Revisiting the Landscape of Matrix Factorization

Hossein Valavi **Sulin Liu** **Peter J. Ramadge**
Department of Electrical Engineering, Princeton University

Abstract

Prior work has shown that low-rank matrix factorization has infinitely many critical points, each of which is either a global minimum or a (strict) saddle point. We revisit this problem and provide simple, intuitive proofs of a set of extended results for low-rank and general-rank problems. We couple our investigation with a known invariant manifold $\mathcal{M}_{\mathbf{0}}$ of gradient flow. This restriction admits a uniform negative upper bound on the least eigenvalue of the Hessian map at all strict saddles in $\mathcal{M}_{\mathbf{0}}$. The bound depends on the size of the nonzero singular values and the separation between distinct singular values of the matrix to be factorized.

1 Introduction

Matrix factorization (MF) is a well known and important problem. Reflecting this, there are many approaches to posing and solving MF problems. Some of the best known include Positive Matrix Factorization (PMF) [Paatero and Tapper, 1994], Principal Component Analysis (PCA) [Hotelling, 1933, 1936, Jolliffe, 2011], Canonical Correlation Analysis (CCA) [Hardoon et al., 2004], Independent Component Analysis (ICA) [Hyvärinen and Oja, 2000], and Dictionary Learning [Mairal et al., 2009]. The MF problem, is inherently a non-convex problem over matrix arguments. It is typically solved using an iterative method (power iteration, gradient descent, etc.). Hence one is very interested in the landscape of the non-convex objective function being used.

One of the first results on the landscape of MF is due to Baldi and Hornik [1989]. This paper examined the landscape of training a one hidden layer rank k linear neural network and proved that all critical points are either global minima or (strict) saddle points. Further, the global minimum value is determined by the projection residual of the training data onto the subspace generated by the first k principal vectors of a covariance matrix associated with the training patterns. This connection to PCA was established by Bourlard and Kamp [1988]. As explained in [Baldi and Hornik, 1989], the training objective used in a linear neural network corresponds to an matrix factorization problem.

There is a recent surge of interest in the landscape of nonconvex objective functions. Li et al. [2019] examine the landscape problem under the lens of invariant groups and show that the parameter space can be divided into three regions: neighborhoods of all strict saddle points, neighborhoods of all global minima, and the complement of the first two regions. Ge et al. [2017a] study the matrix sensing, matrix completion and the robust PCA problems, all of which have a formulation as matrix factorization problems. Similarly, [Ge et al., 2015] considers the tensor decomposition problem, [Sun et al., 2018, Boumal, 2016] study the phase synchronization and retrieval problem and [Sun et al., 2015] examines the dictionary recovery problem.

Much effort has also been devoted to studying the landscape of neural networks under simplified assumptions [Kawaguchi, 2016, Nguyen and Hein, 2017, Hardt and Ma, 2016, Ge et al., 2017b].

Other papers have investigated the implicit constraints imposed by gradient flow in training over-parameterized models such as deep neural networks. For example, [Arora et al., 2018] shows that for over-parameterized multi-layer linear neural networks gradient flow implicitly balances the underlying factors, while [Du et al., 2018] extends this result to fully-connected and convolutional linear sections of neural networks.

This paper revisits the landscape of the MF problem and derives extended results in a contemporary format. This yields more intuitive proofs which offer the possibility of further generalization. Moreover, our results are applicable to both low rank and general-rank factorization. We additionally analyze how an invariance

Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS) 2020, Palermo, Italy. PMLR: Volume 108.

property of gradient flow impacts the strict saddles that can be encountered. Gradient flow is constrained to a manifold $\mathcal{M}_C$, where C is set by the initial condition. The choice $C = \mathbf{0}$ models the common practice of selecting random initialization close to the origin. For MF, once initialized in $\mathcal{M}_{\mathbf{0}}$, gradient flow ensures that the factors remain in $\mathcal{M}_{\mathbf{0}}$. We identify the set of critical points in $\mathcal{M}_{\mathbf{0}}$ and derive an negative upper-bound on the minimum eigenvalue of the Hessian map at all strict saddles. This negative upper bound depends on both the size and separation of the nonzero singular values of the matrix to be factorized. Hence for a given data matrix X, our results suggest that the ability of gradient flow to escape the neighborhood of strict saddles on $\mathcal{M}_{\mathbf{0}}$ depends on the size and the separation of distinct nonzero singular values of X. Due to the continuity, properties exhibited on $\mathcal{M}_{\mathbf{0}}$ will degrade gracefully as we move away from $\mathcal{M}_{\mathbf{0}}$.

Our new contributions arise from: (1) an orbit-based analysis of the problem and the identification of a natural canonical point on each orbit, (2) examining the natural invariant $\mathcal{M}_{\mathbf{0}}$, and (3) showing that the minimum eigenvalue of the Hessian map at all strict saddles on $\mathcal{M}_{\mathbf{0}}$ is uniformly bounded below zero. In addition, we use the natural setting of the problem, and avoid both symmetry assumptions and vectorization of the relevant differentials.

2 Preliminaries

For positive integers m, n, let $\mathbb{R}^{m\times n}$ denote the set of $m \times n$ real matrices, $\mathrm{GL}_n \subset \mathbb{R}^{n\times n}$ denote the general linear group of invertible $n \times n$ matrices, $\mathcal{O}_n \subset \mathrm{GL}_n$ denote the group of orthogonal $n \times n$ matrices, and for $k \leq m$, $\mathrm{St}_{m,k}$ denote the subset of $m \times k$ real matrices with orthonormal columns (the Stiefel manifold). For $A, B \in \mathbb{R}^{m\times n}$, $A_{:,k}$ denotes the k-th column of A, $\langle A, B\rangle$ denotes the standard inner product of A and B, and $\|A\|_F$ denotes the Frobenius norm of A.

Let $f\colon \mathbb{R}^{m\times n} \to \mathbb{R}$, with $X \mapsto f(X)$, be a twice continuously differentiable function. The derivative of f with respect to X evaluated at a given point X_0 is a linear map from $\mathbb{R}^{m\times n}$ to $\mathbb{R}$. Its action on $H \in \mathbb{R}^{m\times n}$, denoted by $D_X f(X_0)[H]$, satisfies

$$D_X f(X_0)[H] = \lim_{\alpha\to 0} \frac{f(X_0 + \alpha H) - f(X_0)}{\alpha}.$$

The gradient of f at X_0, denoted $\nabla_X f(X_0)$, is the unique point in $\mathbb{R}^{m\times n}$ such that

$$D_X f(X_0)[H] = <\nabla_X f(X_0), H>.$$

When no confusion is possible we omit the subscript X and simply write $Df(X_0)$ and $\nabla f(X_0)$.

The gradient function maps X to $\nabla f(X)$. The derivative of this function at X_0 is a linear map from $\mathbb{R}^{m\times n}$ to $\mathbb{R}^{m\times n}$. This linear map, and its action on $H \in \mathbb{R}^{m\times n}$, are denoted by $\nabla^2_X f(X_0)$ and $\nabla^2_X f(X_0)[H]$, respectively. When no confusion is possible we omit the subscript X and simply write $\nabla^2 f(X_0)$. The linear map $\nabla^2 f(X_0)$ corresponds to the Hessian matrix for functions from $\mathbb{R}^n$ to $\mathbb{R}$. The second derivative of f is

$$D^2_X f(X_0)[H, K] = <\nabla^2_X f(X_0)[K], H>. \tag{1}$$

We will always evaluate the second derivative with $K = H$. In this case, we simply write $D^2 f(X_0)[H]$. The second derivative is a scalar valued function from $\mathbb{R}^{m\times n}$ to $\mathbb{R}$. The linear function $\nabla^2_X f(X_0)[H]$ embedded in the second derivative brings in eigenvalues and eigenvectors associated with the second derivative.

