GLOBAL DYNAMICS IN PRINCIPAL SINGULAR SUBSPACE NETWORKS

Ibrahim M. Elfadel

Masimo Corporation
Laguna Hills, CA 92653

ABSTRACT

A left (resp. right) principal singular subspace of dimension p is the subspace spanned by the p left (resp. right) singular vectors corresponding to the p largest singular values of the cross-correlation matrix of two stochastic processes. In this paper, we study the global dynamics of a system of nonlinear ordinary differential equations (ODEs) that govern the unsupervised Hebbian learning of left and right principal singular subspaces from samples of the two stochastic processes. In particular, we show that these equations admit a simple Lyapunov function when they are restricted to a well defined smooth, compact manifold, and that they are related to a matrix Riccati differential equation. Moreover, we show that in the case $p = 1$, the solutions of these ODEs can be given in closed form.

1. INTRODUCTION

Let $\mathbf{x} \in \Re^n$ and $\mathbf{y} \in \Re^m$ be two zero-mean random vectors with a cross-correlation matrix $\mathbf{C} = E[\mathbf{y}\mathbf{x}^T] \in \Re^{m\times n}$. We will assume that $m \geq n$. The i-th left and right singular vectors, $\mathbf{u}_i$ and $\mathbf{v}_i$, of $\mathbf{C}$ are defined as follows [Golub and Van Loan, 1983, p. 17]

$$\mathbf{C}\mathbf{v}_i = s_i\mathbf{u}_i \quad \text{and} \quad \mathbf{C}^T\mathbf{u}_i = s_i\mathbf{v}_i$$

where s_i is the i-th singular value of the matrix $\mathbf{C}$. We will assume that these singular values are indexed in the descending order $s_1 \geq s_2 \geq \ldots \geq s_r \geq s_r = 0 \ldots = s_m$, where r is the rank of the matrix $\mathbf{C}$. An easy computation shows that the left (resp. right) singular vectors of $\mathbf{C}$ are nothing but the eigenvectors of the symmetric, positive, semidefinite matrix $\mathbf{C}^T\mathbf{C} \in \Re^{n\times n}$ (resp. $\mathbf{C}\mathbf{C}^T \in \Re^{m\times m}$). A left (resp. right) singular subspace of dimension p is a subspace of $\Re^n$ (resp. $\Re^m$) spanned by p left (resp. right) singular vectors. The *principal* singular subspaces are those spanned by the p singular vectors corresponding to the p largest singular values. Based on the left and right singular vectors, we can write the singular value decomposition (SVD) of the rectangular matrix $\mathbf{C}$ as

$$\mathbf{C} = \sum_{i=1}^{r} s_i\mathbf{u}_i\mathbf{v}_i^T. \tag{1}$$

This decomposition is such that the matrices $\mathbf{V}_r = [\mathbf{v}_1, \ldots, \mathbf{v}_r]$ and $\mathbf{U}_r = [\mathbf{u}_1, \ldots, \mathbf{u}_r]$ have orthonormal columns.

The principal singular vectors of the cross-correlation matrix encode the directions, in both the space of the $\mathbf{x}$ signal and the space of the $\mathbf{y}$ signal, that support the major "common" features of both signals. Knowing the matrix $\mathbf{C}$, one can use numerical algorithms of SVD type (see, for instance, Golub and Van Loan, [1], Chapter 2) to compute the left and right singular vectors.

In many signal processing and automatic control applications, e.g., adaptive filtering and adaptive control, we have no direct access to the matrix $\mathbf{C}$, either because this matrix is too large to fit in a real-time application, or because, as in a nonstationary environment, it is time-dependent, which makes its real-time computation burdensome. In some adaptive control applications [2], the matrix $\mathbf{C}$ represents the unknown plant "transfer function" from inputs to outputs, and the problem, couched in a control language, becomes how to learn and control the major "modes" of the plant without having to identify the plant's full transfer matrix. Hence, the problem of computing the singular vectors using directly data samples is both of theoretical interest and of practical importance.

