Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Joint Computation of Principal and Minor Components Using Gradient Dynamical Systems Over Stiefel Manifolds

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

Abstract

This paper presents several dynamical systems for simultaneous computation of principal and minor subspaces of a symmetric matrix. The proposed methods are derived from optimizing cost functions which are chosen to have optimal values at vectors that are linear combinations of extreme eigenvectors of a given matrix. Necessary optimality conditions are given in terms of a gradient of certain cost functions over a Stiefel manifold. Stability analysis of equilibrium points of six algorithms is established using Liapunov direct method.

Keywords: Eigenvalue spread, Gradient dynamical systems; Stiefel manifold, Joint PCA-MCA, Joint PSA-MSA, Oja's Rule

1 Introduction

Principal subspace analysis (PSA), and minor subspace analysis (MSA), are essential for many signal processing applications including direction estimation in antenna arrays, data compression, and multiuser detection in wireless communications. Both PSA and MSA require the computation of a few extremal eigenpairs and corresponding eigenspaces of positive definite matrices. Designing learning rules for PSA and MSA has been the focus of many research efforts, see [1]-[6] and numerous references therein. A well-known tool for computing the principal and minor subspace of a data matrix is Oja's rule and several variations of it. A variety of adaptive (on-line) algorithms for PCA or PSA can be found in neural networks literature, see also [7] and references therein.

Many other methods are derived from optimizing Rayleigh and inverse Rayleigh quotients [8]-[10]. In most known methods, either a principal or a minor subspace (or component) but not both are computed. Methods for joint computing eigenspaces corresponding to both maximum and minimum eigenvalues are given in [11]. In this paper, additional methods that expand those of [11] are proposed. These involve algorithms for joint computation of both (PSA and MSA) or (PCA and MCA). Specifically, iterative methods are presented for determining the largest and smallest eigenvalues of a symmetric matrix, and their corresponding eigenvectors, simultaneously.

Let $A \in \mathbb{R}^{n\times n}$ be an arbitrary symmetric matrix, i.e., $A = A^T$, where ^{T} denotes matrix stranspose, and $\mathbb{R}$ is the set of real numbers. Let $\{\lambda_i\}_{i=1}^n$ be the set of eigenvalues of A with associated eigenvectors $\{q_i\}_{i=1}^n$. It will be assumed that q_i is a unit norm eigenvector associated with the eigenvalue λ_i, i.e., $Aq_i = \lambda_i q_i$, $q_i^T q_j = \delta_{ij}$, where δ_{ij} is the Kronecker delta function. Since A is symmetric, all its eigenvalues are real and A has a complete set of orthogonal eigenvectors, i.e., the set $\{q_i\}_{i=1}^n$ is a bais for R^n. It will be assumed throughout that the λ_i's are in decreasing order so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. It will be assumed that $\lambda_1 > \lambda_n$ for otherwise A is an identity matrix.

The quantity $\lambda_1 - \lambda_n$ is sometimes called the eigen-spread of the matrix A. The eigen-spread of a symmetric matrix A may be charaterized by Mirsky result [12] (Theorem 3). Another charaterization of the eigen-spread of the matrix A is given in [11] and is shown to be equivalent to Mirsky result. In this paper, several variations will be derived by generalizing the methods presented in [11]. It will be assumed that $\lambda_1 > \lambda_n$ for otherwise A is an identity matrix.

The following notation will be used throughout. The notation $\mathbb{R}$, and $\mathcal{C}$ denote the set of real numbers, and the set of complex numbers, respectively. The identity matrix of dimension k is expressed with the symbol I_k. The vector e_i denotes the ith column of an identity matix. The magnitude of a vector x will be denoted by $||x|| = \sqrt{x^T x}$. The notation I denotes an identity matrix of appropriate size.The transpose of a real matrix x 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 $tr(B)$, and det(B) denote the trace of B and the determinant of B respectively. Finally, the time derivative of $V(x, y)$ is denoted by $\dot{V}$.

2 Higher Order Eigenvalue Problems

In this section, several results that will be used in the subsequent sections are presented. These results include solving quadratic and higher order eigenvalue problems, and some results regarding the eigen-spread of a symmetric matrix.

Theorem 1. *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix and consider the equation*

$$A^2 z + Az\alpha + x\beta = 0, \qquad (1)$$

978-1-4244-3124-3/08/$25.00 ©2008 IEEE

for some numbers $\alpha, \beta \in \mathbb{R}$, and $z \in \mathbb{R}^{n\times 1}$. Then each nonzero solution of (1) is of the form $z = \gamma_1 q_i + \gamma_2 q_j$ for some numbers γ_1, γ_2.

Proof: Clearly, A has a complete set of eigenvectors $q_1, \cdots, q_n$ since A is symmetric. Thus assume that $z = \sum_{k=1}^{n} \gamma_k q_k$ for some $\gamma_k \in \mathcal{C}$ and $k = 1, \cdots, n$. Hence by substituting in (1) we obtain

$$A^2 z + Az\alpha + x\beta = A^2 \sum_{k=1}^{n} \gamma_k q_k + \alpha A \sum_{k=1}^{n} \gamma_k q_k + (\sum_{k=1}^{n} \gamma_k q_k)\beta = 0.$$

This implies that

$$\gamma_k \lambda_k^2 + \alpha \lambda_k \gamma_k + \gamma_k \beta = \gamma_k(\lambda_k^2 + \alpha\gamma_k + \beta) = 0,$$

for $k = 1, \cdots, n$. Since a quadratic equation has only two zeros (counting multiplicities), it follows that $\gamma_k = 0$ for each $k \in \{1, 2, \cdots, n\}$ except for two indices. Thus assume that $\gamma_i \neq 0$ and $\gamma_j \neq 0$. Consequently, $z = \gamma_i q_i + \gamma_j q_j$, $\beta = \lambda_i \lambda_j$ and $\alpha = -(\lambda_i + \lambda_j)$.