A point X_0 is a critical point of f if $\nabla J(X_0) = \mathbf{0}$. The second order Taylor series of f about a critical point X_0 in the direction of H is

$$f(X_0 + tH) = f(X_0) + {}^1\!/_2 t^2 D^2 f(X_0)[H]. \tag{2}$$

If $\forall H,\ D^2 f(X_0)[H] > 0$, X_0 is a local minimum, if $\forall H,\ D^2 f(X_0)[H] < 0$, it is a local maximum, and if $\forall H,\ D^2 f(X_0)[H] \neq 0$ but is sign indefinite, it is a saddle point. However, if there exists nonzero H such that $D^2 f(X_0)[H] = 0$, the second derivative test is inconclusive, and X_0 is termed a degenerate critical point. The eigenvalues of the Hessian map $\nabla^2 f(X_0)$ are fundamental to making these classifications. Motivated by the observation that many non-convex optimization problems have degenerate critical points, the concepts of a strict saddle and a strict saddle function have been introduced [Ge et al., 2015]. A strict saddle is a critical point for which the Hessian map has at least one negative eigenvalue (this includes local maxima). For $\gamma > 0$, a γ-strict saddle is a critical point for which the least eigenvalue of its Hessian map is bounded above by $-\gamma$, i.e., $\lambda_{\min}(\nabla^2 f(X_0)) \leq -\gamma$.

2.1 Basic Matrix Factorization

In unconstrained matrix factorization, given $X \in \mathbb{R}^{m\times n}$ with $\mathrm{rank}(X) = r$, and a factorization dimension k, we seek $W \in \mathbb{R}^{m\times k}$ and $S \in \mathbb{R}^{k\times n}$ to minimize

$$J(W, S) = {}^1\!/_2 \, \|X - WS\|_F^2. \tag{3}$$

When $k \leq r$, this problem has a well known solution derived from any compact SVD $X = U\Sigma V^T$. Let U_k (resp. V_k) consist of the first k columns of U (resp. V) and Σ_k be the top left $k \times k$ submatrix of Σ. Then $(W, S) = (U_k\sqrt{\Sigma_k}, \sqrt{\Sigma_k}V_k^T)$ is a global minimum of (3). Since computing an SVD can become prohibitive for large matrices, we consider minimizing (3) using gradient descent methods. Hence we are interested in the local properties of the critical points of J.

The variable (W,S) lies in the product space $\mathcal{X} \triangleq \mathbb{R}^{m\times k} \times \mathbb{R}^{k\times n}$. The inner product on $\mathcal{X}$ is $<(G,H),(G',H')> = <G,G'> + <H,H'>$, with associated norm $\|(G,H)\|_F^2 = \|G\|_F^2 + \|H\|_F^2$. We will work extensively with the gradient, Hessian, and second derivative of J defined in (3). Setting $E \triangleq WS - S$, we list the equations for these functions below.

$$\nabla J(W,S) = \left(ES^T, \quad W^T E\right), \tag{4}$$

$$\nabla^2 J(W,S)[(G,H)] = (GSS^T + WHS^T + EH^T, \\ W^T WH + W^T GS + G^T E), \tag{5}$$

$$D^2 J(W,S)[(G,H)] = \|GS\|_F^2 + \|WH\|_F^2 \\ + 2\,\mathrm{trace}(H^T W^T GS + H^T G^T E), \tag{6}$$

A point (W,S) is a critical point of J if $\nabla J(W,S) = \mathbf{0}$. Using (4), this is equivalent to

$$(WS - X)S^T = \mathbf{0} \quad \text{and} \quad W^T(WS - X) = \mathbf{0}. \tag{7}$$

3 Landscape Analysis

We begin by introducing the orbits in $\mathcal{X}$ under GL_k and identifying a natural canonical point on each orbit. This facilitates our analysis of the landscape of J. Specifically, we prove that every critical point of J lies in an orbit of a canonical critical point under GL_k, and that all critical points on the orbit inherent important properties from the canonical point. Our analysis covers both $k \le r$ (low rank matrix factorization) and $k > r$ (an over parameterized linear neural network).

3.1 Orbits and Their Properties

For $A \in \mathrm{GL}_k$, let $L_A\colon \mathcal{X} \to \mathcal{X}$ denote the linear map

$$L_A\colon (G,H) \mapsto (GA, A^{-1}H). \tag{8}$$

Then for given $(W,S) \in \mathcal{X}$ let

$$\Theta(W,S) \triangleq \{L_A(W,S)\colon A \in \mathrm{GL}_k\} \qquad \text{and}$$
$$\mathcal{O}(W,S) \triangleq \{L_Q(W,S)\colon Q \in \mathcal{O}_k\}.$$

We call $\Theta(W,S)$ the orbit of (W,S) under GL_k, and $\mathcal{O}(W,S)$ the suborbit of (W,S) under $\mathcal{O}_k$. The value $J(W,S)$ is constant on each orbit. This has two important consequences. First, if any point on an orbit is a global minima, all points on the orbit are global minima. Second, the gradient $\nabla J(W,S)$ must be orthogonal to $\Theta(W,S)$ at (W,S). This is intuitively clear: moving along $\Theta(W,S)$ the value of J constant, but moving in the direction $\nabla J(W,S)$ yields the greatest change in J. To prove the orthogonality, take the derivative of $(WA, A^{-1}S)$ with respect to A, and then set $A = I_k$. This yields the set of tangents to $\Theta(W,S)$ at (W,S):

$$T_{W,S} = \{(WK, -KS)\colon K \in \mathbb{R}^{k\times k}\}. \tag{9}$$

Orthogonality is then verified by taking the inner product of any element in (9) with (4). Other quantities that are constant on $\Theta(W,S)$ include $\mathrm{rank}(W)$, $\mathrm{rank}(S)$, and $\mathrm{rank}(WS)$.

Qualitative properties of point neighborhoods are also "preserved" along the orbit. To see this let $(G,H) \in \mathcal{X}$ and $t > 0$. Then for any $A \in \mathrm{GL}_k$,

$$J(W+tG, S+tH) = J(WA+t(GA), A^{-1}S+t(A^{-1}H)).$$

So the values of J traced out by moving from (W,S) along a line in the direction of (G,H) are the same as the values traced out when moving from $(WA, A^{-1}S)$ along a line in the direction of $(GA, A^{-1}H)$. The following lemma formalizes these observations.

Lemma 3.1. *For all $(W,S) \in \mathcal{X}$ and $A \in \mathrm{GL}_k$:*

(a) $\nabla J(L_A(W,S)) = L_{A^{-T}} \nabla J(W,S)$,

(b) $DJ(L_A(W,S))[L_A(G,H)] = DJ(W,S)[(G,H)]$,

(c) $\nabla^2 J(L_A(W,S))[L_A(G,H)] = L_{A^{-T}}(\nabla^2 J(W,S)[(G,H)])$,

(d) $D^2 J(L_A(W,S))[L_A(G,H)] = D^2 J(W,S)[(G,H)]$.

