A Logarithmic Cost Function for Principal Singular Component Analysis

Mohammed A. Hasan
Department of Electrical & Computer Engineering
University of Minnesota Duluth
E.mail:mhasan@d.umn.edu

Abstract: *An un-unconstrained optimization problem involving logarithmic cost function that incorporates a diagonal matrix is utilized for deriving gradient dynamical systems that converge to the principal singular components of arbitrary matrix. The equilibrium points of the resulting gradient systems are determined and their stability is thoroughly analyzed. Qualitative properties of the proposed systems are analyzed in detail including the limit of solutions as time approaches infinity. The performance of this system is also examined.*

Keywords: SVD, asymptotic stability, principal singular flow, global convergence, unconstrained optimization

1 Introduction

The need for computing only a few singular vectors of a data matrix arises in many algorithmic development in scientific and engineering applications. Many methods are available for computing the singular value decomposition (SVD) of a rectangular data matrix [1]. These methods compute the whole set of singular or eigenvectors when only a few vectors are desired. Thus the main objective of this paper is to develop dynamical systems for solving the principal singular component analysis (PSCA) problems. Additionally, understanding the properties and features of such dynamical systems is helpful in determining domains of attractions and invariant sets of many principal singular subspace (PSS) and principal and minor components (PCA/MCA) dynamical systems.

There are many adaptive methods in the literature to obtain the PSA, MSA and PSS from a given data. SVD dynamical systems are developed in [2]-[8]. Algorithms for computing smallest singular triplets are proposed in [9]. Generalization of Oja's algorithm for obtaining the SVD of a rectangular matrix is considered in [10, 11]. Cross-correlation neural network for extracting the cross-correlation features between two high-dimensional data streams is developed in [12] and [8].

There are a number of methods for extracting principal or the minor subspaces of a positive definite matrix, however, there appears to be fewer algorithms for PSS in the literature. In this paper, several dynamical systems for computing PSS are derived and analyzed. Some of these algorithms may be considered as generalizations of principal components flows. The proposed dynamical systems converge to individual singular vectors by incorporating a diagonal matrix. Additionally, these systems are stable and self-normalized.

The following notation will be used throughout. The notation $\mathbb{R}$, and $\mathbb{N}$ denote the set of real numbers, and the set of positive integers. The transpose of a real matrix is denoted by x^T, and the derivative of x with respect to time is written as x'. If B is a square matrix, then $tr(B)$ denotes the trace of B. The identity matrix of appropriate dimension is expressed with the symbol I. Finally, the derivative of $V(x,y)$ with respect to time is denoted by $\dot{V}$.

2 Preliminary Results

For completeness, basic concepts from dynamical system theory are summarized in this section. These include Lyaunov and Lagrange stability.

2.1 Stability of Dynamical Systems

The Lyapunov direct method provides a convenient way of proving stability of equilibria, as Lyapunov's theorem can be used without solving the associated differential equations. However, it is not always easy to construct Lyapunov functions or test their time derivatives for non-negative definiteness.

Let $g(x) : \mathbb{R}^{n\times p} \to \mathbb{R}^{n\times p}$, $p \leq n$, be continuously differentiable function and consider the dynamical system

$$x' = g(x). \tag{1}$$

The point $\bar{x}$ is an equilibrium point for the system (1) if $g(\bar{x}) = 0$. Let $\Omega \subset \mathbb{R}^{n\times p}$ be a region containing $\bar{x}$ and $V : \Omega \to \mathbb{R}$ be continuously differentiable function such that $V(\bar{x}) = 0$ and $V(x) > 0$ for each $\bar{x} \neq x \in \Omega$, i.e., V is positive definite. Assume also that $\dot{V}(x) \leq 0$ for each $x \in \Omega$, i.e., $\dot{V}$ is negative semi-definite. Then $\bar{x}$ is stable and V is called a Lyapunov function for the system (1) at $\bar{x} \in \Omega$. If $\dot{V}(x) < 0$ for each $\bar{x} \neq x \in \Omega$, then $\bar{x}$ is asymptotically stable. If in addition to these conditions, we have the function V is radially unbounded, i.e., $V(x) \to \infty$ as $||x|| \to \infty$, then the system is globally stable. The main advantage of using Lyapunov direct method is that Lyapunov theorem can be used to prove stability of equilibria without solving the differential equations. However, constructing Lyapunov functions is not always an easy task. It should be noted that many Lyapunov functions may exist for the same problem. However, a specific choice of Lypunov functions may provide more useful results about the system than others.

