State space models on special manifolds

Yasuko Chikuse

Faculty of Engineering, Kagawa University, 2217-20 Hayashi-cho, Takamatsu-shi, Kagawa-ken, 761-0396 Japan

Received 22 June 2004
Available online 27 April 2006

Abstract

This paper concerns modeling time series observations in state space forms considered on the Stiefel and Grassmann manifolds. We develop a state space model relating the time series observations to a sequence of unobserved state or parameter matrices assuming the matrix Langevin noise processes on the Stiefel manifolds. We show a Bayes method for estimating the state matrices by the posterior modes. We consider a further extended state space model where two sequences of unobserved state matrices are involved. A simple state space model on the Grassmann manifolds with matrix Langevin noise processes is also investigated.

© 2006 Elsevier Inc. All rights reserved.

AMS 1991 subject classification: 62M10; 62H11; 60G20; 62H12

Keywords: State space model; Stiefel manifold; Grassmann manifold; Matrix Langevin distributions; Posterior mode estimation; Iterative method

1. Introduction

There exists a large literature of time series analysis discussed on the Euclidean spaces, including the multivariate (multiple) time series analysis when the time-dependent observations are vector-variate. See e.g., books by Hannan [14], Anderson [1], Koopman [20], and Brockwell and Davis [3].

Breuklig [2] studied directional models in time series analysis on the unit hypersphere. Chikuse [6] developed autoregressive stochastic models based on the distributions defined on the Stiefel manifold. In this paper, we concern modeling time series observations in state space forms considered on the Stiefel and Grassmann manifolds.

Discrete-time state space models relate time series (vector-variate) observations \( Y_1, Y_2, \ldots \), to a sequence \( X_0, X_1, \ldots \), of unobserved (vector-variate) states or parameters. The problem
involves estimating the states $X_0, X_1, \ldots$, given the observations $Y_1, Y_2, \ldots$. The normal state space models assuming the normal noise processes were much discussed in the literature and the solutions are well-known as the Kalman filter. See e.g., Durbin and Koopman [10] for a discussion of the state space models on the usual Euclidean spaces; see also Meinhold and Singpurwalla [22]. There exist related works in the state space models. For example, Bucy [4] developed a theory of non-linear filtering, Kitagawa [19] gave a non-Gaussian state space modeling of nonstationary time series, and Fahrmeir [11] discussed an extension of the Kalman filter by estimating posterior modes of states. Furthermore, Naik-Nimbalkar and Rajarshi [24] suggested an approach to the problem based on the theory of estimating functions.

We consider the state space models assuming the matrix Langevin noise processes on the Stiefel and Grassmann manifolds. The estimation of states via posterior modes is suggested.

The Stiefel manifold $V_{k,m}$ is the space a point of which is a set of $k$ orthonormal vectors in $R^m (k \leq m)$, so that $V_{k,m} = \{X (m \times k); X'X = I_k\}$, where $I_k$ is the $k \times k$ identity matrix. For $m = k$, $V_{k,m}$ is the orthogonal group $O(m)$ of $m \times m$ orthonormal matrices. A random matrix $X$ on $V_{k,m}$ is said to have the matrix Langevin (or von Mises–Fisher) distribution, denoted by $L(m, k; F)$, if its density function is given by [9]

$$\text{etr}(F'X)/\Gamma_1 F_1(1/2m; 1/4F'F) \text{ with } F \text{ an } m \times k \text{ matrix},$$

(1.1)

where $\text{etr}(A) = \exp(\text{tr} A)$, and the $_pF_q$ is a hypergeometric function with matrix argument due to e.g., Herz [15], James [17], Constantine [8] and Muirhead [23]. Here, assuming the rank of $F$ being $k$ for the simplicity of argument, we write the singular value decomposition of $F$ as

$$F = \Gamma \Lambda \Theta' \quad \text{with } \Gamma \in V_{k,m}, \Theta \in O(k) \quad \text{and}$$

$$\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_k), \quad \lambda_1 \geq \cdots \geq \lambda_k > 0$$

(1.2)

which is also expressed as $\Gamma \Theta' \cdot \Theta \Lambda \Theta' = M \cdot C$, say. The distribution has the unique modal orientation $M = \Gamma \Theta' \in V_{k,m}$ and the $\lambda_i$’s (the latent roots of the $k \times k$ positive definite matrix $C$) control the concentrations about the mode in the directions determined by the orientations $\Gamma$ and $\Theta$. These parameters, mode and concentrations, of the matrix Langevin distribution may be considered as the counterparts of the parameters, mean and variance–covariances, of the (multivariate) normal distribution.

