This ICCV workshop paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the version available on IEEE Xplore.

Variational Robust Subspace Clustering with Mean Update Algorithm

Sergej Dogadov
TU Berlin
Berlin, Germany
sergej.dogadov@campus.tu-berlin.de

Andrés Masegosa
University of Almería
Almería, Spain
andresmasegosa@ual.es

Shinichi Nakajima
TU Berlin
Berlin, Germany
nakajima@tu-berlin.de

Abstract

In this paper, we propose an efficient variational Bayesian (VB) solver for a robust variant of low-rank subspace clustering (LRSC). VB learning offers automatic model selection without parameter tuning. However, it is typically performed by local search with update rules derived from conditional conjugacy, and therefore prone to local minima problem. Instead, we use an approximate global solver for LRSC with an element-wise sparse term to make it robust against spiky noise. In experiment, our method (mean update solver for robust LRSC), outperforms the original LRSC, as well as the robust LRSC with the standard VB solver.

1. Introduction

Subspace clustering (SC) finds a clustering structure based not on the adjacency between samples but on the distance to the subspaces spanned by the member samples. The method was proposed in non-Bayesian way [6, 21, 7, 11, 13, 12] and the low-rank SC (LRSC), a variant of SC, was subsequently casted into the Bayesian framework [3, 16].

A notable advantage of Bayesian learning is that it can be free from parameter tuning. More specifically, one can perform model selection by minimizing the marginal likelihood, computed from training data. Moreover, by using automatic relevance determination (ARD) prior, a single inference provides the inference result after model selection by eliminating the irrelevant components automatically [4, 18].

On the other hand, one of the drawbacks of VB learning is that the inference algorithm is typically local search, which is prone to local minima problem. However, there are a few cases where an efficient (approximate) global solver is available.

In probabilictic PCA or fully-observed matrix factorization, it was shown that the VB solution can be expressed as a truncated and shrunken singular value decomposition, where the shrunken singular values can be analytically computed from the original singular values [17]. A similar approach was applied to LRSC, and an efficient approximate global solver has been derived [16].

The analytic solution for the fully-observed matrix factorization was used for developing efficient algorithms for more general cases, including matrix factorization with missing entries and non-conjugate likelihood [19], and Sparse Additive Matrix Factorization (SAMF) where the observed matrix is expressed as a sum of various types of factorizations and noise such that different sparsity structures are captured [14].

In this paper, we use the approximate global solver for LRSC to solve its robust variant, i.e., robust LRSC. Robust LRSC consists of the LRSC term and an element-wise sparse term that captures spiky observation noise. Since the global solver is available for each of the LRSC term and the element-wise sparse term, we can incorporate the framework of *mean update* algorithm developed for SAMF [14] with slight modifications.

In our experiment, our mean update solver for robust LRSC, outperforms the plain LRSC as well as robust LRSC solved by standard algorithm for VB learning [3].

2. Background

In this section, we summarize previous work, which are necessary to derive our method in Section 3.

2.1. Subspace Clustering Methods

Let $\boldsymbol{V} \in \mathbb{R}^{L \times M} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M)$ be L-dimensional observed samples of size M. We generally denote a column vector of a matrix by a bold-faced small letter. We assume that each $\boldsymbol{y}_m$ is approximately expressed as a linear combination of M' *words* in a dictionary, $\boldsymbol{D} = (\boldsymbol{d}_1, \ldots, \boldsymbol{d}_{M'}) \in \mathbb{R}^{L \times M'}$, i.e.,

$$\boldsymbol{V} = \boldsymbol{D}\boldsymbol{U} + \boldsymbol{\mathcal{E}},$$

where $\boldsymbol{U} \in \mathbb{R}^{M' \times M}$ is unknown coefficients, and $\boldsymbol{\mathcal{E}} \in \mathbb{R}^{L \times M}$ is noise. In subspace clustering, the observed matrix $\boldsymbol{V}$ itself is often used as a dictionary $\boldsymbol{D}$. The convex

formulation of the sparse subspace clustering (SSC) [6, 21] is given by

$$\min_{\boldsymbol{U}} \|\boldsymbol{V} - \boldsymbol{V}\boldsymbol{U}\|_{\text{Fro}}^2 + \lambda \|\boldsymbol{U}\|_1, \text{ s.t. } \mathbf{diag}(\boldsymbol{U}) = \boldsymbol{0}, \quad (1)$$

where $\boldsymbol{U} \in \mathbb{R}^{M \times M}$ is a parameter to be estimated, $\lambda > 0$ is a regularization coefficient to be manually tuned. $\|\cdot\|_{\text{Fro}}$ and $\|\cdot\|_1$ are the Frobenius norm and the (element-wise) ℓ_1-norm of a matrix, respectively. $\mathbf{diag}(\cdot)$ extracts the diagonal entries of a matrix, and form a vector. The first term in Eq.(1) requires that each data point $\boldsymbol{v}_m$ can be expressed as a linear combination of a *small* set of other data points $\{\boldsymbol{d}_{m'}\}$ for $m' \neq m$. This *smallness* of the set is enforced by the second (ℓ_1-regularization) term, and leads to the low-dimensionality of each obtained subspace. After the minimizer $\widehat{\boldsymbol{U}}$ is obtained, $\text{abs}(\widehat{\boldsymbol{U}}) + \text{abs}(\widehat{\boldsymbol{U}}^\top)$, where $\text{abs}(\cdot)$ takes the absolute value element-wise, is regarded as an affinity matrix, and a spectral clustering algorithm, such as normalized cuts [20], is applied to obtain clusters.

In another variant, called the low-rank subspace clustering (LRSC) or low-rank representation [7, 11, 13, 12], low-dimensional subspaces are sought by enforcing the low-rankness of $\boldsymbol{U}$:

$$\min_{\boldsymbol{U}} \|\boldsymbol{V} - \boldsymbol{V}\boldsymbol{U}\|_{\text{Fro}}^2 + \lambda \|\boldsymbol{U}\|_{\text{tr}}, \quad (2)$$

where $\|\cdot\|_{\text{tr}}$ denotes the trace norm of a matrix. Thanks to the simplicity, the global solution of Eq.(2) has been analytically obtained [7].

