

Probabilistic Matrix Factorization from Quantized Measurements

Giulio Bottegal
ESAT-Stadius
KU Leuven

Leuven (Heverlee), Belgium, 3001
Email: giulio.bottegal@esat.kuleuven.be

Johan A.K. Suykens
ESAT-Stadius
KU Leuven

Leuven (Heverlee), Belgium, 3001
Email: johan.suykens@esat.kuleuven.be

Abstract—We consider the problem of factorizing a matrix with discrete-valued entries as a product of two low-rank matrices. Under a probabilistic framework, we seek for the minimum mean-square error estimates of these matrices, using full Bayes and empirical Bayes approaches. In the first case, we devise an integration scheme based on the Gibbs sampler that accounts also for hyperparameter and noise variance estimation. A similar technique is used also for the latter case, where we combine Gibbs sampling with the expectation-maximization (EM) algorithm to estimate the model parameters via marginal likelihood maximization. Extension to the case of missing values is also discussed. The proposed methods are evaluated on simulated data, and on a real data set for recommender systems.

I. INTRODUCTION

Matrix factorization –and matrix completion– is a fundamental topic in data science. It finds application in various fields, e.g., face recognition [5], image annotation [13], bioinformatics [9], and recommender systems [10], [20].

Typically, matrix factorization problems aim at decomposing a data matrix in terms of two low-rank factors that best explain the data according to some similarity measure. Usually, the problem is cast under a suitable regularization framework, e.g., to enforce regularity in the solution [10] and to prevent overfitting. However, the optimization problem associated with matrix factorization is in general non-convex, so specific optimization techniques have to be devised to solve it. For instance, under ℓ_2 -type regularization assumptions, alternated least-squares (ALS) techniques are usually adopted to compute the matrix factors [27], [6]. Other popular types of regularization involve the use of the ℓ_1 norm to induce sparsity [19], or the nuclear norm, to enforce low-rank solutions [25].

Another approach to overcome the ill-posedness of the problem is to assign a prior distribution to the unknown factors. This approach, usually referred to as Bayesian or probabilistic matrix factorization (PMF), has been first proposed in [21], [14], and later generalized in [22]. Extension of this framework, e.g., to deal with outliers, are discussed in [24], [17], where non-Gaussian noise models are adopted. Note that one can recover the regularization framework by choosing suitable priors and computing the maximum-a-posteriori estimates of the matrix factors.

In many applications, the entries of the data matrix can only assume discrete or quantized values. For instance, matrices reporting movie ratings assume integer values (e.g., ranging from 1 to 5). Accounting for this constraint is important, especially when the task is to predict missing values in the data

matrix (e.g., in the case of recommender systems). The problem of matrix completion from quantized entries has received increasing attention in recent years [3], [12]. In particular, [1] proposes a probabilistic approach for matrix completion from quantized entries. However, to our best knowledge, except for a few contributions [4], [11], quantized matrix factorization has received little attention.

In this paper, we propose a new method for matrix factorization with quantized entries. Similarly to [1], we assume that the data matrix is the result of a quantization process of an underlying matrix of interest, plus a noise matrix that has unknown variance. The underlying matrix is in turn the result of the product of two unknown low-rank matrices. Our task is to recover this structure from the data matrix, which may have missing entries.

We adopt a probabilistic approach, that is, we assign Gaussian prior distributions to the unknown matrix factors. Then, we estimate these factors by computing their minimum mean-square error (MMSE) estimates. To accomplish this task, we propose two methods. The first is based on a full Bayes approach. Here, we consider every parameter as a stochastic quantity. Similarly to [21], the MMSE estimates are computed using Markov Chain Monte Carlo (MCMC) methods, and, in particular, with a new procedure based on the Gibbs sampler [23]. The second method we proposed is based on an empirical Bayes approach. In this case, the prior hyperparameter and the noise variance are treated as deterministic parameters, which we estimate by marginal likelihood maximization. Also for this case, we devise a novel integration scheme based on the Gibbs sampler. Moreover, to solve the marginal likelihood problem, we combine the Gibbs sampler with the expectation-maximization (EM) algorithm [16], obtaining an effective iterative parameter estimation approach (sometimes called *empirical Bayes Gibbs sampling*, see [2]). The effectiveness of the proposed methods is evaluated both on synthetic data and on a real data set for recommender systems. In particular, we show the advantage of the proposed method compared to other probabilistic matrix factorization techniques (e.g., [21]).

A. Notation

Bold letters indicate matrices and vectors. \mathbf{I}_n denotes the identity matrix of size n ; “ \otimes ” and “ \odot ” denote the Kronecker and the Hadamard products, respectively. Given a matrix \mathbf{A} , $\|\mathbf{A}\|_F$ denotes its Frobenius norm, $\text{vec}[\mathbf{A}]$ its column-wise vectorization, and \mathbf{a}_i^T its i -th row. $\mathcal{N}(\mathbf{m}, \Sigma)$ denotes a

Gaussian random vector with mean \mathbf{m} and covariance matrix Σ ; $\mathcal{N}_a^b(m, \sigma^2)$ denotes a scalar truncated Gaussian random variable in the interval $[a, b]$; $\text{Gamma}(a, b)$ denotes a Gamma random variable with parameters a and b .

II. PROBLEM STATEMENT

We consider a matrix $\mathbf{Z} \in \mathbb{R}^{n \times m}$ resulting from the product of two unknown matrices $\mathbf{U} \in \mathbb{R}^{n \times r}$ and $\mathbf{V} \in \mathbb{R}^{m \times r}$, plus an additional noise matrix $\mathbf{E} \in \mathbb{R}^{n \times m}$:

$$\mathbf{Z} = \mathbf{U}\mathbf{V}^T + \mathbf{E}. \quad (1)$$

We assume that the entries of \mathbf{E} are i.i.d. realizations of Gaussian r.v.'s, that is

$$e_{ij} \sim \mathcal{N}(0, \sigma^2), \quad (2)$$

where the variance σ^2 is unknown.

The matrix \mathbf{Z} is not directly measurable by the experimenter; only an entry-wise quantized version is available. Denoting by \mathbf{Y} this quantized version, we have that

$$y_{ij} = \mathcal{Q}[z_{ij}], \quad (3)$$

where \mathcal{Q} is a known map of the type

$$\mathcal{Q}[x] = s_k \quad \text{if } x \in (q_{k-1}, q_k], \quad (4)$$

with $s_k \in \{s_1, \dots, s_Q\}$ and $q_k \in \{q_0, \dots, q_Q\}$ (and typically $q_0 = -\infty$ and $q_Q = +\infty$).

The main problem we address in this paper can be formulated as follows.

Problem 1 (Matrix factorization): From the data matrix \mathbf{Y} , reconstruct the matrices $(\mathbf{Z}, \mathbf{U}, \mathbf{V})$ that best match the model (1).

