2008 American Control Conference
Westin Seattle Hotel, Seattle, Washington, USA
June 11-13, 2008

Gradient Dynamical Systems for Principal Singular Subspace Analysis

Mohammed A. Hasan† and Jawad A. K. Hasan‡

†Department of Electrical & Computer Engineering, University of Minnesota Duluth

‡Department of Electrical Engineering, University of Baghdad, Baghdad, Iraq

E.mail:mhasan@d.umn.edu

Abstract

Principal singular component analysis has recently been proposed and analyzed by the author. It is a generalization of the principal singular subspace analysis which has been investigated in the literature. In this paper an unconstrained weighted cost function is utilized to develop dynamical systems that converge to the actual principal singular vectors of a given matrix. Stability analysis that reveals the domains of attraction of these systems is also given.

Keywords: polynomial dynamical systems, asymptotic stability, global stability, global convergence, invariant set, Lyapunov stability, Newton method, Halley method, Lasalle invariance principle,

1 Introduction

Computing a few number of singular triplets of a matrix is an essential task in many algorithmic development for various fields in signal processing, control theory, and applied mathematics. By a singular triplet for a matrix $A \in \mathbb{R}^{n\times m}$, we mean (u, σ, v) , where u, v are unit vectors such that $u^T Av = \sigma$, where $u^T u = 1$, $v^T v = 1$, and σ is a positive number. These vectors are equilibrium points for a gradient system obtained from optimizing a quadratic cost function. These equilibrium points are shown to be stable with a domain of attraction that can be enlarged by incorporating an additional parameter.

Conventional methods for computing the singular value decomposition are given in [1]-[2]. Generalization of Oja's algorithm for obtaining the SVD of a rectangular matrix is considered in [3, 4]. Cross-correlation neural network for extracting the cross-correlation features between two high-dimensional data streams is developed in [5]-[6]. Algorithms that are based on optimization over unitary constraints are developed in [7]-[8]. In the aforementioned methods, either the whole set of singular or eigenvectors are computed or a basis for subspace spanned by the principal singular vectors rather than the actual singular vectors.

Thus the main objective of this paper is to develop and study dynamical systems for solving the principal singular component analysis (PSCA) problems, and to provide both discrete and analog neural systems for solving computational problems in real-time. Additionally, understanding the properties and features of such dynamical systems is helpful in determining domains of attractions and invariant sets of dynamical systems of many principal singular subspace (PSS) and principal/minor component analyzers (PCA/MCA).

The following notation will be used throughout. The symbols $\mathbb{R}$, and $\mathbb{N}$ denote the set of real numbers, and the set of positive integers, respectively. The derivative of x with respect to time is written as x'. The identity matrix of appropriate dimension is expressed with the symbol I. Finally, the derivative of a Lyapunov function $V(x)$ with respect to time is denoted by $\dot{V}$. The notation $||x||$ denotes the Euclidean norm of x. It will be assumed in this work that all matrices are real.

2 Preliminary Results

For completeness, basic concepts from dynamical system theory are summarized in this section. These include Liapunov 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 $g(x) : \mathbb{R}^{n\times p} \to \mathbb{R}^{n\times p}$, $p \le n$, 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 $\bar{x} \ne x \in \Omega$, i.e., V is positive definite. Assume also that $\dot{V}(x) \le 0$ 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 $V(x) < 0$ for each $\bar{x} \ne x \in \Omega$, 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 $\dot{V} \le 0$ implies that when a trajectory crosses the level surface $V(x) = c$, it moves inside the set $\Omega_2 = \{x \in \mathbb{R}^{n\times p} : V(x) \le c\}$ and remains there. Since V is positive definite, then Ω_2 is bounded and closed, thus the system must converge to some limiting value.

The *domain of attraction* of an equilibrium point $\hat{x}$ of the system (1) is defined as an open set D containing $\hat{x}$ such that for any initial point $x_0 \in D$, the sequence generated by the dynamical system according to (1) with an arbitrarily small step-size $\alpha > 0$ and satisfying $x_k \in D$, for all i) remains in D and ii) x_k converges to $\hat{x}$.

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.

We state next a few stability results for nonlinear autonomous systems. The invariant set theorems reflect the intuition that the decrease of a Liapunov 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 [9]-[11].

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. *$\dot{V}(x) \leq 0$ for all x in Ω_l.*

Let R be the set of all points within Ω_l where $\dot{V}(x) = 0$ 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) [9] and [10].

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.

We state next a well known result about Lagrange stability. A dynamical system is Lagrange stable if the continuous state remains bounded from any initial condition. For example, if the continuous state converges to a stationary set, the dynamical system is Lagrange stable.

Theorem 2 (Lagrange Stability Theorem)[10]. *Let W be a bounded neighborhood of the origin and let W^c be its complement (W^c is the set of all points outside W). Assume that $V(x)$ is a scalar function with continuous first partial derivatives in W^c and satisfying:*

1. *$V(x) > 0$ for all $x \in W^c$,*
2. *$\dot{V}(x) \leq 0$ for all $x \in W^c$,*
3. *$V(x) \to \infty$ as $||x|| \to \infty$.*

Then each solution of $x' = g(x)$, is bounded for all $t > 0$.

