Generalised Scalable Robust Principal Component Analysis

Georgios Papamakarios
georgios.papamakarios13@imperial.ac.uk

Yannis Panagakis
i.panagakis@imperial.ac.uk

Stefanos Zafeiriou
s.zafeiriou@imperial.ac.uk

Department of Computing
Imperial College London
London, UK

Real world visual data, while typically being very high-dimensional, often lie on a *low-dimensional* subspace. *Low-rank* is an attribute capturing the intrinsic low-dimensional structure of the data, when they are represented as column vectors of a matrix. Therefore, a natural approach in low-dimensional subspace recovery is to minimise the rank of the target matrix, subject to a constraint on the error in fitting the data.

By adopting the least squares error metric in fitting (i.e., assuming that the errors follow Gaussian distribution with small variance), the solution of the above mentioned rank minimisation problem is the classical Principal Component Analysis (PCA) [3]. However, visual data obeying postulated low-rank models may also contain gross errors and outliers to which the least squares metric is known to be sensitive.

To overcome the aforementioned drawbacks of the PCA, robust to gross but sparsely supported errors/outliers variants of the PCA have been proposed. With $\mathbf{X} \in \mathbb{R}^{F\times N}$ representing the data matrix, such methods aim to solve the following rank minimisation problem

$$\min_{\mathbf{E},\mathbf{A}} \operatorname{rank}(\mathbf{A}) + \lambda \|\mathbf{E}\|_0 \quad \text{s.t.} \quad \mathbf{X} = \mathbf{A} + \mathbf{E}, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{F\times N}$ is low-rank, $\mathbf{E} \in \mathbb{R}^{F\times N}$ is sparsely supported and accounts for gross errors/outliers and $\lambda > 0$ is a regularisation parameter.

Due to the discrete nature of the rank and the ℓ_0 quasi-norm, problem (1) is NP-hard and thus intractable. To overcome this, a convex relaxation is typically adopted, by surrogating the ℓ_0 quasi-norm of the fitting error matrix and the rank of the target matrix with their closest convex approximants, namely the ℓ_1-norm and the nuclear norm respectively. For instance, the RPCA [2] minimises $\|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_1$ subject to $\mathbf{X} = \mathbf{A} + \mathbf{E}$. The IRPCA [1] rewrites $\mathbf{A} = \mathbf{P}\mathbf{X}$ and minimises $\|\mathbf{P}\|_* + \lambda \|\mathbf{E}\|_1$ subject to $\mathbf{X} = \mathbf{P}\mathbf{X} + \mathbf{E}$. The active subspace RPCA [4] factorises $\mathbf{A} = \mathbf{U}\mathbf{V}$ with $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ and minimises $\|\mathbf{V}\|_* + \lambda \|\mathbf{E}\|_1$ subject to $\mathbf{X} = \mathbf{U}\mathbf{V} + \mathbf{E}$.

Although the aforementioned nuclear/ℓ_1 norm-based methods mainly involve convex problems with global solutions, the relaxation may make the solutions seriously deviate from the original ones. Consequently, a better approximation to the original ℓ_0 quasi-norm-regularised rank minimisation problem (1) is necessary. In this paper, the *Generalised Scalable Robust PCA* (GSRPCA) is proposed, by reformulating the robust PCA problem using the Schatten p-norm $\|\cdot\|_{S_p}$ and the ℓ_q-norm $\|\cdot\|_q$ subject to orthonormality constraints. Let $\mathbf{U} \in \mathbb{R}^{F\times k}$ be column-orthogonal, such that $k \le F$ and $\mathbf{U}^T\mathbf{U} = \mathbf{I}$, and rewrite $\mathbf{A} = \mathbf{U}\mathbf{V}$. GSRPCA is formulated as the following non-convex optimisation problem

$$\min_{\mathbf{E},\mathbf{V},\mathbf{U}} \|\mathbf{V}\|_{S_p}^p + \lambda \|\mathbf{E}\|_q^q \quad \text{s.t.} \quad \begin{array}{l} \mathbf{X} = \mathbf{U}\mathbf{V} + \mathbf{E} \\ \mathbf{U}^T\mathbf{U} = \mathbf{I}. \end{array} \tag{2}$$

The column vectors of $\mathbf{U}$ can be interpreted as the principal components (base vectors) spanning the principal subspace and $\mathbf{V}$ as the projection of $\mathbf{X}$ onto the principal subspace. The state-of-the-art robust variants of the PCA in [1, 2, 4] are all special cases of the GSRPCA when $p = q = 1$ and by properly choosing the number k of principal components. The advantage of (2) is that, for $p \to 0$ and $q \to 0$, a closer approximation to the original rank minimisation problem in (1) can be achieved, by allowing the optimisation function to become non-convex, while retaining the scalability benefit introduced with the factorisation of $\mathbf{A}$.