This result can be generalized as follows:

Theorem 2. *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix and consider the equation*

$$A^m z + A^{m-1} z\alpha_1 + \cdots + z\alpha_m = 0, \qquad (2)$$

for some numbers $\alpha_k \in \mathbb{R}$, $k = 1, \cdots, n$, and $z \in \mathbb{R}^{n\times 1}$. Then each nonzero solution of (2) is of the form $z = \sum_{k=1}^{m} \gamma_{i_k} q_{i_k}$ for some numbers γ_{i_k}, $i_k = i_1, i_2, \cdots, i_m \in \{1, \cdots, n\}$.

Next we state and proof a well-known result of Mirsky [12] regarding the eigen-spread of a symmetric matrix. The proof, which is simple and algebraic in nature, is given here since the ideas in the proof can be adopted to derive dynamical systems that converges to linear combinations of minimum and maximum eigenvectors.

Theorem 3(Mirsky [12]). *Let $A \in \mathbb{R}^{n\times n}$ be a positive definite matrix and consider the optimization problem:*

$$\begin{aligned} &\text{Maximize } F_1(x, y) = x^T Ay \\ &\text{subject to} \\ &x, y \in \mathbb{R}^{n\times 1}, x^T x = y^T y = 1, x^T y = 0. \end{aligned} \qquad (3)$$

Then (3) attains its maximum $\lambda_1 - \lambda_n$ at $(x, y) = (\pm\frac{q_1+q_n}{\sqrt{2}}, \pm\frac{q_1-q_n}{\sqrt{2}})$. Similarly, the minimum is $\lambda_n - \lambda_1$ and is attained when $(x, y) = (\pm\frac{q_1+q_n}{\sqrt{2}}, \mp\frac{q_1-q_n}{\sqrt{2}})$.

Proof: Let us define the Stiefel manifold S as

$$S = \{x \in \mathbb{R}^{n\times 2} : x^T x = I_2\}. \qquad (4)$$

As shown in [13], the gradient of F_1 with respect to S is

$$\begin{aligned} \nabla_N F_1 &= \nabla F_1(x) - x(\nabla F_1(x))^T x \\ &= [\,Ay \quad Ax\,] - [\,x \quad y\,][\,y \quad x\,]^T A[\,x \quad y\,]. \end{aligned}$$

Clearly, $\nabla_N F_1 = 0$ implies that $U = Q_{ij}\alpha$, where $Q_{ij} = [\,q_i \quad q_j\,]$, $i \neq j$, and $\alpha \in \mathbb{R}^{2\times 2}$ is orthogonal, i.e., $\alpha^T \alpha = I_2$. Consequently, the equation $\nabla_N F_1 = 0$ yields

$$\Sigma\alpha = \alpha J \alpha^T \Sigma \alpha J,$$

where $J = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} \lambda_i & 0 \\ 0 & \lambda_j \end{bmatrix}$. By pre-multiplying by α^T we obtain

$$\alpha^T \Sigma \alpha = \alpha^T \alpha J \alpha^T \Sigma \alpha J = J \alpha^T \Sigma \alpha J.$$

Since $J^2 = I_2$, it follows that

$$\alpha^T \Sigma \alpha J = J \alpha^T \Sigma \alpha,$$

and therefore, $\alpha^T \Sigma \alpha$ has the matrix form

$$\alpha^T \Sigma \alpha = \begin{bmatrix} r & s \\ s & r \end{bmatrix},$$

for some real numbers r and s. The eigenvalues of the last matrix are $r - s$ and $r + s$ with eigenvectors $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Let α be given as

$$\alpha = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} = [\,\alpha_1 \quad \alpha_2\,]$$

where $\alpha_1 = \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \end{bmatrix}$ and $\alpha_2 = \begin{bmatrix} \alpha_{12} \\ \alpha_{22} \end{bmatrix}$. Then,

$$x^T Ay = \alpha_1^T \Sigma \alpha_2 = \alpha_{11}\alpha_{12}\lambda_1 + \alpha_{21}\alpha_{22}\lambda_2 = \alpha_{11}\alpha_{12}(\lambda_i - \lambda_j).$$

Thus the maximum of $x^T Ay$ occurs when $i = 1$ and $j = n$. It remains to determine the maximum of $\alpha_{11}\alpha_{12}$. Maximizing $\alpha_{11}\alpha_{12}$ leads to the following problem:

$$\begin{aligned} &\text{Max } e_1^T \alpha_1 \alpha_2^T e_1 \\ &\text{subject to } \alpha \in \mathbb{R}^{n\times 2},\ \alpha^T \alpha = I_2 \end{aligned}$$

The necessary condition of optimatity is that:

$$e_1 e_1^T \alpha J - \alpha J \alpha^T e_1 e_1^T \alpha = 0,$$

i.e.,

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix}.$$

After simplifications we obtain

$$\begin{bmatrix} \alpha_{12} & \alpha_{11} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \alpha_{11}^2(\alpha_{21} + \alpha_{12}) & \alpha_{11}\alpha_{12}(\alpha_{21} + \alpha_{12}) \\ \alpha_{11}(\alpha_{21}^2 + \alpha_{11}\alpha_{22}) & \alpha_{12}(\alpha_{21}^2 + \alpha_{11}\alpha_{22}) \end{bmatrix}.$$

