Online Robust Principle Component Analysis With Huber Loss for Low-Rank Image Recovery

Jacky Lau

May 23, 2021

Abstract

Robust PCA (RPCA) and the recovery of low rank matrices have been studied extensively. While most RPCA algorithms rely on batch optimization, some online methods have since been proposed which are more computationally feasible for large datasets. In this report, we introduce a modified online RPCA method which uses the Huber loss function, as well as discuss some preliminary results on image inpainting and background separation.

1 Introduction

The topic of exact recovery of low-rank matrices has been widely studied with many real work applications ranging from bioinformatics to image processing and knowledge discovery. In recent years, many methods have been explored which provide extensions to the singular value thresholding algorithm [1] and the robust principle component analysis problem [2]. One of the shortcomings of various robust PCA methods is that these methods are generally based on batch optimization which requires the entire set of sample data to be loaded into memory to perform SVD which can be computationally infeasible for sufficiently large datasets. In [3], an online implementation of principle component pursuit [2] using stochastic optimization was proposed which can be applied to recover subspaces of dynamic sample sets. This online robust PCA method is also able to handle large scale datasets by revealing each sample sequentially and thus removing the large memory requirement.

In this report, we will explore an extension of the online robust PCA method using the Huber loss function [4] with applications to image recovery.

2 Online Robust PCA

Suppose the data matrix $D \in \mathbb{R}^{m\times n}$, is generated by corrupting a low rank matrix $A \in \mathbb{R}^{m\times n}$ with some sparse, additive error $E \in \mathbb{R}^{m\times n}$, i.e., $D = A + E$. The robust PCA problem [2] is to recover A given D such that A is low rank and E is a sparse matrix which can be expressed as:

$$\min_{A,\,E} \quad \mathrm{rank}(A) + \gamma\|\mathrm{vec}(E)\|_0, \quad \text{s.t.} \quad D = A + E, \tag{1}$$

where γ is a Lagrangian multiplier. Considering the rank operator and ℓ_0 norm are non-convex functions, they can be substituted for their respective convex surrogate functions, i.e., the nuclear norm and ℓ_1 norm, and expressed in the equivalent formulation:

$$\min_{A,\,E} \quad \|D - A - E\|_F^2 + \lambda_1\|A\|_* + \lambda_2\|\mathrm{vec}(E)\|_1. \tag{2}$$

In [3], in order to be able formulate the online optimization problem, the rank of A is upper bounded by r such that A can be decomposed into the product of two matrices L, R and the nuclear norm can be expressed in the equivalent form [5]:

$$\|A\|_* = \inf_{L\in\mathbb{R}^{p\times r}, R\in\mathbb{R}^{n\times r}} \left\{ \frac{1}{2}\|L\|_F^2 + \frac{1}{2}\|R\|_F^2 : A = LR^T \right\}. \tag{3}$$

Using this, the robust PCA problem can be reformulated as follows,

$$\min_{L,R,E} \quad \|D - LR^T - E\|_F^2 + \frac{\lambda_1}{2}\left(\|L\|_F^2 + \|R\|_F^2\right) + \lambda_2\|\text{vec}(E)\|_1. \tag{4}$$

Given a set of sample data $D = [d_1, \ldots, d_n] \in \mathbb{R}^{p\times n}$, [3] shows that minimizing the following cost function corresponds to the optimal solution for problem 4,

$$g_t(L) = \frac{1}{t}\sum_{i=1}^{t} \ell(\mathbf{d}_i) + \frac{\lambda_1}{2t}\|L\|_F^2 \tag{5}$$

where ℓ denotes the loss function over each sample,

$$\ell(\mathbf{d}_i, L) = \min_{\mathbf{r},\mathbf{e}} \frac{1}{2}\|\mathbf{d}_i - L\mathbf{r} - \mathbf{e}\|_2^2 + \frac{\lambda_1}{2}\|\mathbf{r}\|_2^2 + \lambda_2\|\mathbf{e}\|_1. \tag{6}$$

The stochastic optimization algorithm for optimizing the cost function as proposed in [3] is outlined in algorithm 1, the details of which regarding convergence guarantees are explained in the original paper.