2.2. Variational Bayesian Low-rank Subspace Clustering

We formulate the probabilistic model of LRSC, so that the maximum a posteriori (MAP) estimator coincides with the solution of the problem (2) under a certain hyperparameter setting:

$$p(\boldsymbol{V}|\boldsymbol{A}', \boldsymbol{B}') \propto \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{V} - \boldsymbol{D}\boldsymbol{B}'\boldsymbol{A}'^\top\|_{\text{Fro}}^2\right), \quad (3)$$

$$p(\boldsymbol{A}') \propto \exp\left(-\frac{1}{2}\text{tr}(\boldsymbol{A}'\boldsymbol{C}_A^{-1}\boldsymbol{A}'^\top)\right), \quad (4)$$

$$p(\boldsymbol{B}') \propto \exp\left(-\frac{1}{2}\text{tr}(\boldsymbol{B}'\boldsymbol{C}_B^{-1}\boldsymbol{B}'^\top)\right). \quad (5)$$

Here, we factorized $\boldsymbol{U}$ as $\boldsymbol{U} = \boldsymbol{B}'\boldsymbol{A}'^\top$, as in [3], to induce low-rankness through the *model-induced regularization* mechanism [15]. In this formulation, $\boldsymbol{A}' \in \mathbb{R}^{M \times H}$ and $\boldsymbol{B}' \in \mathbb{R}^{M \times H}$ for $H \leq \min(L, M)$ are the parameters to be estimated. We assume that hyperparameters

$$\boldsymbol{C}_A = \mathbf{diag}(c_{a_1}^2, \ldots, c_{a_H}^2), \quad \boldsymbol{C}_B = \mathbf{diag}(c_{b_1}^2, \ldots, c_{b_H}^2).$$

are diagonal and positive definite. The dictionary $\boldsymbol{D}$ is treated as a constant, and set to $\boldsymbol{D} = \boldsymbol{V}$, once $\boldsymbol{V}$ is observed.

2.3. Variational Bayesian (VB) Learning

The Bayes posterior of the LRSC model (3)–(5) can be written as

$$p(\boldsymbol{A}', \boldsymbol{B}'|\boldsymbol{V}) = \frac{p(\boldsymbol{V}|\boldsymbol{A}', \boldsymbol{B}')p(\boldsymbol{A}')p(\boldsymbol{B}')}{p(\boldsymbol{V})}, \quad (6)$$

which is intractable because the marginal likelihood $p(\boldsymbol{V}) = \langle p(\boldsymbol{V}|\boldsymbol{A}', \boldsymbol{B}')\rangle_{p(\boldsymbol{A}')p(\boldsymbol{B}')}$ is hard to compute. Here, $\langle\cdot\rangle_p$ denotes the expectation over the distribution p.

The variational Bayesian (VB) learning [1, 4] approximates the Bayes posterior with $r(\boldsymbol{A}', \boldsymbol{B}')$ or r for short, which minimizes the free energy

$$F(r) = \left\langle \log \frac{r(\boldsymbol{A}', \boldsymbol{B}')}{p(\boldsymbol{V}|\boldsymbol{A}', \boldsymbol{B}'), p(\boldsymbol{A}')p(\boldsymbol{B}')} \right\rangle_{r(\boldsymbol{A}', \boldsymbol{B}')} \quad (7)$$

under the independence constraint

$$r(\boldsymbol{A}', \boldsymbol{B}') = r(\boldsymbol{A}')r(\boldsymbol{B}'). \quad (8)$$

Note that the free energy (7) can be written as

$$F(r) = \left\langle \log \frac{r(\boldsymbol{A}', \boldsymbol{B}')}{p(\boldsymbol{A}', \boldsymbol{B}'|\boldsymbol{V})} \right\rangle_{r(\boldsymbol{A}', \boldsymbol{B}')} - \log p(\boldsymbol{V}),$$

where the first term is the Kullback-Leibler (KL) divergence from $r(\boldsymbol{A}', \boldsymbol{B}')$ to the Bayes posterior, and the second term is a constant. Therefore, minimizing the free energy (7) amounts to finding a distribution closest to the Bayes posterior from the possible functions specified by the independence constraint (8).

2.4. Approximate VB Global Solver for LRSC

The VB solution to the LRSC model (3)–(5) under the constraint (8) were analyzed [16], and some properties have been revealed.

Let us transform the parameters as

$$\boldsymbol{A} = \boldsymbol{\Omega}_V^{\text{right}\top}\boldsymbol{A}', \qquad \boldsymbol{B} = \boldsymbol{\Omega}_V^{\text{right}\top}\boldsymbol{B}', \quad (9)$$

where

$$\boldsymbol{V} = \boldsymbol{\Omega}_V^{\text{left}}\boldsymbol{\Gamma}_V\boldsymbol{\Omega}_V^{\text{right}\top}$$

is the singular value decomposition (SVD) of the observed matrix $\boldsymbol{V}$. Then, the following holds:

Proposition 1 *[16] The VB posterior that (globally) minimizes the free energy is written as*

$$r(\boldsymbol{A}) \propto \exp\left(-\frac{tr\left((\boldsymbol{A} - \widehat{\boldsymbol{A}})\widehat{\boldsymbol{\Sigma}}_A^{-1}(\boldsymbol{A} - \widehat{\boldsymbol{A}})^\top\right)}{2}\right), \quad (10)$$

$$r(\boldsymbol{B}) \propto \exp\left(-\frac{(\breve{\boldsymbol{b}}' - \widehat{\breve{\boldsymbol{b}}}')^\top \widehat{\breve{\boldsymbol{\Sigma}}}_B^{-1}(\breve{\boldsymbol{b}} - \widehat{\breve{\boldsymbol{b}}})}{2}\right), \quad (11)$$

where $\breve{\boldsymbol{b}} = \boldsymbol{vec}(\boldsymbol{B}) \in \mathbb{R}^{M \cdot H}$*, and any global solution has its equivalent solution with diagonal means and covariances, i.e., all of* $\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\Sigma}}_A, \widehat{\breve{\boldsymbol{\Sigma}}}_B$ *are diagonal.*

This proposition allows us to focus on the solutions where $(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\Sigma}}_A, \widehat{\breve{\boldsymbol{\Sigma}}}_B)$ are diagonal. Let $J(\leq \min(L, M))$ be the rank of the observed matrix $\boldsymbol{V}$, γ_m be the m-th largest singular value of $\boldsymbol{V}$, and $(\widehat{a}_1, \ldots, \widehat{a}_H)$, $(\sigma^2_{a_1}, \ldots, \sigma^2_{a_H})$, $(\widehat{b}_1, \ldots, \widehat{b}_H)$, $((\sigma^2_{B_{1,1}}, \ldots, \sigma^2_{B_{M,1}}), \ldots, (\sigma^2_{B_{1,H}}, \ldots, \sigma^2_{B_{M,H}}))$ be the diagonal entries of $\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{\Sigma}}_A, \widehat{\boldsymbol{B}}, \widehat{\breve{\boldsymbol{\Sigma}}}_B$, respectively. Then, the following holds:

Proposition 2 *[16] The free energy (7) is written as*

$$F = \frac{1}{2}\left(LM\log(2\pi\sigma^2) + \frac{\sum_{h=1}^{J}\gamma_h^2}{\sigma^2} + \sum_{h=1}^{H} 2F_h\right), \quad (12)$$

where

$$\begin{aligned} 2F_h = M\log\frac{c^2_{a_h}}{\sigma^2_{a_h}} + \sum_{m=1}^{J}\log\frac{c^2_{b_h}}{\sigma^2_{B_{m,h}}} - (M+J) \\ + \frac{\widehat{a}_h^2 + M\sigma^2_{a_h}}{c^2_{a_h}} + \frac{\widehat{b}_h^2 + \sum_{m=1}^{J}\sigma^2_{B_{m,h}}}{c^2_{b_h}} \\ + \frac{1}{\sigma^2}\Big\{\gamma_h^2\left(-2\widehat{a}_h\widehat{b}_h + \widehat{b}_h^2(\widehat{a}_h^2 + M\sigma^2_{a_h})\right) \\ + \sum_{m=1}^{J}\gamma_m^2\sigma^2_{B_{m,h}}(\widehat{a}_h^2 + M\sigma^2_{a_h})\Big\}. \end{aligned} \quad (13)$$

Based on the propositions above, an exact global solver based on *homotopy method* [10, 8] has been derived, which however has exponential complexity with respect to J. In this paper, we focus on an approximate global solver described below.