The Lyapunov linearization method explores the relation between the stability of the linearized system with that of the original nonlinear system.

Theorem 3 (Liapunov's Linearization Method). *Let $x = \hat{x}$ be an equilibrium point for the nonlinear system $\dot{x} = g(x)$, where $g : D \to \mathbb{R}^n$ is continuously differentiable and D is a neighborhood of $\hat{x}$. Let the Jabobian matrix A at $x = \hat{x}$ be:*

$$A = \frac{\partial g}{\partial x}|_{x=\hat{x}}. \tag{2}$$

Let λ_i, $i = 1, \cdots, n$ be the eigenvalues of A. Then,

1. *The point $\hat{x}$ is asymptotically stable if $Re(\lambda_i) < 0$ for all eigenvalues of A.*
2. *The point $\hat{x}$ is unstable if $Re(\lambda_i) > 0$ for any of the eigenvalues of A.*

Here $Re(\lambda)$ denotes the real part of λ.

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

$$A = u\Sigma v^T + u_2\Sigma_2 v_2^T, \tag{3}$$

where $\Sigma = \text{diag}(\sigma_1, \cdots, \sigma_p)$ and $\Sigma_2 = \text{diag}(\sigma_{p+1}, \cdots, \sigma_n)$ are diagonal matrices so that $\sigma_i > \sigma_j$ for $i = 1, \cdots, p$ and $j = p+1, \cdots, n$. The matrices $u, v \in \mathbb{R}^{n\times p}$ and $u_2, v_2 \in \mathbb{R}^{n\times n-p}$ 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

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} u & -u \\ v & v \end{bmatrix}, \tag{4a}$$

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

$$U^T \bar{A} U = \begin{bmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{bmatrix}, \tag{4b}$$

where

$$\bar{A} = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}. \tag{4c}$$

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

$$\bar{A} = U\bar{\Sigma}U^T + U_2\bar{\Sigma}_2 U_2^T, \tag{4d}$$

where

$$\bar{\Sigma} = \begin{bmatrix} 0 & \Sigma \\ \Sigma & 0 \end{bmatrix}, \bar{\Sigma}_2 = \begin{bmatrix} 0 & \Sigma_2 \\ \Sigma_2 & 0 \end{bmatrix}, \quad U_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} u_2 & -u_2 \\ v_2 & v_2 \end{bmatrix}. \tag{4e}$$

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

$$U_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} u_2 & -u_2 \\ v_2 & v_2 \end{bmatrix}, \tag{4d}$$

which 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 be twice continously differentiable function, the first and second order differentials of F are defined by

$$dF = \frac{\partial F(x + \epsilon dx)}{\partial \epsilon}|_{\epsilon=0}, \tag{5a}$$

and

$$d^2F = \frac{\partial^2 F(x + \epsilon dx)}{\partial \epsilon^2}|_{\epsilon=0}. \tag{5b}$$

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.

Lemma 4. *Let $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^{n\times p}$, $D \in \mathbb{R}^{p\times p}$ and consider the functions defined over $\mathbb{R}^{n\times p}$ by*

$$\begin{aligned} F_1(z) &= tr(b^T z), \\ F_2(z) &= tr(z^T \bar{A} z D), \\ F_3(z) &= tr((z^T z)^2). \end{aligned} \tag{6}$$

Then the first and second order differentials of F_1, F_2 and F_3 are given by:

$$dF_1(z) = tr(b^T dz), \; d^2F_1 = 0, \tag{7a}$$

$$\begin{aligned} dF_2 &= tr\{dz^T \bar{A} z D + z^T \bar{A} dz D)\}, \\ d^2F_2(z) &= tr\{2dz^T \bar{A} dz D\}, \end{aligned} \tag{7b}$$

$$\begin{aligned} dF_3 &= 4tr\{dz^T z z^T z\}, \\ d^2F_3(z) &= 4tr\{dz^T dz z^T z + dz^T z dz^T z + dz^T z z^T dz)\}. \end{aligned} \tag{7c}$$

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

2.3 Gradient and Hessian Matrices

The computation of derivatives can be performed simply based on the following lemma [12].

Lemma 5. *Let ϕ be a twice differentiable real-valued function of an $n \times p$ matrix. Then, the following relationships hold:*

$$d\phi(X) = tr(A^T dX) \Leftrightarrow \nabla\phi(X) = A \tag{8a}$$

$$d^2\phi(X) = tr(B(dX)^T C dX) \Leftrightarrow H\phi(X) = \frac{1}{2}(B^T \otimes C + B \otimes C^T) \tag{8b}$$

$$d^2\phi(X) = tr(B(dX)CdX) \Leftrightarrow H\phi(X) = \frac{1}{2}K_{rn}(B^T \otimes C + C^T \otimes B) \quad (8c)$$

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

$$\nabla\phi(X) = \frac{\partial\phi(X)}{\partial X}$$
$$H\phi(X) = \frac{\partial}{\partial(vecX)^T}\left(\frac{\partial\phi(X)}{\partial(vecX)^T}\right)^T \quad (8d)$$

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; $K_{pn}^T = K_{pn}^{-1} = K_{pn}$ and $K_{pm}(A \otimes C) = (C \otimes A)K_{qn}$, where $A \in \mathbb{R}^{m\times n}$ and $C \in \mathbb{R}^{r\times q}$.

