Recovering Joint and Individual Components in Facial Data

Christos Sagonas, Evangelos Ververas, Yannis Panagakis, and Stefanos Zafeiriou

Abstract—A set of images depicting faces with different expressions or in various ages consists of components that are shared across all images (i.e., joint components) and imparts to the depicted object the properties of human faces and individual components that are related to different expressions or age groups. Discovering the common (joint) and individual components in facial images is crucial for applications such as facial expression transfer. The problem is rather challenging when dealing with images captured in unconstrained conditions and thus are possibly contaminated by sparse non-Gaussian errors of large magnitude (i.e., sparse gross errors) and contain missing data. In this paper, we investigate the use of a method recently introduced in statistics, the so-called Joint and Individual Variance Explained (JIVE) method, for the robust recovery of joint and individual components in visual facial data consisting of an arbitrary number of views. Since, the JIVE is not robust to sparse gross errors, we propose alternatives, which are 1) robust to sparse gross, non-Gaussian noise, 2) able to automatically find the individual components rank, and 3) can handle missing data. We demonstrate the effectiveness of the proposed methods to several computer vision applications, namely facial expression synthesis and 2D and 3D face age progression in-the-wild.

Index Terms—Low-Rank, Sparsity, Facial Expression Synthesis, Face Age Progression, Joint and Individual Components.

1 INTRODUCTION

Facial images convey rich information, which can be perceived as a superposition of components associated with attributes, such as facial identity, expression, age etc. For instance, a set of images depicting expressive faces consists of components that are shared across all images (i.e., joint components) and imparts to the depicted object the properties of human faces. Besides joint components, an expressive face consists of individual components that are related to different expressions. Such individual components can be expression-specific deformation of face, i.e., deformations around lips and eyes in case of smiles. Similarly, a set of images depicting faces in different ages can be seen as a superposition of joint components that are invariant to the age and age-specific components that are individual to each age group (e.g., wrinkles). Consequently, being able to extract such joint and individual components from facial images is crucial for applications such as facial expression synthesis and face age progression [1], [2], [3], [4], [5], [6], among other visual data analysis tasks.

Extracting the joint components among data has created a wealth of research in statistics, signal processing, and computer vision. Two mathematically similar but conceptually different models underlie the bulk of the methodologies. In particular, the Canonical Correlation Analysis (CCA) [7] and its variants e.g., [8], [9] have been proposed for extracting linear correlated components among two or more sets of variables. Similarly, inter-battery factor analysis [10] and its extensions e.g., [11] determines the common factors among two sets of variables. The main limitation of the aforementioned methods is that they only recover the most correlated linear subspace of the data, ignoring the individual components among the different views or datasets.

The above mentioned limitation is alleviated by recent methods such as the Joint and Individual Variation Explained (JIVE) [12], the Common Orthogonal Basis Extraction (COBE) [13], and the Robust Correlated and Individual Component Analysis (RCICA) [14], which are briefly described in Section 2.

Besides the rich structure in facial visual data, images are subject to various types of errors, distortions, and noise. Common dense distortions such as ambient noise or quantization noise are of small magnitude and it is natural to assume that they follow a Gaussian distribution of small variance. Methods such as the CCA and its variants, the JIVE, and the COBE are stable in the presence of Gaussian noise.

Apart from these small but dense noises, there are gross errors that are sparsely supported but of large or even unbounded magnitude, such as the salt-and-pepper noise in imaging devices, occlusions in facial images, registration errors, or errors due incorrect localization and tracking. These errors rarely follow a Gaussian distribution and due to their sparse nature (i.e., the number of errors is bounded below some constant) are collectively referred to as sparse gross errors or noise. Except for the most recent RCICA, the COBE and JIVE rely on least squares error minimization and thus they are prone to gross errors and outliers [15]. That is, the estimated components can be arbitrarily away from the true ones. Hence, the problem of joint and individual components recovery is rather challenging when dealing with facial images and in general visual data captured under unconstrained (i.e., in-the-wild) conditions.

In this paper, we investigate the problem of recovering the joint and individual components from facial (and in general visual) data consisting of an arbitrary number of views, captured in-the-wild. Such data are therefore contaminated by sparse, gross, non-Gaussian noise and possibly contain missing values. To this end, we propose robust alternatives to the JIVE (coined collectively as Robust JIVE- RJIVE), where the components are estimated by employing the ℓ₁-norm. The ℓ₁-norm is suitable for robust estimation in the presence of sparse gross errors [15]. The contributions of the paper are summarized as follows:

• We propose a novel, general framework, the RJIVE in Section 3, for the robust recovering of joint and individual components from facial images.
components from multi-view data in the presence of sparse gross errors and possibly missing values. The proposed RJIVE decomposes the data into three terms: a low-rank matrix that captures the joint variation across views, low-rank matrices accounting for structured variation individual to each view, and a sparse matrix collecting the sparse gross errors. In particular, the RJIVE consists of 4 different models, namely, $\ell_1$-RJIVE, NN-$\ell_1$-RJIVE, SRJIVE, and RJIVE-M. In the $\ell_1$-RJIVE, the rank of both joint and individual components are user-defined, while in the NN-$\ell_1$-RJIVE the rank of each one of the individual components is automatically estimated via nuclear norm minimization. As opposed to the previous two models, the SRJIVE directly extracts the orthonormal bases of joint and individual components and improves their scalability. Finally, the RJIVE-M extends the SRJIVE in order to handle missing values. Based on the recovered joint and individual components from training data, two suitable optimization problems that extracts the corresponding modes of variation (i.e., joint and individual components) of unsee test samples, are proposed in Section 4.

To tackle the proposed optimization problems, algorithms based on the Alternating-Directions Method of Multipliers (ADMM) [17] are developed in Sections 3 and 4.

We demonstrate the applicability of the proposed methods in three challenging computer vision tasks, namely facial expression synthesis, face age progression in 2D images and 3D data captured in-the-wild. Experimental results corroborate the effectiveness of the proposed approach in Section 7.

Furthermore, a new challenging data-set of 19,000 images captured in-the-wild with annotations in terms of age, is introduced in Section 7.3 for age-invariant face verification.

Notation: Throughout the paper, scalars are denoted by lowercase letters, vectors (matrices) are denoted by lowercase (uppercase) boldface letters i.e., $x$, ($X$). $I$ denotes the identity matrix. The $j$-th column of $X$ is denoted by $x_j$.

Several norms and metrics will be used. The $\ell_1$ and the $\ell_2$ norms of $x$ are defined as $\|x\|_1 = \sum_i |x_i|$ and $\|x\|_2 = \sqrt{\sum_i x_i^2}$, respectively. $|\cdot|$ denotes the absolute value operator. The matrix $\ell_1$ norm is defined as $\|X\|_1 = \sum_i \sum_j |x_{ij}|$, and the Frobenius norm is defined as $\|X\|_F = \sqrt{\sum_{ij} x_{ij}^2}$, and the nuclear norm of $X$ (i.e., the sum of singular values of a matrix) is denoted by $\|X\|_*$. The vector (matrix) $\ell_0$ (quasi)-norm returns the total number of non-zero elements in a vector (matrix). The rank function is denoted by $\text{rank}()$.

The minimization of both the rank function and the $\ell_0$-norm are NP-hard [18], [19] problems. Consequently, the rank function and the $\ell_0$-norm are typically replaced by their convex surrogates [20], [21].

Operators: The solution of the several problems appeared in the paper rely on different (proximal) operators which are defined next. Let, for any matrix $X = U\Sigma V^T$ be the Singular Value Decomposition.

- Shrinkage operator [22]: $S_{\tau}[\sigma] = \text{sgn}(\sigma) \max(|\sigma| - \tau, 0)$.
- Singular Value Thresholding (SVT) operator [23]: $D_\tau = U S_{\tau} V^T$.
- Rank-svd operator: $Q_r [X] = [U(:, 1 : r) \Sigma(1 : r, 1 : r) V(:, 1 : r)^T]$.
- Procrustes operator: $P[D] = PR^T$ (given the rank-svd SVD of a matrix $D = \text{GPR}^r$).

## 2 Background

To make the paper self-contained, this section includes a brief review of the JIVE [12], COBE [13], and the RCICA [14].