An efficient alternating directions algorithm for GSRPCA is developed (Algorithm 1), based on the method of augmented Lagrange multipliers. The computational cost per iteration is dominated by 2 SVDs of size $k \times N$ and $F \times k$. Since for most applications typically $k \ll \min(F, N)$, the 2 SVDs can be computed in $\mathcal{O}\left(kN^2 + k^3\right)$ and $\mathcal{O}\left(kF^2 + k^3\right)$ respectively. In contrast, the RPCA [2] requires one SVD of size $F \times N$, which is $\mathcal{O}\left(NF^2 + N^3\right)$ per iteration (assuming $F \ge N$) and the IRPCA [1] requires one SVD of size $F \times F$, which is $\mathcal{O}\left(F^3\right)$ per iteration. Therefore, as long as k remains low, GSRPCA scales well to problems where F and/or N become large, contrary to RPCA and IRPCA.

Algorithm 1: Generalised Scalable Robust PCA

Input: Data matrix $\mathbf{X}$, number of components k, parameters p, q
Initialise: $\mathbf{U} = [k$ first singular vectors of $\mathbf{X}]$, $\mathbf{E} = \mathbf{Y} = \mathbf{0}$, $\mu = \frac{FN}{4\|\mathbf{X}\|_1}$

while *not converged* **do**
  Compute the SVD: $\mathbf{U}^T\left(\mathbf{X} - \mathbf{E} + \mu^{-1}\mathbf{Y}\right) = \mathbf{U}_S\mathbf{D}_S\mathbf{V}_S^T$.
  Update: $\mathbf{V} \leftarrow \mathbf{U}_S\mathcal{S}^p_{\mu^{-1}}\{\mathbf{D}_S\}\mathbf{V}_S^T$.
  Update: $\mathbf{E} \leftarrow \mathcal{S}^q_{\lambda\mu^{-1}}\left\{\mathbf{X} - \mathbf{U}\mathbf{V} + \mu^{-1}\mathbf{Y}\right\}$.
  Compute the SVD: $\left(\mathbf{X} - \mathbf{E} + \mu^{-1}\mathbf{Y}\right)\mathbf{V}^T = \mathbf{U}_S\mathbf{D}_S\mathbf{V}_S^T$.
  Update: $\mathbf{U} \leftarrow \mathbf{U}_S\mathbf{V}_S^T$.
  Update: $\mathbf{Y} \leftarrow \mathbf{Y} + \mu\left(\mathbf{X} - \mathbf{U}\mathbf{V} - \mathbf{E}\right)$.
  Update: $\mu \leftarrow \min(\mu\xi, \mu_{max})$.
  Check convergence: $\|\mathbf{X} - \mathbf{U}\mathbf{V} - \mathbf{E}\|_F \le \varepsilon \|\mathbf{X}\|_F$.
end while

Output: Principal components $\mathbf{U}$, projections $\mathbf{V}$, sparse errors $\mathbf{E}$

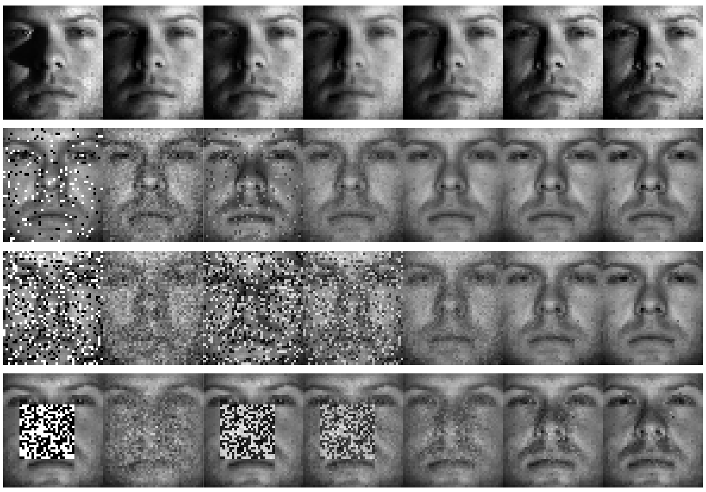

Figure 1: Denoising on the Extended Yale B database. 1st row: shadow removal; 2nd row: 10% salt & pepper noise; 3rd row: 30% salt & pepper noise; 4th row: random patch of maximum size 40×40. 1st column: original image; 2nd column: PCA; 3rd column: RPCA; 4th column: IRPCA; 5th–7th columns: GSRPCA with $p = q \in \{1,\ 0.5,\ 0.1\}$.

The performance of the GSRPCA is assessed by conducting experiments on both synthetic and real data (see for instance Fig. 1). The experimental results indicate that the GSRPCA outperforms the robust PCA methods [1, 2, 4] to which it is compared, without introducing much extra computational cost.

[1] B.-K. Bao, G. Liu, C. Xu, and S. Yan. Inductive robust principal component analysis. *IEEE Trans. Image Processing*, 21(8):3794–3800, 2012.

[2] E. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *J. ACM*, 58(3):11–37, 2011.

[3] H. Hotelling. Analysis of a complex of statistical variables into principal components. *J. Educational Psychology*, 24:417–441,498–520, 1933.

[4] G. Liu and S. Yan. Active subspace: Toward scalable low-rank learning. *Neural Comput.*, 24(12):3371–3394, 2012.