Robust Principal Component Pursuit for Low-Rank Matrix Recovery

Nipun Wahi
A59024794

Sudhansh Peddabomma
A59023825

Priyanka Nagasamudra
A15996374

Saketh Madhuvarusu
A59023232

***Abstract*—The following report summarises the traditional state-of-the-art approaches for the Low-Rank Matrix Recovery Problem. This problem is described using the so called Robust Principal Pursuit formulation. The optimization procedures in the literature approximate the problem and use proximal gradient and alternating minimization methods to obtain the solution. We analyse the performance of these procedures on a dual formulation for the optimization problem and extend these techniques to applications such as foreground separation and time-series analysis as our contribution. Furthermore, we comment on the implementation issues for each of these procedures and their practicality for the discussed applications. The GitHub repository can be found here**

I. Introduction

A. Motivation

Accurately reconstructing low-rank matrices is a fundamental problem that shows up in many applications such as feature detection in video footage [1], structure from motion for 3D object tracking [2], camera calibration with distortion [3], etc. In this paper we will approach this problem along with its extensions using primal-dual formulations and applications.

Formally, given a data matrix D, we aim to decompose it to a low-rank matrix L and a sparse matrix S. For example, in the application of foreground and background separation in a video, the data matrix is formed by concatenating the column vectors corresponding flattened frames. The background in the frames would correspond to the low-rank matrix because there are minimal changes in the background. On the other hand, foreground, which occupies very few pixels in the video feed corresponds to the sparse matrix.

The optimization procedure associated with Low-Rank matrix recovery is referred to as *Robust Principal Component Analysis* problem given by -

$$\begin{gathered}\min_{L,S} \|L\|_* + \lambda |S|_1, \\ \text{such that } D = L + S\end{gathered} \tag{1}$$

where $\|L\|_*$ refers to the *Nuclear norm* of L equivalent to sum of all of its singular values. Using rank of L directly makes this problem NP-hard, and these relaxations make the problem feasible with standard optimization techniques. Furthermore, $|S|_1$ is the l_1 norm of S representing the sum of absolute values in S. The actual low-rank matrix recovery problem involving the rank of the matrix L is NP-hard, and this relaxation is shown to be good approximation for the problem in [4].

At its core, RPCP addresses the challenge of recovering a low-rank matrix from an set of observations, along with retrieving linear functionals based on a limited subset of available data. This methodology is pivotal in scenarios where the underlying data exhibits a mixture of low-rank structure and sparse outliers, facilitating effective data representation and analysis. Unlike traditional Principal Component Analysis (PCA), which assumes that data is clean and follows a Gaussian distribution.

Approaches such as Proximal-Gradient and ADMM are used to solve this problem efficiently. In this report we delve into an optimization procedure based on the dual formulation and applications of this problem.

B. Previous Works

The study conducted in [4], gives the main idea of convex optimization problem and delves into the part of the precise recovery of a corrupted low-rank matrix. It is considered as a robust extension of classical Principal Component Analysis (PCA), and finds applications across diverse domains. The study illustrates the superior performance and comparative advantages of these novel algorithms through comprehensive simulations, thus laying the groundwork for advancements in robust low-rank matrix recovery techniques. Our work uses these techniques as baselines for comparisons in different applications.

RPCP extends the approach mentioned in [5] to handle outliers and corrupted data. It aims to decompose an observed matrix into two components: a low-rank matrix representing the underlying structure of the data and a sparse matrix capturing noise, outliers, or errors. We extended their work to the dual formulation and refined in such a way that it works on the real-world data by taking input as RGB images. The papers cited above all prove the mathematical validity and convex analysis of the approach, and our mathematical analysis borrows ideas from this literature and extended the work from the ideas derived by [6].

In contrast to these traditional techniques, deep learning approaches such as the one explored in [7] propose CNN based techniques to detect the foreground in videos. The problem with these approaches is that they have to be trained specifically to a certain domain and are not versatile across applications. The deep-learning methods, when trained on representative data, outperform conventional models in scenarios with dynamic backgrounds and limited annotated data, highlighting their robustness and effectiveness in real-world applications. The conventional methods have the advantage that they do not need any prior training and work in real-time for zero-shot inference.

C. Contribution

We aim to analyse the primal-dual formulation and steepest ascent method for the *RPCP* problem as done in [5].

