Middlesex University Research Repository

An open access repository of

Middlesex University research

http://eprints.mdx.ac.uk

Copyright:

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

Abstract

The robust estimation of the low-dimensional subspace that spans the data from a set of high-dimensional, possibly corrupted by gross errors and outliers observations is fundamental in many computer vision problems. The state-of-the-art robust principal component analysis (PCA) methods adopt convex relaxations of ℓ_0 quasi-norm-regularised rank minimisation problems. That is, the nuclear norm and the ℓ_1-norm are employed. However, this convex relaxation may make the solutions deviate from the original ones. To this end, the Generalised Scalable Robust PCA (GSRPCA) is proposed, by reformulating the robust PCA problem using the Schatten p-norm and the ℓ_q-norm subject to orthonormality constraints, resulting in a better non-convex approximation of the original sparsity regularised rank minimisation problem. It is worth noting that the common robust PCA variants are special cases of the GSRPCA when $p = q = 1$ and by properly choosing the upper bound of the number of the principal components. An efficient algorithm for the GSRPCA is developed. The performance of the GSRPCA is assessed by conducting experiments on both synthetic and real data. The experimental results indicate that the GSRPCA outperforms the common state-of-the-art robust PCA methods without introducing much extra computational cost.

1 Introduction

Real world visual data, while typically being very high-dimensional, often lie on a *low-dimensional* subspace. This prompts the recovery of the low-dimensional subspace, spanning the data. *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. For instance, 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) [11].

The notion of low-rank is vital in many computer vision and image analysis problems. A typical example is Lambertian reflectance. That is, the rank of the matrix which contains

in its columns a set of images of a Lambertian surface under various lighting conditions is low [2]. In background subtraction (i.e., the problem of modelling the background in a video and simultaneously detecting the objects that stand out from the background), the low-rank constraint is employed to recover the highly correlated content of the background across the video frames [1, 4]. Furthermore, in image alignment (i.e., the problem of transforming a set of different images into the same coordinate system), a batch of aligned images should form a low-rank matrix [15]. The low-rank constraint has also been enforced in image restoration [5] and visual salience detection [16].

Visual data obeying postulated low-rank models contain also gross errors and outliers. Gross errors are often in abundance due to incorrect localisation and tracking, presence of partial occlusion, etc. and rarely follow a Gaussian distribution [1, 4]. Unfortunately, the least squares metric is very sensitive to gross errors and outliers [11] and thus the estimation obtained by the PCA could be arbitrarily away from the true subspace which the data are sampled from. To overcome the aforementioned drawbacks of the PCA, robust to gross but sparsely supported errors/outliers variants of the PCA have been proposed. Such methods include the Robust PCA (RPCA) [4], the Inductive RPCA (IRPCA) [1], the active subspace RPCA [13]. In particular, the aforementioned methods adopt convex relaxation of suitable ℓ_0 quasi-norm-regularised rank minimisation problems. That is, by surrogating, the ℓ_0 quasi-norm of the fitting error matrix and the rank of the target matrix are replaced with their closest convex approximants, namely the ℓ_1-norm [6] and the nuclear norm [7], respectively.

Although the above mentioned nuclear norm minimisation-based RPCA methods involve convex problems with global solutions, the relaxation may make the solutions seriously deviate from the original ones. Consequently, a better approximation of the ℓ_0 quasi-norm-regularised rank minimisation problem is necessary. In this paper, the *Generalised Scalable Robust PCA* (GSRPCA) is proposed, by reformulating the robust PCA problem using the Schatten p-norm and the ℓ_q-norm subject to orthonormality constraints. When p is chosen to be close to zero, the Schatten p-norm is a better approximation to the rank function than the nuclear norm. Also, the choice of ℓ_q-norm ($0 < q \leq 1$) as a measure for fitting error, improves the robustness of the method to sparse errors/outliers. Furthermore, by utilising the unitary invariance of the Schatten p-norm, the low-rank target matrix is represented in a factorised form (i.e., a product of an orthonormal matrix with a low-rank one) which reduces the memory and the computational time complexity of learning (especially when $p = 1$). It is worth mentioning that the state-of-the-art robust variants of the PCA in [1, 4, 13], are all special cases of the GSRPCA when $p = 1$ and by properly choosing the upper bound of the number of the principal components. Although the GSRPCA involves a non-convex objective function, an efficient alternating directions-based algorithm is developed. The performance of the GSRPCA is assessed by conducting experiments on both synthetic and real data. The experimental results indicate that the GSRPCA outperforms the robust PCA methods [1, 4] to which it is compared, without introducing much extra computational cost.