This leads to the following four equations:

$$\begin{aligned} \alpha_{12} &= \alpha_{11}^2(\alpha_{21} + \alpha_{12}), \\ \alpha_{11} &= \alpha_{11}\alpha_{12}(\alpha_{21} + \alpha_{12}), \\ 0 &= \alpha_{11}(\alpha_{21}^2 + \alpha_{11}\alpha_{22}), \\ 0 &= \alpha_{12}(\alpha_{21}^2 + \alpha_{11}\alpha_{22}). \end{aligned}$$

From these equations, one concludes that if $\alpha_{21}^2 + \alpha_{11}\alpha_{22} \neq 0$ then $\alpha_{11} = 0$ and $\alpha_{21} = 0$. This contradicts the orthogonality of α. Hence $\alpha_{21}^2 + \alpha_{11}\alpha_{22} = 0$. Similarly, if $\alpha_{21} + \alpha_{12} = 0$, then $\alpha_{11} = 0$ and $\alpha_{21} = 0$. This contradicts the orthogonality of α. It follows from the first two equations that if $\alpha_{11} = 0$, then $\alpha_{12} = 0$, and if $\alpha_{12} = 0$, then $\alpha_{11} = 0$. Hence $\alpha_{11} \neq 0$, $\alpha_{12} \neq 0$, and $\alpha_{12} + \alpha_{21} \neq 0$. It follows from the first two equations that

$$\frac{\alpha_{12}}{\alpha_{11}} = \frac{\alpha_{11}}{\alpha_{12}}.$$

This implies that $\alpha_{12}^2 = \alpha_{11}^2$, or $\alpha_{12} = \pm\alpha_{11}$. Also, $\alpha_{22}^2 = \alpha_{21}^2$, or $\alpha_{21} = \pm\alpha_{22}$. Thus $\alpha_{12}\alpha_{11}$ is maximum when $\alpha_{12} = \alpha_{11}$, in which case $\alpha_{21} = -\alpha_{22}$. Hence α is equal one of the following matrices:

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}. \qquad (5)$$

Note that $\alpha^T J\alpha = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ and $J\alpha^T\Sigma\alpha = \begin{bmatrix} \lambda_1 & 0 \\ 0 & -\lambda_n \end{bmatrix}$ Thus the maximum of $\alpha_{11}\alpha_{12}$ is $\frac{1}{2}$ and subsequently the maximum of $x^T Ay$ is $\frac{\lambda_1-\lambda_n}{2}$ with the maximizers given by

$$[\,x \quad y\,] = Q\alpha = \pm \left[\, \frac{q_1+q_n}{\sqrt{2}} \quad \frac{q_1-q_n}{\sqrt{2}} \,\right].$$

It follows that $q_1 = \pm\frac{x+y}{\sqrt{2}}$ and $q_n = \pm\frac{x-y}{\sqrt{2}}$.

Next we show that the dynamical system based on Mirsky theorem is stable.

2.0.1 Stability Analysis

A gradient dynamical system for solving the optimization problem (3) is

$$\begin{aligned} x' &= \nabla_N F_1 = \nabla F_1(x) - x(\nabla F_1(x))^T x \\ &= [\,Ay \quad Ax\,] - [\,x \quad y\,]\,[\,y \quad x\,]^T A\,[\,x \quad y\,], \end{aligned} \tag{6a}$$

which can be expressed as

$$U' = AUJ - UJU^T AU, \tag{6b}$$

where $U = [\,x \quad y\,]$, and $J = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. The equilibrium points for this system are contained in the set

$$\Omega_{ij} = \{U : U = [\,q_i \quad q_j\,]\,\alpha,\ \alpha \in \mathbb{R}^{2\times 2},\ \alpha^T\alpha = I_2\}. \tag{7}$$

Assume that $\hat{U}$ is a stationary point for the system (6a), then $\hat{U}^T A\hat{U}J = J\hat{U}^T A\hat{U}$ or equivalently, $\alpha^T\Sigma\alpha J = J\alpha^T\Sigma\alpha$, i.e.,$\alpha^T\Sigma\alpha$ is centro-symmetric since $J\alpha^T\Sigma\alpha J = \alpha^T\Sigma\alpha$.

To examine stability of equilibrium points of the system (6a), we first rewrite the equivalent system (6b) as follows:

$$\bar{U}' = (J \otimes A - JU^T AU \otimes I_n)\bar{U}, \tag{8}$$

where $\bar{U}$ is a vector obtained from concatenating the columns of U, and $\otimes$ denotes the Kronecker product operation. The matrix $J \otimes A - JU^T AU \otimes I_n$ can be written as

$$\begin{aligned} &\begin{bmatrix} 0 & A \\ A & 0 \end{bmatrix} - \begin{bmatrix} \frac{\lambda_1-\lambda_n}{2} & \frac{\lambda_1+\lambda_n}{2} \\ \frac{\lambda_1+\lambda_n}{2} & \frac{\lambda_1-\lambda_n}{2} \end{bmatrix} \otimes I_n \\ &= \begin{bmatrix} -\frac{\lambda_1-\lambda_n}{2} I_n & A - \frac{\lambda_1+\lambda_n}{2} I_n \\ A - \frac{\lambda_1+\lambda_n}{2} I_n & -\frac{\lambda_1-\lambda_n}{2} I_n \end{bmatrix}. \end{aligned} \tag{9}$$

which is negative semidefinite since $-\frac{\lambda_1-\lambda_n}{2} < 0$ with the Schur complement [14]

$$R = -\frac{\lambda_1 - \lambda_n}{2} I_n + 2\frac{(A - \frac{\lambda_1+\lambda_n}{2} I_n)^2}{\lambda_1 - \lambda_n}.$$

It can be verified that if λ_i is an eigenvalue of A, then $\frac{2(\lambda_i-\lambda_1)(\lambda_i-\lambda_n)}{\lambda_1-\lambda_n}$ is an eigenvalue of R. Hence all eigenvalues of the matrix of (9) are negative or zeros. Since the matrix $J \otimes A - JU^T AU \otimes I_n$ is symmetric, the geometric and algebraic multiplicity of each eigenvalue are the same. This implies that the system (6a) is stable over the set Ω_{ij} only if $\Omega_{ij} = \Omega_{1n}$.