We introduce an additional constraint

$$\gamma_m^2\sigma^2_{B_{m,h}} = \overline{\sigma}^2_{b_h} \quad \text{for all} \quad m \leq J, \quad (14)$$

which makes the objective function separable for each singular component. The following holds:

Proposition 3 *[16] Under the additional constraint (14), any stationary point of the free energy (13) for each h satisfies the following polynomial equation with a single variable $\widehat{\widehat{\gamma}}_h$:*

$$\xi_6\widehat{\widehat{\gamma}}_h^6 + \xi_5\widehat{\widehat{\gamma}}_h^5 + \xi_4\widehat{\widehat{\gamma}}_h^4 + \xi_3\widehat{\widehat{\gamma}}_h^3 + \xi_2\widehat{\widehat{\gamma}}_h^2 + \xi_1\widehat{\widehat{\gamma}}_h + \xi_0 = 0, \quad (15)$$

where

$$\xi_6 = \frac{\phi_h^2}{\gamma_h^2}, \quad (16)$$

$$\xi_5 = -2\frac{\phi_h^2 M\sigma^2}{\gamma_h^3} + \frac{2\phi_h}{\gamma_h}, \quad (17)$$

$$\xi_4 = \frac{\phi_h^2 M^2\sigma^4}{\gamma_h^4} - \frac{2\phi_h(2M-J)\sigma^2}{\gamma_h^2} + 1 + \frac{\phi_h^2(M\sigma^2-\gamma_h^2)}{\gamma_h^2}, \quad (18)$$

$$\begin{aligned} \xi_3 = \frac{2\phi_h M(M-J)\sigma^4}{\gamma_h^3} - \frac{2(M-J)\sigma^2}{\gamma_h} + \frac{\phi_h((M+J)\sigma^2-\gamma_h^2)}{\gamma_h} \\ - \frac{\phi_h^2 M\sigma^2(M\sigma^2-\gamma_h^2)}{\gamma_h^3} + \frac{\phi_h(M\sigma^2-\gamma_h^2)}{\gamma_h}, \end{aligned} \quad (19)$$

$$\begin{aligned} \xi_2 = \frac{(M-J)^2\sigma^4}{\gamma_h^2} - \frac{\phi_h M\sigma^2((M+J)\sigma^2-\gamma_h^2)}{\gamma_h^2} \\ + ((M+J)\sigma^2 - \gamma_h^2) - \frac{\phi_h(M-J)\sigma^2(M\sigma^2-\gamma_h^2)}{\gamma_h^2}, \end{aligned} \quad (20)$$

$$\xi_1 = -\frac{(M-J)\sigma^2((M+J)\sigma^2-\gamma_h^2)}{\gamma_h} + \frac{\phi_h MJ\sigma^4}{\gamma_h}, \quad (21)$$

$$\xi_0 = MJ\sigma^4. \quad (22)$$

Here, $\phi_h = \left(1 - \frac{\gamma_h^2}{\underline{\gamma}^2}\right)$ *for* $\underline{\gamma}^2 = (\sum_{m=1}^{J}\gamma_m^{-2}/J)^{-1}$*. For each real solution* $\widehat{\widehat{\gamma}}_h$ *such that*

$$\widehat{\gamma}_h = \widehat{\widehat{\gamma}}_h + \gamma_h - \frac{M\sigma^2}{\gamma_h}, \quad (23)$$

$$\widehat{\kappa} = \gamma_h^2 - (M+J)\sigma^2 - \left(M\sigma^2 - \gamma_h^2\right)\phi_h\frac{\widehat{\widehat{\gamma}}_h}{\gamma_h}, \quad (24)$$

$$\widehat{\tau} = \frac{1}{2MJ}\left(\widehat{\kappa} + \sqrt{\widehat{\kappa}^2 - 4MJ\sigma^4\left(1 + \phi_h\frac{\widehat{\widehat{\gamma}}_h}{\gamma_h}\right)}\right), \quad (25)$$

$$\widehat{\delta}_h = \frac{\sigma^2}{\sqrt{\widehat{\tau}}}\left(\gamma_h - \frac{M\sigma^2}{\gamma_h} - \widehat{\gamma}_h\right)^{-1}, \quad (26)$$

are real and positive, the corresponding stationary point candidate is given by

$$\begin{aligned} \left(\widehat{a}_h, \sigma^2_{a_h}, c^2_{a_h}, \widehat{b}_h, \overline{\sigma}^2_{b_h}, c^2_{b_h}\right) = \left(\sqrt{\widehat{\gamma}\widehat{\delta}}, \frac{\sigma^2\widehat{\delta}_h}{\gamma_h}, \sqrt{\widehat{\tau}}, \right. \\ \left. \sqrt{\widehat{\gamma}/\widehat{\delta}}/\gamma_h, \frac{\sigma^2}{\gamma_h\widehat{\delta}_h - \phi_h\frac{\sigma^2}{\sqrt{\widehat{\tau}}}}, \sqrt{\widehat{\tau}}/\gamma_h^2\right). \end{aligned} \quad (27)$$

Given the noise variance σ^2, computing the coefficients (16)–(22) is straightforward. The approximate global solver (Algorithm 1) solves the sixth-order polynomial equation (15), e.g., by the '`roots`' function in MATLAB®, and obtain all candidate stationary points by using Eqs.(23)–(27). Then, it selects the one giving the smallest F_h, and the global solution is the selected stationary point if it satisfies $F_h < 0$, otherwise the null solution given by

$$\widehat{a}_h = \widehat{b}_h = 0, \qquad \sigma^2_{a_h}, \sigma^2_{B_{m,h}}, c^2_{a_h}, c^2_{b_h} \to 0 \quad \text{for} \quad m = 1, \ldots, M. \quad (28)$$

Note that, although a solution of Eq.(15) is not necessarily a stationary point, selection based on the free energy discards all non-stationary points and local maxima.

Algorithm 1 Approximate Global Solver for LRSC.

