# Minimum-Volume Nonnegative Matrix Completion



Nicolas Gillis
Université de Mons

# Minimum-Volume Nonnegative Matrix Completion

Olivier Vu Thanh    Nicolas Gillis
University of Mons
Rue de Houdain 9, 7000 Belgium
{olivier.vuthanh, nicolas.gillis}@umons.ac.be

***Abstract*—Low-rank matrix approximation is a standard, yet powerful, embedding technique that can be used to tackle a broad range of problems, including the recovery of missing data. In this paper, we focus on the performance of nonnegative matrix factorization (NMF) with minimum-volume (MinVol) regularization on the task of nonnegative data imputation. The particular choice of the MinVol regularization is justified by its interesting identifiability property and by its link with the nuclear norm. We show experimentally that MinVol NMF is a relevant model for nonnegative data recovery, especially when the recovery of a unique embedding is desired. Additionally, we introduce a new version of MinVol NMF that exhibits some promising results.**

***Index Terms*—matrix completion, nonnegative matrix factorization, minimum-volume**

## I. Introduction

Given a data matrix $X \in \mathbb{R}^{m\times n}$, there exist many scenarios where only a few entries of $X$ are observed, e.g., in recommender systems illustrated by the famous Netflix problem [1]. Recovering these missing entries is often tackled by assuming that the fully observed data follow a certain structure. If the structuring assumption is meaningful, by fitting a model that follows the same structure on the observed entries, it is possible to recover the missing entries; see, e.g., [2]–[4]. The low-rank assumption is meaningful in many scenarios [5]. If $X \in \mathbb{R}^{m\times n}$ is low rank, we can express it as the product of two smaller matrices, $W \in \mathbb{R}^{m\times r}$ and $H \in \mathbb{R}^{r\times n}$, as $X = WH$ where $r \ll \min(m,n)$. Let us denote $\Omega \subseteq \{1,\dots,m\}\times\{1,\dots,n\}$ the set containing the indices of the observed entries in $X$. If the rank of $X$ is equal to $r$, we can look for $W \in \mathbb{R}^{m\times r}$ and $H \in \mathbb{R}^{r\times n}$ such that $X(i,j) = W(i,:)H(:,j)$ for all $(i,j) \in \Omega$. Then, for every missing entry at $(i,j) \in \overline{\Omega}$, $X(i,j)$ can be estimated by computing $W(i,:)H(:,j)$. If $X$ is noisy and does not follow the low-rank assumption, it might still be relevant to approximate it through a low-rank structure, because low-rank matrix approximations can identify patterns in the data via the extraction of common features among data points.

When the rank is unknown, a common tractable strategy is to minimize the nuclear norm, that is the sum of the singular values, of the estimation $\tilde{X}$ of $X$:

$$\min_{\tilde{X}} \|\tilde{X}\|_* \quad \text{such that} \quad \mathcal{P}_\Omega(\tilde{X}) = \mathcal{P}_\Omega(X),$$

where $\mathcal{P}_\Omega(Y)$ sets $Y(i,j)$ to zero if $(i,j) \notin \Omega$, or does not change it otherwise.

In this paper, we consider the rank to be known, and our goal is not only to recover the missing entries in $X$, but also to recover the unique matrices $W$ and $H$ that generated the data $X = WH$. This could be useful in hyperspectral unmixing with missing data for instance, where the columns of $W$ are expected to be the spectral signatures of the underlying materials, and where the $j$-th column of $H$ contains the abundance in the $j$-th pixel of each extracted material. In this scenario, it is of course preferable to recover a unique set $(W,H)$. To perform this task, it is possible to first use a data completion algorithm, and then use a constrained matrix factorization algorithm to estimate the sought factors $W$ and $H$. Here, we focus on performing both tasks together, since estimating correctly $W$ and $H$ on $\Omega$ implies a correct recovery of the missing entries in $X = WH$. We assume that the data and the factors are nonnegative, that is, $X \geq 0$, $W \geq 0$ and $H \geq 0$, where $\geq$ is applied element wise. In the fully observed case, estimating nonnegative $W$ and $H$ from $X$ is known as nonnegative matrix factorization (NMF) [6]. Hence, our goal is to perform NMF with missing data while recovering a unique decomposition. To do so, minimum-volume (MinVol) NMF is a relevant option, and its performances on matrix completion have never been explored before. In this paper, we show that when correctly tuned, MinVol NMF performs well on the matrix completion task and is also able to retrieve the true underlying factors using only a few observed entries.