We shall first focus on this problem. As an extension of Problem 1, we also study the case where \mathbf{Y} has some missing entries that we wish to predict. To this end, we introduce the matrix $\mathbf{W} \in \mathbb{R}^{n \times m}$ such that

$$w_{ij} = \begin{cases} 1 & \text{if } y_{ij} \text{ is available} \\ 0 & \text{if } y_{ij} \text{ is not available} \end{cases}. \quad (5)$$

Then we formulate the following problem:

Problem 2 (Matrix completion): From the data matrix $\mathbf{W} \odot \mathbf{Y}$, reconstruct the matrices $(\mathbf{Z}, \mathbf{U}, \mathbf{V})$ –and consequently the missing entries of \mathbf{Y} – that best match the model (1).

In the next section, we introduce the probabilistic approach adopted to tackle the two problems.

III. PROBABILISTIC MODELING AND ESTIMATION FOR QUANTIZED MATRIX FACTORIZATION

A. Priors on \mathbf{U} and \mathbf{V}

Without further assumptions on the unknown matrices \mathbf{U} and \mathbf{V} , the problem is ill-posed. Assuming \mathbf{Z} given, for any noise matrix \mathbf{E} there exists an infinite amount of pairs (\mathbf{U}, \mathbf{V}) such that (1) holds, because $\mathbf{U}\mathbf{A}\mathbf{A}^{-1}\mathbf{V}^T = \mathbf{U}\mathbf{V}^T$ holds for any full-rank matrix $\mathbf{A} \in \mathbb{R}^{r \times r}$. To overcome this issue, we pose prior distributions on \mathbf{U} and \mathbf{V} given by

$$u_{ij} \sim \mathcal{N}(0, \lambda) \quad (6)$$

$$v_{ij} \sim \mathcal{N}(0, 1). \quad (7)$$

Therefore, each entry of the matrices \mathbf{U} and \mathbf{V} is a Gaussian r.v. independent of the other entries. Furthermore, the variance of the entries of \mathbf{U} is regulated by the hyperparameter $\lambda > 0$, which we assume unknown. The variance of the v_{ij} is fixed to the arbitrary value 1, because the bilinear relation between \mathbf{U} and \mathbf{V} implies the impossibility of determining their magnitude (there is always a scaling factor that can be moved from \mathbf{U} to \mathbf{V} and vice versa). We do not impose any prior on λ and σ^2 .

Other priors for \mathbf{U} and \mathbf{V} that may lead to different model assumptions are, e.g., the Laplacian prior, which enforces sparsity, or the Bernoulli prior, to account for binary entries.

The related Bayesian network describing the system is depicted in Figure 1.

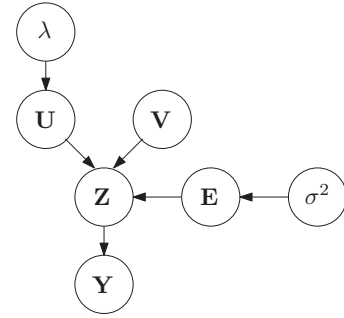


Figure 1. Bayesian network describing the model adopted for quantized matrix factorization.

B. Estimation of the model (1)

Given the probabilistic description of the previous section, for fixed values of λ and σ^2 the minimum mean-square error (MMSE) estimate of $(\mathbf{Z}, \mathbf{U}, \mathbf{V})$ can be expressed in terms of the integral

$$(\hat{\mathbf{Z}}, \hat{\mathbf{U}}, \hat{\mathbf{V}}) = \int (\mathbf{Z}, \mathbf{U}, \mathbf{V}) p(\mathbf{Z}, \mathbf{U}, \mathbf{V} | \mathbf{Y}) d\mathbf{Z} d\mathbf{U} d\mathbf{V}, \quad (8)$$

where $p(\mathbf{Z}, \mathbf{U}, \mathbf{V} | \mathbf{Y})$ is the posterior distribution of $(\mathbf{Z}, \mathbf{U}, \mathbf{V})$ given the data \mathbf{Y} . To account for the estimation of the parameters λ and σ^2 , we propose the following two approaches.

1) *Full Bayes approach:* The first approach treats also λ and σ^2 as random variables, mutually independent and independent of $(\mathbf{Z}, \mathbf{U}, \mathbf{V})$, with flat priors accounting of their positivity. Then, the MMSE estimate of $(\mathbf{Z}, \mathbf{U}, \mathbf{V})$ becomes

$$(\hat{\mathbf{Z}}, \hat{\mathbf{U}}, \hat{\mathbf{V}}) = \int (\mathbf{Z}, \mathbf{U}, \mathbf{V}) p(\mathbf{Z}, \mathbf{U}, \mathbf{V}, \lambda, \sigma^2 | \mathbf{Y}) \times d\mathbf{Z} d\mathbf{U} d\mathbf{V} d\lambda d\sigma^2, \quad (9)$$

where $p(\mathbf{Z}, \mathbf{U}, \mathbf{V}, \lambda, \sigma^2 | \mathbf{Y})$ is the joint posterior distribution of $(\mathbf{Z}, \mathbf{U}, \mathbf{V}, \lambda, \sigma^2)$.

2) *Empirical Bayes approach*: The second approach treats λ and σ^2 as deterministic quantities. Their estimation is carried out via marginal likelihood maximization, i.e.

$$\hat{\lambda}, \hat{\sigma}^2 = \arg \max_{\lambda \geq 0, \sigma^2 \geq 0} p(\mathbf{Y}; \lambda, \sigma^2), \quad (10)$$

where $p(\mathbf{Y}; \lambda, \sigma^2)$ is the marginal distribution of \mathbf{Y} (parameterized by λ and σ^2), defined as

$$p(\mathbf{Y}; \lambda, \sigma^2) = \int p(\mathbf{Y}, \mathbf{Z}, \mathbf{U}, \mathbf{V}; \lambda, \sigma^2) d\mathbf{Z} d\mathbf{U} d\mathbf{V}. \quad (11)$$

Once λ and σ^2 are fixed, the estimate of $(\mathbf{Z}, \mathbf{U}, \mathbf{V})$ can be computed using (8).

Both of the two approaches require computation of integrals that are not analytically tractable. Therefore, we shall make use of Monte Carlo approximations of the type

$$\hat{\mathbf{Z}} = \frac{1}{M} \sum_{k=1}^M \mathbf{Z}^k, \hat{\mathbf{U}} = \frac{1}{M} \sum_{k=1}^M \mathbf{U}^k, \hat{\mathbf{V}} = \frac{1}{M} \sum_{k=1}^M \mathbf{V}^k, \quad (12)$$

where $\mathbf{Z}^k, \mathbf{U}^k, \mathbf{V}^k$, are samples drawn from the posterior density of $(\mathbf{Z}, \mathbf{U}, \mathbf{V})$. In particular, the structure of the problem under study and the modeling assumptions of Section III allow for the use of the Gibbs sampler as detailed in the next sections.

