AVOIDING SADDLE POINTS IN NON-NEGATIVE MATRIX FACTORIZATION VIA SIMPLEX CONSTRAINTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Provably avoiding saddle points in Non-negative Matrix Factorization (NMF) has been an open question since the influential work of Lin (2007a). While algorithms like Lee-Seung Lee & Seung (1999) and SNAP (Lu et al., 2020) empirically perform well, there are no guarantees on whether they avoid the many (suboptimal) saddle points present in the non-convex optimization landscape of NMF. Unfortunately, the extensive literature on escaping saddle points in non-convex settings does not apply to NMF, as NMF is defined over a *non-compact domain* (non-negative orthants) and the loss is non-Lipschitz (both gradient and Hessian are non-Lipschitz). This paper shows how to avoid saddle points in NMF, obtaining convergence to second-order stationary points (SOSPs). Specifically: a) *Asymptotic Convergence*: we show that a natural modification of Multiplicative Weights Updates (MWU) asymptotically converges to *exact* SOSPs of NMF, with probability 1 for any random initialization in the simplex; b) *Rates*: applying existing results from SNAP/SNAP$^+$ also yields non-asymptotic guarantees for approximate stationarity of Simplex-NMF. Our novelty lies in analyzing the landscape of a carefully modified NMF problem with one additional constraint, showing that this modification *preserves SOSPs of the original NMF landscape and their quality*.

1 INTRODUCTION

Non-negative Matrix Factorization (NMF) is a challenging case of a *constrained* non-convex optimization problem, where we are given a non-negative $n \times m$ matrix V that we want to represent as the product of two non-negative matrices W, H so as to approximately minimize the reconstruction error in Frobenius norm.

Formally, to set some notation, NMF asks for the following minimization:

$$\min_{W \in \mathbb{R}_+^{n \times r}, H \in \mathbb{R}_+^{r \times m}} F(W, H) := \|V - W \cdot H\|_F^2. \tag{NMF}$$

In contrast to Matrix Factorization which can be solved efficiently in polynomial time (e.g., by Singular Value Decomposition), it is well-known that NMF is NP-hard to optimize (Vavasis, 2009). As a result, it is impossible to find global minima, and so one typically settles for the milder goal of finding *local* optima.

However, finding local optima has been a challenging task by itself, due to the abundance of low-quality saddle points, leading to many algorithms with guaranteed convergence only to first-order stationary points (FOSPs), which include not only local minima, but also local maxima and saddles.

Unfortunately, for NMF there is currently no analysis showing how to provably avoid saddle points and converge to *Second-Order Stationary Points* (SOSPs), and obtaining such analyses has been posed as an open question in the influential works of Lin (2007b;a). In practice however, there are empirical successes with the early work of Lee & Seung (1999; 2001) and later with the SNAP algorithm by Lu, Razaviyayn, Yang, Huang, and Hong (2020); once again, it is not known whether such algorithms will converge to SOSPs of the NMF problem.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

1.1 Our Main Result: Provable Convergence to Second-Order Stationarity of NMF

Given the NP-hardness and related complexity results for recovering low-rank approximations for NMF (Moitra, 2013; Arora et al., 2016), usually one tries to run some first-order or second-order method (typically a variant of gradient descent or projected gradient descent) and find a sufficiently satisfactory decomposition. The main question that motivated our work is:

Can we escape saddle points in NMF? Are there simple iterative algorithms converging to SOSPs?

Literature on "Avoiding Saddles" does not apply for NMF: Observe that NMF is a case of non-smooth optimization, as both W and H should lie in the non-negative orthant ($W, H \geq 0$), which is an unbounded domain. Because the domain is unbounded, the loss function $\|V - W \cdot H\|_F^2$ is not Lipschitz, its gradient is not Lipschitz and its Hessian is also not Lipschitz. We emphasize this as prior work for optimizing non-convex functions and avoiding saddle points (Jin et al., 2017; 2021; Lu et al., 2020; Mokhtari et al., 2018; Khamaru & Wainwright, 2018; Sun & Fazel, 2018; Zhang et al., 2020; Reddi et al., 2018), operate under strong Lipschitzness and/or smoothness assumptions for the loss, the gradient and/or the Hessian, or they require the constraints to be equality ("$=$") rather than inequality ("$\geq$") constraints that are present in NMF. Hence, NMF does not fall into the existing frameworks. Notably, the SNAP algorithm of Lu et al. (2020) is experimentally demonstrated to perform well on NMF tasks, however the theory developed in that paper requires the loss function to be gradient-Lipschitz and Hessian-Lipschitz (Assumption 1, pg 3 of Lu et al. (2020)), which is not the case for NMF. Please see the Appendix A.2 for in-depth discussion about the related work on NMF.

Theoretical Results (Informal). This paper proves how to escape saddle points in NMF, obtaining for the first time, convergence to second-order stationary points (SOSPs) of NMF. We get:

- (Asymptotic Convergence) We show that a natural modification of Multiplicative Weights Updates (MWU) asymptotically converges to *exact* SOSPs of NMF, with probability 1 for any random initialization in the simplex. See Theorem 1, Theorem 3 for formal statements.
- (Rates of Convergence) We show that a natural modification of the previous SNAP algorithm from Lu et al. (2020) will converge to *approximate* SOSPs of simplex-NMF with the same non-asymptotic rates of convergence as those established in Lu et al. (2020). See Theorem 1, Theorem 3 for formal statements.

Briefly, our work shows that appropriately modified versions of two commonly-used approaches (MWU and SNAP) will provably converge to SOSPs for NMF. As we will see, our proposed modification is to **add one extra constraint** to the optimization (see Simplex-NMF below) and then show that **this will not essentially affect the landscape of NMF**. We emphasize that the main contribution of our work and the novelty of our result lies in analyzing the landscape of a carefully modified NMF problem under additional constraints on W, H, and showing that the landscape after the modification process preserves a two-way relationship between the quality of SOSPs of the original NMF landscape and the modified NMF landscape (after adding the constraint). Please see Theorem 1 for the formal statement of our structural result.

After we have established this structural theorem about the SOSPs (and their quality) of the original NMF and the modified NMF problem, only then can we apply the results of Lu et al. (2020) and Panageas et al. (2019) to obtain convergence to second-order stationarity. Indeed, without our structural theorem, we cannot know a priori that the modified NMF problem will be of any use (see our toy example in Section 3), as extra constraints may severely affect the optimization landscape. Here, we rely on heavily symmetric nature of NMF solutions: if (W, H) is a solution, then so is $(WQ, Q^{-1}H)$ with same loss, for any invertible diagonal matrix Q with non-negative entries, which allows to prove our structural Th. 1.

Experimental Verification. While our paper is theoretical, we perform experiments both on synthetic and real-world datasets (movie recommendations, MNIST digits, face recognition) and showcase that our modified version of SNAP, called nSNAP+ (from normalized SNAP+) performs well in practice while being more stable during training compared to Lee-Seung and SNAP algorithms (see Fig. 1 and Experiments section).

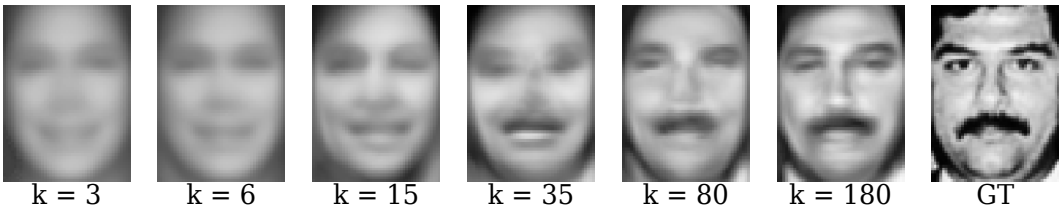


Figure 1: Example of our nSNAP+ NMF Face Reconstruction on 13,233 images from the Labeled Faces in the Wild (LfW) dataset Huang et al. (2007) with decompositions of different rank (k).

2 Main Idea: Add Extra Constraint and Analyze NMF Landscape

The domain in NMF is unbounded since matrices $W, H \geq 0$ (entrywise non-negative). This creates challenges as the loss function $\|V - W \cdot H\|_F^2$ is not Lipschitz, its gradient is not Lipschitz and its Hessian is also not Lipschitz, and prior works on avoiding saddle points do not apply.

How can we circumvent these challenges for Eq. (NMF)? A quick fix for circumventing them is an obvious one: *make the domain bounded by imposing upper bounds on matrix W and on H*, i.e., our main idea is to enforce the optimization to take place over a bounded domain, which leads to our slightly modified NMF formulation below (which we later show how we can conveniently use it to inform about the original NMF landscape):

New NMF Formulation. Concretely, we propose a new formulation which we call **Simplex-NMF** that we will be able to analyze w.r.t. to stationary points and convergence of known iterative methods:

$$\begin{aligned} \min\ & \|V - W \cdot H\|_F^2, \\ \text{s.t.}\ & \sum_{i,k} W_{ik} + \sum_{k,j} H_{kj} = C \\ & W \in \mathbb{R}_+^{n \times r}, H \in \mathbb{R}_+^{r \times m}, \end{aligned} \tag{Simplex-NMF}$$

We sometimes refer to $\sum_{i,k} W_{ik} + \sum_{k,j} H_{kj} = C \iff \sum_{i,k} \frac{W_{ik}}{C} + \sum_{k,j} \frac{H_{kj}}{C} = 1$ as the "normalization" or the "simplex" constraint, because it bounds the domain and forces the loss, its gradient and its Hessian to become Lipschitz. The parameter C is a constant that needs to be carefully chosen: obviously, setting it too low (e.g., $C = 0$) will result in both $W = 0$ and $H = 0$, thus incurring huge loss of $\|V\|_F^2$ for the original (NMF). Hence, a poorly chosen C can completely ruin the optimization landscape in a way that solutions to (Simplex-NMF) are extremely suboptimal for the original (NMF) (they could be saddles as here (0,0) is).

New Analysis: Addressing the Effects to NMF Landscape. The bulk of the proofs go into analyzing the properties of the landscape of (Simplex-NMF) versus the landscape of (NMF). As we want to prove convergence to SOSPs, we need to analyze how the constraint $\sum_{i,k} W_{ik} + \sum_{k,j} H_{kj} = C$ affects the landscape.

We show two things, both of which are necessary for (Simplex-NMF) and highlight its usefulness towards understanding the original NMF landscape.

First, even though new stationary points will be created, these are actually *strict* saddles (first-order stationary points that are not second-order), hence algorithms from prior works that converge to second-order stationarity can now be used, such as Nouiehed & Razaviyayn (2020); Lu et al. (2020); Panageas et al. (2019). However, it could still be the case that the simplex constraint destroyed SOSPs of the original formulation. In fact this does happen, because upper bounding the variables in W, H destroys *infinitely* many solutions, those where entries either in WQ or $Q^{-1}H$ are too large for the simplex constraint to be satisfied.

Luckily, the second thing **we prove is that for every stationary point of the original (NMF), there exists a stationary point in our formulation (Simplex-NMF), attaining the same value in objective**. In other words, the simplex constraint with *appropriate choice* of C has many benefits as

it makes NMF amenable to techniques from smooth non-convex optimization, and in addition, the stationary points found, are also SOSPs of (NMF) and of the same quality as (NMF).

Essentially, our work is a reduction stating that the problem of finding SOSPs with respect to (NMF) reduces into finding a SOSP of (Simplex-NMF). Our main structural result is stated next:

Theorem 1. *In Simplex-NMF, let the total sum of entries be bounded by* $\sum_{i,k} W_{ik} + \sum_{k,j} H_{kj} = C$.

1. ***Case 1:*** *In the case that the parameter* C *is chosen to be* $C > 2r(nm)^{1/4}\sqrt{\|V\|_F}$*, then we have:*
   - ***(Part I)*** *Any SOSP of* (Simplex-NMF) *is also a SOSP of* (NMF).
   - ***(Part II)*** *For every stationary point of* (NMF)*, there exists a stationary point of* (Simplex-NMF) *in which the objective attains the same value (with the exception of (0,0) which is a saddle). (See Motivating Example, Figure 2)*
2. ***Case 2:*** *We complement the above case, by showing that a bad choice of* C *(*$C \leq 2r(nm)^{1/4}\sqrt{\|V\|_F}$*) can result in the creation of new low-quality SOSPs in the landscape of* (Simplex-NMF)*, that were not SOSPs for the original landscape of* (NMF)*. (See Motivating Example, Figure 2)*

Remark 1. *For our initial goal of finding SOSPs of* (NMF) *by finding SOSPs of* (Simplex-NMF)*, the second part of Theorem 1 is not necessary. However it is necessary to argue about the quality of the factorization produced by our reduction: stationary points of* (NMF) *that are simultaneously stationary points of* (Simplex-NMF) *could correspond to higher values of the objective. For example, the optimal value of* (Simplex-NMF) *could be much greater than the optimal value of* (NMF)*, and in fact, our analysis indicates that this can happen when* $C < 2r(nm)^{1/4}\sqrt{\|V\|_F}$.

