LOW-RANK MATRIX COMPLETION BY VARIATIONAL SPARSE BAYESIAN LEARNING

S. Derin Babacan[1], *Martin Luessi*[2], *Rafael Molina*[3], *Aggelos K. Katsaggelos*[2]

[1] Beckman Institute
University of Illinois
at Urbana-Champaign, USA
dbabacan@illinois.edu

[2]Department of Electrical Engineering
and Computer Science
Northwestern University, USA
mluessi@northwestern.edu
aggk@eecs.northwestern.edu

[3]Departamento de Ciencias
de la Computación e I.A.
Universidad de Granada, Spain
rms@decsai.ugr.es

ABSTRACT

There has been a significant interest in the recovery of low-rank matrices from an incomplete of measurements, due to both theoretical and practical developments demonstrating the wide applicability of the problem. A number of methods have been developed for this recovery problem, however, a principled method for choosing the unknown target rank is generally missing. In this paper, we present a recovery algorithm based on sparse Bayesian learning (SBL) and automatic relevance determination principles. Starting from a matrix factorization formulation and enforcing the low-rank constraint in the estimates as a sparsity constraint, we develop an approach that is very effective in determining the correct rank while providing high recovery performance. We provide empirical results and comparisons with current state-of-the-art methods that illustrate the potential of this approach.

***Index Terms*—** Low-rank matrix completion, Bayesian methods, automatic relevance determination.

1. INTRODUCTION

The problem of low-rank matrix completion (and approximation) recently received significant interest due to new theoretical advances [1,2] as well as interesting practical problems (e.g., the *Netflix prize*). Matrix completion finds application in many areas of engineering, including system identification [3], sensor networks [4], machine learning [5], computer vision [6], and medical imaging [7].

The matrix completion problem is formulated as follows. Let $\mathbf{X} \in \mathbb{R}^{m\times n}$ be an unknown matrix with rank $r \ll \min(m,n)$. Suppose that we only observe a subset Ω of its entries, that is, $\{Y_{ij} = X_{ij} : (i,j) \in \Omega\}$. The cardinality of Ω is pmn with $0 < p \leq 1$. It has been shown in [1] that most matrices $\mathbf{X}$ can be recovered very accurately under certain conditions by solving the affine rank minimization problem

$$\begin{array}{ll} \text{minimize} & \operatorname{rank}(\mathbf{X}) \\ \text{subject to} & \mathcal{P}_\Omega(\mathbf{Y}) = \mathcal{P}_\Omega(\mathbf{X}), \end{array} \tag{1}$$

where $\mathcal{P}_\Omega$ is the projection operator such that the $(i,j)^{\text{th}}$ component of $\mathcal{P}_\Omega(\mathbf{X})$ is equal to X_{ij} if $(i,j)\in\Omega$ and zero otherwise, and $\mathbf{Y}$ are the observations. Since this problem is NP-hard, a popular approach is to utilize the convex relaxation based on the nuclear norm. When the observations are corrupted with noise, this problem can be stated as

$$\begin{array}{ll} \text{minimize} & \|\mathbf{X}\|_* \\ \text{subject to} & \| \mathcal{P}_\Omega(\mathbf{Y}) - \mathcal{P}_\Omega(\mathbf{X}) \|_{\mathcal{F}}^2 < \epsilon, \end{array} \tag{2}$$

where $\|\mathbf{X}\|_*$ is equal to the sum of the singular values of $\mathbf{X}$ and $\|\cdot\|_{\mathcal{F}}$ denotes the Frobenius norm.

A number of methods have been proposed for the low-rank matrix recovery problem. The nuclear norm based optimization problem can be recast as a semidefinite program, and can be solved with interior-point solvers [3]. Singular value thresholding [8] provides an attractive alternative in terms of computation. FPCA [9] introduced an efficient nuclear norm-based regularized least-squares method, whereas OPTSPACE [10] developed a method based on optimization over the Grasmann manifold with a theoretical performance guarantee for the noiseless case. A greedy approach is proposed in ADMIRA [11]. Finally, Bayesian methods have also been developed: a nonparametric approach for symmetric positive definite matrices is proposed in [12], and a variational Bayes method is developed for collaborative filtering in [13].