The paper is organised as follows. In Section 2, notation conventions are introduced. The GSRPCA is detailed in Section 3. The experimental results are presented in Section 4 while conclusions are drawn in Section 5.

2 Notations

Throughout the paper, vectors (matrices) are denoted by lowercase (uppercase) boldface letters, *e.g.* $\mathbf{x}$ ($\mathbf{X}$). $\mathbf{X}^T$ is the transpose of $\mathbf{X}$ and X_{ij} is its entry at position (i, j). $\mathbf{I}$ is the

identity matrix of compatible dimensions. The set of real numbers is denoted by $\mathbb{R}$. The sign function is denoted by $\text{sgn}(x)$.

A variety of norms on real-valued matrices will be used. The ℓ_0 quasi-norm counting the number of nonzero entries in $\mathbf{X}$ is denoted by $\|\mathbf{X}\|_0$. If $|\cdot|$ denotes the absolute value operator, $\|\mathbf{X}\|_q = \left(\sum_i \sum_j |X_{ij}|^q\right)^{1/q}$ is the elementwise ℓ_q-norm, of which the Frobenius norm $\|\mathbf{X}\|_F$ is a special case for $q = 2$. If $\sigma_i(\mathbf{X})$ is the ith singular value of $\mathbf{X}$, $\|\mathbf{X}\|_{S_p} = \left(\sum_i \sigma_i(\mathbf{X})^p\right)^{1/p}$ is the Schatten p-norm of $\mathbf{X}$, of which the nuclear norm[1] $\|\mathbf{X}\|_*$ is a special case for $p = 1$. Finally, $\langle \mathbf{X}, \mathbf{Y} \rangle = \sum_i \sum_j X_{ij} Y_{ij}$ is the standard matrix inner product between $\mathbf{X}$ and $\mathbf{Y}$. Note that $\langle \mathbf{X}, \mathbf{X} \rangle = \|\mathbf{X}\|_F^2$.

3 Generalised Scalable Robust Principal Component Analysis

3.1 Formulation as optimisation problem

Let $\mathbf{X} \in \mathbb{R}^{F \times N}$ be the data matrix, with each column being a data point in $\mathbb{R}^F$ (*e.g.* a vectorised image) and assume that it can be decomposed into a low-rank matrix $\mathbf{A} \in \mathbb{R}^{F \times N}$ and a sparse matrix accounting for gross errors/outliers $\mathbf{E} \in \mathbb{R}^{F \times N}$. Motivated by the RPCA [4], the objective is to recover $\mathbf{A}$ and $\mathbf{E}$ from $\mathbf{X}$ by solving:

$$\min_{\mathbf{E},\mathbf{A}} \text{rank}(\mathbf{A}) + \lambda \|\mathbf{E}\|_0 \quad \text{s.t.} \quad \mathbf{X} = \mathbf{A} + \mathbf{E}, \tag{3.1}$$

where $\lambda > 0$ is a regularisation parameter. Due to the discrete nature of the rank and the ℓ_0 quasi-norm, problem (3.1) is NP-hard and thus intractable. To overcome this, the RPCA [4] attempts a convex relaxation and minimises $\|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_1$ instead. Similarly, the IRPCA [9] rewrites $\mathbf{A} = \mathbf{PX}$ and minimises $\|\mathbf{P}\|_* + \lambda \|\mathbf{E}\|_1$ subject to $\mathbf{X} = \mathbf{PX} + \mathbf{E}$. The active subspace RPCA [13] follows an interesting approach by factorising $\mathbf{A}$ into an orthogonal matrix and a low-rank one, which is exploited to improve scalability. Let $\mathbf{U} \in \mathbb{R}^{F \times k}$ be column-orthogonal, such that $k \leq F$ and $\mathbf{U}^T \mathbf{U} = \mathbf{I}$, and rewrite $\mathbf{A} = \mathbf{UV}$. The column vectors of $\mathbf{U}$ can be thought of as the principal components (base vectors) spanning the principal subspace and $\mathbf{V}$ as the projection of $\mathbf{X}$ onto the principal subspace. Due to the unitary invariance of the nuclear norm, $\|\mathbf{UV}\|_* = \|\mathbf{V}\|_*$ and (3.1) is relaxed to

$$\min_{\mathbf{E},\mathbf{V},\mathbf{U}} \|\mathbf{V}\|_* + \lambda \|\mathbf{E}\|_1 \quad \text{s.t.} \quad \begin{matrix} \mathbf{X} = \mathbf{UV} + \mathbf{E} \\ \mathbf{U}^T \mathbf{U} = \mathbf{I}. \end{matrix} \tag{3.2}$$