The Grassmann manifold $G_{k,m-k}$ is the space whose points are $k$-planes $v$, that is, $k$-dimensional hyperplanes in $R^m$ containing the origin. To each $k$-plane $v$ in $G_{k,m-k}$, corresponds a unique $m \times m$ orthogonal projection matrix $P$ idempotent of rank $k$ onto $v$. If the $k$ columns of an $m \times k$ matrix $Y$ in $V_{k,m}$ span $v$, we have $YY' = P$. Let $P_{k,m-k}$ denote the set of all $m \times m$ orthogonal projection matrices idempotent of rank $k$. We shall conduct our statistical analysis on the manifold $P_{k,m-k}$ which is equivalent to the Grassmann manifold $G_{k,m-k}$. For a random matrix $P$ on $P_{k,m-k}$, the distribution having the density function

$$\text{etr}(BP)/\Gamma_1 F_1(1/2k; 1/2m; B) \text{ with } B \text{ an } m \times m \text{ symmetric matrix}$$

(1.3)

is a slight modification of the Downs’ [9] distribution (1.1) on the Stiefel manifold, and may be called the matrix Langevin distribution on $P_{k,m-k}$, which is denoted by $L^{(P)}(m, k; B)$. Here, assuming the rank of $B$ being $k$ for the sake of simplicity, we write the spectral decomposition of $B$ as

$$B = \Gamma \Lambda \Gamma' \quad \text{with } \Gamma \in V_{k,m} \text{ and} \quad \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_k), \quad \lambda_1 \geq \cdots \geq \lambda_k$$

(1.4)
and we impose identifiability restrictions on \( B \),

\[
\text{tr } B = b \quad \text{being fixed} \quad \text{or} \quad \text{rank } B = k < m. \tag{1.5}
\]

The distribution has the mode \( \Gamma \Gamma' \), and when \( B \) is positive semi-definite, the \( \lambda_i \)'s may control the concentrations of the distribution about the mode. Chikuse and Watson [7] defined and discussed the distribution (1.3) as a special case of a more general family of distributions expressed in terms of zonal polynomials or hypergeometric functions with matrix argument.

We note that the distributions on \( V_{k,m} \) and \( P_{k,m-k} \) are expressed with respect to the normalized invariant measures (uniform distributions) on the respective manifolds; see James [16], Farrell [12], and Chikuse and Watson [7] for a discussion of these invariant measures. The distribution (1.1) for \( k = 1 \) is called the von Mises (when \( m = 2 \)) or the Fisher (when \( m = 3 \)) distribution. The density function (1.1) is obtained from the joint density of the elements of an \( m \times k \) random matrix \( X \), where the \( k \times 1 \) row vectors of \( X \) are independently distributed as \( k \)-variate normal with mean matrix \( E(X) = M \) and \( k \times k \) covariance matrix \( \Sigma \), and we put \( F = M \Sigma^{-1} \), imposing the condition \( X'X = I_k \). The density function (1.3) is obtained from the joint density of the elements of an \( m \times m \) matrix \( P = XX' \) with an \( m \times k \) random matrix \( X \), where the \( m \times 1 \) column vectors of \( X \) are independently distributed as \( m \)-variate normal with zero means and \( m \times m \) covariance matrix \( \Sigma \), and we put \( B = \frac{1}{2} \Sigma^{-1} \), imposing the condition \( X'X = I_k \). The distributions (1.1) and (1.3) are exponential-type distributions and are useful for statistical analyses on the Stiefel and Grassmann manifolds, respectively.

For the special case \( k = 1 \), the observations from the unit hypersphere \( V_{1,m} \) are directed unit vectors, i.e., directions, and those from the real projective space \( G_{1,m-1} \) are axes or undirected lines through the origin, i.e., one-dimensional subspaces. There exists a large literature of applications of these directional statistics and its statistical analysis. Most directional statistics in practice occur in two or three dimensions, i.e., on the circle \( (m = 2) \) and the sphere \( (m = 3) \). Directional analysis of data played important roles in the Earth Sciences, Astrophysics, Biology, Meteorology, Animal Behavior and many other fields; see e.g., Watson [25], Fisher et al. [13], and Mardia and Jupp [21].