2. MAJOR SINGULAR VECTORS

2.1. Single-Neuron Equations

The major singular vectors are the left and right singular vectors corresponding to the largest singular value. Denote by $\mathbf{l}(k)$ and $\mathbf{r}(k)$ the respective estimates at time k of the left and right major singular vectors. The neuron of the major singular vectors has two components, left and right, that are coupled to each other. The vectors $\mathbf{l}(k)$ and $\mathbf{r}(k)$ represent the weights of left

0-7803-2431-5/95 $4.00

and right connection layers. The output of the right component is $\rho(k) \triangleq \mathbf{r}(k)^T\mathbf{x}(k)$, and that of the left component is $\lambda(k) \triangleq \mathbf{l}(k)^T\mathbf{y}(k)$. The update equations for the right and left weight vectors are given by

$$\begin{aligned} \mathbf{l}(k+1) &= \mathbf{l}(k) + \eta(\mathbf{y}(k) - \mathbf{l}(k)\lambda(k))\rho(k) \quad (2) \\ \mathbf{r}(k+1) &= \mathbf{r}(k) + \eta(\mathbf{x}(k) - \mathbf{r}(k)\rho(k))\lambda(k) \quad (3) \end{aligned}$$

where, as in Oja's Hebbian rule for principal component analysis (PCA) [4], we have subtracted the nonlinear terms $\mathbf{l}(k)\lambda(k)\rho(k)$ and $\mathbf{r}(k)\rho(k)\lambda(k)$ to stabilize the growth of the estimates $\mathbf{l}(k)$ and $\mathbf{r}(k)$, respectively. In the above update rules, η is a positive parameter controlling the step size. The linear terms in the update equations correspond to a form of "mutual" Hebbian learning between the left and right estimates that was called, in a recent paper [3], the "cross-coupled Hebbian rule." Taking the conditional means across equations (3) and (2) given $\mathbf{l}(k)$ and $\mathbf{r}(k)$ and letting the time step $\Delta t = \eta \rightarrow 0$ give the following system of nonlinear ordinary differential equations:

$$\begin{aligned} \dot{\mathbf{l}} &= \mathbf{C}\mathbf{r} - \mathbf{l}\mathbf{l}^T\mathbf{C}\mathbf{r} \quad (4) \\ \dot{\mathbf{r}} &= \mathbf{C}^T\mathbf{l} - \mathbf{r}\mathbf{r}^T\mathbf{C}^T\mathbf{l} \quad (5) \end{aligned}$$

Note that if $\mathbf{x}(k) = \mathbf{y}(k)$, the matrix $\mathbf{C}$ will be symmetric, positive, semidefinite, and each one of the ODEs above generates a trajectory identical to the one generated by the well-known, single-neuron Oja equation if the initial conditions are such that $\mathbf{l}_0 = \mathbf{r}_0$.

2.2. Closed-Form Solutions

Oja's single-neuron equation converges exponentially to the major principal subspace from *any* initial condition. This fact has been known for a long time via simulations and has been proved rigorously only very recently [5, 6]. Diamantaras and Kung [3] stated a similar observation for equations (4) and (5), also based on simulations. This exponential behavior should not be surprising, since these equations look very similar to Oja's, and the method used in [5] to find a closed-form solution of Oja' equation can be adapted to this context as well. Let $\mathbf{l}_0$ and $\mathbf{r}_0$ be two arbitrary initial conditions, and define the square, symmetric, positive, semidefinite matrices $\mathbf{A} \triangleq \mathbf{C}\mathbf{C}^T, \mathbf{B} \triangleq \mathbf{C}^T\mathbf{C}$. Then we have, for (4) and (5), the following trajectories

$$\begin{aligned} \mathbf{l}(t) &= \left(f(t^2\mathbf{A})\mathbf{l}_0 + t\mathbf{C}g(t^2\mathbf{B})\mathbf{r}_0\right)/n(t) \quad (6) \\ \mathbf{r}(t) &= \left(t\mathbf{C}^T g(t^2\mathbf{A})\mathbf{l}_0 + f(t^2\mathbf{B})\mathbf{r}_0\right)/n(t) \quad (7) \end{aligned}$$

where $n(t)$ is a positive, continuous function of time that can be explicitly given in terms of $\mathbf{C}$ and the initial conditions, while the functions f and g are defined by the absolutely convergent series

$$f(x) = \sum_{i=0}^{\infty} \frac{x^i}{(2i)!} \text{ and } g(x) = \sum_{i=0}^{\infty} \frac{x^i}{(2i+1)!}.$$