Although several methods have been developed for this problem, a principled method for choosing the unknown target rank is generally not motivated. In this paper, we present a recovery algorithm based on sparse Bayesian learning (SBL) principles. Based on the low-rank factorization of the unknown matrix, we employ independent sparsity priors on the individual factors with a common sparsity profile which favors low-rank solutions and simultaneously explain the observed data. Our formulation offers a few advantages over other approaches. Firstly, prior knowledge on the rank of the matrix is not required; the proposed formulation implicitly estimates the rank of the unknown matrix similarly to the automatic relevance determination principle in machine learning [14]. This property is not present in most of the proposed approaches (for instance, [10, 11]). Second, algorithmic parameters are treated as stochastic quantities in the proposed approach, and are handled with the combination of prior distributions and a fully-Bayesian inference procedure. In this regard, this type of formulation frees the user from extensive parameter-tuning and data- and application-dependent supervision. Finally, empirical results demonstrate that the proposed method provides very good reconstruction performance compared to existing methods while accurately estimating the unknown effective rank.

This paper is organized as follows. We present the proposed modeling of the problem in Section 2. Section 3 develops the estimation algorithm based on variational Bayesian inference. We provide empirical results in Section 4, and conclude in Section 5.

2. PROPOSED MODELING

Assume that the unknown $m \times n$ matrix $\mathbf{X}$ is of rank r. Our modeling is based on the following parametrization of $\mathbf{X}$

$$\mathbf{X} = \mathbf{A}\mathbf{B}^T\,, \tag{3}$$

where $\mathbf{A}$ is an $m \times r$ matrix, and $\mathbf{B}$ an $n \times r$ matrix, such that $\mathrm{rank}(\mathbf{X}) = r \leq \min(m, n)$. The factors $\mathbf{A}$ and $\mathbf{B}$ can then be estimated using

$$\begin{array}{ll} \text{minimize} & \| \mathbf{A} \|_{\mathcal{F}}^2 + \| \mathbf{B} \|_{\mathcal{F}}^2 \\ \text{subject to} & \| \mathcal{P}_{\Omega}(\mathbf{Y}) - \mathcal{P}_{\Omega}(\mathbf{X}) \|_{\mathcal{F}}^2 < \epsilon, \end{array} \tag{4}$$

The equivalence of this optimization problem to (2) is easy to show (see [15]). We formulate the problem (4) using the Bayesian methodology as follows. $\mathbf{X}$ is the sum of outer-products of the columns of $\mathbf{A}$ and $\mathbf{B}$, that is,

$$\mathbf{X} = \sum_{i}^{k} \mathbf{a}_{\cdot i} \mathbf{b}_{\cdot i}{}^{T}, \tag{5}$$

where we use $\mathbf{a}_{\cdot i}$ and $\mathbf{a}_{i \cdot}$ to denote the i^{th} column and row of $\mathbf{A}$, respectively. Notice that each outer-product contributes at most one to the rank to $\mathbf{X}$. Since a low-rank estimate of $\mathbf{X}$ is sought, our goal is to achieve column sparsity in $\mathbf{A}$ and $\mathbf{B}$, such that most columns in $\mathbf{A}$ and in $\mathbf{B}$ are set equal to zero. To this end, we associate the columns of $\mathbf{A}$ and $\mathbf{B}$ with Gaussian priors of variances γ_i, that is,

$$\mathrm{p}(\mathbf{A}|\boldsymbol{\gamma}) = \prod_{i=1}^{k} \mathcal{N}\left(\mathbf{a}_{\cdot i}|\mathbf{0}, \gamma_i \mathbf{I}\right), \tag{6}$$

$$\mathrm{p}(\mathbf{B}|\boldsymbol{\gamma}) = \prod_{i=1}^{k} \mathcal{N}\left(\mathbf{b}_{\cdot i}|\mathbf{0}, \gamma_i \mathbf{I}\right). \tag{7}$$

Thus, the columns of $\mathbf{A}$ and $\mathbf{B}$ have the same sparsity profile enforced by the common variances γ_i. As shown later, many of the variances γ_i will assume very small values during inference, which effectively removes the corresponding outer-products from $\mathbf{X}$, and hence reduces the rank of the estimate. This formulation therefore is the analog of sparse Bayesian learning formulation (or automatic relevance determination) [14] successfully utilized for compressive sensing reconstruction, where sparsity-inducing Gaussian priors are employed on each of the coefficients of the unknown vector.