We go one step further, noticing that the nuclear norm is a special case of the Schatten p-norm (which is also unitary invariant) and that the elementwise ℓ_1-norm is a special case of the elementwise ℓ_q-norm, for $p = 1$ and $q = 1$ respectively. We generalise (3.2) to

$$\min_{\mathbf{E},\mathbf{V},\mathbf{U}} \|\mathbf{V}\|_{S_p}^p + \lambda \|\mathbf{E}\|_q^q \quad \text{s.t.} \quad \begin{matrix} \mathbf{X} = \mathbf{UV} + \mathbf{E} \\ \mathbf{U}^T \mathbf{U} = \mathbf{I}. \end{matrix} \tag{3.3}$$

The advantage of (3.3) is that, for $p \to 0$ and $q \to 0$, one can achieve a closer approximation to the original rank minimisation problem in (3.1), by allowing the optimisation function to

---

[1]Also known as the trace norm.

become non-convex, while retaining the scalability benefit introduced with the factorisation of $\mathbf{A}$. For this reason, we refer to (3.3) as the *Generalised Scalable Robust PCA problem*. In the following, we describe an algorithm for solving (3.3).

3.2   Solution based on the method of Augmented Lagrange Multipliers

We develop an efficient algorithm for solving the GSRPCA problem, based on the method of Augmented Lagrange Multipliers (ALM) [3]. The (partial) augmented Lagrangian of the GSRPCA problem in (3.3) is

$$\mathcal{L}(\mathbf{E}, \mathbf{V}, \mathbf{U}, \mathbf{Y}, \mu) = \|\mathbf{V}\|_{S_p}^p + \lambda \|\mathbf{E}\|_q^q + \langle \mathbf{X} - \mathbf{U}\mathbf{V} - \mathbf{E}, \mathbf{Y} \rangle + \frac{\mu}{2} \|\mathbf{X} - \mathbf{U}\mathbf{V} - \mathbf{E}\|_F^2 \qquad (3.4)$$

where $\mathbf{Y} \in \mathbb{R}^{F \times N}$ is the Lagrange multiplier matrix for the first equality constraint in (3.3) and $\mu > 0$ is a positive parameter. Consequently, to solve (3.3), we need to minimise $\mathcal{L}$ with respect to $\mathbf{E}$, $\mathbf{V}$, and $\mathbf{U}$, subject to $\mathbf{U}^T\mathbf{U} = \mathbf{I}$. It is possible to take advantage of the separability of $\mathcal{L}$ with respect to its arguments and employ an Alternating Directions approach, i.e. minimise $\mathcal{L}$ with respect to each argument separately, keeping all other arguments fixed. The procedure is detailed in the following subsections.

3.2.1   Minimisation with respect to E

The minimisation of $\mathcal{L}$ with respect to $\mathbf{E}$ is written as follows

$$\begin{aligned}
\mathbf{E}^* &= \arg\min_{\mathbf{E}} \lambda \|\mathbf{E}\|_q^q + \langle \mathbf{X} - \mathbf{U}\mathbf{V} - \mathbf{E}, \mathbf{Y} \rangle + \frac{\mu}{2} \|\mathbf{X} - \mathbf{U}\mathbf{V} - \mathbf{E}\|_F^2 \\
&= \arg\min_{\mathbf{E}} \alpha \|\mathbf{E}\|_q^q + \frac{1}{2} \|\mathbf{E} - \mathbf{Z}\|_F^2,
\end{aligned} \qquad (3.5)$$

where $\alpha = \lambda \mu^{-1}$ and $\mathbf{Z} = \mathbf{X} - \mathbf{U}\mathbf{V} + \mu^{-1}\mathbf{Y}$. We note that (3.5) is separable with respect to the elements of $\mathbf{E}$ and therefore can be decomposed into $F \times N$ problems of the form

$$\min_{E_{ij}} \alpha |E_{ij}|^q + \frac{1}{2}(E_{ij} - Z_{ij})^2. \qquad (3.6)$$