Remark 1: Another estimation approach that we will not treat in this paper is based on the maximum-a-posteriori (MAP) technique, that is

$$\begin{aligned} \hat{\mathbf{Z}}, \hat{\mathbf{U}}, \hat{\mathbf{V}} &= \arg \max p(\mathbf{Z}, \mathbf{U}, \mathbf{V} | \mathbf{Y}) \\ &= \arg \max p(\mathbf{Y} | \mathbf{Z}) p(\mathbf{Z} | \mathbf{U}, \mathbf{V}) p(\mathbf{U}) p(\mathbf{V}). \end{aligned} \quad (13)$$

Taking the logarithm of the above problem, it can be shown that it is equivalent to solving

$$\begin{aligned} \hat{\mathbf{Z}}, \hat{\mathbf{U}}, \hat{\mathbf{V}} &= \arg \min \frac{1}{\sigma^2} \|\mathbf{Z} - \mathbf{U}\mathbf{V}^T\|_F^2 + \frac{1}{\lambda} \|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2 \\ \text{s.t. } & z_{ij}^{\min} < z_{ij} \leq z_{ij}^{\max} \end{aligned} \quad (14)$$

where z_{ij}^{\min} and z_{ij}^{\max} are determined by (3). The resulting problem is considerably harder than standard matrix factorization problems, so ALS are not directly applicable in this case.

IV. THE FULL BAYES APPROACH

We first describe the full Bayes approach to solve Problem 1. Recall that in this case, also the quantities λ and σ^2 are considered as random. Thus, they are to be considered in the integration scheme.

To use the Gibbs sampler, we must be able to sample from the conditional densities of the joint posterior $p(\mathbf{Z}, \mathbf{U}, \mathbf{V}, \lambda, \sigma^2 | \mathbf{Y})$. Let us introduce the vectors

$$\mathbf{z} := \text{vec}[\mathbf{Z}], \mathbf{u} := \text{vec}[\mathbf{U}], \mathbf{v} := \text{vec}[\mathbf{V}^T], \mathbf{e} := \text{vec}[\mathbf{E}]; \quad (15)$$

then we can write (1) either as

$$\mathbf{z} = (\mathbf{V} \otimes \mathbf{I}_n) \mathbf{u} + \mathbf{e}, \quad (16)$$

or as

$$\mathbf{z} = (\mathbf{I}_m \otimes \mathbf{U}) \mathbf{v} + \mathbf{e}. \quad (17)$$

Note that $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \lambda \mathbf{I}_{nr})$ and $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{mr})$. Then the conditionals are as follows:

- $z_{ij} | \{z_{hk}\}_{(h,k) \neq (i,j)}, \mathbf{U}, \mathbf{V}, \lambda, \sigma^2, \mathbf{Y}$. First note that, since

$$z_{ij} = \mathbf{u}_i^T \mathbf{v}_j + e_{ij}, \quad (18)$$

given \mathbf{U}, \mathbf{V} , for any i, j z_{ij} is independent of $\{z_{hk}\}_{(h,k) \neq (i,j)}$ and as such it can be sampled alone. Furthermore, from (18) it follows that $z_{ij} | \mathbf{U}, \mathbf{V}, \sigma^2 \sim \mathcal{N}(\mathbf{u}_i^T \mathbf{v}_j, \sigma^2)$. Knowing also that $y_{ij} = s_k$ for some $k \in \{1, \dots, Q\}$ permits to narrow the possible values taken by z_{ij} to the range $(q_{k-1}, q_k]$. Thus we obtain that

$$z_{ij} | \{z_{hk}\}_{(h,k) \neq (i,j)}, \mathbf{U}, \mathbf{V}, \lambda, \sigma^2, \mathbf{Y} \sim \mathcal{N}_{q_k - q_{k-1}}^{q_k}(\mathbf{u}_i^T \mathbf{v}_j, \sigma^2).$$

- $\mathbf{U} | \mathbf{Z}, \mathbf{V}, \lambda, \sigma^2, \mathbf{Y}$. Given \mathbf{V} , from (16) we see that there is a linear relation between \mathbf{Z} and \mathbf{U} , which are then jointly Gaussian. Therefore, using the vector notation for \mathbf{U} and \mathbf{Z} we have

$$\mathbf{u} | \mathbf{Z}, \mathbf{V}, \lambda, \sigma^2, \mathbf{Y} \sim \mathcal{N}(\mathbf{m}_u, \mathbf{P}_u), \quad (19)$$

where

$$\begin{aligned} \mathbf{P}_u &= \left(\frac{1}{\sigma^2} (\mathbf{V}^T \otimes \mathbf{I}_n) (\mathbf{V} \otimes \mathbf{I}_n) + \lambda^{-1} \mathbf{I}_{nr} \right)^{-1}, \\ \mathbf{m}_u &= \mathbf{P}_u (\mathbf{V}^T \otimes \mathbf{I}_n) \frac{\mathbf{z}}{\sigma^2}. \end{aligned} \quad (20)$$

- $\mathbf{V} | \mathbf{Z}, \mathbf{U}, \lambda, \sigma^2, \mathbf{Y}$. Given \mathbf{U} , from (17) we see that there is a linear relation between \mathbf{Z} and \mathbf{V} . Similarly to the previous case, we obtain

$$\mathbf{v} | \mathbf{Z}, \mathbf{U}, \lambda, \sigma^2, \mathbf{Y} \sim \mathcal{N}(\mathbf{m}_v, \mathbf{P}_v), \quad (21)$$

where

$$\begin{aligned} \mathbf{P}_v &= \left(\frac{1}{\sigma^2} (\mathbf{I}_m \otimes \mathbf{U}^T) (\mathbf{I}_m \otimes \mathbf{U}) + \mathbf{I}_{mr} \right)^{-1}, \\ \mathbf{m}_v &= \mathbf{P}_v (\mathbf{I}_m \otimes \mathbf{U}^T) \frac{\mathbf{z}}{\sigma^2}. \end{aligned} \quad (22)$$

- $\lambda | \mathbf{Z}, \mathbf{U}, \mathbf{V}, \sigma^2, \mathbf{Y}$. The posterior of λ depends only on \mathbf{U} . It is well-known [15] that, given the flat prior given to this parameter, the posterior of its inverse becomes

$$\lambda^{-1} | \mathbf{Z}, \mathbf{U}, \mathbf{V}, \sigma^2, \mathbf{Y} \sim \text{Gamma} \left(\frac{nr}{2}, \frac{\|\mathbf{U}\|_F^2}{2} \right).$$

- $\sigma^2 | \mathbf{Z}, \mathbf{U}, \mathbf{V}, \lambda, \mathbf{Y}$. From (1), given $\mathbf{Z}, \mathbf{U}, \mathbf{V}$, we can construct \mathbf{E} , which is sufficient to determine the conditional of σ^2 . Similarly to the previous case, we have

$$\sigma^{-2} | \mathbf{Z}, \mathbf{U}, \mathbf{V}, \lambda, \mathbf{Y} \sim \text{Gamma} \left(\frac{nm}{2}, \frac{\|\mathbf{Z} - \mathbf{U}\mathbf{V}^T\|_F^2}{2} \right).$$

Given the availability of all the conditional densities of the posterior $p(\mathbf{Z}, \mathbf{U}, \mathbf{V}, \lambda, \sigma^2 | \mathbf{Y})$ it is straightforward to compute (12) by sequentially drawing the relative samples from the above conditional densities. The resulting procedure is