Our **novel** contribution is to extend the ADMM and steepest ascent for the dual formulation algorithms to applications such as video foreground-background separation anomaly detection in time-series data. We also extend these algorithms to obtain consistent results in the presence of additive Gaussian noise.

D. Organization of the Paper

In section II we discuss the modified optimization procedure, derive its dual formulation and discuss the KKT conditions. Following this, in section III we discuss the applied techniques and their respective performance and convergence rates. In this section, we also present the novel extension of these procedures to new domain of applications. Finally, we discuss the results in section IV and provide guidelines for future work in section V.

II. Problem Statement

Given a data matrix $D \in \mathbb{R}^{m\times n}$ with potentially corrupted entries, we aim to find a low-rank matrix $L \in \mathbb{R}^{m\times n}$ and a sparse matrix $S \in \mathbb{R}^{m\times n}$ such that $D = L + S$. More often, D is corrupted with additive noise such as Gaussian, and the exact reconstruction is not possible. In such cases, we resort to a related optimization problem described below.

A. Primal Formulation

The Robust Principal Component Analysis as stated in Eq. 1 is given by

$$\min_{L,S} \|L\|_* + \lambda|S|_1, \qquad \text{such that } D = L + S \tag{2}$$

where $\|.\|_F$ represents the Frobenius norm and ϵ is a parameter chosen based on the variance of the noise. Since the nuclear norm $\|.\|_*$ and the l_1 norm $\|.\|_1$ are difficult to handle with potentially undefined gradients, we consider the dual norms to reformulate the primal problem.

The work in [6] obtains an augmented Lagrangian for this primal formulation as

$$\begin{aligned}\mathcal{L}(L,S,Y) = \|L\|_* + \lambda\|S\|_1 + &\langle Y, D-L-S\rangle \\ &+\frac{\mu}{2}\|D-L-S\|_F\end{aligned} \tag{3}$$

They use an alternating minimizing procedure to solve this optimization problem, and we summarized these results in Section IV.

B. Dual Formulation

We use the results in [5] to derive the dual formulation using dual norms. Using the definition of dual norm,

$$\|L\|_* = \max_Y \mathsf{Tr}(Y^T L) \text{ s.t. } \|Y\|_2 \le 1$$
$$\lambda\|S\|_1 = \max_Y \mathsf{Tr}(Y^T S) \text{ s.t. } \lambda^{-1}\|Y'\|_\infty \le 1$$

Through the results in [5], these dual variables will have the same value at the optimum. The optimization problems can be represented using Semi-definite programs, and using the results in [8] (Appendix A.1), we can formulate the dual as

$$\max_Y \min_{L,S} \quad \mathsf{Tr}(Y^T L) + \mathsf{Tr}(Y^T S) \qquad \text{such that } \|Y\|_2 \le 1, \lambda^{-1}\|Y\|_\infty \le 1 \tag{4}$$

Consequently, we get

$$\max_Y \quad \mathsf{Tr}(Y^T D) \qquad \text{such that } J(Y) \le 1 \tag{5}$$

since $L + S = D$. The matrices L, S can be recovered from Y using the definition of dual norm.

C. KKT Conditions

The Lagrangian for the optimization problem can be derived as

$$\mathcal{L}(L,S,Y) = \|L\|_* + \lambda|S|_1 + \langle Y, D-L-S\rangle \tag{6}$$

Now we can derive the stationary conditions by taking the gradient of our Lagrangian and setting it equal to 0.

$$\nabla_L \mathcal{L}(L,S,Y) = \|L\|_* - Y = 0 \implies Y^* \in \partial\|L^*\|_*$$
$$\nabla_S \mathcal{L}(L,S,Y) = \nabla\lambda|S|_1 - Y = 0 \implies Y^* \in \partial\lambda|S^*|_1$$

The KKT conditions are summarized as follows for a solution (Y^*, L^*, S^*)

1) Primal Feasibility - $D = L^* + S^*$ from Equation 2
2) Dual Feasibility - $J(Y^*) \le 1$ from Equation 5
3) Stationary Point - $Y^* \in \partial\|L^*\|_*$ and $\lambda^{-1}Y^* \in \partial\|S^*\|_1$
4) Complementary Slackness - This is inapplicable in our case since there are no inequality constraints in our primal problem.

In conclusion, we must find an optimal solution (Y^*, L^*, S^*) that obey these conditions in order for strong duality to hold.

III. Approaches

These optimization problems can be solved using approaches such as Accelerated Proximal Gradient (APG), Augmented Lagrange Minimization (ALM), and Steepest Ascent on the Dual Formulation.