### 2.1 Joint and Individual Variation Explained (JIVE)

The JIVE recovers the joint and individual components among $M \geq 2$ data-sets $\{X^{(i)} \in \mathbb{R}^{d^{(i)} \times J}, i = 1, 2, \ldots, M\}$, where $J$ is the number of samples of each data-set. In particular, each matrix is decomposed into two terms: a low-rank matrix $J^{(i)} \in \mathbb{R}^{d^{(i)} \times J}$ capturing joint structure between data-sets and a low-rank matrix capturing individual structure $A^{(i)} \in \mathbb{R}^{d^{(i)} \times J}$ to each data-set. That is, $X^{(i)} = J^{(i)} + A^{(i)}$, $i = 1, 2, \ldots, M$. Let $X$ and $J$, be $\sum_{i=1}^M d^{(i)} \times J$ matrices constructed by concatenation of the corresponding matrices i.e., $X = \{X^{(1)}T, X^{(2)}T, \ldots, X^{(M)}T\}T$, $J = \{J^{(1)}T, J^{(2)}T, \ldots, J^{(M)}T\}T$, the JIVE solves the rank-constrained least-squares problem [12]:

$$
\min_{\{A^{(i)}\}_{i=1}^M} \frac{1}{2} \left\| X - J - \left[ A^{(1)}T, \ldots, A^{(M)}T \right]T \right\|_F^2,
$$

s.t. $\text{rank}(J) = r$, $\text{rank}(A^{(i)}) = r^{(i)}$, $J A^{(i)} T = 0$ $i = 1, \ldots, M$

Problem (1) imposes rank constraints on joint and individual components and requires the rows of $J$ and $\{A^{(i)}\}_{i=1}^M$ to be orthogonal. The intuition behind the orthogonality constraint is that, sample patterns responsible for joint structure between data types are unrelated to sample patterns responsible for individual structure [12]. By adopting the least squares error, the JIVE assumes Gaussian distributions with small variance [15]. Such an assumption rarely holds in real word data, where gross non-Gaussian corruptions are in abundance. Consequently, the components obtained by employing the JIVE in the analysis of grossly corrupted data may be arbitrarily away from the true ones, degenerating their performance.

### 2.2 Common Orthogonal Basis Extraction (COBE)

A closely related method to the JIVE is the COBE which extracts the common and the individual components $M$ data-sets of the same dimensions by solving a set of least-squares minimization problems [13]. More specifically, each data-set $X^{(i)} \in \mathbb{R}^{d^{(i)} \times d^{(i)}}$ is factorized as $\Xi^{(i)} A^{(i)} T$ where a column of $\Xi^{(i)}$ signifies a latent variable to be found and $A^{(i)}$ signifies a matrix of weights. $\Xi^{(i)}$ is assumed to be decomposable in blocks as $[\Xi \Xi^{(i)}]$ where $\Xi \in \mathbb{R}^{n \times m}$, $\Xi^{(i)} \in \mathbb{R}^{n \times (d^{(i)} - m)}$ and $m \leq \min\{d^{(i)}, i = 1, \ldots, M\}$. In other words, $\Xi$ is assumed to be common to all...
factorizations and hence it presents joint structure while \( \tilde{\Xi}(i) \) is assumed to represent individual structure. Similarly, \( \tilde{A}(i) \) splits as \( \tilde{\Lambda}(i) \) and \( \tilde{\Theta}(i) \). The optimization problem of the COBE takes the following form:

\[
\min_{\Xi, \Theta(i)} \sum_{i=1}^{M} \left\| X(i) - \Xi \tilde{\Lambda}(i)^T - \Xi \tilde{\Theta}(i)^T \right\|_F^2, \\
\text{s.t.} \quad \Xi^T \Xi = I, \{ \Xi(i)^T \tilde{\Xi}(i) = I, \Xi(i)^T \tilde{\Xi}(i)^T = 0 \}_{i=1}^{M}.
\]  

(2)

Similarly to the JIVE, the usage of the least square error makes the COBE non-robust against sparse, non-Gaussian errors.

### 2.3 Robust Correlated and Individual Component Analysis (RCICA)

The goal of the RCICA [14] is to extract both the correlated and individual components between two known high-dimensional datasets or views namely, \( \{ X(i) \in \mathbb{R}^{d(i) \times J} \}_{i=1}^{M} \), in the presence of sparse noise (or errors).

To this end, the RCICA seeks a decomposition of each data matrix \( \{ X(i) \} \) into three terms: \( X(i) = C(i) + A(i) + E(i) \), \( i = 1, 2, C(i) \in \mathbb{R}^{d(i) \times J} \) and \( A(i) \in \mathbb{R}^{d(i) \times J} \) are low-rank matrices, with \( \text{rank}(C(i)) \leq k_c \) and \( \text{rank}(A(i)) \leq k_i \) and mutually independent columns, capturing the correlated and individual components, respectively and \( E(i) \in \mathbb{R}^{d(i) \times J} \) is a sparse matrix accounting for the sparse noise.

To find the correlated components \( C(i) \in \mathbb{R}^{d(i) \times J} \), the cost function of the Canonical Correlation Analysis (CCA) [7] is adopted. That is, by further decomposing the matrix \( \{ C(i) \}_{i=1}^{2} \) as: \( C(i) = U(i) V(i)^T X(i) \), the maximally correlated components are derived by minimizing the CCA cost, namely

\[
\frac{\lambda_i}{2} \| V(i)^T X(i) - V(2)^T X(2) \|_F^2.
\]

Here, \( U(i) \) are orthonormal basis, transforming the correlated components back to the observation space \( X(i) \). Since, the column space of the individual components \( A(i) \) is desired to be orthogonal to the one of the correlated components we have to enforce \( \{ Q(i) U(i)^T \}_{i=1}^{M} = 0 \), where \( Q(i) \) are column orthonormal basis spanning the column space of the individual components \( A(i) \), that is \( A(i) = Q(i) H(i) \).

Consequently, a natural estimator accounting for the upper-bounded rank of the correlated and independent components and the sparsity of \( \{ E(i) \}_{i=1}^{2} \) is to minimize the objective function of CCA, i.e., \( \frac{1}{2} \| V(i)^T X(i) - V(2)^T X(2) \|_F^2 \) as well as the rank of \( \{ C(i) = U(i) V(i)^T X(i), A(i) = Q(i) H(i)^T \}_{i=1}^{M} \) and the number of nonzero entries of \( \{ E(i) \}_{i=1}^{M} \), measured by the \( \ell_0 \)-quasi norm, e.g., [22]. To avoid the NP-hardness of rank and \( \ell_0 \)-norm minimization, the nuclear- and the \( \ell_1 \)-norms are typically adopted as surrogates to rank and \( \ell_0 \)-norm, respectively [20], [21]. By employing the unitary invariance of the nuclear norm e.g., \( \| V(i)^T V(i)^T \|_1 = \| V(i)^T \|_1 \), the optimization problem of RCICA is formulated as the following constrained non-linear one:

\[
\min_{V} \sum_{i=1}^{2} \left\| V(i)^T \right\|_1 + \lambda_1 \| H(i) \|_1 + \lambda_i \| E(i) \|_1, \\
\text{s.t.} \quad (i) \quad X(i) = U(i) V(i)^T X(i) + Q(i) H(i) + E(i), \\
(ii) \quad V(i)^T X(i) X(i)^T V(i) = I, \\
(iii) \quad U(i)^T U(i) = I, \quad Q(i)^T Q(i) = I, \\
(iv) \quad Q(i)^T U(i) = 0, \quad i = 1, 2,
\]  

(3)

where the positive parameters \( \lambda_c, \lambda_1, \lambda_2, \lambda_i \) and \( \lambda_i \) control the correlation, rank and sparsity of the derived spaces and \( \mathcal{V} = \{ U(i), V(i), Q(i), H(i), E(i) \}_{i=1}^{M} \) collects the optimization variables. The constraints (ii) in (3) have been adopted from the CCA [7] while the constraints (iii) and (iv) ensure that both the recovered correlated and individual components are linearly independent.

Although the RCICA is robust to sparse, non-Gaussian error, its extension to more than two data-sets is not trivial due to the orthogonality among the correlated and individual components and column of orthonormality of the basis matrices \( U(i) \) and \( Q(i) \), \( i = 1, 2, \ldots, M \), with \( M \) being the number of different views. This makes the resulting optimization problem highly-nonlinear and hence difficult to solve.