Remark 2 (On the Choice of Parameter C)**.** *We emphasize that our bound for* C *being larger than* $C_0 = 2r(nm)^{1/4}\sqrt{\|V\|_F}$ *is sufficient, and not necessarily sharp. For example, a concrete starting choice for experimental purposes would be* $C = 1.1C_0$*. As proved in our theorems, above the threshold, exact SOSPs are preserved (Part I in our Th. 1). This does not imply identical optimization trajectories, convergence speed, or downstream performance across choices of* C.

New NMF Convergence Results. As a consequence of our main structural result, we get two main corollaries, stated informally as follows:

1. **Limit points of MWU are SOSPs:** Using prior work of Panageas et al. (2019) as a black box, we conclude that the natural Multiplicative Weights Updates (MWU) dynamics for (Simplex-NMF) will asymptotically converge to exact SOSPs of (NMF), hence avoiding saddle points. This holds with probability 1 for any random initialization in the simplex.
2. **Rates of Convergence using SNAP, SNAP$^+$:** Prior work of Lu et al. (2020) defined the Successive Negative-curvature grAdient Projection (SNAP) algorithm and its first-order analogue SNAP$^+$ to obtain non-asymptotic convergence to SOSPs of problems with smooth non-convex objective and linear constraints. In light of our structural result Theorem 1, these algorithms become applicable, and we can find $(\epsilon, \sqrt{\epsilon})$-SOSPs after $O(\frac{1}{\epsilon^{2.5}})$ iterations for (Simplex-NMF).[1]

3 Motivating Example in 1-dimension: Simplex-NMF Informs NMF

To highlight our structural Theorem 1, it is instructive to look at a very simple toy NMF example where $n = m = r = 1$, shown in Figure 2:

$$\min f(x, y) := (1 - xy)^2, \text{ subject to } x \geq 0, y \geq 0 \qquad \text{(Ex. 1)}$$

[1] In Lu et al. (2020), even though they numerically evaluated the performance of SNAP for NMF, their theoretical results required Lipschitzness for the gradient and Hessian (Assumption 1, page 3 of their Arxiv version). Such assumptions do not always hold for NMF. In order to use Lu et al. (2020), we first calculate the Lipschitzness of the gradient and Hessian of the objective (Simplex-NMF) under the constraint $\sum_{i,k} W_{ik} + \sum_{k,j} H_{kj} = C$. Now Hessian is Lipschitz with constant $2C^2(r^2(n + nm + m) + rm^2n)$ (see Appendix).

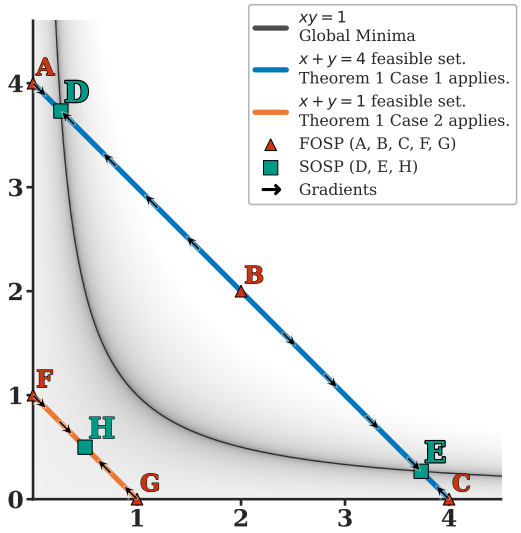


Figure 2: Toy example to illustrate Theorem 1, Case 1 and Case 2, demonstrating that the constant C needs to be appropriately chosen: with $C = 1$ (Case 2), points F, G are FOSPs and H is SOSP for (Simplex-NMF), and known algorithms Panageas et al. (2019); Lu et al. (2020) (initialized anywhere except F,G) will converge to F= $(\frac{1}{2}, \frac{1}{2})$ which is a very bad solution for the original (NMF). With $C = 4$ (Case 1), points A, B, C are FOSPs, while D, E are SOSPs for (Simplex-NMF) and are also SOSPs for (NMF) (Part I of Theorem 1) with the same loss (Part II of Theorem 1). The aforementioned algorithms (initialized anywhere except A, B, C) converge to either D or E. Notice the arrows indicating that dynamics are attracting near SOSPs (D, E, H) and repelling near FOSPs (A, B, C, F, G).

Observe that the heavily symmetric nature of NMF makes it an ill-posed problem in the sense that there is not a unique solution but rather a continuum of solutions. Note that if (W, H) is a solution, then so is $(WQ, Q^{-1}H)$ for any invertible diagonal matrix Q with non-negative entries ($WQ \cdot Q^{-1}H = W \cdot H$ so they obtain the same loss). In our toy example, observe that $x = 1, y = 1$ is a solution and so is $x = t, y = \frac{1}{t}$ for any positive constant t. In general, for more complicated examples, this continuum of solutions may create infinitely many saddle points which become an obstacle to finding local optima.

Let us see how our (Simplex-NMF) formulation and its landscape would look like for different values of the parameter C and importantly, how the (Simplex-NMF) landscape compares to the (NMF) landscape. To do so, we first pick $C = 1$ and then $C = 4$, and we compare the landscapes. We refer the reader to the Preliminaries for definitions of FOSPs or SOSPs, if necessary.

We examine 3 cases:

- Unconstrained Landscape: The set of FOSPs is the union of the sets: $\{(x, 0) : x \geq 0\}, \{(0, y) : y \geq 0\}, \{(x, y) : xy = 1, x \geq 0, y \geq 0\}$, while the set of SOSPs is just the set $\{(x, y) : xy = 1, x \geq 0, y \geq 0\}$. This simple example illustrates the interest in finding SOSPs: observe that the set of SOSPs coincides with the set of global minima (given by $xy = 1$), while FOSPs contain very bad solutions, including the saddle point at $(0, 0)$.
- Case $C = 1$: Here, $C < 2r(nm)^{1/4}\sqrt{\|V\|_F} \leq 2$ and thus, none of our results applies. The problem becomes $\min\{(1 - xy)^2 : x + y = 1, x \geq 0, y \geq 0\}$. The set of FOSPs are now $\{(0, 1), (1, 0), (\frac{1}{2}, \frac{1}{2})\}$, while SOSPs only contain $(\frac{1}{2}, \frac{1}{2})$.[2] We conclude that if we were to run MWU Panageas et al. (2019) or SNAP Lu et al. (2020) then the produced sequence of solutions would converge to $(\frac{1}{2}, \frac{1}{2})$, since this is the only SOSP. Unfortunately, this would be a very bad solution (of value $\frac{9}{16}$) for the initial optimization problem whose optimal value is actually 0. We remark that $(\frac{1}{2}, \frac{1}{2})$ is not a SOSP for the unconstrained

[2]The constraint $x + y = 1$ removes an infinite set of FOSPs and SOSPs compared to the unconstrained case.

case $\min\{(1-xy)^2 : x \geq 0, y \geq 0\}$, which was our initial algorithmic goal. This is not a surprise: C must be selected greater than 2 for the first part of our Theorem 1 to apply.

- Case $C = 4$: Here, $C > 2r(nm)^{1/4}\sqrt{\|V\|_F} \geq 2$ and thus, our landscape analysis applies. The problem becomes $\min\{(1-xy)^2 : x+y=4, x \geq 0, y \geq 0\}$. The set of FOSPs are now $\{(2,2), (2-\sqrt{3}, 2+\sqrt{3}), (2+\sqrt{3}, 2-\sqrt{3})\}$, while the set of SOSPs contains two points $\{(2-\sqrt{3}, 2+\sqrt{3}), (2+\sqrt{3}, 2-\sqrt{3})\}$. It is easy to check that both of these SOSPs are in fact global minima of the original example $\min\{(1-xy)^2 : x \geq 0, y \geq 0\}$. Also, since $(2-\sqrt{3}) \cdot (2+\sqrt{3}) = 1$, they also belong to the set of SOSPs of the original example. This is not a surprise: C was correctly selected greater than 2 and so both parts of our Theorem 1 apply.

From the above, we conclude that the problem of finding SOSPs for the original NMF instance reduces to finding SOSPs for the constrained instance with $C = 4$ (or just $C > 2$), as stated in Theorem 1. Moreover, algorithms of Panageas et al. (2019); Lu et al. (2020) become applicable and correctly find the global minima $(2-\sqrt{3}, 2+\sqrt{3})$ or $(2+\sqrt{3}, 2-\sqrt{3})$.

4 Comparing the Two NMF Landscapes: Proof of our Structural Theorem 1

In this section, we prove both parts of our main structural result (Theorem 1) and two of its algorithmic implications for MWU and SNAP based on prior works of Panageas et al. (2019); Lu et al. (2020) respectively.

Notation. W^i denotes the i-th column of matrix W and H_i denotes the i-th row of H. We denote by $\text{vec}(A)$ the standard vectorization of matrix A, by $[n]$ the set $\{1, 2, ..., n\}$ and by Δ_n the simplex of size n, that is $\Delta_n = \{x \in \mathbb{R}^n : x \geq 0, \sum_{i=1}^n x_i = 1\}$. Finally, we use $\|\cdot\|_F$ to denote the Frobenius norm of a matrix and $\|.\|_1$ to denote the ℓ_1 norm of a vector. Also, bold $\mathbf{0}$ denotes the all-zeros matrix with appropriate dimensions.

Stationary Points. There are two types of stationary points we are interested in: first-order stationary points, abbreviated as FOSPs, and second-order stationary points, abbreviated as SOSPs (we defer some definitions to the Appendix as writing down the KKT conditions becomes lengthy). We also use the term "saddle points" and "equivalence" throughout the paper.

Definition 4.1 (Saddle points)**.** *Saddle points of a problem are those FOSPs that are not SOSPs.*

Definition 4.2 (Equivalence)**.** *Let (W_1, H_1) be a FOSP of (NMF) and (W_2, H_2) be a FOSP of (Simplex-NMF). (W_1, H_1) and (W_2, H_2) are called* equivalent *if there exists a positive scalar $t > 0$ $W_2 = tW_1$ and $H_2 = \frac{1}{t}H_1$. Observe that $W_1H_1 = W_2H_2$ and hence $F(W_1, H_1) = F(W_2, H_2)$ (the two pairs of matrices give the same objective value). Moreover, since $t > 0$ is a scalar, (W_1, H_1) is SOSP for (NMF) if and only if (W_2, H_2) is SOSP for (Simplex-NMF).*

Finally, approximate stationary points, e.g., (ϵ_g, ϵ_H)-SOSPs, can also be defined, where inequalities just pick up an additive factor (as used in Lu et al. (2020)). We defer this to the Appendix.

4.1 Choice of Simplex Constraint Parameter C

As illustrated before, the parameter C needs to be carefully chosen. Here we explain why we have to pick $C > 2r(nm)^{1/4}\sqrt{\|V\|_F}$, in order to preserve interesting properties of the optimization landscape of (NMF), in particular the equivalence of FOSPs:

Lemma 4.3. *Let $(W^*, H^*) \neq (\mathbf{0}, \mathbf{0})$ be a FOSP for (NMF), and w.l.o.g. let's assume that $W^* \neq \mathbf{0}$. As long as $C > 2r(nm)^{1/4}\sqrt{\|V\|_F}$, there exists another pair of matrices $(\hat{W}, \hat{H})$ which is a FOSP for (Simplex-NMF) and is equivalent to (W^*, H^*).*[3]

[3]Though not needed here, one can easily extend this to show that every SOSP of (NMF) is also a SOSP of (Simplex-NMF). This holds since SOSPs are always FOSPs and (Simplex-NMF) has even more constraints than (NMF), and so they will be preserved up to equivalence.

The proof is given in Appendix A.4.2.

Remark 3. *To better understand the lemma, we instantiate it on our toy example with loss* $(1-xy)^2$ *where* $C=4$. *Let the initial (unconstrained) FOSP be* $(W^*,H^*)=(2,\frac{1}{2})$ *satisfying* $W^*\cdot H^*=1$, *but with total sum of entries* $\frac{5}{2}<4$. *The lemma first moves towards* $(\tilde{W},\tilde{H})=(1,1)$ *with same loss* $\tilde{W}\tilde{H}=1$, *but with minimum total sum (2 in this case), and then rescales this pair to obtain* $(t\tilde{W},\frac{1}{t}\tilde{H})=(t,\frac{1}{t})$ *such that* $t+\frac{1}{t}=4$. *The solution to this is* $(2-\sqrt{3},2+\sqrt{3})$ *or symmetrically* $(2+\sqrt{3},2-\sqrt{3})$ *which we know is FOSP for* (Simplex-NMF) *(in Figure 2, these correspond to points A,B). Moreover, note that (0,0) is no longer a FOSP since it violates the simplex constraint.*

4.2 SOSPs of Simplex-NMF are SOSPs of NMF

As we saw, adding the simplex constraint on W,H may create new stationary points. For our toy example with $C=4$, this indeed happens as point $(2,2)$ becomes stationary: $(2,2)$ is a FOSP for (Simplex-NMF), but not for (NMF). Luckily, it does not belong to the set of SOSPs of (Simplex-NMF) and this is not a coincidence as we show with the following theorem:

Theorem 2. *As long as* $C>2r(nm)^{\frac{1}{4}}\sqrt{\|V\|_F}$, *any SOSP of* (Simplex-NMF) *is also a SOSP of* (NMF).