1: Calculate the SVD of $\boldsymbol{V} = \boldsymbol{\Omega}_V^{\text{left}} \Gamma_V \boldsymbol{\Omega}_V^{\text{right}\top}$.
2: **for** $h = 1$ to H **do**
3: Find all real solutions of the sixth-order polynomial equation (15).
4: Discard prohibitive solutions such that any of Eqs.(23)–(26) gives complex or negative number.
5: Compute the corresponding stationary point by Eq.(27) and its free energy contribution F_h by Eq.(13) for each of the retained solutions.
6: Select the stationary point giving the minimum free energy contribution F_h.
7: The global solution for h is the selected stationary point if it satisfies $F_h < 0$, otherwise the null local solution (28).
8: **end for**
9: Compute $\widehat{\boldsymbol{A}}' = \boldsymbol{\Omega}_V^{\text{right}} \widehat{\boldsymbol{A}}$ and $\widehat{\boldsymbol{B}}' = \boldsymbol{\Omega}_V^{\text{right}} \widehat{\boldsymbol{B}}$.
10: Apply spectral clustering with the affinity matrix equal to $\text{abs}(\widehat{\boldsymbol{B}}' \widehat{\boldsymbol{A}}'^{\top}) + \text{abs}(\widehat{\boldsymbol{A}}' \widehat{\boldsymbol{B}}'^{\top})$.

2.5. Mean Update Algorithm

Sparse additive matrix factorization (SAMF) [14] model consists of multiple (additive) terms

$$\boldsymbol{V} = \sum_{s=1}^{S} \boldsymbol{U}^{(s)} + \boldsymbol{\mathcal{E}},$$

where each term $\boldsymbol{U}^{(s)}$ is designed as a specific type of factorization to induce sparsity. For example, the Bayesian robust PCA [5, 2] that consists of a low-rank term and an element-wise sparse term can be modelled with

$$\boldsymbol{U}^{(1)} = \boldsymbol{U}^{\text{low-rank}} = \boldsymbol{B}\boldsymbol{A}^{\top}, \quad \boldsymbol{U}^{(2)} = \boldsymbol{U}^{\text{element}} = \boldsymbol{E} * \boldsymbol{G}, \tag{29}$$

where $*$ denotes the Hadamard product.

The standard VB algorithm updates the parameters $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{E}, \boldsymbol{G})$ one by one in addition to the hyperparameters. However, a better algorithm, called *mean update* (MU), has been proposed by using the analytic-form VB solution obtained for fully-observed matrix factorization [17]. In the MU algorithm, the set of parameters and hyperparameters contained in each s-th term is updated based on the global analytic solution, which shows faster convergence to a better local solution than the standard VB algorithm [14]. In the next section, we apply this idea for a robust variant of LRSC.

3. Proposed method

In this section, we introduce robust LRSC and its efficient VB solver.

3.1. Robust Low-rank Subspace Clustering

We build robust LRSC in the SAMF frame work with an LRSC term and an element-wise sparse term:

$$\boldsymbol{V} = \boldsymbol{U}^{\text{LRSC}} + \boldsymbol{U}^{\text{element}} + \boldsymbol{\mathcal{E}}, \tag{30}$$

where each term factorizes as

$$\boldsymbol{U}^{\text{LRSC}} = \boldsymbol{D}\boldsymbol{B}'\boldsymbol{A}'^{\top}, \qquad \boldsymbol{U}^{\text{element}} = \boldsymbol{E} * \boldsymbol{G}. \tag{31}$$

For the dictionary $\boldsymbol{D} \in \mathbb{R}^{L \times M}$, we use a denoised version of observed matrix based on the estimated spiky noise, i.e.,

$$\boldsymbol{D} = \boldsymbol{V} - \widehat{\boldsymbol{U}}^{\text{element}}. \tag{32}$$

The corresponding probabilistic model is given as

$$p(\boldsymbol{V}|\boldsymbol{A}', \boldsymbol{B}', \boldsymbol{E}, \boldsymbol{G}) \propto \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{V} - \boldsymbol{D}\boldsymbol{B}'\boldsymbol{A}'^{\top} - \boldsymbol{E} * \boldsymbol{G}\|_{\text{Fro}}^2\right), \tag{33}$$

$$p(\boldsymbol{A}') \propto \exp\left(-\frac{1}{2}\text{tr}(\boldsymbol{A}'\boldsymbol{C}_A^{-1}\boldsymbol{A}'^{\top})\right), \tag{34}$$

$$p(\boldsymbol{B}') \propto \exp\left(-\frac{1}{2}\text{tr}(\boldsymbol{B}'\boldsymbol{C}_B^{-1}\boldsymbol{B}'^{\top})\right), \tag{35}$$

$$p(\boldsymbol{E}) \propto \exp\left(-\frac{\sum_{l=1}^{L}\sum_{m=1}^{M} E_{l,m}^2}{2(\boldsymbol{C}_E)_{l,m}}\right), \tag{36}$$

$$p(\boldsymbol{G}) \propto \exp\left(-\frac{\sum_{l=1}^{L}\sum_{m=1}^{M} G_{l,m}^2}{2(\boldsymbol{C}_G)_{l,m}}\right), \tag{37}$$

where each entry of $\boldsymbol{C}_E, \boldsymbol{C}_G \in \mathbb{R}^{L \times M}$ corresponds to the prior variance of the entry of $\boldsymbol{E}$ and $\boldsymbol{G}$, respectively. We treat $(\boldsymbol{A}', \boldsymbol{B}', \boldsymbol{E}, \boldsymbol{G})$ as (unknown) parameters and $(\boldsymbol{C}_A, \boldsymbol{C}_B, \boldsymbol{C}_E, \boldsymbol{C}_G, \sigma^2)$ as hyperparameters, and apply VB learning to estimate all of them from the observed matrix $\boldsymbol{V}$.

3.2. Mean Update Algorithm for robust LRSC

Let $\boldsymbol{\Theta} = (\boldsymbol{A}', \boldsymbol{B}', \boldsymbol{E}, \boldsymbol{G})$ summarize the parameters. We approximate the Bayes posteterior with the VB posterior that minimizes the free energy

$$F(r) = \left\langle \log \frac{r(\boldsymbol{\Theta})}{p(\boldsymbol{V}, \boldsymbol{\Theta})} \right\rangle_{r(\boldsymbol{\Theta})} \tag{38}$$

under the independence constraint:

$$r(\boldsymbol{\Theta}) = r(\boldsymbol{A}')r(\boldsymbol{B}')r(\boldsymbol{E})r(\boldsymbol{G}). \tag{39}$$

Because of the imposed independence between $(\boldsymbol{A}', \boldsymbol{B}')$ and $(\boldsymbol{E}, \boldsymbol{G})$, the VB posterior for $(\boldsymbol{A}', \boldsymbol{B}')$ depends on the VB posterior for $(\boldsymbol{E}, \boldsymbol{G})$ only through the *means*, i.e., $(\widehat{\boldsymbol{E}}, \widehat{\boldsymbol{G}})$. The opposite also holds. Consequently, we can easily obtain the following theorems (the proof is given in Appendix A):