Define $h(E_{ij}) = \alpha |E_{ij}|^q + \frac{1}{2}(E_{ij} - Z_{ij})^2$, $c_1 = [\alpha q(1-q)]^{\frac{1}{2-q}}$ and $c_2 = c_1 + \alpha q |c_1|^{q-1}$. Then, according to [14], the minimiser of (3.6) for $q \in (0,1]$ is given analytically by

$$E_{ij}^* = \begin{cases} 0 & \text{if } |Z_{ij}| \leq c_2 \\ \arg\min_{E_{ij} \in \{0,\rho_1\}} h(E_{ij}) & \text{if } Z_{ij} > c_2 \\ \arg\min_{E_{ij} \in \{0,\rho_2\}} h(E_{ij}) & \text{if } Z_{ij} < -c_2, \end{cases} \qquad (3.7)$$

where ρ_1 and ρ_2 are the roots of $h'(E_{ij}) = \alpha q |E_{ij}|^{q-1} \text{sgn}(E_{ij}) + E_{ij} - Z_{ij} = 0$ in $[c_1, Z_{ij}]$ and $[Z_{ij}, -c_1]$ respectively. The roots ρ_1 and ρ_2 can easily be obtained using the iterative Newton-Raphson root-finding method initialised at Z_{ij} [14]. We will refer to the element-wise application of (3.7) as *generalised q-shrinkage* and denote it by $\mathcal{S}_a^q\{\cdot\}$. Therefore, the solution of (3.5) is given by

$$\mathbf{E}^* = \mathcal{S}_{\lambda\mu^{-1}}^q \{\mathbf{X} - \mathbf{U}\mathbf{V} + \mu^{-1}\mathbf{Y}\} \qquad (3.8)$$

3.2.2   Minimisation with respect to V

Taking into account the orthonormality of $\mathbf{U}$ and the fact that the Frobenius norm is unitary-invariant, the minimisation of $\mathcal{L}$ with respect to $\mathbf{V}$ is written as follows

$$\mathbf{V}^* = \arg\min_{\mathbf{V}} \|\mathbf{V}\|_{S_p}^p + \langle \mathbf{X} - \mathbf{UV} - \mathbf{E}, \mathbf{Y} \rangle + \frac{\mu}{2} \|\mathbf{X} - \mathbf{UV} - \mathbf{E}\|_F^2$$

$$= \arg\min_{\mathbf{V}} \mu^{-1} \|\mathbf{V}\|_{S_p}^p + \frac{1}{2} \|\mathbf{UV}\|_F^2 - \langle \mathbf{UV}, \mathbf{X} - \mathbf{E} + \mu^{-1}\mathbf{Y} \rangle$$

$$= \arg\min_{\mathbf{V}} \mu^{-1} \|\mathbf{V}\|_{S_p}^p + \frac{1}{2} \|\mathbf{V}\|_F^2 - \langle \mathbf{V}, \mathbf{U}^T (\mathbf{X} - \mathbf{E} + \mu^{-1}\mathbf{Y}) \rangle$$

$$= \arg\min_{\mathbf{V}} \mu^{-1} \|\mathbf{V}\|_{S_p}^p + \frac{1}{2} \|\mathbf{V} - \mathbf{U}^T (\mathbf{X} - \mathbf{E} + \mu^{-1}\mathbf{Y})\|_F^2 \qquad (3.9)$$

Assume the Singular Value Decomposition (SVD) $\mathbf{U}^T (\mathbf{X} - \mathbf{E} + \mu^{-1}\mathbf{Y}) = \mathbf{U}_S \mathbf{D}_S \mathbf{V}_S^T$. Then, as it has been shown in [14], the optimiser of $\mathcal{L}$ for $p \in (0,1]$ is given by

$$\mathbf{V}^* = \mathbf{U}_S \mathcal{S}_{\mu^{-1}}^p \{\mathbf{D}_S\} \mathbf{V}_S^T. \qquad (3.10)$$

3.2.3   Minimisation with respect to U

Minimising $\mathcal{L}$ with respect to $\mathbf{U}$, subject to $\mathbf{U}^T \mathbf{U} = \mathbf{I}$, can be rewritten as follows

$$\mathbf{U}^* = \arg\min_{\mathbf{U}} \langle \mathbf{X} - \mathbf{UV} - \mathbf{E}, \mathbf{Y} \rangle + \frac{\mu}{2} \|\mathbf{X} - \mathbf{UV} - \mathbf{E}\|_F^2$$

$$= \arg\min_{\mathbf{U}} \frac{1}{2} \|(\mathbf{X} - \mathbf{E} + \mu^{-1}\mathbf{Y}) - \mathbf{UV}\|_F^2 \quad \text{s.t.} \quad \mathbf{U}^T \mathbf{U} = \mathbf{I} \qquad (3.11)$$

Assume the following SVD

$$(\mathbf{X} - \mathbf{E} + \mu^{-1}\mathbf{Y}) \mathbf{V}^T = \mathbf{U}_S \mathbf{D}_S \mathbf{V}_S^T \qquad (3.12)$$