As the observation model we follow the standard assumption and incorporate white Gaussian noise on the observations $\mathcal{P}_{\Omega}(\mathbf{Y})$, such that

$$\mathrm{p}(\mathcal{P}_{\Omega}(\mathbf{Y})|\mathbf{A}, \mathbf{B}, \beta) = \prod_{(i,j)\in\Omega} \mathcal{N}\left(y_{ij}|x_{ij}, \beta^{-1}\right), \tag{8}$$

with $\beta = 1/\epsilon$ the noise precision. The joint distribution, therefore, is expressed as

$$\begin{aligned} \mathrm{p}(\mathcal{P}_{\Omega}(\mathbf{Y}), \mathbf{A}, \mathbf{B}, \boldsymbol{\gamma}, \beta) = {} & \mathrm{p}(\mathcal{P}_{\Omega}(\mathbf{Y})|\mathbf{A}, \mathbf{B}, \beta) \\ & \times \mathrm{p}(\mathbf{A}|\boldsymbol{\gamma})\, \mathrm{p}(\mathbf{B}|\boldsymbol{\gamma})\, \mathrm{p}(\boldsymbol{\gamma})\, \mathrm{p}(\beta). \end{aligned} \tag{9}$$

In addition to (6) and (7), we incorporate the conjugate inverse Gamma hyperprior on the variances γ_i

$$\mathrm{p}(\gamma_i) \propto \left(\frac{1}{\gamma_i}\right)^{a+1} \exp\left(-\frac{b}{\gamma_i}\right). \tag{10}$$

In this work, the parameters a and b are treated as deterministic whose values are set by the user.

3. APPROXIMATE BAYESIAN INFERENCE

Exact Bayesian inference for all unknowns $\mathbf{A}$,$\mathbf{B}$,$\boldsymbol{\gamma}$ and β using the joint distribution in (9) is intractable, since $\mathrm{p}(\mathbf{y})$ cannot be computed. Therefore, approximation methods must be utilized. In this work, we present an inference procedure based on mean field variational Bayes [16]. Our goal is to compute posterior distribution approximations by minimizing the Kullback-Leibler (KL) divergence in an alternating fashion for each latent variable. Let $\mathbf{z} = (\mathbf{A}, \mathbf{B}, \boldsymbol{\gamma}, \gamma, \beta)$ be the vector of all latent variables. The posterior approximation $\mathrm{q}(\mathbf{z}_k)$ of each latent variable $\mathbf{z}_k \in \mathbf{z}$ is found using

$$\log \mathrm{q}(\mathbf{z}_k) = \langle \log \mathrm{p}(\mathcal{P}_{\Omega}(\mathbf{Y}), \mathbf{A}, \mathbf{B}, \boldsymbol{\gamma}, \beta) \rangle_{\mathbf{z}\setminus\mathbf{z}_k} + \mathrm{const}, \tag{11}$$

where we have employed the factorization $\mathrm{q}(\mathbf{z}) = \prod \mathrm{q}(\mathbf{z}_k)$ and $\mathbf{z}\setminus\mathbf{z}_k$ denotes the set $\mathbf{z}$ with $\mathbf{z}_k$ removed. Thus, for each latent variable, the expectations of all parameters (excluding the current one) in the joint distribution (9) are taken with respect to their most recent distributions, and the result is normalized to find the approximate posterior distribution. Since all distributions in the hierarchical model presented in the previous section are in the conjugate exponential family, the calculation of each posterior approximation is relatively straightforward. We present the update rules resulting from this inference scheme in the following subsections.