Another version of Mirsky theorem is derived in [11] which is stated in the next result.

Theorem 4 [11]. *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix and consider the optimization problem:*

$$\begin{aligned} &\text{Maximize } F_2(x, y) = x^T Ax - y^T Ay \\ &\text{subject to } x, y \in \mathbb{R}^{n\times 1},\ x^T x = y^T y = 1, x^T y = 0. \end{aligned} \tag{10}$$

Then (10) attains its maximum at $x = \pm q_n$ and $y = \pm q_1$, and the maximum is $\lambda_n - \lambda_1$. Similarly, the minimum is $\lambda_1 - \lambda_n$ and is attained when $x = \pm q_1$ and $y = \pm q_n$.

This result has been stated and analyzed in [11] using constrained optimization techniques. In this section, we examine the stability of a gradient dynamical system that is based on the gradient of F_2 on the Stiefel's manifold S defined in (4):

$$\begin{aligned} x' &= Ax - xx^T Ax + yy^T Ax \\ y' &= -Ay - xx^T Ay + yy^T Ay. \end{aligned} \tag{11}$$

In more compact form, the system (11) is equivalent to

$$\bar{U}' = (J \otimes A - J\Sigma \otimes I_n)\bar{U}, \tag{12}$$

where $J = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. This system is stable over the set Ω_{1n} since the matrix $J \otimes A - J\Sigma \otimes I_n$ can be shown to be negative semidefinite. Clearly, the matrix

$$J \otimes A - J\Sigma \otimes I_n = \begin{bmatrix} A - \lambda_1 I_n & 0 \\ 0 & -A + \lambda_n I_n \end{bmatrix},$$

is negative semidefinite since its eigenvalues are all negative or zeros.

3 The Propsed Methods

To increase convergence rate, several variations of F_1 and F_2 are considered in the next few sections.

3.1 Algorithm 1

A logarithmic version of Mirsky's theorem is given in the next result.

Theorem 5. *Let $A \in \mathbb{R}^{n\times n}$ be a positive definite matrix and consider the optimization problem:*

$$\begin{aligned} &\text{Maximize } F_3(x, y) = \log(x^T Ay) \\ &\text{subject to} \\ &x, y \in \mathbb{R}^{n\times 1},\ x^T x = y^T y = 1, x^T y = 0. \end{aligned} \tag{13}$$

Then (13) attains its maximum at $x = \pm q_n$ and $y = \pm q_1$, and the maximum is $\log(\frac{\lambda_1}{\lambda_n})$. Similarly, the minimum is $\log(\frac{\lambda_n}{\lambda_1})$ and is attained when $x = \pm q_1$ and $y = \pm q_n$.

A gradient dynamical system that is based on the gradient of F_3 on the Stiefel's manifold S defined in (3) is

$$\begin{aligned} x' &= Ay(x^T Ay)^{-1} - x - y(x^T Ay)^{-1} x^T Ax, \\ y' &= Ax(y^T Ax)^{-1} - y - x(y^T Ay)^{-1} y^T Ay. \end{aligned} \tag{14}$$

The equilibrium points, i.e., the solutions of $\nabla_N F_3 = 0$, are contained in the set Ω_{ij}, $1 \le i, j \le n$.

3.1.1 Stability Analysis

The dynamical system (14) may be rewritten in concise form as

$$U' = AUJ - UJU^T AU, \tag{15}$$

where $U = [\,x \quad y\,]$, and $J_1 = \begin{bmatrix} 0 & \frac{1}{x^T Ay} \\ \frac{1}{y^T Ax} & 0 \end{bmatrix}$. Next we verify that Ω_{1n} is a stable equilibrium set for the system (14). If $\hat{U}$ is a steady state solution of the system (15), then since $x^T Ay = y^T Ax$, then

$$\hat{U}^T A\hat{U} J_1 = J_1 \hat{U}^T A\hat{U},$$

or equivalently,

$$\alpha^T \Sigma \alpha J_1 = J_1 \alpha^T \Sigma \alpha,$$

where $\Sigma = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_n \end{bmatrix}$. This implies that $\alpha^T \Sigma \alpha = \begin{bmatrix} r & s \\ s & r \end{bmatrix}$, i.e., $\alpha^T \Sigma \alpha$ is centro-symmetric. Clearly, the eigenvalues of the matrix $J_1 \otimes A - J_1 U^T AU \otimes I_n$ are same as those of the matrix $\frac{1}{x^T Ay}(J \otimes A - JU^T AU \otimes I_n)$ which are non-positive provided that $x^T Ay > 0$ which is the case since $[x \quad y]$ is a maximizer of F_3.

3.2 Algorithm 2

In this section, a logarithmic version of Algorithm 1 is developed.

