On the Fixed Points of Gradient Flows on Orthogonal Groups

Gen Hori

Brain Science Institute, RIKEN, Saitama 351-0198, Japan
hori@bsp.brain.riken.go.jp

Abstract: The fixed points of two known gradient flows defined on adjoint orbits of orthogonal groups are analyzed through the critical point analysis of the potential functions. The results show that some known properties of these gradient flows are shared with the gradient flows of the same potential functions with respect to other metrics.

1. Introduction

Over the past decade, several attempts have been made on solving matrix optimization problem appeared in some aspects of engineering using continuous matrix dynamical systems, especially matrix gradient flows[9]. Such matrix optimization problem include eigenvalue decomposition (EVD), independent component analysis (ICA), blind source separation (BSS) and joint approximate diagonalization (JAD). Comparing to other kinds of numerical methods, the approach using continuous dynamical system is suitable for adaptive applications, such as signal processing methods for tracking moving targets.

The most of dynamical systems associated with the eigenvalue decomposition of real symmetric matrices or complex Hermitian matrices are defined on the orthogonal group or the unitary group. For example, Brockett's flow which solve eigenvalue decomposition of real symmetric matrices can be regarded as a dynamical system on the orthogonal group. ICA problem for prewhitened observed data is to find orthogonal demixing matrix which maximize the independence between channels. Therefore the matrix dynamical systems for ICA are defined on the orthogonal group.

The aim of present work is to analyze the fixed points of gradient flows defined on orthogonal groups through the critical point analysis of the potential functions, using local coordinate systems defined by plane rotations. The fixed points are classified according to the index numbers which represent the stabilities of the dynamics around the points.

2. Fixed Point Analysis

In this section, we analyze the fixed points of two gradient flows defined on the adjoint orbit of the orthogonal group $SO(n)$,

$$\Omega = \{G^{-1}A_0G \mid G \in SO(n)\}.$$

The gradient equation of the potential function

$$\phi(A) = \phi(G^tA_0G)$$

restricted on the adjoint orbit Ω is given as

$$\dot{A} = -[A, \pi[A^t, (\frac{\partial\phi(A)}{\partial A})]] \qquad (1)$$

where $[A, B] = AB-BA$ is a commutator product and π means projection to the space of skew-symmetric matrices. It is known that the right hand side of (1) vanishes if and only if

$$\pi[A^t, (\frac{\partial\phi(A)}{\partial A})] = 0$$

holds. (See [1][5].) The solution of (1) evolves inside the adjoint orbit Ω and $A(t)$ at each time point t is orthogonally similar to the initial value $A(0) = A_0$.

We analyze the fixed points of gradient flows on the adjoint orbit using the following coordinate systems. Let E_{ij} denote an $n \times n$ matrix whose (i, j)-th element is 1 and all the other elements are 0's and put

$$X_{ij} = E_{ji} - E_{ij} \quad (i < j).$$

Then a plane rotation is expressed as an exponential of X_{ij},

$$e^{\vartheta X_{ij}} = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & \cos\vartheta & & -\sin\vartheta & & \\ & & & \ddots & & & \\ & & \sin\vartheta & & \cos\vartheta & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}.$$

We introduce a local coordinate system

$$\theta = (\theta_{12}, \theta_{13}, \dots, \theta_{1n}, \theta_{23}, \dots, \theta_{2n}, \dots, \theta_{n-1,n})$$

on $SO(n)$ around the identity by

$$\begin{aligned} G(\theta) &= e^{\theta_{12}X_{12}} \cdot e^{\theta_{13}X_{13}} \dots e^{\theta_{1n}X_{1n}} \\ &\quad \cdot e^{\theta_{23}X_{23}} \dots e^{\theta_{2n}X_{2n}} \\ &\quad \dots \\ &\quad \cdot e^{\theta_{n-1,n}X_{n-1,n}} \end{aligned}$$

and on the adjoint orbit Ω around A_0 by

$$A(\theta) = G(\theta)^{-1}A_0G(\theta).$$

Then the first order and second order derivatives on the adjoint orbit Ω at A_0 is calculated as

$$\partial_{ij}A(\theta)|_{\theta=0} = [A_0, X_{ij}],$$

$$\partial_{ij}\partial_{kl}A(\theta)|_{\theta=0} = [[A_0, X_{ij}], X_{kl}]$$

where ∂_{ij} denotes the differentiation with respect to θ_{ij},

$$\partial_{ij} = \frac{\partial}{\partial \theta_{ij}}.$$

Note that the order of X_{ij} and X_{kl} in the right hand side of the second order derivative is the same as the order appeared in the definition of $G(\theta)$ and does not depend on the order of differentiation. We introduce local coordinates systems around arbitrary points A on the adjoint orbit Ω through the left action of the group, in which we have

$$\partial_{ij}A = [A, X_{ij}],$$
$$\partial_{ij}\partial_{kl}A = [[A, X_{ij}], X_{kl}].$$