### 3 Robust JIVE

Consider data consisting of \( M \) views \( \{ X(i) \in \mathbb{R}^{d(i) \times J} \}_{i=1}^{M} \), with \( X_j \in \mathbb{R}^{d(i)}, \ j = 1, \ldots, J \) being a vectorized (visual) data sample, possibly contaminated by gross, sparse errors. The goal of the RJIVE is to robustly recover the joint components which are shared across all views as well as the components which are deemed individual for each view. That is:

\[
X = J + [A(1)^T, \ldots, A(M)^T]^T + E, \\
\]  

(4)

where \( X = [X(1)^T, \ldots, X(M)^T]^T \in \mathbb{R}^{q \times J}, \ J = [J(1)^T, \ldots, J(M)^T]^T \in \mathbb{R}^{q \times J}, \ A(i) \in \mathbb{R}^{d(i) \times J} \in_{i=1}^{M} \), \( q = d(1)+\ldots+d(M) \), are low-rank matrices capturing the joint and individual variations, respectively and \( E \in \mathbb{R}^{q \times J} \) denotes the error matrix accounting for the gross, but sparse, non-Gaussian noise. In order to ensure the identifiability of (4), the joint and common components should be mutual incoherent, i.e., \( JA(i)^T = 0 \) \in_{i=1}^{M}. Assuming that the number of errors is bounded below some constant, the number of errors in the estimated components is similarly bounded and hence a natural estimator accounting for the sparsity of the error matrix \( E \), is to minimize the number of the nonzero entries of \( E \) measured by the \( \ell_0 \)-quasi norm [22]. However as in case of the RCICA, to make the problem computationally tractable the \( \ell_0 \)-norm is replaced by its convex surrogate, namely the \( \ell_1 \)-norm. Thus, the joint and individual components as well as the sparse error are recovered by solving the following constrained non-linear optimization problem:

\[
\min_{J,\{A(i)\}_{i=1}^{M}} \left\| X - J - [A(1)^T, \ldots, A(M)^T]^T \right\|_1, \\
\text{s.t.} \quad \text{rank}(J) = r, \text{rank}(A(i)) = r(i), \text{rank}(E) = 0 \in_{i=1}^{M},
\]  

(5)

Clearly, (5) is a robust extension to JIVE [12] and requires an estimation for the rank of both joint and individual components. However, in practice those \( (M+1) \) values are unknown and difficult to estimate since an extensive tuning procedure is required. To alleviate this issue, we propose a variant of (5) which is able to determine the optimal ranks of individual components directly. By assuming that the actual ranks of individual components are upper bounded i.e., \( \text{rank}(A(i)) \leq K(i) \in_{i=1}^{M} \), problem (5) is relaxed to the following one:

\[
\min_{J,\{A(i)\}_{i=1}^{M}} \lambda \left\| X - J - [A(1)^T, \ldots, A(M)^T]^T \right\|_1 + \sum_{i=1}^{M} \left\| A(i) \right\|_1, \text{s.t.} \quad \text{rank}(J) = r, \text{rank}(E) = 0 \in_{i=1}^{M} \\
\]  

(6)
where the rank function is replaced by its convex envelope, namely the nuclear norm and $\lambda > 0$ is a regularizer.

### 3.1 Optimization Algorithms

In this section, algorithms for solving (5) and (6) are developed.

To solve (5), the Alternating-Direction Method of Multipliers (ADMM) [17] is employed. To this end, problem (5) is reformulated to the following separable one:

$$\min_{J, \{A(i)\}_{i=1}^{M}, E} \|E\|_1,$$

s.t. $X = J + \left[A(1)^T, \ldots, A(M)^T\right]^T + E,$

$$\text{rank}(J) = r, \{\text{rank}(A(i)) = r(i), JA(i)^T = 0\}_{i=1}^{M},$$

where $E$ is an auxiliary variable. To solve (7), the corresponding augmented Lagrangian function is given by:

$$L(J, \{A(i)\}_{i=1}^{M}, E, L) = \|E\|_1 - \frac{1}{2\mu} \|L\|_F^2 + \frac{\mu}{2} \left\|X - J - \left[A(1)^T, \ldots, A(M)^T\right]^T - E - \frac{L}{\mu}\right\|_F^2,$$

where $L$ is the Lagrange multiplier matrix related to the equality constraint in (7), and $\mu$ is a positive parameter. Then, by employing the ADMM, (8) is minimized with respect to each variable in an alternating fashion and finally the Lagrange multipliers $L$ are updated. The ADMM solver of (9) is outlined in Algorithm 1. Algorithm 1 terminates when $\|X - J_{t+1} - \left[A(1)^T, \ldots, A(M)^T\right]^T - E_t - \frac{L_t}{\mu}\|_F^2$ is less than a predefined threshold $\epsilon$ or the number of iterations reach a maximum value.

**Algorithm 1: ADMM solver for (7) ($\ell_1$-RJIVE).**

**Input:** Data $\{X(i)\} \in \mathbb{R}^{d(i) \times J} \_{i=1}^{M}$, Rank of joint component $r$. Ranks of individual components $\{r(i)\}_{i=1}^{M}$, Parameter $\rho$.

**Output:** Joint component $J$, individual components $\{A(i)\}_{i=1}^{M}$, Parameter $\mu$.

**Initialize:** Set $J_0$, $\{A(i)\}_{i=1}^{M}$, $E_0$, $L_0$ to zero matrices, $t = 0$, $\mu_0 > 0$.

1. $X = \left[X(1)^T, \ldots, X(M)^T\right]^T$.
2. **while not converged**
   3. $J_{t+1} = \mathbb{Q}_r[M], |U|, \Sigma, V] = \text{svd}(M)$;
   4. $P = I - V(:, 1 : r)V(:, 1 : r)^T$;
   5. for $i = 1 : M$ do
      6. $A(i)_{t+1} = \mathbb{Q}_r[A(i)_t - E_t + \mu_t^{-1}L_t]$;
   7. end
   8. $E = S_{\frac{\mu_t}{\rho}} \left[X - J_{t+1} - \left[A(1)^T, \ldots, A(M)^T\right]^T - \frac{L_{t+1}}{\mu_t}\right]$;
   9. $L_{t+1} = \mu_t \left[X - J_{t+1} - \left[A(1)^T, \ldots, A(M)^T\right]^T - E_t + X\right]$;
   10. $\mu_{t+1} = \min(\mu_t, 10^{-3})$;
   11. $t = t + 1$;
   12. end

**To solve problem (6) via ADMM, we firstly reformulate it as:**

$$\min_{J, \{A(i)\}_{i=1}^{M}, E} \sum_{i=1}^{M} \left\|A(i)^T\right\|_F + \lambda \|E\|_1,$$

s.t. $X = J + \left[A(1)^T, \ldots, A(M)^T\right]^T + E,$

$$\text{rank}(J) = r, \{\text{rank}(A(i)) = r(i), JA(i)^T = 0\}_{i=1}^{M},$$

where $\{A(i)\}_{i=1}^{M} \in \mathbb{R}^{d(i) \times J}$ are auxiliary variables and the corresponding constraints, respectively. The ADMM solver of (9) is wrapped up in Algorithm 2 where $F$, $\{Y(i)\}_{i=1}^{M}$ are the Lagrange multipliers related to the equality constraints in (9), and $\mu$ is a positive parameter. A similar to Algorithm 1 convergence criterion is employed. The augmented Lagrangian function of (9) as well as the derivation of proposed Algorithm can be found in supplementary material.

**Algorithm 2: ADMM solver of (9) ($\ell_1$-RJIVE).**

**Input:** Data $\{X(i)\} \in \mathbb{R}^{d(i) \times J} \_{i=1}^{M}$, Rank of joint component $r$. Ranks of individual components $\{r(i)\}_{i=1}^{M}$, Parameter $\rho$.

**Output:** Joint component $J$, individual components $\{A(i)\}_{i=1}^{M}$, Parameter $\mu$.

**Initialize:** Set $J_0$, $\{A(i)\}_{i=1}^{M}$, $R(i)_0$, $Y(i)_0$, $M_0$, $F_0$, $S_0$ to zero matrices, $t = 0$, $\mu_0 > 0$.

1. $X = \left[X(1)^T, \ldots, X(M)^T\right]^T$;
2. **while not converged**
   3. $J_{t+1} = \mathbb{Q}_r[M], |U|, \Sigma, V] = \text{svd}(M)$;
   4. for $i = 1 : M$ do
      5. $A(i)_{t+1} = \frac{X(i) - J_{t+1} - Y(i)_{t+1} - F(i) - \mu_t R(i)_{t+1}}{\mu_t}$;
   6. $R(i)_{t+1} = D_{\frac{\mu_t}{\rho}} \left[A(i)_{t+1} - Y(i)_{t+1}\right]$;
   7. $Y(i)_{t+1} = X(i)_{t+1} + \mu_t (R(i)_{t+1} - A(i)_{t+1})$;
   8. end
   9. $E_{t+1} = \frac{X - \left[A(1)^T, \ldots, A(M)^T\right]^T - E_t + F_t}{\mu_t}$;
   10. $F_{t+1} = \frac{X - \left[A(1)^T, \ldots, A(M)^T\right]^T - E_t + X}{\mu_t}$;
   11. $\mu_{t+1} = \min(\mu_t, 10^{-3})$;
   12. $t = t + 1$;
   13. end