Theorem 6. *Let $A \in \mathbb{R}^{n\times n}$ be a positive definite matrix and consider the optimization problem:*

$$\begin{aligned} &\text{Maximize } F_3(x,y) = \log(x^T Ax) - \log(y^T Ay) \\ &\text{subject to } x, y \in \mathbb{R}^{n\times 1},\ x^T x = y^T y = 1, x^T y = 0. \end{aligned} \tag{16}$$

Then (16) attains its maximum at $x = \pm q_n$ and $y = \pm q_1$, and the maximum is $\log(\frac{\lambda_1}{\lambda_n})$. Similarly, the minimum is $\log(\frac{\lambda_n}{\lambda_1})$ and is attained when $x = \pm q_1$ and $y = \pm q_n$.

A gradient dynamical system that is based on the gradient of F_3 on the Stiefel's manifold Ω is

$$\begin{aligned} x' &= Ax(x^T Ax)^{-1} - x + y(y^T Ay)^{-1} y^T Ax \\ y' &= -Ay(y^T Ay)^{-1} + y + x(x^T Ax)^{-1} x^T Ay. \end{aligned} \tag{17a}$$

3.2.1 Stability Analysis

The system (17a) may be written as:

$$U' = AUJ - UJU^T AU, \tag{17b}$$

where $J = \begin{bmatrix} \frac{1}{x^T Ax} & 0 \\ 0 & \frac{-1}{y^T Ay} \end{bmatrix}$. At optimality, the matrix $J = \begin{bmatrix} \frac{1}{\lambda_1} & 0 \\ 0 & \frac{-1}{\lambda_n} \end{bmatrix}$.

It can be shown that the system (17b) is stable over the set Ω_{1n} since it is equivalent to

$$\bar{U}' = (J \otimes A - JU^T AU \otimes I_n)\bar{U}. \tag{18}$$

The eigenvalues of this system are those of the matrix

$$\begin{bmatrix} \frac{A}{\lambda_1} & 0 \\ 0 & \frac{-A}{\lambda_n} \end{bmatrix} - \begin{bmatrix} I_n & 0 \\ 0 & -I_n \end{bmatrix} = \begin{bmatrix} \frac{A}{\lambda_1} - I_n & 0 \\ 0 & \frac{-A}{\lambda_n} + I_n \end{bmatrix}. \tag{19}$$

The eigenvalues of the last matrix are $\frac{\lambda_i}{\lambda_1} - 1$ or $1 - \frac{\lambda_i}{\lambda_n}$ which are both non-positive. There are exactly two zero eigenvalues which occur when $i = 1$ or $i = n$.

3.3 Algorithm 3

If in Theorem 3, the vector y is replaced by $y = \frac{y_1}{\sqrt{y_1^T y_1}}$, where $y_1 = Ax - xx^T Ax$, then $y^T y = 1$, $x^T y = 0$, and $x^T x = 1$. This leads to the following optimization problem:

$$\begin{aligned} &\text{Maximize } F_4(x) = \frac{1}{2} trace\{x^T A^2 x - (x^T Ax)^2\} \\ &\text{subject to } x \in \mathbb{R}^{n\times 1},\ x^T x = 1, \end{aligned} \tag{20}$$

The gradient of F_4 with respect to the Stiefel manifold $\{x \in \mathbb{R}^{n\times 1} : x^T x = 1\}$ is

$$\begin{aligned} \nabla_N F_4 &= \nabla F_4(x) - x(\nabla F_4(x))^T x, \\ &= A^2 x - 2Axx^T Ax - xx^T A^2 x + 2x(x^T Ax)^2. \end{aligned} \tag{21}$$

From Theorem 1, any solution of $\nabla_N F_4 = 0$ is of the form $x = \gamma_i q_i + \gamma_j q_j$, for some numbers γ_i, γ_j such that $\gamma_i^2 + \gamma_j^2 = 1$. Now,

$$\begin{aligned} F_4(x) &= x^T A^2 x - (x^T Ax)^2 \\ &= (\gamma_i^2 - \gamma_i^4)\lambda_i^2 - 2\gamma_i^2\gamma_j^2\lambda_i\lambda_j + (\gamma_j^2 - \gamma_j^4)\lambda_j^2 \\ &= \gamma_i^2\gamma_j^2\lambda_i^2 - 2\gamma_i^2\gamma_j^2\lambda_i\lambda_j + \gamma_i^2\gamma_j^2\lambda_j^2 \\ &= \gamma_i^2\gamma_j^2(\lambda_i - \lambda_j)^2. \end{aligned} \tag{22}$$

Since $\gamma_i^2 + \gamma_j^2 = 1$, the maximum of $\gamma_i^2\gamma_j^2$ occurs when $\gamma_i^2 = \gamma_j^2 = \frac{1}{2}$. Also $(\lambda_i - \lambda_j)^2$ is maximum when $i = 1$ and $j = n$. Thus the maximum of the function F is

$$\frac{(\lambda_1 - \lambda_n)^2}{4},$$

and this maximum occurs when $z = \pm\frac{q_1+q_n}{\sqrt{2}}$ and $z = \pm\frac{q_1-q_n}{\sqrt{2}}$.

Thus depending on initial condition $x(0)$, i.e., whether $x(0)$ has components along the eigenvectors q_1 and q_n, the dynamical system

$$x' = A^2 x - 2Axx^T Ax - xx^T A^2 x + 2x(x^T Ax)^2, \tag{23}$$

may converge to either $z_1 = \frac{q_1+q_n}{\sqrt{2}}$, $z_2 = -\frac{q_1+q_n}{\sqrt{2}}$, $z_3 = \frac{q_1-q_n}{\sqrt{2}}$ or $z_4 = -\frac{q_1-q_n}{\sqrt{2}}$.

Note $\nabla_N F_4(x) = 0$ if x is any normalized eigenvector of A. In this case $F_4(x) = 0$, which is the minimum of F_4 since $F(x) \geq 0$ for all $x \in S$.