Proof. Expand of each side of the stated equality. See the supplementary material for the full proof. □

We use Lemma 3.1 to prove additional orbit properties.

Theorem 3.1. *Let $(W,S) \in \mathcal{X}$ and $A \in \mathrm{GL}_k$. Then*

(a) If (W,S) is a global minimum, so is every point in $\Theta(W,S)$.

(b) If (W,S) is a critical point (resp. strict saddle), so is every point in $\Theta(W,S)$.

(c) The eigenvalues of $\nabla^2 J(W,S)$ are invariant in $\mathcal{O}(W,S)$.

Proof. (a) By (3), $J(W,S) = J(WA, A^{-1}S)$. So if (W,S) is a global minimum, so is $L_A(W,S)$.
(b) By Lemma 3.1 part (a) and the linearity of L_A, if $\nabla J(W,S) = \mathbf{0}$, then $\nabla J(L_A(W,S)) = \mathbf{0}$. So if (W,S) is a critical point, so is every point on $\Theta(W,S)$. Now let (W,S) be a strict saddle and the minimum eigenvalue of $\nabla^2 J(W,S)$ be $-\lambda < 0$. Then for some unit norm (G,H), $<\nabla^2(W,S)[(G,H)], (G,H)> = -\lambda$. By Lemma 3.1 part (d), $D^2 J(L_A(W,S))[L_A(G,H)] =$

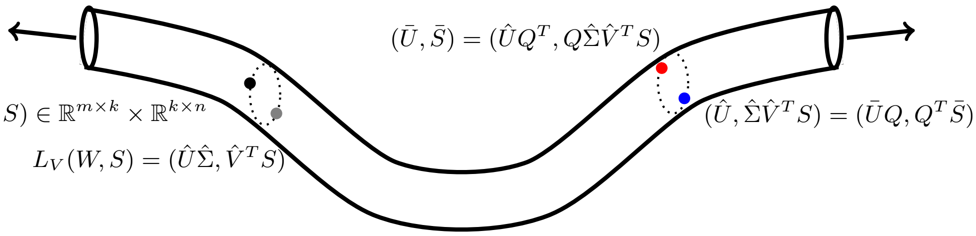

Figure 1: A conceptual view of the orbit $\Theta(W, S)$. (W, S) is a critical point with $\text{rank}(W) = \min\{k, m\}$ and compact SVD $W = \hat{U}\hat{\Sigma}\hat{V}^T$. $L_{\hat{V}}(W, S) = (\hat{U}\hat{\Sigma}, \hat{V}^T S)$ is a critical point on the same sub-orbit under $\mathcal{O}_k$. This point is can be transported along the orbit $\Theta(W, S)$ using $L_{\Sigma^{-1}}$ to $(\hat{U}, \hat{\Sigma}\hat{V}^T S)$. There exists an orthogonal Q that $\hat{U} = \bar{U}Q$ where $\bar{U}$ formed from k orthonormal eigenvectors of XX^T. This gives a canonical point $(\bar{U}, \Lambda\bar{V}^T)$ on the same sub-orbit under $\mathcal{O}_k$ as $(\hat{U}, \hat{\Sigma}\hat{V}^T S)$.

$D^2(W, S)[(G, H)] = -\lambda$. Thus $\nabla^2 J(L_A(W, S))$ has a negative eigenvalue. So if (W, S) is a strict saddle, so is every point on $\Theta(W, S)$.
(c) Consider an eigenvector $(G, H) \in \mathcal{X}$ of $\nabla^2 J(W, S)$ with eigenvalue λ. By Lemma 3.1 part (c),

$$\nabla^2 J(L_A(W, S))[L_A(G, H)] = \lambda L_{A^{-T}}((G, H)) = \lambda(GA^{-T}, A^T H).$$

If $A \in \mathcal{O}_k$ then $A^{-1} = A^T$. Hence $\nabla^2 J(W, S)[L_A(G, H)] = \lambda L_A((G, H))$. Thus λ is also an eigenvector of $\nabla^2 J(WA, A^{-1}S)$. A symmetric argument proves the converse also holds. ☐

3.2 Canonical Points

It will be convenient to identify a canonical point on each orbit of critical points. The existence of this point and its properties will be used to prove results for all critical points in the same orbit. The canonical point itself need not be computed. We construct the canonical points using the left singular vectors of X. This exploits the known connection to PCA (and SVD).

WLOG we assume $m < n$ and consider the $m \times m$ matrix XX^T. This has r positive eigenvalues and $(m-r)$ zero eigenvalues. Denote these by

$$\sigma_1^2 \geq \sigma_2^2 \geq \cdots \sigma_r^2 > 0 \quad \sigma_{r+1}^2 = \cdots = \sigma_m^2 = 0. \tag{10}$$

Let $u_1, \dots, u_m$ denote a set of m corresponding orthonormal eigenvectors with $XX^T u_j = \sigma_j^2 u_j$ for $j \in [1:r]$, and $XX^T u_j = \mathbf{0}$ for $j \in [r+1:m]$. The first r orthonormal eigenvectors need not be unique since the nonzero eigenvalues may be repeated. Place the eigenvectors in the columns of $U \in \mathbb{R}^{m\times m}$. This basis can then used to form a compatible full SVD

$$X = U\Sigma V^T = \textstyle\sum_{i=1}^m \sigma_i u_i v_i^T. \tag{11}$$

Fix a factorization dimension k and select an integer

$$q \leq \min\{k, m\}.$$

Then place a subset of q eigenvalues of XX^T, denoted by $\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_q^2$, in decreasing order in a diagonal matrix $\Lambda^2 \in \mathbb{R}^{q\times q}$. Place the corresponding set of q orthonormal eigenvectors of XX^T into the columns of $\bar{U} \in \mathbb{R}^{m\times q}$ such that $XX^T\bar{U} = \bar{U}\Lambda^2$. Let $\bar{V}$ denote the matrix of corresponding columns of V. So

$$\bar{U}^T X = \Lambda\bar{V}^T \quad \text{and} \quad X\bar{V} = \bar{U}\Lambda.$$

Select $S_2 \in \mathbb{R}^{(k-q)\times n}$ to satisfy

$$(\bar{U}\bar{U}^T - I)XS_2^T = \mathbf{0}. \tag{12}$$

By Lemma 3.2 below, S_2^T has the form

$$S_2^T = \bar{V}\bar{C} + V_0 C_0,$$

where $V_0 = [v_{r+1}, \dots, v_n]$, $\bar{C} \in \mathbb{R}^{q\times(k-q)}$, and $C_0 \in \mathbb{R}^{(n-r)\times(k-q)}$. It follows that the set of S_2 satisfying (12) contains nonzero elements and is unbounded.

Lemma 3.2. *If S_2 satisfies* (12)*, then:*

(a) $S_2^T = \bar{V}\bar{C} + V_0 C_0$ for some $\bar{C} \in \mathbb{R}^{q\times(k-q)}$, $C_0 \in \mathbb{R}^{(n-r)\times(k-q)}$.

(b) If Λ^2 is invertible, $\bar{C}$ and C_0 are unique and $S_2 S_2^T = \bar{C}^T\bar{C} + C_0^T C_0$.