Geometrically, the condition $\dot{V} \leq 0$ implies that when a trajectory crosses the level surface $V(x) = c$, it moves inside the set $\Omega_2 = \{x \in \mathbb{R}^{n\times p} : V(x) \leq c\}$ and remains there. Since V is positive definite, then Ω_2 is bounded and closed, thus the system must converge to some limiting value.

A set S is an *invariant set* for the system (1) if every trajectory $x(t)$ which starts from a point in S remains in S for all time. For example, any equilibrium point is an invariant set. The domain of attraction of an equilibrium point is also an invariant set.

Next, we state a few stability results for nonlinear autonomous systems. The invariant set theorems reflect the intuition that the decrease of a Lyaunov function V has to gradually vanish. In other words $\dot{V}$ has to converge to zero because V is lower bounded. Proofs of the results below can be found in [13]-[15].

Theorem 1 (Local Invariant Set Theorem). *Consider an autonomous system of the form $x' = g(x)$, with g continuous and let $V(x) : \mathbb{R}^n \to \mathbb{R}$ be a scalar function with continuous first partial derivatives. Assume that*

1. *for some $l > 0$, the set Ω_l defined by $V(x) \leq l$ is bounded.*
2. *$V'(x) \leq 0$ for all x in Ω_l.*

Let R be the set of all points within Ω_l where $V'(x) = 0$ and M be the largest invariant set in R. Then, every solution $x(t)$ originating in Ω_l tends to M as $t \to \infty$.

Proof. See Slotine and Li (1991) [13] and [14].

In Theorem 1, the word largest means that M is the union of all invariant sets within R. Notice that R is not necessarily connected, nor is the set M.

To analyze systems involving a matrix $A \in \mathbb{R}^{n\times m}$, $m \leq n$, it will be assumed that the singular value decomposition of A is

$$A = u\Sigma v^T + u_2\Sigma_2 v_2^T, \tag{2}$$

where $\Sigma = \text{diag}(\sigma_1, \cdots, \sigma_p)$ and $\Sigma_2 = \text{diag}(\sigma_{p+1}, \cdots, \sigma_m)$ are diagonal matrices so that $\sigma_i > \sigma_j$ for $i = 1, \cdots, p$ and $j = p+1, \cdots, m$. The matrices $u \in \mathbb{R}^{n\times p}$, $v \in \mathbb{R}^{m\times p}$, $u_2 \in \mathbb{R}^{n\times n-p}$, and $v_2 \in \mathbb{R}^{m\times n-p}$ are orthogonal, i.e., $u^T u = I, v^T v = I$ and $u_2^T u_2 = I, v_2^T v_2 = I$, $u^T u_2 = 0$, $v^T v_2 = 0$. It can be easily verified that the matrix

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} u & -u \\ v & v \end{bmatrix}, \tag{3a}$$

is orthogonal, i.e., $U^T U = I$, and that

$$U^T \bar{A} U = \begin{bmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{bmatrix}, \tag{3b}$$

where

$$\bar{A} = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}. \tag{3c}$$

Thus $\bar{A}$ can be expressed as

$$\bar{A} = U\bar{\Sigma}U^T + U_2\bar{\Sigma}_2 U_2^T, \tag{3d}$$

where

$$\begin{aligned} \bar{\Sigma} &= \begin{bmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{bmatrix}, \bar{\Sigma}_2 = \begin{bmatrix} \Sigma_2 & 0 \\ 0 & -\Sigma_2 \end{bmatrix}, \\ U_2 &= \frac{1}{\sqrt{2}} \begin{bmatrix} u_2 & -u_2 \\ v_2 & v_2 \end{bmatrix}. \end{aligned} \tag{3e}$$

Note that U_2 is orthogonal, i.e., $U_2^T U_2 = I$.