3.1. Estimation of factors A and B

With some algebra, it follows from (11) that the approximation to the posterior distributions of $\mathbf{A}$ and $\mathbf{B}$ decompose as independent distributions of their rows. By combining the prior in (6) and the observation model in (8), the posterior density of the i^{th} row $\mathbf{a}_{i\cdot}$ of $\mathbf{A}$ is found as

$$\mathrm{q}(\mathbf{a}_{i\cdot}) = \mathcal{N}\left(\mathbf{a}_{i\cdot}|\langle\mathbf{a}_{i\cdot}\rangle, \boldsymbol{\Sigma}_i^a\right), \tag{12}$$

with mean and covariance

$$\langle\mathbf{a}_{i\cdot}\rangle^T = \beta\, \boldsymbol{\Sigma}_i^a \,\langle\mathbf{B}_i\rangle^T\ \mathbf{y}_{i\cdot}{}^T, \tag{13}$$

$$\boldsymbol{\Sigma}_i^a = \left(\beta \,\langle\mathbf{B}_i^T\mathbf{B}_i\rangle + \boldsymbol{\Gamma}\right)^{-1}, \tag{14}$$

where $\boldsymbol{\Gamma} = \mathrm{diag}(\gamma_i^{-1})$ and the matrix $\mathbf{B}_i$ contains only the j^{th} rows of $\mathbf{B}$ for which $(i, j) \in \Omega$, such that,

$$\langle\mathbf{B}_i^T\mathbf{B}_i\rangle = \sum_{j:(i,j)\in\Omega} \langle\mathbf{b}_{j\cdot}{}^T\,\mathbf{b}_{j\cdot}\rangle = \sum_{j:(i,j)\in\Omega} \langle\mathbf{b}_{j\cdot}{}^T\rangle\langle\mathbf{b}_{j\cdot}\rangle + \boldsymbol{\Sigma}_j^b,$$

with $\boldsymbol{\Sigma}_j^b$ the posterior covariance of j^{th} row of $\mathbf{B}$. Additionally, the row vector $\mathbf{y}_{i\cdot}$ contains the observed entries in the i^{th} row of $\mathbf{Y}$. Similarly, by combining the prior in (7) and the observation model in (8), the posterior density of the j^{th} row $\mathbf{b}_{j\cdot}$ of $\mathbf{B}$ is found as a normal distribution

$$\mathrm{q}(\mathbf{b}_{j\cdot}) = \mathcal{N}\left(\mathbf{b}_{j\cdot}|\langle\mathbf{b}_{j\cdot}\rangle, \boldsymbol{\Sigma}_j^b\right) \tag{15}$$

with parameters

$$\langle\mathbf{b}_{j\cdot}\rangle^T = \beta\, \boldsymbol{\Sigma}_j^b \,\langle\mathbf{A}_j\rangle^T \mathbf{y}_{\cdot j}, \tag{16}$$

$$\boldsymbol{\Sigma}_j^b = \left(\beta \,\langle\mathbf{A}_j^T\mathbf{A}_j\rangle + \boldsymbol{\Gamma}\right)^{-1}, \tag{17}$$

where $\mathbf{A}_j$ contains the i^{th} rows of $\mathbf{A}$ for which $(i, j) \in \Omega$, and the vector $\mathbf{y}_{\cdot j}$ is constructed from the observed entries in the j^{th} column of $\mathbf{Y}$. It can be observed that the uncertainty in the estimate of $\mathbf{B}$ is incorporated in the estimation of $\mathbf{A}$ through the covariance matrices $\boldsymbol{\Sigma}_i^b$ (and vice versa).

3.2. Estimation of hyperparameters $\boldsymbol{\gamma}$

By combining $\mathrm{p}(\mathbf{A}|\boldsymbol{\gamma})$, $\mathrm{p}(\mathbf{B}|\boldsymbol{\gamma})$ and $\mathrm{p}(\gamma_i)$, the posterior density of γ_i becomes an inverse Gamma distribution

$$\mathrm{q}(\gamma_i) \propto \left(\frac{1}{\gamma_i}\right)^{a+1+\frac{m+n}{2}} \exp\left(-\frac{2b + \langle \mathbf{a}_{\cdot i}{}^T \mathbf{a}_{\cdot i}\rangle + \langle \mathbf{b}_{\cdot i}{}^T \mathbf{b}_{\cdot i}\rangle}{2\gamma_i}\right) \tag{18}$$

with mean

$$\langle \gamma_i \rangle = \frac{2b + \langle \mathbf{a}_{\cdot i}{}^T \mathbf{a}_{\cdot i}\rangle + \langle \mathbf{b}_{\cdot i}{}^T \mathbf{b}_{\cdot i}\rangle}{2a + m + n}\,. \tag{19}$$