Theorem 1 *Given the means* $(\widehat{\boldsymbol{E}}, \widehat{\boldsymbol{G}})$ *of the posteriors* $r(\boldsymbol{E}), r(\boldsymbol{G})$ *and the noise variance* σ^2*, the VB posteriors* $r(\boldsymbol{A}'), r(\boldsymbol{B}')$ *for* $(\boldsymbol{A}', \boldsymbol{B}')$ *along with the hyperparameters* $(\boldsymbol{C}_A, \boldsymbol{C}_B)$ *can be obtained by minimizing the free energy* (7) *for (original) LRSC with* $\widehat{\boldsymbol{D}} = \boldsymbol{V} - \widehat{\boldsymbol{E}} * \widehat{\boldsymbol{G}}$ *substituted for the observed matrix* $\boldsymbol{V}$*.*

Theorem 2 *Given the means* $(\widehat{\boldsymbol{A}}', \widehat{\boldsymbol{B}}')$ *of the posteriors* $r(\boldsymbol{A}'), r(\boldsymbol{B}')$*, the dictionary* $\widehat{\boldsymbol{D}}$*, and the noise variance* σ^2*, the VB posteriors* $r(\boldsymbol{E}), r(\boldsymbol{G})$ *for* $(\boldsymbol{E}, \boldsymbol{G})$ *along with the hyperparameters* $(\boldsymbol{C}_E, \boldsymbol{C}_G)$ *can be obtained by minimizing the free energy for the element-wise matrix factorization with* $\boldsymbol{V} - \widehat{\boldsymbol{D}}\widehat{\boldsymbol{B}}'\widehat{\boldsymbol{A}}'$ *substituted for the observed matrix* $\boldsymbol{V}$*.*

Here, the element-wise matrix factorization denotes

$$\boldsymbol{V} = \boldsymbol{E} * \boldsymbol{G} + \boldsymbol{\mathcal{E}} \tag{40}$$

with the corresponding priors (36) and (37), for which an analytic-form VB solution is available:

Proposition 4 *[17] The VB solution for the element-wise matrix factorization* (40)*,* (36)*,* (37) *(where* $\boldsymbol{\mathcal{E}}$ *is independent Gaussian noise with mean 0 and variance* σ^2*) is given by*

$$(\widehat{\boldsymbol{U}}^{\text{element}})_{l,m} = (\widehat{\boldsymbol{E}} * \widehat{\boldsymbol{G}})_{l,m} = \begin{cases} 0 \text{ if } |V_{l,m}| < \sigma\sqrt{(1+\underline{\tau})(1+\underline{\tau}^{-1})}, \\ \dfrac{V_{l,m}\left(1 - \frac{2\sigma^2}{V_{l,m}^2} + \sqrt{1 - \frac{4\sigma^2}{V_{l,m}^2}}\right)}{2} \text{ otherwise}, \end{cases}$$

where $\underline{\tau}$ *can be approximated by* $\underline{\tau} \approx 2.5129$*.*

Based on the theorems above, we propose to compute the VB solution of RLRSC by iterating the update for the LRSC part $r(\boldsymbol{A}'), r(\boldsymbol{B}')$ with the approximate global solver (Algorithm 1) and for the element-wise sparse part $r(\boldsymbol{E}), r(\boldsymbol{G})$ with its analytic-form solution (Proposition 4). We can also estimate the noise variance σ^2 by using the following lemma:

Lemma 1 *Given the VB posteriors* $r(\boldsymbol{A}'), r(\boldsymbol{B}'), r(\boldsymbol{E})$*,* $r(\boldsymbol{G})$*,*[1] *the noise parameter* σ^2 *that minimizes the free energy is given by*

$$\sigma^2 = \frac{1}{LM}\Big\{\|\boldsymbol{V} - \boldsymbol{D}\widehat{\boldsymbol{B}}'\widehat{\boldsymbol{A}}'^{\top} - \widehat{\boldsymbol{E}} * \widehat{\boldsymbol{G}}\|^2 + \langle\|\boldsymbol{D}\boldsymbol{B}'\boldsymbol{A}'^{\top} - \boldsymbol{D}\widehat{\boldsymbol{B}}'\widehat{\boldsymbol{A}}'^{\top}\|^2\rangle_{r(\boldsymbol{A}')r(\boldsymbol{B}')} + \langle\|\boldsymbol{E} * \boldsymbol{G} - \widehat{\boldsymbol{E}} * \widehat{\boldsymbol{G}})\|^2\rangle_{r(\boldsymbol{E})r(\boldsymbol{G})}\Big\}. \tag{41}$$

[1]Based on Theorem 1 and Theorem 2, we can completely specify the VB posteriors [16, 17].

Algorithm 2 Mean Update Solver for Robust LRSC.

1: Initialize $\widehat{\boldsymbol{U}}^{\text{element}} = \boldsymbol{0}$, $\sigma^2 = \|\boldsymbol{V}\|_{\text{Fro}}^2/(LM)$.
2: **while** until convergence **do**
3: Compute $\widehat{\boldsymbol{A}}'$ and $\widehat{\boldsymbol{B}}'$ by the approximate global solver (Algrithm 1 up to Step 9) with the denoised dictionary $\widehat{\boldsymbol{D}} = \boldsymbol{V} - \widehat{\boldsymbol{U}}^{\text{element}}$ substituted for the observed matrix $\boldsymbol{V}$.
4: Set $\widehat{\boldsymbol{U}}^{\text{LRSC}} = \widehat{\boldsymbol{D}}\widehat{\boldsymbol{B}}'\widehat{\boldsymbol{A}}'^{\top}$.
5: Compute $\widehat{\boldsymbol{E}}$ and $\widehat{\boldsymbol{G}}$ by using the analytic-form solution (Proposition 4) with $\boldsymbol{V} - \widehat{\boldsymbol{U}}^{\text{LRSC}}$ substituted for the observed matrix $\boldsymbol{V}$.
6: Set $\widehat{\boldsymbol{U}}^{\text{element}} = \widehat{\boldsymbol{E}} * \widehat{\boldsymbol{G}}$.
7: Update the noise variance σ^2 by Lemma 1.
8: **end while**
9: Apply spectral clustering with the affinity matrix equal to $\text{abs}(\widehat{\boldsymbol{B}}'\widehat{\boldsymbol{A}}'^{\top}) + \text{abs}(\widehat{\boldsymbol{A}}'\widehat{\boldsymbol{B}}'^{\top})$.