2.1 Brockett's equation

Let $N = \mathrm{diag}(n_1, n_2, \ldots, n_n)$ be a fixed real diagonal matrix with distinct diagonal elements, the potential function be defined as

$$\phi_1(A) = -\sum_p n_p a_{pp}$$

where a_{ij} denotes the (i,j)-th element of A and the initial value $A(0)$ be a real symmetric matrix with distinct eigenvalues in (1). Then we obtain

$$\dot{A} = [A, [A, N]]. \qquad (2)$$

This dynamical system was introduced by Brockett[4]. (Note that the projection to the space of skew symmetric matrices π can be omitted in (2) because $[A, N]$ is skew symmetric for $^\forall t$.)

We first re-derive basic properties of the dynamical system on its fixed points and their stabilities directly from the critical point analysis of the potential function. After that, we derive our result on the number of the fixed point with each index number (degree of instability).

The critical points of the potential function $\phi_1(A)$ satisfy

$$\partial_{ij}\phi_1(A) = -\sum_p n_p\, \partial_{ij} a_{pp} = 0 \quad (i < j).$$

Using

$$\partial_{ij}A = [A, X_{ij}] = \begin{pmatrix} & a_{1j} & -a_{1i} & \\ & a_{2j} & -a_{2i} & \\ & \vdots & \vdots & \\ & & & \\ & a_{nj} & -a_{ni} & \end{pmatrix} - \begin{pmatrix} & & & \\ -a_{j1} & -a_{j2} & \cdots & -a_{jn} \\ & & & \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ & & & \end{pmatrix}$$

and the symmetry of A, we obtain

$$-2(n_i - n_j)a_{ij} = 0 \quad (i < j).$$

Because n_i's are distinct, it follows that all the critical points of the potential function $\phi_1(A)$, which are the fixed points of (2), are diagonal matrices whose diagonal elements are the eigenvalues of the initial matrix $A(0)$. Then the number of the fixed points of (2) is $n!$ if the eigenvalues of the initial matrix $A(0)$ are distinct.

The Hessian of the potential function

$$\partial_{ij}\partial_{kl}\phi_1(A) = -\sum_p n_p\, \partial_{ij}\partial_{kl} a_{pp}$$

at a critical point $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is calculated as

$$\begin{cases} 2(\lambda_i - \lambda_j)(n_i - n_j) & (i = k, j = l) \\ 0 & (\text{otherwise}) \end{cases}$$

using

$$\begin{aligned} \partial_{ij}\partial_{kl}A &= [[\Lambda, X_{ij}], X_{kl}] \\ &= \begin{cases} -2(\lambda_i - \lambda_j)(E_{ii} - E_{jj}) & (i = k, j = l) \\ (\lambda_i - \lambda_j)(E_{jl} + E_{lj}) & (i = k, j \neq l) \\ -(\lambda_i - \lambda_j)(E_{ik} + E_{ki}) & (i \neq k, j = l) \\ (\lambda_i - \lambda_j)(E_{il} + E_{li}) & (k = j, i < l) \\ 0 & (\text{otherwise}) \end{cases} \end{aligned}.$$

The obtained Hessian is already diagonalized and its $n(n-1)/2$ diagonal elements are given as

$$2(\lambda_i - \lambda_j)(n_i - n_j) \quad (i < j).$$

A critical point is a stable fixed point of (2) if all the diagonal elements of the Hessian at the critical point are positive when it is diagonalized, which is satisfied when λ_i's and n_i's are similarly ordered. Therefore, exactly one among $n!$ diagonal matrices is stable fixed point of (2).

The number of the negative diagonal elements of the Hessian at a critical point when it is diagonalized is the index number which indicates the degree of instability of the fixed point. The following theorem is the basis of our result on the number of the fixed point of (2) with each index number.