In the next section, dynamical systems for principal singular component analysis for a general rectangular matrix are derived from the optimization of two cost functions, G_1 and G_2 defined as follow:

$$G_1(x,y) = tr(x^T AyD) - \frac{\alpha}{4}tr\{(x^Tx + y^Ty)^2\}, \quad (9)$$

$$G_2(x,y) = tr(x^T AyD) - \frac{\alpha}{4}tr\{(x^Tx)^2 + (y^Ty)^2\}, \quad (10)$$

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.

3 Properties of the Cost Functions

The cost functions G_1 and G_2 can be shown to be upper bounded and both $-G_1$ and $-G_2$ are radially unbounded. Thus gradient systems that converge to the PSC of the given matrix A can be derived.

To gain some insight of the above cost function defined in (2), we consider the scalar case as in the following example.

Example 1: Let $F(x,y) = axyd - \frac{1}{4}(x^2+y^2)^2$, where $a > 0$ and $d > 0$. The objective is to find the minima and maxima of F. The gradient and the Hessian matrix of F can be verified to be

$$\nabla F(x,y) = \begin{bmatrix} ayd - x(x^2+y^2) \\ axd - y(x^2+y^2) \end{bmatrix}, \quad (11a)$$

and

$$\nabla^2 F(x,y) = \begin{bmatrix} -3x^2 - y^2 & ad \\ ad & -3y^2 - x^2 \end{bmatrix}. \quad (11b)$$

The equilibrium points of F are solutions of the equations

$$\begin{aligned} ady &= x(x^2+y^2) \\ adx &= y(x^2+y^2). \end{aligned} \quad (11c)$$

Clearly, the point $(x,y) = (0,0)$ 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 $x = y = \pm\sqrt{ad}$. Note that $y = -x$ yields $da = -2x^2$ which is not possible since ad is positive. This implies that $\nabla F = 0$ at $(x,y) = (0,0)$ and $(x,y) = (\sqrt{\frac{ad}{2}}, \sqrt{\frac{ad}{2}})$. Now, the Hessian matrix evaluated at the equilibrium point $(x,y) = \pm(\sqrt{\frac{ad}{2}}, \sqrt{\frac{ad}{2}})$.

$$\nabla^2 F = \begin{bmatrix} -4x^2 & ad \\ ad & -4y^2 \end{bmatrix} = \begin{bmatrix} -4ad & ad \\ ad & -4ad \end{bmatrix}. \quad (12)$$

The last matrix is negative definite provided that ad is positive. If $(x,y) = (0,0)$, then $HF(0,0) = \begin{bmatrix} 0 & ad \\ ad & 0 \end{bmatrix}$ which is indefinite matrix. Thus $(0,0)$ is a saddle point.

One can make the observation that the function F has one saddle point, and two maximizers $(\pm\hat{x}, \pm\hat{y})$, i.e., if $(\hat{x}, \hat{y})$ is a maximizer then $-(\hat{x}, \hat{y})$ is also a maximizer.

To compute the gradient and Hessian matrices of G_1 and G_2, let $z = \begin{bmatrix} x \\ y \end{bmatrix}$, then G_1 and G_2 can be expressed as

$$G_1(x,y) = G_1(z) = \frac{1}{2}tr(z^T\bar{A}zD) - \frac{\alpha}{4}tr\{(z^Tz)^2\},$$

where $\bar{A} = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}$ is symmetric, and D is diagonal matrix and all its eigenvalues are positive and distinct.

Similarly,

$$G_2(x,y) = G_2(z) = \frac{1}{2}tr(z^T\bar{A}zD) - \frac{\alpha}{4}tr\{(z^T\bar{B}z)^2 + (z^T\bar{C}z)^2\},$$

where $\bar{B} = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$, and $\bar{C} = \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}$.

It follows from Lemma 4 that the first and second order differentials of G_1 are

$$dG_1 = \frac{1}{2}tr\{dz^T\bar{A}dzD + z^T\bar{A}dzD) - \alpha dz^Tzz^Tz\}, \quad (13)$$

$$\begin{aligned} d^2G_1 = \frac{1}{2}tr\{&2dz^TAdzD - \alpha dz^Tdzz^Tz - \alpha dz^Tzdz^Tz \\ &- \alpha dz^Tzz^Tdz\}, \end{aligned} \quad (14a)$$

and those of G_2 are

$$\begin{aligned} dG_2 = \frac{1}{2}tr\{&dz^T\bar{A}dz + z^T\bar{A}dz) - \alpha dz^T\bar{B}zz^T\bar{B}z \\ &- \alpha dz^T\bar{C}zz^T\bar{C}z\}, \end{aligned} \quad (14b)$$

$$\begin{aligned} d^2G_2 = \frac{1}{2}tr\{&2dz^T\bar{A}dz - \alpha dz^T\bar{B}dzz^T\bar{B}z \\ &- \alpha dz^T\bar{B}zdz^T\bar{B}z - \alpha dz^T\bar{B}zz^T\bar{B}dz - \alpha dz^T\bar{C}dzz^T\bar{C}z \\ &- \alpha dz^T\bar{C}zdz^T\bar{C}z - \alpha dz^T\bar{C}zz^T\bar{C}dz\}. \end{aligned} \quad (14c)$$