The derivation of Equations (6) and (7) is lengthy and is omitted here because of limited space. Just to partially check the correctness of the above result, we note that when $\mathbf{C}$ is square and symmetric (Oja's single-neuron equation), and the initial conditions are such that $\mathbf{l}_0 = \mathbf{r}_0 \triangleq \mathbf{w}_0$, we have

$$\begin{aligned} f(t^2\mathbf{A})\mathbf{l}_0 + t\mathbf{C}g(t^2\mathbf{B})\mathbf{r}_0 &= \\ t\mathbf{C}^T g(t^2\mathbf{A})\mathbf{l}_0 + f(t^2\mathbf{B})\mathbf{r}_0 &= e^{\mathbf{C}t}\mathbf{w}_0, \end{aligned}$$

which is the numerator of Oja's single-neuron solution as found in [5, 6]. It can be easily seen that the numerators in equations (6) and (7) correspond to the solutions of the "unstable" cross-coupled Hebbian equations

$$\dot{\mathbf{l}} = \mathbf{C}\mathbf{r} \quad \text{and} \quad \dot{\mathbf{r}} = \mathbf{C}^T\mathbf{l},$$

while the normalizing function $n(t)$ corresponds to the stabilizing nonlinearities. When the matrix $\mathbf{C}$ has full rank, the matrix $\mathbf{B}$ is invertible, and one can show that the numerator of (7) takes the more familiar form

$$cosh(t\mathbf{B})\mathbf{r}_0 + sinh(t\mathbf{B})\mathbf{B}^{-1}\mathbf{C}^T\mathbf{l}_0$$

which is consistent with the experimentally observed [3] exponential behavior of (4) and (5).

3. SINGULAR SUBSPACE DYNAMICS

3.1. Multineuron Equations

For the case of a system of p interconnected neurons processing two sequences of random vectors $\mathbf{x}(k) \in \Re^n$, and $\mathbf{y}(k) \in \Re^m$ with $p \le n \le m$, the left (resp. right) connection weights are represented by the matrix $\mathbf{L} \in \Re^{m\times p}$ (resp. $\mathbf{R} \in \Re^{n\times p}$), where the j-th column of $\mathbf{L}$ (resp. $\mathbf{R}$) represents the left (resp. right) weight vector for the j-th neuron. The vectors of left (resp. right) outputs at time k are given by $\Lambda(k) \in \Re^p$ (resp. $\mathcal{R}(k) \in \Re^p$), where

$$\Lambda(k) = \mathbf{L}^T(k)\mathbf{y}(k) \qquad \mathcal{R}(k) = \mathbf{R}^T(k)\mathbf{x}(k)$$

The left and right weight matrices are updated according to

$$\begin{aligned} \mathbf{L}(k+1) &= \mathbf{L}(k) + \Delta\mathbf{L}(k) \\ \mathbf{R}(k+1) &= \mathbf{R}(k) + \Delta\mathbf{R}(k) \end{aligned}$$

where

$$\Delta\mathbf{L}(k) = \eta(\mathbf{y}(k)\mathcal{R}(k)^T - \mathbf{L}(k)\Lambda(k)\mathcal{R}(k)^T)\mathbf{y}(k)^T)$$
$$\Delta\mathbf{R}(k) = \eta(\mathbf{x}(k)\Lambda(k)^T - \mathbf{R}(k)\mathcal{R}(k)\Lambda(k)^T)\mathbf{x}(k)^T)$$

If $\mathbf{x}$ and $\mathbf{y}$ have zero mean, the conditional means of $\Delta\mathbf{L}(k)$ and $\Delta\mathbf{R}(k)$ given $\mathbf{L}(k)$ and $\mathbf{R}(k)$ evolve in the continuous-time limit according to the matrix equations

$$\dot{\mathbf{L}} = \mathbf{CR} - \mathbf{LL}^T\mathbf{CR} \quad (8)$$
$$\dot{\mathbf{R}} = \mathbf{C}^T\mathbf{L} - \mathbf{RR}^T\mathbf{C}^T\mathbf{L} \quad (9)$$

Equations (8) and (9), which extend Oja's matrix equation [4] to the case where the matrix $\mathbf{C}$ is arbitrary, reduce to Equations (4) and (5) in the special case $p = 1$.