Algorithm 1 Full-Bayes Quantized Probabilistic Matrix Factorization (F-Q-PMF)

```

1: procedure F-Q-PMF
2:   Input:  $\mathbf{Y}$ 
3:   Output:  $\hat{\mathbf{Z}}, \hat{\mathbf{U}}, \hat{\mathbf{V}}$ 
4:   for  $k = 1$  to  $M + M_0$  do
5:     for  $i = 1$  to  $n$  do
6:       for  $j = 1$  to  $m$  do
7:         draw  $z_{ij}^{(k)} \sim \mathcal{N}_{q_{k-1}}^{q_k}(\mathbf{u}_i^T \mathbf{v}_j, \sigma^2)$ 
8:         draw  $\mathbf{U}^{(k)} \sim \mathcal{N}(\mathbf{m}_u, \mathbf{P}_u)$ 
9:         draw  $\mathbf{V}^{(k)} \sim \mathcal{N}(\mathbf{m}_v, \mathbf{P}_v)$ 
10:        draw  $\lambda^{-1(k)} \sim \text{Gamma}\left(\frac{nr}{2}, \frac{\|\mathbf{U}\|_F^2}{2}\right)$ 
11:        draw  $\sigma^{-2(k)} \sim \text{Gamma}\left(\frac{nm}{2}, \frac{\|\mathbf{Z} - \mathbf{U}\mathbf{V}^T\|_F^2}{2}\right)$ 
12:         $\hat{\mathbf{Z}} \leftarrow \frac{1}{M} \sum_{k=1}^M \mathbf{Z}^k$ 
13:         $\hat{\mathbf{U}} \leftarrow \frac{1}{M} \sum_{k=1}^M \mathbf{U}^k$ 
14:         $\hat{\mathbf{V}} \leftarrow \frac{1}{M} \sum_{k=1}^M \mathbf{V}^k$ 

```

reported in Algorithm 1, where the integer M_0 is introduced for accounting for the *burn-in* phase of the Gibbs sampler.

In Appendix, we describe an efficient method for sampling from (19) and (21).

V. THE EMPIRICAL BAYES APPROACH

In this section we introduce the empirical Bayes approach to quantized probabilistic matrix factorization. We consider the parameters λ and σ^2 as deterministic. The whole procedure amount to solving (10) to get estimates of λ and σ^2 , which are in turn inserted into (11) to obtain $(\hat{\mathbf{Z}}, \hat{\mathbf{U}}, \hat{\mathbf{V}})$.

Assume λ and σ^2 fixed; similarly to the full Bayes case, computing (11) requires to devise a sampling technique to sample from $p(\mathbf{Y}, \mathbf{Z}, \mathbf{U}, \mathbf{V}; \lambda, \sigma^2)$. This can be done using the Gibbs sampler and following the same reasoning of the previous section. We get that:

- $z_{ij} | \{z_{hk}\}_{(h,k) \neq (i,j)}, \mathbf{U}, \mathbf{V}, \mathbf{Y}; \lambda, \sigma^2 \sim \mathcal{N}_{q_{k-1}}^{q_k}(\mathbf{u}_i \mathbf{v}_j^T, \sigma^2)$;
- $\mathbf{u} | \mathbf{Z}, \mathbf{V}, \mathbf{Y}; \lambda, \sigma^2, \mathbf{Y} \sim \mathcal{N}(\mathbf{m}_u, \mathbf{P}_u)$, where the first and second moments correspond to (20);
- $\mathbf{v} | \mathbf{Z}, \mathbf{U}, \mathbf{Y}; \lambda, \sigma^2, \mathbf{Y} \sim \mathcal{N}(\mathbf{m}_v, \mathbf{P}_v)$, where the first and second moments correspond to (22).

The corresponding procedure is reported in Algorithm 2. Now, we show how to use Algorithm 2 also to perform the marginal likelihood maximization (10). To this end, we introduce and use the EM algorithm. Let $\lambda^{(n)}, \sigma^{2(n)}$ be the parameter estimate obtained at the n -th iteration of the EM method. Then, $n + 1$ -th update is obtained with the following steps:

- (E-step) Compute

$$Q^{(n)}(\lambda, \sigma^2) := \mathbb{E} [\log p(\mathbf{Y}, \mathbf{Z}, \mathbf{U}, \mathbf{V}; \lambda, \sigma^2)] , \quad (23)$$

where the expectation is taken with respect to the posterior density $p(\mathbf{Y}, \mathbf{Z}, \mathbf{U}, \mathbf{V}; \lambda^{(n)}, \sigma^{2(n)})$, with λ and σ^2 fixed at the value $\lambda^{(n)}, \sigma^{2(n)}$;

Algorithm 2 Gibbs Sampler for empirical-Bayes Quantized Probabilistic Matrix factorization (GS4-E-Q-PMF)

```

1: procedure GS4-E-Q-PMF
2:   Input:  $\mathbf{Y}, \lambda, \sigma^2$ 
3:   Output:  $\{\mathbf{Z}^{(k)}\}_{k=1}^M, \{\mathbf{U}^{(k)}\}_{k=1}^M, \{\mathbf{V}^{(k)}\}_{k=1}^M$ 
4:   for  $k = 1$  to  $M + M_0$  do
5:     for  $i = 1$  to  $n$  do
6:       for  $j = 1$  to  $m$  do
7:         draw  $z_{ij}^{(k)} \sim \mathcal{N}_{q_{k-1}}^{q_k}(\mathbf{u}_i^T \mathbf{v}_j, \sigma^2)$ 
8:         draw  $\mathbf{U}^{(k)} \sim \mathcal{N}(\mathbf{m}_u, \mathbf{P}_u)$ 
9:         draw  $\mathbf{V}^{(k)} \sim \mathcal{N}(\mathbf{m}_v, \mathbf{P}_v)$ 

```

- (M-step) Compute

$$\hat{\lambda}^{(n+1)}, \hat{\sigma}^{2(n+1)} = \arg \max_{\lambda, \sigma^2} Q^{(n)}(\lambda, \sigma^2). \quad (24)$$

The parameter updates $\hat{\lambda}^{(n+1)}, \hat{\sigma}^{2(n+1)}$ increase the marginal likelihood compared to the estimates at the n -th step of the method [26], guaranteeing convergence to a local solution of (10).

We now turn our attention to the computation of the E-step. Using the derivation reported in Appendix, we see that

$$\begin{aligned} -2Q^{(n)}(\lambda, \sigma^2) &:= nm \log \sigma^2 + \frac{1}{\sigma^2} \mathbb{E} [\|\mathbf{Z} - \mathbf{U}\mathbf{V}^T\|_F^2] \\ &\quad + nr \log \lambda + \frac{1}{\lambda} \mathbb{E} [\|\mathbf{U}\|_F^2] . \end{aligned} \quad (25)$$

