Kejun Huang and Nicholas D. Sidiropoulos

Putting Nonnegative Matrix Factorization to the Test

A tutorial derivation of pertinent Cramér–Rao bounds and performance benchmarking

Nonnegative matrix factorization (NMF) is a useful tool in a broad range of applications, from signal separation to computer vision and machine learning. NMF is a hard (NP-hard) computational problem for which various approximate solutions have been developed over the years. Given the widespread interest in NMF and its applications, it is perhaps surprising that the pertinent Cramér–Rao lower bound (CRLB) on the accuracy of the nonnegative latent factor estimates has not been worked out in the literature. In hindsight, one reason may be that the required computations are more subtle than usual: the problem involves constraints and ambiguities that must be dealt with, and the Fisher information matrix is always singular. We provide a concise tutorial derivation of the CRLB for both symmetric NMF and asymmetric NMF, using the latest CRLB tools, which should be of broad interest for analogous derivations in related factor analysis problems. We illustrate the behavior of these bounds with respect to model parameters and put some of the best NMF algorithms to the test against one another and the CRLB. The results help illuminate what can be expected from the current state of art in NMF algorithms, and they are reassuring in that the gap to optimality is small in relatively sparse and low rank scenarios.

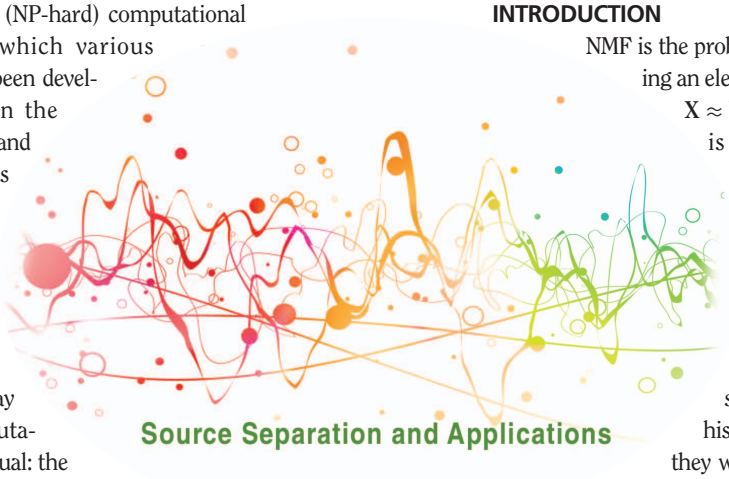


IMAGE LICENSED BY INGRAM PUBLISHING

Digital Object Identifier 10.1109/MSP.2013.2296172
Date of publication: 7 April 2014

INTRODUCTION

NMF is the problem of (approximately) factoring an element-wise nonnegative matrix $\mathbf{X} \approx \mathbf{W}\mathbf{H}^T$, where $\mathbf{W}$ is $I \times K$, $\mathbf{H}$ is $J \times K$, $K < \min(I, J)$, and $\mathbf{W} \geq 0$, $\mathbf{H} \geq 0$ element-wise [1], [2]. Symmetric NMF is the problem of factoring a square matrix $\mathbf{X} \approx \mathbf{W}\mathbf{W}^T$, where the $I \times K$ matrix $\mathbf{W} \geq 0$ element-wise. Both general (asymmetric) and symmetric NMF have a long history and various applications; they were more recently introduced to the signal processing community, primarily as means to restore identifiability in bilinear matrix factorization/blind source separation (BSS).

The CRLB [3, Ch. 3] is the most widely used estimation benchmark in signal processing. In many cases it is relatively easy to compute, and it is asymptotically achievable by maximum likelihood (ML) estimators in high signal-to-noise ratio (SNR) scenarios [3, pp. 164]. In other cases, there may be technical difficulties in deriving (or complexity issues in computing) the pertinent CRLB; but due to the central role of this bound in signal processing research, work on developing CRLB tools continues [4]–[7], thereby enlarging the set of problems for which the CRLB can be used in practice.

1053-5888/14/$31.00©2014IEEE

Interestingly, despite the popularity of NMF, the pertinent CRLB on the latent factors has not been studied, to the best of our knowledge. This is surprising, especially because ML NMF is NP-hard, and it is natural to wonder how far from the best achievable estimation performance existing (suboptimal) NMF algorithms operate, under different scenarios. The missing link can perhaps be explained by the fact that most NMF researchers come from different communities, and, even for someone versed in statistical signal processing, the CRLB computations for NMF are subtle, requiring modern tools, as we will see. The aim of this article is threefold: first, to fill this gap; second, to put some of the leading NMF algorithms to the test using the CRLB as a benchmark; and third, to do so in an easily accessible way that can serve as a starting point for analogous derivations in related constrained matrix and tensor factorization problems.

FUNDAMENTALS

IDENTIFIABILITY

Rank-constrained matrix factorization is highly unidentifiable without additional constraints. For any given factorization $\mathbf{X} = \mathbf{W}\mathbf{H}^T$ and any invertible $\mathbf{Q}$, $\mathbf{X} = \hat{\mathbf{W}}\hat{\mathbf{H}}^T$ with $\hat{\mathbf{W}} = \mathbf{W}\mathbf{Q}^T$ and $\hat{\mathbf{H}} = \mathbf{H}\mathbf{Q}^{-1}$. For symmetric factorization $\mathbf{X} = \mathbf{W}\mathbf{W}^T$, we need only further require $\mathbf{Q}$ to be unitary. To force the factorization to be unique, one must put additional constraints on the latent factors (the columns of $\mathbf{W}$ and $\mathbf{H}$), e.g., orthogonality in the case of singular value decomposition (SVD). With $\mathbf{W} := [\mathbf{w}_1, \cdots, \mathbf{w}_K]$, $\mathbf{H} := [\mathbf{h}_1, \cdots, \mathbf{h}_K]$, $\mathbf{W}\mathbf{H}^T = \mathbf{w}_1\mathbf{h}_1^T + \cdots + \mathbf{w}_K\mathbf{h}_K^T$; hence we may permute the rank-one outer products $\{\mathbf{w}_k\mathbf{h}_k^T\}_{k=1}^K$, and/or scale $\mathbf{w}_k$ by $s > 0$ and counterscale $\mathbf{h}_k$ by $1/s$ without changing $\mathbf{W}\mathbf{H}^T$. These ambiguities are inherent to NMF, requiring additional conventions (as opposed to conditions) to resolve, similar to ordering the singular values in the SVD. These inherent ambiguities are often inconsequential in applications, and we will say that a model is essentially identifiable or essentially unique when it can be identified up to these inherently unresolvable ambiguities. Still, these ambiguities are reflected in, and in fact dominate the CRLB, unless they are properly accounted for. In this article, for asymmetric NMF, we assume the columns of $\mathbf{W}$ are scaled to sum up to one, i.e.,

$$\sum_{i=1}^{I} w_{i1} = \sum_{i=1}^{I} w_{i2} = \cdots = \sum_{i=1}^{I} w_{iK} = 1 \tag{1}$$

to overcome the scaling ambiguity. Once we get estimates of $\mathbf{W}$ and $\mathbf{H}$, denoted $\hat{\mathbf{W}}$ and $\hat{\mathbf{H}}$, respectively, using any NMF algorithm, we scale the columns of $\hat{\mathbf{W}}$ to satisfy (1), and counterscale the corresponding columns of $\hat{\mathbf{H}}$. Then least-squares matching of the columns of $\hat{\mathbf{W}}$ to those of $\mathbf{W}$ is equivalent to the so-called linear assignment problem [8], whose solution can be found by the Hungarian algorithm [9], [10]. The MATLAB code is available at http://www.mathworks.com/matlabcentral/fileexchange/11609-hungarian-algorithm. In the symmetric case, there is no scaling ambiguity, so we directly use the Hungarian algorithm to find the best column permutation.

Conditions for (essential) uniqueness of NMF (ensuring that $\mathbf{Q}$ can only be a positively scaled permutation matrix in the asymmetric case, or simply a permutation matrix in the symmetric case) have previously been studied in [11]–[13], and are summarized in [14]. In a nutshell, NMF is not always unique, and pertinent conditions ensuring uniqueness are complicated (e.g., a sufficient condition for uniqueness requires the conic hull of the row vectors of $\mathbf{W}$ to be a superset of a specific second-order cone [14]). The following corollary is a useful rule of thumb: if the sufficient condition given in [14, Th. 4] is satisfied for the symmetric NMF $\mathbf{X} = \mathbf{W}\mathbf{W}^T$, then

- the supports (sets of indices of nonzero entries) of any two columns of $\mathbf{W}$ are not contained in one another.
- each column of $\mathbf{W}$ contains at least $K - 1$ zeros.

The same holds for both $\mathbf{W}$ and $\mathbf{H}$ in the asymmetric case $\mathbf{X} = \mathbf{W}\mathbf{H}^T$. These two properties together are neither sufficient nor necessary for uniqueness; in practice however, as shown empirically in [14, Examples 3 and 4], it is very likely that NMF will give an essentially unique solution if these two conditions are both satisfied. Notice that if we set the zero entries of $\mathbf{W}$ (and $\mathbf{H}$ in the asymmetric case) randomly, with density (number of nonzero entries over the number of entries) less than $(I - K)/I$, then for large I these conditions will be met with high probability.