The paper is organized as follows. In Section II, we introduce MinVol NMF and its identifiability properties. In Section III, we introduce the nonnegative matrix completion problem and discuss on the relevance of MinVol NMF for this problem. We also introduce a new variant of MinVol NMF. In Section IV, we describe the algorithms used in the experiment presented in Section V. Finally, we conclude and discuss future work in Section VI.

## II. Minimum-volume NMF (MinVol NMF)

### A. Existing declinations of MinVol NMF

Given a matrix $X \in \mathbb{R}^{m\times n}_+$ and a factorization rank $r$, in the exact case NMF consists in finding two smaller matrices $W \in \mathbb{R}^{m\times r}_+$ and $H \in \mathbb{R}^{r\times n}_+$ such that $X = WH$. Geometrically, this implies that $\text{cone}(X) \subseteq \text{cone}(W)$[1]. Distinctively, with MinVol NMF the cone of $W$ should enclose the cone of $X$ as

This work was supported by the Fonds de la Recherche Scientifique - FNRS (F.R.S.-FNRS) under PDR no T.0097.22 and under a FRIA Ph.D. grant for OVT, and by the European Union (ERC consolidator, eLinoR, no 101085607).

[1]Given $A \in \mathbb{R}^{m\times n}$, $\text{cone}(A) = \{y \mid y = Ax \text{ for } x \in \mathbb{R}^n_+\}$.

tightly as possible, hence the expression "minimum-volume". In other words, MinVol NMF consists in finding a couple of factors $(W, H) \in \mathbb{R}_+^{m\times r} \times \mathbb{R}_+^{r\times n}$ such that $X = WH$ while minimizing the volume of the convex hull of the columns of $W$ and the origin, which is given by $\frac{1}{r!}\sqrt{\det(W^\top W)}$. This improves the interpretability of the features (the columns of $W$) while prioritizing a unique decomposition of the data under relatively mild assumptions, that are given in Theorem 1. Additionally, one of the factors should be constrained such that the scaling ambiguity between $W$ and $H$ coupled with the minimized volume does not make $W$ tend to zero at optimality. Identifiable MinVol NMFs typically use simplex structuring constraints, namely $W \in \Delta^{m\times r}$ [7] or $H \in \Delta^{r\times n}$ [8] or $H^\top \in \Delta^{n\times r}$ [9], where $\Delta^{m\times r} = \{Y \in \mathbb{R}_+^{m\times r}, e^\top Y = e^\top\}$ and $e$ is the all-one vector of appropriate dimension. Among the three mentioned variants of MinVol, we will only consider the following formulation in the remainder of this paper:

$$\begin{aligned} \min_{W,H} \quad & \frac{1}{2}\|X - WH\|_F^2 + \frac{\lambda}{2}\operatorname{logdet}(W^\top W + \delta I) \\ \text{s.t.} \quad & W \in \Delta^{m\times r},\ H \in \mathbb{R}_+^{r\times n}, \end{aligned} \tag{1}$$

where $\delta$ is a parameter that prevents the logdet from going to $-\infty$ when $W$ is rank deficient, and $\lambda \geq 0$ balances the two terms. Note that the true volume spanned by the columns of $W$ and the origin is equal to $\frac{1}{r!}\sqrt{\det(W^\top W)}$, but minimizing $\operatorname{logdet}(W^\top W)$ is equivalent in the exact case and makes the problem numerically easier to solve because the function $\operatorname{logdet}(\cdot)$ is concave and it is easier to design a "nice" majorizer for it [10].

### *B. Identifiability of MinVol NMF*

We say that $(W, H)$ is an exact MinVol NMF of size $r$ of $X$ if it solves

$$\min_{W\in\Delta^{m\times r}, H\in\mathbb{R}_+^{r\times n}} \operatorname{logdet}(W^\top W) \text{ such that } X = WH.$$

**Definition 1** (Essential uniqueness of MinVol NMF)**.** *The exact MinVol NMF $(W, H)$ of $X = WH$ of size $r$ is unique if, and only if, for any other exact MinVol NMF $(\tilde{W}, \tilde{H})$ of $X = \tilde{W}\tilde{H}$ of size $r$, there exists a scaled permutation matrix $Q$ such that $\tilde{W} = WQ^{-1}$ and $\tilde{H} = QH$.*

Essential uniqueness is also referred as "identifiability". MinVol NMF with column wise simplex structured $H$ has first been proved to be identifiable in [8], under the so called sufficiently scattered conditions (SSC).