In the next section, the gradient and the Hessian matrices for some matrix functions are given using the first and second order differentials.

2.2 First and Second Order Differentials

Let $F : U \to \mathbb{R}$, where $U \in \mathbb{R}^{n\times p}$, be twice continously differentiable function, and let $x \in U$ and $dx \in \mathbb{R}^{n\times p}$. Then the first and second order differentials of F at x in the direction of dx are defined by

$$dF(x, dx) = \frac{dF(x + \epsilon dx)}{d\epsilon}|_{\epsilon=0}, \tag{4a}$$

and

$$d^2F = \frac{d^2F(x + \epsilon dx)}{d\epsilon^2}|_{\epsilon=0}. \tag{4b}$$

The quantity $dF(x, dx)$ is sometimes called the Gateaux derivative of F at x in the direction of dx.

To compute the gradient and the Hessian matrix for a cost function F, the first and second order differentials need to be derived first. In the next result, the first and second order differentials for linear, quadratic, and quartic functions are computed.

Theorem 2. *Let $E \in \mathbb{R}^{n\times n}$, and let $F : \mathbb{R}^{n\times p} \to \mathbb{R}$ be twice continuously differentiable function such that $F(z) = tr\{\log(z^T Ez)\}$, where $z \in \mathbb{R}^{n\times p}$. Then the first order differential of F is*

$$dF = tr\{(z^T Ez)^{-1}(dz^T Ez + z^T Edz)\}, \tag{5a}$$

and the second order differential of F is

$$\begin{aligned} d^2F &= 2tr\{(z^T Ez)^{-1} dz^T Edz\} \\ &- tr\{(z^T E^T z)^{-1} z^T E^T dz(z^T E^T z)^{-1} z^T E^T dz \\ &- 2tr\{(z^T Ez)^{-1} dz^T Ez(z^T Ez)^{-1} z^T Edz \\ &- tr\{(z^T Ez)^{-1} z^T Edz(z^T Ez)^{-1} z^T Edz. \end{aligned} \tag{5b}$$

Proof: The proof is a direct application of the definitions (5a), (5b) and properties of the trace operator.

2.3 Gradient and Hessian Matrices

The gradient and Hessian matrices can be obtained from first and second order differentials as the following lemma [16].

Lemma 3. *Let ϕ be a twice differentiable real-valued function of an $n \times p$ matrix. Then, the following relationships hold:*

$$d\phi(X) = tr(A^T dX) \Leftrightarrow \nabla\phi(X) = A \tag{6a}$$

$$d^2\phi(X) = tr(B(dX)^T CdX) \Leftrightarrow H\phi(X) = \frac{1}{2}(B^T \otimes C + B \otimes C^T) \tag{6b}$$

$$d^2\phi(X) = tr(B(dX)CdX) \Leftrightarrow H\phi(X) = \frac{1}{2}K_{rn}(B^T \otimes C + C^T \otimes B) \tag{6c}$$

where d denotes the differential, and A, B, and C are matrices, each of which may be a function of X. The gradient of ϕ with respect to X and the Hessian matrix of ϕ at X are defined as

$$\begin{aligned} \nabla\phi(X) &= \frac{\partial\phi(X)}{\partial X} \\ H\phi(X) &= \frac{\partial}{(vecX)^T}\left(\frac{\partial\phi(X)}{\partial(vecX)^T}\right)^T \end{aligned} \tag{6d}$$

where vec is the vector operator and stands for the operation of stacking the columns of a matrix into one column, and $\otimes$ denotes the Kronecker product. The matrix K_{pn} denotes the $pn \times pn$ commutation matrix; $K_{pn}^T = K_{pn}^{-1} = K_{np}$ and $K_{rm}(A \otimes C) = (C \otimes A)K_{qn}$, where $A \in \mathbb{R}^{m\times n}$ and $C \in \mathbb{R}^{r\times q}$.