ALGORITHMS

Owing to the NP-hardness of asymmetric NMF [15], numerous approximation algorithms have been developed (cf. [16] and references therein). On the contrary, there are relatively few algorithms available for symmetric NMF (cf. [17] and references therein and [14]). If a symmetric matrix admits an exact symmetric NMF (not necessarily low rank), it is called *completely positive* (*CP*) [18]. It was recently proven that checking whether a matrix is CP is also NP-hard [19].

He et al. [17] summarized existing algorithms for symmetric NMF, which turned out being very similar (all based on so-called multiplicative updates). They concluded that those algorithms all belong to two basic kinds of algorithms: α-symmetric NMF and β-symmetric NMF, where α and β are tuning parameters that moderate performance (e.g., the algorithm in [20] belongs to α-symmetric NMF with $\alpha = 1/4$, and the algorithm in [21] belongs to β-symmetric NMF with $\beta = 1/2$). A very different algorithm based on Procrustes rotation was proposed in [14].

The algorithms for asymmetric NMF can be broadly classified as optimization-based and geometry-based. The cost function in optimization-based methods usually measures the quality of factorization, e.g., in terms of Euclidian distance, K-L divergence, etc., and may include regularization terms that capture presumed properties of the sought latent factors, e.g., sparsity, smoothness, etc. None of these formulations is jointly convex in $\mathbf{W}$ and $\mathbf{H}$ ($\mathbf{W}\mathbf{H}^T$ is a bilinear form); but in most cases they are conditionally convex over one factor given the other. Most optimization-based methods therefore adopt an alternating optimization approach—a few algorithms employ all-at-once (joint) parameter updates using gradient or Newton steps, but these require careful parameter tuning to ensure convergence to a local optimum. In the context of alternating optimization algorithms, for the update of one factor, one can take a

gradient direction but with a very conservative step-size such that positivity is always satisfied; this can be reduced to a multiplicative update [22], [23]. Alternatively, a more aggressive step-size can be used, but then a projection back to the nonnegative orthant is required [24]. A less popular way is to take the second-order derivative into account [25].

The most commonly used cost function is Euclidean distance. One reason for this is that when one factor is fixed, and if we ignore the nonnegativity constraint, the problem reduces to linear least squares, in which case we know the solution in closed form. Therefore, a straightforward way is to simply replace the negative entries of the least squares result with zeros in each update [26]—which is, however, suboptimum, and not guaranteed to converge. Taking the nonnegativity constraints back into consideration, the conditional update problem is nonnegative least squares, which is convex but the solution is not in closed form. Existing methods use quadratic programming [27], active set [28], [29], and coordinate descent [30].

Geometry-based methods stem from the geometric interpretation of NMF by Donoho [11]. The basic idea is to find a simplicial cone, with a certain number of extreme rays, that is contained in the nonnegative orthant and contains all the data points. The effectiveness of geometry-based methods is application dependent; in cases where the so-called separability assumption [11] is reasonable, the extreme rays of the simplicial cone can be found by selecting from the data vectors per se [31], [32]. In other cases, nonnegativity is not strictly required for one factor, and the aim is to find the minimum volume simplicial cone that contains the data points [33], [34]. A polytope approximation method [35] seems to be more general compared to the others in this genre.

MODERN CRLB TOOLS

Suppose a set of measurements $\mathbf{X}$ is drawn from a probability density function $p(\mathbf{X};\theta)$ parameterized by θ, and our goal is to estimate θ given the realizations of $\mathbf{X}$. If the regularity condition $\mathrm{E}_{\mathbf{X}}\{\nabla_\theta \ln p(\mathbf{X};\theta)\} = 0$ is satisfied, then we can define the Fisher information matrix (FIM) as $\mathbf{F}_\theta \triangleq \mathrm{E}_{\mathbf{X}}\left\{[\nabla_\theta \ln p(\mathbf{X};\theta)][\nabla_\theta \ln p(\mathbf{X};\theta)]^T\right\}$, and the CRLB on the covariance matrix of any unbiased estimator of θ on the basis of $\mathbf{X}$ is the inverse of the FIM [3, Ch. 3], i.e., the difference between the estimator covariance matrix and the inverse of the FIM is positive semidefinite. From this, it follows that $\mathrm{E}_{\mathbf{X}}\{\|\theta - \hat{\theta}\|_2^2\} \geq \mathrm{tr}\{\mathbf{F}_\theta^{-1}\}$, where $\hat{\theta}$ is any unbiased estimator of θ on the basis of $\mathbf{X}$. More detailed discussion of the CRLB, including conditions under which there exists an estimator that can attain the bound, can be found in classic textbooks on estimation theory, e.g., [3, Ch. 3].

When the FIM is singular, Stoica and Marzetta [6] have shown that we can use the Moore–Penrose pseudoinverse instead (in hindsight, this can be deduced from the Schur complement generalized to singular matrices [36, p. 651]). The pseudoinverse is still a lower bound, albeit it is generally looser, and more difficult to attain. Important references on the CRLB for problems with constraints on the unknown parameters, represented by equalities and inequalities, include [4], [5], and [7]. Their results show that inequality constraints do not affect the CRLB, whereas equality constraints do. (Strictly speaking, inequalities do not affect the CRLB if they are not equivalent to equalities. For example, the two inequality constraints $\theta \geq 0$ and $\theta \leq 0$, are equivalent to $\theta = 0$. See the definition of a regular point in [4] for details.) Suppose the equality constraints are $\mathrm{g}(\theta) = 0$, then we can define $\mathbf{U}$ as an orthonormal matrix whose columns span the null space of $\nabla_\theta \mathrm{g}(\theta)$, the Jacobian matrix of $\mathrm{g}(\theta)$, i.e., $\nabla_\theta \mathrm{g}(\theta)\mathbf{U} = 0$ and $\mathbf{U}^T\mathbf{U} = \mathrm{I}$. Then the constrained CRLB is modified as

$$\mathrm{E}_{\mathbf{X}}\{\|\theta - \hat{\theta}\|_2^2\} \geq \mathrm{tr}\left\{\mathbf{U}(\mathbf{U}^T\mathbf{F}_\theta\mathbf{U})^\dagger\mathbf{U}^T\right\},$$

where the superscript "†" denotes the pseudoinverse. A simple derivation of the CRLB under affine equality constraints is given in "Cross-Checking the Constrained CRLB."

CROSS-CHECKING THE CONSTRAINED CRLB

It is instructive to check the constrained CRLB for the special case of affine $\mathrm{g}(\theta)$ via the CRLB under transformation [3, Sec. 3.8]. Suppose $\mathrm{g}(\theta) = \mathbf{G}\theta - \mathrm{b} = 0$, and suppose $\mathbf{U}$ satisfies that it is an orthonormal basis of the nullspace of $\mathbf{G}$. Then any feasible θ can be represented by the unconstrained variable α as $\theta = \mathbf{U}\alpha + \boldsymbol{\theta}_0 \Rightarrow \alpha = \mathbf{U}^T(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$, where $\boldsymbol{\theta}_0$ is one feasible point. Thus,

$$\nabla_\theta \ln p(\mathrm{x};\boldsymbol{\theta}) = \mathbf{U}^T\nabla_\alpha \ln p(\mathrm{x};\boldsymbol{\alpha}) \;\Rightarrow\; \mathbf{F}_\alpha = \mathbf{U}^T\mathbf{F}_\theta\mathbf{U}.$$

Now α is an unconstrained parameter to estimate, and the CRLB of θ via transformation of α is

$$(\nabla_\theta\alpha)\mathbf{F}_\alpha^\dagger(\nabla_\theta\alpha)^T = \mathbf{U}(\mathbf{U}^T\mathbf{F}_\theta\mathbf{U})^\dagger\mathbf{U}^T.$$

CRAMÉR–RAO BOUNDS FOR NMF

In this section, we derive the CRLB for both symmetric and asymmetric NMF, under an additive white Gaussian noise (AWGN) model. Note that at low SNRs, Gaussian noise may generate observations having negative values, albeit the probability that this happens is negligible at higher SNRs. Yet the same is true for any additive noise model that is not one sided. A multiplicative noise model can capture two-sided perturbations with nonnegative noise, but if the signal elements are ≥ 1, then taking the logarithm one obtains a NMF model with two-sided additive noise in the log domain. Hence the possibility of having negative data is unavoidable. Furthermore, Gaussian noise is implicitly assumed in all NMF applications where least squares is adopted for model fitting—including, e.g., the hierarchical alternating least squares (HALS) algorithm [30]. This is so because the least squares criterion can be interpreted as ML under a Gaussian noise model. Beyond this, it is interesting to note that for general signal models observed in independent and identically distributed (i.i.d.) additive noise, the CRLB under any noise distribution that possesses everywhere continuous first and second derivatives is the same as the corresponding Gaussian CRLB up to a constant multiplicative factor that depends on the noise distribution [37]. Hence, our results are more general than meets the eye.

