrices with independent entries," *arXiv Preprint arXiv1408.6185*, Aug. 2014, Available online: `http://arxiv.org/abs/1408.6185v3`.

[28] A S Georghiades, P N Belhumeur, and D J Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[29] L Li, W Huang, I Y H Gu, and Q Tian, "Statistical Modeling of Complex Backgrounds for Foreground Object Detection," *IEEE Transactions on Image Processing*, vol. 13, no. 11, pp. 1459–1472, Nov. 2004.

[30] Ricardo Otazo, Emmanuel Candès, and Daniel K Sodickson, "Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components.," *Magnetic Resonance in Medicine*, vol. 73, no. 3, pp. 1125–1136, Mar. 2015.