Algorithm 1: Online Robust PCA via Stochastic Optimization

Input: Observed data $\{\mathbf{d}_1, \ldots, \mathbf{d}_T\}$, regularization parameters $\lambda_1, \lambda_2 > 0$, $L_0 \in \mathbb{R}^{p\times r}$, $\mathbf{r}_0 \in \mathbb{R}^r, \mathbf{e}_0 \in \mathbb{R}^p$

for $t = 1, \ldots, T$ **do**

Project new sample $\mathbf{d}_t$:

$$\{\mathbf{r}_i, \mathbf{e}_i\} \leftarrow \arg\min_{\mathbf{r},\mathbf{e}} \frac{1}{2}\|\mathbf{d}_t - L_{t-1}\mathbf{r} - \mathbf{e}\|_2^2 + \frac{\lambda_1}{2}\|\mathbf{r}\|_2^2 + \lambda_2\|\mathbf{e}\|_1. \tag{7}$$

$Y_t \leftarrow Y_{t-1} + \mathbf{r}_t\mathbf{r}_t^T,\ Z_t \leftarrow Z_{t-1} + (\mathbf{d}_t - \mathbf{e}_t)\mathbf{r}_t^T$

Update L_t via algorithm 2:

$$L_t \leftarrow \arg\min_{L} \frac{1}{2}\text{tr}\left(L^T(Y_t + \lambda_1 I)L\right) - \text{tr}(L^T Z_t) \tag{8}$$

end

return $A = L_T R_T^T,\ E_T$

Algorithm 2: Basis Update

Input: $L = [\mathbf{l}_1, \ldots, \mathbf{l}_r] \in \mathbb{R}^{p\times r},\ Y = [\mathbf{y}_1, \ldots, \mathbf{y}_r] \in \mathbb{R}^{r\times r},\ Z = [\mathbf{z}_1, \ldots, \mathbf{z}_r] \in \mathbb{R}^{p\times r}$, $\tilde{Y} = Y + \lambda_1 I$.

for $j = 1, \ldots, r$ **do**

Update the j-th column vector of L:

$$\mathbf{l}_j \leftarrow \frac{1}{\tilde{Y}_{j,j}}(\mathbf{z}_j - L\mathbf{y}_j) + \mathbf{l}_j. \tag{9}$$

end

return L.

---

3 OR-PCA with Huber Loss

In this report we propose a modification to the online robust PCA algorithm which employs the Huber loss function [4] as opposed to the squared ℓ_2 norm in 7. The Huber loss function is piecewise function given by,

$$H_\delta(x) = \begin{cases} \frac{1}{2}x^2 & \text{for } |x| \leq \delta, \\ \delta|x| - \frac{1}{2}\delta^2 & \text{otherwise,} \end{cases} \tag{10}$$

which is quadratic for small values and linear otherwise which is commonly used for introducing robustness to parameter estimation problems. For applying this to the problem above, we use the sum of the elementwise Huber loss as the loss function,

$$F_\delta(\mathbf{x}) = \sum_{i=1}^{n} H_\delta(x_i), \quad \mathbf{x} \in \mathbb{R}^n. \tag{11}$$

The modified algorithm is outlined below.

Algorithm 3: Online Robust PCA via Stochastic Optimization with Huber Loss

Input: Observed data $\{\mathbf{d}_1, \ldots, \mathbf{d}_T\}$, regularization parameters $\lambda_1, \lambda_2 > 0$, $L_0 \in \mathbb{R}^{p\times r}$, $\mathbf{r}_0 \in \mathbb{R}^r, \mathbf{e}_0 \in \mathbb{R}^p$

for $t = 1, \ldots, T$ **do**

Project new sample $\mathbf{d}_t$:

$$\{\mathbf{r}_i, \mathbf{e}_i\} \leftarrow \arg\min_{\mathbf{r},\mathbf{e}} \; F_\delta(\mathbf{d}_t - L_{t-1}\mathbf{r} - \mathbf{e}) + \frac{\lambda_1}{2}\|\mathbf{r}\|_2^2 + \lambda_2\|\mathbf{e}\|_1. \tag{12}$$

$Y_t \leftarrow Y_{t-1} + \mathbf{r}_t\mathbf{r}_t^T, \; Z_t \leftarrow Z_{t-1} + (\mathbf{d}_t - \mathbf{e}_t)\mathbf{r}_t^T$