IDENTIFIABILITY, FIM, AND CRLB FOR THE SCALAR CASE

Before we delve into FIM and CRLB computations for NMF, it is instructive to consider the scalar case first, particularly $x = wh + n$, where w and h are nonnegative reals. This is clearly unidentifiable unless, e.g., we fix $w = 1$. Then this is equivalent to the linear estimation problem $x = h + n$, and if n is Gaussian with variance σ^2, the CRLB is σ^2. But for now, let us treat it as an estimation problem with two unknown parameters $[w\,h]^T$, with the constraint $w = 1$. Then the FIM is

$$\mathbf{F}_{w,h} = \frac{1}{\sigma^2}\begin{bmatrix} h^2 & wh \\ hw & w^2 \end{bmatrix} = \frac{1}{\sigma^2}\begin{bmatrix} h^2 & h \\ h & 1 \end{bmatrix},$$

while $\mathbf{u} = [0\,1]^T$ spans the null space of the Jacobian of the equality constraint. Therefore, the CRLB is

$$\mathbf{u}\left(\mathbf{u}^T\mathbf{F}_{w,h}\mathbf{u}\right)^{-1}\mathbf{u}^T = \begin{bmatrix} 0 & 0 \\ 0 & \sigma^2 \end{bmatrix},$$

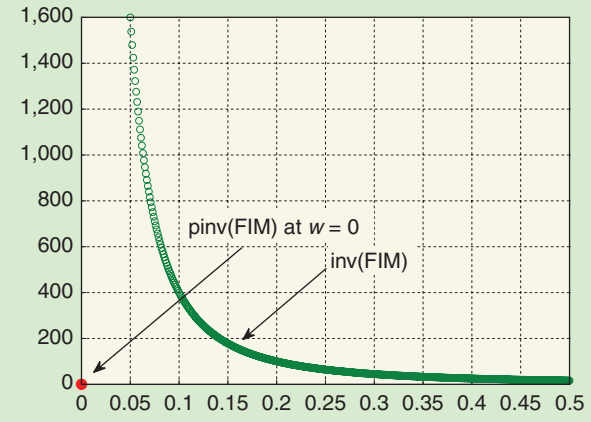
which is consistent with what we get by treating it as a single parameter problem. The symmetric scalar model $x = w^2 + n$ is sign-unidentifiable, but with the nonnegativity constraint $w \geq 0$ it becomes identifiable. For n zero-mean Gaussian with variance σ^2, it is easy to compute the Fisher information for w, which is

$$F_w = \frac{4}{\sigma^2}w^2.$$

Notice that the Fisher information is zero if $w = 0$, and as a special case of pseudoinverse, $0^\dagger = 0$. Since the parametric constraint is an inequality, the CRLB is unaffected according to [4], so for any unbiased estimator $\hat{w}$,

$$\mathbf{E}_x\{(w - \hat{w})^2\} \geq \begin{cases} 0 & w = 0, \\ \dfrac{\sigma^2}{4}w^{-2} & w \neq 0. \end{cases}$$

This is illustrated in Figure S1. Notice that the pseudoinverse of the FIM is a legitimate bound, albeit far from being attainable when $w = 0$. The situation is not as bad in the matrix case—in fact, we will see that existing algorithms come close to attaining the optimistic CRLB obtained from the pseudoinverse, under certain conditions.



[FIGS1] The CRLB for scalar symmetric NMF.

As a warm-up, a derivation of the CRLB for scalar NMF is presented in "Identifiability, FIM, and CRLB for the Scalar Case."

A CRLB FOR SYMMETRIC NMF

Consider the $I \times I$ symmetric matrix $\mathbf{X}$ generated as

$$\mathbf{X} = \mathbf{W}\mathbf{W}^T + \mathbf{N}, \tag{2}$$

where $\mathbf{W}$ is $I \times K$, $\mathbf{W} \geq 0$, and the elements of $\mathbf{N}$ are drawn from an i.i.d. Gaussian distribution with zero-mean and variance σ^2. The $IK \times IK$ Fisher information matrix for $\mathbf{W}$ is

$$\mathbf{F}_{\mathbf{W}} = \frac{2}{\sigma^2}(\mathbf{W}^T\mathbf{W} \otimes \mathbf{I}_I + (\mathbf{I}_K \otimes \mathbf{W})\,\mathbf{P}\,(\mathbf{I}_K \otimes \mathbf{W})^T), \tag{3}$$

where $\mathbf{I}_I$ is the identity matrix of size $I \times I$, and likewise for $\mathbf{I}_K$ and all the boldface $\mathbf{I}$ with a subscript indicating its size in the rest of the article, "$\otimes$" indicates matrix Kronecker product [38, Sec. 10.2.1], and $\mathbf{P}$ is a specific permutation matrix; see the supporting supplementary material that accompanies this article in IEEE *Xplore*. Here the constraints are $\mathbf{W} \geq 0$, which do not affect the CRLB. In addition, $\mathbf{F}_{\mathbf{W}}$ is rank deficient (see the supporting supplementary material), so we need to compute its pseudoinverse to get the CRLB.

In practice, when the size of $\mathbf{W}$ is large, we are usually interested in the overall reconstruction error $\|\mathbf{W} - \hat{\mathbf{W}}\|_F^2$, and the CRLB implies that $\mathrm{E}_{\mathbf{X}}\{\|\mathbf{W} - \hat{\mathbf{W}}\|_F^2\} \geq \mathrm{tr}\{\mathbf{F}_{\mathbf{W}}^\dagger\}$. We also look at the relative error, normalized by $\|\mathbf{W}\|_F^2$, so that the scale and the size of $\mathbf{W}$ are taken into account. Thus, the normalized aggregate CRLB for symmetric NMF is given by

$$\frac{\mathrm{E}_{\mathbf{X}}\{\|\mathbf{W} - \hat{\mathbf{W}}\|_F^2\}}{\|\mathbf{W}\|_F^2} \geq \frac{\mathrm{tr}\{\mathbf{F}_{\mathbf{W}}^\dagger\}}{\|\mathbf{W}\|_F^2}. \tag{4}$$

For $K = 1$, the symmetric decomposition is unique even without nonnegativity constraints, and the FIM is invertible. The CRLB can be calculated in closed form, as provided in "Identifiability, FIM, and CRLB for the Symmetric Vector Case."

Figure 1 illustrates how this normalized CRLB changes as a function of the outer dimension I (the number of rows of $\mathbf{W}$), the inner dimension K (the number of columns of $\mathbf{W}$), and the density (the amount of nonzero entries). The pattern of (non)zeros in $\mathbf{W}$ were drawn from an i.i.d. Bernoulli distribution, and the nonzero entries of $\mathbf{W}$ were drawn from an i.i.d. exponential distribution. In Figure 1(a), the inner dimension is fixed to be ten, while the outer dimension increases from 50 to 150, for different densities; in (b), the outer dimension is fixed at 100, while the inner dimension increases from five to 25, with different densities. In all cases, the SNR

$$\mathrm{SNR} = 10\log_{10}\frac{\|\mathbf{W}\mathbf{W}^T\|_F^2}{I^2\sigma^2}$$

IDENTIFIABILITY, FIM, AND CRLB FOR THE SYMMETRIC VECTOR CASE

Consider the vector case

$$\mathbf{X} = \mathbf{w}\mathbf{w}^T + \mathbf{N}.$$

Obviously, this problem is also identifiable if $\mathbf{N} = 0$, apart from a sign ambiguity. We do not need to impose nonnegativity constraints on all the elements of w to resolve the ambiguity, but only on one element, e.g., $w_1 \geq 0$. The FIM can be computed as a special case of the formula (3), whose derivation can be found in the supplementary material in IEEE *Xplore*, yielding

$$\mathbf{F}_w = \frac{2}{\sigma^2}(\|\mathbf{w}\|^2 I_I + \mathbf{w}\mathbf{w}^T),$$

which is nonsingular for $\mathbf{w} \neq \mathbf{0}$, and we can calculate its inverse in closed form, using the matrix inversion lemma [38],

$$\mathbf{F}_{\mathbf{w}}^{-1} = \frac{\sigma^2}{2}\left(\|\mathbf{w}\|^{-2}\mathbf{I}_I - \frac{1}{2}\|\mathbf{w}\|^{-4}\mathbf{w}\mathbf{w}^T\right).$$

Thus,

$$\frac{\mathbf{E}_{\mathbf{X}}\{\|\mathbf{w} - \hat{\mathbf{w}}\|^2\}}{\|\mathbf{w}\|^2} \geq \frac{\mathrm{tr}\{\mathbf{F}_{\mathbf{w}}^{-1}\}}{\|\mathbf{w}\|^2} = \frac{\sigma^2}{2}\left(I - \frac{1}{2}\right)\|\mathbf{w}\|^{-4}.$$

Notice here that italic I is the dimension of w (not to be confused with the identity matrix I).

is fixed at 10 dB. Each CRLB with the specified size and density is calculated as the average of 100 Monte Carlo draws of W. Note how the density of W affects the CRLB—the sparser the latent factors, the lower the CRLB. Not surprisingly, the CRLB increases as the ratio between the outer dimension and the inner dimension decreases.