Theorem 1. Assume that λ_i's are in ascending order, $\lambda_1 < \lambda_2 < \cdots < \lambda_n$. Let $C_k^{(n)}$ denote the number of the permutations on $\{1, 2, \ldots, n\}$ for which there are k index pairs (i,j) with $i < j$ and $\lambda_{\pi(i)} > \lambda_{\pi(j)}$, where π denotes the permutation. Then the polynomial of t whose k-th coefficient is $C_k^{(n)}$ is given as,

$$\sum_{k=0}^{n(n-1)/2} C_k^{(n)} t^k = \prod_{i=1}^{n-1}\left(\sum_{j=0}^{i} t^j\right). \qquad (3)$$

Proof. We prove by induction.
i) For the case of $n = 2$, obviously we have $C_0^{(2)} = 1$ and

$C_1^{(2)} = 1$ then the polynomial is $1+t$. This shows that (3) holds for $n = 2$.
ii) We assume that (3) holds for $\lambda_1 < \cdots < \lambda_{n-1}$ and the permutations on $\{1, \ldots, n-1\}$, and prove that (3) holds also for $\lambda_1 < \cdots < \lambda_n$ and the permutations on $\{1, \ldots, n\}$. There are n positions to insert the largest value λ_n to $\lambda_{\pi(1)}, \ldots, \lambda_{\pi(n-1)}$. Inserting λ_n to the right end causes nothing on the numbers of the index pairs with reversed order. Inserting λ_n between $\lambda_{\pi(n-1)}$ and $\lambda_{\pi(n-2)}$ adds 1 to the numbers of the index pairs with reversed order. From these observations, we see that the right hand side of (3) for n is obtained by multiplying the right hand side for $n-1$ by $(1+t+\cdots+t^{n-1})$. ■

Using the above theorem, the number of the fixed points of (2) whose index number is k, denoted by C_k, can be obtained by differentiation or straightforward calculation of the polynomial of t. For example, the polynomial for $n = 4$ is calculated as

$$\begin{aligned}(1+t)(1+t+t^2)(1+t+t^2+t^3)\\ = 1+3t+5t^2+6t^3+5t^4+3t^5+t^6\end{aligned}$$

which yields $C_0 = 1$, $C_1 = 3$, $C_2 = 5$, $C_3 = 6$, $C_4 = 5$, $C_5 = 3$ and $C_6 = 1$. Substituting $t = 1$ in the right hand side of (3) gives the total number of the permutations on $\{1, 2, \ldots, n\}$. It is easy to see that C_k's remain the same for the case that n_i's are not in ascending order, as long as both n_i's and λ_i's have distinct values respectively.

2.2 Gradient equation of off-diagonals

Let the potential function defined as

$$\phi_2(A) = \mathbf{off}(A) = \sum_{1 \le i \ne j \le n} {a_{ij}}^2$$

and the initial value $A(0)$ be a real symmetric matrix with distinct eigenvalues in (1). Then we obtain

$$\dot{A} = [A, [A, \mathrm{diag}(A)]]. \qquad (4)$$

This dynamical system was studied by Faybusovich[7] and Nakamura[8]. (The projection operator π can be omitted because $[A, \mathrm{diag}(A)]$ is skew symmetric for $^\forall t$.) It is shown that the solution of (4) converges to a diagonal matrix whose diagonal elements are the eigenvalues of the initial value $A(0)$, for almost all initial values.

We first identify the fixed points of the dynamical system through the critical point analysis of the potential function and then examine the properties of the fixed points.

We note that the gradient equation of the potential function $\phi_2(A)$ is equivalent to the gradient equation of the following potential function

$$\tilde{\phi}_2(A) = -\sum_p {a_{pp}}^2$$

since the squared sum of all the elements of A is invariant under similar transformation by orthogonal matrices. The critical points of the potential function $\tilde{\phi}_2(A)$ satisfy

$$\partial_{ij}\tilde{\phi}_2(A) = -2\sum_p a_{pp}\,\partial_{ij}a_{pp} \quad (i<j).$$

From this, using

$$\partial_{ij}A = [A, X_{ij}]$$

and the symmetry of A, we obtain

$$-4(a_{ii} - a_{jj})a_{ij} = 0 \quad (i<j).$$

It follows that all the critical points of the potential function $\tilde{\phi}_2(A)$, which are the fixed points of the dynamical system (4), satisfy $a_{ij} = 0$ or $a_{ii} = a_{jj}$ for $i < j$.

From the above observation, the fixed points of (4) are classified into the following four categories :

1. a diagonal matrix
2. a matrix whose diagonal elements are all equal
3. a block diagonal matrix whose diagonal blocks are of type 1 or type 2
4. a matrix obtained by applying the same permutation to the rows and the columns of a matrix of type 3

We denote the variance of the diagonal elements of A (multiplied by n) by $\mathbf{vardiag}(A)$,

$$\begin{aligned}\mathbf{vardiag}(A) &= \sum_i (a_{ii} - \mathrm{tr}(A)/n)^2\\ &= \sum_i {a_{ii}}^2 - \mathrm{tr}(A)^2/n.\end{aligned}$$

Obviously we have the following,

$$\mathbf{vardiag}(A) \ge 0, \quad \mathbf{off}(A) \ge 0$$

$$\mathbf{vardiag}(A) + \mathbf{off}(A) = \sum_{i,j} {a_{ij}}^2 - \mathrm{tr}(A)^2/n.$$

The last equality shows that $\mathbf{vardiag}(A) + \mathbf{off}(A)$ is invariant under similar transformations by orthogonal matrices. Therefore the dynamical system (4) can be regarded as the steepest ascent equation of $\mathbf{vardiag}(A)$ as well as the steepest descent equation of $\mathbf{off}(A)$.