Our proposed algorithm, called mean update solver (MUS) for robust LRSC is summarized in Algorithm 2. The mean update algorithm finds the global solution for $\widehat{\boldsymbol{U}}^{\text{LRSC}} = \widehat{\boldsymbol{D}}\widehat{\boldsymbol{B}}'\widehat{\boldsymbol{A}}'^{\top}$ given $\widehat{\boldsymbol{U}}^{\text{element}} = \widehat{\boldsymbol{E}} * \widehat{\boldsymbol{G}}$, and vice versa, in each iteration. Although this procedure is not guaranteed to find the joint global solution over all unknowns, it has empirically shown significantly better performance in SAMF than the standard VB iteration where each parameter is updated one by one [14]. In the next section, we show its good performance in robust LRSC.

4. Experiment

In this section, we experimentally evaluate the performance of our method, mean update solver for robust LRSC (MUS-RobustLRSC). We compare it with the original LRSC solved by approximate global solver (AGS-LRSC), and the robust LRSC solved by the matrix variate Gaussian approximation algorithm (MVGA-RobustLRSC), which is the practical standard VB algorithm developed for LRSC [3, 16].

We apply the methods to the *Hopkins 155 motion* database [22], where each sample corresponds to a trajectory of a point in a video, and clusteirng the trajectories amounts to finding a set of rigid bodies. We always use the full-rank model, i.e., $H = \min(L, M)$, and expect VB learning to automatically find the true rank without any parameter tuning.

Figure 1 shows the clustering errors over the first 20 sequences. We observed improvement with our MUS-RobustLRSC against the baselines—the average accuracies are 0.0575 (MUS-RobustLRSC), 0.0837 (AGS-LRSC), and 0.3297(MVGA-RobustLRSC), respectively. This shows that, although it is said that outliers have been removed

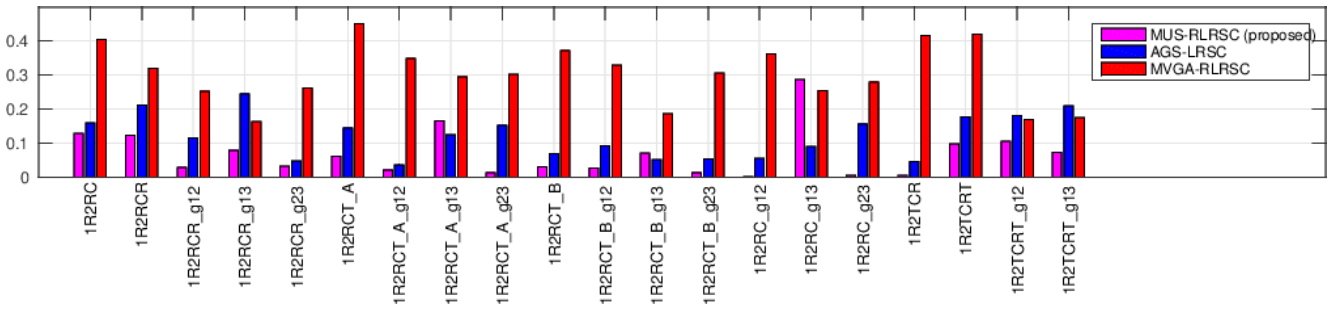

Figure 1. Clustering errors on the first 20 sequences of Hopkins 155 dataset.

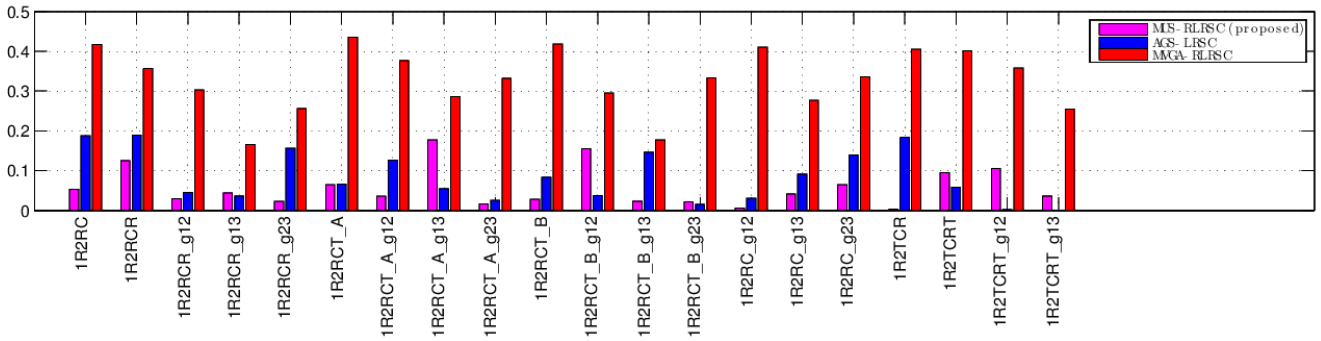

Figure 2. Clustering errors on the first 20 sequences of Hopkins 155 dataset with additional spiky noise.

from the *Hopkins 155 motion* dataset, they still contain some spiky noise, and removing the effect of them improves the accuracy. To see the performance dependent on the spiky noise, we also conducted the same experiment with artificially added spiky noise, which is created by $\boldsymbol{V} = \boldsymbol{V}^* + \boldsymbol{U}^{\text{element}*}$, where $\boldsymbol{V}^*$ is the original Hopkins motion data, and the additional spiky noise $\boldsymbol{U}^{\text{element}*}$ has ρLM for $\rho = 0.1$ non-zero entries, and each non-zero entry follows the Gaussian distribution with mean zero and variance $\eta^2 \|\boldsymbol{V}\|_{\text{Fro}}^2/(LM)$ for $\eta^2 = 0.1$.

Figure 2 shows the clustering errors. As expected, the non-robust variant, AGS-LRSC, performs worse in this case, while our MUS-RobustLRSC shows good performance. MVGA-RobustLRSC cannot be trained well in both cases. The average accuracies are 0.0681 (MUS-RobustLRSC), 0.1206 (AGS-LRSC), and 0.3034 (MVGA-RobustLRSC), respectively.

As a visual experimental example, we applied the methods to the Yale Database [9].[2] We selected first 5 classes (persons), and for each class 18 bright frontal images. Each image is resized to $M = 48 \times 42 = 2016$. After that, we added 5 different spiky noise for each image. In total, our data set contains $L = 5 \times 18 \times 5 = 450$ images. We investigated two cases for spiky noise, i.e., $\eta = 0.5, \rho = 0.1$, and $\eta = 1, \rho = 0.2$, which correspond to the top and the bottom rows in Figure 3, respectively.

[2] The Extended Yale Face Database B, used in previous works [9, 3], is no longer available. We made a data set by adding spiky noise.

The observed matrix is shown in the left most, and the decompositions by MVGA-RLRSC, AGS-LRSC, and MUS-RLRSC are shown in the other columns. As expected, our proposed MUS-RLRSC can cope with the spiky noise, and extracts the noiseless face image as the low-rank component.