CRLB FOR ASYMMETRIC NMF

Consider the $I \times J$ asymmetric matrix generated as

$$\mathrm{X} = \mathrm{W}\mathrm{H}^T + \mathrm{N}, \tag{5}$$

where W is $I \times K$, $\mathrm{W} \geq 0$, H is $J \times K$, $\mathrm{H} \geq 0$, and the elements of N are drawn from an i.i.d. Gaussian distribution with zero-mean and variance σ^2. The $(I+J)K \times (I+J)K$ Fisher information matrix of W and H is (cf. supporting supplementary material in IEEE *Xplore*, which also shows that $\mathrm{F}_{\mathrm{W,H}}$ is rank deficient)

$$\mathrm{F}_{\mathrm{W,H}} = \frac{1}{\sigma^2}\begin{bmatrix} \mathrm{H}^T\mathrm{H} \otimes \mathrm{I}_I & (\mathrm{I}_K \otimes \mathrm{W})\,\mathrm{P}\,(\mathrm{I}_K \otimes \mathrm{H})^T \\ (\mathrm{I}_K \otimes \mathrm{H})\,\mathrm{P}\,(\mathrm{I}_K \otimes \mathrm{W})^T & \mathrm{W}^T\mathrm{W} \otimes \mathrm{I}_J \end{bmatrix}. \tag{6}$$

Here, the constraints on the parameters are $\mathrm{W} \geq 0$, $\mathrm{H} \geq 0$, and (1). In calculating the CRLB, we only need to take into account the equality constraints. The Jacobian of the equality constraints over W is

$$\nabla_{\mathrm{vec(W)}}\begin{bmatrix} \sum_{i=1}^{I} w_{i1} - 1 \\ \vdots \\ \sum_{i=1}^{I} w_{iK} - 1 \end{bmatrix} = \mathrm{I}_K \otimes 1^T,$$

where 1 is the all 1 vector with dimension I. Upon defining

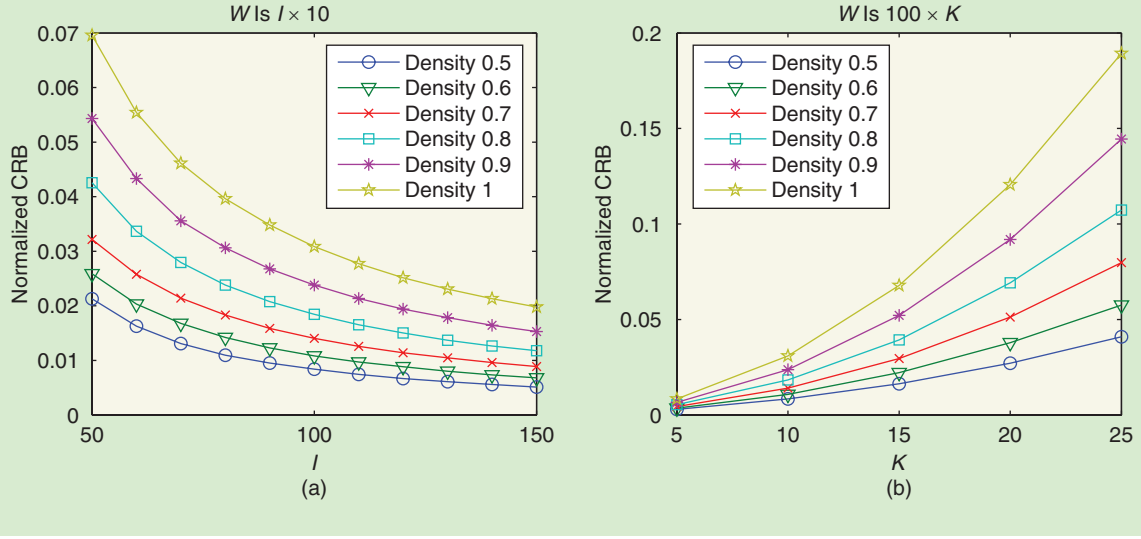
$$\mathrm{v}_i = \frac{1}{\sqrt{i+i^2}}\left(\sum_{l=1}^{i} e_l - i e_{i+1}\right), \quad \mathrm{V} = [\mathrm{v}_1 \;\; \mathrm{v}_2 \;\; \cdots \;\; \mathrm{v}_{I-1}], \tag{7}$$

we have $\mathrm{V}^T 1 = 0$ and $\mathrm{V}^T\mathrm{V} = \mathrm{I}_{I-1}$. Therefore, let

$$\mathrm{U} = \begin{bmatrix} \mathrm{I}_K \otimes \mathrm{V} & 0 \\ 0 & \mathrm{I}_{JK} \end{bmatrix},$$

satisfying

$$\left(\nabla_{\begin{bmatrix}\mathrm{vec(W)} \\ \mathrm{vec(H)}\end{bmatrix}}\begin{bmatrix} \sum_{i=1}^{I} w_{i1} - 1 \\ \vdots \\ \sum_{i=1}^{I} w_{iK} - 1 \end{bmatrix}\right)\mathrm{U} = 0, \;\; \mathrm{U}^T\mathrm{U} = \mathrm{I}_{(I+J-1)K}.$$



[FIG1] (a) and (b) The symmetric NMF CRLB—how the outer dimension, inner dimension, and density affects the CRLB, for SNR=10 dB.

Using the FIM $\mathbf{F}_{\mathbf{W},\mathbf{H}}$ and the null basis $\mathbf{U}$ above, we obtain the CRLB for $\mathbf{W}$ and $\mathbf{H}$ as $\mathbf{U}(\mathbf{U}^T\mathbf{F}_{\mathbf{W},\mathbf{H}}\mathbf{U})^{\dagger}\mathbf{U}^T$.

In practice, the reconstruction errors $\|\mathbf{W}-\hat{\mathbf{W}}\|_F^2$ and $\|\mathbf{H}-\hat{\mathbf{H}}\|_F^2$ are usually assessed separately since $\mathbf{W}$ and $\mathbf{H}$ model different entities (e.g., loadings and scores). Partition $(\mathbf{U}^T\mathbf{F}_{\mathbf{W},\mathbf{H}}\mathbf{U})^{\dagger}$ into blocks

$$(\mathbf{U}^T\mathbf{F}_{\mathbf{W},\mathbf{H}}\mathbf{U})^{\dagger}=\begin{bmatrix}\boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2\\ \boldsymbol{\Phi}_2^T & \boldsymbol{\Phi}_3\end{bmatrix},$$

where $\boldsymbol{\Phi}_1$ is $IK\times IK$ and $\boldsymbol{\Phi}_3$ is $JK\times JK$. Then

$$\frac{\mathrm{E}_{\mathbf{X}}\{\|\mathbf{W}-\hat{\mathbf{W}}\|_F^2\}}{\|\mathbf{W}\|_F^2}\geq\frac{\mathrm{tr}\{(\mathbf{I}_K\otimes\mathbf{V})\boldsymbol{\Phi}_1(\mathbf{I}_K\otimes\mathbf{V})^T\}}{\|\mathbf{W}\|_F^2},\tag{8a}$$

$$\frac{\mathrm{E}_{\mathbf{X}}\{\|\mathbf{H}-\hat{\mathbf{H}}\|_F^2\}}{\|\mathbf{H}\|_F^2}\geq\frac{\mathrm{tr}\{\boldsymbol{\Phi}_3\}}{\|\mathbf{H}\|_F^2},\tag{8b}$$

with similar normalization as in the symmetric case.

Similar to the symmetric case, for $K=1$ the asymmetric decomposition is essentially unique, and the matrix we need to pseudoinvert for calculating the CRLB is actually nonsingular. The closed form CRLB for this case is given in "Identifiability, FIM, and CRLB for the Asymmetric Vector Case."

Figure 2 plots the CRLB for asymmetric NMF for various sizes and densities. Figure 2(a) and (b) shows the CRLB for $\mathbf{W}$, which is constrained such that each column sums up to one, while (c) and (d) show the CRLB for $\mathbf{H}$, which does not have any scaling constraints. Figure 2(a) and (c) shows the CRLB when the size of $\mathbf{W}$ is fixed at 100×10, and the number of rows in $\mathbf{H}$ increases from 50 to 150, with different densities. Figure 2(b) and (d) shows the CRLB when the number of rows in $\mathbf{W}$ and $\mathbf{H}$ is fixed at 100 and 120, respectively, and the number of columns in $\mathbf{W}$ and $\mathbf{H}$ increases from five to 25, with different densities. As usual, SNR

$$\mathrm{SNR}=10\log_{10}\frac{\|\mathbf{W}\mathbf{H}^T\|_F^2}{IJ\sigma^2}$$

is fixed at 10 dB. Each CRLB point for a specified size and density is calculated as the average of 100 Monte Carlo draws. Figure 2(c) may seem curious: it shows the normalized CRLB with respect to $\mathbf{H}$ when we fix $\mathbf{W}$ and gradually increase the number of rows of $\mathbf{H}$, and we observe that the normalized CRLB does not change very much. It slowly increases as the outer-dimension of $\mathbf{H}$ increases, as opposed to the normalized CRLB for $\mathbf{W}$, which seems to decrease exponentially. This is because the block in the FIM $\mathbf{F}_{\mathbf{W},\mathbf{H}}$ that corresponds to $\mathbf{H}$ is $\mathbf{W}^T\mathbf{W}\otimes\mathbf{I}_J$, where the dimension of $\mathbf{I}_J$ changes according to the dimension of $\mathbf{H}$, which contributes the most to the block of the CRLB that corresponds to $\mathbf{H}$. The $\mathbf{W}^T\mathbf{W}$ part is fixed, and the size of $\mathbf{I}_J$ grows approximately linearly with $\|\mathbf{H}\|_F^2$, which explains intuitively why the normalized CRLB for $\mathbf{H}$ does not change very much. Apart from that, the overall tendency of the CRLB versus the size is similar to the symmetric case: it goes down as one of the outer dimensions increases, and it goes up as the common inner dimension increases, as intuitively expected from "equations versus unknowns" considerations. Note, however, that here as the number of observations increases, so does the number of unknown parameters. For example, if a new column is appended to $\mathbf{X}$ then a new row is appended to $\mathbf{H}$ as well, and the CRLB may worsen, depending on the new entries and other factors [e.g., the way we resolve the scaling ambiguity; see Figure 2(a) and (c)].

What is more, the sparser $\mathbf{W}$ and $\mathbf{H}$, the lower the CRLB in all cases.