3.2. Invariant Sets

In this subsection, we study some of the invariant sets under the flow of Equations (8) and (9). For any pair of integers $(q, r), q \geq r$, define the set

$$S_{q,r} \triangleq \{\mathbf{X} \in \Re^{q\times r} | \mathbf{X}^T\mathbf{X} = \mathbf{I}_r\}, \quad (10)$$

where $\mathbf{I}_r$ is the identity matrix of order r. The above set is a smooth, compact manifold of dimension $qr - \frac{1}{2}q(q+1)$ known as the Stiefel manifold [6]. The product set $S_{m,p} \times S_{n,p}$ is also a smooth, compact manifold for which we have the following

Proposition 1 *The product manifold $S_{m,p} \times S_{n,p}$ is invariant under the flow of equations (8) and (9).*

Proof: Easy algebra shows that

$$\frac{d}{dt}(\mathbf{L}^T\mathbf{L}) = \mathbf{R}^T\mathbf{C}^T\mathbf{L}(\mathbf{I}_m - \mathbf{L}^T\mathbf{L}) + (\mathbf{I}_m - \mathbf{L}^T\mathbf{L})\mathbf{L}^T\mathbf{CR}$$

$$\frac{d}{dt}(\mathbf{R}^T\mathbf{R}) = \mathbf{L}^T\mathbf{CR}(\mathbf{I}_n - \mathbf{R}^T\mathbf{R}) + (\mathbf{I}_n - \mathbf{R}^T\mathbf{R})\mathbf{R}^T\mathbf{C}^T\mathbf{L}$$

The above equations show that if the initial condition $(\mathbf{L}_0, \mathbf{R}_0) \in S_{m,p} \times S_{n,p}$ then the trajectory $(\mathbf{L}(t), \mathbf{R}(t)) \in S_{m,p} \times S_{n,p}$ for all $t > 0$. ■

The invariance of $S_{m,p} \times S_{n,p}$ justifies the restriction of singular subspace dynamics to this set. This can be easily done by picking up the initial condition so that the columns of the initial left and right matrix are orthonormal.

Another invariant set that will be essential for the derivation of the Riccati flow corresponding to Equations (8) and (9) is the following

$$\Sigma_p(\mathbf{C}) \triangleq \{(\mathbf{X}, \mathbf{Y}) \in \Re^{m,p} \times \Re^{n,p} | \mathbf{X}^T\mathbf{CY} = \mathbf{Y}^T\mathbf{C}^T\mathbf{X}\}.$$

To prove that $\Sigma_p(\mathbf{C})$ is indeed invariant, denote by $\mathbf{Z} \triangleq \mathbf{L}^T\mathbf{CR}$, and compute

$$\frac{d\mathbf{Z}}{dt} = \mathbf{R}^T\mathbf{C}^T\mathbf{CR} + \mathbf{L}^T\mathbf{CC}^T\mathbf{L} - \mathbf{ZZ}^T - \mathbf{Z}^T\mathbf{Z},$$

Now note that the right-hand side is symmetric, which means that

$$\frac{d}{dt}(\mathbf{L}^T\mathbf{CR}) = \frac{d}{dt}(\mathbf{R}^T\mathbf{C}^T\mathbf{L})$$

Therefore if $(\mathbf{L}_0, \mathbf{R}_0) \in \Sigma_p(\mathbf{C})$ then $(\mathbf{L}(t), \mathbf{R}(t)) \in \Sigma_p(\mathbf{C})$ for all $t > 0$.