3.3.1 Stability Analysis

The system (23) can be shown to be globally stable using the Lyapunov function $V(x) = \frac{1}{4}(x^T x - 1)^2$. The time derivative of V along a trajectory of (23) is

$$\begin{aligned} V' &= x^T A^2 x - 2(x^T Ax)^2 - x^T xx^T A^2 x + 2x^T x(x^T Ax)^2 \\ &= -(x^T x - 1)^2(x^T A^2 x - 2(x^T Ax)^2) \leq 0, \end{aligned} \tag{24}$$

over the set $\{x \in \mathbb{R}^{n\times 1} : x^T x \leq \frac{1}{2}\}$. Hence (23) is stable [15,16].

Let $x(t)$ be a solution of (23) in the interval $[0, \infty)$, where $x(0) = x_0$ is given and has nonzero components along q_1 and q_n. Assume that $x(t) \to z_1$ as $t \to \infty$. It can be verified that $Az_1 - z_1 z_1^T Az_1 = \frac{\lambda_1-\lambda_n}{2\sqrt{2}}(q_1 - q_n) = \frac{\lambda_1-\lambda_n}{2\sqrt{2}} z_3$, and $Az_3 - z_3 z_1^T Az_3 = \frac{\lambda_1-\lambda_n}{2\sqrt{2}}(q_1 + q_n) = \frac{\lambda_1-\lambda_n}{2\sqrt{2}} z_1$. Thus to extract q_1 and q_n, we have $q_1 = \frac{z_1+z_3}{\sqrt{2}}$, and $q_n = \frac{z_1-z_3}{\sqrt{2}}$.

Remark: It can be shown that the system (23) is stable on the set $S_1 = \{x : x^T x = 1, x^T Ax = \frac{\lambda_1+\lambda_n}{2}, \quad x^T A^2 x = \frac{\lambda_1^2+\lambda_n^2}{2}\}$. On the set S_1, the system (23) transforms into

$$\begin{aligned} x' &= A^2 x - (\lambda_1 + \lambda_n)Ax + \lambda_1\lambda_n x \\ &= (A^2 - (\lambda_1 + \lambda_n)A + \lambda_1\lambda_n I_n)x \end{aligned} \tag{25}$$

The matrix $A^2 - (\lambda_1 + \lambda_n)A + \lambda_1\lambda_n I_n$ is negative semi-definite since its eigenvalues are given by $\lambda_i^2 - (\lambda_1 + \lambda_n)\lambda_i + \lambda_1\lambda_n = (\lambda_i - \lambda_1)(\lambda_i - \lambda_n) \leq 0$. Equality holds only when $\lambda_i = \lambda_1$ or $\lambda_i = \lambda_n$.

3.4 Algorithm 4

This algorithm is based on the following maximization problem:

$$\begin{aligned}&\text{Maximize } F_5(x) = \frac{1}{2} trace(x^T A^2 x) - det(x^T Ax)\\ &\text{subject to } x \in \mathbb{R}^{n\times 2},\ x^T x = I_2.\end{aligned} \tag{26}$$

The gradient of F_5 with respect to the Stiefel manifold S defined in (4) is

$$\begin{aligned}&\nabla_N F_5 = \nabla F_5(x) - x(\nabla F_5(x))^T x\\ &= A^2 x - 2det(x^T Ax)Ax(x^T Ax)^{-1} - xx^T A^2 x\\ &+ 2xdet(x^T Ax).\end{aligned} \tag{27}$$

One can show through Lagrange multipliers theory that $\nabla_N F_5 = 0$ if $x = [\,q_i \quad q_j\,]\,\alpha$, where α is orthogonal. However, this solution is not stable unless $\{i, j\} = \{1, n\}$.

3.4.1 Stability Analysis

Now consider the dynamical system

$$x' = \nabla_N F_5(x), \tag{28}$$

or equivalently,

$$\bar{x}' = H\bar{x},$$

where

$$\begin{aligned}&H = I \otimes A^2 - 2det(x^T Ax)(x^T Ax)^{-1} \otimes A - x^T A^2 x \otimes I_n\\ &+ 2det(x^T Ax)I_2 \otimes I_n\end{aligned}$$

$$= I_2 \otimes A^2 - 2det(x^T Ax)(x^T Ax)^{-1} \otimes A - x^T A^2 x \otimes I_n + 2det(x^T Ax)I_2 \otimes I_n$$

It can be shown that the system (28) is stable over the set Ω_{1n}. The matrix H can be expressed as

$$\begin{aligned}H =&(\alpha^T \otimes I_n)\{I_2 \otimes A^2 - 2det(\Sigma)\Sigma^{-1} \otimes A - \Sigma^2 \otimes I_n\\ &+ 2det(\Sigma)I_2 \otimes I_n\}(\alpha \otimes I_n).\end{aligned}$$

Hence the eigenvalues of H are the same as those of the matrix:

$$\begin{bmatrix} A^2 - 2\lambda_n A - \lambda_1^2 I_n + 2\lambda_1\lambda_n I_n & 0\\ 0 & A^2 - 2\lambda_1 A - \lambda_n^2 I_n + 2\lambda_1\lambda_n I_n\end{bmatrix}.$$

Let λ_i be an eigenvalue of A, then each eigenvalue of H is of the form

$$\lambda_i^2 - 2\lambda_1\lambda_i - \lambda_n^2 + 2\lambda_1\lambda_n = (\lambda_i - \lambda_n)(\lambda_i + \lambda_n - 2\lambda_1),$$

or the form

$$\lambda_i^2 - 2\lambda_n\lambda_i - \lambda_1^2 + 2\lambda_1\lambda_n = (\lambda_i - \lambda_1)(\lambda_i + \lambda_1 - 2\lambda_n).$$

Clearly, each eigenvalue of H is negative except when $i = 1$ or $i = n$, in which case, these quantities are zero.