PUTTING NMF ALGORITHMS TO THE TEST

SYMMETRIC NMF

We compared three algorithms for symmetric NMF with the CRLB derived in the section "Cramer–Rao Bounds for NMF." These are α-symmetric NMF and β-symmetric NMF with $\alpha=\beta=0.99$ [17], and the algorithm recently proposed in [14]. The true $\mathbf{W}$ is generated such that a certain proportion of its entries are randomly set to zero, and the rest are drawn from an i.i.d. exponential distribution. Using the generative model (2) the

IDENTIFIABILITY, FIM, AND CRLB FOR THE ASYMMETRIC VECTOR CASE

For $K=1$, i.e., when $\mathbf{w}$ and $\mathbf{h}$ are vectors, asymmetric factorization is identifiable from noiseless (rank-one) data, similar to the symmetric case. There is still a scaling issue, and we can resolve this by fixing the scaling of one factor, e.g., setting $\mathbf{1}^T\mathbf{w}=1$ as we did in the matrix case. Then, using (6), the FIM is

$$\mathbf{F}_{\mathbf{w},\mathbf{h}}=\frac{1}{\sigma^2}\begin{bmatrix}\|\mathbf{h}\|^2\mathbf{I}_I & \mathbf{w}\mathbf{h}^T\\ \mathbf{h}\mathbf{w}^T & \|\mathbf{w}\|^2\mathbf{I}_J\end{bmatrix}.$$

The corresponding $\mathbf{U}$ matrix is

$$\mathbf{U}=\begin{bmatrix}\mathbf{V} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_J\end{bmatrix},$$

with the same $\mathbf{V}$ as defined in (7). Let us first try to calculate the following inversion

$$(\mathbf{U}^T\mathbf{F}_{\mathbf{w},\mathbf{h}}\mathbf{U})^{-1}=\sigma^2\begin{bmatrix}\|\mathbf{h}\|^2\mathbf{I}_{I-1} & \mathbf{V}^T\mathbf{w}\mathbf{h}^T\\ \mathbf{h}\mathbf{w}^T\mathbf{V} & \|\mathbf{w}\|^2\mathbf{I}_J\end{bmatrix}^{-1}=\begin{bmatrix}\boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2\\ \boldsymbol{\Phi}_2^T & \boldsymbol{\Phi}_3\end{bmatrix},$$

where $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_3$ are the inverse of the Schur complement [36, p. 650] of $\sigma^{-2}\|h\|^2\mathbf{I}_{I-1}$ and $\sigma^{-2}\|\mathbf{w}\|^2 I_J$, respectively, in $\mathbf{U}^T\mathbf{F}_{\mathbf{w},\mathbf{h}}\mathbf{U}$, i.e.,

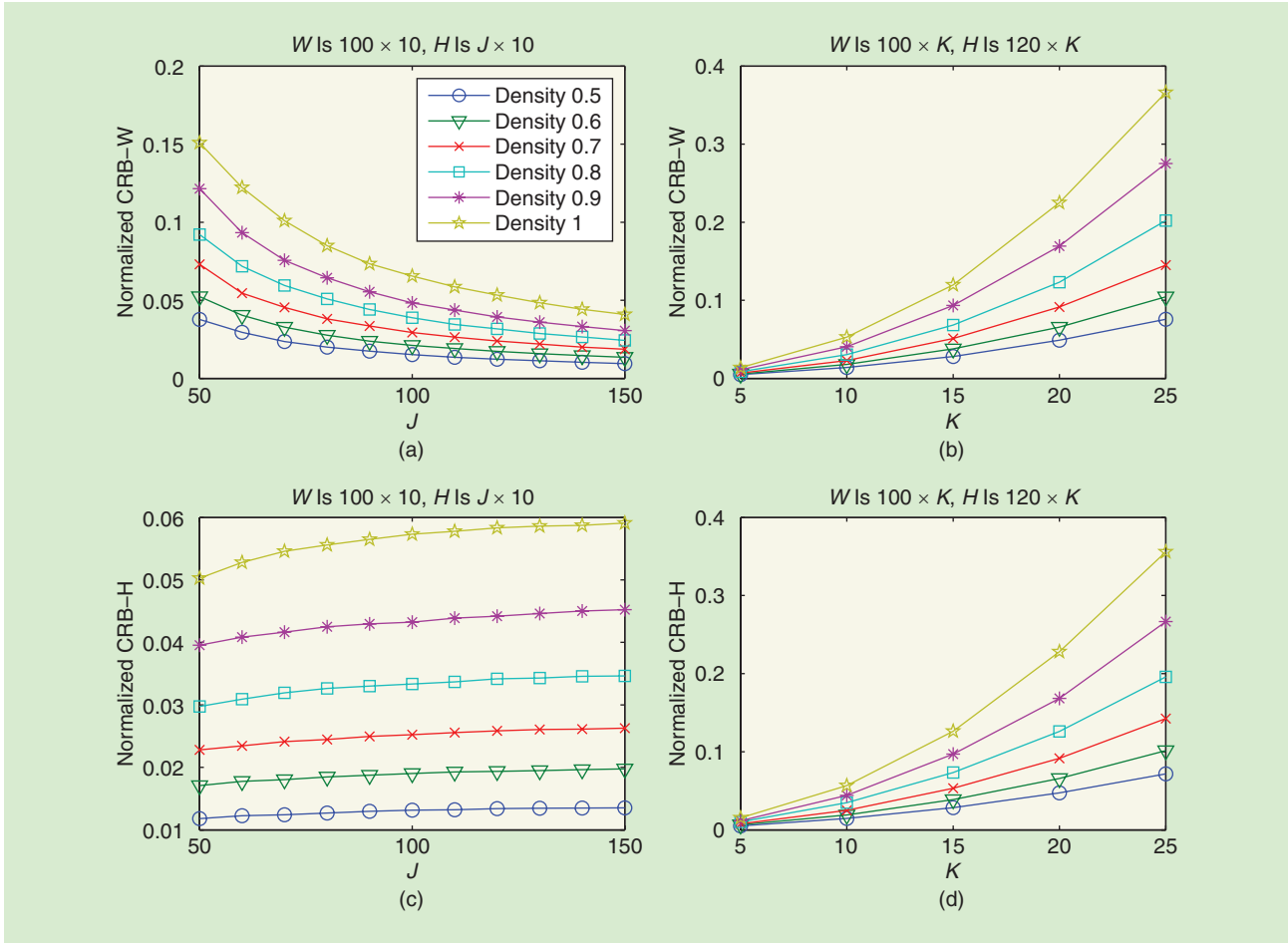
$$\boldsymbol{\Phi}_1=\sigma^2\left(\|\mathbf{h}\|^2\mathbf{I}_{I-1}-\frac{\|\mathbf{h}\|^2}{\|\mathbf{w}\|^2}\mathbf{V}^T\mathbf{w}\mathbf{w}^T\mathbf{V}\right)^{-1},$$

$$\boldsymbol{\Phi}_3=\sigma^2\left(\|\mathbf{w}\|^2\mathbf{I}_J-\frac{\|\mathbf{V}^T\mathbf{w}\|^2}{\|\mathbf{h}\|^2}\mathbf{h}\mathbf{h}^T\right)^{-1}.$$

Again, the inverses can be calculated in closed form by using the matrix inversion lemma. Using the Pythagorean theorem $\|\mathbf{w}\|^2=\|\mathbf{V}^T\mathbf{w}\|^2+(1)/(I)\|\mathbf{1}^T\mathbf{w}\|^2$ (details omitted), we obtain

$$\frac{\mathbf{E}_{\mathbf{X}}\{\|\mathbf{w}-\hat{\mathbf{w}}\|^2\}}{\|\mathbf{w}\|^2}\geq\frac{\mathrm{tr}\{\mathbf{V}\boldsymbol{\Phi}_1\mathbf{V}^T\}}{\|\mathbf{w}\|^2}=\sigma^2\|\mathbf{w}\|^{-2}\|\mathbf{h}\|^{-2}(I-1+I(\mathbf{V}^T\mathbf{w}(^2),$$

$$\frac{\mathbf{E}_{\mathbf{X}}\{\|\mathbf{h}-\hat{\mathbf{h}}\|^2\}}{\|\mathbf{h}\|^2}\geq\frac{\mathrm{tr}\{\boldsymbol{\Phi}_3\}}{\|\mathbf{h}\|^2}=\sigma^2\|\mathbf{w}\|^{-2}\|\mathbf{h}\|^{-2}(J+I\|\mathbf{V}^T\mathbf{w}\|^2).$$



[FIG2] (a)–(d) The asymmetric NMF CRLB—how the outer dimensions, inner dimension, and density affects the CRLB, for SNR=10 dB.