Therefore, by applying Lemma 5, and assuming that $\alpha = 1$ for convenience, the gradient and the Hessian matrix for G_1 are respectively given by

$$\begin{aligned} \nabla G_1 &= \bar{A}zD + \bar{A}^TzD - z(z^Tz) \\ &= \begin{bmatrix} AyD - x(x^Tx + y^Ty) \\ A^TxD - y(x^Tx + y^Ty) \end{bmatrix}, \end{aligned} \quad (15a)$$

and

$$\begin{aligned} HG_1 = D \otimes &\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} - (x^Tx + y^Ty) \otimes \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \\ &- I \otimes \begin{bmatrix} xx^T & xy^T \\ yx^T & yy^T \end{bmatrix} - K\begin{bmatrix} x \\ y \end{bmatrix} \otimes [x^T \quad y^T]. \end{aligned} \quad (15b)$$

Here K is a permutation matrix such that $K\begin{bmatrix} x \\ y \end{bmatrix} \otimes [x^T \quad y^T] = [x^T \quad y^T] \otimes \begin{bmatrix} x \\ y \end{bmatrix} K_2$, for some permutation matrix K_2.

Similarly, the gradient and the Hessian matrix for G_2 are respectively given by

$$\nabla G_2 = \begin{bmatrix} AyD - x(x^Tx) \\ A^TxD - y(y^Ty) \end{bmatrix}, \quad (16a)$$

and

$$\begin{aligned} HG_2 = D \otimes &\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} - (x^Tx) \otimes \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \\ &- (y^Ty) \otimes \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix} - I \otimes \begin{bmatrix} xx^T & 0 \\ 0 & 0 \end{bmatrix} \\ &- K\begin{bmatrix} x \\ 0 \end{bmatrix} \otimes \begin{bmatrix} x \\ 0 \end{bmatrix} - K\begin{bmatrix} 0 \\ y \end{bmatrix} \otimes [0 \quad y^T], \end{aligned} \quad (16b)$$

for some permutation matrix K.

Thus, gradient dynamical systems for maximizing G_1 and G_2 may be expressed by

$$\begin{aligned} x' &= AyD - \frac{1}{2}x(x^Tx + y^Ty), \\ y' &= A^TxD - \frac{1}{2}y(x^Tx + y^Ty), \end{aligned} \tag{17}$$

and

$$\begin{aligned} x' &= AyD - \frac{1}{2}xx^Tx, \\ y' &= A^TxD - \frac{1}{2}yy^Ty, \end{aligned} \tag{18}$$

respectively.

In the next few sections, the dynamical systems (17) and (18) will be analyzed in terms of stability, convergence, and the limiting behavior as $t \to \infty$.

4 Stability Analysis of the System (17)

The stability of the system (17) can be established as in the following theorem:

Theorem 6. *Let $B \in \mathbb{R}^{N\times N}$ be symmetric invertible matrix and consider the following system*

$$z' = BzD - zz^Tz, \tag{19}$$

where $z \in \mathbb{R}^{N\times p}$, and $D \in \mathbb{R}^{p\times p}$ is diagonal matrix and all its eigenvalues are distinct and positive. Then the equilibrium points $z = \pm u\sqrt{\Sigma D}$ are stable with domain of attraction $\Omega = \{z \in \mathbb{R}^{n\times p} : z \neq 0\}$. The columns of the matrix $u \in \mathbb{R}^{n\times p}$ consist of eigenvectors corresponding to the p largest eigenvalues of B, i.e., $u^TBu = \Sigma$ is diagonal and that the diagonal elements of Σ are ordered so that $tr(\Sigma D)$ is maximum. The zero equilibrium point is unstable. Additionally, the set $\Omega = \{z : z^TB^{-1}z = D\}$ is invariant set for the system (19).

Outline of Proof: To show that the system is stable, let $V(z) = \frac{1}{4}tr\{(z^TB^{-1}z - D)^2\}$, then V is lower bounded and radially unbounded. The time derivative of V is given by

$$\dot{V} = -tr\{(z^TB^{-1}z - D)^2z^Tz\} \leq 0. \tag{20}$$

Since V is radially unbounded, the system (19) is globally stable. Note that $\dot{V} = 0$ if and only if $(z^TB^{-1}z - D)z^T = 0$ or equivalently, $(z^TB^{-1}z)^2 = Dz^TB^{-1}z$. Since D is diagonal and all its eigenvalues are distinct, then $z^TB^{-1}z$ is diagonal. Consequently, if B is positive definite and z is full rank, then $z^TB^{-1}z = D$.