Corollary 4. *Let let $F : \mathbb{R}^{n\times p} \to \mathbb{R}$ be as in Theorem 4. Then the gradient and the Hessian matrix of F are given by*

$$\nabla F = Ez(z^T Ez)^{-1} + E^T z(z^T E^T z)^{-1}, \tag{7a}$$

$$\begin{aligned} HF &= (z^T E^T z)^{-1} \otimes E + (z^T Ez)^{-1} \otimes E^T \\ &- (z^T E^T z)^{-1} \otimes Ez(z^T Ez)^{-1} z^T E \\ &- (z^T Ez)^{-1} \otimes E^T z(z^T E^T z)^{-1} z^T E^T \\ &- KEz(z^T Ez)^{-1} \otimes (z^T E^T z)^{-1} z^T E^T \\ &- KE^T z(z^T E^T z)^{-1} \otimes (z^T Ez)^{-1} z^T E. \end{aligned} \tag{7b}$$

for some permutation matrix K.

This result can be generalized for computing the first and second order differentials of $F(z) = tr\{\log(z^T Ez + D)\}$ as shown in the following result.

Corollary 5. *Let $E \in \mathbb{R}^{n\times n}$, and $D \in \mathbb{R}^{p\times p}$ such that $D^T = D$. Let $F : \mathbb{R}^{n\times p} \to \mathbb{R}$ be defined by $F(z) = tr\{\log(z^T Ez + D)\}$, then the first and second order differentials of F are*

$$dF = tr\{(z^T Ez + D)^{-1}(dz^T Ez + z^T Edz)\}, \tag{8a}$$

and

$$\begin{aligned} d^2F &= 2tr\{(z^T Ez + D)^{-1} dz^T Edz\} \\ &- tr\{(z^T E^T z + D)^{-1} z^T E^T dz(z^T E^T z + D)^{-1} z^T E^T \\ &- 2tr\{(z^T Ez + D)^{-1} dz^T Ez(z^T Ez + D)^{-1} z^T Edz \\ &- tr\{(z^T Ez + D)^{-1} z^T Edz(z^T Ez + D)^{-1} z^T E. \end{aligned} \tag{8b}$$

Therefore, the gradient and the Hessian matrix of F are

$$\nabla F = Ez(z^T Ez + D)^{-1} + E^T z(z^T E^T z + D)^{-1}, \tag{9a}$$

$$\begin{aligned} HF &= (z^T E^T z + D)^{-1} \otimes E + (z^T Ez + D)^{-1} \otimes E^T \\ &- (z^T E^T z + D)^{-1} \otimes Ez(z^T Ez + D)^{-1} z^T E \\ &- (z^T Ez + D)^{-1} \otimes E^T z(z^T E^T z + D)^{-1} z^T E^T \\ &- KEz(z^T Ez + D)^{-1} \otimes (z^T E^T z + D)^{-1} z^T E^T \\ &- KE^T z(z^T E^T z + D)^{-1} \otimes (z^T Ez + D)^{-1} z^T E, \end{aligned} \tag{9b}$$

for some permutation matrix K.

Proof: The proof is a direct application of the definitions (5a), (5b), and Lemma 5.

3 A Logarithmic Cost Function

In this section, a logarithmic cost function is introduced. Based on the gradient of this function, dynamical systems for principal singular component analysis for a general rectangular matrix are derived. The cost function that will be considered is defined as

$$G(x, y) = tr\{\log(x^T Ay + D)\} - \frac{\alpha}{2} tr\{(x^T x + y^T y)\}, \tag{10}$$

where $A \in \mathbb{R}^{n\times m}$, $D \in \mathbb{R}^{p\times p}$, $x \in \mathbb{R}^{n\times p}$, $y \in \mathbb{R}^{m\times p}$, and $\alpha > 0$ is sufficiently large number.

Remark: The natural logarithm of a square matrix C, denoted by $\log(C)$, is defined if and only if C is invertible. This means that $\log(C)$ is defined as long as the spectrum of C does not contain the origin.

The cost function G can be shown to be upper bounded and $-G$ is radially unbounded. Thus gradient systems that converge to the principal singular components (PSC) of the given matrix A can be derived. To gain some insight of the above cost function defined in (10), we consider the scalar case as in the following example.