As we will see later, for (Simplex-NMF), the known algorithms Panageas et al. (2019); Lu et al. (2020) converge to SOSPs. The role of Part I of Theorem 1, stated also as Theorem 2 above, is to guarantee that (Simplex-NMF) has no new SOSPs other than those that are already SOSPs for (NMF), and hence our algorithms will essentially find SOSPs for (NMF), even though they were run on a bounded domain. Once we prove this, the guarantees of Panageas et al. (2019); Lu et al. (2020) (that require smoothness) actually hold, as now the gradient and Hessian have become Lipschitz.

Proving Theorem 2 requires some intermediate lemmas and claims (omitted proofs are in Appendix):

Claim 4.4. *Let* (W^*,H^*) *be a FOSP for* (Simplex-NMF), *as per Def. A.3, for some constant* $c\neq 0$. *Then for all* $k\in\{1,\dots,r\}$, *we have*

$$\sum_{i=1}^{n}W^*_{ik}=\sum_{j=1}^{m}H^*_{kj}.$$

Lemma 4.5. *When* $C>2r(nm)^{\frac{1}{4}}\sqrt{\|V\|_F}$, *any SOSP of* (Simplex-NMF) *is also a FOSP of* (NMF)

Proof. We will equivalently show that as long as $C>2r(nm)^{\frac{1}{4}}\sqrt{\|V\|_F}$, any SOSP of (Simplex-NMF) satisfies the Def. A.3 with a constant $c=0$, and hence it is also a FOSP for (NMF), as per Def. A.1.

To arrive at a contradiction, let (W^*,H^*) be a SOSP of (Simplex-NMF) as per Def. A.4, with the constant $c\neq 0$.

Towards this, we carefully construct a pair of matrices (W^s,H^s) with $\sum_{i,k}W^s_{ik}+\sum_{k,j}H^s_{kj}=0$ satisfying:

$$\frac{\partial F(W^*,H^*)}{\partial W^*_{ik}}>c\implies W^s_{ik}=0,\text{ and also}\tag{1}$$

$$\frac{\partial F(W^*,H^*)}{\partial H^*_{kj}}>c\implies H^s_{kj}=0\tag{2}$$

and yet, Equation (4) is violated, i.e.,

$$s^\top\nabla^2F(W^*,H^*)s<0,$$

where we set $s^\top=(\text{vec}(W^s)^\top,\text{vec}(H^s)^\top)$ for ease of notation. This inequality contradicts with (W^*,H^*) being a SOSP.

We now construct the desired (W^s,H^s). Set $k:=\mathbf{argmax}_{1\leq\ell\leq r}\left(\sum_{i=1}^{n}W^*_{i\ell}+\sum_{j=1}^{m}H^*_{\ell j}\right)$.

Then, we set only the k^{th} column of W^s to coincide with the k^{th} column of W^* (i.e., $W^s_{ik} = W^*_{ik}$), and all other entries in W^s are set to zero (i.e., $W^s_{il} = 0$).

For H^s we set $H^s_{kj} = -H^*_{kj}$, and all its other entries to 0.

Using Claim 4.4, we get that the following sums of matrices cancel out:

$$\sum_{i,k} W^s_{ik} + \sum_{k,j} H^s_{kj} = \sum_{i=1}^{n} W^*_{ik} - \sum_{j=1}^{m} H^*_{kj} = 0$$

so it is only left to verify Eq. (1) and that $s^\top \nabla^2 F(W^*, H^*)s < 0$. This requires several calculations and is deferred to the Appendix. □

To finish off the proof of Theorem 2, we again start with a SOSP of (Simplex-NMF) and assume for contradiction that it is not a SOSP of (NMF). Testing the KKT conditions in this case however requires a more elaborate construction for (W^s, H^s) that we defer to the Appendix.

5 Putting it all together: SOSPs Preserved, Saddles Avoided

To put everything together and summarize our approach, our Theorem 2 proved above corresponds to Part I from Theorem 1, whereas the Lemma 4.3 is Part II from Theorem 1. Part I guarantees that even after making the transformation from the original NMF problem (NMF) to the modified version (Simplex-NMF) by adding constraints, any SOSP of (Simplex-NMF) is also a SOSP of (NMF). This means that our algorithm that is guaranteed to converge (asymptotically) to a SOSP of (Simplex-NMF) will indeed find a SOSP of (NMF) (or converge approximately to a SOSP of (Simplex-NMF)). The next part is about quality: Part II of from Theorem 1 guarantees that for every stationary point of (NMF) (except the saddle point at the origin which is uninteresting), there exists a stationary point of (Simplex-NMF) in which the objective attains the same value.

As a consequence, *the proposed transformation will not affect the value of the objective at the SOSPs found*. In light of this discussion, we can view our result as a reduction that guarantees the following:

Theorem 3. *Finding SOSPs for the original (NMF), whose domain is unbounded and gradient/Hessian is non-Lipschitz, reduces into finding SOSPs for our formulation (Simplex-NMF) which is over a bounded domain and has Lipschitz gradient/Hessian.*[4] *The work of Lu et al. (2020) gives non-asymptotic convergence rates to approximate SOSPs (specifically SNAP converges to $(\epsilon, \sqrt{\epsilon})$-SOSPs after $O(\frac{1}{\epsilon^{2.5}})$ iterations) and Panageas et al. (2019) give asymptotic convergence (for MWU as described in the supplement).*

6 Experiments

For our numerical experiments, we implement a version of SNAP+ Lu et al. (2020) which we call *Normalized SNAP+* (nSNAP+). nSNAP+ adds a normalization condition such that the target matrix is $\tilde{V} = \frac{V}{C}$ where $C = 2r(nm)^{1/4}\sqrt{\|V\|_F}$ and ensures that the solution always satisfies $\sum_{i,k} W_{ik} + \sum_{k,j} H_{kj} \leq 1$. For random initializations, the initial W and H are scaled so $\sum_i WH = \sum_i \tilde{V}$. For results in Fig. 3, we automatically select the learning rate based on a few trial steps of descent. During optimization, we apply backtracking if the descent step increases the norm by $\geq 1\%$.

We construct a sample NMF problem by sampling a random 12 x 12 matrix from $U(0, 1)$, performing its SVD, and reconstructing the matrix with all singular values equal to 1, and clipping negative values. We report the results of performing $rank(NM) = k = 3$ reconstructions in Fig. 3. On this problem, we verified that nSNAP+ results converged to points which are not saddle points.

Due to space constraints, please see Appendix for more experiments and figures.

[4] In our case, the Lipschitz constant for the Hessian can be upper-bounded by K, where $K = 2C^2(r^2(n + nm + m) + rm^2n)$. See Appendix for proof.

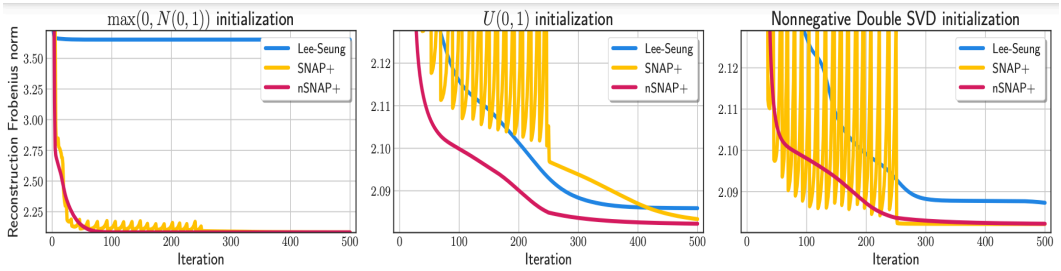


Figure 3: **Convergence behavior for a sample NMF problem.** We show results with three different initializations: sampled from a truncated Gaussian distribution, sampled from a uniform normal distribution, and the popular NNDSVD (Boutsidis & Gallopoulos, 2008) often used in practice. Results are shown for Lee-Seung (Lee & Seung, 1999), SNAP+ (Lu et al., 2020) and our normalized SNAP+. Lee-Seung can get stuck in bad initializations, such those with many 0s. SNAP+ works well, but can often oscillate at high learning rates. nSNAP+ converges to good results, and the normalization can provide smoother convergence.

Conclusion

Recent progress on escaping saddle points in non-convex optimization relies on strong smoothness assumptions that do not always hold. Here we study a popular such problem, Non-negative Matrix Factorization, and we propose a small modification to the constraint set, by adding an appropriate simplex constraints, that renders the problem smooth and amenable to results from prior works. While, a priori, such modification may severely alter the optimization landscape of the problem, through a series of steps that are particular to NMF, we establish that the new landscape is indeed similar (in terms of SOSPs) to the original NMF landscape, and amenable to convergence results. In this way, we obtain, to the best of our knowledge, the first analysis showing convergence to second-order stationary points of the original NMF problem. This extends a long list of works for NMF studying (first-order) convergence properties of iterative algorithms (Lee & Seung, 1999; 2001; Lin, 2007b;a; Gonzalez & Zhang, 2005) and specifically addresses a question of Lin (2007a) on how an upper bound on the NMF variables would affect optimization. We believe such modifications and analyses could find more applications to other problems in the future.

AI Use Statement

Generative AI (ChatGPT) was used to assist with converting the existing LaTeX source from overleaf to the current ICLR 2027 template, and for checking formatting and compilation.

Generative AI (ChatGPT) was used to assist with the calculation of the Hessian Lipschitz bound provided in Appendix A.4.1. Also for polishing the phrasing of Remark 2. All claims in this paper were checked by the authors, and we take full responsibility.

References

S.I. Amari A. Cichocki, R. Zdunek. Hierarchical als algorithms for nonnegative matrix and 3d tensor factorization. In *Lecture Notes in Computer Science*, pp. 169–176, 2007.

Sanjeev Arora, Rong Ge, Ravi Kannan, and Ankur Moitra. Computing a nonnegative matrix factorization—provably. *SIAM Journal on Computing*, 45(4):1582–1611, 2016.

Johan Bjorck, Anmol Kabra, Kilian Q Weinberger, and Carla Gomes. Characterizing the loss landscape in non-negative matrix factorization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 6768–6776, 2021.

C. Boutsidis and E. Gallopoulos. Svd based initialization: A head start for nonnegative matrix factorization. *Pattern Recognition*, 41(4):1350–1362, 2008. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2007.09.010.

Yann N Dauphin, Razvan Pascanu, Caglar Gulcehre, Kyunghyun Cho, Surya Ganguli, and Yoshua Bengio. Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. *Advances in neural information processing systems*, 27, 2014.

C. Févotte, N. Bertin, and J.L. Durrieu. Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis. *Neural Computation*, 21(3):793–830, 2009.

Yuan Gao and George Church. Improving molecular cancer class discovery through sparse non-negative matrix factorization. *Bioinformatics*, 21(21):3970–3975, 2005.

Rong Ge, Furong Huang, Chi Jin, and Yang Yuan. Escaping from saddle points - online stochastic gradient for tensor decomposition. In *Proceedings of The 28th Conference on Learning Theory, COLT 2015, Paris, France, July 3-6, 2015*, pp. 797–842, 2015.

Rong Ge, Jason D Lee, and Tengyu Ma. Matrix completion has no spurious local minimum. *Advances in neural information processing systems*, 29, 2016.

N. Gillis. The why and how of nonnegative matrix factorization". In M. Signoretto J.A.K. Suykens and A. Argyriou (eds.), *Regularization, Optimization, Kernels, and Support Vector Machines*. Chapman & Hall/CRC, 2014.

Edward F. Gonzalez and Yin Zhang. Accelerating the lee-seung algorithm for nonnegative matrix factorization, 2005.

F. Maxwell Harper and Joseph A. Konstan. The movielens datasets: History and context. *ACM Trans. Interact. Intell. Syst.*, 5(4), dec 2015. ISSN 2160-6455. URL https://doi.org/10.1145/2827872.

N.D. Ho. *Nonnegative matrix factorization algorithms and applications*. PhD thesis, Univ. Catholique de Louvain, 2008.

Ngoc-Diep Ho, Paul Van Dooren, and Vincent D Blondel. Descent methods for nonnegative matrix factorization. In *Numerical Linear Algebra in Signals, Systems and Control*, pp. 251–293. Springer, 2011.

Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

Chi Jin, Rong Ge, Praneeth Netrapalli, Sham M Kakade, and Michael I Jordan. How to escape saddle points efficiently. In *International Conference on Machine Learning*, pp. 1724–1732. PMLR, 2017.

Chi Jin, Praneeth Netrapalli, Rong Ge, Sham M Kakade, and Michael I Jordan. On nonconvex optimization for machine learning: Gradients, stochasticity, and saddle points. *Journal of the ACM (JACM)*, 68(2):1–29, 2021.