**Definition 2** (SSC)**.** *The matrix $H \in \mathbb{R}_+^{r\times n}$ is sufficiently scattered if the following two conditions are satisfied:*

*[SSC1]* $\mathcal{C} = \{x \in \mathbb{R}_+^r \mid e^\top x \geq \sqrt{r-1}\|x\|_2\} \subseteq \operatorname{cone}(H)$.

*[SSC2] There does not exist any orthogonal matrix $Q$ such that $\operatorname{cone}(H) \subseteq \operatorname{cone}(Q)$, except for permutation matrices.*

These conditions require that the aperture of $\operatorname{cone}(H)$ is "large enough" by making sure $\operatorname{cone}(H)$ contains the second-order cone $\mathcal{C}$ which is tangent to every facet of the nonnegative orthant (SSC1), and not to tightly (SSC2). Geometrical interpretations of SSC1 is provided in Fig. 1 and Fig. 2, where $\Delta^r = \{y \in \mathbb{R}_+^r, e^\top y = 1\}$.

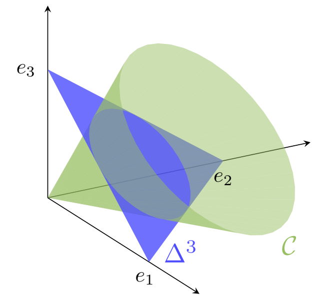


Fig. 1: Visualization of $\mathcal{C}$ with $r = 3$. $e_i$'s are the canonical vectors.

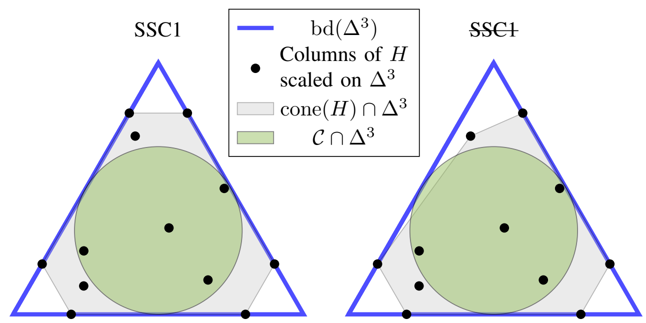


Fig. 2: Geometrical interpretation of SSC1 with $r = 3$.

Under the SSC, we have the following identifiability theorem.

**Theorem 1** ([7])**.** *If $W \in \mathbb{R}^{m\times r}$ is full column rank and $H \in \mathbb{R}^{r\times n}$ satisfies the SSC, then the exact MinVol NMF of size $r$ of $X = WH$ is essentially unique.*

Theorem 1 is fulfilled under relatively mild conditions, as the requirements are almost only on the factor $H$. The only assumption on $W$ is that it should be full rank.

## III. Nonnegative matrix completion (NMC)

In this section, we justify the choice of the minimum-volume criterion for the task of nonnegative matrix completion. Matrix completion in general has been well studied, especially by the compressed sensing community. Among the techniques to perform matrix completion, the low-rank approach often arises, because the low-rank structure has been observed to be quite powerful in this setting, as it is able to identify hidden (linear) features in data. However, minimizing the rank of the estimation matrix while guaranteeing the equality constraints on the set of observed entries is NP-hard in general. A good convex relaxation that promotes low-rank structures is the nuclear norm minimization; see [11]. This is coming from the fact that the rank is the $\ell_0$ norm of the vector of the singular values, while the nuclear norm is the $\ell_1$ norm of this vector. Still, this requires to store the whole estimation $\tilde{X}$ of $X$, and it also becomes harder to impose additional structuring constraints. When the rank is known, we can fully exploit the low-rank structure by working with the low-rank factors $W$ and $H$ instead. It is then easier to add some structuring constraints on $W$ and $H$. Also, this allows one to deal with larger problems. Since