Proof. (i) (12) implies $\bar{U}\bar{U}^T XS_2^T = XS_2^T$. Hence for some matrix $\bar{B} = [b_{i,j}] \in \mathbb{R}^{q\times(k-q)}$, $XS_2 = \bar{U}\bar{B}$. Since $\bar{U}$ has orthonormal columns, $\bar{B}$ is unique.
(ii) Using part (i) and the SVD of X, $U\Sigma V^T S_2^T = \bar{U}B$. For any $1 \leq j \leq r$, let u_j be an eigenvector that is not a column of $\bar{U}$ and v_j denote the corresponding column of V. The condition $1 \leq j \leq r$, ensures $\sigma_j^2 > 0$. Multiplying both sides of the previous equation by u_j^T yields $\sigma_j v_j^T S_2^T = u_j^T \bar{U}B = \mathbf{0}$. Hence $v_j^T S_2^T = \mathbf{0}$.
(iii) By part (ii), for every v_j that is not a column in $\bar{V}$ and its corresponding singular value is nonzero, $v_j^T S_2^T = 0$. Thus the columns of S_2^T must lie in the span of the v_j that are either columns of $\bar{V}$ or have zero eigenvalues. Hence there exist $\bar{C}, C_0$ such that $S_2^T = \bar{V}\bar{C} + V_0 C_0$. This proves (a).

(iv) The assumption that Λ^2 is invertible excludes v_j with zero eigenvalues from the columns of $\bar{V}$. In this case, the combined columns of $\bar{V}$ and V_0 form an orthonormal set. Hence for a given S_2, $\bar{C}$ and C_0 are unique. The orthonormality of the columns in $\bar{V}$ and V_0 ensures $S_2 S_2^T = \bar{C}^T\bar{C} + C_0^T C_0$. This proves (b). □

Finally, we form the canonical point (W, S) as

$$(W, S) = \left(\begin{bmatrix} \bar{U} & \mathbf{0}_{m\times(k-q)} \end{bmatrix}, \begin{bmatrix} \Lambda\bar{V}^T \\ \bar{C}^T\bar{V}^T + C_0^T V_0^T \end{bmatrix} \right). \quad (13)$$

Theorem 3.2. *For $1 \le q \le \min\{k, m\}$, (W, S) specified by (13) is a critical point with*

$$J(W, S) = 1/2 \left(\sum_{i=1}^{r} \sigma_i^2 - \sum_{j=1}^{q} \lambda_j^2 \right).$$

Conversely, if (W', S') is a critical point with $\operatorname{rank}(W') = q$, then there exists a critical point (W, S) of the form (13) and $A \in \mathrm{GL}_k$ such that $(W', S') = L_A(W, S)$. □

Proof. (a) By construction W has rank q. Moreover, (W, S) satisfies (7) and is hence a critical point:

$$(WS - X)S^T = (\bar{U}\Lambda\bar{V}^T - U\Sigma V^T)[\bar{V}\Lambda, \bar{V}\bar{C} + V_0 C_0] = \mathbf{0},$$

$$W^T(WS - X) = \begin{bmatrix} \bar{U}^T \\ \mathbf{0}_{(k-q)\times m} \end{bmatrix} (\bar{U}\Lambda\bar{V}^T - U\Sigma V^T) = \mathbf{0}.$$

In addition, $J(W, S) = 1/2 \|\bar{U}\Lambda\bar{V}^T - U\Sigma V^T\|_F^2 = 1/2 \left(\sum_{i=1}^r \sigma_i^2 - \sum_{j=1}^q \lambda_j^2 \right)$.

(b) Since (W', S') is a rank q critical point, there is a permutation matrix $P \in \mathrm{GL}_k$ such that the first q columns of $W'P$ are linearly independent. Let $\hat{W}$ denote the matrix of the first q columns of $W'P$. Hence

$$W' = \begin{bmatrix} \hat{W} & \mathbf{0}_{m\times(k-q)} \end{bmatrix} \begin{bmatrix} I_q & F \\ \mathbf{0}_{(k-q)\times q} & I_{(k-q)} \end{bmatrix} P^T, \quad (14)$$

where F is determined by the last $k - q$ columns of $W'P$. Now let $\hat{U}\hat{\Sigma}\hat{V}^T$ be a compact SVD of $\hat{W}$. Noting that $\hat{U} \in \mathrm{St}_{m,q}$ and $\hat{\Sigma}, \hat{V} \in \mathrm{GL}_q$, modify (14) to

$$W' = \begin{bmatrix} \hat{U} & \mathbf{0}_{m\times(k-q)} \end{bmatrix} \begin{bmatrix} \hat{\Sigma}\hat{V}^T & \hat{\Sigma}\hat{V}^T F \\ \mathbf{0}_{(k-q)\times q} & I_{(k-q)} \end{bmatrix} P^T, \quad (15)$$

Let C denote the product of the two rightmost matrices in (15), and $\tilde{W} = \begin{bmatrix} \hat{U} & \mathbf{0}_{m\times(k-q)} \end{bmatrix}$. Then $C \in \mathrm{GL}_k$ and by (15), $\tilde{W} = W'C^{-1}$. Setting $\tilde{S} = CS'$, and using Theorem 3.1, we see that $(\tilde{W}, \tilde{S})$ is a critical point of J. Hence $(\tilde{W}, \tilde{S})$ must satisfy (7). Write $\tilde{S}^T = \begin{bmatrix} \tilde{S}_1^T & \tilde{S}_2^T \end{bmatrix}$. Then by (7),

$$(\hat{U}\tilde{S}_1 - X)\begin{bmatrix} \tilde{S}_1^T & \tilde{S}_2^T \end{bmatrix} = \mathbf{0} \quad \text{and}$$

$$\begin{bmatrix} \hat{U}^T \\ \mathbf{0}_{(k-q)\times m} \end{bmatrix} [\hat{U}\tilde{S}_1 - X] = \mathbf{0}.$$

The second condition implies $\tilde{S}_1 = \hat{U}^T X$. Then the first implies: (i) $(\hat{U}\hat{U}^T - I)XX^T\hat{U} = \mathbf{0}$ and (ii) $(\hat{U}\hat{U}^T - I)X\tilde{S}_2^T = \mathbf{0}$. Condition (i) shows that the image of the range of $\hat{U}$ under XX^T is contained in the range of $\hat{U}$, i.e., $XX^T\mathcal{R}(\hat{U}) \subseteq \mathcal{R}(\hat{U})$. Since the dimension of $\mathcal{R}(\hat{U})$ is q, there exist q orthonormal eigenvectors of XX^T that form a basis for $\mathcal{R}(\hat{U})$. Let $\bar{U}$ be the matrix with the elements of this basis as its columns. Every column of $\hat{U}$ has a representation in this basis. So for some $Q \in \mathrm{GL}_q$, $\hat{U} = \bar{U}Q$. In addition, $I_k = \hat{U}^T\hat{U} = Q^T Q$. Hence $Q \in \mathcal{O}_k$. We can use this to rewrite (15) as

$$W' = \begin{bmatrix} \bar{U} & \mathbf{0}_{m\times(k-q)} \end{bmatrix} \begin{bmatrix} Q & \mathbf{0} \\ \mathbf{0}_{(k-q)\times q} & I_{(k-q)} \end{bmatrix} \begin{bmatrix} \hat{\Sigma}\hat{V}^T & \hat{\Sigma}\hat{V}^T F \\ \mathbf{0}_{(k-q)\times q} & I_{(k-q)} \end{bmatrix} P^T, \quad (16)$$