Example: Let $F(x, y) = \log(axy+d) - \frac{1}{2}x^2 - \frac{1}{2}y^2$, where $a > 0$ and $d > 0$. The objective is to find the minima and maxima of F over $\mathbb{R}^2$. The gradient and the Hessian matrix of F can be verified to be

$$\nabla F = \begin{bmatrix} \frac{ay}{axy+d} - x \\ \frac{ax}{axy+d} - y \end{bmatrix}, \tag{11a}$$

and

$$\nabla^2 F = \begin{bmatrix} \frac{-(ay)^2}{(axy+d)^2} - 1 & \frac{ad}{(axy+d)^2} \\ \frac{ad}{(axy+d)^2} & \frac{-(ax)^2}{(axy+d)^2} - 1 \end{bmatrix}. \tag{11b}$$

The equilibrium points of F are solutions of the equations

$$\begin{aligned} ay &= x(axy+d), \\ ax &= y(axy+d). \end{aligned} \tag{12a}$$

Clearly, $(x, y) = (0, 0)$ is one of the solutions for $\nabla F = 0$. If $x \neq 0$, then $y \neq 0$ and $\frac{y}{x} = \frac{x}{y}$. Thus $x^2 = y^2$, or equivalently $y = \pm x$. Now if $y = x$, then $a = ax^2 + d$ or $x^2 = 1 - \frac{d}{a}$. The last equation has real solutions only if $\frac{d}{a} \leq 1$, in which case $x = \pm\sqrt{1 - \frac{d}{a}}$. This yields the following solutions: $\{(x, y) = \pm(\sqrt{1-\frac{d}{a}}, \sqrt{1-\frac{d}{a}})\}$. Similarly if $y = -x$, we obtain the following solutions: $x = \pm\sqrt{1+\frac{d}{a}}$. This yields the following solutions: $\{(x, y) = (\pm\sqrt{1+\frac{d}{a}}, \mp\sqrt{1+\frac{d}{a}})\}$. Hence there are at least five solutions.

At non-zero equilibrium points we have a) $y^2 = x^2 = 1 - \frac{d}{a}$, or b) $y^2 = x^2 = 1 + \frac{d}{a}$. For the first case (a), $\nabla^2 F$ simplifies to

$$\begin{aligned} \nabla^2 F &= \begin{bmatrix} -x^2 - 1 & \frac{d^2}{a^2} \\ \frac{d^2}{a^2} & -y^2 - 1 \end{bmatrix} \\ &= \begin{bmatrix} -2 + \frac{d}{a} & \frac{d}{a} \\ \frac{d}{a} & -2 + \frac{d}{a} \end{bmatrix}. \end{aligned} \tag{12b}$$

Thus $\nabla^2 F$ is negative definite provided that $\frac{d}{a} \leq 2$ and $det(\nabla^2 F) = 4(1 - \frac{d}{a}) > 0$. Consequently, $\nabla^2 F$ is negative definite if and only if $\frac{d}{a} < 1$.

For Case (b), similar analysis shows that $\nabla^2 F = \begin{bmatrix} -2 - \frac{d}{a} & \frac{d}{a} \\ \frac{d}{a} & -2 - \frac{d}{a} \end{bmatrix}$ which is negative definite if and only if $\frac{d}{a} > -1$. This condition is always satisfied if a is positive and d is non-negative.

3.1 A Gradient Dynamical System

Let $z = \begin{bmatrix} x \\ y \end{bmatrix}$, and $E = \begin{bmatrix} 0 & A \\ 0 & 0 \end{bmatrix}$, then $z^T E z = x^T A y$, $z^T E^T z = y^T A^T x$, and therefore, the cost function $G(x, y)$ defined in (11) may be expressed as $G(x, y) = G(z) = tr\{\log(z^T E z + D) - \frac{\alpha}{2} z^T z\}$. From Corollary 7, it follows that the gradient and the Hessian matrix HG can be expressed in terms of the matrices A, D, x, y as follows:

$$\nabla G = \begin{bmatrix} Ay(x^T A y + D)^{-1} - \alpha x \\ A^T x(y^T A^T x + D)^{-1} - \alpha y \end{bmatrix}. \tag{13a}$$