$$\|X\|_* = \min_{X=WH} \frac{1}{2}\left(\|W\|_F^2 + \|H\|_F^2\right),$$

a good alternative to the nuclear norm regularization is then the regularizer $\frac{1}{2}\left(\|W\|_F^2 + \|H\|_F^2\right)$ [12]. If the rank is unknown, an overestimated rank coupled with a proper penalization of $\frac{1}{2}\left(\|W\|_F^2 + \|H\|_F^2\right)$ can yield state-of-the-art results. For example, in [13], a properly tuned matrix factorization model using the above regularizer can outperform deep neural networks on recommendation systems. In [14], they showed that the sightly different regularizer $\|W\|_* + \frac{1}{2}\|H\|_F^2$ yields better results than $\frac{1}{2}\left(\|W\|_F^2 + \|H\|_F^2\right)$, both with uniform or non-uniform samplings. Going back to our point of interest, it is interesting to observe that the MinVol regularizer provides more adaptability as a (non-convex) relaxation of the rank [15], since $\operatorname{logdet}(W^\top W + \delta I) = \sum_i \log(\sigma_i^2(W) + \delta)$. As it can

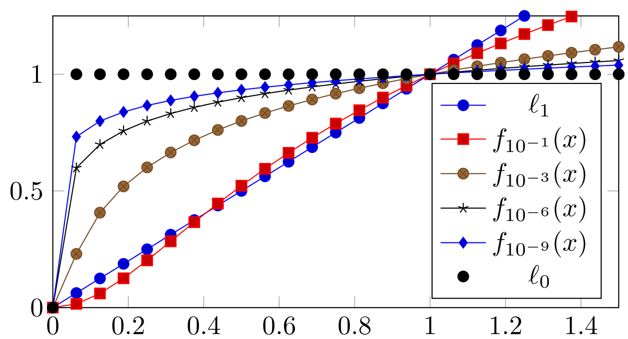


Fig. 3: Function $f_\delta(x) = \frac{\ln(x^2+\delta)-\ln(\delta)}{\ln(1+\delta)-\ln(\delta)}$ for various values of $\delta$, along the $\ell_0$ and $\ell_1$ norm.

be seen in Fig. 3, $\operatorname{logdet}(W^\top W + \delta I)$ approximates a range of behaviors between the $\ell_0$ and the $\ell_1$ norms. In particular, as $\delta$ goes to zero, $\operatorname{logdet}(W^\top W + \delta I)$ converges to the $\ell_0$ norm of the vector of singular values of $X$, up to a constant factor. Hence the MinVol criterion $\operatorname{logdet}(W^\top W + \delta I)$ is clearly a good candidate as a regularizer for NMC.

Let us now propose two models to tackle NMC. The first one is to adapt (1) to the NMC problem, which yields

$$\begin{aligned} \min_{W,H} \quad & \frac{1}{2}\|\mathcal{P}_\Omega(X - WH)\|_F^2 + \frac{\lambda}{2}\operatorname{logdet}(W^\top W + \delta I) \\ \text{s.t.} \quad & W \in \Delta^{m\times r},\ H \in \mathbb{R}_+^{r\times n}. \end{aligned} \tag{2}$$

Theorem 1 does not extend to the case where some values are missing. If the matrix completion is not unique, then it is impossible to guarantee a unique recovery of the matrices $W$ and $H$. Hence, a trivial way to adapt Theorem 1 to missing values is to add the condition that matrix completion under MinVol should be unique. However, conditions under which solving (2) recovers a unique completion are, up to now, unknown.

The second one introduces a new variant of MinVol NMF which is not simplex structured. Inspired by the regularizer $\|W\|_* + \frac{1}{2}\|H\|_F^2$ and motivated by the link between the behavior of the nuclear norm and the MinVol criterion, here we consider $\operatorname{logdet}(W^\top W + \delta I) + \|H\|_F^2$ as a regularizer. The resulting new MinVol NMF adapted for NMC is

$$\begin{aligned} \min_{W,H} & \frac{1}{2}\|\mathcal{P}_\Omega(X-WH)\|_F^2 + \frac{\lambda}{2}\operatorname{logdet}(W^\top W+\delta I) + \frac{\gamma}{2}\|H\|_F^2 \\ & \text{s.t. } W \in \mathbb{R}_+^{m\times r},\ H \in \mathbb{R}_+^{r\times n}. \end{aligned} \tag{3}$$