Denote the center matrix in (16) by D. Then we have $W'(DC)^{-1} = \begin{bmatrix} \bar{U} & \mathbf{0}_{m\times(k-q)} \end{bmatrix}$ and

$$(DC)S' = D\begin{bmatrix} \tilde{S}_1 \\ \tilde{S}_2 \end{bmatrix} = \begin{bmatrix} Q\hat{U}^T X \\ \tilde{S}_2 \end{bmatrix} = \begin{bmatrix} \bar{U}^T X \\ \tilde{S}_2 \end{bmatrix} = \begin{bmatrix} \Lambda\bar{V}^T \\ \tilde{S}_2 \end{bmatrix}.$$

So $(W'(DC)^{-1}, (WD)S')$ is a critical point and by construction, $XX^T\bar{U} = \bar{U}\Lambda^2$ for diagonal Λ^2, and by (ii) $(\bar{U}\bar{U}^T - I)X\tilde{S}_2^T = \mathbf{0}$. By Lemma 3.2, $\tilde{S}_2^T$ takes the form $\tilde{S}_2^T = \bar{V}\bar{C} + V_0 C_0$ where $V_0 = [v_{r+1}, \ldots, v_n]$, $\bar{C} \in \mathbb{R}^{q\times(k-q)}$, and $C_0 \in \mathbb{R}^{(n-r)\times(k-q)}$. □

The key result of Theorem 3.2 is that every critical point with W of positive rank lies on the orbit of a canonical point. This canonical point can be used to reason about all points on its orbit. See Fig. 1.

The results and proof of Theorem 3.2 simplify when $q = k \le m$. In this situation, there is no need for the term S_2 and the condition (12). The additional complexity when $q < k$ arises because W can then have a nontrivial null space. In that case, S can be decomposed into the sum of a term S_1 in $\mathcal{N}(W)^\perp$ and a term S_2 in $\mathcal{N}(W)$. The term S_2 is redundant since it does not impact the value of J, but we need to account for its possible presence.

Using (13) we can write a canonical point in the form

$$(W, S) = (W, S_{\bar{C}} + Z_0), \quad (17)$$

where

$$W = \begin{bmatrix} \bar{U} & \mathbf{0}_{m\times(k-q)} \end{bmatrix}, S_{\bar{C}} = \begin{bmatrix} \Lambda\bar{V}^T \\ \bar{C}^T\bar{V}^T \end{bmatrix}, Z_0 = \begin{bmatrix} \mathbf{0} \\ C_0^T V_0^T \end{bmatrix}.$$

When $\bar{C} = \mathbf{0}$, denote $S_{\bar{C}}$ by $S_{\mathbf{0}}$. We now show that for fixed $\bar{U}$ and Z_0, the family of canonical points (17) as $\bar{C}$ ranges over $\mathbb{R}^{(k-q)\times n}$ are all on the orbit $\Theta(W, S_{\mathbf{0}} + Z_0)$.

Lemma 3.3. *For each canonical point of the form* (17)*,* $\Theta(W, S_{\bar{C}} + Z_0) = \Theta(W, S_{\mathbf{0}} + Z_0)$.

Proof. It is enough to show that there exists an $A \in \mathrm{GL}_k$ such that $L_A(W, S_{\mathbf{0}} + Z_0) = (W, S_{\bar{C}} + Z_0)$. Then any point in the orbit of $(W, S_{\bar{C}} + Z_0)$ is in the orbit of $(W, S_{\mathbf{0}} + Z_0)$. Similarly, using $L_{A^{-1}}$ one sees that any point in the orbit of $(W, S_{\mathbf{0}} + Z_0)$ is in the orbit of $(W, S_{\bar{C}} + Z_0)$.

The matrix $A = \begin{bmatrix} I_q & \mathbf{0}_{(k-q)\times q} \\ -\bar{C}^T\Lambda^{-1} & I_{k-q} \end{bmatrix}$ has inverse $A^{-1} = \begin{bmatrix} I_q & \mathbf{0}_{(k-q)\times q} \\ -\bar{C}^T\Lambda^{-1} & I_{k-q} \end{bmatrix}$. Hence $A \in \mathrm{GL}_k$, and

$$\begin{bmatrix} \bar{U} & \mathbf{0} \end{bmatrix} A = \begin{bmatrix} \bar{U} & \mathbf{0} \end{bmatrix},$$
$$A^{-1}\left(\begin{bmatrix} \Lambda\bar{V}^T \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ C_0^T V_0^T \end{bmatrix}\right) = \begin{bmatrix} \Lambda\bar{V}^T \\ \bar{C}^T\bar{V}^T \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ C_0^T V_0^T \end{bmatrix}.$$

Hence for fixed $\bar{U}$ and Z_0, all points of the form (17) are on the same orbit. ☐

Lemma 3.3 allows us to make a useful simplification. Since the canonical points are representative elements of orbits, we need only consider canonical points of the form $(W, S_{\mathbf{0}} + Z_0)$, i.e., $\bar{C} = \mathbf{0}$. So, henceforth we only consider canonical points of the form

$$(W, S) = \left(\begin{bmatrix} \bar{U} & \mathbf{0}_{m\times(k-q)} \end{bmatrix}, \begin{bmatrix} \Lambda\bar{V}^T \\ C_0^T V_0^T \end{bmatrix}\right). \tag{18}$$

We say that the critical point (18) is maximal if $\lambda_i^2 = \sigma_i^2$, $i \in [1:q]$. So (W, S) is maximal if and only if the q columns of $\bar{U}$ are eigenvectors of XX^T for a set of its q largest eigenvalues. If this does not hold, then there exists a least integer $p \in [1:q]$ such that $\lambda_p < \sigma_p$. In this case we say that is (W, S) not maximal at p.

Since every canonical point (18) is a critical point, it is either a global minimum or a strict saddle. For the strict saddles one can derive an expression for $\lambda_{\min}(\nabla^2 J(W, S))$ and use this to show that over the orbit of a canonical strict saddle $\lambda_{\min}(\nabla^2 J(WA, A^{-1}S))$ is negative but may not uniformly bounded below zero [Valavi et al., 2020]. To address this issue, in the next section we exploit a known manifold of balanced factorizations to show that there exists $\gamma > 0$ such that $\lambda_{\min}(\nabla^2 J(W, S)) \le -\gamma$ for every strict saddle (W, S) on the manifold.