To show that the system (17) is stable we first note that this system follows from (19) by setting $B = \bar{A}$ and $z = \begin{bmatrix} x \\ y \end{bmatrix}$. However, Theorem 6 may not apply since the matrix $\bar{A}$ is symmetric but may not be invertible unless A is a square invertible matrix. To alleviate this problem, the matrix B may be modified so that $B = \bar{A} + \alpha I$ for sufficiently large α, in which case B is positive definite. Hence the new system is

$$\begin{aligned} x' &= AyD - \frac{1}{2}x(x^Tx + y^Ty - \alpha D), \\ y' &= A^TxD - \frac{1}{2}y(x^Tx + y^Ty - \alpha D). \end{aligned} \tag{21}$$

It should also be noticed that the system (21) can be shown to be Lagrange stable for any matrix A by using the function $V(x, y) = \frac{1}{2}tr(x^Tx + y^Ty)$ in which case the time derivative of V along the trajectory of (21) is

$$\dot{V} = \frac{1}{2}tr\{x^TAyD + Dy^TA^Tx - 2(x^Tx)^2 - 2(y^Ty)^2\}. \tag{22}$$

Since the term $tr\{(x^Tx)^2 + (y^Ty)^2\}$ dominates $\frac{1}{2}tr\{x^TAyD + Dy^TA^Tx\}$ for large $||x|| + ||y||$, there exists a number R such that $\dot{V} \leq 0$ for all $(x, y) \in W^c$ where $W = \{(x, y) : x \in \mathbb{R}^{n\times p}, y \in \mathbb{R}^{m\times p} : ||x|| + ||y|| \leq R\}$. Theorem 2 implies that the system is Lagrange stable.

The limiting behavior of the systems (21) may be analyzed as in the following results.

Proposition 7. *Let $A \in \mathbb{R}^{n\times m}$, $D \in \mathbb{R}^{p\times p}$, $x \in \mathbb{R}^{n\times p}$, and $y \in \mathbb{R}^{m\times p}$ and consider the dynamical system (21). Then, the equilibrium points of (17), i.e., the solutions of $\nabla G_1(x, y) = 0$ are*

$$\begin{aligned} \hat{x} &= u\sqrt{\frac{\Sigma D}{2}}, \\ \hat{y} &= v\sqrt{\frac{\Sigma D}{2}}, \end{aligned} \tag{23}$$

where u, v and Σ are as defined in (3) and (4). Let $P = \lim_{t\to\infty} x(t)^Tx(t)$, $Q = \lim_{t\to\infty} y(t)^Ty(t)$, and $B = \lim_{t\to\infty} x(t)^TAy(t)$. Assume that ΣD has distinct eigenvalues, then $P = Q = \Sigma D$ and $B = \Sigma^2 D$. The maximum of G_1 is

$$\frac{1}{2}tr(\Sigma D) = \frac{1}{2}\sum_{i=1}^{p} \sigma_i d_i. \tag{24}$$

Hence the elements of Σ are ordered so that if $d_i < d_j$ then $\sigma_i < \sigma_j$.

Outline of Proof: Let $x(t)$ and $y(t)$ be solutions of (21) for $t \geq 0$, and let P, Q, and B be as defined above. We will prove the result assuming that P and Q are invertible. As $t \to \infty$ we have

$$\begin{aligned} BD &= P(P + Q - \alpha D), \\ B^TD &= Q(P + Q - \alpha D). \end{aligned} \tag{25}$$

For convenience we assumed that $\alpha = 1$. From these equations, we have $P^{-1}B = Q^{-1}B^T$, and hence $B = PQ^{-1}B^T$. Let $S = PQ^{-1}$, then each eigenvalue of S is real and positive, (see Appendix Proposition 11). This implies that $B = SB^T$ and $BS^T = B^T$. By subtracting the last two equations, we obtain $B(I - S^T) = (S - I)B^T$. Let w be an eigenvalue of $I - S$ with corresponding eigenvector λ, then since the eigenvalues of S are real we obtain $\lambda w^T(B + B^T)w = 0$. Since $B + B^T$ is positive definite, then $\lambda = 0$. This shows that the matrix $I - S$ is nilpotent, i.e., $(I - S)^k = 0$ for some k. One can show by induction that $(I - S)^{k-1} = 0$, and therefore $S = I$, or equivalently, $P = Q$. It follows from the equations (25) that $BD = 2P^2$ and $B^TD = 2Q^2 = 2P^2$, and therefore, $B = B^T$, and $BD = DB$. From Proposition 9 (see Appendix), $B = D_1$ for some diagonal matrix D_1. This also shows that P^2 is diagonal. Since it is assumed that all eigenvalue of ΣD are distinct, then $P = Q = \sqrt{\Sigma D}$, and $B = \Sigma^2 D$.