3.3. Equilibrium Points

To find the equilibrium points $(\mathbf{L}^*, \mathbf{R}^*)$ of Equations (8) and (9), we define the two symmetric matrices

$$\mathbf{P} \triangleq \mathbf{R}^*\mathbf{R}^{*T} \quad \text{and} \quad \mathbf{Q} \triangleq \mathbf{L}^*\mathbf{L}^{*T}.$$

Then at equilibrium, we have

$$\mathbf{CR}^* - \mathbf{QCR}^* = 0 \quad \text{and} \quad \mathbf{C}^T\mathbf{L}^* - \mathbf{PC}^T\mathbf{L}^* = 0$$

from which we can deduce the equalities:

$$\mathbf{CP} = \mathbf{QC}, \;\; \mathbf{C}^T\mathbf{CP} = \mathbf{PC}^T\mathbf{C}, \;\; \mathbf{CC}^T\mathbf{Q} = \mathbf{QCC}^T.$$

Since we are interested in equilibrium points for which the matrices $\mathbf{R}^*$ and $\mathbf{L}^*$ have orthonormal columns, we consider only the flow on the product manifold $S_{m,p} \times S_{n,p}$.

Proposition 2 *The equilibrium points of (8) and (9) on $S_{m,p} \times S_{n,p}$ correspond to matrices $\mathbf{P}$ and $\mathbf{Q}$ such that $\mathbf{P}$ is an orthogonal projection on a p-dimensional right singular subspace of $\mathbf{C}$ and $\mathbf{Q}$ is the orthogonal projection on the corresponding p-dimensional left singular subspace. Moreover, the only stable equilibrium points are those for which $\mathbf{P}$ and $\mathbf{Q}$ are orthogonal projections on the right and left* principal *singular subspaces of dimension p.*

The first part of this proposition can be proved using the above equalities satisfied by $\mathbf{P}$ and $\mathbf{Q}$ and Lemma A1 in [4]. The stability part can be proved using a perturbation argument around equilibrium matrices $(\mathbf{L}^*, \mathbf{R}^*)$ whose columns do *not* span left and right *principal* singular subspaces. An alternative is to use the Lyapunov-like function defined in the next section. As a corollary of the above proposition, one can state that the number of equilibrium points of (8) and (9) is at least $\frac{n!}{p!(n-p)!}$.

3.4. Lyapunov-like Function

In [3], it was shown that the function

$$V(\mathbf{l},\mathbf{r}) = s_1 - \frac{\mathbf{l}^T\mathbf{C}\mathbf{r}}{||\mathbf{l}||||\mathbf{r}||}$$

is a Lyapunov function for the coupled equations (4) and (5), from which it was concluded that the major left and right singular subspace are globally asymptotically stable if $s_1 > s_2 \geq \ldots \geq s_r$. Global asymptotic stability is much harder to show in the multineuron case. However, if the flow of the multineuron singular subspace dynamics is restricted to the invariant product set $S_{m,p} \times S_{n,p}$, then we can show the following

Proposition 3 *The scalar function*

$$\mathcal{V}(\mathbf{L},\mathbf{R}) = -tr(\mathbf{L}^T\mathbf{C}\mathbf{R})$$

is strictly decreasing along the trajectories of (8) and (9) on the invariant set $S_{m,p} \times S_{n,p}$.

Proof: Along the trajectories of (8) and (9), we have

$$\begin{aligned}\frac{d}{dt}\mathcal{V}(\mathbf{L},\mathbf{R}) &= -tr(\dot{\mathbf{L}}^T\mathbf{C}\mathbf{R}) - tr(\mathbf{L}^T\mathbf{C}\dot{\mathbf{R}}) \\ &= -tr(\mathbf{R}^T\mathbf{C}^T(\mathbf{I}_m - \mathbf{L}\mathbf{L}^T)\mathbf{C}\mathbf{R}) \\ &- tr(\mathbf{L}^T\mathbf{C}(\mathbf{I}_n - \mathbf{R}\mathbf{R}^T)\mathbf{C}^T\mathbf{L}).\end{aligned}$$