The analysis of data on the general Stiefel manifold \( V_{k,m} \) is required in particular for \( k \leq m \leq 3 \) in practical applications in the Medical Sciences, Astronomy and other fields. See Downs [9], Jupp and Mardia [18], Watson [26], and Fisher et al. [13] for the analyses of the data of vectorcardiogram orientations and of measurements of orbits of comets, which are described with \( k = 2 \) and \( m = 3 \). One is naturally interested in \( k \)-dimensional subspaces as observations from the general Grassmann manifold \( G_{k,m-k} \). We note that if \( X \) is an observation on the Stiefel manifold \( V_{k,m} \), \( XX' \) is an observation on the manifold \( P_{k,m-k} \) equivalent to the Grassmann manifold \( G_{k,m-k} \). Examples of observations on \( G_{k,m-k} \) arise in the signal processing of radar with \( m \) elements observing \( k \) targets. The Grassmann manifold is a rather new subject treated as a statistical sample space. See Chikuse [5] for statistical analyses on the Stiefel and Grassmann manifolds.

In Section 2, we develop a state space model relating the time series observations \( \{Y_1, Y_2, \ldots, Y_t\} \) on the Stiefel manifold \( V_{k,m} \) to a sequence of unobserved state modal orientation matrices \( \{X_0, X_1, \ldots, X_t\} \) on \( V_{k,m} \) of noise processes distributed as matrix Langevin. We show a Bayes method for estimating the states \( \{X_1, X_2, \ldots, X_t\} \) by the posterior modes assuming \( X_0 \) given. An iterative procedure for the estimation is suggested. Further, we consider an extended state space model on Stiefel manifolds, where two sequences of unobserved state modal orientation matrices \( \{X_0, X_1, \ldots, X_t\} \) on \( V_{k,m} \) and unobserved state regression matrices \( \{Y_0, Y_1, \ldots, Y_t\} \) on \( O(m) \), in consideration of orientational regressions, are involved.
Section 3 investigates a simple state space model on the manifold $P_{k,m-k}$, where the time series observations $\{Q_1, Q_2, \ldots, Q_t\}$ on $P_{k,m-k}$ are related to a sequence of unobserved state matrices $\{P_0, P_1, \ldots, P_t\}$ on $P_{k,m-k}$ of noise processes distributed as matrix Langevin.

2. State space models on $V_{k,m}$

2.1. A simple state space model

We may be interested in the time series represented in a state space model considered on Stiefel manifolds. The model relates the matrix-valued time series observations $\{Y_1, Y_2, \ldots, Y_t\}$ on the Stiefel manifold $V_{k,m}$ to a sequence of $m \times k$ unobservable state matrices $\{X_0, X_1, \ldots, X_t\}$ on $V_{k,m}$ as follows. The observations $Y_s$ given $X_s$ are independently distributed as matrix Langevin $L(m, k; X_s \Phi_s)$ with the modal orientation $X_s$. The unobserved states $X_s$ given $X_{s-1}$ are independently distributed as matrix Langevin $L(m, k; X_{s-1} \Phi_s)$ (of Markovian type) with the modal orientation $X_{s-1}$. Here, we assume that the concentration matrices $\Phi_s > 0$ and $\Phi_s > 0$ are known $k \times k$ positive definite. These conditional densities $p(\cdot | \cdot)$ are

\[
p(Y_s | X_s) = \text{etr}(Y_s' X_s \Omega_s)/\sqrt{\frac{1}{2} \Omega_s^2} F_1(\frac{1}{2} m; \frac{1}{4} \Omega_s^2)
\]

and

\[
p(X_s | X_{s-1}) = \text{etr}(X_s' X_{s-1} \Phi_s)/\sqrt{\frac{1}{2} \Phi_s^2} F_1(\frac{1}{2} m; \frac{1}{4} \Phi_s^2) \quad \text{for } s = 1, 2, \ldots, t.
\]

(2.1)

We assume that $X_0$ is given.

We show a Bayes method for estimating the states $\{X_1, X_2, \ldots, X_t\}$ by the posterior modes. We put

\[
Y_s = \{Y_0, Y_1, \ldots, Y_s\} \quad \text{with } Y_0 = \emptyset, \text{ the null set and}
\]

\[
X_s = \{X_0, X_1, \ldots, X_s\}.
\]

(2.2)

Given $Y_s$, we estimate $X_s$ based on the posterior density $p(X_s | Y_s)$.