Thus to determine the equilibrium points we may assume that $\hat{x} = u\sqrt{D_1}$ and $\hat{y} = v\sqrt{D_1}$ for some diagonal matrix D_1. From the equation $\nabla G_1(\hat{x}, \hat{y}) = 0$ it follows that

$$A\hat{y}D = \hat{x}(\hat{x}^T\hat{x} + \hat{y}^T\hat{y} - \alpha D), \tag{26a}$$

$$A^T\hat{x}D = \hat{y}(\hat{x}^T\hat{x} + \hat{y}^T\hat{y} - \alpha D). \tag{26b}$$

Therefore,

$$u\Sigma\sqrt{D_1}D = u\sqrt{D_1}(D_1 + D_1 - \alpha D), \tag{27a}$$

$$v\Sigma\sqrt{D_1}D = v\sqrt{D_1}(D_1 + D_1 - \alpha D). \tag{27b}$$

Hence $2D_1 - \alpha D = \Sigma D$, or

$$D_1 = \frac{\Sigma D + \alpha D}{2}.$$

Consequently,

$$\hat{x} = u\sqrt{\frac{\Sigma D + \alpha D}{2}}, \qquad \hat{y} = v\sqrt{\frac{\Sigma D + \alpha D}{2}}. \tag{28}$$

Next, we show that the eigenvalues of HG_1 are non-positive. It will be assumed that $\alpha = 0$. Thus the Hessian matrix at the equilibrium point (23) is

$$\begin{aligned} HG_1 = D \otimes \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} - D\Sigma \otimes \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \\ - I \otimes \begin{bmatrix} uD\Sigma u^T & uD\Sigma v^T \\ vD\Sigma u^T & vD\Sigma v^T \end{bmatrix} \\ - K \begin{bmatrix} u\sqrt{D\Sigma} \\ v\sqrt{D\Sigma} \end{bmatrix} \otimes \begin{bmatrix} \sqrt{D\Sigma}u^T & \sqrt{D\Sigma}v^T \end{bmatrix}. \end{aligned} \tag{29}$$

The eigenvalues associated with the block eigenvector matrix $I \otimes \begin{bmatrix} u_2 & -u_2 \\ v_2 & v_2 \end{bmatrix}$ are the eigenvalues of $D \otimes \begin{bmatrix} 0 & \Sigma_2 \\ -\Sigma_2 & 0 \end{bmatrix} - D\Sigma \otimes \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$ which are either of the form $d_i\lambda_j - d_i\lambda_i = d_i(\lambda_j - \lambda_i)$ or $-d_i\lambda_j - d_i\lambda_i = d_i(\lambda_j + \lambda_i)$. Here λ_i and λ_j are eigenvalues of Σ and Σ_2, respectively. Since $\lambda_j < \lambda_i$, it follows that eigenvalues of (29) corresponding to these eigenvectors are negative. The eigenvalues of the matrix $D \otimes \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}$ corresponding to the eigenvectors $I \otimes \begin{bmatrix} u \\ v \end{bmatrix}$ are those of the matrix $D \otimes \Sigma$. Similarly, the eigenvalues of the matrix $D \otimes \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}$ corresponding to the eigenvectors $I \otimes \begin{bmatrix} -u \\ v \end{bmatrix}$ are those of the matrix $-D \otimes \Sigma$.

Thus to determine the eigenvalues of HG_1 corresponding to the eigenvectors $I \otimes \begin{bmatrix} u \\ v \end{bmatrix}$ and $I \otimes \begin{bmatrix} -u \\ v \end{bmatrix}$, we have the following cases:

1. The eigenvalues of the matrix $D\Sigma \otimes \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$ corresponding to the eigenvectors $I \otimes \begin{bmatrix} u \\ v \end{bmatrix}$ and $I \otimes \begin{bmatrix} -u \\ v \end{bmatrix}$ are those of the matrix $D\Sigma \otimes I$.
2. The eigenvalues of the matrix $I \otimes \begin{bmatrix} u\Sigma Du^T & u\Sigma Dv^T \\ v\Sigma Du^T & v\Sigma Dv^T \end{bmatrix}$ corresponding to the eigenvectors $I \otimes \begin{bmatrix} u \\ v \end{bmatrix}$, and $I \otimes \begin{bmatrix} -u \\ v \end{bmatrix}$ are those of the matrix $I \otimes D\Sigma$.
3. The eigenvalues of the matrix $K \begin{bmatrix} u\sqrt{D\Sigma} \\ v\sqrt{D\Sigma} \end{bmatrix} \otimes \begin{bmatrix} \sqrt{D\Sigma}u^T & \sqrt{D\Sigma}v^T \end{bmatrix}$ corresponding to the eigenvectors $I \otimes \begin{bmatrix} u \\ v \end{bmatrix}$, and $I \otimes \begin{bmatrix} -u \\ v \end{bmatrix}$ are those of the matrices $\pm\sqrt{D\Sigma} \otimes \sqrt{D\Sigma}$.