3.5 Algorithm 5

This algorithm is derived from the following optimization problem:

$$\begin{aligned}&\text{Maximize } F_6(x) = \frac{1}{2}\{trace(x^T Ax)\}^2 - 2det(x^T Ax)\\ &\text{subject to } x \in \mathbb{R}^{n\times 2},\ x^T x = I_2.\end{aligned} \tag{29}$$

The gradient of F_6 with respect to the Stiefel manifold $S = \{x \in \mathbb{R}^{n\times r} : x^T x = I_2\}$ is

$$\begin{aligned}&\nabla_N F_6 = \nabla F_6(x) - x(\nabla F_6(x))^T x\\ &= A^2 x - 2det(x^T Ax)Ax(x^T Ax)^{-1} - xx^T A^2 x + 2xdet(x^T Ax)\end{aligned} \tag{30}$$

Note $\nabla_N F_6 = 0$ if $x = [\,q_i \quad q_j\,]\,\alpha$ for some orthogonal matrix α. Stability of stationary points can be analyzed using the following dynamical system:

$$\begin{aligned}&x' = \nabla F_6(x) - x(\nabla F_6(x))^T x\\ &= trace(x^T Ax)Ax - 2Axdet(x^T Ax)Ax(x^T Ax)^{-1}\\ &- xx^T Axtrace(x^T Ax) + 2xdet(x^T Ax).\end{aligned} \tag{31a}$$

3.5.1 Stability Analysis

One way to show that the system (31a) is stable over the set $S = \{x \in \mathbb{R}^{n\times 2} : x = [\,q_1 \quad q_n\,]\,\alpha,\ \alpha^T\alpha = I_2\}$, is to examine the associated matrix

$$\begin{aligned}&I_2 \otimes A^2 - 2det(x^T Ax)(x^T Ax)^{-1} \otimes A - x^T A^2 x \otimes I_n\\ &+ 2det(x^T Ax)I_2 \otimes I_n,\end{aligned} \tag{31b}$$

for definiteness. Local stability of this system over the set Ω_{1n} follows from analyzing the linearized system. The matrix version of (31b) is

$$\begin{bmatrix} tr(x^T Ax)A & 0\\ 0 & tr(x^T Ax)A\end{bmatrix} - 2\begin{bmatrix} \frac{\lambda_1\lambda_n}{\lambda_1}A & 0\\ 0 & \frac{\lambda_1\lambda_n}{\lambda_1}A\end{bmatrix}$$
$$- \begin{bmatrix} \lambda_1 tr(x^T Ax)I_n & 0\\ 0 & \lambda_n tr(x^T Ax)I_n\end{bmatrix} + 2\begin{bmatrix} \lambda_1\lambda_n I_n & 0\\ 0 & \lambda_1\lambda_n I_n\end{bmatrix}. \tag{31c}$$

For each eigenvalue λ_1 of the matrix A, there are two eigenvalues of the matrix (31c) given by:

$$(\lambda_1+\lambda_n)\lambda_i - 2\lambda_n\lambda_i - \lambda_1(\lambda_1+\lambda_n) + 2\lambda_1\lambda_n = (\lambda_i - \lambda_1)(\lambda_1 - \lambda_n) \le 0,$$

and

$$(\lambda_1+\lambda_n)\lambda_i - 2\lambda_1\lambda_i - \lambda_n(\lambda_1+\lambda_n) + 2\lambda_1\lambda_2 = (\lambda_i - \lambda_n)(\lambda_n - \lambda_1) \le 0.$$

Thus all eigenvalues of the linearized system are non-positive and each zero eigenvalue has geometric multiplicity 1. This indicates that the system (31a) is stable over Ω_{1n}.

3.6 Algorithm 6

Finally, the last algorithm is derived from the gradient system associated with the following optimization problem:

$$\begin{aligned}&\text{Maximize } F_7(x) = \frac{1}{2} trace\{(x^T Ax)^2\} - 2det(x^T Ax)\\ &\text{subject to } x \in \mathbb{R}^{n\times 2}\ x^T x = I_2.\end{aligned} \tag{32}$$

The gradient of F_7 with respect to the Stiefel manifold $S = \{x \in \mathbb{R}^{n\times r} : x^T x = I_2\}$ is

$$\begin{aligned}&\nabla_N F_7 = \nabla F_7(x) - x(\nabla F_7(x))^T x\\ &= Axx^T Ax - 2det(x^T Ax)Ax(x^T Ax)^{-1} - x(x^T Ax)^2\\ &+ 2xdet(x^T Ax)\end{aligned} \tag{33}$$

The associated dynamical system is

$$\begin{aligned}&x' = \nabla F_7(x) - x(\nabla F_7(x))^T x\\ &= Axx^T Ax - det(x^T Ax)Ax(x^T Ax)^{-1} - x(x^T Ax)^2\\ &+ xdet(x^T Ax)\end{aligned} \tag{34}$$

Clearly, if x is a solution of the equation $\nabla_N F_7 = 0$, then $x \in \Omega_{ij}$, $1 \le i, j \le n$. In the next section Ω_{1n} is shown to be stable equilibrium set for the system (34).