Then, due to the *Reduced Rank Procrustes Theorem* [17], the solution of the aforementioned problem is given by

$$\mathbf{U}^* = \mathbf{U}_S \mathbf{V}_S^T. \qquad (3.13)$$

3.3   Algorithm and implementation

Algorithm 3.1 describes the GSRPCA in detail. Following [4], we use $\lambda = 1/\sqrt{\max(F,N)}$ and we initialise $\mu = FN/4\|\mathbf{X}\|_1$. Also, for updating μ, we use $\xi = 1.2$ and a maximum value of $\mu_{max} = 10^9$. In order to determine convergence, we use $\varepsilon = 10^{-7}$. Finally, we use a "warm" initialisation for $\mathbf{U}$, setting it to the k leftmost singular vectors of $\mathbf{X}$, which is equivalent to applying classical PCA on $\mathbf{X}$ and initialising $\mathbf{U}$ with the resulting principal components.

3.4   Discussion and connections to previous works

Algorithm 3.1 is particularly attractive in that, in addition to decomposing a data matrix $\mathbf{X}$ into a low-rank and a sparse component, it also recovers the principal subspace $\mathbf{U}$. As such, unlike the RPCA [4], it is essentially an inductive method. To see this, note that $\mathbf{V} = \mathbf{U}^T \mathbf{X}$ and take $\mathbf{P} = \mathbf{U}\mathbf{U}^T$ to be the projection matrix onto the principal subspace. Since the low-rank component of $\mathbf{X}$ can be written as $\mathbf{A} = \mathbf{P}\mathbf{X}$, GSRPCA subsumes both the RPCA [4]

Algorithm 3.1: Generalised Scalable Robust PCA

Input: Data matrix $\mathbf{X}$, number of principal components k, norm parameters p and q.
Initialise: $\mathbf{U} = [k$ leftmost singular vectors of $\mathbf{X}]$, $\mathbf{E} = \mathbf{Y} = 0$, $\mu = {}^{FN}/4\|\mathbf{X}\|_1$.

1 **while** *not converged* **do**
2   Compute the SVD: $\mathbf{U}^T \left(\mathbf{X} - \mathbf{E} + \mu^{-1}\mathbf{Y} \right) = \mathbf{U}_S \mathbf{D}_S \mathbf{V}_S^T$.
3   Update: $\mathbf{V} \leftarrow \mathbf{U}_S \mathcal{S}_{\mu^{-1}}^p \{\mathbf{D}_S\} \mathbf{V}_S^T$.
4   Update: $\mathbf{E} \leftarrow \mathcal{S}_{\lambda\mu^{-1}}^q \{\mathbf{X} - \mathbf{UV} + \mu^{-1}\mathbf{Y}\}$.
5   Compute the SVD: $\left(\mathbf{X} - \mathbf{E} + \mu^{-1}\mathbf{Y} \right) \mathbf{V}^T = \mathbf{U}_S \mathbf{D}_S \mathbf{V}_S^T$.
6   Update: $\mathbf{U} \leftarrow \mathbf{U}_S \mathbf{V}_S^T$.
7   Update: $\mathbf{Y} \leftarrow \mathbf{Y} + \mu \left(\mathbf{X} - \mathbf{UV} - \mathbf{E} \right)$.
8   Update: $\mu \leftarrow \min \left(\mu \xi, \mu_{max} \right)$.
9   Check convergence: $\|\mathbf{X} - \mathbf{UV} - \mathbf{E}\|_F \leq \varepsilon \|\mathbf{X}\|_F$.
10 **end while**
Output: Principal components $\mathbf{U}$, projections $\mathbf{V}$ and sparse errors $\mathbf{E}$.

---

and the IRPCA [1], as it both recovers the low-rank component $\mathbf{A}$ and the projection matrix $\mathbf{P}$. Note that the latter can be subsequently used for efficiently projecting previously unseen data points onto the principal subspace, without the need for repeating the decomposition procedure. This is an additional advantage of the GSRPCA compared to the RPCA.

Furthermore, our method shows flexibility in the choice of the number of principal components k. For $k = F$, the result of algorithm 3.1 for $p = q = 1$ is equivalent to that of the RPCA and the IRPCA. However, since $\mathrm{rank}(A) \leq \mathrm{rank}(P) = k$, k serves as a controllable upper bound to the rank of the low-rank component. This property can be exploited in applications where an upper bound of rank value is desirable or the rank is known a priori.