Thus by combining (1) through (3) it is sufficient to show that the eigenvalues of the matrices

$$D \otimes \Sigma - D\Sigma \otimes I - I \otimes \Sigma D - K\sqrt{\Sigma D} \otimes \sqrt{\Sigma D}, \tag{30}$$

and

$$-D \otimes \Sigma - D\Sigma \otimes -I \otimes \Sigma D - K\sqrt{\Sigma D} \otimes \sqrt{\Sigma D}, \tag{31}$$

are negative or zero. Since $K\sqrt{\Sigma D} \otimes \sqrt{\Sigma D} = \sqrt{\Sigma D} \otimes \sqrt{\Sigma D}K$, and $K^T = K = K^{-1}$, then the eigenvalues of the matrix (30) are of the form $\lambda_j d_i - \lambda_i d_i - \lambda_j d_j \pm \sqrt{\lambda_j d_j \lambda_i d_i} \leq 0$. Similarly, the eigenvalues of the matrix (31) are of the form $-\lambda_j d_i - \lambda_i d_i - \lambda_j d_j \pm \sqrt{\lambda_j d_j \lambda_i d_i} < 0$. This shows that each eigenvalue of HG_1 is negative or zero.

5 Stability Analysis of the System (18)

Similar analysis may be applied to prove stability of the system (18). as indicated in a previous section, this system is based on the gradient of the function G_2 defined in (10). The gradient and the Hessian matrix are given in (16a) and (16b). It will be shown that the system converges to the principal singular components of a matrix A by showing that the eigenvalues of the Hessian matrix HG_2 are negative as in the following result.

Proposition 8. *Let $A \in \mathbb{R}^{n\times m}$, $D \in \mathbb{R}^{p\times p}$, $x \in \mathbb{R}^{n\times p}$, and $y \in \mathbb{R}^{m\times p}$ and consider the function $G_2 : \mathbb{R}^{n\times p} \times \mathbb{R}^{m\times p} \to \mathbb{R}$ defined by $G_2(x,y) = tr\{x^T AyD) - \frac{1}{4}tr\{(x^T x)^2 + (y^T y)^2\}$. Then, the equilibrium points of (18), i.e., the solutions of $\nabla G_2(x,y) = 0$ are*

$$\hat{x} = u\sqrt{\Sigma D}, \qquad \hat{y} = v\sqrt{\Sigma D}, \tag{32}$$

where u, v and Σ are as defined in (3). Let $P = \lim_{t\to\infty} x(t)^T x(t)$, $Q = \lim_{t\to\infty} y(t)^T y(t)$, and $B = \lim_{t\to\infty} x(t)^T Ay(t)$. Then $P = Q = \Sigma D$ and $B = \Sigma^2 D$. The maximum of G_2 is

$$\frac{1}{2}tr(\Sigma D) = \frac{1}{2}\sum_{i=1}^{p} \sigma_i d_i. \tag{33}$$

Hence the elements of Σ are ordered so that if $d_i < d_j$ then $\sigma_i < \sigma_j$.

Proof: We first show that P, Q and B are diagonal and that $P = Q$. As $t \to \infty$ we have

$$BD = P^2, \qquad B^T D = Q^2. \tag{34}$$

Since P and Q are symmetric, then $BD = DB^T$ and $B^T D = DB$. By adding the two equations, we obtain $(B + B^T)D = D(B + B^T)$. From Proposition 9 (see Appendix) it follows that $B + B^T = D_1$ for some diagonal matrix D_1. If the eigenvalues of DD_1 are all distinct, then $P = Q = D_2$ for some diagonal matrix D_2. Hence $BD = DB^T = D(D_1 - B)$ or $BD + DB = DD_1$. From Proposition 10 (see Appendix), we obtain that B is diagonal. Since the eigenvalues of ΣD are assumed distinct, it follows that $P = Q$ is diagonal. To find the equilibrium points of the system, i.e., $\nabla G_2(\hat{x}, \hat{y}) = 0$, assume that $P = Q = D_2$, $\hat{x} = u\sqrt{D_2}$ and $\hat{y} = v\sqrt{D_2}$ for some diagonal matrix D_2. From the equation $\nabla G_2(\hat{x}, \hat{y}) = 0$ it follows that

$$A\hat{y}D = \hat{x}\hat{x}^T\hat{x},$$

$$A^T\hat{x}D = \hat{y}\hat{y}^T\hat{y}.$$

Hence,

$$u\Sigma\sqrt{D_2}D = u\sqrt{D_2}D_2,$$

$$v\Sigma\sqrt{D_2}D = v\sqrt{D_2}D_2,$$

from which it follows that $D_2 = \Sigma D$. This shows that the equilibrium points of the system (18) are of the form $\hat{x} = u\sqrt{\Sigma D}$ and $\hat{y} = v\sqrt{\Sigma D}$.

The Hessian matrix at the equilibrium point (32) is

$$\begin{aligned} HG_2 = D \otimes \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} - \Sigma D \otimes \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \\ - I \otimes \begin{bmatrix} u\Sigma DuT & 0 \\ 0 & v\Sigma Dv^T \end{bmatrix} \\ - K \begin{bmatrix} u\sqrt{\Sigma D} \\ 0 \end{bmatrix} \otimes \begin{bmatrix} \sqrt{\Sigma D}u^T & 0 \end{bmatrix} \\ - K \begin{bmatrix} 0 \\ v\sqrt{\Sigma D} \end{bmatrix} \otimes \begin{bmatrix} 0 & \sqrt{\Sigma D}v^T \end{bmatrix}. \end{aligned} \tag{35}$$

To compute the eigenvalues of $HG_2(\hat{x}, \hat{y})$, one may use an analysis similar to that in the proof of Proposition 7, which shows that all eigenvalues of HG_2 are negative or zero.