Because the flow is restricted to $S_{m,p} \times S_{n,p}$, both $\Pi_m \triangleq \mathbf{I}_m - \mathbf{L}\mathbf{L}^T$ and $\Pi_n \triangleq \mathbf{I}_n - \mathbf{R}\mathbf{R}^T$ are orthogonal projections, and therefore

$$\begin{aligned}tr(\mathbf{R}^T\mathbf{C}^T\Pi_m\mathbf{C}\mathbf{R}) &= ||\Pi_m\mathbf{C}\mathbf{R}||_F^2 \geq 0 \\ tr(\mathbf{L}^T\mathbf{C}\Pi_n\mathbf{C}^T\mathbf{L}) &= ||\Pi_n\mathbf{C}^T\mathbf{L}||_F^2 \geq 0\end{aligned}$$

where the subscript F denotes the matrix Frobenius norm. Note now that $\frac{d}{dt}\mathcal{V}(\mathbf{L}',\mathbf{R}') = 0$ if and only if $(\mathbf{L}',\mathbf{R}')$ is an equilibrium point of (8) and (9). The result follows immediately from the above trace inequalities. ∎

An important consequence of the above proposition is that the flow on $S_{m,p} \times S_{n,p}$ cannot exhibit sustained oscillations.

3.5. Riccati Equation

Finally, define the two $(m+n) \times (m+n)$ symmetric matrices

$$\mathbf{K} \triangleq \begin{bmatrix} \mathbf{O} & \mathbf{C} \\ \mathbf{C}^T & \mathbf{O} \end{bmatrix} \text{ and } \mathbf{M} \triangleq \begin{bmatrix} \mathbf{L}\mathbf{L}^T & \mathbf{R}\mathbf{L}^T \\ \mathbf{L}\mathbf{R}^T & \mathbf{R}\mathbf{R}^T \end{bmatrix},$$

and assume that the initial condition $(\mathbf{L}_0,\mathbf{R}_0) \in \Sigma_p(\mathbf{C})$. Then $\mathbf{M}$ satisfies the following matrix differential Riccati equation (DRE)

$$\dot{\mathbf{M}} = \mathbf{M}\mathbf{K} + \mathbf{K}\mathbf{M} - \mathbf{M}\mathbf{K}\mathbf{M}.$$

As shown in [6], the matrix DRE corresponding to the multineuron Oja equation plays an essential role in the derivation of closed-form solutions for Oja's principal subspace networks. The DRE derived above for singular subspace networks should play a similar role, at least in the case when the flow of (8) and (9) is restricted to the invariant set $\Sigma_p(\mathbf{C})$.

4. CONCLUSION

In this paper, we have investigated the dynamics of left and right singular subspace unsupervised learning as defined by a straightforward extension of Oja's multineuron equation for principal subspace learning. Among our results are a characterization of the closed-form solution for the major left and right singular subspace and a characterization of the equilibrium points on the invariant manifold corresponding to left and right matrices with orthonormal columns. The flow on this manifold admits a Lyapunov-like function that is strictly decreasing along any trajectory that is not an equilibrium point. Finally, we have shown that the singular subspace dynamics is, like the eigen subspace dynamics, related to a matrix differential Riccati equation. There is a wealth of information [6] about the Riccati equation that can be readily used to derive other fundamental results on the dynamics of singular subspace networks.

5. REFERENCES

[1] Golub, G.H., and C.F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, 1983.

[2] Sanger, T., "Theoretical Elements of Hierarchical Control in Vertebrate Motor Systems," Ph. D. Thesis, MIT, 1993

[3] Diamantaras, K.I., and S. Kung, "Cross-Correlation Neural Networks Models," *IEEE Trans. on Signal Processing,* Vol. 42, pp. 3218 - 3223, 1994.

[4] E. Oja "Principal Components, Minor Components, and Linear Neural Networks," *Neural Networks*, Vol. 5, pp. 927–935, 1992.

[5] Wyatt, J.L., and I. M Elfadel, "Time-Domain Solutions of Oja's Equations," to appear in *Neural Computations.* 1994.

[6] Yan, W., U. Helmke, and J.B. Moore,"Global Analysis of Oja's Flow for Neural Networks," *IEEE Trans. on Neural Networks,* Vol. 5, pp. 674 - 683, 1994.