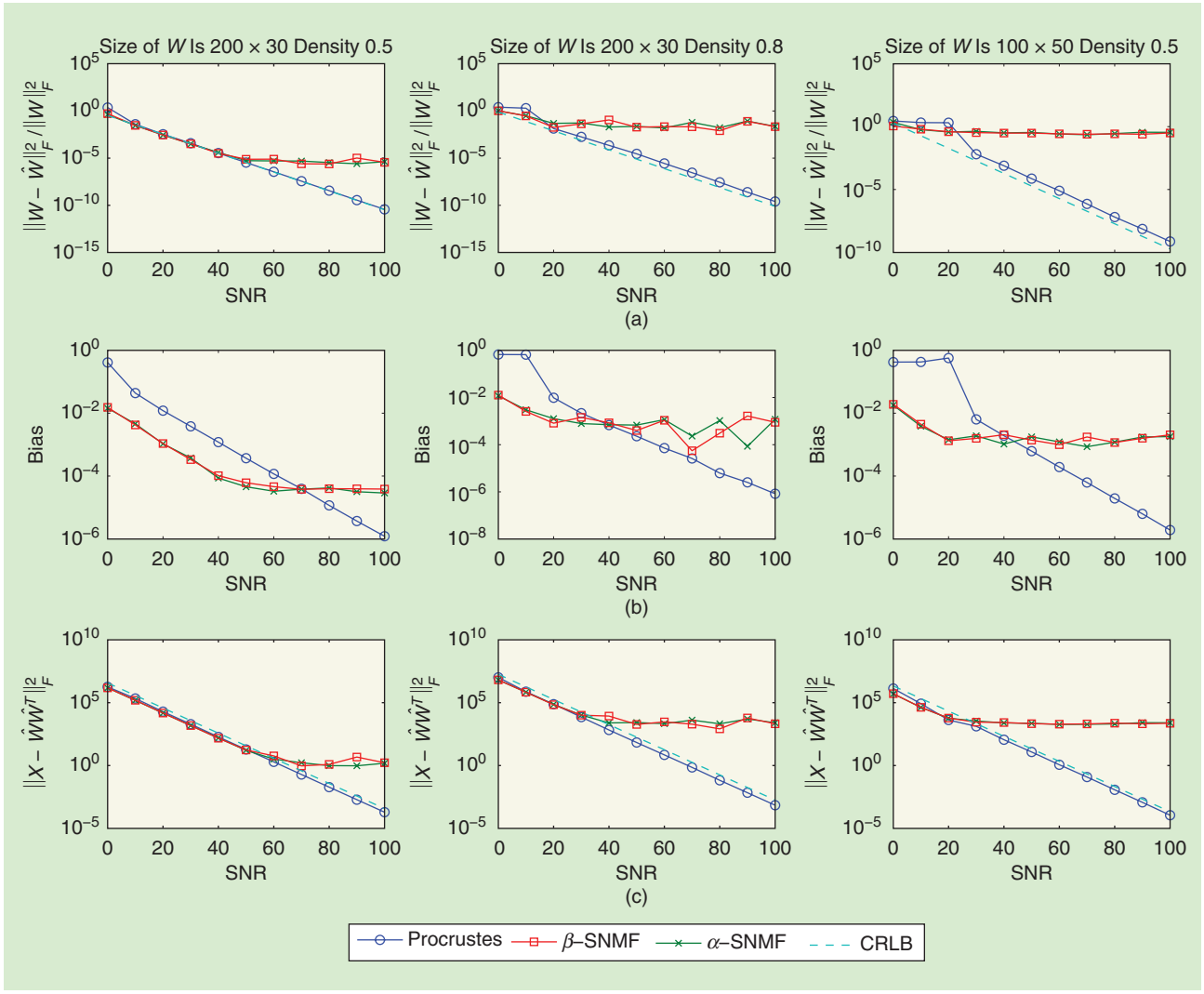
resulting $\mathbf{X}$ will not be symmetric, so we use $(1/2)(\mathbf{X}+\mathbf{X}^T)$, since all algorithms are designed specifically for symmetric nonnegative matrices. Reference [17] did not provide a termination criterion, so both α-symmetric NMF and β-symmetric NMF are left to run for a large number of iterations (10^4), to ensure the best possible results. For the algorithm in [14], we used the termination criterion described in [14, Fig. 4] with the tolerance set to machine precision eps. We used a single draw of $\mathbf{W}$ for each (size, density) combination reported. Under various SNRs, the normalized squared error $(\|\hat{\mathbf{W}}-\mathbf{W}_F^2\|)/(\|\mathbf{W}\|_F^2)$ is calculated and averaged over 100 Monte Carlo tests, so that we can get a better approximation to the expected error $\mathbf{E}_X\{(\|\hat{\mathbf{W}}-\mathbf{W}\|_F^2)/(\|\mathbf{W}\|_F^2)\}$.

The results are plotted in Figure 3, where (a) shows the normalized squared error benchmarked by the CRLB, (b) shows the (aggregate) bias for each estimate, defined as

$$\text{bias} = \left\| \frac{1}{T}\sum_{t=1}^{T}(\mathbf{W}-\hat{\mathbf{W}}_t) \right\|_F, \tag{9}$$

where T is the number of trials, in this case 100, and (c) shows the model fitting error for each algorithm. The dashed lines in (c) show the total noise power; a good approximation should yield a fitting error close to the noise power. The plots in the left column show a case where the symmetric NMF problem is relatively "overdetermined," since the inner dimension (30) is small compared to the outer dimension (200), and the latent factors are quite sparse (density 0.5). The two other columns show more difficult cases—low rank (30 versus 200) but relatively dense latent factors for the middle column, not-so-low rank (50 versus 100) but relatively sparse latent factors for the right column. Recall the discussion in the section "Fundamentals" for the rule of thumb for when identifiability can be expected—the middle and right columns illustrate cases where this requirement is barely satisfied.

In all cases, the aggregate bias is small and goes to zero as SNR increases, indicating that the estimates provided by these algorithms are asymptotically unbiased, and we can use the CRLB to approximately bound the performance. Generally speaking, α/β-symmetric NMF slightly outperform the Procrustes rotation algorithm [14] in the low SNR regime but fail to reach the CRLB in the high SNR regime. The algorithm in [14] exhibits classic threshold behavior—for SNR higher than some threshold, the mean square error (MSE) stays close to the CRLB. The reason is that it employs eigenanalysis to estimate the column space of $\mathbf{W}$ as a first step and then applies Procrustes rotations in the estimated subspace. On the other hand, both



[FIG3] (a) The normalized squared error of three existing symmetric NMF algorithms versus the CRLB; similarly, (b) shows the (aggregate) bias, and (c) shows the fitting error.

symmetric NMF variants are modifications of the multiplicative update algorithm using $\|\mathrm{X} - \mathrm{WW}^T\|_F^2$ (Gaussian log-likelihood) as the objective, so that it is not surprising that they perform better in the low-SNR regime. We can also see this from Figure 3(b), as the biases of α/β-symmetric NMF are lower than that of the Procrustes method under low SNR.

ASYMMETRIC NMF

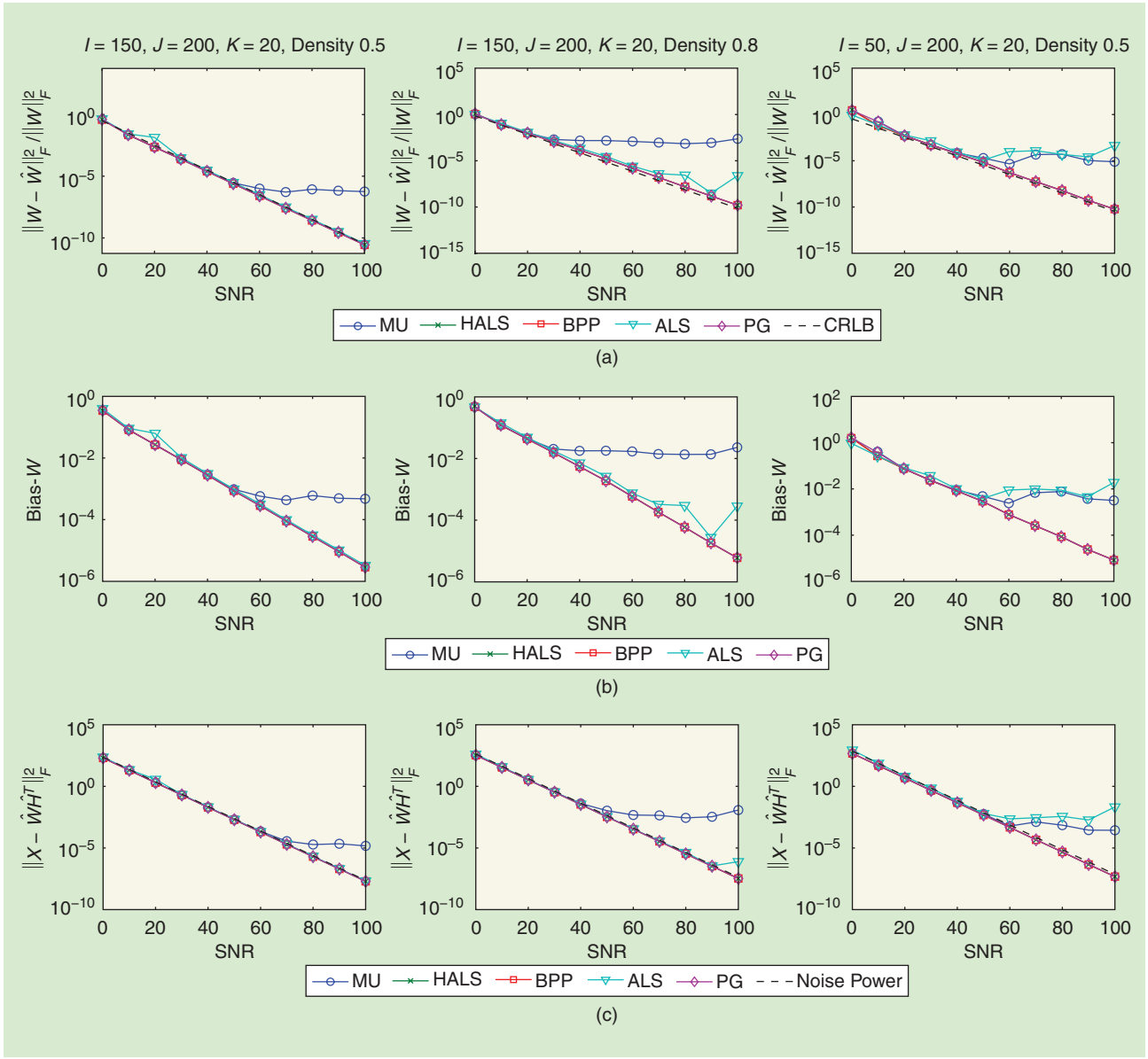
In this section, we compare several asymmetric NMF algorithms aiming to minimize the Euclidian distance. Notice that the data we synthetically generated were corrupted by additive i.i.d. Gaussian noise, so using Euclidian distance as the objective actually gives us the ML estimate. This is why algorithms that use other divergence functions as the objective were not considered here. The algorithms tested are:

- multiplicative update (MU) proposed by Lee and Seung [22]
- alternating least squares (ALS) proposed by Berry et al. [26]
- projected gradient (PG) proposed by Lin [24] (the MATLAB code can be downloaded from http://www.csie.ntu.edu.tw/~cjlin/nmf/index.html)
- fast HALS proposed by Cichocki and Phan [30, Algor. 2]
- block principle pivoting (BPP) alternating nonnegative least squares using BPP proposed by Kim and Park [29] (the MATLAB code can be downloaded from http://www.cc.gatech.edu/~hpark/nmfsoftware.php).

For all algorithms, we used the optimality condition in [39] to check for termination, i.e., calculate

$$\left\| \begin{bmatrix} \mathrm{W} \circledast ((\mathrm{X} - \mathrm{WH}^T)\mathrm{H}) \\ \mathrm{H} \circledast ((\mathrm{X}^T - \mathrm{HW}^T)\mathrm{W}) \end{bmatrix} \right\|_F$$

in each iteration and terminate when it is smaller than the machine precision eps, with a maximum number of iteration set as

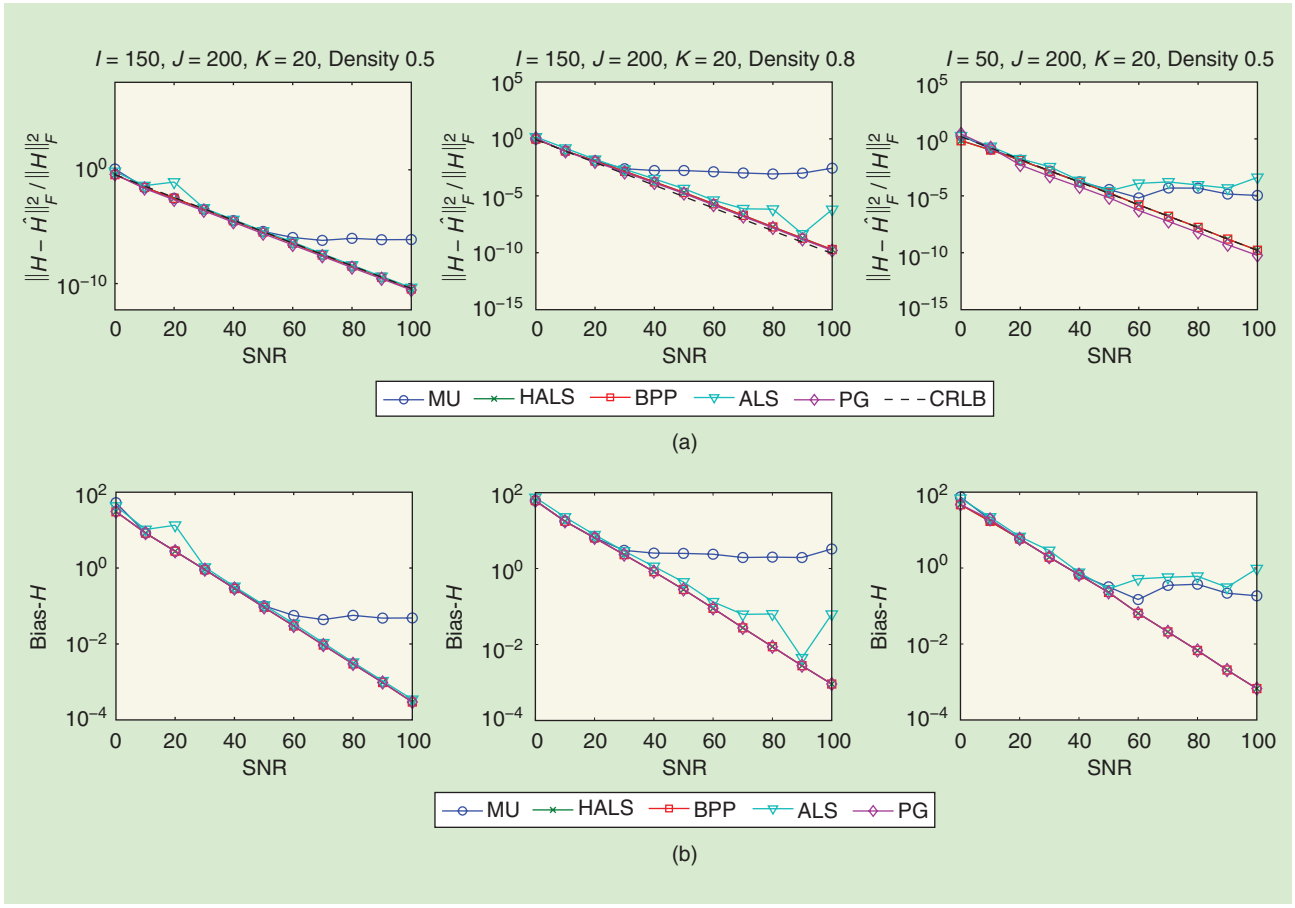


[FIG4] In (a), the normalized squared error for W using various asymmetric NMF algorithms versus the CRLB is shown; similarly, (b) shows the (aggregate) bias for W and (c) the fitting error.

10^4. In the expression, $\circledast$ stands for the Hadamard (element-wise) matrix product. Similar to the symmetric case, the entries of $\mathbf{W}$ and $\mathbf{H}$ were generated such that a certain proportion of them are randomly set to 0, and the rest are drawn from an i.i.d. exponential distribution. Then the columns of $\mathbf{W}$ are scaled to sum up to one.

Three tests were conducted and illustrated in Figures 4 and 5 for $\mathbf{W}$ and $\mathbf{H}$, respectively—low-rank and sparse latent factors on the left, low rank but moderately dense in the middle, and an unbalanced case (J much larger than I) where the rank is not small compared to the smaller outer dimension, with density set relatively small to ensure identifiability. Similar to Figure 3, Figures 4(a) and 5(a) show the normalized squared error for each algorithm benchmarked by the CRLB, Figures 4(b) and 5(b) show the (aggregate) bias of $\mathbf{W}$ as defined in (9), and similarly for $\mathbf{H}$, and Figure 4(c) shows the fitting error for each algorithm.

As we can see from Figures 4(b) and 5(b) the biases are generally small and approach zero with increasing SNR, indicating that we can use the CRLB to approximately bound performance. In all three cases, HALS, BPP, and PG were able to provide a good estimate with MSE close to the CRLB, under all SNRs tested. On the other hand, MU and ALS are not guaranteed to work well even under very high SNR. All methods separate the variables into blocks, and HALS, BPP, and PG aim to find the conditionally optimal point before moving to the next block, whereas the updates of MU and ALS cannot guarantee this. Interestingly, in the "well-posed" case shown in the left columns of Figures 4 and 5, ALS



[FIG5] (a) The normalized squared error for H using various asymmetric NMF algorithms versus the CRLB; similarly (b) shows the (aggregate) bias for H.

gave similar results to those three methods, indicating that if we know a priori that the latent factors are both low rank and sparse, it is worth trying ALS, since its updates rules only require linear least-squares followed by simple projection to the nonnegative orthant, which is much simpler than the rest.

RECAP AND TAKE-HOME POINTS

WHAT WE LEARNED

NMF entails a singular FIM as well as constraints and ambiguities that must be dealt with in the computation of the pertinent CRLB. We learned how to tackle those and used the results to benchmark and develop insights on what can be expected from some of the best available algorithms. For symmetric NMF, the CRLB can be approached using the Procrustes rotation algorithm [14] in the high SNR regime, or α/β-symmetric NMF in low SNR cases. For asymmetric NMF, the best-performing algorithms were able to give results with MSE close to the CRLB. In both cases, approaching the CRLB is possible when the signal rank is small and the latent factors are not dense, i.e., when there is a small number of latent components whose loadings contain sufficiently many zeros. This is quite remarkable given that the CRLB with a singular FIM is generally unattainable; see Figure S1. There may be room for improvement in cases involving moderate SNR and/or moderate rank and/or moderate density.

WHY IT IS IMPORTANT

Beyond NMF, the approach and techniques we learned can be used to facilitate analogous derivations for related factor analysis problems. For example, the FIMs provided here can be applied to more general bilinear matrix factorizations, e.g., using other types of constraints on W. The FIM will remain the same, but the U matrix will be different. Also, we can exploit a basis of the nullspace of the FIM to reduce the complexity of computing its pseudoinverse, and this idea is more broadly applicable to other bilinear matrix factorizations. The results can also be extended toward, e.g., nonnegative tensor factorization.