Clearly, tThe eigenvalues of the matrix $I \otimes \begin{bmatrix} u\Sigma Du^T & 0 \\ 0 & v\Sigma Dv^T \end{bmatrix}$ corresponding to the eigenvectors $I \otimes \begin{bmatrix} u \\ v \end{bmatrix}$, and $I \otimes \begin{bmatrix} -u \\ v \end{bmatrix}$ are those of the matrix $\pm I \otimes D\Sigma$.

Also, the eigenvalues of the matrix $K \begin{bmatrix} u\sqrt{\Sigma D} \\ 0 \end{bmatrix} \otimes \begin{bmatrix} \sqrt{\Sigma D}u^T & 0 \end{bmatrix}$ corresponding to the eigenvectors $I \otimes \begin{bmatrix} u \\ v \end{bmatrix}$ and $I \otimes \begin{bmatrix} -u \\ v \end{bmatrix}$ are those of the matrix $\sqrt{\Sigma D} \otimes \sqrt{\Sigma D}$.

The eigenvalues of the matrix $K \begin{bmatrix} 0 \\ v\sqrt{\Sigma D} \end{bmatrix} \otimes \begin{bmatrix} 0 \\ \sqrt{\Sigma D}v^T \end{bmatrix}$ corresponding to the eigenvectors $I \otimes \begin{bmatrix} u \\ v \end{bmatrix}$ are those of the matrix $\pm\sqrt{\Sigma D} \otimes \sqrt{\Sigma D}$.

As in Proposition 7, since $K\sqrt{\Sigma D} \otimes \sqrt{\Sigma D} = \sqrt{\Sigma D} \otimes \sqrt{\Sigma D} K$, and $K^T = K = K^{-1}$, then the eigenvalues of the Hessian matrix HG_2 are either of the form $\lambda_j d_i - \lambda_i d_i - \lambda_j d_j \pm \sqrt{\lambda_j d_j \lambda_i d_i} \leq 0$, or $-\lambda_j d_i - \lambda_i d_i - \lambda_j d_j \pm \sqrt{\lambda_j d_j \lambda_i d_i} < 0$. This shows that each eigenvalue of HG_2 is negative or zero.

Finally, the eigenvalues associated with the block eigenvector matrix $I \otimes \begin{bmatrix} u_2 & -u_2 \\ v_2 & v_2 \end{bmatrix}$ are the eigenvalues of $D \otimes \begin{bmatrix} 0 & \Sigma_2 \\ -\Sigma_2 & 0 \end{bmatrix} - D\Sigma \otimes \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$ which are either of the form $d_i\lambda_j - d_i\lambda_i = d_i(\lambda_j - \lambda_i)$ or $-d_i\lambda_j - d_i\lambda_i = d_i(\lambda_j + \lambda_i)$. Here λ_i and λ_j are eigenvalues of Σ and Σ_2, respectively. Since $\lambda_j < \lambda_i$, it follows that eigenvalues of $HG_2(\hat{x}, \hat{y})$ in (35) corresponding to these eigenvectors are negative or zero.

Now, the value of G_2 at the equilibrium points are of the form $\frac{1}{2}\sum_{k=1}^{p} d_k\sigma_k$, which is maximum only if the diagonal elements of Σ are ordered so that if $d_i < d_j$ then $\sigma_i < \sigma_j$.

6 Conclusion

This paper represents an attempt to utilize dynamical system theory for deriving systems that converge to the principal singular components of a given matrix. Specifically, using an unconstrained optimization problems, gradient dynamical systems for computing the principal singular components of arbitrary matrix are derived. Invariant sets and domain of attractions of these systems are determined. Also it shown that these systems are globally convergent in that they converge to the actual singular triplets starting from any full rank initial conditions. The limiting behavior of these systems is influenced by incorporating a diagonal matrix D. It should be stated that the work presented here requires more detailed analysis and generalization. Extension of the proposed rules to complex data and matrices can be achieved with minor modifications.

7 Appendix

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

Proposition 9. *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 10 [13]. *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 11. *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] Hori, G., "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] 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.

[7] A. Edelman, T. A. Arias and S. T. Smith, "The geometry of algorithms with orthogonality constraints" SIAM J. Matrix Anal. Appl., 20(2):303-353, 1998.

[8] Hasan, M.A., "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] J. J. E. Slotine and W. Li. Applied nonlinear Comtrol. Prentice Hall, 1991.

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

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

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

[13] Hasan, M.A., "Natural Gradient for Minor Component Extraction," to be presented at the 2005 International Symposium on Circuits and Systems, May 2005.

[14] Hasan, M.A., "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.