3.3 The Manifold $\mathbf{W^T W - S S^T = 0}$

Let $\mathcal{M}_C \triangleq \{(W, S)\colon WW^T - SS^T = C\}$, where $C \in \mathbb{R}^{k\times k}$ is a symmetric matrix. These manifolds are interesting for several reasons. First, the factorization problem permits imbalance between W and S in the sense that by selecting $A \in \mathrm{GL}_k$, one can make WA very large (resp. small) while making $A^{-1}S$ very small (resp. large) without changing the value of the objective. However, if $(W, S) \in \mathcal{M}_C$, the difference between the norms of W and S is bounded: $\|W\|_F^2 - \|S\|_F^2 = \mathrm{trace}(C)$. In particular, if $C = \mathbf{0}$, $\|W\|_F^2 = \|S\|_F^2$. In the neural network literature this is referred to as a balance condition [Arora et al., 2018, Du et al., 2018]. Second, the term $C_0^T V_0^T$ in (18) is redundant and we have no a priori bound on its value. We will show that if the search for a global minimum is constrained to $\mathcal{M}_{\mathbf{0}}$, then $C_0 = \mathbf{0}$. Third, it is known that the manifold $\mathcal{M}_C$ is invariant under gradient flow. An initial value for (W, S) specifies C, and the gradient flow o.d.e.

$$\frac{d}{dt}(W_t, S_t) = -\nabla J(W_t, S_t), \quad (W_0, S_0) \in \mathcal{X}, \tag{19}$$

then ensures $(W_t, S_t) \in \mathcal{M}_C$ for $t \ge 0$ [Arora et al., 2018, Theorem 1], [Du et al., 2018, Lemma 3.1]. Lemma 3.4 gives a self-contained proof of this result.

Lemma 3.4. *Along every solution of the gradient flow o.d.e.,* $W_t^T W_t - S_t S_t^T$ *is a constant symmetric* $k \times k$ *matrix and* $\|W_t\|_F^2 - \|S_t\|_F^2$ *is a constant.*

Proof. Taking the inner product of both sides of (19) with $T_{W,S}(H)$ in (9), and using $<T_{W,S}(H), \nabla J(W, S)> = 0$, yields $H^T W_t^T \dot{W}_t - S_t^T H^T \dot{S}_t = <H, W_t^T \dot{W}_t - \dot{S}_t S_t^T> = \mathbf{0}$, $\forall H \in \mathbb{R}^{k\times k}$. Thus $W_t^T \dot{W}_t - \dot{S}_t S_t^T = \mathbf{0}$. Adding this to its transpose gives $\frac{d}{dt}(W_t^T W_t - S_t S_t^T) = \mathbf{0}$. Thus $W_t^T W_t - S_t S_t^T = W_0^T W_0 - S_0 S_0^T$. $\|W_t\|_F^2 - \|S_t\|_F^2 = \mathrm{trace}(W_t^T W_t) - \mathrm{trace}(S_t^T S_t) = \mathrm{trace}(W_t^T W_t - S_t S_t^T)$, and $\mathrm{trace}(W_t^T W_t - S_t S_t^T)$ is a constant. ☐

Motivated by the discussion above, we focus on the critical points in $\mathcal{M}_{\mathbf{0}}$. The choice $\mathcal{M}_{\mathbf{0}}$ reflects an idealization of the common practice of selecting random initializations close to the origin. Gradient flow then ensures $(W_t, S_t) \in \mathcal{M}_{\mathbf{0}}$ thereafter. Note that $\mathcal{M}_{\mathbf{0}}$ is also natural choice since it provides perfect balance between $\|W\|_F$ and $\|S\|_F$. In addition, properties in $\mathcal{M}_{\mathbf{0}}$ will degrade gracefully as we move away from $\mathcal{M}_{\mathbf{0}}$. This is assured by continuity of the functions involved w.r.t. the matrix parameter C.

Each critical point (W, S) of J is on the orbit of some canonical point. Hence this holds for critical points in $\mathcal{M}_{\mathbf{0}}$. Below, we identify these canonical points by showing that the orbit of a general canonical point (18) intersects the manifold $\mathcal{M}_{\mathbf{0}}$ if and only if Λ^2 is invertible and $C_0 = \mathbf{0}$.

Theorem 3.3. *Fix* $\bar{U} \in \mathrm{St}_{m,q}$ *with* $XX^T\bar{U} = \bar{U}\Lambda^2$ *and consider the family of canonical points* (W, S) *of the form* (18)*. Then* *(a) There exists $A \in \mathrm{GL}_k$ such that $L_A(W,S) \in \mathcal{M}_{\mathbf{0}}$ if and only if Λ^2 is invertible and $C_0 = \mathbf{0}$.*

(b) If $L_A(W,S) \in \mathcal{M}_{\mathbf{0}}$, then for each $Q \in \mathcal{O}_k$, $L_{AQ}(W,S) \in \mathcal{M}_{\mathbf{0}}$.

Proof. Let (W,S) be a canonical point of the form (18). First note that there exists $A \in \mathrm{GL}_k$ such that $L_A(W,S) \in \mathcal{M}_{\mathbf{0}}$ if and only if $AA^T(W^TW)AA^T = SS^T$. This follows by noting that $A^TW^TWA = A^{-1}SS^TA^{-T}$ if and only if the above condition holds.

(If) Assume Λ^2 is invertible and $C_0 = \mathbf{0}$. Since $C_0 = \mathbf{0}$,

$$W^TW = \begin{bmatrix} I_q & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \text{and} \quad SS^T = \begin{bmatrix} \Lambda^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

For $A \in \mathrm{GL}_k$, let $R = AA^T$. Note that $R \succ \mathbf{0}$. Write

$$R = \begin{bmatrix} R_1 & R_3 \\ R_3^T & R_2 \end{bmatrix}. \tag{20}$$

We seek the set of symmetric positive definite R with

$$R \begin{bmatrix} I_q & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} R = \begin{bmatrix} R_1^2 & R_1R_3 \\ R_3^TR_1 & R_3^TR_3 \end{bmatrix} = \begin{bmatrix} \Lambda^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{21}$$

Clearly we must have $R_1^2 = \Lambda^2$, $R_1R_3 = \mathbf{0}$, and $R_3^TR_3 = \mathbf{0}$. The third requirement implies $R_3 = \mathbf{0}$, and this ensures the second requirement is also satisfied. Since Λ^2 is diagonal and positive definite, the first requirement gives $R_1 = \Lambda$. Hence the set of solutions for R, and the corresponding solutions for A are

$$\left\{ R = \begin{bmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & R_2 \end{bmatrix} : R_2 \succ \mathbf{0} \right\},$$
$$\left\{ A = \begin{bmatrix} \sqrt{\Lambda} & \mathbf{0} \\ \mathbf{0} & \sqrt{R_2} \end{bmatrix} Q \colon Q \in \mathcal{O}_k \right\}.$$

Thus there exists $A \in \mathrm{GL}_k$ with $L_A(W,S) \in \mathcal{M}_{\mathbf{0}}$. Note that $WA = [\bar{U}, \mathbf{0}]A = [\bar{U}\Lambda^{1/2}, \mathbf{0}]Q$ and $A^{-1}S = Q^T \begin{bmatrix} \Lambda^{1/2}\bar{V}^T \\ \mathbf{0} \end{bmatrix}$. So the free parameter R_2 plays no role in determining $L_A(W,S)$. For each solution A, AQ is also a solution for every $Q \in \mathcal{O}_k$. This proves (b).