Since these expectations are with respect to the posterior $p(\mathbf{Y}, \mathbf{Z}, \mathbf{U}, \mathbf{V}; \lambda^{(n)}, \sigma^{2(n)})$, we can use Algorithm 2 to compute them numerically. Replacing

$$\mathbb{E} [\|\mathbf{Z} - \mathbf{U}\mathbf{V}^T\|_F^2] \simeq \frac{1}{M} \sum_{k=1}^M \|\mathbf{Z}^{(k)} - \mathbf{U}^{(k)} \mathbf{V}^{T(k)}\|_F^2 , \quad (26)$$

$$\mathbb{E} [\|\mathbf{U}\|_F^2] \simeq \frac{1}{M} \sum_{k=1}^M \|\mathbf{U}^{(k)}\|_F^2 , \quad (27)$$

we obtain that the M-step consists of the following two updates:

$$\sigma^{2(n+1)} = \frac{1}{Mnm} \sum_{k=1}^M \|\mathbf{Z}^{(k)} - \mathbf{U}^{(k)} \mathbf{V}^{T(k)}\|_F^2 , \quad (28)$$

$$\lambda^{(n+1)} = \frac{1}{Mnr} \sum_{k=1}^M \|\mathbf{U}^{(k)}\|_F^2 . \quad (29)$$

The whole procedure for the empirical Bayes approach to quantized probabilistic matrix factorization is reported in Algorithm 3. It is seen that, at each iteration of the EM method, the procedure calls 2 until the EM method has converged or has reached a fixed amount of iterations n_{\max} . Convergence can be established for instance when the relative change in the parameter estimates is below a given threshold. As a final step, the procedure calls Algorithm 2 to get the final estimates of $(\mathbf{Z}, \mathbf{U}, \mathbf{V})$. The number of samples drawn by the Gibbs sampler during the EM routine and during the final step may differ. In particular, it is recommended to use a higher number of samples at the final step [2].

Algorithm 3 Empirical-Bayes Quantized Probabilistic Matrix Factorization (E-Q-PMF)

```

1: procedure E-Q-PMF
2:   Input:  $\mathbf{Y}$ 
3:   Output:  $\hat{\mathbf{Z}}, \hat{\mathbf{U}}, \hat{\mathbf{V}}$ 
4:   while EM has not converged or  $n < n_{\max}$  do
5:     call Algorithm 2
6:      $\sigma^{2(n+1)} \leftarrow \frac{1}{M_1 n m} \sum_{k=1}^{M_1} \|\mathbf{Z}^{(k)} - \mathbf{U}^{(k)} \mathbf{V}^{T(k)}\|_F^2$ 
7:      $\lambda^{(n+1)} \leftarrow \frac{1}{M_1 n r} \sum_{k=1}^{M_1} \|\mathbf{U}^{(k)}\|_F^2$ 
8:      $n \leftarrow n + 1$ 
9:   call Algorithm 2
10:   $\hat{\mathbf{Z}} \leftarrow \frac{1}{M_2} \sum_{k=1}^{M_2} \mathbf{Z}^k$ 
11:   $\hat{\mathbf{U}} \leftarrow \frac{1}{M_2} \sum_{k=1}^{M_2} \mathbf{U}^k$ 
12:   $\hat{\mathbf{V}} \leftarrow \frac{1}{M_2} \sum_{k=1}^{M_2} \mathbf{V}^k$ 

```

VI. EXTENSION TO MISSING VALUES

In this section we show how to extend the methods for probabilistic matrix factorization from quantized entries that we presented in Section IV and V to the case where not all the entries of \mathbf{Y} are available. For sake of brevity, we focus only on the full Bayes approach, the extension to the empirical Bayes approach can be done using the same arguments.

We recall that now we have available the matrix \mathbf{W} defined in (5) telling us which entries of \mathbf{Y} are available. Having an entry y_{ij} unavailable means that, when we have to draw a sample from the conditional of the corresponding entry z_{ij} , we can not determine the interval to which the drawn value must lay. In other words, the conditional distribution of an entry z_{ij} such that $w_{ij} = 0$ will be a standard Gaussian with no truncations. We thus have:

- $z_{ij} | \{z_{hk}\}_{(h,k) \neq (i,j)}, \mathbf{U}, \mathbf{V}, \lambda, \sigma^2, \mathbf{Y}, w_{ij} = 1$. In this case we are in the same conditions as of Section IV. Therefore

$$z_{ij} | \{z_{hk}\}_{(h,k) \neq (i,j)}, \mathbf{U}, \mathbf{V}, \lambda, \sigma^2, \mathbf{Y}, w_{ij}=1 \sim \mathcal{N}_{q_{k-1}}^{q_k}(\mathbf{u}_i \mathbf{v}_j^T, \sigma^2).$$

- $z_{ij} | \{z_{hk}\}_{(h,k) \neq (i,j)}, \mathbf{U}, \mathbf{V}, \lambda, \sigma^2, \mathbf{Y}, w_{ij} = 0$. In this case we cannot bound z_{ij} and thus

$$z_{ij} | \{z_{hk}\}_{(h,k) \neq (i,j)}, \mathbf{U}, \mathbf{V}, \lambda, \sigma^2, \mathbf{Y}, w_{ij}=0 \sim \mathcal{N}(\mathbf{u}_i \mathbf{v}_j^T, \sigma^2).$$

The other conditional distributions remain unchanged with respect to those in Section IV. The whole procedure is reported in Algorithm 4.

VII. NUMERICAL EXPERIMENTS

In this section we report the results of experiments made both on synthetic and real data.

A. Simulations

First, we assess the performance of the proposed methods on several data sets of synthetic data. To this end, we perform several numerical experiments, each featuring different experimental conditions. Each numerical experiment consists of independent 100 Monte Carlo simulations where random data are generated according to model (1). In particular, the

Algorithm 4 Full-Bayes Quantized Probabilistic Matrix Factorization with missing entries (F-Q-PMF2)

```

1: procedure F-Q-PMF2
2:   Input:  $\mathbf{W} \odot \mathbf{Y}$ ,
3:   Output:  $\hat{\mathbf{Z}}, \hat{\mathbf{U}}, \hat{\mathbf{V}}$ 
4:   for  $k = 1$  to  $M + M_0$  do
5:     for  $i = 1$  to  $n$  do
6:       for  $j = 1$  to  $m$  do
7:         if  $w_{ij} = 1$  then
8:           draw  $z_{ij}^{(k)} \sim \mathcal{N}_{q_{k-1}}^{q_k}(\mathbf{u}_i^T \mathbf{v}_j, \sigma^2)$ 
9:         else
10:          draw  $z_{ij}^{(k)} \sim \mathcal{N}(\mathbf{u}_i^T \mathbf{v}_j, \sigma^2)$ 
11:        draw  $\mathbf{U}^{(k)} \sim \mathcal{N}(\mathbf{m}_u, \mathbf{P}_u)$ 
12:        draw  $\mathbf{V}^{(k)} \sim \mathcal{N}(\mathbf{m}_v, \mathbf{P}_v)$ 
13:        draw  $\lambda^{-1(k)} \sim \text{Gamma}\left(\frac{nr}{2}, \frac{\|\mathbf{U}\|_F^2}{2}\right)$ 
14:        draw  $\sigma^{-2(k)} \sim \text{Gamma}\left(\frac{nm}{2}, \frac{\|\mathbf{Z} - \mathbf{U}\mathbf{V}^T\|_F^2}{2}\right)$ 
15:       $\hat{\mathbf{Z}} \leftarrow \frac{1}{M} \sum_{k=1}^M \mathbf{Z}^k$ 
16:       $\hat{\mathbf{U}} \leftarrow \frac{1}{M} \sum_{k=1}^M \mathbf{U}^k$ 
17:       $\hat{\mathbf{V}} \leftarrow \frac{1}{M} \sum_{k=1}^M \mathbf{V}^k$ 

```

generating models for \mathbf{E} , \mathbf{U} , and \mathbf{V} are as in (2), (6) (with $\lambda = 1$) and (7), respectively. The size of the data matrix \mathbf{Y} is set to 50×50 , and it is obtained by rounding \mathbf{Z} to the next integer (i.e., the map \mathcal{Q} it is a ceil-type quantizer). Depending on the experiment, the matrices \mathbf{U} and \mathbf{V} have rank $r = 3, 5, 10, 20$; similarly, we run experiments with different noise variance $\sigma^2 = 0.1, 1, 10$. Combining these experimental conditions we obtain 12 numerical experiments in total. In our experiments, the following methods are tested:

F-Q-PMF: the full Bayes method for quantized probabilistic matrix factorization, described in Section IV. We set $M = 500$, $M_0 = 50$.