where $\lambda \geq 0$ and $\gamma \geq 0$ balance the regularizers. Note that neither $W$ nor $H$ is simplex structured. The scaling ambiguity coupled with the volume penalization is counter balanced by the penalization of $\|H\|_F^2$. In fact, in the exact case and when $\delta = 0$, every row of $H$ has the same norm at optimality. Consider a feasible $(W, H)$ for (3) such that $X = WH$ and let $f(D) = \frac{\lambda}{2}\operatorname{logdet}(D^{-1}W^\top W D^{-1}) + \frac{\gamma}{2}\|DH\|_F^2$ where $D = \operatorname{Diag}(d_1, \dots, d_r)$ is a positive diagonal matrix that can be seen as the scaling ambiguity between $W$ and $H$. Nullifying the gradient of $f$ relatively to each $d_i$, we have that $d_i^2 = \frac{\lambda}{\gamma\|H(i,:)\|_F^2}$, meaning that at optimality $\|H(i,:)\|_F^2 = \frac{\lambda}{\gamma}$ for all $i$.

## IV. Algorithms

In Section V, we compare NMF, MinVol (2) and new MinVol (3). For a fair comparison, these models are fit with the same algorithmic scheme, adapted from [16], which is an extrapolated alternating block majorization-minimization method. Our adaptation is described in Algorithm 1, where $\mathcal{P}_{\Delta^{m\times r}}$ (respectively $\mathcal{P}_{\mathbb{R}_+^{m\times r}}$) projects a matrix of size $m \times r$ onto $\Delta^{m\times r}$ (respectively $\mathbb{R}_+^{m\times r}$). See [17] for the details on the projection onto $\Delta^{m\times r}$. Essentially, the updates for $W$ and $H$ are several projected gradient descent steps, performed with a step size equal to the inverse of the Lipschitz constant. The updates for each model and each factor, as well as the corresponding Lipschitz constant, are given in Table I and Table II. The used Lipschitz constants are deliberately not tight. Consider the MinVol NMF update of $H$ for instance. Let $M \in \{0,1\}^{m\times n}$ be such that $M(i,j) = 1$ if $(i,j) \in \Omega$, $M(i,j) = 0$ otherwise. A tighter Lipschitz constant is $\max_j \left\|W^\top((M(:,j)e^\top) \circ W)\right\|_2$. We deliberately keep $\|W^\top W\|_2$ as it is less costly to compute and the additional cost might not be worth it. Moreover, if at least one column of $X$ is fully observed, then $\max_j \left\|W^\top((M(:,j)e^\top) \circ W)\right\|_2 = \|W^\top W\|_2 = \|W\|_2^2$.

## V. Experiments

The goal of this section is to highlight the performance of the MinVol criterion for NMC. All experiments are run with Julia on a PC with an Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz and 16GB RAM. All displayed measurements are averaged out of 20 runs. The code is available at https://gitlab.com/vuthanho/minvol-nmc. The compared models are NMF (to provide a baseline of a non-regularized model), MinVol (2), and the new proposed MinVol (3). For all models, the stopping criteria of the **while** loop in line 2 is just a number of outer iterations equal to 50, and the stopping criteria of the two **while** loops in lines 3 and 8 is a number of inner iterations equal to 20. All models are also initialized with the same

**Algorithm 1:** Main algorithm scheme

**input:** data matrix $X \in \mathbb{R}^{m \times n}$, initial factors $W \in \mathbb{R}_+^{m \times r}$ and $H \in \mathbb{R}_+^{r \times n}$

$\alpha_1 = \alpha_2 = 1,\ W_o = W,\ H_o = H$

**while** *stopping criteria not satisfied* **do**

  **while** *stopping criteria not satisfied* **do**

    $\alpha_0 = \alpha_1, \quad \alpha_1 = \frac{1}{2}(1 + \sqrt{1 + 4\alpha_0^2})$

    $\overline{W} = W + \frac{\alpha_0 - 1}{\alpha_1}(W - W_o)$

    $W_o = W$

    Update $W$ according to Table I

  **while** *stopping criteria not satisfied* **do**

    $\alpha_0 = \alpha_2, \quad \alpha_2 = \frac{1}{2}(1 + \sqrt{1 + 4\alpha_0^2})$

    $\overline{H} = H + \frac{\alpha_0 - 1}{\alpha_2}(H - H_o)$

    $H_o = H$

    Update $H$ according to Table II

| | Update |
|---|---|
| MinVol | $\mathcal{P}_{\Delta^{m \times r}}\left(\overline{W} - \frac{1}{L}\nabla_W\right)$ |
| new MinVol / NMF (with $\lambda = 0$) | $\mathcal{P}_{\mathbb{R}_+^{m \times r}}\left(\overline{W} - \frac{1}{L}\nabla_W\right)$ |