5. Conclusion

In this paper, we developed mean update algorithm for variational Bayesian (VB) learning in robust low-rank subspace clustering (LRSC), based on the global solvers for LRSC and matrix factorization. In the experiment on *Hopkins 155 motion* database, our proposed method outperformed baseline methods, and demonstrated its usefulness. In our future work, we further explore efficient algorithms for approximate Bayesian learning.

Acknowledgments

The authors thank the reviewers for helpful comments. This work was supported by the German Ministry for Education and Research as Berlin Big Data Center BBDC, funding mark 01IS14013A.

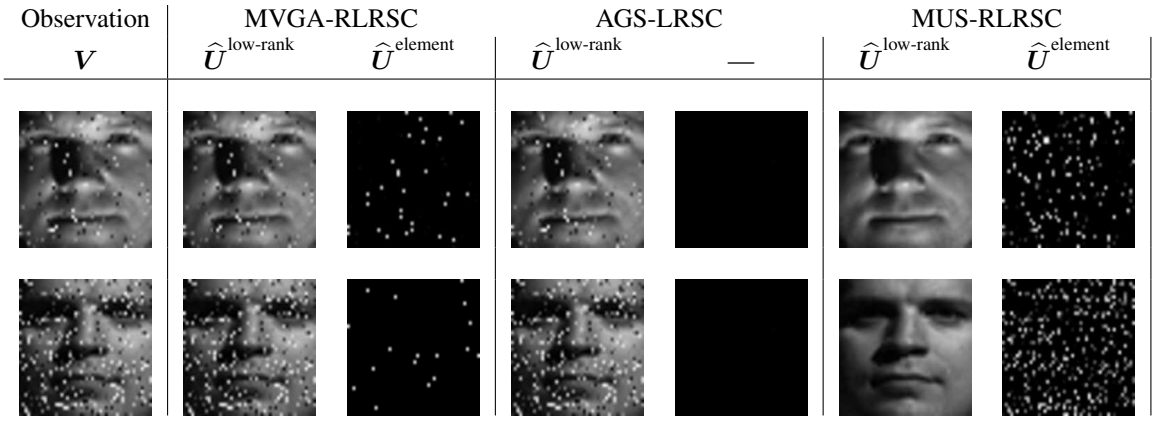

Figure 3. Spiky noise removal on Yale Database.

A. Proof of Theorem 1, Theorem 2 and Lemma 1

By applying the calculus of variations to the free energy (38) with the independence constraint (39), we can obtain the general update rule:

$$r(\boldsymbol{\theta}_i) \propto \exp\langle \log p(\boldsymbol{V}, \boldsymbol{\Theta}) \rangle_{r(\boldsymbol{\Theta} \setminus \boldsymbol{\theta}_i)}. \tag{42}$$

Here, $\boldsymbol{\theta}$ is either $\boldsymbol{A}'$, $\boldsymbol{B}'$, $\boldsymbol{E}$, or $\boldsymbol{G}$.

Since

$$\begin{aligned}\log p(\boldsymbol{V}, \boldsymbol{\Theta}) &= p(\boldsymbol{V}|\boldsymbol{A}', \boldsymbol{B}', \boldsymbol{E}, \boldsymbol{G}) p(\boldsymbol{A}') p(\boldsymbol{B}') p(\boldsymbol{E}) p(\boldsymbol{G}) \\ &= -\tfrac{1}{2\sigma^2} \|\boldsymbol{V} - \boldsymbol{D}\boldsymbol{B}'\boldsymbol{A}'^{\top} - \boldsymbol{E} * \boldsymbol{G}\|_{\text{Fro}}^2 \\ &\quad - \tfrac{1}{2}\text{tr}(\boldsymbol{A}'\boldsymbol{C}_A^{-1}\boldsymbol{A}'^{\top}) - \tfrac{1}{2}\text{tr}(\boldsymbol{B}'\boldsymbol{C}_B^{-1}\boldsymbol{B}'^{\top}) \\ &\quad - \tfrac{\sum_{l=1}^{L}\sum_{m=1}^{M} E_{l,m}^2}{2(\boldsymbol{C}_E)_{l,m}} - \tfrac{\sum_{l=1}^{L}\sum_{m=1}^{M} G_{l,m}^2}{2(\boldsymbol{C}_G)_{l,m}},\end{aligned}$$

Eq.(42) implies that the VB posteriors can be written as follows:

$$r(\boldsymbol{A}') \propto \exp\left(-\tfrac{1}{2}\text{tr}\left(\left(\boldsymbol{A}' - \widehat{\boldsymbol{A}}'\right)\widehat{\boldsymbol{\Sigma}}_{A'}^{-1}\left(\boldsymbol{A}' - \widehat{\boldsymbol{A}}'\right)^{\top}\right)\right), \tag{43}$$

$$r(\boldsymbol{B}') \propto \exp\left(-\tfrac{1}{2}\left(\breve{\boldsymbol{b}}' - \widehat{\breve{\boldsymbol{b}}}'\right)^{\top}\breve{\widehat{\boldsymbol{\Sigma}}}_{B'}^{-1}\left(\breve{\boldsymbol{b}}' - \widehat{\breve{\boldsymbol{b}}}'\right)\right), \tag{44}$$

$$r(\boldsymbol{E}) \propto \exp\left(-\textstyle\sum_{l=1}^{L}\sum_{m=1}^{M} \frac{(E_{l,m} - \widehat{E}_{l,m})^2}{2\widehat{\sigma}_{E_{l,m}}^2}\right), \tag{45}$$

$$r(\boldsymbol{G}) \propto \exp\left(-\textstyle\sum_{l=1}^{L}\sum_{m=1}^{M} \frac{(G_{l,m} - \widehat{G}_{l,m})^2}{2\widehat{\sigma}_{G_{l,m}}^2}\right). \tag{46}$$