An interesting special case of Algorithm 3.1 is for $p = q = 1$, which corresponds to solving the convex relaxation problem in (3.2). It is easy to see that, in this case, the generalised p-shrinkage operator reduces to the elementwise application of the well-known shrinkage operator, defined by

$$\mathcal{S}_a\{x\} = \mathrm{sgn}(x)\max(|x| - a, 0) \tag{3.14}$$

Therefore, our framework subsumes the active subspace RPCA [13] as a special case. Furthermore, our work extends [13] to $p, q < 1$ by removing the need for convexity of the objective function. In fact, perhaps the most attractive feature of GSRPCA is that it can circumvent the intractability of the NP-hard minimisation problem in (3.3) by allowing a tighter approximation as p and q approach zero.

One of the main characteristics of GSRPCA is the non-convexity of its optimisation problem in (3.3). This non-convexity comes from both the factorisation $\mathbf{A} = \mathbf{UV}$ and the usage of the Schatten p-norm and the elementwise ℓ_q-norm, which become non-convex when $p, q < 1$. In contrast, both the RPCA and the IRPCA involve convex optimisation problems. The non-convexity of GSRPCA allows for both recovering the principal subspace and better approximating the original intractable rank minimisation problem in (3.1). On the downside, only local optimality of the solution can be guaranteed, unlike the RPCA and IRPCA which, due to their convexity, attain globally optimal solutions. In GSRPCA, initialisation is therefore critical in obtaining a good solution. Our suggestion is initialising $\mathbf{U}$ with the k leftmost singular vectors of data matrix $\mathbf{X}$, which is equivalent to first performing classical PCA on $\mathbf{X}$ and then initialising $\mathbf{U}$ with the resulting principal components. Empirically, we have found

that using this "warm" initialisation scheme yields respectable results in most cases, as the experiments in Section 4 corroborate.

Finally, it is worth discussing the good scalability properties of algorithm 3.1 and compare it to similar approaches. For the convex case where $p = q = 1$, the generalised p-shrinkage operator becomes particularly efficient as it reduces to elementwise shrinkage and the computational cost per iteration is dominated by 2 SVDs of size $k \times N$ and $F \times k$. Since for most applications it is typical that $k \ll \min(F, N)$, the 2 SVDs can be computed in $\mathcal{O}(kN^2 + k^3)$ and $\mathcal{O}(kF^2 + k^3)$ respectively [9]. In comparison, the RPCA requires one SVD of size $F \times N$, which is $\mathcal{O}(NF^2 + N^3)$ per iteration[2] and the IRPCA requires one SVD of size $F \times F$, which is $\mathcal{O}(F^3)$ 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.

4 Experiments

We compare the GSRPCA with (a) the classical PCA (b) the RPCA [4], and (c) the IRPCA [1]. In all experiments, the regularisation parameter is set for the RPCA equal to $\lambda = 1/\sqrt{\max(F,N)}$, as suggested in [4], and for the IRPCA equal to $\lambda = 0.0001$, similar to the experiments in [1]. Three versions of the GSRPCA are considered, corresponding to $p = q \in \{1, 0.5, 0.1\}$. Results are reported on both real and synthetic data.

4.1 Synthetic data

We first validate our approach on synthetic data. We randomly generate a low-rank matrix $\mathbf{A} \in \mathbb{R}^{1000 \times 1000}$ with $\mathrm{rank}(\mathbf{A}) = 50$ as the product $\mathbf{A} = \mathbf{XY}^T$, where the entries of $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{1000 \times 50}$ are i.i.d. Gaussian variables with mean 0 and variance 10^{-3}. We then form data matrix $\mathbf{X} = \mathbf{A} + \mathbf{E}$, where $\mathbf{E}$ is sparse with independent Bernoulli ± 1 nonzero entries. The percentage of nonzero entries of $\mathbf{E}$ is denoted as $\rho_E = \|\mathbf{E}\|_0 / 10^6$. Fig. 1 shows the reconstruction error for each method calculated as $\|\hat{\mathbf{A}} - \mathbf{A}\|_1 / \|\mathbf{A}\|_1$, where $\hat{\mathbf{A}}$ is the reconstructed low-rank component of $\mathbf{X}$, for (a) $\rho_E = 5\%$, (b) $\rho_E = 15\%$ and (c) $\rho_E = 30\%$. It can be seen that the GSRPCA outperforms all other methods, with the difference in performance becoming particularly evident for large values of ρ_E. It is noteworthy that the performance of the GSRPCA increases as p and q approach 0.

4.2 Real data: reconstruction of facial images

We evaluate our approach on the Extended Yale B database of cropped face images [8, 12], taken under various illumination conditions. This set of images is intrinsically of low rank, making it suitable for the application of rank minimisation techniques. We use $N = 256$ images and we rescale them to size 48×42, i.e. $F = 2016$.