TABLE I: Updates for $W$ according to the model, where $P = (\overline{W}^\top \overline{W} + \delta I)^{-1}$, $L = \|HH^\top + \lambda P\|_2$ and $\nabla_W = \mathcal{P}_\Omega\left(\overline{W}H - X\right)H^\top + \lambda W P$.

| | $L$ | Update |
|---|---|---|
| MinVol / NMF | $\|W^\top W\|_2$ | $\mathcal{P}_{\mathbb{R}_+^{r \times n}}\left(\overline{H} - \frac{1}{L}\nabla_H\right)$ |
| new MinVol | $\|W^\top W + \gamma I\|_2$ | $\mathcal{P}_{\mathbb{R}_+^{r \times n}}\left(\frac{L-\gamma}{L}\overline{H} - \frac{1}{L}\nabla_H\right)$ |

TABLE II: Updates for $H$ according to the model, where $\nabla_H = W^\top \mathcal{P}_\Omega\left(W\overline{H} - X\right)$.

warm start $(W_0, H_0)$, which is the output of 500 iterations of NMF where the columns of $W$ are simplex-structured. In this setting, all methods converge. For both MinVols, $\lambda$ is first set to $\frac{\max(\|\mathcal{P}_\Omega(X - W_0H_0)\|_F^2, 10^{-6})}{|\operatorname{logdet}(W_0^\top W_0 + \delta I)|}$. For the new proposed MinVol, $\gamma$ is first set to $0.01\frac{\max(\|\mathcal{P}_\Omega(X - W_0H_0)\|_F^2, 10^{-6})}{\|H_0\|_F^2}$. On the hyperparameters $\lambda$ and $\gamma$, we adapt the automatic tuning method developed in [18]. The automatic tuning does not introduce a significant additional cost and is triggered when the difference between the current and the last objective values divided by $\|\mathcal{P}_\Omega(X)\|_F^2$ is less than $10^{-3}$.

*a) First experiment: noiseless synthetic data:* The first experiment focuses on both data completion and recovery of the exact generating factors in a noiseless case. For this experiment, for a given rank $r$, we randomly generate two factors $(W, H) = [0,1]^{200 \times r} \times [0,1]^{r \times 200}$ following a uniform distribution. Then, 80% random values of $H$ are set to zeros. This is a reasonable assumption in real scenarios such as hyperspectral unmixing. For the explored range of ranks, this will provide almost surely a sufficiently scattered $H$. Then, we generate the full data matrix $X$ simply by computing $WH$. The average of the elements of $X$ is always set to 1, dividing $X$ by its average. Finally, we create the observed data $\tilde{X}$ by removing a certain percentage of the entries in $X$. We vary the rank from 5 to 10, and the percentage of missing values from 80% to 90%. We report the root-mean-squared error (RMSE) of the missing values according to Def. 3 and the maximum subspace angle between the factor $W$ that took part in generating the data $X$ and its estimation $\tilde{W}$ according to Def. 4.

**Definition 3** (RMSE)**.** *The RMSE on the unobserved set $\overline{\Omega}$ is defined as follows*

$$\mathit{RMSE}(\tilde{X}, WH) = \sqrt{\frac{1}{|\overline{\Omega}|}\|\mathcal{P}_{\overline{\Omega}}(\tilde{X} - WH)\|_F^2}.$$

**Definition 4** (Subspace angle)**.** *Let $USV$ and $\tilde{U}\tilde{S}\tilde{V}$ respectively be the singular value decomposition of $W$ and $\tilde{W}$. Then the angle between the two subspaces specified by the columns of $W$ and $\tilde{W}$ is defined as follows*

$$\mathit{Angle}(W, \tilde{W}) = \arcsin(\min(1, \|\tilde{U} - UU^\top \tilde{U}\|_2)).$$

The RMSEs are reported in Fig. 4 and the subspace angles in Fig. 5. MinVol NMF coupled with the proposed auto-tuning proposed in [18] clearly outperforms the vanilla MinVol NMF with a fixed $\lambda$. The auto-tuned MinVol NMF is itself outperformed by our new proposed variant of MinVol NMF. For 90% missing values and a rank equal to 10 for instance, the average RMSE of the auto-tuned Minvol is 0.52 while it is 0.41 for the new MinVol.