E-Q-PMF: the empirical Bayes method for quantized probabilistic matrix factorization, described in Section IV. We set $M_1 = 50$, $M_2 = 200$, $M_0 = 20$. The parameters λ and σ^2 are initialized to 100 and 1, respectively.

Baseline: This method does not perform any matrix factorization, but it estimates the entries z_{ij} by computing $z_{ij} = y_{ij} - 0.5$. Therefore, it serves as a baseline to check whether the probabilistic methods can reconstruct the matrix \mathbf{Z} better.

F-Oracle: This method performs full Bayes probabilistic matrix factorization. It corresponds to the approach of [21]. Since it uses \mathbf{Z} as data matrix, it is referred to an oracle. We use it to check how close our reconstruction of \mathbf{U} and \mathbf{V} is compared to a method that has more information available.

F-Naive: This method uses the same approach as F-Oracle, but it uses \mathbf{Y} as data matrix, disregarding the fact that this matrix is a quantized version of \mathbf{Z} .

E-Oracle: This method performs empirical Bayes probabilistic matrix factorization, using \mathbf{Z} as data matrix (see also [18] for details).

E-Naive: Same as E-Oracle, but with \mathbf{Y} as data matrix.

To evaluate the performance of the various methods, we

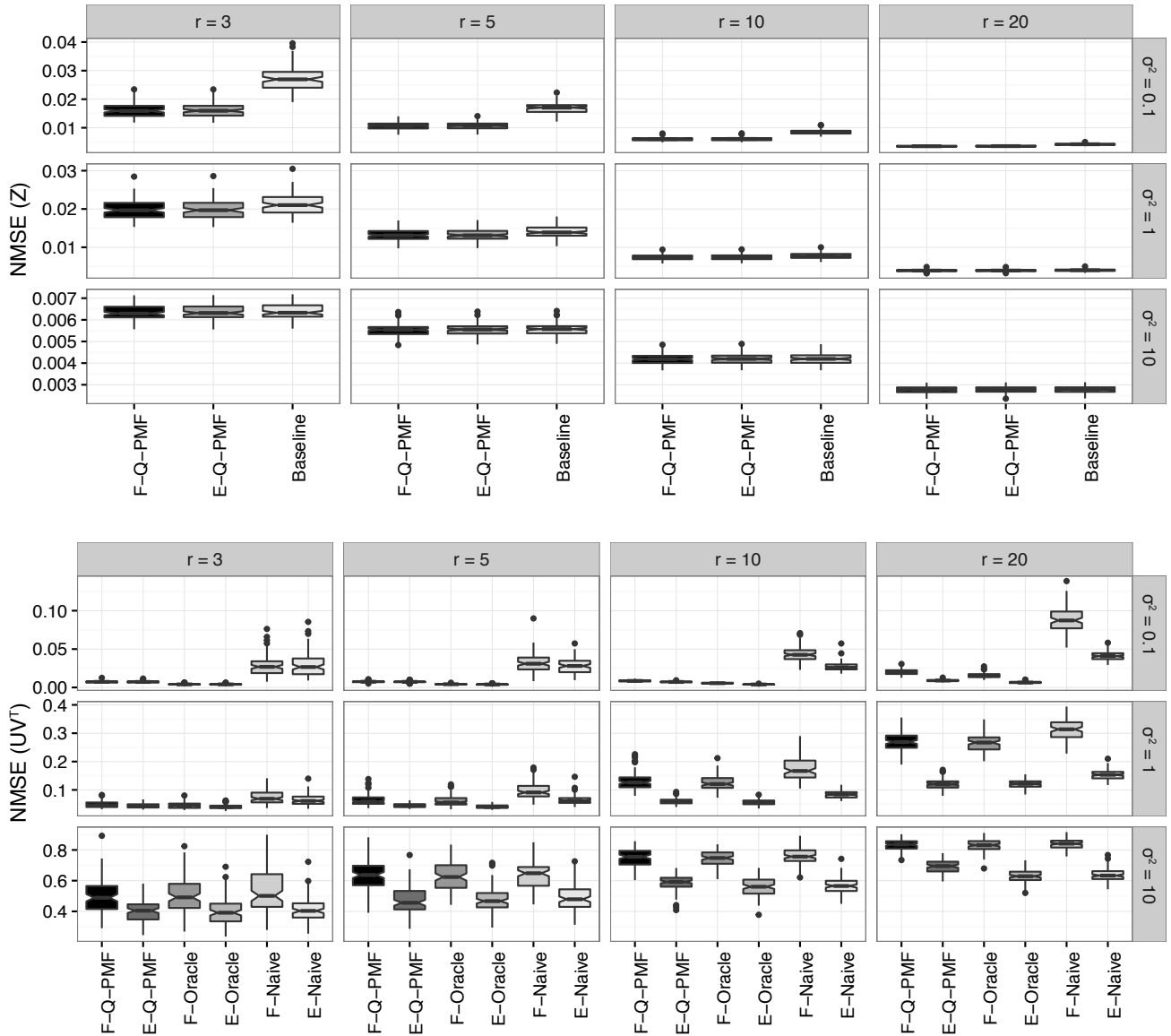


Figure 2. Top panel: Box plots of the performance of the tested methods in terms of the index $NMSE(\mathbf{Z})$. Bottom panel: Box plots of the performance of the tested methods in terms of the index $NMSE(\mathbf{UV}^T)$.

introduce the following indices:

- The normalized mean-square error (NMSE) on \mathbf{Z}

$$NMSE(\mathbf{Z}) = \frac{\|\hat{\mathbf{Z}} - \mathbf{Z}\|_F^2}{\|\mathbf{Z}\|_F^2};$$

- The NMSE on the product \mathbf{UV}^T

$$NMSE(\mathbf{UV}^T) = \frac{\|\hat{\mathbf{U}}\hat{\mathbf{V}}^T - \mathbf{UV}^T\|_F^2}{\|\mathbf{UV}^T\|_F^2}.$$