### 4 RJIVE-Based Reconstruction

Having recovered the individual and common components of the $M$ views or different data-sets during training, we can exploit them in order to extract the joint and individual modes of variations of a test sample. For instance, the components recovered by applying the RJIVE on a set of facial images of $M$ different expressions can be utilized in order reconstruct $M$ expressive images $\{Y(i)\}_{i=1}^{M}$ of an input face $t$. The key motivation here, is that the expression-related patterns of the image $t$ in the expression $(i)$ lie in a linear subspace spanned by $D(i) \in \mathbb{R}^{d(i) \times W(i)}$, where $D(i)$ has been obtained by applying the SVD onto extracted $A(i)$ components. Thus, the expression-related (individual) part of the test image $t$ in expression $(i)$ can be represented as a linear combination of the orthonormal
bases $D^{(i)}$, i.e., $y_{\text{individual}}^{(i)} \approx D^{(i)}c^{(2)}$ with $c^{(2)} \in \mathbb{R}^{W_A^{(i)} \times 1}$ being a sparse coefficient vector. Similarly, the joint part $y_{\text{joint}}^{(i)}$ is expressed as a linear combination of the orthonormal bases $B^{(i)} \in \mathbb{R}^{d^{(i)} \times W_A^{(i)}}$ extracted from the corresponding joint component $J^{(i)}$, i.e., $y_{\text{joint}}^{(i)} \approx B^{(i)}c^{(1)}$, $c^{(1)} \in \mathbb{R}^{W_B^{(i)} \times 1}$. Thus, the expressive image $y^{(i)}$ of the unseen input face $t$ is reconstructed by solving the following constrained optimization problem:

$$\begin{align*}
\min_{\{e^{(n)}_i, v^{(n)}_i\}^{2}_n, \gamma} & \sum_{n=1}^{2} \left\| v^{(n)}_i \right\|_1 + \lambda \left\| e_i \right\|_1, \\
\text{s.t.} & \{v^{(n)}_i = e^{(n)}_i \}^{2}_n, \\
& t = B^{(i)}c^{(1)} + D^{(i)}c^{(2)} + e_i, \quad y = B^{(i)}c^{(1)} + D^{(i)}c^{(2)}
\end{align*} \tag{10}$$

where $\lambda$ is a positive parameter that balances the norms, $v^{(1)}$, $v^{(2)}$ are auxiliary variables which are employed in order to make the problem separable, $\gamma$ corresponds to the non-negative clean reconstruction, and $e_i$ is an error term accounting for the gross, non-Gaussian sparse noise. Equation (10) resembles the dense error correction model proposed in [24], which is suitable for guaranteed recovery of sparse representations from high-dimensional measurements, such as images of high resolution (e.g., 22000 pixels in this paper) in the presence of noise. The ADMM solver of (10) is outlined in Algorithm 3. Algorithm 3 terminates when $\left\| t - U_J(c^{(1)}_0 - U_Ac^{(2)} + e + 1) \right\|_2^2 / \left\| t \right\|_2^2$ is less than a predefined threshold $\epsilon$ or the number of iterations reached. The augmented Lagrangian function of (10) can be found in supplementary material.

**Algorithm 3:** ADMM-based solver of (10).

**Input:** Input sample $t$. Orthonormal bases $B^{(i)} \in \mathbb{R}^{d^{(i)} \times W_A^{(i)}}$, $D^{(i)} \in \mathbb{R}^{d^{(i)} \times W_B^{(i)}}$. Parameters $\lambda, \rho$.

**Output:** Clean reconstructed image $y$, $c^{(1)}$, $c^{(2)}$.

**Initialize:** Set $\{v^{(n)}_i, e^{(n)}_i\}^{2}_n$, $\{h^{(n)}_0\}^{2}_n$, $\{h^{(n)}_0\}^{2}_n$, $\{h^{(n)}_0\}^{2}_n$, $y_0$, and $e_0$ to zero vectors, $t = 0$, $\mu_0 > 0$.

1. **while not converged do**
2. 2 for $n = 1:2$ do
3. 3 \[ v^{(n)}_i = S_{\frac{\mu_0}{\mu_n}} \left[ c^{(n)}_i - \frac{h^{(n)}_i}{\mu_n} \right]; \]
4. 4 end
5. $t_1 = t - D^{(i)}c^{(2)}_i + e + h^{(3)}_i / \mu_t$;
6. $t_2 = y - D^{(i)}c^{(2)}_i + h^{(3)}_i / \mu_t$;
7. $c^{(1)}_t = B^{(i)T}(t_1 + t_2) + v^{(1)}_i + h^{(1)}_i / \mu_t$;
8. $t_1 = t - B^{(i)}c^{(1)}_i - e + h^{(3)}_i / \mu_t$;
9. $t_2 = y - B^{(i)}c^{(1)}_i + h^{(3)}_i / \mu_t$;
10. $c^{(2)}_t = D^{(i)T}(t_1 + t_2) + v^{(2)}_i + h^{(2)}_i / \mu_t$;
11. $y_{t+1} = \max\{B^{(i)}c^{(1)}_i + D^{(i)}c^{(2)}_i - h^{(4)}_i / \mu_t, 0\}$;
12. $e_{t+1} = S_{\frac{\mu_0}{\mu_1}} \left[ t - B^{(i)}c^{(1)}_i - D^{(i)}c^{(2)}_i - h^{(4)}_i / \mu_t \right]$;
13. $h^{(1)}_i = h^{(1)}_i + \mu_t (v^{(1)}_i - c^{(1)}_i)$;
14. $h^{(2)}_i = h^{(2)}_i + \mu_t (v^{(2)}_i - c^{(2)}_i)$;
15. $h^{(3)}_i = h^{(3)}_i + \mu_t (t - B^{(i)}c^{(1)}_i - D^{(i)}c^{(2)}_i - e_{t+1})$;
16. $h^{(4)}_i = h^{(4)}_i + \mu_t (y - B^{(i)}c^{(1)}_i - D^{(i)}c^{(2)}_i)$;
17. $\mu_{t+1} = \min(\mu_0, 10^7)$;
18. **end**

5 **Scalable RJIVE**

The computational complexity of the vanilla JIVe as well as the $\ell_1$-RJIVE and NN-$\ell_1$-RJIVE at each iteration is $O(\max(qJ, qJ^2)) + \sum_{i=1}^{M} O(\max(q\ell_i^2, J, d^{(i)}J^2)) = O(\max(qJ, qJ^2))$, due to the SVD. Clearly, this is computationally prohibitive when dimension of the images $d^{(i)}$ becomes very large, e.g., 22500 in our case. To alleviate the aforementioned computational complexity issue and at the same time learn the orthonormal bases that are used for reconstruction, we propose to factorize the matrices $J\{A^{(i)}\}_{i=1}^{M}$ as products of orthonormal basis matrices $B \in \mathbb{R}^{d_{\infty} \times \cdots \times d(M) \times W_A}$, $B^{T}B = I$, $\{D^{(i)} \in \mathbb{R}^{d^{(i)} \times W_B^{(i)}} \}_{i=1}^{M}$, and low-rank coefficients matrices $G\{C^{(i)}\}_{i=1}^{M}$ such that $J = BG$ and $\{A^{(i)} = D^{(i)}C^{(i)}\}_{i=1}^{M}$. It can be easily shown that the constraints are now written as $\left\{JA^{(i)}\}_{i=1}^{M} = GC^{(i)} = 0$ and rank$(J) = \text{rank}(BG) = \text{rank}(G) = r$. In addition, due to the unitary invariance property of the nuclear norm we have $\left\| \{A^{(i)}\}_{i=1}^{M} \right\|_*$ = $\left\| \{D^{(i)}C^{(i)}\}_{i=1}^{M} \right\|_*$ $\leq \left\| \{C^{(i)}\}_{i=1}^{M} \right\|_*$. Thus, by incorporating the factorizations of joint and individual components the optimization problem now is as follows:

$$\begin{align*}
\min_{\{D^{(i)}, C^{(i)}\}_{i=1}^{M}, E} & \sum_{i=1}^{M} \left\| \Delta^{(i)} \right\|_*, \\
\text{s.t.} & X = BG + \left[ \left( D^{(1)}C^{(1)} \right)^T \cdots \left( D^{(M)}C^{(M)} \right)^T \right]^T + E, \\
& \text{rank}(G) = r, B^{T}B = I, \\
& \{\Delta^{(i)} = C^{(i)}JG^{(i)T} \}_{i=1}^{M} = 0, \\
& \text{rank}(A^{(i)}) = \text{rank}(I) = r \quad \forall \quad i.
\end{align*} \tag{11}$$

where $\{\Delta^{(i)} \in \mathbb{R}^{d^{(i)} \times J} \}_{i=1}^{M}$, $\{\Delta^{(i)} = A^{(i)}\}_{i=1}^{M}$ are auxiliary variables and the corresponding constraints, respectively.