We have

\[
p(X_s | Y_s) = p(X_s, Y_s)/p(Y_s) \propto p(X_s, Y_s),
\]

where the symbol $\propto$ denotes the equality to the terms involving only the states $X_s$. We can write

\[
p(X_s, Y_s) = p(Y_s | Y_{s-1}, X_s) \cdot p(X_s, Y_{s-1})
\]

\[
= p(Y_s | Y_{s-1}, X_s) \cdot p(X_s | X_{s-1}, Y_{s-1}) p(X_{s-1}, Y_{s-1})
\]

\[
= \prod_{s=1}^t p(Y_s | Y_{s-1}, X_s) p(X_s | X_{s-1}, Y_{s-1}).
\]

(2.3)

From our model (2.1), we obtain

\[
p(X_s, Y_s) \propto \prod_{s=1}^t \{\text{etr}(Y_s' X_s \Omega_s)\} \text{etr}(X_s' X_{s-1} \Phi_s)
\]

\[
\propto \prod_{s=1}^t \{\text{etr}(Y_s' X_s \Omega_s)\} \text{etr}(X_s' X_{s-1} \Phi_s)
\]

\[
\propto \prod_{s=1}^t \{\text{etr}(Y_s' X_s \Omega_s)\} \text{etr}(X_s' X_{s-1} \Phi_s)
\]
and hence
\[ L(X_t|Y_t) = \log p(X_t|Y_t) \propto \sum_{s=1}^{t} \text{tr}(Y_s' X_s \Omega_s + X_s' X_{s-1} \Phi_s). \] (2.4)

Now, we suggest an iterative procedure for estimating \( X_s \) by the posterior modes.

(i) At the stage \( s = 1 \), with \( X_0^* = X_0 \) given, we have
\[ L(X_1|Y_1) \propto \text{tr}[X_1' (Y_1 \Omega_1 + X_0^* \Phi_1)] = \text{tr}[X_1' G_1(X_0^*)] \text{ say}. \] (2.5)

The mode of the log-density (2.5) is given by
\[ \hat{X}_1 = H[G_1(X_0^*)], \] (2.6)

where we put the polar decomposition of \( G = G_1(X_0^*) \) as
\[ G = H[G] T^{1/2}[G] \quad \text{with } H[G] = G(G'G)^{-1/2} \in V_{k,m} \quad \text{and} \quad T[G] = G'G. \] (2.7)

(ii) We shall state the procedure for the general stages \( s = 2, 3, \ldots, t-1 \), with \( \hat{X}_{s-2}^* = \{ \hat{X}_1^*, \hat{X}_2^*, \ldots, \hat{X}_{s-2}^* \} \) having been already estimated and
\[ \hat{X}_{s-1} = H[G_{s-1}(\hat{X}_{s-2}^*)] \] (2.8)
given at the end of the stage \( s-1 \), where \( G_{s-1}(\cdot) \) will be defined below. Using the symbol \( \propto \) denoting the equality to the terms involving the states \( X_{s-1} \) and \( X_s \), which are only relevant at the stage \( s \), we have
\[ L(X_s|Y_s) \propto \text{tr}[X_s' G_{s-1}(\hat{X}_{s-2}^*) + X_s' (Y_s \Omega_s + X_{s-1} \Phi_s)] \]
\[ = \text{tr}[X_s' G_s(X_{s-1}) + X_s' G_{s-1}(\hat{X}_{s-2}^*)] \text{ say} \] (2.9)
\[ = \text{tr}[X_s' [G_{s-1}(\hat{X}_{s-2}^*) + X_s \Phi_3] + X_s' Y_s \Omega_s] \]
\[ = \text{tr}[X_s' F_{s-1}(\hat{X}_{s-2}^*, X_s) + X_s' Y_s \Omega_s] \text{ say}. \] (2.10)

With \( \hat{X}_{s-1} \) given by (2.8), we estimate \( X_s \) from (2.9) as
\[ \hat{X}_{s(1)} = H[G_{s}(\hat{X}_{s-1})] \] (2.11)
and then \( X_{s-1} \) from (2.10) as
\[ \hat{X}_{s-1(1)} = H[F_{s-1}(\hat{X}_{s-2}^*, \hat{X}_{s(1)})]. \] (2.12)