Note that $NMSE(\mathbf{Z})$ can be computed only for the algorithms F-Q-PMF, E-Q-PMF, and Baseline, while $NMSE(\mathbf{UV}^T)$ can be computed for F-Q-PMF, E-Q-PMF, and the Oracle-type and naive-type methods. The results of the

simulations are reported in Figure 2, which depicts the box plots of the distributions of $NMSE(\mathbf{Z})$ and $NMSE(\mathbf{UV}^T)$ obtained by the tested algorithms. From the top panel, it can be seen that the full Bayes approach and the empirical Bayes approach give substantially the same performance in reconstructing the hidden matrix \mathbf{Z} . Moreover, both are uniformly better than the baseline method. It should be also noted that the performance seems to increase as the rank increases: this is because the effect of the noise matrix \mathbf{E} becomes more and more negligible compared to the product \mathbf{UV}^T . As for the ability in reconstructing the product \mathbf{UV}^T , i.e., the index $NMSE(\mathbf{UV}^T)$, the results are more surprising. All the methods based on the empirical Bayes seem to perform

better than those based on full Bayes approach. Nevertheless, it can be seen that the proposed methods perform better than the naive methods that do not account for the presence of the quantization map, giving results that get close to the Oracle algorithms (i.e., those algorithms having access to the matrix \mathbf{Z}).

B. Experiment on a real data set

We test the ability of the method proposed in Section VI to predict the missing values of a quantized matrix. To this end, we perform an experiment on the *MovieLens 100k* data set¹. This data set contains the rates given by 943 users to 1682 movies. The rates can assume integer values between 1 and 5. Since each user rated only a few movies, only 10^5 rates are available, so the resulting matrix \mathbf{Y} of size 943×1682 is mostly empty (approximately 6.3% of the entries are available).

To assess the performance of the full Bayes method (precisely, Algorithm 4), we further randomly remove 20% of the available entries to construct a test set. Then, we train the method on the remaining entries, evaluating its performance in predicting the test set. The procedure is repeated 20 times, each time randomly selecting a different test set. The method is compared with OptSpace [8], [7]. The performance is evaluated through the following indices (see also e.g. [1]):

- The root mean-square error (RMSE)

$$\text{RMSE} = \frac{1}{\sqrt{|\Omega|}} \sqrt{\sum_{(i,j) \in \Omega} (\hat{y}_{i,j} - y_{i,j})^2}, \quad (30)$$

where Ω denotes the test set and $|\Omega|$ its cardinality;

- The normalized mean absolute error (NMAE)

$$\text{NMAE} = \frac{1}{|\Omega|(\max(\mathbf{Y}) - \min(\mathbf{Y}))} \sum_{(i,j) \in \Omega} |\hat{y}_{i,j} - y_{i,j}|. \quad (31)$$

Regarding the proposed method, we test different choices of the hidden rank, i.e., $r = 3, 5, 7$. The Gibbs sampler is initialized with $M = 150$ and $M_0 = 30$. Moreover, we assume that a user, when giving a rate, tends to round their evaluation to the next integer; so, the used quantization map is a ceil. The method OptSpace is used with its default settings. Table I reports the mean and standard deviation of the above indices, obtained by the two algorithms. It can be noted that, while OptSpace achieves a better RMSE, the proposed method gives a lower NMAE. The reason for this lays in the fact that the proposed method provides an integer as an output, while OptSpace gives real-valued predictions. Therefore, in general OptSpace commits more errors, but with small magnitude, while, every time the proposed method makes a wrong prediction, the magnitude of the prediction error is larger (greater than or equal to 1). The results of Table VII-B are then explained by the fact that the quadratic index RMSE penalizes large errors more than NMAE. The choice $r = 3$ seems to be the most effective one; note that this correspond to the rank automatically estimated by OptSpace.

¹<http://grouplens.org/datasets/movielens/>

Method	RMSE	NMAE
OptSpace	0.9579 ± 0.0038	0.1875 ± 0.0006
F-Q-PMF ($r = 3$)	0.9872 ± 0.0061	0.1763 ± 0.0015
F-Q-PMF ($r = 5$)	0.9900 ± 0.0060	0.1763 ± 0.0014
F-Q-PMF ($r = 7$)	0.9946 ± 0.0044	0.1770 ± 0.0013

Table I

RESULTS OF THE METHODS TESTED ON THE DATA SET MOVIELENS 100K.

VIII. CONCLUSION

We have introduced a new probabilistic matrix factorization method for matrices having quantized measurements. To estimate the matrix factors, we have proposed two methods. The first is based on a full Bayes approach; the latter is based on an empirical Bayes approach. Both approaches rely on particular instances of the Gibbs sampler. In the empirical Bayes approach, the Gibbs sampler is combined with the expectation maximization algorithm to return also the hyperparameter estimates. Numerical simulations and real data experiments show the effectiveness of the devised methods.

In future work, we plan to integrate an automatic rank detection for the underlying matrix factors. Also, methods based on variational inference are under study.

APPENDIX

A. Drawing samples from (19) and (21)

The algorithms presented in this paper rely on sampling from (19) and (21). Let us focus on (19) –the case (21) is virtually the same. Drawing from such a distribution requires the computation of the covariance matrix $\mathbf{P}_{\mathbf{u}}$, which in turn requires the inversion of a $nr \times nr$ matrix, see (20). In general, this operation becomes cumbersome for high size n ; however, we show that in this case we can exploit the structure of the model to simplify dramatically this inversion. Recalling the identity for Kronecker products $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$, we have that

$$\begin{aligned} \mathbf{P}_{\mathbf{u}} &= \left(\frac{1}{\sigma^2} (\mathbf{V}^T \otimes \mathbf{I}_n) (\mathbf{V} \otimes \mathbf{I}_n) + \lambda^{-1} \mathbf{I}_{nr} \right)^{-1} \\ &= \left(\frac{1}{\sigma^2} (\mathbf{V}^T \mathbf{V} \otimes \mathbf{I}_n) + \lambda^{-1} \mathbf{I}_r \otimes \mathbf{I}_n \right)^{-1} \\ &= \left(\left(\frac{1}{\sigma^2} \mathbf{V}^T \mathbf{V} + \lambda^{-1} \mathbf{I}_r \right) \otimes \mathbf{I}_n \right)^{-1} \\ &= \left(\frac{1}{\sigma^2} \mathbf{V}^T \mathbf{V} + \lambda^{-1} \mathbf{I}_r \right)^{-1} \otimes \mathbf{I}_n. \end{aligned}$$

Thus, obtaining $\mathbf{P}_{\mathbf{u}}$ requires the inversion of a $r \times r$, with great reduction of the computational burden. The identity is also useful for the task of drawing samples because, defining \mathbf{L} as the Cholesky factor of $\left(\frac{1}{\sigma^2} \mathbf{V}^T \mathbf{V} + \lambda^{-1} \mathbf{I}_r \right)^{-1}$, the Cholesky factor of $\mathbf{P}_{\mathbf{u}}$ is equal to $\mathbf{L} \otimes \mathbf{I}_n$.