The required expectations are given by

$$\langle \mathbf{a}_{\cdot i}{}^T \mathbf{a}_{\cdot i}\rangle = \langle \mathbf{a}_{\cdot i}\rangle^T \langle \mathbf{a}_{\cdot i}\rangle + \sum_j \left(\boldsymbol{\Sigma}_j^a\right)_{ii}, \tag{20}$$

$$\langle \mathbf{b}_{\cdot i}{}^T \mathbf{b}_{\cdot i}\rangle = \langle \mathbf{b}_{\cdot i}\rangle^T \langle \mathbf{b}_{\cdot i}\rangle + \sum_j \left(\boldsymbol{\Sigma}_j^b\right)_{ii}. \tag{21}$$

3.3. Estimation of noise precision β

Assuming a conjugate Gamma prior for $\mathrm{p}(\beta)$ with parameters c and d, the posterior approximation assumes a Gamma distribution with the mean

$$\langle \beta \rangle = \frac{2d + pmn}{2c + \langle \parallel \mathcal{P}_\Omega(\mathbf{Y}) - \mathcal{P}_\Omega(\mathbf{A}\mathbf{B}^T) \parallel_{\mathcal{F}}^2 \rangle}\,, \tag{22}$$

However, this estimation may lead to identifiability problems, and the calculations of the expectations are computationally complex and have high memory requirements. In practice, we found out that the algorithm is quite robust to this parameter and setting it to a reasonable value leads to good empirical results.

In summary, the algorithm proceeds as first estimating the rows of $\mathbf{A}$ and $\mathbf{B}$ using (13) and (16), followed by the estimation of the variances γ_i using (19), and (if desired) the noise precision β using (22). Notice that during inference most of the hyperparameters γ_i are driven to zero, which will force the posterior means of the columns to go to zero as well (see (14) and (17)). In our implementation, columns of $\mathbf{A}$ and $\mathbf{B}$ were declared irrelevant if the corresponding $\gamma_i < 10^{-10}$ for $a = 10^5$ and $b = 10^{-5}$. Other selections of a and b (including zero values) resulted in similar reconstruction errors but different convergence speeds. In each case, the threshold value of γ_i should be chosen according to minimum value possible in (19) (i.e., when $\langle \mathbf{a}_{\cdot i}{}^T \mathbf{a}_{\cdot i}\rangle \approx 0$ and $\langle \mathbf{b}_{\cdot i}{}^T \mathbf{b}_{\cdot i}\rangle \approx 0$).

The methodology presented in this work resembles the method in [13] proposed for collaborative filtering, where independent Gaussian priors are placed on the columns of $\mathbf{A}$ and $\mathbf{B}$ with separate sets of variances. Although the modeling is similar, the columns of $\mathbf{A}$ and $\mathbf{B}$ are not coupled through the use of common variances as in our work. Employing common parameters is of crucial importance in removing redundant components from the estimated matrix and determining the effective rank. In theory, the modeling in (6) and (7) with common variances is used to represent the correlation between the columns of $\mathbf{A}$ and $\mathbf{B}$, and it also removes possible scale problems arised due to the use of separate sets of variances.

The computational complexity of the algorithm can be shown to be $\mathrm{O}(m \cdot \min(p^3 n^3, k^3) + n \cdot \min(p^3 m^3, k^3))$ with p the fraction of observed entries (proof not shown for space limitations). The bottleneck of the algorithm is the computations in (14) and (17). Note, however, that for instance (17) can also be calculated using the Woodbury identity as

$$\boldsymbol{\Sigma}_j^b = \boldsymbol{\Gamma}^{-1} - \boldsymbol{\Gamma}^{-1}\langle \mathbf{A}_j\rangle^T \left(\beta^{-1}\mathbf{I} + \langle \mathbf{A}_j\rangle \boldsymbol{\Gamma}^{-1}\langle \mathbf{A}_j\rangle^T\right)^{-1} \langle \mathbf{A}_j\rangle \boldsymbol{\Gamma}^{-1}. \tag{23}$$

Depending on the dimensions of the estimates of $\mathbf{A}$ and $\mathbf{B}$, these two forms can be alternated to achieve faster estimation.