Koulik Khamaru and Martin Wainwright. Convergence guarantees for a class of non-convex and non-smooth optimization problems. In *International Conference on Machine Learning*, pp. 2601–2610. PMLR, 2018.

H. Kim and H. Park. Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method. *SIAM Journal on Matrix Analysis and Applications*, 30(2): 713–730, 2008.

Jingu Kim, Yunlong He, and Haesun Park. Algorithms for nonnegative matrix and tensor factorizations: a unified view based on block coordinate descent framework. *J. Global. Optim.*, 58:285–319, 2014.

Yann LeCun. The mnist database of handwritten digits. *http://yann. lecun. com/exdb/mnist/*, 1998.

Daniel D. Lee and H. Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.

Daniel D. Lee and H. Sebastian Seung. Algorithms for non-negative matrix factorization. In *In NIPS*, pp. 556–562. MIT Press, 2001.

L. Li and Y.-J Zhang. Fastnmf: highly efficient monotonic fixed-point nonnegative matrix factorization algorithm with good applicability. *J. Electron. Imaging*, 18:033004, 2009.

Chih-Jen Lin. Projected gradient methods for nonnegative matrix factorization. *Neural computation*, 19(10):2756–2779, 2007a.

Chih-Jen Lin. On the convergence of multiplicative update algorithms for nonnegative matrix factorization. *IEEE Transactions on Neural Networks*, 18(6):1589–1596, 2007b.

Songtao Lu, Meisam Razaviyayn, Bo Yang, Kejun Huang, and Mingyi Hong. Finding second-order stationary points efficiently in smooth nonconvex linearly constrained optimization problems. *Advances in Neural Information Processing Systems*, 33:2811–2822, 2020.

A. Moitra. An almost optimal algorithm for computing nonnegative rank. In *Proc. of the 24th Annual ACM-SIAM Symp. on Discrete Algorithms*, pp. 1454–1464, 2013.

Aryan Mokhtari, Asuman Ozdaglar, and Ali Jadbabaie. Escaping saddle points in constrained optimization. *Advances in Neural Information Processing Systems*, 31, 2018.

Maher Nouiehed and Meisam Razaviyayn. A trust region method for finding second-order stationarity in linearly constrained nonconvex optimization. *SIAM Journal on Optimization*, 30(3):2501–2529, 2020.

Maher Nouiehed, Jason D. Lee, and Meisam Razaviyayn. Convergence to second-order stationarity for constrained non-convex optimization. *CoRR*, 2019.

Pentti Paatero and Unto Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5(2):111–126, 1994.

Ioannis Panageas, Georgios Piliouras, and Xiao Wang. Multiplicative weights updates as a distributed constrained optimization algorithm: Convergence to second-order stationary points almost always. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, pp. 4961–4969, 2019.

Sashank Reddi, Manzil Zaheer, Suvrit Sra, Barnabas Poczos, Francis Bach, Ruslan Salakhutdinov, and Alex Smola. A generic approach for escaping saddle points. In *International conference on artificial intelligence and statistics*, pp. 1233–1242. PMLR, 2018.

Mikkel N. Schmidt and Rasmus Kongsgaard Olsson. Single-channel speech separation using sparse non-negative matrix factorization. In *Ninth International Conference on Spoken Language Processing*, 2006.

Yue Sun and Maryam Fazel. Escaping saddle points efficiently in equality-constrained optimization problems. In *Workshop on Modern Trends in Nonconvex Optimization for Machine Learning, International Conference on Machine Learning*, 2018.

S. Vavasis. On the complexity of nonnegative matrix factorization. *SIAM J. Optimization*, 20: 1364–1377, 2009.

Wei Xu, Xin Liu, and Yihong Gong. Document clustering based on non-negative matrix factorization. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pp. 267–273, 2003.

M. Yannakakis. Expressing combinatorial optimization problems by linear programs. *J. Comput. Syst. Sci.*, 43(3):441–466, 1991.

Jingzhao Zhang, Hongzhou Lin, Stefanie Jegelka, Ali Jadbabaie, and Suvrit Sra. Complexity of finding stationary points of nonsmooth nonconvex functions. *arXiv preprint arXiv:2002.04130*, 2020.

A APPENDIX

Organization of Appendix. We first present some definitions that we omitted in the Preliminaries section regarding FOSPs and SOSPs of the NMF and the Simplex-NMF problem. Then, we have an in-depth comparison to Related Work, followed by the necessary definitions of Approximate Stationarity by Lu et al. (2020). The remainder is devoted to omitted proofs, restating the MWU algorithm and giving some more examples/experiments. Finally, we also present some further related work.

A.1 DEFINITIONS FOR FOSPS AND SOSPS

Definition A.1 (FOSPs of (NMF))**.** *A pair of matrices (W^*, H^*) is a **first-order stationary point (FOSP)** of problem (NMF) if for all $i \in [n], k \in [r], j \in [m]$ it satisfies the first-order Karush–Kuhn–Tucker conditions presented next. First, $W^*_{ik}, H^*_{kj} \geq 0$ (non-negativity) and also:*

- *Whenever an entry of W or H is positive, then the corresponding gradient entry needs to be 0, i.e., $W^*_{ik} > 0 \implies \frac{\partial F(W^*,H^*)}{\partial W_{ik}} = 0$, and $H^*_{kj} > 0 \implies \frac{\partial F(W^*,H^*)}{\partial H_{kj}} = 0$.*
- *Whenever an entry of W or H is 0, then the corresponding gradient entry should be non-negative, i.e., $W^*_{ik} = 0 \implies \frac{\partial F(W^*,H^*)}{\partial W_{ik}} \geq 0$, and $H^*_{kj} = 0 \implies \frac{\partial F(W^*,H^*)}{\partial H_{kj}} \geq 0$.*

Definition A.2 (SOSPs of (NMF))**.** *A pair of matrices (W^*, H^*) is a **second-order stationary point (SOSP)** of problem (NMF) if it is a FOSP, and moreover the quadratic form of the Hessian:*

$$(vec(W)^\top, vec(H)^\top)\nabla^2 F(W^*, H^*) \begin{pmatrix} vec(W) \\ vec(H) \end{pmatrix} \geq 0, \quad \forall \text{ pairs } (W, H) \text{ satisfying the following:} \tag{3}$$

- *Whenever $\frac{\partial F(W^*,H^*)}{\partial W_{ik}} > 0$ then the entry $W_{ik} = 0$, and whenever $\frac{\partial F(W^*,H^*)}{\partial H_{kj}} > 0$ then the entry $H_{kj} = 0$, for all $i \in [n], k \in [r], j \in [m]$.*
- *Whenever the entry $W^*_{ik} = 0$ then then entry $W_{ik} \geq 0$, and whenever the entry $H^*_{kj} = 0$ then the entry $H_{kj} \geq 0$, for all $i \in [n], k \in [r], j \in [m]$.*

Definition A.3 (FOSPs of (Simplex-NMF))**.** *A pair of matrices (W^*, H^*) is a **first-order stationary point (FOSP)** of problem (Simplex-NMF) if for all $i \in [n], k \in [r], j \in [m]$ it satisfies the first-order Karush–Kuhn–Tucker conditions presented next. First, $W^*_{ik}, H^*_{kj} \geq 0$ (non-negativity), $\sum_{i,k} W^*_{ik} + \sum_{k,j} H^*_{kj} = C$ (normalization constraint) and additionally there exists some constant c such that:*

- *Whenever an entry of W or H is positive, then the corresponding gradient entry needs to equal c, i.e., $W^*_{ik} > 0 \implies \frac{\partial F(W^*,H^*)}{\partial W_{ik}} = c$, and $H^*_{kj} > 0 \implies \frac{\partial F(W^*,H^*)}{\partial H_{kj}} = c$.*
- *Whenever an entry of W or H is 0, then the corresponding gradient entry should be at least c, i.e., $W^*_{ik} = 0 \implies \frac{\partial F(W^*,H^*)}{\partial W_{ik}} \geq c$, and $H^*_{kj} = 0 \implies \frac{\partial F(W^*,H^*)}{\partial H_{kj}} \geq c$.*

Definition A.4 (SOSPs of (Simplex-NMF))**.** *A pair of matrices (W^*, H^*) is a **second-order stationary point** of the problem (Simplex-NMF) if it is a FOSP, and moreover the quadratic form of the Hessian:*

$$(vec(W)^\top, vec(H)^\top)\nabla^2 F(W^*, H^*) \begin{pmatrix} vec(W) \\ vec(H) \end{pmatrix} \geq 0, \quad \forall \text{ pairs } (W, H) \text{ satisfying the following:} \tag{4}$$

- *$\sum_{i=1}^n \sum_{k=1}^r W_{ik} + \sum_{j=1}^m \sum_{k=1}^r H_{kj} = 0$.*
- *Whenever $\frac{\partial F(W^*,H^*)}{\partial W_{ik}} > c$ then the entry $W_{ik} = 0$, and whenever $\frac{\partial F(W^*,H^*)}{\partial H_{kj}} > c$ then the entry $H_{kj} = 0$, for all $i \in [n], k \in [r], j \in [m]$.*
- *Whenever the entry $W^*_{ik} = 0$ then then entry $W_{ik} \geq 0$, and whenever the entry $H^*_{kj} = 0$ then the entry $H_{kj} \geq 0$, for all $i \in [n], k \in [r], j \in [m]$.*

where c is same constant used in the FOSP definition above.

A.2 Comparison to Related Work

There is a vast literature on NMF, so here we try to cover and compare the works that are mostly related to our results.

The NMF problem was introduced in the engineering community in the seminal paper of Paatero & Tapper (1994) and was subsequently popularized further in Lee & Seung (1999; 2001). Approximate factorizations based on non-negative matrices have found numerous applications throughout many fields of science and engineering with notable examples including music analysis Févotte et al. (2009), document clustering Xu et al. (2003), speech-source separation Schmidt & Olsson (2006) and cancer-class identification Gao & Church (2005). NMF is useful as it leads to additive and sparse representations of the input data, while it is a tool to describe the efficacy of linear programming for hard combinatorial problems Yannakakis (1991).

NMF's wide range of applications led to an intense interest towards the design of efficient algorithms to tackle it. Unfortunately even the simpler decision version of NMF, asking to decide whether equation (NMF) has value 0, is NP-hard Vavasis (2009). Going beyond the decision version of NMF, the last years, several *iterative gradient-based methods* have been proposed for the optimization version (NMF). Arguably one of the most popular approaches is the multiplicative update (MU) rule introduced in Lee & Seung (2001).

In this MU framework, the updates are given by

$$W_{t+1} = W_t \circ \frac{V H_t^\top}{W_t H_t H_t^\top} \quad H_{t+1} = H_t \circ \frac{W_{t+1}^\top V}{W_{t+1}^\top W_{t+1} H_t},$$

where $X \circ Y, X/Y$ denotes the componentwise multiplication, division of two matrices respectively. Please note that to compute the entries of H_{t+1}, the entries of W_{t+1} are needed (see the iteration subscripts), so the updates take place in an alternating manner (not concurrently). This sequential nature of the above algorithm is in contrast to the concurrent (parallel) nature of the proposed method in Algorithm 1.

In Lee & Seung (2001) it was shown that under the MU rule, NMF' s objective function does not increase. More recent approaches establish this non-increasing property by applying the *block coordinate descent framework*. A typical example of this kind of methods is the Alternating Non-negative Least Squares (ANLS) method Kim & Park (2008), in which W and H are sequentially updated to the minimizers of the convex programs, $\min_{W \geq 0} \|V - WH\|^2$, and $\min_{H \geq 0} \|V - WH\|^2$. Another example is the Hierarchical ALS (HALS) method A. Cichocki (2007), which has been rediscovered several times, e.g., as the rank-one residue iteration Ho (2008) and FastNMF in Li & Zhang (2009). For additional details concerning iterative methods for NMF and their convergence properties the reader is referred to the surveys Kim et al. (2014); Gillis (2014); Lin (2007a;b) and references therein.

In particular, in Lin (2007a;b), the author describes that it is unclear how to prove convergence to first-order stationarity of the algorithm of Lee & Seung (2001), even if we add some upper bound constraints akin to our normalization constraints. Lin (2007a;b) also describes a variation of the original algorithm of Lee & Seung (2001) (where some small constants are introduced in the denominators to avoid division by 0) that does converge to first-order stationarity. However, none of the proposed methods so far is known to converge to second-order stationary points, as we proved in this work.

Despite the fact that these methods are widely used in practice mostly due to the non-increasing property of the objective, it is possible for them to converge to saddle point solutions. This is because the fixed points of the above presented methods is either not known to converge at all or is only known to converge to the set of *first-order stationary points*, which contains highly sub-optimal solutions corresponding to *saddle points* of (NMF). Saddle points for (NMF) are pair of matrices satisfying the first-order KKT conditions, but not the second-order KKT (3). As Example 1 shows, solutions of extremely bad quality can satisfy the above requirements.