B. Computation of the function $Q^{(n)}(\lambda, \sigma^2)$

Using the various dependencies of the variables in our model, we first decompose the log-complete likelihood as follows

$$\log p(\mathbf{Y}, \mathbf{Z}, \mathbf{U}, \mathbf{V}; \lambda, \sigma^2) = \log p(\mathbf{Y}|\mathbf{Z}) \quad (32)$$

$$+ \log p(\mathbf{Z}|\mathbf{U}, \mathbf{V}; \sigma^2) + \log p(\mathbf{U}; \lambda) + \log p(\mathbf{V}).$$

Furthermore,

$$\begin{aligned} -2 \log p(\mathbf{Z}|\mathbf{U}, \mathbf{V}; \sigma^2) &= nm \log \sigma^2 + \frac{1}{\sigma^2} \|\mathbf{Z} - \mathbf{U}\mathbf{V}^T\|_F^2, \\ -2 \log p(\mathbf{U}; \lambda) &= nr \log \lambda + \frac{1}{\lambda} \|\mathbf{U}\|_F^2, \\ -2 \log p(\mathbf{V}) &= mr + \|\mathbf{V}\|_F^2. \end{aligned}$$

Taking the expectation of (32), we observe the following facts:

1) The first term admits the factorization

$$\log p(\mathbf{Y}|\mathbf{Z}) = \sum_{i=1}^n \sum_{j=1}^m \log p(y_{ij}|z_{ij}), \quad (33)$$

where each factor is of the type

$$p(y_{ij}|z_{ij}) = \begin{cases} 1 & \text{if } y_{ij} = s_k \text{ and } z_{ij} \in (q_{k-1}, q_k] \\ 0 & \text{otherwise} \end{cases}. \quad (34)$$

When computing the expectation of this term using the Gibbs sampler introduced throughout the paper, it is ensured that all the generated samples $z_{ij}^{(k)}$ belong to the interval corresponding to the observed quantized value y_{ij} . Hence, when we compute the expectation of $\log p(\mathbf{Y}|\mathbf{Z})$, it is ensured that each term (33) is always equal to 1 and thus $\log p(\mathbf{Y}|\mathbf{Z})$.

2) The term $\log p(\mathbf{V})$ does not depend on λ or σ^2 , therefore it does not affect the maximum of $Q^{(n)}(\lambda, \sigma^2)$.

Given these observations, it is straightforward to obtain (25).

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC AdG A-DATADRIVE-B (290923). This paper reflects only the authors' views, the Union is not liable for any use that may be made of the contained information; Research Council KUL: GOA/10/09 MaNet, CoE PFV/10/002 (OPTec), BIL12/11T; PhD/Postdoc grants; Flemish Government: FWO: PhD/Postdoc grants, projects: G.0377.12 (Structured systems), G.088114N (Tensor based data similarity); IWT: PhD/Postdoc grants, projects: SBO POM (100031); iMinds Medical Information Technologies SBO 2014; Belgian Federal Science Policy Office: IUAP P7/19 (DYSCO, Dynamical systems, control and optimization, 2012-2017).

REFERENCES

- [1] S.A. Bhaskar. Probabilistic low-rank matrix completion from quantized measurements. *Journal of Machine Learning Research*, 17(60):1–34, 2016.
- [2] G. Casella. Empirical bayes gibbs sampling. *Biostatistics*, 2(4):485–500, 2001.
- [3] M.A. Davenport, Y. Plan, E. van den Berg, and M. Wootters. 1-bit matrix completion. *Information and Inference*, 3(3):189–223, 2014.
- [4] P. Gopalan, F.J. Ruiz, R. Ranganath, and D.M. Blei. Bayesian nonparametric Poisson factorization for recommendation systems. In *AISTATS*, pages 275–283, 2014.
- [5] D. Guillaumet and J. Vitria. Non-negative matrix factorization for face recognition. In *Topics in artificial intelligence*, pages 336–344. Springer, 2002.
- [6] P. Jain, P. Netrapalli, and S. Sanghavi. Low-rank matrix completion using alternating minimization. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 665–674. ACM, 2013.

- [7] R.H. Keshavan, A. Montanari, and S. Oh. Matrix completion from noisy entries. *Journal of Machine Learning Research*, 11(Jul):2057–2078, 2010.
- [8] R.H. Keshavan, S. Oh, and A. Montanari. Matrix completion from a few entries. In *2009 IEEE International Symposium on Information Theory*, pages 324–328. IEEE, 2009.
- [9] S.A. Khan, D. Malani, A. Murumägi, O. Kallioniemi, T. Aittokallio, and S. Kaski. Drug response prediction by inferring pathway-response associations with kernelized Bayesian matrix factorization. *Bioinformatics*, 32(17):i455–i463, 2016.
- [10] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.
- [11] J. Lafond, O. Klopp, E. Moulines, and J. Salmon. Probabilistic low-rank matrix completion on finite alphabets. In *Advances in Neural Information Processing Systems*, pages 1727–1735, 2014.
- [12] A.S. Lan, C. Studer, and R.G. Baraniuk. Matrix recovery from quantized and corrupted measurements. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4973–4977. IEEE, 2014.
- [13] Z. Li, J. Liu, X. Zhu, T. Liu, and H. Lu. Image annotation using multi-correlation probabilistic matrix factorization. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 1187–1190. ACM, 2010.
- [14] H. Ma, H. Yang, M.R. Lyu, and I. King. Sorec: social recommendation using probabilistic matrix factorization. In *Proceedings of the 17th ACM conference on Information and knowledge management*, pages 931–940. ACM, 2008.
- [15] P. Magni, R. Bellazzi, and G. De Nicolao. Bayesian function learning using MCMC methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1319–1331, 1998.
- [16] G. McLachlan and T. Krishnan. *The EM algorithm and extensions*, volume 382. John Wiley & Sons, 2007.
- [17] D. Meng and F. De La Torre. Robust matrix factorization with unknown noise. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1337–1344, 2013.
- [18] S. Nakajima and M. Sugiyama. Theoretical analysis of Bayesian matrix factorization. *Journal of Machine Learning Research*, 12(Sep):2583–2648, 2011.
- [19] R. Remi and K. Schnass. Dictionary identification-Sparse matrix-factorization via-minimization. *IEEE Transactions on Information Theory*, 56(7):3523–3539, 2010.
- [20] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*, pages 452–461. AUAI Press, 2009.
- [21] R. Salakhutdinov and A. Mnih. Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. In *Proceedings of the 25th international conference on Machine learning*, pages 880–887. ACM, 2008.
- [22] H. Shan and A. Banerjee. Generalized probabilistic matrix factorizations for collaborative filtering. In *2010 IEEE International Conference on Data Mining*, pages 1025–1030. IEEE, 2010.
- [23] A.F.M. Smith and G.O. Roberts. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 3–23, 1993.
- [24] N. Wang, T. Yao, J. Wang, and D.-Y. Yeung. A probabilistic approach to robust matrix factorization. In *European Conference on Computer Vision*, pages 126–139. Springer, 2012.
- [25] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In *Advances in neural information processing systems*, pages 2080–2088, 2009.
- [26] C.J. Wu. On the convergence properties of the EM algorithm. *The Annals of statistics*, pages 95–103, 1983.
- [27] D. Zachariah, M. Sundin, M. Jansson, and S. Chatterjee. Alternating least-squares for low-rank matrix reconstruction. *IEEE Signal Processing Letters*, 19(4):231–234, 2012.