SUPPLEMENTARY MATERIAL

The supplementary material that is available through IEEE *Xplore* contains detailed FIM derivations, as well as auxiliary results on FIM rank and efficient numerical computation of its pseudoinverse. These results reduce the complexity of computing the CRLB from $O((IK)^3)$ to $O(IK^5)$ in the symmetric case, and from $O(((I+J)K)^3)$ to $O((I+J)K^5)$ in the asymmetric case (recall $I, J \geq K$, and usually $I, J \gg K$). The supplementary

material also includes streamlined and optimized MATLAB code for computing these CRLBs.

AUTHORS

Kejun Huang (huang663@umn.edu) received the B.Eng. degree in communication engineering from Nanjing University of Information Science and Technology, Nanjing, China, in 2010. He has been working toward his Ph.D. degree in the Department of Electrical and Computer Engineering, University of Minnesota, since 2010. His research interests include signal processing, machine learning, and data analytics. His current research focuses on identifiability, algorithms, and performance analysis for factor analysis of big matrix and tensor data.

Nicholas D. Sidiropoulos (nikos@umn.edu) received the diploma in electrical engineering from the Aristotelian University of Thessaloniki, Greece, and M.S. and Ph.D. degrees in electrical engineering from the University of Maryland, College Park, in 1988, 1990, and 1992, respectively. He was an assistant professor at the University of Virginia (1997–1999); associate professor at the University of Minnesota, Minneapolis (2000–2002); professor at the Technical University of Crete, Greece (2002–2011); and professor at the University of Minnesota, Minneapolis (2011–present). His current research focuses on signal and tensor analytics, with applications in cognitive radio, big data, and preference measurement. He received the National Science Foundation/CAREER Award (1998), the IEEE Signal Processing Society (SPS) Best Paper Award (2001, 2007, 2011), and the IEEE SPS Meritorious Service Award (2010). He has served as an IEEE SPS Distinguished Lecturer (2008–2009) and chair of the IEEE Signal Processing for Communications and Networking Technical Committee (2007–2008). He received the Distinguished Alumni Award of the Department of Electrical and Computer Engineering, University of Maryland, College Park (2013).

REFERENCES

[1] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, no. 2, pp. 111–126, 1994.

[2] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[3] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[4] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. Inform. Theory*, vol. 36, no. 6, pp. 1285–1301, 1990.

[5] P. Stoica and B. C. Ng, "On the Cramér–Rao bound under parametric constraints," *IEEE Signal Processing Lett.*, vol. 5, no. 7, pp. 177–179, 1998.

[6] P. Stoica and T. L. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Trans. Signal Processing*, vol. 49, no. 1, pp. 87–90, 2001.

[7] Z. Ben-Haim and Y. C. Eldar, "On the constrained Cramér–Rao bound with a singular Fisher information matrix," *IEEE Signal Processing Lett.*, vol. 16, no. 6, pp. 453–456, 2009.

[8] P. Tichavský and Z. Koldovský, "Optimal pairing of signal components separated by blind techniques," *IEEE Signal Processing Lett.*, vol. 11, no. 2, pp. 119–122, 2004.

[9] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Res. Logist. Quart.*, vol. 2, no. 1–2, pp. 83–97, 1955.

[10] R.E. Burkard, M. Dell'Amico, and S. Martello, *Assignment Problems*. Philadelphia, PA: SIAM, 2009.

[11] D. L. Donoho and V. C. Stodden, "When does non-negative matrix factorization give a correct decomposition into parts?," in *Advances in Neural Information Processing Systems (NIPS)*. Cambridge, MA: MIT Press, 2003, vol. 16, pp. 1141–1148.

[12] H. Laurberg, M. G. Christensen, M. D. Plumbley, L. K. Hansen, and S. H. Jensen, "Theorems on positive data: On the uniqueness of NMF," *Computat. Intell. Neurosci.*, vol. 2008, Article ID 764206, 9 pages, DOI: 10.1155/2008/764206.

[13] N. Gillis, "Sparse and unique nonnegative matrix factorization through data preprocessing," *J. Mach. Learn. Res.*, vol. 13, pp. 3349–3386, Nov. 2012.

[14] K. Huang, N. D. Sidiropoulos, and A. Swami, "Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric decomposition," *IEEE Trans. Signal Processing*, vol. 62, no. 1, pp. 211–224, Jan. 2014.

[15] S. A. Vavasis, "On the complexity of nonnegative matrix factorization," *SIAM J. Optim.*, vol. 20, no. 3, pp. 1364–1377, 2009.

[16] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation*. Hoboken, NJ: Wiley, 2009.

[17] Z. He, S. Xie, R. Zdunek, G. Zhou, and A. Cichocki, "Symmetric nonnegative matrix factorization: Algorithms and applications to probabilistic clustering," *IEEE Trans. Neural Networks*, vol. 22, no. 12, pp. 2117–2131, 2011.

[18] A. Berman and N. Shaked-Monderer, *Completely Positive Matrices*. Singapore: World Scientific, 2003.

[19] P. J. C. Dickinson and L. Gijben, "On the computational complexity of membership problems for the completely positive cone and its dual," *Computat. Optim. Applicat.*, submitted for publication. DOI:10.1007/s10589-013-9594-z

[20] Z. Yang and E. Oja, "Quadratic nonnegative matrix factorization," *Pattern Recognit.*, vol. 45, no. 4, pp. 1500–1510, 2012.

[21] C. Ding, X. He, and H. D. Simon, "On the equivalence of nonnegative matrix factorization and spectral clustering," in *Proc. SIAM Int. Conf. Data Mining (SDM'05)*, 2005, vol. 5, pp. 606–610.

[22] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Adv. Neural Inform. Process. Syst. (NIPS)*, vol. 13, pp. 556–562, 2001.

[23] A. Cichocki, S. Amari, R. Zdunek, R. Kompass, G. Hori, and Z. He, "Extended SMART algorithms for non-negative matrix factorization," in *Artificial Intelligence and Soft Computing (ICAISC)*. New York: Springer, 2006, pp. 548–562.

[24] C.-J. Lin, "Projected gradient methods for nonnegative matrix factorization," *Neural Computat.*, vol. 19, no. 10, pp. 2756–2779, 2007.

[25] R. Zdunek and A. Cichocki, "Nonnegative matrix factorization with constrained second-order optimization," *Signal Process.*, vol. 87, no. 8, pp. 1904–1916, 2007.

[26] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," *Computat. Stat. Data Anal.*, vol. 52, no. 1, pp. 155–173, 2007.

[27] M. Heiler and C. Schnörr, "Learning sparse representations by non-negative matrix factorization and sequential cone programming," *J. Mach. Learn. Res.*, vol. 7, pp. 1385–1407, July 2006.

[28] H. Kim and H. Park, "Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method," *SIAM J. Matrix Anal. Applicat.*, vol. 30, no. 2, pp. 713–730, 2008.

[29] J. Kim and H. Park, "Fast nonnegative matrix factorization: An active-set-like method and comparisons," *SIAM J. Sci. Comput.*, vol. 33, no. 6, pp. 3261–3281, 2011.

[30] A. Cichocki and A.-H. Phan, "Fast local algorithms for large scale nonnegative matrix and tensor factorizations," *IEICE Trans. Fundam. Electron., Commun. Comput. Sci.*, vol. 92, no. 3, pp. 708–721, 2009.

[31] B. Klingenberg, J. Curry, and A. Dougherty, "Non-negative matrix factorization: Ill-posedness and a geometric algorithm," *Pattern Recognit.*, vol. 42, no. 5, pp. 918–928, 2009.

[32] R. Zdunek, "Initialization of nonnegative matrix factorization with vertices of convex polytope," in *Artificial Intelligence and Soft Computing*. New York: Springer, 2012, pp. 448–455.

[33] J. Li and J. M. Bioucas-Dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, 2008, vol. 3, pp. 250–253.

[34] W. S. B. Ouedraogo, A. Souloumiac, M. Jaidane, and C. Jutten, "Simplicial cone shrinking algorithm for unmixing nonnegative sources," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)* 2012, pp. 2405–2408.

[35] M. T. Chu and M. M. Lin, "Low-dimensional polytope approximation and its applications to nonnegative matrix factorization," *SIAM J. Sci. Comput.*, vol. 30, no. 3, pp. 1131–1155, 2008.

[36] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge, UK: Cambridge Univ. Press, 2004.

[37] A. Swami, "Cramér-Rao bounds for deterministic signals in additive and multiplicative noise," *Signal Process.*, vol. 53, no. 2, pp. 231–244, 1996.

[38] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*. Kongens Lyngby, Denmark: Technical Univ. Denmark, 2006.

[39] M. Chu, F. Diele, R. Plemmons, and S. Ragni. (2004). Optimality, computation, and interpretation of nonnegative matrix factorizations. [Online]. Available: http://www4.ncsu.edu/~mtchu/Research/Papers/nnmf.pdf

[40] C. Hung and T. L. Markham, "The Moore-Penrose inverse of a partitioned matrix $M = \left(\begin{smallmatrix} A & D \\ B & C \end{smallmatrix}\right)$," *Linear Algebra Applica.*, vol. 11, no. 1, pp. 73–86, 1975.

[41] C. Hung and T. L. Markham, "The Moore-Penrose inverse of a sum of matrices," *J. Aust. Math. Soc., Ser. A*, vol. 24, no. 4, pp. 385–392, 1977.

[SP]