Thus gradient dynamical systems for maximizing G may be expressed by

$$\begin{aligned} x' &= Ay(x^T A y + D)^{-1} - \alpha x, \\ y' &= A^T x(y^T A^T x + D)^{-1} - \alpha y. \end{aligned} \tag{13b}$$

To analyze stability of equilibrium points of this system, an expression for the Hessian matrix HG can be verified to be:

$$\begin{aligned} HG = {} & (y^T A^T x + D)^{-1} \otimes \begin{bmatrix} 0 & A \\ 0 & 0 \end{bmatrix} + (x^T A y + D)^{-1} \\ & \otimes \begin{bmatrix} 0 & 0 \\ A^T & 0 \end{bmatrix} - (x^T A y + D)^{-1} \otimes \begin{bmatrix} 0 & 0 \\ A^T & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ & \times (y^T A^T x + D)^{-1} \begin{bmatrix} x^T & y^T \end{bmatrix} \begin{bmatrix} 0 & 0 \\ A^T & 0 \end{bmatrix} \\ & - (y^T A^T x + D)^{-1} \otimes \begin{bmatrix} 0 & A \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} (x^T A y + D)^{-1} \\ & \times \begin{bmatrix} x^T & y^T \end{bmatrix} \begin{bmatrix} 0 & A \\ 0 & 0 \end{bmatrix} \\ & - K \begin{bmatrix} 0 & A \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} (x^T A y + D)^{-1} \otimes (y^T A^T x + D)^{-1} \\ & \times \begin{bmatrix} x^T & y^T \end{bmatrix} \begin{bmatrix} 0 & 0 \\ A^T & 0 \end{bmatrix} \\ & - K \begin{bmatrix} 0 & 0 \\ A^T & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} (y^T A^T x + D)^{-1} \otimes (x^T A y + D)^{-1} \\ & \times \begin{bmatrix} x^T & y^T \end{bmatrix} \begin{bmatrix} 0 & A \\ 0 & 0 \end{bmatrix}, \end{aligned} \tag{14}$$

$$\begin{aligned} = {} & (y^T A^T x + D)^{-1} \otimes \begin{bmatrix} 0 & A \\ 0 & 0 \end{bmatrix} \\ & + (x^T A y + D)^{-1} \otimes \begin{bmatrix} 0 & 0 \\ A^T & 0 \end{bmatrix} \\ & - (x^T A y + D)^{-1} \otimes \begin{bmatrix} 0 & 0 \\ A^T x(y^T A^T x + D)^{-1} y^T A^T & 0 \end{bmatrix} \\ & - (y^T A^T x + D)^{-1} \otimes \begin{bmatrix} 0 & Ay(x^T A y + D)^{-1} x^T A \\ 0 & 0 \end{bmatrix} \\ & - K \begin{bmatrix} Ay \\ 0 \end{bmatrix} (x^T A y + D)^{-1} \otimes (y^T A^T x + D)^{-1} \begin{bmatrix} y^T A^T & 0 \end{bmatrix} \\ & - K \begin{bmatrix} 0 \\ A^T x \end{bmatrix} (y^T A^T x + D)^{-1} \otimes (x^T A y + D)^{-1} \begin{bmatrix} 0 & x^T A \end{bmatrix}. \end{aligned} \tag{15}$$

for some permutation matrix K.

In the next section, the dynamical system (13b) will be analyzed in terms of stability, convergence, and the limiting behavior as $t \to \infty$.