The ADMM solver of the proposed SJIVe method is outlined in Algorithm 4, where $\Gamma$ and $\{Z^{(i)}\}_{i=1}^{M}$ are the Lagrangian multipliers related to the equality constraints of (11) (the Lagrange function corresponds to problem (11) can be found on supplementary material).

The computational complexity of Algorithm 4 is dominated by the cost of the SVD involved in the computation of SVT and Procrustes operators in Steps 4 and 5, respectively. Thus, the computational complexity of each iteration is $O(\max(W_A^2 J, W_A J^2))$ and $O(\max(qW_A^2, qW_A^2))$, respectively. Given that $W_A \ll q = d_1 \cdots d(M)$ (in this paper $q = 225000$ and $W_A \leq 600$), which implies $W_A J + qW_A \ll qJ$, the proposed scalable version of JIVE, i.e., the SJIVe has a significantly reduced computational cost compared to that of JIVE and RJIVE.

Regarding the convergence of the presented Algorithms 2, 1, 4 there is currently no theoretical proof known for the ADMM in problems with more than two blocks of variables. However ADMM has been applied successfully in non-linear optimization problems in practice [14, 25, 26, 27, 28]. In addition, the thorough experimental evaluation of the proposed methods, presented in Section 7, indicates that the obtained solutions are good for the data that RJIVE tested.

6 **RJIVE with missing values and application to face aging using 3D Morphable Models**

3D Morphable Models (3MMs) are statistical deformable models of the 3D shape and appearance of the human face [29]. Typi-
Algorithm 4: ADMM solver of (11) (Scalable NN-ℓ2-RJIVE, SRJIVE).

\textbf{Input} : Data \( \{ \mathbf{X}^{(i)} \} \in \mathbb{R}^{d^{(i)} \times J_{i}} \), Rank of joint component \( r \). Number of bases to be extracted from the Joint and Individual components \( W_{j} \) and \( W_{a}^{(i)} \), respectively. Parameter \( \rho \).

\textbf{Output} : Orthonormal Joint and Individual bases matrices \( \mathbf{B}, \{ \mathbf{D}^{(i)} \}_{i=1}^{M} \). Coefficient matrices \( \mathbf{G}, \{ \mathbf{C}^{(i)} \}_{i=1}^{M} \).

\textbf{Initialize:} Set \( \mathbf{G}_{0}, \mathbf{B}_{0}, \{ \mathbf{D}^{(i)}_{0}, \mathbf{C}^{(i)}_{0}, \mathbf{Z}^{(i)}_{0} \}_{i=1}^{M}, \mathbf{E}_{0}, \mathbf{\Gamma}_{0} \) to zero matrices, \( t = 0, \mu_{0} > 0 \).

\begin{align*}
1 \quad & \mathbf{X} = \left[ \mathbf{X}^{(1)^{T}}, \cdots, \mathbf{X}^{(M)^{T}} \right]^{T}; \\
2 \quad & \text{while not converged do} \\
3 \quad & \quad \mathbf{M} = \mathbf{B}_{t}^{T} \left( \mathbf{X} - \left[ \left( \mathbf{D}^{(1)}_{t} \mathbf{C}^{(1)}_{t} \right)^{T}, \cdots, \left( \mathbf{D}^{(M)}_{t} \mathbf{C}^{(M)}_{t} \right)^{T} \right] - \mathbf{E}_{t} + \mu_{t}^{-1} \mathbf{\Gamma}_{t} \right); \quad \left[ \mathbf{U}, \mathbf{\Sigma}, \mathbf{V} \right] = \text{svd}(\mathbf{M}); \\
4 \quad & \quad \mathbf{G}_{t+1} = \mathcal{Q}; \left[ \mathbf{M} \right]; \\
5 \quad & \quad \mathbf{B}_{t+1} = \mathcal{P} \left[ \left( \mathbf{X} - \left[ \left( \mathbf{D}^{(1)}_{t} \mathbf{C}^{(1)}_{t} \right)^{T}, \cdots, \left( \mathbf{D}^{(M)}_{t} \mathbf{C}^{(M)}_{t} \right)^{T} \right] - \mathbf{E}_{t} + \mu_{t}^{-1} \mathbf{\Gamma}_{t} \right) \mathbf{G}_{t+1}^{T} \right]; \\
6 \quad & \quad \mathbf{M} = \mathbf{X} - \mathbf{B}_{t+1} \mathbf{G}_{t+1} - \mathbf{E}_{t} + \mu_{t}^{-1} \mathbf{\Gamma}_{t}; \\
7 \quad & \text{for } n = 1 : M \text{ do} \\
8 \quad & \quad \mathbf{D}^{(i)}_{t+1} = \mathcal{P} \left[ \mathbf{M}^{(i)} {\mathbf{C}^{(i)}_{t}}^{T} \right]; \\
9 \quad & \quad \mathbf{C}^{(i)}_{t+1} = 0.5 \left( \mathbf{D}^{(i)}_{t+1} \mathbf{M}^{(i)} + \Delta^{(i)}_{t+1} + \mu_{t}^{-1} \mathbf{Z}^{(i)}_{t+1} \right) \left( \mathbf{I} - \mathbf{V} \mathbf{V}^{T} \right); \\
10 \quad & \quad \Delta^{(i)}_{t+1} = \mathcal{D}_{t} \left[ \mathbf{C}^{(i)}_{t+1} - \mu_{t}^{-1} \mathbf{Z}^{(i)}_{t+1} \right]; \\
11 \quad & \quad \mathbf{Z}^{(i)}_{t+1} = \mathbf{Z}^{(i)}_{t+1} + \mu_{t} \left( \Delta^{(i)}_{t+1} - \mathbf{C}^{(i)}_{t+1} \right); \\
12 \quad & \text{end} \\
13 \quad & \mathbf{E} = \frac{\mathbf{S}_{\rho}}{\mu_{t}} \left[ \mathbf{X} - \mathbf{B}_{t+1} \mathbf{G}_{t+1} - \left[ \left( \mathbf{D}^{(1)}_{t+1} \mathbf{C}^{(1)}_{t+1} \right)^{T}, \cdots, \left( \mathbf{D}^{(M)}_{t+1} \mathbf{C}^{(M)}_{t+1} \right)^{T} \right]^{T} + \mu_{t}^{-1} \mathbf{\Gamma}_{t} \right]; \\
14 \quad & \mathbf{\Gamma}_{t+1} = \mathbf{\Gamma}_{t} + \mu_{t} \left( \mathbf{X} - \mathbf{B}_{t+1} \mathbf{G}_{t+1} - \left[ \left( \mathbf{D}^{(1)}_{t+1} \mathbf{C}^{(1)}_{t+1} \right)^{T}, \cdots, \left( \mathbf{D}^{(M)}_{t+1} \mathbf{C}^{(M)}_{t+1} \right)^{T} \right]^{T} - \mathbf{E}_{t+1} \right); \\
15 \quad & \mu_{t+1} = \min(\rho, \mu_{t}, 10^{7}); \\
16 \quad & t = t + 1; \\
17 \quad & \text{end}
\end{align*}

cally, a 3DMM consists of PCA models for shape and appearance, as well as a camera projection model. More specifically, the shape model describes facial features that consist of \( N \) vertexes and is built by applying dense registration on a set of training meshes followed by a PCA [29]. An instance of the shape model can be expressed as the linear combination of a mean shape \( \overline{\mathbf{s}} \) and the subspace \( \mathbf{U}_{s} \) with parameters \( \mathbf{p} \) as \( \mathbf{s} = \overline{\mathbf{s}} + \mathbf{U}_{s} \mathbf{p} \). Similarly, the texture model is a linear PCA model that describes the texture associated with the shape model and can be constructed from captured 3D texture as in [29], or from single 2D images as in [30]. Moreover, the camera model maps a 3D mesh on the image plane, utilizing an orthographic or a perspective transformation \( \mathbf{W} (\mathbf{p}, \mathbf{c}) \), where \( \mathbf{c} \) are the camera parameters. Fitting a 3DMM into a new image is an iterative process, where the current model parameters (regarding shape, texture, and camera) are updated at each iteration. Typically, the fitting procedure is formulated as a Gauss-Newton optimization problem, where the main task is the minimization of the error between the input and the reconstructed image [30].

The extraction of 3D texture from single images commences with fitting a 3DMM on them. Then, a UV texture map is calculated by projecting the reconstructed 3D shape on the image plane and subsequently sampling the image at the locations of the shape’s vertexes. However, extracting the 3D texture from a 2D image in this way leads to incomplete 3D texture representations, mainly due to the presence of self-occlusions, especially when the person depicted in the image is not in a frontal pose. Therefore, data collected with the aforementioned technique include missing values. In order to specify the location (i.e., image coordinates) of the missing values in a UV texture image, a self-occlusion mask for each image is calculated by casting a ray from the camera to each vertex of the reconstructed shape. Each element of the extracted mask denotes whether a value of the UV texture map is missing or not (please see the first column of Figure 11 for an example of the extracted UV space).