Converging to saddle points is a well-known drawback of many gradient-based approaches applied to important optimization problems (far beyond the scope of NMF) such as training of neural networks, matrix completion, community detection Dauphin et al. (2014). To overcome this difficulty a recent line of research examines methods converging to solutions satisfying the second-order KKT

conditions (second-order stationary points (SOSP)), which in most problems of practical interest correspond to high-quality solutions Jin et al. (2017); Ge et al. (2016; 2015). Our work can be seen as one more step towards dealing with saddle points, in this case for the particular problem of NMF optimization.

We remark that in Example 1 the set of second-order stationary points, not only excludes the *all-zero solution* but also has the property that it contains only the *global minima*.

Despite the aforementioned line of research, the above question does not admit an easy answer. Although simple methods such as gradient descent avoid saddle points in the unconstrained setting, there is no such first-order method for general constrained convex sets and in particular for the NMF problem (NMF). On top of that, Nouiehed et al. (2019) provides an example in which projected gradient descent converges to saddle points (does not avoid them), even under simple polytope (linear) constraints. Recently, Bjorck et al. (2021) motivated by the fact that gradient-descent-based methods often find good solutions, they studied the NMF loss landscape in non-worst-case settings. Specifically, in a non-worst case model with a planted solution, they showed that the property of star-convexity (this property implies that the gradients will point towards the final minimizer of the loss) holds with high probability; as this result is non-worst case and assumes a planted solution, it is incomparable to ours.

A.3 Omitted Definitions for (Approximate) Stationarity by Lu et al. (2020)

The work of Lu et al. (2020) defined and used the notion of approximate SOSPs for a similar context where the domain is bounded (instead of unbounded like in NMF). However, after we introduce our simplex (or normalization) constraints, the domain becomes bounded and their definitions and convergence results hold.

In particular, for some given small constants $\epsilon_g > 0$ and $\epsilon_H > 0$, they proposed the SNAP algorithm, which computes some approximate (ϵ_g, ϵ_H)-SOSP with $O(\max(1/\epsilon_g^2, 1/\epsilon_H^3))$ iterations, and with polynomial computational complexity in dimensions as long as projection, gradient, and Hessian can be computed efficiently.

Concretely, they consider the following class of non-convex linearly constrained optimization problems:

$$\min_{\mathbf{x}} f(\mathbf{x}), \quad \mathbf{x} \in \mathcal{X} \coloneqq \{\mathbf{x} | \mathbf{A}\mathbf{x} \leq \mathbf{b}\}$$

where $f : \mathbb{R}^d \to \mathbb{R}$ is twice differentiable (possibly non-convex), $A \in \mathbb{R}^{m \times d}$, and $b \in \mathbb{R}^m$ are some given matrix and vector. It is easy to see that even though the original NMF problem given in (NMF) does not fall under their framework, our novel formulation (Simplex-NMF) does fall under their framework, so we can use their results in terms of convergence to stationary points. However, a priori, the quality of such stationary points has nothing to do with the stationary points of the original NMF, and this is why our main paper and results had to deal with this issue, separately from convergence guarantees.

Definition A.5 ((ϵ_g, ϵ_H)-SOSP)**.** *A point $\mathbf{x}^* \in \mathcal{X}$ is an (ϵ_g, ϵ_H)-second-order stationary point for the above optimization problem of Lu et al. (2020) if the following conditions are satisfied:*

$$\nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq -\epsilon_g, \;\; \forall \mathbf{x} \in \mathcal{X}, \;\; s.t. \;\; \|\mathbf{x} - \mathbf{x}^*\| \leq 1$$

and also,

$$(\mathbf{x} - \mathbf{x}^*)^\top \nabla^2 f(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \geq -\epsilon_H, \forall \mathbf{x} \in \mathcal{X} \;\; s.t. \;\; \nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) = 0$$

Instantiating their definition and results for $\epsilon_g = \epsilon, \epsilon_H = \sqrt{\epsilon}$, we get convergence to a $(\epsilon, \sqrt{\epsilon})$-SOSP in $O(\frac{1}{\epsilon^{2.5}})$ iterations for our problem (Simplex-NMF). Our reduction and analyses here together with our obtained result for non-negative matrix factorization is, to the best of our knowledge, the first analysis proving convergence to second-order stationary points for the original (NMF) with non-asymptotic rates.

A.4 Omitted Proofs

A.4.1 Hessian Lipschitz bound

Lemma A.6. *For a pair $z = (W, H)$, let $\|z\| = (\|W\|_F^2 + \|H\|_F^2)^{1/2}$, and Hessians are equipped with the induced operator norm. For every $C > 0$ and all $z, \widetilde{z} \in \Delta_C$,*

$$\|\nabla F(z) - \nabla F(\widetilde{z})\| \le \left(2\|V\|_F + \frac{5}{2}C^2\right)\|z - \widetilde{z}\|,$$
$$\|\nabla^2 F(z) - \nabla^2 F(\widetilde{z})\|_{\mathrm{op}} \le 6C\|z - \widetilde{z}\|.$$

Proof. Let $P(z) = WH$, $u = (A, B)$, and $v = (E, G)$. Then

$$DP_z[u] = AH + WB, \qquad D^2P_z[u, v] = AG + EB.$$

Nonnegativity and total mass C imply

$$\|DP_z[u]\|_F \le C\|u\|, \qquad \|D^2P_z[u, v]\|_F \le \|u\|\|v\|, \qquad \|WH\|_F \le C^2/4.$$

For example, the second bound follows from $\|A\|_F\|G\|_F + \|E\|_F\|B\|_F \le \|u\|\|v\|$. Using (7) and $\|AB\|_F \le (\|A\|_F^2 + \|B\|_F^2)/2$ gives

$$|D^2F_z[u, u]| \le \left(2C^2 + 2(\|V\|_F + C^2/4)\right)\|u\|^2.$$

This bounds the Hessian operator norm by $L_1 = 2\|V\|_F + 5C^2/2$. Since $D^3P = 0$, for any third direction w,

$$D^3F_z[u, v, w] = 2\big(\langle D^2P_z[u, v], DP_z[w]\rangle_F + \langle D^2P_z[u, w], DP_z[v]\rangle_F + \langle D^2P_z[v, w], DP_z[u]\rangle_F\big).$$

Hence $|D^3F_z[u, v, w]| \le 6C\|u\|\|v\|\|w\|$. The line segment between z and $\widetilde{z}$ lies in the convex simplex. Integrating the Hessian, respectively the third derivative, along that segment proves the two claimed Lipschitz inequalities. □

A.4.2 Proof of Lemma 4.3

Proof of Lemma 4.3. The only difference between the two formulations is the normalization constraint with parameter C. To find the $(\hat{W}, \hat{H})$ that is a FOSP for (Simplex-NMF), we first need to define an intermediate FOSP of (NMF), say $(\tilde{W}, \tilde{H})$, where $\tilde{W} = W^* \cdot D$ and $\tilde{H} = D^{-1} \cdot H^*$ for an appropriately chosen diagonal matrix D with positive entries (so that the inverse D^{-1} is well-defined). The basic idea is to select D in a way that minimizes the total sum of all entries of $\tilde{W}$ plus $\tilde{H}$, and show that:

$$\sum_{i,k} \tilde{W}_{ik} + \sum_{k,j} \tilde{H}_{kj} \le C \tag{5}$$

Then, the equivalence of $(\hat{W}, \hat{H})$ and (W^*, H^*) directly follows as we will eventually select $\hat{W} = t\tilde{W}$ and $\hat{H} = \frac{1}{t}\tilde{H}$ for some sufficiently large scalar $t > 0$ and hence, $\hat{W}\hat{H} = t\tilde{W} \cdot \frac{1}{t}\tilde{H} = W^*D \cdot D^{-1}H^* = W^*H^*$. Next we show that $(\tilde{W}, \tilde{H})$ indeed satisfies Eq. (5).

Consider the pair $(\tilde{W}, \tilde{H})$ where for each $k \in [r]$ (see Preliminaries for notation):

$$\tilde{W}^k := \sqrt{\frac{\|H_k^*\|_1}{\|W^{*k}\|_1}}W^{*k} \text{ and } \tilde{H}_k := \sqrt{\frac{\|W^{*k}\|_1}{\|H_k^*\|_1}}H_k^*$$

By definition,[5]

$$\tilde{W}\tilde{H} = \sum_{k=1}^{r} \tilde{W}^k\tilde{H}_k = \sum_{k=1}^{r} W^{*k}H_k^* = W^*H^* \implies F(W^*, H^*) = F(\tilde{W}, \tilde{H}).$$

[5]Without loss of generality, we can assume that both $\|W^{*k}\|_1$ and $\|H_k^*\|_1$ are not equal to zero (if not, we can ignore the k^{th} column and row and the desired inequality Eq. (5) still holds.

Furthermore, note that

$$\left\|\tilde{W}^k\right\|_1 + \left\|\tilde{H}_k\right\|_1 = 2\left(\left\|\tilde{W}^k\right\|_1 \left\|\tilde{H}_k\right\|_1\right)^{\frac{1}{2}} = 2\left(\sum_{i,j} W^*_{ik}H^*_{kj}\right)^{\frac{1}{2}} = 2\left[(\sum_{i,j} W^*_{ik}H^*_{kj})^2\right]^{\frac{1}{4}}$$

$$\leq 2\left[nm\sum_{i,j} W^{*2}_{ik}H^{*2}_{kj}\right]^{\frac{1}{4}} \leq 2(nm)^{\frac{1}{4}}\left\|W^*H^*\right\|_F^{\frac{1}{2}} \leq 2(nm)^{\frac{1}{4}}\left\|V\right\|_F^{\frac{1}{2}}$$

where we used the fact that any FOSP of (NMF) satisfies $\|W^*H^*\|_F \leq \|V\|_F$ (Theorem 3.3 in Ho et al. (2011)) for the last inequality. As k ranges over $[r]$, we have shown Eq. (5):

$$\sum_{i,k}\tilde{W}_{ik} + \sum_{k,j}\tilde{H}_{kj} = \sum_k\left(\left\|\tilde{W}^k\right\|_1 + \left\|\tilde{H}_k\right\|_1\right) \leq 2r(nm)^{\frac{1}{4}}\left\|V\right\|_F^{\frac{1}{2}} < C$$

Notice how our assumption $C > 2r(nm)^{\frac{1}{4}}\|V\|_F^{\frac{1}{2}}$ was crucial: if C was selected smaller, then $(\tilde{W}, \tilde{H})$ (and hence $(\hat{W}, \hat{H})$) may not exist. Lastly, consider the parametrized family of matrix pairs $(\hat{W}, \hat{H}) = (t\tilde{W}, \frac{1}{t}\tilde{H})$ and increase t until the total sum reaches C:

$$t\left(\sum_{i,k}\tilde{W}_{ik}\right) + \frac{1}{t}\left(\sum_{k,j}\tilde{H}_{kj}\right) = C.$$

This can always be done by continuity with respect to the parameter t. Observe that the expression on the left tends to infinity as t grows. Because only rescalings were involved in the transformations between the matrices, this leads to a pair of matrices $(\hat{W}, \hat{H})$ that correspond to a FOSP of the (Simplex-NMF). □

First of all let's compute the partial derivatives since these will be used throughout the appendix.

The entries of the gradient of $F(W, H) := \|V - WH\|_F^2$ are given by:

$$\begin{aligned}\frac{\partial F}{\partial W_{ik}} &= -2\sum_{j=1}^m\left(V_{ij} - \sum_{\ell=1}^r W_{i\ell}H_{\ell j}\right)H_{kj},\\ \frac{\partial F}{\partial H_{kj}} &= -2\sum_{i=1}^n\left(V_{ij} - \sum_{\ell=1}^r W_{i\ell}H_{\ell j}\right)W_{ik},\end{aligned} \tag{6}$$

whereas the entries of its Hessian are given by:

$$\begin{aligned}&\frac{\partial^2 F}{\partial^2 W_{ik}} = 2\sum_{j=1}^m H^2_{kj} \text{ and } \frac{\partial^2 F}{\partial^2 H_{kj}} = 2\sum_{i=1}^n W^2_{ik}\\ &\frac{\partial^2 F}{\partial W_{ik}\partial W_{i'\ell}} = 0 \text{ and } \frac{\partial^2 F}{\partial H_{kj}\partial H_{\ell j'}} = 0 \text{ for } i \neq i', j \neq j'\\ &\frac{\partial^2 F}{\partial W_{ik}\partial W_{i\ell}} = 2\sum_{j=1}^m H_{\ell j}H_{kj}\\ &\frac{\partial^2 F}{\partial H_{kj}\partial H_{\ell j}} = 2\sum_{i=1}^n W_{i\ell}W_{ik}\\ &\frac{\partial^2 F}{\partial W_{ik}\partial H_{kj}} = -2V_{ij} + 2\sum_{\ell=1}^r W_{i\ell}H_{\ell j} + 2W_{ik}H_{kj}\\ &\frac{\partial^2 F}{\partial W_{ik}\partial H_{\ell j}} = 2W_{i\ell}H_{kj}.\end{aligned} \tag{7}$$