*b) Second experiment: noisy synthetic data:* We keep the same settings as in the first experiment, while fixing the rank to 10, and adding some uniformly distributed noise. The noise level corresponds to the RMSE between the clean data and the noisy data. We vary the noise level from 0 to 1 and the percentage of missing values from 80% to 90%. We report the RMSE in Fig. 6. It is not necessary to report the subspace angle since it is degrading too fast. Perfect matrix completion is a necessary condition to retrieve a low subspace angle, which is already not possible starting from a noise level equal to 0.2. Results in Fig. 6 show that our proposed variant of MinVol NMF is more consistent relatively to the percentage of missing values and more precise than vanilla MinVol NMF in the presence of noise.

## VI. Conclusion and Discussion

This paper argued on the favor of using more the MinVol criterion in the domain of matrix completion, which has never been explored before. Not only the MinVol criterion can emulate a broad of behaviors going from the rank minimization to the nuclear minimization, but it also acts in favor of recovering the unique decomposition of a low-rank matrix if it exists. This paper also introduced a new variant of MinVol NMF which

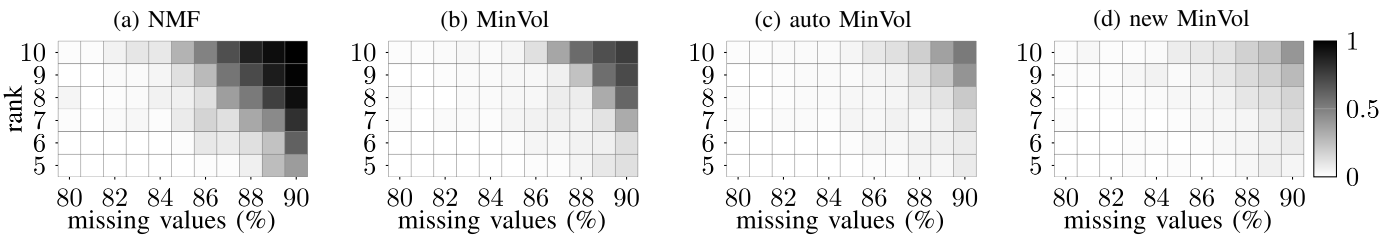


Fig. 4: Average RMSE according to the rank $r$ and to the percentage of missing values over 20 runs.

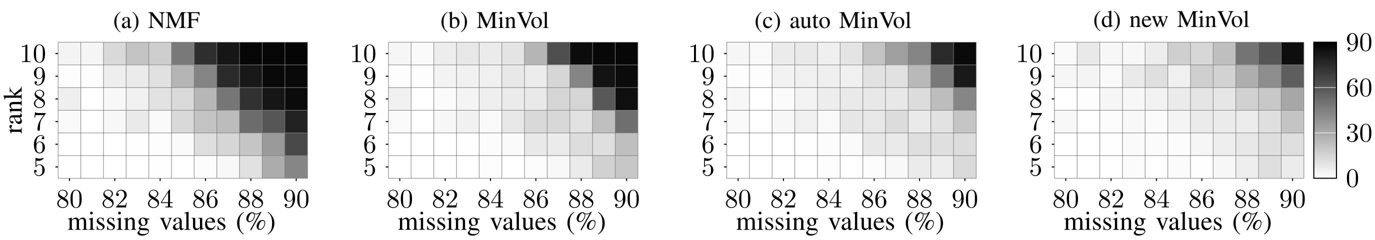


Fig. 5: Average angle according to the rank $r$ and to the percentage of missing values over 20 runs.