4. EMPIRICAL COMPARISONS

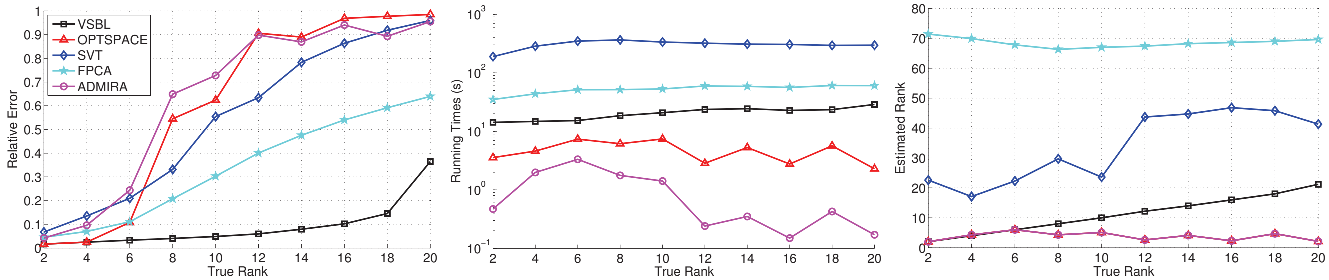
To examine the empirical performance of the proposed method compared to existing algorithms, we performed simulations commonly used in the literature. Our first example illustrates the effectiveness of the proposed approach on determining the correct rank. We generated test matrices $\mathbf{X}$ of size 200×200 of ranks $r = 2, \ldots, 20$ by randomly sampling $200 \times r$ matrices $\mathbf{X}_L$ and $\mathbf{X}_R$ from a standard normal distribution $\mathcal{N}(0, 1)$ and setting $\mathbf{X} = \mathbf{X}_L \mathbf{X}_R^T$. The fraction of observed entries is 0.2, and they are sampled uniformly at random. For each experiment, the relative recovery error is measured as $\parallel \hat{\mathbf{X}} - \mathbf{X} \parallel_{\mathcal{F}} / \parallel \mathbf{X} \parallel_{\mathcal{F}}$, where $\hat{\mathbf{X}}$ is the estimate.

We present comparisons with the following algorithms: OPTSPACE [10], SVT [8], FPCA [9] and ADMIRA [11]. Our method is denoted by VSBL. We used the procedure proposed in [10] to estimate the initial target rank required by ADMIRA and OPT. On the other hand, other methods automatically estimate the rank of the unknown matrix. The observed entries are corrupted by zero-mean white Gaussian noise with standard deviation 0.05. Each simulation result is obtained by averaging 10 random instances. Figure 1 shows the relative reconstruction error, running times (on a 3GHz Core2 Duo CPU) and estimated ranks for each algorithm. Among all algorithms, VSBL provides the highest recovery performance for all ranks, and also estimates the correct rank in each case. As expected, errors in both the recovery and the estimated rank increase as the original rank increases. OPTSPACE and ADMIRA consistently underestimate the rank, whereas FPCA and SVT overestimate it. Overall, VSBL exhibits a better ability to recover the original matrix and the correct rank than other methods.

We next consider another set of experimental conditions where 200×200 matrices of fixed rank of 5 are generated, and the number of observed entries is varied according to different oversampling degrees of freedom. Note that a matrix of size $m \times n$ of rank r depends upon $\mathrm{df} = r(m + n - r)$ degrees of freedom, and the oversampling degrees of freedom (osdf) is defined as pmn/df [17]. Experimental results for $\mathrm{osdf} = 2, 3, \ldots, 10$ are depicted in Figure 2 for the same noise conditions as above. The corresponding sampling ratios are $p \approx 0.1, 0.14, 0.20, 0.24, 0.30, 0.34, 0.40, 0.44, 0.50$. It is clear that VSBL provides very accurate reconstructions even with very low number of observations, for which other algorithms fail to provide meaningful results. In terms of computation time, ADMIRA provided the best performance in most of the simulations, whereas execution times for VSBL were stable throughout the testing conditions and were comparable to compared to the other methods.