Update L_t via algorithm 2:

$$L_t \leftarrow \arg\min_{L} \frac{1}{2}\mathrm{tr}\left(L^T(Y_t + \lambda_1 I)L\right) - \mathrm{tr}(L^T Z_t)$$

end

return $A = L_T R_T^T, \; E_T$

4 Results

For the application of image recovery, we consider a series of images which are similar but consisting of some corruption between each frame, for instance foreground and background objects in a sequence of video frames or shadows in face images. A set of sample images can be expressed as $D = [\mathbf{d}_1, \ldots, \mathbf{d}_n] \in \mathbb{R}^{hw\times n}$ where each column vector $\mathbf{d}_i$ denotes the flattened image with height h and width w.

Figures 1 to 3 show a comparison between the recovered images using PCA/SVD, robust PCA [2], online robust PCA (ORPCA) [3], and the proposed approach using the Huber loss function. Figure 1 shows the results on the Extended Yale Face Database B [6] with the recovered uncorrupted images and the respective sparse error images, all the methods can somewhat successfully subtract the shadows from the original face images; however, there is a slight improvement in the quality using the Huber loss function over the regular online RPCA method, RPCA and regular PCA.

Figure 2 shows the results on a set of roughly 100 frames of stable security footage with similar results between the three robust methods in terms of the quality of the recovered background images as well as the sparse error. Figure 3 depicts a more prominent disparity between the results of the four methods as the original images consisted of camera jitter with quickly moving foreground elements. On this set of sample images, the original RPCA method seems to produce the best result in terms of the quality of the uncorrupted images while the other methods consisted of some ghosting; however, the online RPCA methods produced the best results in terms of the sparse noise components with the Huber loss providing some marginal improvement.

Lastly, in figure 4, we compare the results from inpainting a corrupted pattern image. In this case, we only use the single corrupted image as opposed to several stacked images as the underlying image contains a consistent pattern which is low rank. The online RPCA method with the Huber loss function is able to mostly recover the original image while compromising slightly on some of the contrast and detail of the original image while the other methods still contain a rather prominent darkened area where the original corruption took place.

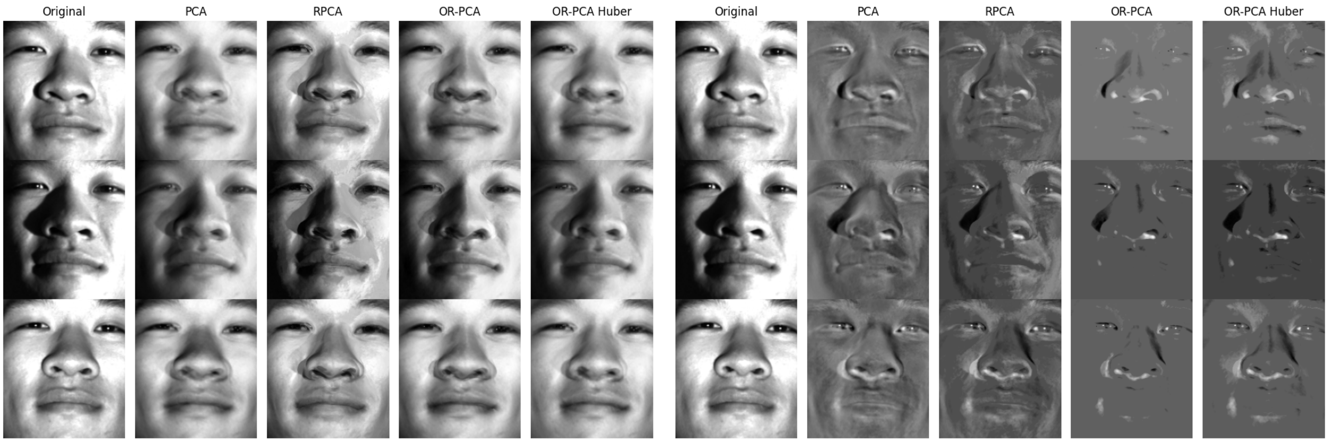

Figure 1: Face images from the Extended Yale Face Database B [6].

Figure 2: Security footage.