4 Convergence Analysis

The stability of the system (13b) can be established as in the following theorem:

Theorem 6. *Consider the dynamical system (14b) with $\alpha = 1$, and assume that $x(t)$ and $y(t)$ are solutions for the system for $t \geq 0$. Assume also that $x(0) = x_0 \in \mathbb{R}^{n\times p}$, $y(0) = y_0 \in \mathbb{R}^{m\times p}$ are full rank. Let $P = \lim_{t\to\infty} x(t)^T x(t)$, $Q = \lim_{t\to\infty} y(t)^T y(t)$, and $B = \lim_{t\to\infty} x(t)^T A x(t)$. Then each equilibrium point of (13b) is of the form*

$$\begin{aligned} x &= u\sqrt{I - D\Sigma^{-1}}, \\ y &= v\sqrt{I - D\Sigma^{-1}}, \end{aligned} \tag{16}$$

where u and v are as defined in (2). Moreover, P, Q, and B are diagonal and $P = Q$. The principal singular values of A are the diagonal elements of the matrix $P^{-1}B$.

Proof: The proof of this result follows from solving $\nabla G = 0$ and applications of Propositions 7, 8, and 9 see Appendix). The details are omitted due to space limitations.

5 Numerical Example

In this example, we examine the convergence of the dynamical system (13b). A matrix A is chosen randomly using the matlab function *rand*. Euler method is used to approximate the solution with a learning parameter $\alpha = 0.62$. This number is chosen randomly. The matrices A, D are as given below. The algorithm is iterated N=7700 times and convergence is measured by the off-diagonal elements of $x^T Ay$, $x^T x$, and $y^T y$. The initial matrices x_0, y_0 are chosen to be full rank using the Matlab function *rand*. As can be seen from the matrices below, in the limit, $x^T x$, $y^T y$, and $x^T Ay$ are all diagonal (as in Matlab syntax x' denotes x^T). It should be noted that the diagonal elements of D and $\hat{x}^T A\hat{y}$ have the same ordering.

```
A=
40.4305  12.0298  13.9160  12.9982   3.8420  13.9184
12.1381  39.3421  14.5428   6.2107  18.0638  11.1726
12.3564  14.5607  24.1947  15.2457  10.5267   5.8570
14.9267   5.6902  12.6775  35.2897   8.8043  10.4545
 6.9882  19.9212  14.1417  10.8331  27.8930   8.9412
16.1348  13.0127   6.4962  14.3504  11.5070  47.2777

x = 0.4297   -0.6360   -0.3960   -0.3002
    0.4375   -0.2463    0.2243    0.5647
    0.3428    0.1050   -0.2929    0.1940
    0.3683    0.5533   -0.4639   -0.2240
    0.3634    0.2527    0.1013    0.4187
    0.4822    0.1026    0.6357   -0.5272

y = 0.4476   -0.6006   -0.3892   -0.3104
    0.4502   -0.2404    0.2251    0.5687
    0.3517    0.0520   -0.2920    0.2204
    0.3916    0.5934   -0.4346   -0.2200
    0.3361    0.2793    0.1664    0.3871
    0.4480    0.0727    0.6469   -0.5325

x'Ay=94.8730    0.0000    0.0000    0.0000
     -0.0000   22.4310   -0.0000    0.0000
     -0.0000   -0.0000   30.5027    0.0050
      0.0000   -0.0000    0.0050   36.1468

x'x=0.9937    0.0000         0   -0.0000
    0.0000    0.8567   -0.0000    0.0000
         0   -0.0000    0.9225    0.0000
   -0.0000    0.0000    0.0000    0.9501

y'y=0.9937    0.0000         0   -0.0000
    0.0000    0.8567   -0.0000   -0.0000
         0   -0.0000    0.9225    0.0000
   -0.0000   -0.0000    0.0000    0.9501

D= 0.5983  0        0         0
   0       3.7533   0         0
   0       0        2.5613    0
   0       0        0         1.8994
```

6 Conclusion

In this paper, unconstrained optimization methods are utilized to derive a dynamical system that converges to the principal singular components of a given matrix. Numerical experiments have shown that the proposed system is fast and the learning parameter is nearly independent of the matrix used. It is also noticed that the system converges to the actual singular triplets starting from any full rank initial conditions. The case where the initial conditions are not full rank remains to be explored. The work presented here requires additional analysis and generalization to complex data and matrices.

7 Appendix

In this appendix, we list a number of results that are used in proving some of the propositions of this work.