We test all methods both on the original images and on synthetically corrupted versions thereof. For synthetic corruption we consider two scenarios. The first consists of corrupting a certain percentage of each image with random salt & pepper noise, in which case the value of the corrupted pixels is set to 0 or 1 with equal probability. The second, more challenging case, involves corrupting each image with a spatially continuous rectangular patch of random noise. The horizontal and vertical patch sizes are uniformly chosen between 1 and 40 pixels independently and the patch itself consists of random pixel values taken from $\{0, 1\}$.

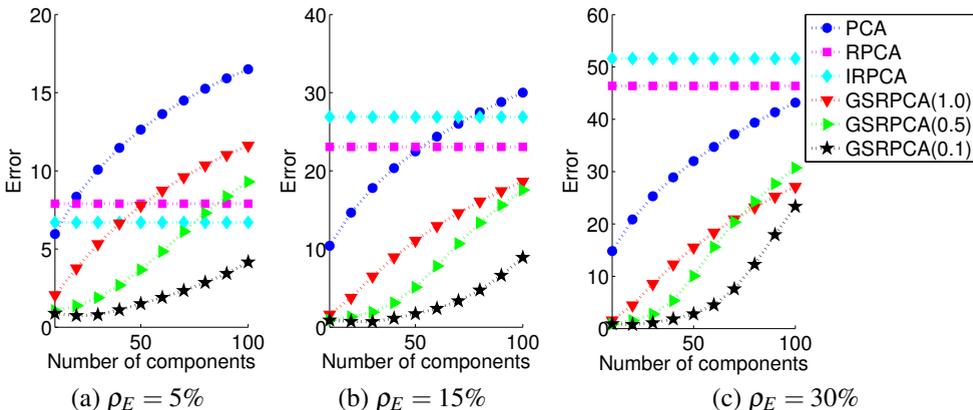
[2]Assuming without loss of generality that $F \geq N$.

Figure 1: Reconstruction error for each method versus the number of principal components k on synthetic data. For RPCA and IRPCA the error is constant since they do not consider k. The percentage ρ_E of nonzero entries in $\mathbf{E}$ is (a) 5%, (b) 15% and (c) 30%.

Fig. 2 shows the result of reconstruction of example images over all methods and for various noise corruptions, both natural (i.e. cast shadows) and synthetic. The number of components k used for the PCA and the GSRPCA is 20 for the 1st row (shadow removal), 50 for the 2nd and 3rd rows (salt & pepper noise) and 30 for the 4th row (corruption with a random patch). It can be seen that the GSRPCA reconstructs the original image with increased fidelity, even in the case of heavy corruption where the other methods fail. The reconstruction becomes sharper and more accurate as p and q approach 0, even though the number of principal components is kept low.

Fig. 3 plots the reconstruction error versus the number of principal components k for the cases where the images were corrupted by (a) 10% salt & pepper noise, (b) 30% salt & pepper noise and (c) a random patch of maximum size 40×40. The error is computed as $\|\hat{\mathbf{X}} - \mathbf{X}\|_1 / N \|\mathbf{X}\|_1$, where $\mathbf{X}$ is the original data matrix with the N vectorised images as columns and $\hat{\mathbf{X}}$ is the reconstructed one. Note that cases (b) and (c) correspond to particularly heavy corruption. Nevertheless, it can be seen that the GSRPCA outperforms other methods, especially when p and q approach 0, achieving good reconstruction rates for small k.

Finally, Table 1 shows the time performance of each algorithm (except PCA which is as fast as a single SVD) for the case where no corruption was considered and the number of components was set to $k = 10$. The implementation was done in MATLAB and the experiments were performed on an 8-core i7 Intel CPU at 3.40 GHz with 16 GB RAM. The GSRPCA with $p = q = 1$ is particularly efficient to compute as it takes only 48 iterations to converge and it outperforms the RPCA (the second fastest) by a factor of $\times 3.5$. This advan-

	RPCA	IRPCA	GSRPCA(1.0)	GSRPCA(0.5)	GSRPCA(0.1)
Time [sec]	3.562	401.968	1.048	21.809	69.017
Iterations	49	82	48	95	375

Table 1: Time in seconds and number of iterations until convergence for each algorithm. Experiment performed on the Extended Yale B database, without added noise. For PCA and GSRPCA, $k = 10$ principal components were used.

| Original | PCA | RPCA | IRPCA | $p = q = 1$ | $p = q = 0.5$ | $p = q = 0.1$ |

Figure 2: Comparison among different methods on the Extended Yale B database corrupted by noise. 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 $p = q \in \{1, 0.5, 0.1\}$ respectively.

tage is both due to the efficient version of the generalised p-shrinkage operator for $p = 1$ and due to the small size of required SVDs for a small k.