(Only If) In general,

$$W^TW = \begin{bmatrix} I_q & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \text{and} \quad SS^T = \begin{bmatrix} \Lambda^2 & \mathbf{0} \\ \mathbf{0} & C_0^TC_0 \end{bmatrix}.$$

For some $A \in \mathrm{GL}_k$, let $AA^T(W^TW)AA^T = SS^T$. Let $R = AA^T$. So $R \succ \mathbf{0}$. Write R in the form (20). Then R satisfies,

$$R \begin{bmatrix} I_q & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} R = \begin{bmatrix} R_1^2 & R_1R_3 \\ R_3^TR_1 & R_3^TR_3 \end{bmatrix} = \begin{bmatrix} \Lambda^2 & \mathbf{0} \\ \mathbf{0} & C_0^TC_0 \end{bmatrix}. \tag{22}$$

It follows that $R_1^2 = \Lambda^2$, $R_1R_3 = \mathbf{0}$, and $R_3^TR_3 = C_0^TC_0$. Since Λ^2 is diagonal with nonzero diagonal entries, the first condition gives $R_1 = \Lambda$. Since $R \succ \mathbf{0}$, $R_1 \succ \mathbf{0}$, and hence $\Lambda \in \mathrm{GL}_q$. The second condition gives $R_3 = \mathbf{0}$, and the third implies $C_0^TC_0 = \mathbf{0}$. □

We henceforth assume that (i) $C_0 = \mathbf{0}$ and (ii) Λ^2 is invertible. Note that (ii) requires $q \le \min\{k, r\}$. Under (ii) we can apply $L_{\Lambda^{1/2}}$ to map the canonical point in (18) (with $C_0 = \mathbf{0}$) to $\mathcal{M}_{\mathbf{0}}$. Consequently, the following corollary holds.

Corollary 3.1. *Fix $\bar{U} \in \mathrm{St}_{m,q}$ with $XX^T\bar{U} = \bar{U}\Lambda^2$ and $\Lambda \in \mathrm{GL}_q$. Then the set of critical points on $\mathcal{M}_{\mathbf{0}}$ is*

$$\mathcal{O}\left(\begin{bmatrix} \bar{U}\Lambda^{1/2} & \mathbf{0}_{m\times(k-q)} \end{bmatrix}, \begin{bmatrix} \Lambda^{1/2}\bar{V}^T \\ \mathbf{0}_{(k-q)\times n} \end{bmatrix} \right). \tag{23}$$

By Theorem 3.2, the behavior around these points is completely determined by the behavior around

$$(W,S) = \left(\begin{bmatrix} \bar{U}\Lambda^{1/2} & \mathbf{0}_{m\times(k-q)} \end{bmatrix}, \begin{bmatrix} \Lambda^{1/2}\bar{V}^T \\ \mathbf{0}_{(k-q)\times n} \end{bmatrix} \right). \tag{24}$$

Hence we only need to examine such critical points. Recall that (W,S) in (24) is not maximal at p if there exists $p \in [1\colon q]$ such that $\lambda_p < \sigma_p$. Alternatively, if $\lambda_i^2 = \sigma_i^2$, $i \in [1\colon q]$, then we say that (W,S) is maximal. Since $\Lambda^2 \in \mathrm{GL}_q$, (W,S) is always maximal if $q = r$.

Theorem 3.4. *For the critical point (W,S) in (24).*

(a) For $q = \min\{k, r\}$, if (W,S) is maximal, it is a global minimum. If it is not maximal at p, $q = k$ and (W,S) is a strict saddle with

$$\lambda_{\min}(\nabla^2 J(W,S)) \le -(\sigma_p - \lambda_k).$$

(b) For $q < \min\{k, r\}$, (W,S) is a strict saddle with

$$\lambda_{\min}(\nabla^2 J(W,S)) \le \begin{cases} -\sigma_{q+1}, & (W,S) \text{ is maximal;} \\ -\sigma_p, & (W,S) \text{ not maximal at } p. \end{cases}$$

Proof. (a) (i) Assume $q = \min\{k, r\}$ and (W,S) is maximal. So the diagonal of $\Lambda^2 \in \mathbb{R}^{q\times q}$ consists of a set of $q \le r$ largest eigenvalues of XX^T. Hence $J(W,S) = {}^1\!/\!_2\|\bar{U}\Lambda\bar{V}^T - U\Sigma V^T\|_F^2 = {}^1\!/\!_2(\sum_{i=1}^r \sigma_i^2 - \sum_{j=1}^q \lambda_j^2) = \sum_{j=q+1}^r \sigma_j^2$. If $q = k$, this is the least loss attainable for k; if $q = r$ the loss is zero and hence minimal. In both cases (W,S) is a global minimum.

(ii) Now let $q = \min\{k, r\}$ with (W,S) not maximal at p. If $q = r \le k$, then $q = \min\{k, r\}$ and (W,S) is maximal. Hence $q = k < r$, and p is the least integer for which $\lambda_p < \sigma_p$. Let $G_{i,j} = u_i\mathbf{e}_j^T \in \mathbb{R}^{m\times k}$ and $H_{j,i} = \mathbf{e}_jv_i^T \in \mathbb{R}^{k\times n}$, where $\mathbf{e}_j$ denotes the j-th standard basis. Consider the action of $D^2J(W,S)$ on

$(G_{p,k}, H_{k,p})$. Using (24) we find

$$
\begin{aligned}
G_{p,k}S &= u_p \mathbf{e}_k^T S = \lambda_k^{1/2} u_p \bar{v}_k^T,\\
WH_{k,p} &= W\mathbf{e}_k v_p^T = \lambda_k^{1/2} \bar{u}_k v_p^T,\\
(WH_{k,p})^T G_{p,k} S &= \lambda_k^{1/2} v_p \bar{u}_k^T \lambda_k^{1/2} u_p \bar{v}_k^T = 0,\\
(G_{p,k}H_{k,p})^T E &= v_p u_p^T (WS - X) = -\sigma_p v_p v_p^T.
\end{aligned}
$$

In the third equation we used $\bar{u}_k \perp u_p$. This holds since the λ_j^2 are listed in decreasing order and σ_p^2 is missing in location p. Hence it can't occur in location $k \geq p$. Thus we have $\|G_{p,k}S\|_F^2 + \|WH_{k,p}\|_F^2 = 2\lambda_k$, and $2\,\mathrm{trace}((G_{p,k}H_{k,p})^T E) = -2\sigma_p$. Substituting these results into (6) yields $D^2J[(G_{p,k}, H_{k,p})] = 2(\lambda_k - \sigma_p)$. Normalizing this result by the squared norm of $(G_{p,k}, H_{k,p})$ one obtains $\lambda_{\min}(\nabla^2 J(W,S)) \leq -(\sigma_p - \lambda_k)$ with $\sigma_p > \lambda_k$.