Even thought, the RJIVE can robustly recover joint and individual components in the presence of sparse non-Gaussian errors of large magnitude, it is not able to handle data with missing values. To overcome this limitation of the RJIVE we propose the RJIVE-Missing (RJIVE-M). Consider \( M \) data-sets of different ages \( \{ \mathbf{X}^{(i)} \} \in \mathbb{R}^{d^{(i)} \times J_{i}} \), with \( \mathbf{x}^{(i)}_{j} \in \mathbb{R}^{d^{(i)}} \), being a vectorized form of the \( j \)-th gross corrupted and incomplete UV texture, \( j = 1, \ldots, J \), that displays a face within the \( i \)-th age group, \( i = 1, \ldots, M \). The goal of the RJIVE-M is not only to recover the joint and individual components but also to perform completion on the UV textures with missing values. To this end, problem (11) is reformulated to the following one:

\[
\min_{\mathbf{B}, \mathbf{G}, \{ \mathbf{D}^{(i)}, \mathbf{C}^{(i)}, \mathbf{\Delta}^{(i)} \}_{i=1}^{M}, \mathbf{E}} \sum_{i=1}^{M} \left\| \mathbf{\Delta}^{(i)} \right\|_{2} + \lambda \left\| \mathbf{W} \circ \mathbf{E} \right\|_{1}, \\
\text{s.t.} \quad \mathbf{X} = \mathbf{BG} + \left[ \left( \mathbf{D}^{(1)} \mathbf{C}^{(1)} \right)^{T}, \cdots, \left( \mathbf{D}^{(M)} \mathbf{C}^{(M)} \right)^{T} \right]^{T} + \mathbf{E}, \\
\text{rank}(\mathbf{G}) = r, \mathbf{B}^{T} \mathbf{B} = \mathbf{I}, \\
\{ \mathbf{\Delta}^{(i)} = \mathbf{C}^{(i)}, \mathbf{GC}^{(i)^{T}} = \mathbf{0}, \mathbf{D}^{(i)^{T}} \mathbf{D}^{(i)} = \mathbf{I} \}_{i=1}^{M},
\]
\[
\{w^{(i)}_1, w^{(i)}_2, \ldots, w^{(i)}_J\} \in \{0, 1\}^{J}, \text{with } w^{(i)}_j \text{ being a vectorized form of the self-occlusion mask that corresponds to the } j\text{-th UV texture of the } i\text{-th data-set. The Algorithm for solving the proposed RJIJE-M problem is similar to the SRJIVE one and has the same complexity and convergence criterion. The only difference is in the updating step of the error matrix } E. \text{ More specifically, the following additional step is performed after executing the step 13 of the Algorithm 4: } E = W \odot E + W \odot [X - B_{t+1}G_{t+1} - \left(\left[D^{(1)}_1C^{(1)}_t\right]^T \cdots \left[D^{(M)}_tC^{(M)}_t\right]^T + \mu_{t}^{-1}I\right] .
\]

Similarly, the presented RJIJE-based reconstruction method can be also extended to handle missing values in a test image. To this end, given a test sample with missing values (e.g., face UV texture) and the vectorized form of the corresponding occlusion mask } w, \text{ problem (10) is extended to the following one:}
\[
\begin{align*}
\min_{\{c^{(n)}, w^{(n)}\}_{n=1}^{2}, y} & \sum_{n=1}^{2} \left\|v^{(n)}\right\|_1 + \lambda \left\|w \odot e\right\|_1, \\
\text{s.t.} & \{v^{(n)} = c^{(n)}\}_{n=1}^{2} \\
& t = B^{(i)}c^{(1)} + D^{(i)} c^{(2)} + e, \quad y = B^{(i)} e^{(1)} + D^{(i)} e^{(2)}
\end{align*}
\]

An ADMM-based solver similar to the Algorithm 3 is employed in order to solve problem (13). More specifically, the update step of the error vector performed in step 12 of the Algorithm 3 is followed by the following one: } e_{t+1} = w \odot e + W \odot \left[t - B^{(i)} c^{(1)}_{t+1} - D^{(i)} c^{(2)}_{t+1} + h^{(3)}_{t+1} \right].

### 7 Experimental Evaluation

The performance of the proposed RJIJE method is assessed on synthetic data corrupted by sparse, non-Gaussian noise (Section 7.1), as well as on data captured under constrained and in-the-wild conditions with applications to (a) facial expression synthesis, (b) 2D and (c) 3D face age progression.

| TABLE 1: Parameters used in the conducted experiments. |
|-------------|-----|-----------|---|---|---|
| Section | r | λ | \(W^{(i)}_1\) | \(W^{(i)}_2\) | λ | ε |
| 7.2.1 | 20 | 1 | 70 | 70 | 0.03 | 10^{-5} |
| 7.2.2 | 150 | \(\sqrt{\text{max}(g,J)}\) | 300 | 300 | 600 | 600 |
| 7.3 | 300 | \(\sqrt{\text{max}(g,J)}\) | 600 | 600 |

### 7.1 Synthetic

In this section, the ability of RJIJE to robustly recover the common and individual components of synthetic data corrupted by sparse non-Gaussian noise is tested. To this end, sets of matrices \(\{X^{(i)}_j = J^{(i)} + A^{(i)}_j + E^{(i)}_j\} \in \mathbb{R}^{g \times d_j} \times J\) of varying dimensions were generated. In more detail, a rank-\(r\) joint component } J is \(\in \mathbb{R}^{(q+d^2)^{J}}\) and \(\text{was created from a random matrix } X = \left[\begin{array}{l} X^{(1)} \quad X^{(2)} \end{array}\right]^T \in \mathbb{R}^{q \times J}. \text{Next, the orthogonal to } J \text{ rank-}\(r^{(1)}\), rank-\(r^{(2)}\) common components } A^{(1)}_j \text{ and } A^{(2)}_j \text{ were computed by } [A^{(1)}_j]^T A^{(2)}_j = (X - J_{j})(I - VV^T), \text{ where } V \text{ is the matrix formed from the first } r \text{ columns of the row space of } X. E^{(i)}_j \text{ is a sparse error matrix with } 20\% \text{ non-zero entries being sampled independently from } \mathcal{N}(0, 1). \text{ The Relative Reconstruction Error (RRE) of the joint and individual components achieved by both } \ell_1\text{-RJIJE and Nuclear-Norm regularized (NN-}\ell_1\text{-RJIJE) for a varying number of dimensions, joint and individual ranks, are reported in Table 2.}

The corresponding RRE obtained by JIVE [12], COBE [13], and RCICA [14] are also presented. As it can be seen, the proposed methods accurately recovered both the joint and individual components. It is worth mentioning that the NN-\(\ell_1\)-RJIJE successfully recovered all components by utilizing only the true rank of the joint component. In contrast, all the other methods required the knowledge regarding the true rank for both joint and individual components. Furthermore, the SRJIVE achieved same results to the NN-\(\ell_1\)-RJIJE by reducing the computation times more that five times. Based on the performance of SRJIVE on the synthetic data, we decided to exploit it in the experiments described bellow and referred to as RJIJE hereafter.

Furthermore, we tested the RJIJE on synthetic data contaminated by Gaussian error. The RJIJE, can implicitly handle data contaminated by Gaussian noise by vanishing the error term. That is by setting the regularizer \(\lambda\) in problems (7), (9), (11) \(\lambda \to \infty\) i.e. \(E = 0\). In such case, the Frobenius norms corresponds to the equality constraints \(X = J + [A^{(1)} \cdots A^{(M)}]^T + E, \quad X = BG + [(D^{(1)}C^{(1)}) \cdots (D^{(M)}C^{(M)})]^T + E\) appearing in the corresponding augmented Lagrangian functions are deemed as the appropriate regularizer for handling Gaussian noise. The RRE of all compared methods are reported in Table 2. As it can be seen, the proposed methods accurately recovered both the joint and individual components.

### 7.2 Facial Expression Synthesis

In this section, we investigate the ability of the RJIJE to synthesize a set of different expressions of a given facial image. Consider } \(M\) data-sets where each one contains images of different subjects that depict a specific expression. In order to effectively recover the joint and common components, the faces of each data-set should be put in correspondence. Thus, their \(N = 68\) facial.

2. Additional results can be found on supplementary material.
TABLE 2: Quantitative recovering results produced by JIVE [12], COBE [13], RCICA [14], $\ell_1$-RJIVE (7), and NN-$\ell_1$-RJIVE (9) under Gaussian and gross non-Gaussian noise. Each compared method was applied on the same data generated by utilizing each set of parameters. The average recovery accuracy and computation time (in CPU seconds) were computed by repeating the experiment 10 times.