Experiments in [6] show that the practical performance of APG methods is not accurate in the absence of good continuation schemes. In contrast, ALM methods achieve good results and converge in very few iterations. We see similar results in our experiments as summarized in IV. However, ALM methods are not easy to scale to larger matrices since the complexity is $\mathcal{O}(m^3)$ (for $S \in \mathbb{R}^{m\times n}$ where $m < n$) due to SVD calculation.

The pseudocode for the ALM algorithm is summarized as follows.

Algorithm 1 RPCP by Alternating Directions

initialize: $S_0 = Y_0 = 0$, $\mu > 0$
do
  compute $L_{k+1} = T_\mu(D - S_k - \mu^{-1} Y_k)$
  compute $S_{k+1} = S_{\lambda\mu}(D - L_{k+1} + \mu^{-1} Y_k)$
  compute $Y_{k+1} = Y_k + \mu(D - L_{k+1} - S_{k+1})$
while not *converged*;
output : L, S

where $T_\mu(X) = US_\tau(\Sigma)V^T$, $X = U\Sigma V^T$ and $S_\tau(x) = sgn(x)\max(\|x\| - \tau, 0)$. This approach works for noisy D as well due to the augmented Lagrangian formulation.

As our contribution, we also implemented the steepest ascent method for the dual formulation we derived in section II. This method is faster in theory in comparison to ALM since it does not require a complete SVD decomposition. However, some implementation issues arise which slow down the approach in practice, and it is not very robust to noise. Nevertheless, the results are more accurate as outlined in the next section. The pseudocode for this approach is given by

Algorithm 2 Robust PCP via the Dual

input : Observation matrix $D \in \mathbb{R}^{m\times n}$, λ.
$Y_0 = \text{sgn}(D)/J(\text{sgn}(D))$ $k \leftarrow 0$ **do**
  if $\|Y_k\|_2 > \lambda^{-1}|Y_k|_\infty$ **then**
    $D_k \leftarrow \pi_2(D);\ L \leftarrow D;\ S \leftarrow 0$
  else if $\lambda^{-1}|Y_k|_\infty > \|Y_k\|_2$ **then**
    $D_k \leftarrow \pi_\infty(D);\ L \leftarrow 0;\ S \leftarrow D$
  else
    $L \leftarrow 0$ $S \leftarrow 0$
    do
      $L \leftarrow \pi_2(D - S);\ S \leftarrow \pi_\infty(D - L)$
    while not converged;
    $D_k \leftarrow L + S$
  $Y_{k+1} \leftarrow Y_k + \delta_k(D - D_k)$ $k \leftarrow k + 1$
while not converged;
output: (L, S)

where $\pi_2(.)$ and $\pi_\infty(.)$ are projections into respective norm spaces as described in [5].

IV. Results

Video Foreground-Background separation

For this application, we linearize frames of a video and concatenate these vectors as columns of the D matrix.

The hypothesis is that the static background of a video forms the low-rank component of the matrix L, whereas the foreground that is not present in all frames of the video forms the sparse component S.

To improve the results on long videos, we form the matrix using alternating frames, i.e, sample every 5th frame, to improve the sparsity of the foreground and also the consistency in the background.

The results using the ADMM approach are summarized in Figure 1 where the optimization procedure successfully separates foreground and background without many artefacts. Performing *hyper-parameter tuning* yields the best $\lambda = 0.001$ and $\mu = 10$. The number of steps in the algorithm is chosen as 20, which gives impressive results in short duration (10 sec on Apple M1 Pro) for 60 frames.

We observe that this approach robustly separates the foreground and background without any training in figure 1b and 1h. As mentioned in section I-B, these algorithms run in real-time in zero-shot modality.

The steepest ascent approach for dual formulation is also implemented for this application, and this also adds to our **novel** contribution. This method has been executed on 30 frames of the video. The best hyperparameters for this algorithm were $\lambda = 0.001$, $\beta = 0.5$ (in the gradient update rule for line search from Pg 304 in [9]) and the steps to converge is 10.

The results for this method as shown in figure 1e are better compared to the ALM method. The results are summarised in table I.