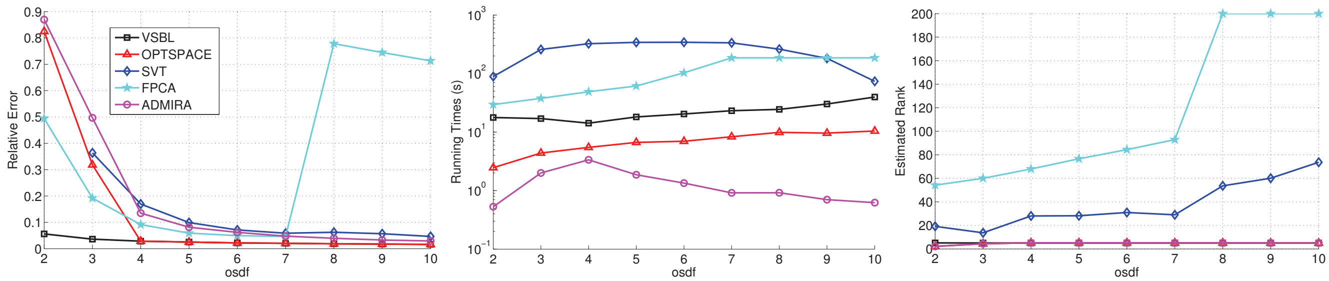
5. CONCLUSIONS

In this paper, we have applied sparse Bayesian learning principles to the low-rank matrix completion problem using a variational Bayesian perspective. We introduced a formulation where the low-rank constraint is imposed on the estimate by using its sparse representation; starting from the factorized form of the unknown matrix, we enforce a common sparsity profile on its underlying components using a probabilistic formulation. We then developed



(a) (b) (c)

Fig. 1. Estimation results with matrices of size 200×200 with varying ranks when 20% of the entries are observed. (a) Relative recovery error, (b) running times, and (c) estimated ranks.



(a) (b) (c)

Fig. 2. Estimation results with matrices of size 200×200 of rank 5 with varying oversampling degrees of freedom. (a) Relative recovery error, (b) running times, and (c) estimated ranks. Error rates of SVT for osdf $= 2$ and of FPC for osdf $= 8, 9, 10$ are very high due to convergence failures.

an inference method based on mean-field variational Bayes approximating the posteriors of interest. Empirical results suggest that the proposed approach is very effective in pruning irrelevant dimensions and recover the correct number of effective components in the matrix estimate, and it outperforms current state-of-the-art approaches in terms of reconstruction performance.

6. REFERENCES

[1] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. of Comput. Math.*, vol. 9, pp. 717–772, 2008.

[2] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans. Inform. Theory*, vol. 56, no. 5, pp. 2053–2080, 2009.

[3] Z. Liu and L. Vandenberghe, "Interior-point method for nuclear norm approximation with application to system identification," *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 3, pp. 1235–1256, 2009.

[4] S. Oh, A. Karbasi, and A. Montanari, "Sensor Network Localization from Local Connectivity : Performance Analysis for the MDS-MAP Algorithm," in *IEEE Information Theory Workshop (ITW 2010)*, 2010.

[5] N. Srebro, "Learning with matrix factorizations," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA, 2004.

[6] H. Ji, C. Liu, Z. Shen, and Y. Xu, "Robust video denoising using low rank matrix completion," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010)*, 2010.

[7] Z.-P. Liang, "Spatiotemporal imaging with partially separable functions," in *NFSI-ICFBI 2007*, 2007, pp. 181 –182.

[8] J.-F. Cai, E. J. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2008.

[9] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and Bregman iterative methods for matrix rank minimization," *arXiv:0905.1643v2*, 2009.

[10] R. H. Keshavan, S. Oh, and A. Montanari, "Matrix completion from a few entries," *submitted to IEEE Trans. Inf. Theory, arXiv:0901.3150v2*, 2009.

[11] K. Lee and Y. Bresler, "ADMiRA: Atomic decomposition for minimum rank approximation," *arXiv:0905.0044*, 2009.

[12] J. Paisley and L. Carin, "A nonparametric Bayesian model for kernel matrix completion," in *ICASSP 2010*, Dallas, USA.

[13] Y. J. Lim and Y. W. Teh, "Variational Bayesian approach to movie rating prediction," in *Proceedings of KDD Cup and Workshop*, 2007.

[14] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, pp. 211–244, 2001.

[15] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *to appear in SIAM review, arXiv:0706.4138v1*, 2007.

[16] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer-Verlag, 2006.

[17] E. J. Candès and Y. Plan, "Matrix completion with noise," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 925 – 936, 2009.