Observing $\mathbf{off}(A) = 0$ holds for the type 1 fixed points and $\mathbf{vardiag}(A) = 0$ does for the type 2 fixed points, we see that the type 1 fixed points correspond to the global minima of the potential function $\phi_2(A)$ and the type 2 fixed points do to the global maxima. The type 1 fixed points exist for any initial matrix $A(0)$ and are isolated points. On the other hand, the following lemma guarantees the existence of the type 2 fixed points for any initial matrix.

Lemma 1. Any diagonal matrix can be transformed to a matrix with equal diagonal elements by an orthogonally similar transformation,

$$G^{-1}\Lambda G = \begin{pmatrix} \bar{\lambda} & * & * & \cdots & * \\ * & \ddots & * & \cdots & * \\ * & * & & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & * \\ * & * & \cdots & * & \bar{\lambda} \end{pmatrix}$$

where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$, $G \in SO(n)$, $\bar{\lambda} = (\lambda_1 + \lambda_2 + \cdots + \lambda_n)/n$ and $*$ means an arbitrary value.

The type 3 and type 4 fixed points are not global minima or global maxima of $\phi_2(A)$. It is expected for the Hessian of the potential function at the type 3 and type 4 fixed points to have both of negative and positive diagonal elements when it is diagonalized.

3. Concluding Remarks

The properties of the fixed points of two gradient equations on orthogonal groups have been studied through the critical point analysis of their potential functions. These observations show that some properties of these gradient equations are not specific to the equations but are shared with other gradient equations of the same potential functions with respect to other metrics on orthogonal groups. The fixed points of Brockett's equation (2) have been classified according to the index number. The fixed points of the gradient equation of off-diagonals (4) have been classified into four categories and the stabilities of the fixed points of each category are discussed. We note that there is a certain inequality between the number of the fixed points of the gradient equations with specified index numbers and the rank of the free part of the homology groups of the orthogonal groups on which the gradient equations are defined.

References

[1] G. Hori, "Isospectral gradient flows for non-symmetric eigenvalue problem", *Japan J. Indust. Appl. Math.*, 17, pp.27–42, 2000.

[2] G. Hori, "Isospectral flows expressed in multiple bracket forms", *Japan J. Indust. Appl. Math.*, 14, pp.315–327, 1997.

[3] G. Hori, "A new approach to joint diagonalization", *Proc. 2nd Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)*, pp.151–155, Helsinki, Finland, 2000.

[4] R. W. Brockett, "Dynamical systems that sort lists, diagonalize matrices and solve linear programming problems", *Linear Algebra Appl.*, 146, pp.79–91, 1991.

5] R. W. Brockett, "Differential geometry and the design of gradient algorithms", in *Differential Geometry*, 1, R. E. Green and S.-T. Yau, eds., Amer. Math. Soc., Providence, pp.69–91, 1993.

6] P. Deift, T. Nanda and C. Tomei, "Ordinary differential equations for the symmetric eigenvalue problem", *SIAM J. Numer. Anal.*, 20, pp.1–22, 1983.

7] L. Faybusovich, "Dynamical systems which solve optimization problems with linear constraints", *IMA J. Math. Control Infomation*, 8, pp.135–149, 1991.

] Y. Nakamura, "A new nonlinear dynamical system that leads to eigenvalues", *Japan J. Indust. Appl. Math.*, 9, pp.133–139, 1992.

[9] U. Helmke and J. B. Moore, "Optimization and dynamical systems", Springer-Verlag, London, 1994.

[10] A. Bunse-Gerstner, R. Byers and V. Mehrmann, "Numerical methods for simultaneous diagonalization", *SIAM J. Mat. Anal. Appl.*, 14, pp.927–949, 1993.

[11] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals", *IEEE Proc.-F*, 140, pp.362–370, 1993.

[12] J.-F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization", *SIAM J. Mat. Anal. Appl.*, 17, pp.161-164, 1996.

[13] L. Morgedey and H. G. Schuster, "Separation of a mixture of independent signals using time delayed correlations", *Phys. Rev. Lett.*, 72, pp.3634-3637, 1994.