(b) We consider three cases:
(i) If $q = 0$, then $(W,S) = (\mathbf{0},\mathbf{0})$. Consider the action of $D^2J(\mathbf{0},\mathbf{0})$ on $(G_{1,1}, H_{1,1})$. Using (6),

$$
\begin{aligned}
D^2J(\mathbf{0},\mathbf{0})[G_{1,1}, H_{1,1}] &= -2\,\mathrm{trace}(\textstyle\sum_{i=1}^r \sigma_i v_i u_i^T u_1 v_1^T)\\
&= -2\sigma_1.
\end{aligned}
$$

Normalizing by the squared norm of $(G_{1,1}, H_{1,1})$ gives

$$
\lambda_{\min}(\nabla^2 J(\mathbf{0},\mathbf{0})) \leq -\sigma_1. \tag{25}
$$

(ii) Assume $1 \leq q < \min\{k, r-1\}$ and let $p \in [1\!:\!q]$, be the least p with $\lambda_p < \sigma_p$. Hence (W,S) is not maximal at p. Consider the action of $D^2J(W,S)$ on

$$
(G_{p,q+1}, H_{q+1,p}) = (u_p \mathbf{e}_{q+1}^T, \mathbf{e}_{q+1} v_p^T). \tag{26}
$$

Using (26) and (24), $G_{p,q+1}S = u_p \mathbf{e}_{q+1}^T S = \mathbf{0}$, and $WH_{q+1,p} = W\mathbf{e}_{q+1} v_p^T = \mathbf{0}$. Thus $\|G_{p,q+1}S\|_F^2 = 0$, $\|WH_{q+1,p}\|_F^2 = 0$, and $H^T W^T G S = \mathbf{0}$. Finally,

$$
\begin{aligned}
H_{q+1,p}^T G_{p,q+1}^T E = v_p \mathbf{e}_{q+1}^T \mathbf{e}_{q+1} u_p^T E &= v_p u_p^T (WS - X)\\
&= -\sigma_p v_p v_p^T.
\end{aligned}
$$

Substituting the above results into (6) yields $D^2J[(G,H)] = -2\sigma_p$. Finally, normalizing this equation by the squared norm of (G,H) we obtain $\lambda_{\min}(\nabla^2 J(W,S)) \leq -\sigma_p$.
(iii) Assume $1 \leq q < \min\{k, r-1\}$ and (W,S) is maximal. In this situation we consider the action of $D^2J(W,S)$ on $(G_{q+1,q+1}, H_{q+1,q+1}) = (u_{q+1} e_{q+1}^T, e_{q+1} v_{q+1}^T)$. The proof proceeds exactly as in case (ii) and yields $\lambda_{\min}(\nabla^2 J(W,S)) \leq -\sigma_{q+1}$. □

Theorem 3.4 has two key contributions. First, it provides a negative upper bound on the minimum eigenvalue of the Hessian map at every strict saddle. Second, for a given X, the negative upper bound can take only one of a finite set of negative values. Hence there exists $\gamma > 0$ such that $\lambda_{\min}(\nabla^2 J(W,S)) \leq -\gamma$ for every strict saddle (W,S) in $\mathcal{M}_{\mathbf{0}}$.

The bounds discussed above take two forms. If the rank of W is $\min\{k, r\}$, the bound depends on the separation of two distinct nonzero singular values. Otherwise, it depends on the size of a single nonzero singular value. The uniform bound for all strict saddles on $\mathcal{M}_{\mathbf{0}}$, is the minimum of the smallest nonzero singular value and the smallest gap between distinct singular values. Using a more technical analysis it is possible to derive expressions for all of the Hessian eigenvalues at a canonical strict saddle. This indicates the number of negative eigenvalues and gives an expression for the minimum eigenvalue [Valavi et al., 2020].

4 Conclusion

Our contribution is to provide a more complete understanding of the landscape of matrix factorization. From prior work, all critical points are either global minima or (strict) saddles. However, the minimum eigenvalue of the Hessian map over the strict saddle points is arbitrarily close to zero.

Using our approach of restricting attention to the manifold $\mathcal{M}_{\mathbf{0}}$, the family of strict saddles on $\mathcal{M}_{\mathbf{0}}$, has a uniform negative upper bound on the minimum eigenvalue of the Hessian map. At each strict saddle on $\mathcal{M}_{\mathbf{0}}$, the bound depends on the particular structure of the saddle. Generally, the larger the nonzero singular values and the greater the separation of distinct values assumed, the more negative is the obtained upper bound. We expect this to correlate with faster escape times from a neighborhood of a strict saddle. In addition, our development has used the natural setting of the problem, avoided assuming or introducing symmetry, and avoided vectorization of the relevant differentials. We believe that this yields greater clarity and hence more insight, and will be more amenable to generalization to related problems.

References

Sanjeev Arora, Nadav Cohen, and Elad Hazan. On the optimization of deep networks: Implicit acceleration by overparameterization. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 244–253, Stockholmsmassan, Stockholm Sweden, 10–15 Jul 2018. PMLR. URL `http://proceedings.mlr.press/v80/arora18a.html`.

Pierre Baldi and Kurt Hornik. Neural networks and principal component analysis: Learning from examples without local minima. *Neural networks*, 2(1): 53–58, 1989.

Nicolas Boumal. Nonconvex phase synchronization. *SIAM Journal on Optimization*, 26(4):2355–2377, 2016.

H. Bourlard and Y. Kamp. Auto-association by multilayer perceptrons and singular value decomposition. *Biological Cybernetics*, 59(4):291–294, Sep 1988. ISSN 1432-0770. doi: 10.1007/BF00332918. URL `https://doi.org/10.1007/BF00332918`.

Simon S. Du, Wei Hu, and Jason D. Lee. Algorithmic regularization in learning deep homogeneous models: Layers are automatically balanced. In *Advances in Neural Information Processing Systems 31*, pages 384–395. 2018.

Rong Ge, Furong Huang, Chi Jin, and Yang Yuan. Escaping from saddle points? online stochastic gradient for tensor decomposition. In *Conference on Learning Theory*, pages 797–842, 2015.

Rong Ge, Chi Jin, and Yi Zheng. No spurious local minima in nonconvex low rank problems: A unified geometric analysis. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1233–1242. JMLR. org, 2017a.

Rong Ge, Jason D. Lee, and Tengyu Ma. Learning one-hidden-layer neural networks with landscape design. *arXiv preprint arXiv:1711.00501*, 2017b.

David R. Hardoon, Sandor Szedmak, and John Shawe-Taylor. Canonical correlation analysis: An overview with application to learning methods. *Neural computation*, 16(12):2639–2664, 2004.

Moritz Hardt and Tengyu Ma. Identity matters in deep learning. *arXiv preprint arXiv:1611.04231*, 2016.

Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.

Harold Hotelling. Relations between two sets of variates. *Biometrika*, 28(3/4):321–377, 1936. ISSN 00063444. URL `http://www.jstor.org/stable/2333955`.

Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430, 2000.

Ian Jolliffe. *Principal component analysis*. Springer, 2011.

Kenji Kawaguchi. Deep learning without poor local minima. In *Advances in neural information processing systems*, pages 586–594, 2016.

Xingguo Li, Junwei Lu, Raman Arora, Jarvis Haupt, Han Liu, Zhaoran Wang, and Tuo Zhao. Symmetry, saddle points, and global optimization landscape of nonconvex matrix factorization. *IEEE Transactions on Information Theory*, 2019.

Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online dictionary learning for sparse coding. In *Proceedings of the 26th annual international conference on machine learning*, pages 689–696. ACM, 2009.

Quynh Nguyen and Matthias Hein. The loss surface of deep and wide neural networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 2603–2612. JMLR. org, 2017.

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.

Ju Sun, Qing Qu, and John Wright. Complete dictionary recovery over the sphere. *arXiv preprint arXiv:1504.06785*, 2015.

Ju Sun, Qing Qu, and John Wright. A geometric analysis of phase retrieval. *Foundations of Computational Mathematics*, 18(5):1131–1198, 2018.

Hossein Valavi, Sulin Liu, and Peter Ramadge. The landscape of matrix factorization revisted. *arXiv preprint, submitted 2/26/2020 arXiv:submit/3063416*, 2020.