Proof of Claim 4.4. Direct calculation reveals that:

$$
\begin{aligned}
c\sum_{i=1}^{n} W_{ik}^* &= \sum_{i:W_{ik}^*>0}^{n} W_{ik}^* \cdot c = \sum_{i:W_{ik}^*>0} W_{ik}^* \cdot \frac{\partial F(W^*,H^*)}{\partial W_{ik}} \\
&= \sum_{i=1}^{n} W_{ik}^* \left[-2\sum_{j=1}^{m} \left(V_{ij} - \sum_{\ell=1}^{r} W_{i\ell}^* H_{\ell j}^* \right) H_{kj}^* \right] \\
&= \sum_{j=1}^{m} H_{kj}^* \left[-2\sum_{i=1}^{n} \left(V_{ij} - \sum_{\ell=1}^{r} W_{i\ell}^* H_{\ell j}^* \right) W_{ik}^* \right] \\
&= \sum_{j:H_{kj}^*>0}^{m} H_{kj}^* \left[-2\sum_{i=1}^{n} \left(V_{ij} - \sum_{\ell=1}^{r} W_{i\ell}^* H_{\ell j}^* \right) W_{ik}^* \right] = c\sum_{j=1}^{m} H_{kj}^*.
\end{aligned}
$$

□

Lemma A.7. *For any pair of matrices (W,H) and index $k \in \{1,\dots,r\}$, let the pair of matrices $(\tilde{W},\tilde{H})$ such that **1)** $\tilde{W}_{ik} = W_{ik}$ and $\tilde{W}_{i\ell} = 0$, **2)** $\tilde{H}_{kj} = -H_{kj}$ and $\tilde{H}_{\ell j} = 0$. Then*

$$
(\mathrm{vec}(\tilde{W})^\top, \mathrm{vec}(\tilde{H})^\top) \cdot \nabla^2 F(W,H) = \begin{pmatrix} \mathbf{0}^\top \\ -\nabla_{W^k} F \\ \mathbf{0}^\top \\ \nabla_{H_k} F \\ \mathbf{0}^\top \end{pmatrix},
$$

where $\mathbf{0}$ denotes the zero column vector of appropriate size.

Proof. Let us start by proving that for all the entries of the vector corresponding to W_{ik} and H_{kj}, we have (respectively):

- $\left[(\mathrm{vec}(\tilde{W})^\top, \mathrm{vec}(\tilde{H})^\top) \cdot \nabla^2 F(W,H)\right]_{W_{ik}} = -\frac{\partial F}{\partial W_{ik}}$
- $\left[(\mathrm{vec}(\tilde{W})^\top, \mathrm{vec}(\tilde{H})^\top) \cdot \nabla^2 F(W,H)\right]_{H_{kj}} = \frac{\partial F}{\partial H_{kj}}$

By direct calculation we get that,

$$
\begin{aligned}
&\left[(\mathrm{vec}(\tilde{W})^\top, \mathrm{vec}(\tilde{H})^\top) \cdot \nabla^2 F(W,H)\right]_{W_{ik}} \\
&= \frac{\partial^2 F}{\partial^2 W_{ik}} \tilde{W}_{ik} + 2\sum_{\ell\neq k} \frac{\partial^2 F}{\partial W_{ik}\partial W_{i\ell}} \tilde{W}_{i\ell} \\
&\quad + 2\sum_{j=1}^{m} \frac{\partial^2 F}{\partial W_{ik}\partial H_{kj}} \tilde{H}_{kj} + 2\sum_{j=1}^{m}\sum_{\ell\neq k} \frac{\partial^2 F}{\partial W_{ik}\partial H_{\ell j}} \tilde{H}_{\ell j} \\
&= \frac{\partial^2 F}{\partial^2 W_{ik}} \tilde{W}_{ik} + 2\sum_{j=1}^{m} \frac{\partial^2 F}{\partial W_{ik}\partial H_{kj}} \tilde{H}_{kj} \\
&= \frac{\partial^2 F}{\partial^2 W_{ik}} W_{ik} + 2\sum_{j=1}^{m} \frac{\partial^2 F}{\partial W_{ik}\partial H_{kj}} (-H_{kj}) \\
&= 2\sum_{j=1}^{m} W_{ik} H_{kj}^2 + 2\sum_{j=1}^{m} \left[V_{ij} - \sum_{\ell=1}^{r} W_{i\ell} H_{\ell j} \right] H_{kj} - 2\sum_{j=1}^{m} W_{ik} H_{kj}^2 \\
&= -\frac{\partial F}{\partial W_{ik}}
\end{aligned}
$$

where the last equality follows by (6). Respectively for H_{kj}. Up next we prove that

$$\left[(\mathrm{vec}(\tilde{W})^\top, \mathrm{vec}(\tilde{H})^\top) \cdot \nabla^2 F(W,H)\right]_{W_{ik'}} = \left[(\mathrm{vec}(\tilde{W})^\top, \mathrm{vec}(\tilde{H})^\top) \cdot \nabla^2 F(W,H)\right]_{H_{k'j}} = 0$$

$$\begin{aligned}
&\left[(\mathrm{vec}(\tilde{W})^\top, \mathrm{vec}(\tilde{H})^\top) \cdot \nabla^2 F(W,H)\right]_{W_{ik'}} \\
&= \frac{\partial^2 F}{\partial^2 W_{ik'}} \tilde{W}_{ik'} + 2\sum_{\ell \neq k'} \frac{\partial^2 F}{\partial W_{ik'} \partial W_{i\ell}} \tilde{W}_{i\ell} \\
&\quad + 2\sum_{j=1}^{m} \frac{\partial^2 F}{\partial W_{ik'} \partial H_{k'j}} \tilde{H}_{k'j} + 2\sum_{j=1}^{m}\sum_{\ell \neq k'} \frac{\partial^2 F}{\partial W_{ik'} \partial H_{\ell j}} \tilde{H}_{\ell j} \\
&= 2\frac{\partial^2 F}{\partial W_{ik'} \partial W_{ik}} \tilde{W}_{ik} + 2\sum_{j=1}^{m} \frac{\partial^2 F}{\partial W_{ik'} \partial H_{kj}} \tilde{H}_{kj} \\
&= 2\frac{\partial^2 F}{\partial W_{ik'} \partial W_{ik}} W_{ik} + 2\sum_{j=1}^{m} \frac{\partial^2 F}{\partial W_{ik'} \partial H_{kj}} (-H_{kj}) \\
&= 2\sum_{j=1}^{m} H_{k'j} H_{kj} W_{ik} - 2\sum_{j=1}^{m} W_{ik} H_{k'j} H_{kj} = 0.
\end{aligned}$$

Respectively for $H_{k'j}$. □

Proof of Lemma 4.5, continued. Recall that we constructed the desired (W^s, H^s) based on a particular index $k := \mathbf{argmax}_{1 \leq \ell \leq r} \left(\sum_{i=1}^n W^*_{i\ell} + \sum_{j=1}^m H^*_{\ell j}\right)$. By the choice of k and averaging, it follows immediately using Claim 4.4 that

$$\left\|W^{*k}\right\|_1 = \|H^*_k\|_1 \geq C/2r. \tag{8}$$

By the choice of (W^s, H^s), it is immediate that Eq. (1) are satisfied, since $\frac{\partial F(W^*,H^*)}{\partial W_{ki}} > c$ implies $W^*_{ik} = 0$ and thus $W^s_{ik} = 0$, by the definition of W^s (and analogously for H^*_{kj}).

It is left to show that $s^\top \nabla^2 F(W^*, H^*) s < 0$. By Lemma A.7, we have that

$$\begin{aligned}
&s^\top \nabla^2 F(W^*,H^*) s = \\
&(\mathrm{vec}(W^s)^\top, \mathrm{vec}(H^s)^\top) \cdot \nabla^2 F(W^*,H^*) \cdot (\mathrm{vec}(W^s), \mathrm{vec}(H^s)) \\
&= (\mathrm{vec}(W^s)^\top, \mathrm{vec}(H^s)^\top) \cdot \begin{pmatrix} 0 \\ -\nabla_{W^k} F(W^*,H^*) \\ 0 \\ \nabla_{H_k} F(W^*,H^*) \\ 0 \end{pmatrix} \\
&= \sum_{i=1}^n \left(-\frac{\partial F(W^*,H^*)}{\partial W_{ik}}\right) W^s_{ik} + \sum_{i=1}^n \frac{\partial F(W^*,H^*)}{\partial H_{kj}} H^s_{kj} \\
&= \sum_{i=1}^n \left(-\frac{\partial F(W^*,H^*)}{\partial W_{ik}}\right) W^*_{ik} + \sum_{i=1}^n \frac{\partial F(W^*,H^*)}{\partial H_{kj}} (-H^*_{kj}) \\
&= 2\sum_{i=1}^n \sum_{j=1}^m V_{ij} W^*_{ik} H^*_{kj} - 2\sum_{i=1}^n \sum_{j=1}^m W^*_{ik} H^*_{kj} \sum_{\ell=1}^r W^*_{i\ell} H^*_{\ell j} \\
&\leq 2\sum_{i=1}^n \sum_{j=1}^m V_{ij} W^*_{ik} H^*_{kj} - 2\sum_{i=1}^n \sum_{j=1}^m W^{*2}_{ik} H^{*2}_{kj} \\
&\leq 2\|V\|_F \left\|W^{*k} H^*_k\right\|_F - 2\left\|W^{*k} H^*_k\right\|_F^2
\end{aligned}$$

where the first inequality follows from the fact that all the entries $W^*_{i\ell}, H^*_{\ell j}$ are positive and the second from the Cauchy-Schwarz inequality. In order to show $s^\top \nabla^2 F(W^*, H^*)s < 0$, it remains to show that $\|V\|_F - \left\|W^{*k}H^*_k\right\|_F < 0$. Indeed,

$$\begin{aligned}
\left\|W^{*k}H^*_k\right\|_F^2 &= \sum_{i=1}^{n}\sum_{j=1}^{m} W^{*2}_{ik}H^{*2}_{kj} \\
&\geq \frac{1}{nm}\left(\sum_{i=1}^{n}\sum_{j=1}^{m} W^*_{ik}H^*_{kj}\right)^2 \\
&= \frac{1}{nm}\left(\sum_{i=1}^{n} W^*_{ik}\right)^2\left(\sum_{j=1}^{m} H^*_{kj}\right)^2 \\
&\geq \frac{1}{16nm}\left(\left\|W^{*k}\right\|_1\right)^4,
\end{aligned}$$

where the last inequality follows by Claim 4.4. Combining the above with (8) we arrive at,

$$\left\|W^{*k}H^*_k\right\|_F \geq \frac{C^2}{4r^2\sqrt{nm}} > \|V\|_F,$$

where we used that $C > 2r(nm)^{\frac{1}{4}}\sqrt{\|V\|_F}$.

□

We now continue to finish off the proof of Theorem 2, where we again start with a SOSP of (Simplex-NMF) and assume for contradiction that it is not a SOSP of (NMF). Testing the KKT conditions in this case will require a more elaborate construction for (W^s, H^s) that we describe next.

Proof of Theorem 2, continued. Let (W^*, H^*) be a second-order stationary point of the problem (Simplex-NMF) which is not a second-order stationary point of the unconstrained problem (NMF). Since (W^*, H^*) is a SOSP of (Simplex-NMF), Lemma 4.5 implies that $c = 0$.

As in the proof of Lemma 4.5, we will arrive at a contradiction by identifying a non-zero pair of matrices (W^s, H^s) that satisfy the following 4 conditions:

$$\frac{\partial F(W^*, H^*)}{\partial W_{ik}} > 0 \implies W^s_{ik} = 0, \tag{9}$$

$$\frac{\partial F(W^*, H^*)}{\partial H_{kj}} > 0 \implies H^s_{kj} = 0, \tag{10}$$

$$\sum_{i,k} W^s_{ik} + \sum_{k,j} H^s_{kj} = 0, \text{ and yet,} \tag{11}$$

$$s^\top \nabla^2 F(W^*, H^*)s < 0, \tag{12}$$

where we set $s^\top = (\text{vec}(W^s)^\top, \text{vec}(H^s)^\top)$.