<table>
<thead>
<tr>
<th>$(d^{(1)}, d^{(2)}, j, r, r^{(1)}, r^{(2)})$</th>
<th>Method</th>
<th>$\left{ \frac{|A^{(1)} - A^{(1)}|^2_F}{|A^{(1)}|^2_F} \right}$</th>
<th>$\left{ \frac{|A^{(2)} - A^{(2)}|^2_F}{|A^{(2)}|^2_F} \right}$</th>
<th>Time (in CPU seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>non-Gaussian</td>
<td>Gaussian</td>
<td>non-Gaussian</td>
<td>Gaussian</td>
</tr>
<tr>
<td></td>
<td>COBE</td>
<td>3.6403</td>
<td>1.0927</td>
<td>1.0975</td>
</tr>
<tr>
<td></td>
<td>JIVE</td>
<td>0.5424</td>
<td>1.3558e-04</td>
<td>0.9349</td>
</tr>
<tr>
<td></td>
<td>RCICA</td>
<td>5.5028e-08</td>
<td>3.5073e-08</td>
<td>6.7179e-07</td>
</tr>
<tr>
<td></td>
<td>$\ell_1$-RJIVE</td>
<td>5.1515e-08</td>
<td>4.3416e-08</td>
<td>3.3904e-08</td>
</tr>
<tr>
<td></td>
<td>SKJIVE</td>
<td>2.7770e-08</td>
<td>1.6564e-04</td>
<td>3.8706e-08</td>
</tr>
<tr>
<td>(500, 500, 500, 5, 10, 10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>COBE</td>
<td>4.9982</td>
<td>1.08555</td>
<td>1.1890</td>
</tr>
<tr>
<td></td>
<td>JIVE</td>
<td>0.8398</td>
<td>1.8880e-04</td>
<td>1.4810</td>
</tr>
<tr>
<td></td>
<td>RCICA</td>
<td>-</td>
<td>-</td>
<td>6.7260e-07</td>
</tr>
<tr>
<td></td>
<td>$\ell_1$-RJIVE</td>
<td>8.5033e-08</td>
<td>5.5423e-08</td>
<td>2.9260e-04</td>
</tr>
<tr>
<td></td>
<td>NN-$\ell_1$-RJIVE</td>
<td>9.3804e-08</td>
<td>7.6262e-08</td>
<td>1.1801e-04</td>
</tr>
<tr>
<td></td>
<td>SKJIVE</td>
<td>6.8905e-08</td>
<td>2.3496e-04</td>
<td>6.0017e-08</td>
</tr>
<tr>
<td>(1000, 1000, 1000, 10, 20, 20)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>COBE</td>
<td>6.9981</td>
<td>1.3564</td>
<td>1.3469</td>
</tr>
<tr>
<td></td>
<td>JIVE</td>
<td>1.3961</td>
<td>2.6525e-04</td>
<td>2.1977</td>
</tr>
<tr>
<td></td>
<td>RCICA</td>
<td>-</td>
<td>-</td>
<td>5.9359e-05</td>
</tr>
<tr>
<td></td>
<td>$\ell_1$-RJIVE</td>
<td>1.2305e-07</td>
<td>1.0512e-07</td>
<td>4.1133e-04</td>
</tr>
<tr>
<td></td>
<td>NN-$\ell_1$-RJIVE</td>
<td>8.8570e-08</td>
<td>9.1058e-08</td>
<td>5.6000e-04</td>
</tr>
<tr>
<td></td>
<td>SKJIVE</td>
<td>9.7434e-08</td>
<td>2.7074e-04</td>
<td>1.0117e-07</td>
</tr>
</tbody>
</table>

Fig. 2: Joint, individual components and error matrices produced by the compared JIVE and RJIVE methods.

landmark points are localized using the detector [31], [32] and subsequently employed to compute a mean reference shape. Then, the faces of each data-set are warped into corresponding reference shape by using the piecewise affine warp function $W(\cdot)$ [33]. After applying the RJIVE on the warped data-sets, the recovered components can be used for synthesizing $M$ different expressions of an unseen subject. To do that, the new (unseen) facial image is warped to reference frame corresponds to expression that we want to synthesize and subsequently is given as input to the solve of (10).

The performance of RJIVE in FES task is assessed by conducting inner- and cross-databases experiments on MPIE [34], CK+ [35], and in-the-wild facial images collected from the internet (ITW). The synthesized expressions obtained by RJIVE are compared to those obtained by the state-of-the-art BKRRR [36] method. In particular, the BKRRR is a regression-based method that learns a mapping from the ‘Neutral’ expression to the target ones. Then, given the ‘Neutral’ face of an unseen subject, new expressions are synthesized by employing the corresponding learnt regression functions. The performance of the compared methods is measured by computing the correlation between the vectorized forms of true images and the reconstructed ones.

7.2.1 Controlled Conditions

In the first experiment, 534 frontal images of MPIE database that depict 89 subjects under six expressions (i.e., ‘Neutral’, ‘Scream’, ‘Squint’, ‘Surprise’, ‘Smile’, ‘Disgust’) were employed to train both RJIVE and BKRRR. Then, all expressions of 58 unseen subjects from the same database were synthesized by using their images correspond to ‘Neutral’ expressions. In Figure 3(a) the average correlations obtained by the compared methods for the different expressions are visualized. As it can be seen the proposed RJIVE method achieves the same accuracy to BKRRR without learning any kind of mappings between the different expressions of the same subject. Specifically, the RJIVE extracts only the individual components of each expression and the common one.

Furthermore, the performance of both methods is compared by performing a cross-database experiment on CK+ database. More specifically, we employed the ‘Neutral’, ‘Smile’, and ‘Surprised’ images of MPIE for training purposes while images of 69 subjects (three images per subject) of CK+ were used as test ones. In Figure 3(b) we can see that RJIVE outperforms by a large margin the BKRRR. This is due to the fact that the BKRRR performs the regression based on how close is the unseen ‘Neutral’ face to the training ones. Thus, in cases that the unseen subjects (e.g., subjects of CK+) presents enough differences compared to the training ones (e.g., subjects of MPIE), the synthesized expressions are characterized as non-accurate. The synthesized expressions of subjects ‘014’ and ‘015’ from MPIE produced by the BKRRR and RJIVE are visualized in Figure 4. Clearly, the proposed method produces expressed images of higher quality compared to the BKRRR.

Finally, the recovering accuracy of JIVE and RJIVE in FES was also qualitatively assessed. To this end, the images used in the previous experiments contaminated by sparse error and subsequently given to the compared methods.

Figure 5 displays the obtained components and the corresponding error matrices. Clearly, the proposed RJIVE method successfully recovered all the components. It is worth mentioning that the RJIVE removes the sparse noise and outliers e.g., occlusions due to eyeglasses (please see red dotted boxes of Figure 5). Clearly, the JIVE is not able to cope with the additive noise and occlusions.

7.2.2 In-The-Wild Conditions

As an additional experiment, we collected from the internet 180 images depicting 60 subjects with ‘Surprise’, ‘Smile’, and
Fig. 3: Mean average correlation achieved by JIVE and BKRRR methods on (a) MPIE, (b) CK+, and (c) ITW databases.

Fig. 4: Synthesized expressions of MPIE’s subject (a) ‘014’ and (b) ‘015’ produced by the BKRRR and RJIVE methods.

Fig. 5: Joint, individual components and error matrices produced by the compared JIVE and RJIVE methods.

7.3 Face Age Progression In-The-Wild
7.3.1 2D age progression of an unseen subject
Face age progression consists in synthesizing plausible faces of subjects at different ages. It is considered as a very challenging task due to the fact that the face is a highly deformable object and its appearance drastically changes under different illumination conditions, expressions, and poses. Various databases that contain faces at different ages have been collected in the last couple of years [39], [40]. Although these databases contain huge number of images, they have some limitations including limited images for each subject that cover a narrow range of ages and noisy age labels, since most of them have been collected by employing automatic procedures (crawlers). In order to overcome the aforementioned problems, we collected a new data-set called Age In-the-Wild (ATW). More specifically, 19,000 images that depict 540 subjects from 0 to 100 years old were collected from the internet. Subsequently, each image was manually annotated in terms of age and identity of the depicted subject. On average, there are 36 images that span 55 years for each subject.