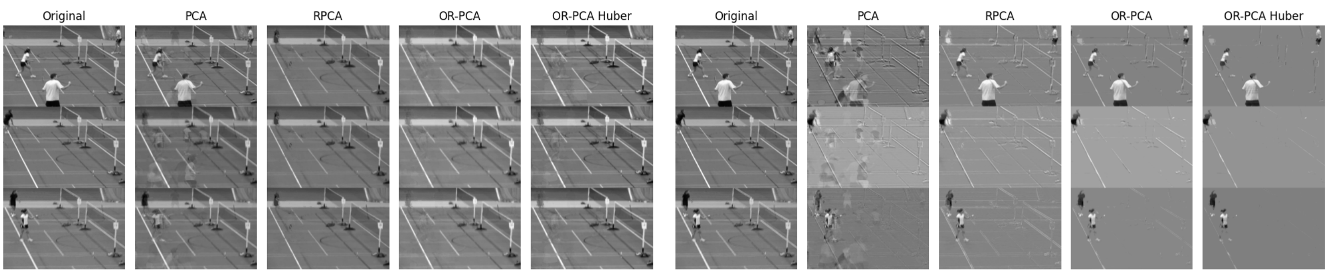

Figure 3: Badminton footage with camera jitter from the ChangeDetection dataset [7].

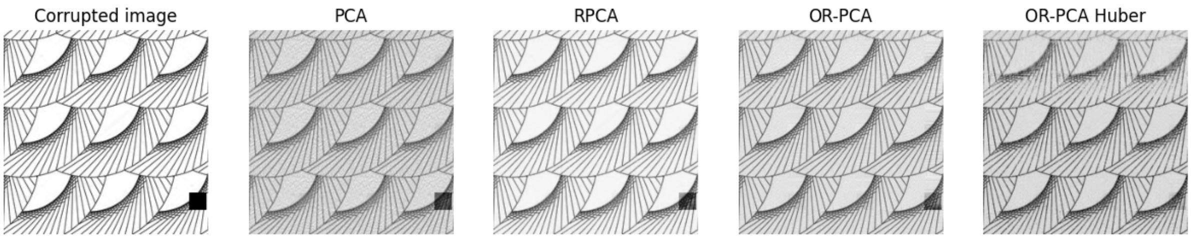

Figure 4: Inpainting a corrupted pattern image.

5 Conclusion

In this report, we explored an extension of the online robust PCA method using the Huber loss function which is more robust to large error terms. Empirically, this modification to the original method is able to more appropriately separate the sparse error from a given set of samples, primarily in the context of background-foreground subtraction as well as inpainting pattern images. Future work may consider the theoretical properties of this modified loss on subspace recovery.

References

[1] Jian-Feng Cai, Emmanuel J. Candes, and Zuowei Shen. *A Singular Value Thresholding Algorithm for Matrix Completion*. 2008. arXiv: `0810.3286 [math.OC]`.

[2] John Wright et al. "Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices via Convex Optimization". In: *Advances in Neural Information Processing Systems*. Ed. by Y. Bengio et al. Vol. 22. Curran Associates, Inc., 2009.

[3] Jiashi Feng, Huan Xu, and Shuicheng Yan. "Online Robust PCA via Stochastic Optimization". In: *Advances in Neural Information Processing Systems*. Ed. by C. J. C. Burges et al. Vol. 26. Curran Associates, Inc., 2013.

[4] Peter J. Huber. "Robust Estimation of a Location Parameter". In: *The Annals of Mathematical Statistics* 35.1 (1964), pp. 73–101. DOI: `10.1214/aoms/1177703732`. URL: `https://doi.org/10.1214/aoms/1177703732`.

[5] Benjamin Recht, Maryam Fazel, and Pablo A. Parrilo. "Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization". In: *SIAM Review* 52.3 (Jan. 2010), pp. 471–501. ISSN: 1095-7200. DOI: `10.1137/070697835`. URL: `http://dx.doi.org/10.1137/070697835`.

[6] 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". In: *IEEE Trans. Pattern Anal. Mach. Intelligence* 23.6 (2001), pp. 643–660.

[7] Nil Goyette et al. "Changedetection.net: A new change detection benchmark dataset". In: *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. 2012, pp. 1–8. DOI: `10.1109/CVPRW.2012.6238919`.