As (W^*, H^*) is not a SOSP of (NMF), there exist a pair of matrices $(\hat{W}, \hat{H})$ satisfying conditions (9), (10) and (12). In the remainder of the proof, we use the pair $(\hat{W}, \hat{H})$ to construct a new pair (W^s, H^s) that also satisfies (11). In the construction of the pair (W^s, H^s), we will also use the pair $(\tilde{W}, \tilde{H})$ of Lemma A.7 for $k = \mathbf{argmax}_{1\leq\ell\leq r}\{\sum_{i,\ell} W^*_{i\ell} + \sum_{\ell,j} H^*_{\ell j}\}$. We remind that $\tilde{W}_{ik} = W^*_{ik}, \tilde{H}_{kj} = -H^*_{kj}$ and $\tilde{W}_{i\ell} = \tilde{H}_{\ell j} = 0$. Moreover ($\mathbf{0}$ denotes the zero column vector of

appropriate size), $(\mathrm{vec}(\tilde{W})^\top, \mathrm{vec}(\tilde{H})^\top)$·

$$\nabla^2 F(W^*, H^*) = \begin{pmatrix} \mathbf{0}^\top \\ -\nabla_{W^k} F(W^*, H^*) \\ \mathbf{0}^\top \\ \nabla_{H_k} F(W^*, H^*) \\ \mathbf{0}^\top \end{pmatrix}.$$

We are now ready to describe our construction. Consider the following family of matrix pairs:

$$(W^t, H^t) = (\tilde{W}, \tilde{H}) + t(\hat{W}, \hat{H})^6,$$

which we show satisfies Conditions (9), (10) and (12). Indeed, assuming that $\frac{\partial F(W^*, H^*)}{\partial W_{ik}} > 0$, it follows by (9) that $\hat{W}_{ik} = 0$, and furthermore, as (W^*, H^*) is a SOSP for (Simplex-NMF), we also have that $W^*_{ik} = 0$ and thus $\tilde{W}_{ik} = 0$. Combining these we get that $W^t_{ik} = 0$, i.e., (W^t, H^t) satisfies (9) for all t. Analogously, if $\frac{\partial F(W^*, H^*)}{\partial H_{kj}} > 0$ it follows that $H^t_{kj} = 0$.

Lastly, we show that for all $t \neq 0$, the pair of matrices (W^t, H^t) satisfy (12). To simplify notation let,

- $s_t^\top = (\mathrm{vec}(W^t)^\top, \mathrm{vec}(H^t)^\top)$, respectively $\hat{s}$ for $(\hat{W}, \hat{H})$ and $\tilde{s}$ for $(\tilde{W}, \tilde{H})$.
- $\nabla^2 = \nabla^2 F(W^*, H^*)$.

Then, we have that

$$\begin{aligned} s_t^\top \nabla^2 s_t &= (\tilde{s} + t\hat{s})^\top \nabla^2 (\tilde{s} + t\hat{s}) \\ &= \tilde{s}^\top \nabla^2 \tilde{s} + 2t\tilde{s}^\top \nabla^2 \hat{s} + t^2 \hat{s}^\top \nabla^2 \hat{s} \\ &= \underbrace{(0, -\nabla_{W^k} F(W^*, H^*), 0, \nabla_{H_k} F(W^*, H^*), 0)^\top \cdot \tilde{s}}_{=0} \\ &+ \underbrace{2t(0, -\nabla_{W^k} F(W^*, H^*), 0, \nabla_{H_k} F(W^*, H^*), 0)^\top \cdot \hat{s}}_{=0} \\ &+ t^2 \underbrace{\hat{s}^\top \nabla^2 \hat{s}}_{<0} \end{aligned}$$

where for the second equality we use Lemma A.7. The first term is zero since if $\tilde{W}_{ik} \neq 0$, it implies that $W^*_{ik} \neq 0$, which in turn implies that $\frac{\partial F(W^*, H^*)}{\partial W_{ik}} = 0$. This is because (W^*, H^*) is assumed to be a SOSP (analogously for H_{kj}). By Conditions (9)-(10), we know that if $\frac{\partial F(W^*, H^*)}{\partial W_{ik}} > 0$ then $\hat{W}_{ik} = 0$ (analogously for H_{kj}), meaning that the second term is also zero. Finally, the third term is strictly negative as $(\hat{W}, \hat{H})$ satisfies (12).

We complete the proof of Theorem 2 by noting that for

$$\begin{aligned} t &= -\frac{\sum_{i=1}^n \sum_{k=1}^r \tilde{W}_{ik} + \sum_{k=1}^r \sum_{j=1}^m \tilde{H}_{kj}}{\sum_{i=1}^n \sum_{k=1}^r \hat{W}_{ik} + \sum_{k=1}^r \sum_{j=1}^m \hat{H}_{kj}} \\ &= -\frac{\sum_{i=1}^n W^*_{ik} - \sum_{j=1}^m H^*_{kj}}{\sum_{i=1}^n \sum_{k=1}^r \hat{W}_{ik} + \sum_{k=1}^r \sum_{j=1}^m \hat{H}_{kj}}, \end{aligned}$$

the remaining Condition (11) is satisfied. An important detail is that $t \neq 0$ since $\sum_{i=1}^n W^*_{ik} = \sum_{j=1}^m H^*_{kj}$ contradicts with the assumption that (W^*, H^*) is SOSP (recall the proof of Lemma 4.5 presented above). □

[6] We slightly abuse the notation here since, W^t up until now meant the t-th column, whereas here it is used to denote a parameterized family of matrices where the parameter is t.

A.4.3 COMPUTING LIPSCHITZNESS FOR HESSIAN

An upper bound on the Lipschitz constant The entries of the Hessian matrix of F are bounded. Recall C is a constant such that $C > 2r(nm)^{\frac{1}{4}}\sqrt{\|V\|_F}$.

$$\begin{aligned}\frac{\partial^2 F}{\partial W_{i\beta}\partial W_{i\alpha}} &= 2\langle H_\alpha, H_\beta\rangle \le 2\|H_\alpha\|_2\|H_\beta\|_2 \\ &\le 2\|H_\alpha\|_{1,1}\|H_\beta\|_{1,1} \le 2(\|H\|_{1,1}+\|W\|_{1,1})(\|H\|_{1,1}+\|W\|_{1,1}) \\ &= 2C^2\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 F}{\partial H_{\alpha j}\partial W_{i\alpha}} &= 2(W_i H^j - V_{ij}) + 2W_{i\alpha}H_{\alpha j} \\ &\le 2\left|W_i H^j\right| + 2|V_{ij}| + 2|W_{i\alpha}H_{\alpha j}| \\ &\le 2\|WH\|_{1,1} + 2\|V\|_{1,1} + 2\|WH\|_{1,1} \\ &= 4\|WH\|_{1,1} + 2\|V\|_{1,1} \\ &\le (\|W\|_{1,1}+\|H\|_{1,1})^2 + 2\|V\|_{1,1} = C^2 + 2\sqrt{nm}\|V\|_F \\ &\le 2C^2 \ \ (\text{since } 2\sqrt{nm}\|V\|_F \le C^2)\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 F}{\partial H_{\alpha' j}\partial W_i\alpha} &= 2H_{\alpha j}W_{i\alpha'} \\ &\le 2\|H\|_{1,1}\|W\|_{1,1} \le \frac{1}{2}(\|H\|_{1,1}+\|W\|_{1,1})^2 = \frac{1}{2}C^2\end{aligned}$$

So the entry-wise l_1 norm of Hessian,i.e. $\left\|\nabla^2 F\right\|_{1,1}$, is bounded by $2C^2$. Recall we denote $\|A\|_{1,1}$ the sum of absolute values of all entries of A, i.e. $\|A\|_{1,1} = \sum_{i,j}|A_{ij}|$, we have the Cauchy-Schwarz inequality by vectorizing the matrix A:

$$\|A\|_F \le \|A\|_{1,1}.$$

Then combining with the matrix norm inequality between the l_2 norm and the Frobenious norm, we have

$$\left\|\nabla^2 F\right\|_2 \le \left\|\nabla^2 F\right\|_F \le \left\|\nabla^2 F\right\|_{1,1} \le K,$$

where $K = 2C^2(r^2(n+nm+m)+rm^2n)$. The esitmate of $\left\|\nabla^2 F\right\|_{1,1}$ is completed by direct calculation. Using the notation of matrix differentiation, we can write the Hessian in the following form

$$\nabla^2 F(W,H) = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where

$$A = \begin{bmatrix} \frac{\partial^2 F}{\partial W_1 \partial W_1^\top} & \cdots & \frac{\partial^2 F}{\partial W_n \partial W_1^\top} \\ \vdots & & \vdots \\ \frac{\partial^2 F}{\partial W_1 \partial W_n^\top} & \cdots & \frac{\partial^2 F}{\partial W_n \partial W_n^\top} \end{bmatrix}, B = \begin{bmatrix} \frac{\partial^2 F}{\partial H_1 \partial W_1^\top} & \cdots & \frac{\partial^2 F}{\partial H_r \partial W_1^\top} \\ \vdots & & \vdots \\ \frac{\partial^2 F}{\partial H_1 \partial W_n^\top} & \cdots & \frac{\partial^2 F}{\partial H_r \partial W_n^\top} \end{bmatrix}$$

$$C = \begin{bmatrix} \frac{\partial^2 F}{\partial W_1 \partial H_1^\top} & \cdots & \frac{\partial^2 F}{\partial W_n \partial H_1^\top} \\ \vdots & & \vdots \\ \frac{\partial^2 F}{\partial W_1 \partial H_r^\top} & \cdots & \frac{\partial^2 F}{\partial W_n \partial H_r^\top} \end{bmatrix} \text{ and } D = \begin{bmatrix} \frac{\partial^2 F}{\partial H_1 \partial H_1^\top} & \cdots & \frac{\partial^2 F}{\partial H_r \partial H_1^\top} \\ \vdots & & \vdots \\ \frac{\partial^2 F}{\partial H_1 \partial H_r^\top} & \cdots & \frac{\partial^2 F}{\partial H_r \partial H_r^\top} \end{bmatrix}.$$

Moreover, the matrix A, B, C, D are still blocked matrices. For matrix A, the diagonal blocks have the form of

$$\begin{bmatrix} \langle H_1, H_1 \rangle & \cdots & \langle H_1, H_r \rangle \\ \vdots & & \vdots \\ \langle H_r, H_1 \rangle & \cdots & \langle H_r, H_r \rangle \end{bmatrix}$$

and the other blocks of A are $\mathbf{0}_{r\times r}$. Thus we have $\|A\|_{1,1} \leq r^2 n 2C^2$. To obtain upper bound of $\|B\|_{1,1}$ and $\|C\|_{1,1}$, we look at the form of the blocked entries of B and C. The blocked entries of B have the same form of $\frac{\partial^2 F}{\partial H_1 \partial W_1^\top}$, and we have that

$$\frac{\partial^2 F}{\partial H_1 \partial W_1^\top} = \begin{bmatrix} (W_1 H^1 - V_{11}) & \cdots & (W_1 H^m - V_{1m}) \\ \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} + \begin{bmatrix} W_{11} H_{11} & \cdots & W_{11} H_{1m} \\ \vdots & & \vdots \\ W_{11} H_{r1} & \cdots & W_{11} H_{rm} \end{bmatrix},$$

and the blocked entries of C have the form of

$$\frac{\partial^2 F}{\partial W_1 \partial H_1^\top} = \begin{bmatrix} (W_1 H^1 - V_{11}) & \mathbf{0} \\ \vdots & \vdots \\ (W_1 H^m - V_{1m}) & \mathbf{0} \end{bmatrix} + \begin{bmatrix} W_{11} H_{11} & \cdots & W_{11} H_{r1} \\ \vdots & & \vdots \\ W_{11} H_{1m} & \cdots & W_{11} H_{rm} \end{bmatrix}.$$

We can see that all the entries of B and C are bounded by $2C^2$, so $\|B\|_{1,1} \leq 2C^2 m^2 rn$ and $\|C\|_{1,1} \leq 2C^2 r^2 mn$. Each block entry of D is a diagonal matrix,

$$\frac{\partial^2 F}{\partial H_{\alpha'} \partial H_\alpha^\top} = \mathbf{diag}\{\langle W^{\alpha'}, W^\alpha \rangle\}_{m\times m},$$

and then we have $\|D\|_{1,1} \leq 2C^2 mr^2$. Adding these bounds up, we can obtain the upper bound of $\left\|\nabla^2 F\right\|_{1,1}$.

A.5 Omitted Iterative Algorithm

In Algorithm 1 we describe the MWU method applied to simplex-NMF (Simplex-NMF). As already mentioned, the recent result in Panageas et al. (2019) ensures that Algorithm 1 converges with probability 1 to a second-order stationary point (SOSP) of (Simplex-NMF), which in case $C > 2r(nm)^{1/4}\sqrt{\|V\|_F}$ is also a second-order stationary point (SOSP) of (NMF). The latter is formally stated in Corollary A.8. A simple implementation of the algorithm in Matlab is provided in the Supplement.

Corollary A.8. *Assume the sequence of matrices produced by Algorithm 1, $(W^T, H^T) \to (W^*, H^*)$ as $T \to \infty$ for some initialization (W^0, H^0). The probability (W^*, H^*) is a second-order stationary point (SOSP) for simplex-NMF (Simplex-NMF) is one Panageas et al. (2019). By our Theorem 2, (W^*, H^*) is also a second-order stationary point (SOSP) for (NMF).*

We also note a perhaps peculiar property of the Algorithm 1, compared to all prior algorithms of NMF (see also Further Related Work section). Algorithm 1 can be run in parallel, i.e., the updates of matrix W and of matrix H can be performed in parallel, without the need to first update one of them (let's say W) and then using the updated values of W, to update the other one (H in this case). In practice, we didn't observe significant runtime benefits from this property, but in the future it could yield fast implementations without the need of sequential updates. For an example where the seminal algorithm of Lee & Seung (1999; 2001) can go wrong if run in parallel (instead of sequentially), please check Figure 4.