In order to train the RJIVE, the ATW was divided into $M = 10$ age groups: $0 − 3, 4 − 7, 8 − 15, 16 − 20, 21 − 30, 31 − 40, 41 − 50, 51 − 60, 61 − 70, and 71 − 100$. Then, following the same procedure as in FES task, the RJIVE was
employed to extract the joint and common components from the warped images. The performance of RJIVE in face age progression in-the-wild is qualitatively assessed conducting experiments on images from the FG-NET database [41]. To this end, we compare the performance of RJIVE with the Illumination Aware Age Progression (IAAP) method [1], Coupled Dictionary Learning (CDL) method [2], Deep Ageing with Restricted Boltzmann Machines (DARB) method [3], Craniofacial Growth (CG) [4] model, Exemplar-based Age Progression (EAP) [5] method, Face Transformer (FT Demo) [42], and Recurrent Face Aging (RFA) method [6]. In Figures 7, 8 progressed images produced by the compared methods are depicted. Note, that all the progressed faces have been warped back and fused with the actual ones.

Figure 9 depicts faces synthesized by the DARB, IAAP, and RJIVE methods. By observing the results, it can be clearly seen that the identity information is not preserved in case of DARB. In particular, the progressed faces of all subjects for a specific age group are very similar between them. It looks like all of them were created by transferring the skin colour from the input image to the same mean appearance. Instead, the identity information remains in the faces produced by the proposed RJIVE method. Finally, progressed example faces in all the age-groups produced the RJIVE are visualized in Figure 10.

7.3.2 3D age progression of an unseen subject

Here, the ability of the proposed RJIVE-M method to perform 3D face age progression is demonstrated. Similarly to the 2D face age progression experiments presented previously, the ATW database was divided into $M = 6$ age groups (21-30, 31-40, 41-50, 51-60, 61-70, 71+) and used to train the RJIVE-M. In order to acquire the 3D training data for this task the 3DMM-ITW [30] was employed. The optimal shape and camera parameters were extracted by fitting the model to each one of the images of all age groups. In order to recover 3D shapes of high quality, we used the age and gender specific version of the LSFM shape model introduced in [43] in order to describe identity and the blendshapes of [44] in order to describe facial expressions. After recovering the 3D shape of each face, we computed the self-occlusion mask by using ray-tracing (see first row of Figure 11). Then, the completed joint and individual components of the grossly corrupted and incomplete UV textures were obtained by employing the RJIVE-M. The joint components obtained by applying a variant of JIVE with missing values i.e., JIVE-M and the RJIVE-M on UV textures are displayed in Figure 11. By observing the results, we can clearly see that the RJIVE was successfully removed the occlusions produced from eyeglasses and fingers in all images. This is attributed to the fact that the matrix $\ell_1$ was adopted in RJIVE, which effectively handles sparse noise of possibly large magnitude.

Similarly, to the 2D face aging experiment we can apply the RJIVE-M to the recovered UV maps to learn components that can be used to age the UV texture of a test unseen subject. Since, the 3D shapes are produced by the LSFM model they neither have missing values nor are contaminated by noise. Hence, for training aging components for the 3D shape we used standard JIVE.

In the test phase, the 3D shape of the test face is obtained by using the 3DMM-ITW algorithm [30]. Then, the UV texture and the corresponding self-occlusion mask are computed by em-
ploying the recovered 3D shape. The progression of the texture of the test subject in an age group is obtained by solving the problem (13) (for the shape we use the problem in (10)). Progressed unseen subjects in all age groups, projected back in the image plane, are visualized in Figure 12. After calculating a progressed 3D texture image and 3D shape the result face model is projected back in the image plane using the camera parameters initially acquired by fitting the 3DMM-ITW in the test image.

Figure 13 presents additional results that demonstrate the ability of the RJIVE-M to perform not only age progression but also completion. For each subject the original and two side poses are depicted. The extracted by the 3DMM-ITW 3D face model of the input image is displayed on the first row. By observing the results it becomes obvious that due to the self-occlusions, the

Fig. 8: Progressed faces produced by the compared methods on the FG-NET database.

Fig. 9: Comparisons between the IAAP, DARB, and RJIVE methods.

Fig. 10: Progressed faces produced by the proposed RJIVE method.

Fig. 11: Input images and corresponding joint components produced by the compared JIVE-M and RJIVE-M methods. As it can be observed the proposed method is able to remove occlusions produced by fingers and glasses.
instance of the 3D model with pose different to the input one, contains huge areas of missing values (black color). This is not the case for the progressed and completed results produced by the RJIVE-M (second row). As it can be seen, the completion of the regions with missing data is accurate which proves the significant representational power of the bases extracted from RJIVE-M.

Fig. 13: Progressed and completed 3D texture images, produced by the proposed RJIVE-M method. The 3D face models are visualized in the original and two side poses, so as the missing and the completed data to be visible.

7.3.3 Age-invariant face verification in-the-wild

The performance of the RJIVE is also quantitatively assessed by conducting age-invariant face verification experiments. Following the successfully used verification protocol of the LFW database [45], we propose four new age-invariant face verification protocols based on the proposed ATW database. Each one of the protocols was created by splitting the ATW database into 10 folds, with each fold consisting of 300 intra-class pairs and 300 inter-class pairs. The essential difference between these protocols is that in each protocol the age difference of each pair’s faces is equal to a predefined value i.e., {5 ages, 10 ages, 20 ages, 30 ages}.

In order to assess the performance of RJIVE, the following procedure was performed. For each fold of a specific protocol the training images were split into $M = 10$ age-groups and subsequently the RJIVE was employed on their warped version in order to extract the joint and individual components. All images of each training pair were then progressed into $M = 10$ age groups resulting into 10 new pairs. The progressed images of six subjects are depicted in Figure 10. As we wanted to represent each pair by using a single feature, gradients orientations were extracted from the corresponding images and subsequently the mean value of their cosine difference was employed as the pair’s feature. $M$ different Support Vector Machines (SVM) were trained by utilizing the extracted features. Finally, the scores produced by all the SVMs were lately fused by using SVM. In Figure 14, Receiver Operating Characteristic (ROC) curves computed based on the 10 folds of each one of the proposed protocols are depicted. The corresponding mean classification accuracy and Area Under Curve (AUC) are reported in Table 3. In order to assess the

<table>
<thead>
<tr>
<th>Protocol</th>
<th>AUC</th>
<th>Accuracy</th>
<th>AUC</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 years</td>
<td>0.686</td>
<td>0.637</td>
<td>0.646</td>
<td>0.609</td>
</tr>
<tr>
<td>10 years</td>
<td>0.654</td>
<td>0.621</td>
<td>0.624</td>
<td>0.591</td>
</tr>
<tr>
<td>20 years</td>
<td>0.633</td>
<td>0.598</td>
<td>0.585</td>
<td>0.552</td>
</tr>
<tr>
<td>30 years</td>
<td>0.584</td>
<td>0.552</td>
<td>0.484</td>
<td>0.419</td>
</tr>
</tbody>
</table>
effect of progression, the results obtained by utilizing only the original images are also provided. Some interesting observations are drawn from the results. Firstly, the improvement in accuracy validates that the identity information of the face remains after the RJIVE-based progression. Furthermore, the improvement in accuracy is higher when the age difference of images of each pair is big enough. For instance, the improvement in accuracy in ‘Protocol 30 years’ is higher than the corresponding in ‘Protocol 5 years’. Finally, the produced results justify that the problem of age-invariant face verification becomes more difficult when the age difference is very large (e.g., 30 years).

The performance of RJIVE in age-invariant face verification is also compared against the IAAP [1] by conducting experiment on the FG-NET database. The experimental protocol employed is as follows. By selecting images where the depicted subjects are older than the age of 18 years, we created a subset of the FG-NET database consists of 518 images. Then, based on the selected images we created 1250 intra-class pairs i.e., the images of each pair depict the same subject under different ages, and another 1250 inter-class pairs. The experiment protocol was finally created by dividing the pairs on 5 folds with each fold contains 250 intra-class pairs and 250 inter-class ones. All images were then progressed by employing the RJIVE and IAAP methods. A similar to previous experiment procedure was followed in order to perform the age-invariant verification. The produced ROC curves are displayed in Figure 15. As it can be observed the proposed RJIVE method outperforms the IAAP by a large margin indicating that the RJIVE produces progressed images of high quality without removing the identity information.

Fig. 15: ROC curve of the RJIVE and IAAP on FG-NET database.

8 Conclusions

A general framework for robust recovering of joint and individual variance among several data-sets possibly contaminated by gross non-Gaussian errors and incomplete has been proposed in this paper. Four different models namely, $\ell - 1$-RJIVE, NN-$\ell - 1$-RJIVE, SRJIVE, and RJIVE-M have been proposed. Furthermore, based on the recovered components from training data, two novel optimization problems that extracts the joint and individual components of an unseen test sample, are introduced. The effectiveness of the RJIVE was first tested by conducting experiments on synthetic data. Then, extensive experiments were conducted on facial expression synthesis and 2D an 3D face age progression by utilizing five data-sets captured under both controlled and in-the-wild conditions. The experimental results validate the effectiveness of the proposed RJIVE method over the state-of-the-art.

References