Therefore, the free energy (38) can be explicitly written as

$$\begin{aligned}2F &= LM\log\sigma^2 + M\log\tfrac{|\boldsymbol{C}_{A'}|}{|\widehat{\boldsymbol{\Sigma}}_{A'}|} + \log\tfrac{|\boldsymbol{C}_{B'}\otimes\boldsymbol{I}_M|}{|\breve{\widehat{\boldsymbol{\Sigma}}}_{B'}|} \\ &\quad + \textstyle\sum_{l=1}^{L}\sum_{m=1}^{M}\left(\log\frac{(\boldsymbol{C}_E)_{l,m}}{\widehat{\sigma}_{E_{l,m}}^2} + \log\frac{(\boldsymbol{C}_G)_{l,m}}{\widehat{\sigma}_{G_{l,m}}^2}\right) \\ &\quad + \text{tr}(\boldsymbol{C}_{A'}^{-1}(\widehat{\boldsymbol{A}}'^{\top}\widehat{\boldsymbol{A}}' + M\widehat{\boldsymbol{\Sigma}}_{A'})) \\ &\quad + \text{tr}((\boldsymbol{C}_{B'}^{-1}\otimes\boldsymbol{I}_M)(\widehat{\breve{\boldsymbol{b}}}'\widehat{\breve{\boldsymbol{b}}}'^{\top} + \breve{\widehat{\boldsymbol{\Sigma}}}_{B'})) \\ &\quad + \textstyle\sum_{l=1}^{L}\sum_{m=1}^{M}\left(\frac{\widehat{E}_{l,m}^2 + \widehat{\sigma}_{E_{l,m}}^2}{(\boldsymbol{C}_E)_{l,m}} + \frac{\widehat{G}_{l,m}^2 + \widehat{\sigma}_{G_{l,m}}^2}{(\boldsymbol{C}_G)_{l,m}}\right) \\ &\quad + \tfrac{\|\boldsymbol{V} - \boldsymbol{D}\widehat{\boldsymbol{B}}'\widehat{\boldsymbol{A}}'^{\top} - \widehat{\boldsymbol{E}} * \widehat{\boldsymbol{G}}\|^2}{\sigma^2} \\ &\quad + \tfrac{\langle\|\boldsymbol{D}\boldsymbol{B}'\boldsymbol{A}'^{\top} - \boldsymbol{D}\widehat{\boldsymbol{B}}'\widehat{\boldsymbol{A}}'^{\top}\|^2\rangle_{r(\boldsymbol{A}')r(\boldsymbol{B}')}}{\sigma^2} \\ &\quad + \tfrac{\langle\|\boldsymbol{E}*\boldsymbol{G} - \widehat{\boldsymbol{E}}*\widehat{\boldsymbol{G}})\|^2\rangle_{r(\boldsymbol{E})r(\boldsymbol{G})}}{\sigma^2} + \text{const.}.\end{aligned} \tag{47}$$

If σ^2, $\widehat{\boldsymbol{E}}$, and $\widehat{\boldsymbol{G}}$ are given, and the dictionary is set by Eq.(32), then, the free energy (47) coincides with the free energy (7) for original LRSC (see [16] for the explicit form) as a function of the means and the covariances of the VB posteriors (43) and (44) with $\boldsymbol{V} - \widehat{\boldsymbol{E}} * \widehat{\boldsymbol{G}}$ substituted for $\boldsymbol{V}$. This completes the proof of Theorem 1.

Similarly, if σ^2, $\widehat{\boldsymbol{A}}'$, and $\widehat{\boldsymbol{B}}'$ are given, then, the free energy (47) coincides with the free energy for element-wise matrix factorization (see [17] for the explicit form) as a function of the means and the covariances of the VB posteriors (45) and (46) with $\boldsymbol{V} - \boldsymbol{D}\widehat{\boldsymbol{B}}'\widehat{\boldsymbol{A}}'^{\top}$ substituted for $\boldsymbol{V}$. This completes the proof of Theorem 2.

Finally, by taking the derivative of (47) with respect to σ^2, we obtain the update rule (41), which completes the proof of Lemma 1. □

References

[1] H. Attias. Inferring parameters and structure of latent variable models by variational Bayes. In *Proc. of UAI*, pages 21–30, 1999. 2

[2] S. D. Babacan, M. Luessi, R. Molina, and A. K. Katsaggelos. Sparse Bayesian methods for low-rank matrix estimation. *IEEE Trans. on Signal Processing*, 60(8):3964–3977, 2012. 4

[3] S. D. Babacan, S. Nakajima, and M. N. Do. Probabilistic low-rank subspace clustering. In *Advances in Neural Infor-*

mation Processing Systems 25, pages 2753–2761, 2012. 1, 2, 5, 6

[4] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, New York, NY, USA, 2006. 1, 2

[5] X. Ding, L. He, and L. Carin. Bayesian robust principal component analysis. *IEEE Transactions on Image Processing*, 20(12):3419–3430, 2011. 4

[6] E. Elhamifar and R. Vidal. Sparse subspace clustering. In *Proc. of CVPR*, pages 2790–2797, 2009. 1, 2

[7] P. Favaro, R. Vidal, and A. Ravichandran. A closed form solution to robust subspace estimation and clustering. In *Proceedings of CVPR*, pages 1801–1807, 2011. 1, 2

[8] T. Gunji, S. Kim, M. Kojima, A. Takeda, K. Fujisawa, and T. Mizutani. Phom—a polyhedral homotopy continuation method. *Computing*, 73:57–77, 2004. 3

[9] K. C. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. Pattern Anal. Machine Intell.*, 27:684–698, 2005. 6

[10] T. L. Lee, T. Y. Li, and C. H. Tsai. Hom4ps-2.0: a software package for solving polynomial systems by the polyhedral homotopy continuation method. *Computing*, 83:109–133, 2008. 3

[11] G. Liu, Z. Lin, and Y. Yu. Robust subspace segmentation by low-rank representation. In *Proc. of ICML*, pages 663–670, 2010. 1, 2

[12] G. Liu, H. Xu, and S. Yan. Exact subspace segmentation and outlier detection by low-rank representation. In *Proc. of AISTATS*, 2012. 1, 2

[13] G. Liu and S. Yan. Latent low-rank representation for subspace segmentation and feature extraction. In *Proc. of ICCV*, 2011. 1, 2

[14] S. Nakajima, M. Sugiyama, and S. D. Babacan. Variational Bayesian sparse additive matrix factorization. *Machine Learning*, 92:319–1347, 2013. 1, 4, 5

[15] S. Nakajima, M. Sugiyama, S. D. Babacan, and R. Tomioka. Global analytic solution of fully-observed variational Bayesian matrix factorization. *Journal of Machine Learning Research*, 14:1–37, 2013. 2

[16] S. Nakajima, A. Takeda, S. D. Babacan, M. Sugiyama, and I. Takeuchi. Global solver and its efficient approximation for variational Bayesian low-rank subspace clustering. In *Advances in Neural Information Processing Systems 26*, 2013. 1, 2, 3, 5, 7

[17] S. Nakajima, R. Tomioka, M. Sugiyama, and S. D. Babacan. Condition for perfect dimensionality recovery by variational Bayesian PCA. *Journal of Machine Learning Research*, 16:3757–3811, 2015. 1, 4, 5, 7

[18] R. M. Neal. *Bayesian Learning for Neural Networks*. Springer, 1996. 1

[19] M. Seeger and G. Bouchard. Fast variational Bayesian inference for non-conjugate matrix factorization models. In *Proceedings of International Conference on Artificial Intelligence and Statistics*, La Palma, Spain, 2012. 1

[20] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Machine Intell.*, 22(8):888–905, 2000. 2

[21] M. Soltanolkotabi and E. J. Candès. A geometric analysis of subspace clustering with outliers. *CoRR*, 2011. 1, 2

[22] R. Tron and R. Vidal. A benchmark for the comparison of 3-D motion segmentation algorithms. In *Proc. of CVPR*, 2007. 5