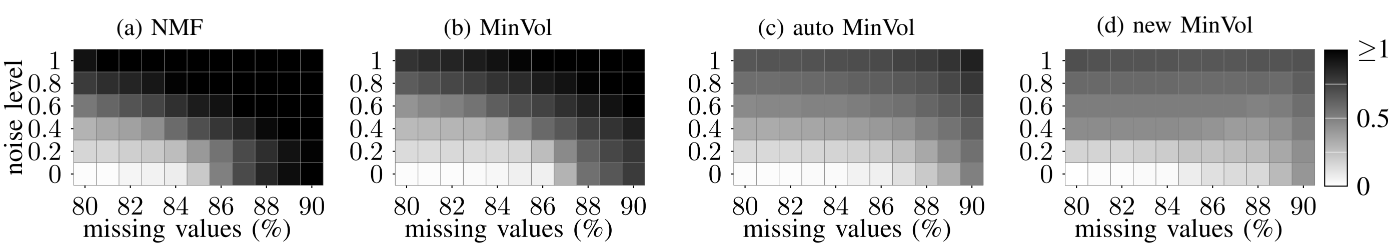


Fig. 6: Average RMSE according to the noise level and to the percentage of missing values over 20 runs.

is not simplex-structured. Experiments show that a properly tuned MinVol NMF provides encouraging results, both on the task of matrix completion and unique factors recovery. Last but not least, experiments show that our new proposed variant of MinVol NMF outperforms vanilla MinVol NMF. Future work will focus on the potential identifiability of this new variant and on comparing with other matrix completion algorithms.

## REFERENCES


[1] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.

[2] E. J. Candes and Y. Plan, "Matrix completion with noise," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 925–936, 2010.

[3] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1548–1566, 2011.

[4] E. Candes and B. Recht, "Exact matrix completion via convex optimization," *Communications of the ACM*, vol. 55, no. 6, pp. 111–119, 2012.

[5] M. Udell and A. Townsend, "Why are big data matrices approximately low rank?" *SIAM Journal on Mathematics of Data Science*, vol. 1, no. 1, pp. 144–160, 2019.

[6] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[7] V. Leplat, N. Gillis, and M. S. Ang, "Blind audio source separation with minimum-volume beta-divergence NMF," *IEEE Trans. Signal Process.*, vol. 68, pp. 3400–3410, 2020.

[8] X. Fu, W.-K. Ma, K. Huang, and N. D. Sidiropoulos, "Blind separation of quasi-stationary sources: Exploiting convex geometry in covariance domain," *IEEE Trans. Signal Process.*, vol. 63 (9), pp. 2306–2320, 2015.

[9] X. Fu, K. Huang, and N. D. Sidiropoulos, "On identifiability of non-negative matrix factorization," *IEEE Signal Processing Letters*, vol. 25, no. 3, pp. 328–332, 2018.

[10] X. Fu, K. Huang, B. Yang, W.-K. Ma, and N. D. Sidiropoulos, "Robust volume minimization-based matrix factorization for remote sensing and document clustering," *IEEE Trans. Signal Process.*, vol. 64 (23), pp. 6254–6268, 2016.

[11] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.

[12] N. Srebro, J. Rennie, and T. S. Jaakkola, "Maximum-margin matrix factorization," in *Advances in Neural Information Processing Systems (NIPS)*, 2005, pp. 1329–1336.

[13] S. Rendle, W. Krichene, L. Zhang, and Y. Koren, "Revisiting the performance of ials on item recommendation benchmarks," in *Proceedings of the 16th ACM Conference on Recommender Systems*, 2022, pp. 427–435.

[14] G. Liu, Q. Liu, and X. Yuan, "A new theory for matrix completion," in *Advances in Neural Information Processing Systems*, 2017.

[15] V. Leplat, A. M. Ang, and N. Gillis, "Minimum-volume rank-deficient nonnegative matrix factorizations," in *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP).* IEEE, 2019, pp. 3402–3406.

[16] O. Vu Thanh, A. Ang, N. Gillis *et al.*, "Inertial majorization-minimization algorithm for minimum-volume nmf," in *European Signal Processing Conference (EUSIPCO)*, 2021, pp. 1065–1069.

[17] L. Condat, "Fast projection onto the simplex and the l 1 ball," *Mathematical Programming*, vol. 158, no. 1-2, pp. 575–585, 2016.

[18] D. T. Nguyen and E. C. Chi, "Towards tuning-free minimum-volume nonnegative matrix factorization," *arXiv:2309.13733*, 2023.