5 Conclusions

This work identified two key elements of robust low rank subspace recovery. Firstly, overcoming the convexity constraint for the objective function can lead to closer approximation of the original robust rank minimisation problem. Secondly, recovering the principal subspace places an upper bound to the rank constraint and allows for efficient algorithmic implementation which scales well to large problems. Exploiting the above, we introduced Generalised Scalable Robust PCA. We verified that the GSRPCA can achieve respectable results in computer vision applications such as facial image reconstruction, even under heavy corruptions that weaken the performance of state-of-the-art methods.

Acknowledgements

(a) 10% S&P noise   (b) 30% S&P noise   (c) Random patch of max 40×40
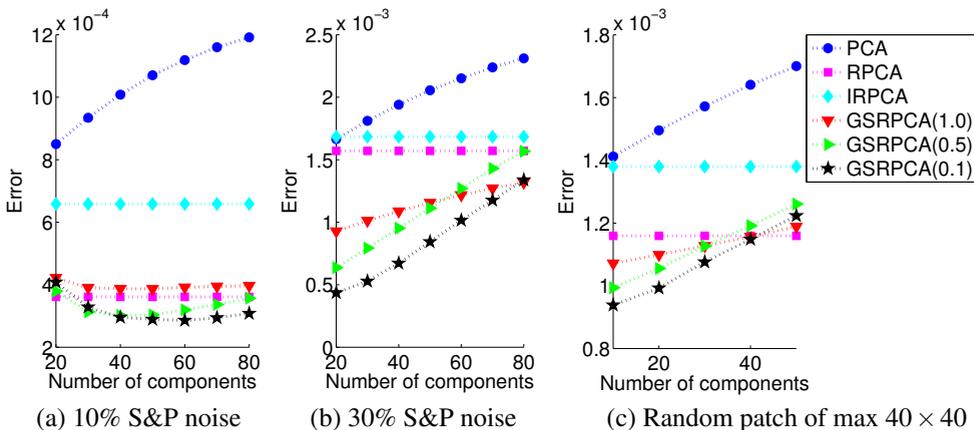
Figure 3: Reconstruction error for each method versus the number of principal components k on the Extended YALE B database with synthetic corruption. For RPCA and IRPCA the error is constant since they do not consider k. (a) 10% salt & pepper noise, (b) 30% salt & pepper noise and (c) random patch of maximum size 40×40 pixels.

References

[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] R. Basri and D. W. Jacobs. Lambertian reflectance and linear subspaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(2):218–233, 2003.

[3] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Athena Scientific, Belmont, MA, 2nd edition, 1996.

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

[5] D. Chen and L.-Z. Cheng. Image inpainting based on low-rank and joint-sparse matrix recovery. *Electronics Letters*, 49(1):35–36, 2013.

[6] D. Donoho. For most large underdetermined systems of equations, the minimal ℓ_1-norm near-solution approximates the sparsest near-solution. *Communications on Pure and Applied Mathematics*, 59(7):907–934, 2006.

[7] M. Fazel. *Matrix Rank Minimization with Applications*. PhD thesis, Dept. Electrical Engineering, Stanford University, CA, USA, 2002.

[8] 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 Trans. Pattern Anal. Mach. Intelligence*, pages 643–660.

[9] Gene H. Golub and Charles F. Van Loan. *Matrix Computations (3rd Ed.)*. Johns Hopkins University Press, Baltimore, MD, USA, 1996. ISBN 0-8018-5414-8.

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

[11] P.J. Huber, J. Wiley, and W. InterScience. *Robust statistics*. Wiley New York, 1981.

[12] K. C. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. Pattern Anal. Mach. Intelligence*, pages 684–698.

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

[14] F. Nie, H. Wang, H. Huang, and Ch. Ding. Joint schatten p-norm and ℓ_p-norm robust matrix completion for missing value recovery. *Knowledge and Information Systems*, pages 1–20, 2013.

[15] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma. Toward a practical face recognition system: Robust alignment and illumination by sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(2):372–386, 2012.

[16] J. Yan, M. Zhu, H. Liu, and Y. Liu. Visual saliency detection via sparsity pursuit. *IEEE Signal Processing Letters*, 17(8):739–742, 2010.

[17] H. Zou, T. Hastie, and R. Tibshirani. Sparse principal component analysis.