3.6.1 Stability Analysis

The system (34) can be expressed as

$$\bar{x}' = H\bar{x},$$

where

$$H = x^T Ax \otimes A - det(x^T Ax)(x^T Ax)^{-1} \otimes A - (x^T Ax)^2 \otimes I_n + det(x^T Ax) \otimes I_n.$$

If $x \in \Omega_{1n}$, then the matrix H can be equivalently written as

$$\begin{aligned} H &= \alpha^T \Sigma\alpha \otimes A - (\alpha^T \Sigma\alpha)^{-1} det(\alpha^T \Sigma\alpha) \otimes A \\ &- \alpha^T \Sigma^2 \alpha \otimes I_n + det(\alpha^T \Sigma\alpha) I_2 \otimes I_n \\ &= (\alpha^T \otimes I_n)\{\Sigma \otimes A - \Sigma^{-1} det(\Sigma) \otimes A \\ &- \Sigma^2 \otimes I_n + det(\Sigma) I_2 \otimes I_n\}(\alpha \otimes I_n). \end{aligned}$$

This implies that the eigenvalues of H are those of the matrix

$$\begin{bmatrix} \lambda_1 A & 0 \\ 0 & \lambda_n A \end{bmatrix} - \begin{bmatrix} \frac{\lambda_1\lambda_n}{\lambda_1} A & 0 \\ 0 & \frac{\lambda_1\lambda_n}{\lambda_1} A \end{bmatrix} - \begin{bmatrix} \lambda_1^2 I_n & 0 \\ 0 & \lambda_n^2 I_n \end{bmatrix} + \begin{bmatrix} \lambda_1\lambda_n I_n & 0 \\ 0 & \lambda_1\lambda_n I_n \end{bmatrix},$$

$$= \begin{bmatrix} \lambda_1 A - \lambda_n A - \lambda_1^1 I_n + \lambda_1\lambda_n I_n & 0 \\ 0 & \lambda_n A - \lambda_1 A - \lambda_n^1 I_n + \lambda_1\lambda_n I_n \end{bmatrix}.$$

Corresponding to each eigenvalue λ_i of A, there exist two eignvalues of H given by

$$\lambda_1\lambda_i - \lambda_n\lambda_i - \lambda_1^2 + \lambda_1\lambda_n = (\lambda_1 - \lambda_n)(\lambda_i - \lambda_1) \le 0,$$

or

$$\lambda_n\lambda_i - \lambda_1\lambda_i - \lambda_n^2 + \lambda_1\lambda_n = (\lambda_n - \lambda_1)(\lambda_i - \lambda_n) \le 0.$$

This shows that all eigenvalues of the symmetric matrix H are negative except two which occur when $\lambda_i = 1$ or $\lambda_i = \lambda_n$.

4 Conclusion

Several dynamical systems for joint computation of minimum and maximum eigenvalues and corresponding eigenvectors are developed. These systems are extensions of the author work [11] for extracting extreme eigenvectors simultaneously. The main challenge in deriving these systems is to come up with the appropriate cost function. Local stability of the proposed systems is established using Lyapunov direct method. There remain many issues to be examined including global convergence and stability. It should be stated that more rigorous proofs of stability can be established using results from central manifold theory. Preliminay simulations, which are not shown here due to space limits, have shown that these systems converge from a wide range of intitial conditions demonstrating global convergence. Also, it is observed that incorporating penalty terms in the cost functions F_1-F_4 leads to significant acceleration in convergence.

References

[1] E. Oja, and J. Karhunen, "On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix", J. Math. Anal. Appl., 1985, 106, pp. 69-84.

[2] E. Oja, "Principal components, minor components, and linear neural networks. Neural Networks," 5:927-935, November 1992.

[3] G. Cirrincione, M. Cirrincione, J. Herault, and S. Van Huffel, "The MCA EXIN Neuron for the Minor Component Analysis," IEEE Trans. on Neural Networks, Vol. 13, No. 1, pp. 160-187, January 2002.

[4] T. D. Sanger, "Optimal unsupervised learning in a singlelayer linear feedforward network," Neural Networks, 2:459-473, 1989.

[5] L. Xu, "Least mean square error recognition principle for self organizing neural nets," Neural Networks, 6:627-648, 1993.

[6] J. H. Manton, Uwe Helmke, and Iven M. Y. Mareels, "Dynamical Systems for Principal and Minor Component Analysis," Proceedings of the 42nd IEEE Conference on Decision and Control, Maui, Hawaii USA, pp. 1863-1868, December 2003.

[7] K. I. Diamantaras and S. Y. Kung, Principal Component Neural Networks: Theory and Applications, John Wiely & Sons, INC, 1996.

[8] M.A. Hasan, "Natural Gradient for Minor Component Extraction," proceeding of the 2005 International Symposium on Circuits and Systems, May 2005.

[9] 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 vol. 2.

[10] M.A. Hasan, "Revisiting Weighted Inverse Rayleigh Quotient for Minor Component Extraction," Proceedings of the Fifth International Conference on Information, Communications and Signal Processing (ICICS 2005), December 6 - 9, 2005.

[11] M.A. Hasan, "Dynamical systems for joint principal and minor component analysis," American Control Conference, 14-16 June 2006, Page(s):6 pp.

[12] L. Mirsky. The spread of a matrix. Mathematika, 3:127-130, 1956.

[13] A. Edelman, T.A. Arias, and S.T. Smith, "The geometry of algorithms with orthogonality constraints," SIAM J. Matrix Anal. Appl., vol. 20, pp. 303-353, 1998.

[14] Arthur Albert, Conditions for positive and nonnegative definiteness in terms of pseudoinverses, SIAM Journal on Applied Mathematics, 17 (1969), pp. 434-440.

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

[16] J. J. E. Soltine and W. Li. Applied nonlinear Comtrol. Prentice Hall, 1991.