Method	SSIM	PSNR
Steepest Ascent for Dual	0.9296	33.34
Augmented Lagrange Minimization	0.7856	32.15

TABLE I: Foreground-Background Separation Results

Stock Price Analysis

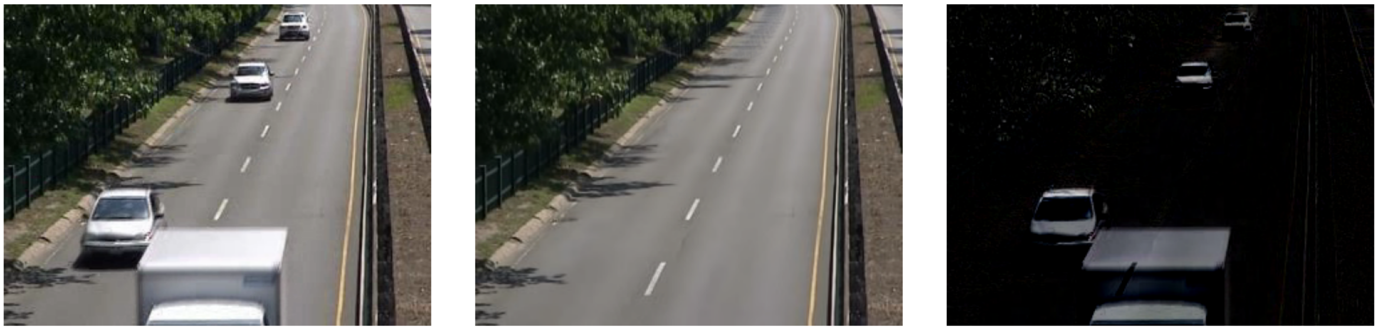
In this application, we propose a **novel** hypothesis that equity prices are linearly correlated and form a low-rank matrix and the anomalies in the data form the sparse-noise. Each row of our data matrix D represents the *normalized* value of a stock over a fixed period. We claim that RPCP can be used to find anomalies in stock prices to determine the global sentiment in the economy. The impact of this formulation is realised in medical applications such as despeckling ultrasound data [10].

We compare this approach with Principal Component Analysis, and show thatw the low-rank recovery approach gives superior results. We performed the analysis on commodities such as Crude Oil, Natural Gas, equities such as Netflix, Google, and crypto data such as Bitcoin obtained from [11]. In figure 2, we see that the orange line indicates the anomalies in the data.

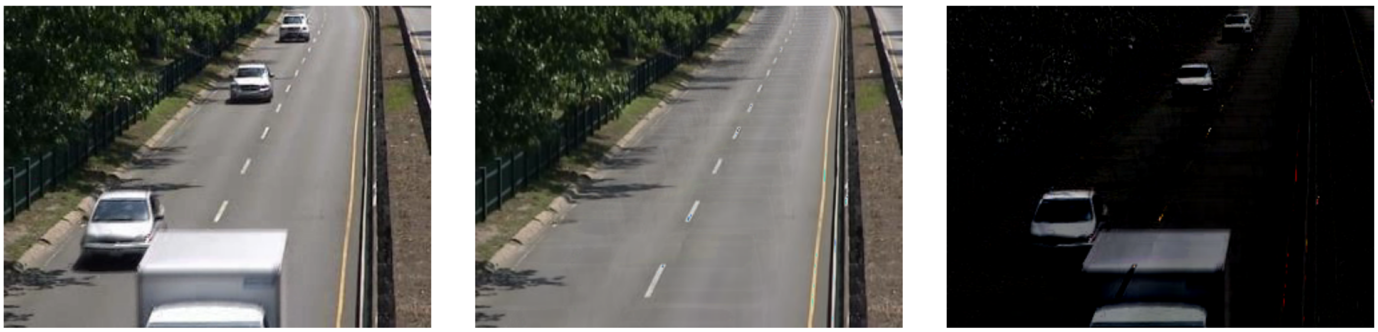
In the analysis for gold price in figure 2b, the RPCP procedure indicates an error (an upward spike) in the price on March 17 2023. Articles show that the Federal bank paused the interest rate around this time, resulting in people buying gold as a hedge strategy.

Similar analysis for crude oil in figure 2a detected a spike due to Russia-Ukraine War on 8 March 2022. The method also identified the surge in natural gas prices in August due to the war in figure 2c.