Proposition 7. *Let $D, A \in \mathbb{R}^{n\times n}$ be positive definite matrices and assume that D is diagonal having distinct eigenvalues. If $AD = DA$, then A is diagonal.*

Proof: Assume that $A = [a_{ij}]$ and $D = \text{diag}(\mu_1, \cdots, \mu_n)$, then for each i, j we have $a_{ij}\mu_j = \mu_i a_{ij}$ or $(\mu_j - \mu_i)a_{ij} = 0$. Thus $a_{ij} = 0$ for $i \neq j$, i.e., A is diagonal.

Proposition 8 [17]. *Let $B, D \in \mathbb{R}^{p\times p}$ and assume that D is diagonal and all eigenvalues of D are distinct. If $BD + DB$ is diagonal, then B is diagonal.*

Proposition 9. *Let $A, B, C \in \mathbb{R}^{n\times n}$, then the matrices ABC, BCA, CAB are similar and thus have the same set of eigenvalues.*

References

[1] G. H. Golub and C. F. van Loan, Matrix Computations. Baltimore, MD: The Johns Hopkins Univ. Press, 1989.

[2] P. Comon and G. H. Golub, "Tracking a few extreme singular values and vectors in signal processing," Proceedings of the IEEE, 78(8):1327-1343, August 1990.

[3] G. Hori, "A general framework for SVD flows and joint SVD flows," Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003 (ICASSP '03), Volume: 2, 6-10, April 2003, pp:II-693-696.

[4] A. Cichocki, "Neural network for singular value decomposition," Electron. Lett., vol. 28, pp. 784-786, 1992.

[5] K. I. Diamantaras and S. Y. Kung, "Cross-correlation neural network models," IEEE Trans. Signal Processing, vol. 42, pp. 32183223, Nov. 1994.

[6] U. Helmke and J. B. Moore, "Singular value decomposition via gradient and self-equivalent flows," Linear Algebra Appl., vol. 169, pp. 223-248, 1992.

[7] J. B. Moore, R. E. Mahony, and U. Helmke, "Numerical gradient algorithms for eigenvalues and singular value calculations," SIAM J. Matrix Anal. Appl., vol. 15, pp. 881902, 1994.

[8] M.A. Hasan,"Dynamical Systems for Principal Singular Subspace Analysis," Fourth IEEE workshop on Sensor Array and Multichannel Processing, July 12-14, 2006, Page(s):142-146.

[9] E. Kokiopoulou, C. Bekas, E. Gallopoulos, "Computing smallest singular triplets with implicitly restarted Lanczos bidiagonalization," Applied Numerical Mathematics archive, Volume 49 , Issue 1, April 2004, Pages: 39-61

[10] A. L. Yuile, D. M. Kammen, and D. S. Cohen, Quadrature and develop-ment of orientation selective cortical cells by Hebb rules, Biol. Cybern., vol. 61, pp. 183194, 1989.

[11] N. Samardzija and R. L. Waterland, "A neural network for computing eigenvectors and eigenvalues," Biol. Cybern., vol. 65, no. 4, pp. 211214.

[12] Da-Zheng Feng; Xian-Da Zhang; Zheng Bao, "A neural network learning for adaptively extracting cross-correlation features between two high-dimensional data streams," IEEE Transactions on Neural Networks, Volume: 15, Issue: 6, pp. 1541-1554, Nov. 2004.

[13] J. J. E. Slotine and W. Li. Applied nonlinear Comtrol. Prentice Hall, 1991.

[14] J. LaSalle, "Some Extensions of Lyaunov's Second Method," IRE Transactions on Circuit Theory, Volume 7, Issue 4, Dec 1960 Page(s):520-527.

[15] Hassan K. Khalil, Nonlinear Systems, 3rd edition. Prentice Hall, 2002.

[16] J. R. Magnus and H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, 2nd ed. New York: Wiley, 1991.

[17] M.A Hasan,"Diagonally weighted and shifted criteria for minor and principal component extraction," IEEE International Joint Conference on Neural Networks, 2005. IJCNN '05. Proceedings. Volume 2, 31 July-4 Aug. 2005 Page(s):1251-1256.