A.6 Another Simple NMF Example

Example 1. *Let the NMF instance with $V = \begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix}$ and $r = 2$. An optimal solution is $W^* = H^* = \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}$ achieving 0 in the objective function. The pair of matrices $W^s = H^s = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$*

Algorithm 1 Concurrent Multiplicative Weights Update

Input: A matrix $V_{n\times m}$ with positive elements
Output: Matrices $W_{n\times r}, H_{r\times m}$ s.t. $V \simeq W \cdot H$.

$C = 4r(nm)^{1/4}\sqrt{\|V\|_F}$, $\epsilon = \Theta(\frac{1}{C^2})$
$(W^0, H^0) \leftarrow$ a random point in simplex
Set $V \leftarrow V/C^2$ /* Rescaling V*/
for $t = 0$ to T **do**
  $Z = 1 - \epsilon\big(\sum_{i,k} W^t_{ik} \cdot \frac{\partial F(W^t,H^t)}{\partial W_{ik}} +$
  $+ \sum_{k,j} H^t_{kj} \cdot \frac{\partial F(W^t,H^t)}{\partial H_{kj}}\big)$ /*Rescaling term*/
  $W^{t+1}_{ik} = W^t_{ik} \cdot (1 - \epsilon \cdot \frac{\partial F(W^t,H^t)}{\partial W_{ik}})/Z$
  $H^{t+1}_{kj} = H^t_{kj}(1 - \epsilon \cdot \frac{\partial F(W^t,H^t)}{\partial H_{kj}})/Z$
end for
$W \leftarrow C \cdot W^{T+1}, H \leftarrow C \cdot H^{T+1}$
return (W, H).

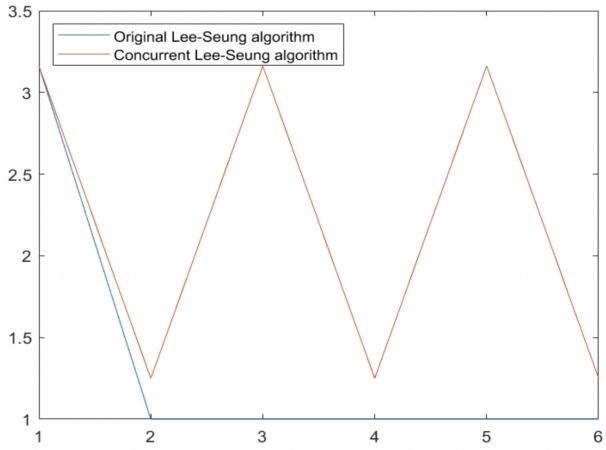


Figure 4: y-axis is the frobenius norm error, x-axis is the number of iterations. If we run the concurrent version of the Lee-Seung algorithm, then the algorithm oscillates between bad solutions. However, the sequential version of Lee-Seung as it was designed quickly finds a good solution of small error. The specific example is easy to describe formally. Matrix V is the 2×2 identity matrix, whereas the initial matrix $W^0 = H^0 = \mathbf{J}$ where $\mathbf{J}$ is the 2×2 all-ones matrix.

with objective value $F(W^s, H^s) = 4 \cdot 10^4$ is a saddle point since it satisfies the first-order KKT but not the second-order KKT conditions.

A.7 A Simple Experiment

Our paper does not propose new algorithms, but rather it is a reduction with provable guarantees showing that the (non-smooth) problem of NMF can be modified to fit the framework of smooth non-convex optimization. Here, we implemented, in our context, the MWU algorithm from Panageas et al. (2019) and we randomly initialize in the simplex. We used both synthetic and real data. For synthetic data, we create random matrices V, similarly to what was done in Gonzalez & Zhang (2005); Lu et al. (2020), for $n = 30, m = 20$. Our goal was to run MWU, where the inner dimension r is slowly increased, and observe convergence properties of the algorithm and to check that choosing C carefully is important. In particular, we care about convergence. To test it, we can either check the eigenvalues of the Hessian or approximately verify the KKT conditions. For small $n, m, r \leq 6$, we used symbolic Matlab to verify that the points returned by our algorithm are not saddle points (Hessian has only positive eigenvalues). For larger values, we check the residual norm of the KKT conditions as was done in Gonzalez & Zhang (2005): $\min(H, W^TWH - W^TV) = 0$ where min is

taken component-wise. This condition is necessary for every local minimum of NMF. For real data, we used MIT's CBC Faces Dataset face (images of 115×115 pixels),[7] common in NMF literature.

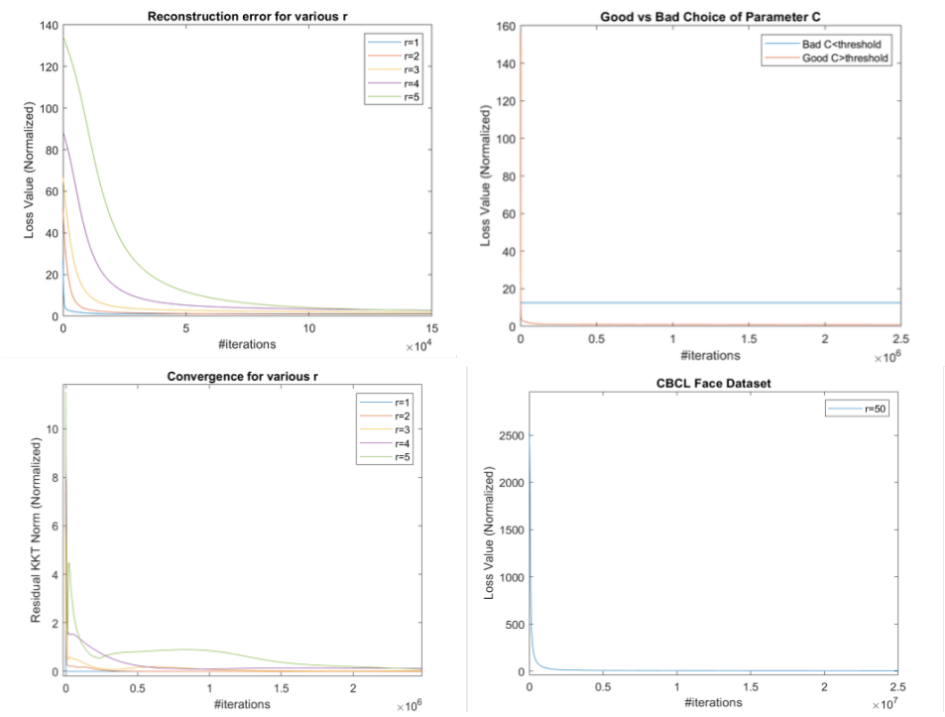


Figure 5: Top Left: We plot Frobenius norm error for various k. Top right: Frobenius norm error for two choices of C vs $2r(nm)^{\frac{1}{4}} \|V\|_F^{\frac{1}{2}}$. Bottom Left: To verify convergence we check the KKT condition which is necessary. Bottom Right: Convergence of MWU for real data from CBCL.

A.8 Further Related Work

Brief History of NMF. The NMF problem was introduced in the engineering community in the seminal paper of Paatero & Tapper (1994) and was subsequently popularized further in Lee & Seung (1999; 2001). Approximate factorizations based on non-negative matrices have found numerous applications throughout many fields of science and engineering with notable examples including music analysis Févotte et al. (2009), document clustering Xu et al. (2003), speech-source separation Schmidt & Olsson (2006) and cancer-class identification Gao & Church (2005). NMF is useful as it leads to additive and sparse representations of the input data, while it is a tool to describe the efficacy of linear programming for hard combinatorial problems Yannakakis (1991).

NMF's wide range of applications led to an intense interest towards the design of efficient algorithms to tackle it. Unfortunately even the simpler decision version of NMF, asking to decide whether equation (NMF) has value 0, is NP-hard Vavasis (2009). Going beyond the decision version of NMF, the last years, several *iterative gradient-based methods* have been proposed for the optimization version (NMF). Arguably one of the most popular approaches is the multiplicative update (MU) rule introduced in Lee & Seung (2001).

In this MU framework, the updates are given by

$$W_{t+1} = W_t \circ \frac{V H_t^\top}{W_t H_t H_t^\top} \quad H_{t+1} = H_t \circ \frac{W_{t+1}^\top V}{W_{t+1}^\top W_{t+1} H_t},$$

where $X \circ Y, X/Y$ denotes the componentwise multiplication, division of two matrices respectively. Please note that to compute the entries of H_{t+1}, the entries of W_{t+1} are needed (see the iteration

[7] http://cbcl.mit.edu/software-datasets/heisele/facerecognition-database.html.

Table 1: MovieLens Small (Harper & Konstan, 2015) with $k = 5$. (R) uses a random $U(0, 1)$ init. and (S) uses NNDSVD. We report the Frobenius Norm on the training data and the RMSE on the test data.

Method	Norm	RMSE
Lee-Seung (R)	384	1.33
Lee-Seung (S)	386	1.34
nSNAP+ (R)	385	1.33
nSNAP+ (S)	387	1.34

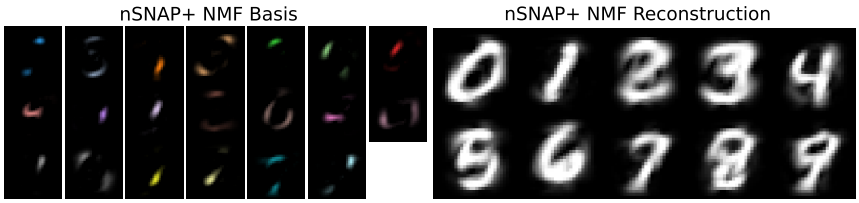


Figure 6: MNIST digits (LeCun, 1998) with $k = 20$, with [0-9] exemplars; $U(0, 1)$ init.

subscripts), so the updates take place in an alternating manner (not concurrently). This sequential nature of the above algorithm is in contrast to the concurrent (parallel) nature of the proposed method in Algorithm 1.

In Lee & Seung (2001) it was shown that under the MU rule, NMF' s objective function does not increase. More recent approaches establish this non-increasing property by applying the *block coordinate descent framework*. A typical example of this kind of methods is the Alternating Non-negative Least Squares (ANLS) method Kim & Park (2008), in which W and H are sequentially updated to the minimizers of the convex programs, $\min_{W\geq 0} \|V - WH\|^2$, and $\min_{H\geq 0} \|V - WH\|^2$. Another example is the Hierarchical ALS (HALS) method A. Cichocki (2007), which has been rediscovered several times, e.g., as the rank-one residue iteration Ho (2008) and FastNMF in Li & Zhang (2009). For additional details concerning iterative methods for NMF and their convergence properties the reader is referred to the surveys Kim et al. (2014); Gillis (2014); Lin (2007a;b) and references therein.

In particular, in Lin (2007a;b), the author describes that it is unclear how to prove convergence to first-order stationarity of the algorithm of Lee & Seung (2001), even if we add some upper bound constraints akin to our normalization constraints. Lin (2007a;b) also describes a variation of the original algorithm of Lee & Seung (2001) (where some small constants are introduced in the denominators to avoid division by 0) that does converge to first-order stationarity. However, none of the proposed methods so far is known to converge to second-order stationary points, as we proved in this work.

A.9 Omitted Figures from Section 6

We also show results on real NMF problems (see Table 1 and Figures 6–8), demonstrating that not only do Simplex-NMF constructions have nice theoretical benefits, but that practical algorithms exist with reasonable results. On image data, using a moving average gradient ("momentum") improved convergence rates.

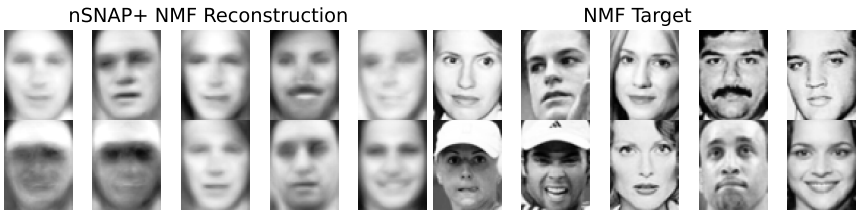


Figure 7: Labeled Faces in the Wild Huang et al. (2007) NMF with $k = 25$; $U(0, 1)$ init.

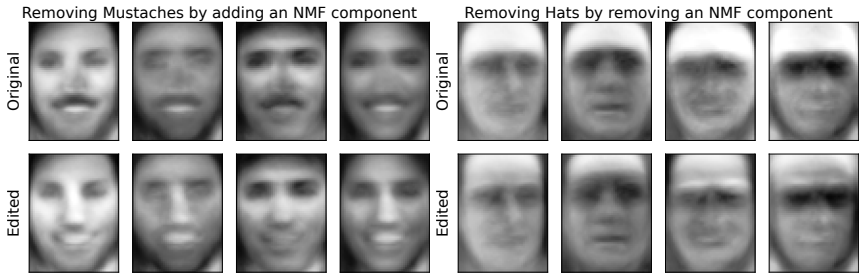


Figure 8: Editing faces by adding or subtracting NMF components.