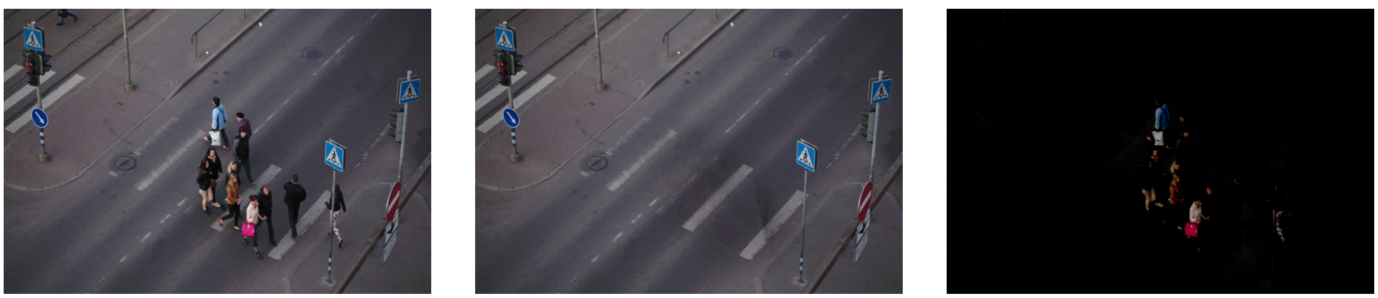
In figure 2d, we compare our method with PCA, and the latter fails to detect any abnormalities in the data. This



(a) True Frame (b) Separated background using ALM (c) Separated foreground using ALM



(d) True Frame (e) Background using Dual steepest ascent (f) Foreground using Dual steepest ascent



(g) True Frame (h) Separated background (i) Separated foreground

Fig. 1: Foreground-Background using Low-Rank Matrix Recovery

approach can be used by portfolio managers to identify *risk factors* in their respective portfolio.

V. Conclusion

In summary, we present the dual formulation for the Robust Principal Component Analysis problem, and extend the algorithms to novel applications such as anomaly detection. We contrast the Augmented Lagrange Multiplier Approach and Steepest Ascent methods for the foreground-background separation task.

Future Work

The RPCP problem formulation assumes S is sparse in the canonical basis. This can be enhanced by considering sparsity in other bases. For example, we can use this approach to extract specific features from videos, e.g, faces, using trained dictionaries. In this case, the problem formulation becomes

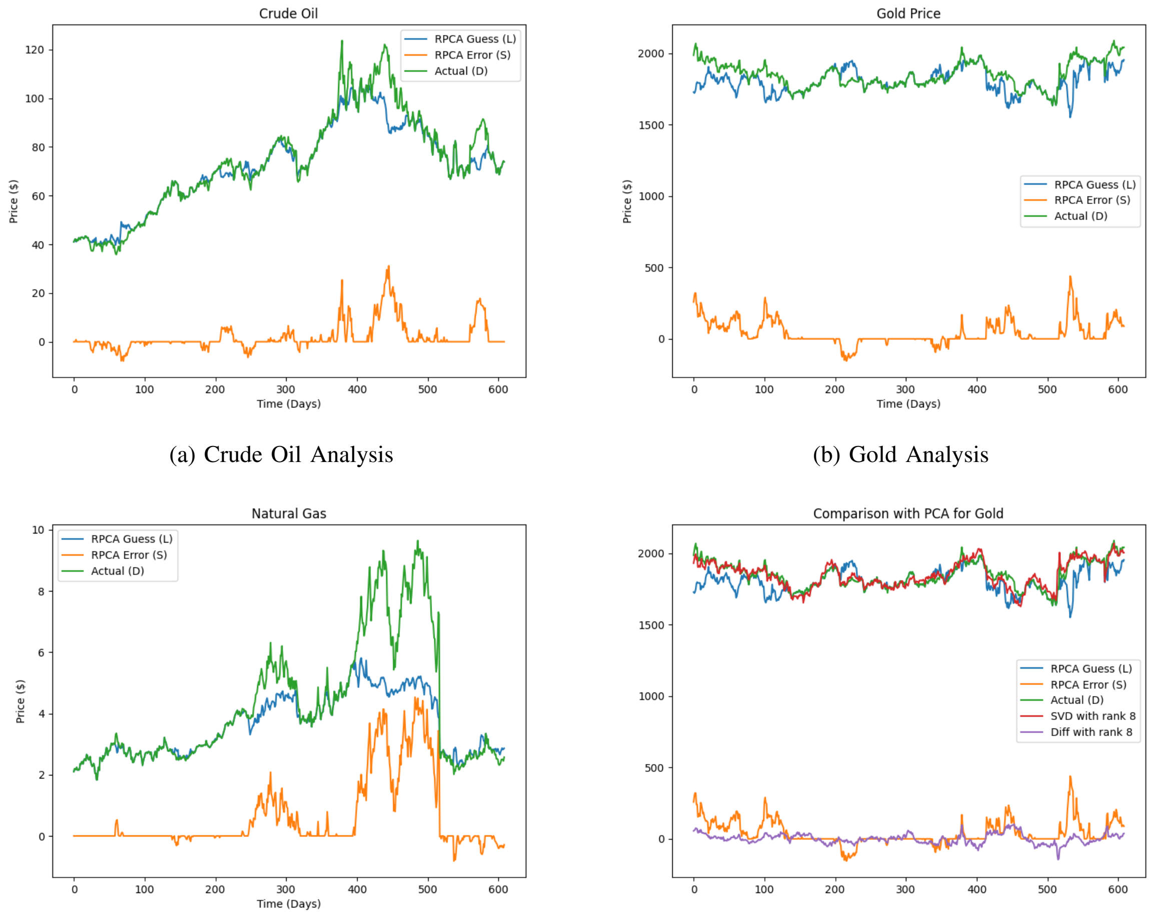
$$\min_{L,S} \|L\|_* + \|S\|_1$$
$$\text{such that } D = L + \mathcal{D}S$$

for a pre-trained dictionary $\mathcal{D}$. Bases such as DFT or DCT can be used to remove low-frequency or high-frequncy noise from frames of a video. To the best of our knowledge, this formulation has not been explored and is worth looking into.

Apart from the theoretical aspect, these algorithms can also be tested for applications such as audio denoising and image inpainting.

Tasks Assignment and Fulfillment

The tasks were distributed and fulfilled equally. Nipun has worked on the experiments including ADMM and Time-Series implementation. Sudhansh focused on implementing steepest ascent for Dual formulation, analysing the performance of algorithms and report organization. Priyanka has worked on the theoretical formulations including KKT conditions and rigorous derivation for the dual formulation. Saketh has performed literature review and report write-up.



(a) Crude Oil Analysis

(b) Gold Analysis

(c) Natural Gas Analysis

(d) Comparison with classical PCA

Fig. 2: Stock Price Analysis for anomaly detection

References

[1] Dongjing Shan and Zhang Chao. Improved 1-tracker using robust pca and random projection. *Mach. Vision Appl.*, 27(4):577–583, may 2016.

[2] Thierry Bouwmans, Sajid Javed, Hongyang Zhang, Zhouchen Lin, and Ricardo Otazo. On the applications of robust pca in image and video processing. *Proceedings of the IEEE*, 106(8):1427–1457, 2018.

[3] Zhengdong Zhang, Yasuyuki Matsushita, and Yi Ma. Camera calibration with lens distortion from low-rank textures. In *CVPR 2011*, pages 2321–2328, 2011.

[4] Emmanuel J. Candes and Benjamin Recht. Exact matrix completion via convex optimization, 2008.

[5] Zhouchen Lin, Arvind Ganesh, John Wright, Leqin Wu, Minming Chen, and Yi Ma. Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix. 2009.

[6] Emmanuel J. Candes, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis?, 2009.

[7] Thangarajah Akilan, Q. M. Jonathan Wu, and Wandong Zhang. Video foreground extraction using multi-view receptive field and encoder–decoder dcnn for traffic and surveillance applications. *IEEE Transactions on Vehicular Technology*, 68(10):9478–9493, 2019.

[8] Venkat Chandrasekaran, Sujay Sanghavi, Pablo A. Parrilo, and Alan S. Willsky. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21(2):572–596, April 2011.

[9] Appendix c a guide to implementable algorithms. In E. Polak, editor, *Computational Methods in Optimization: A Unified Approach*, volume 77 of *Mathematics in Science and Engineering*, pages 299–316. Elsevier, 1971.

[10] Sameera V. Mohd Sagheer and Sudhish N. George. Ultrasound image despeckling using low rank matrix approximation approach. *Biomedical Signal Processing and Control*, 38:236–249, 2017.

[11] US stock market 2020 to 2024.

[12] Jan Lellmann. Convex optimization with applications to image processing. *Lecture Notes from https://www.lellmann.net/work/_media/teaching/coip-mt13-web.pdf, University of Cambridge*, 2013.

[13] Shiqian Ma and Necdet Serhat Aybat. Efficient optimization algorithms for robust principal component analysis and its variants. *Proceedings of the IEEE*, 106(8):1411–1426, 2018.