Repeating the process yields the estimate \( \hat{X}_{s(2)} \) of \( X_s \) given by (2.11) with \( \hat{X}_{s-1(1)} \) replacing \( \hat{X}_{s-1} \), and then the estimate \( \hat{X}_{s-1(2)} \) of \( X_{s-1} \) given by (2.12) with \( \hat{X}_{s(2)} \) replacing \( \hat{X}_{s(1)} \), and we proceed similarly. For some \( N_s \) large enough to ensure the convergence of the estimates \( \hat{X}_{s-1(j)} \), we set the final estimate of \( X_{s-1} \) as
\[ \hat{X}_{s-1}^* = \hat{X}_{s-1(N_s)} = H[F_{s-1}(\hat{X}_{s-2}^*, \hat{X}_{s(N_s)})] \] (2.13)
and put
\[ \hat{X}_s = H[G_{s}(\hat{X}_{s-1}^*)]. \] (2.14)
(iii) At the final stage $s = t$, with $\hat{X}^*_{t-2}$ having been already estimated and $\hat{X}_{t-1} = H[G_{t-1}(\hat{X}^*_{t-2})]$ given at the end of the stage $t - 1$, we carry out the procedure stated in (ii) with $s = t$. We obtain, for some $N_t$, the final estimate of $X_{t-1}$ as

$$\hat{X}^*_{t-1} = \hat{X}_{t-1(N_t)} = H[F_{t-1}(\hat{X}^*_{t-2}, \hat{X}_{t(N_t)})]$$

and we write the final estimate of $X_t$ as

$$\hat{X}^*_t = H[G_t(\hat{X}^*_{t-1})].$$

Thus, we established the desired estimates $\hat{X}^* = \{\hat{X}^*_1, \hat{X}^*_2, \ldots, \hat{X}^*_t\}$.

**Remark 2.1.** By a similar procedure, we can treat the state space model, where $\{Y_1, Y_2, \ldots, Y_t\}$ are $m \times k$ matrix-valued time series observations, and the conditional distribution of $Y_s$ given $X_s$ is such that $m$ rows of $Y_s$ are independently distributed as $k$-variate normal with covariance matrix $\Sigma_s$ known $k \times k$ positive definite and $E(Y_s|X_s) = X_s\Omega_s$ and the rest of the conditions are the same as for our present model.

**Remark 2.2.** It is seen from (2.1) that the conditional distributions $p(Y_s|X_s)$ and $p(X_s|X_{s-1})$ may be of the more general forms $p(Y_s|Y_{s-1}, X_s)$ and $p(X_s|X_{s-1}, Y_{s-1})$, respectively, so that the matrix coefficients $\Omega_s$ and $\Phi_s$ may depend on the previous observations $Y_{s-1}$.

**Remark 2.3.** We note a relationship of our procedure to a prediction problem. We can write

$$p(X_{t+1}|Y_t) = p(X_{t+1}|X_t, Y_t)p(X_t|Y_t)$$

and

$$p(X_t|Y_t) = p(X_t, Y_t|Y_{t-1})/p(Y_t|Y_{t-1}) \propto p(X_t, Y_t|Y_{t-1}) = p(Y_t|X_t, Y_{t-1})p(X_t|Y_{t-1}),$$

which, from our model (2.1), yields

$$p(X_{t+1}|Y_t) = p(X_{t+1}|X_t)p(X_t|Y_t)$$

(2.17)

and

$$p(X_t|Y_t) \propto p(Y_t|X_t)p(X_t|Y_{t-1}).$$

(2.18)

These indicate that the prediction problem can be treated by using prior information of the future parameter and that a new observation is used adding to the information available at the prediction stage.

### 2.2. An extended state space model

We may be interested in an extended state space model on Stiefel manifolds, where two sequences of unobservable state modal orientation matrices $\{X_0, X_1, \ldots, X_t\}$ on $V_{k,m}$ and unobservable state regression matrices $\{Y_0, Y_1, \ldots, Y_t\}$ on $O(m)$, in consideration of orientational regressions, are involved. The time series observations $\{Z_1, Z_2, \ldots, Z_t\}$ on $V_{k,m}$ are related to the state matrices as follows. The observations $Z_s$ given $X_s$ are independently distributed as matrix Langevin $L(m, k; X_s\Omega_s)$ with the modal orientation $X_s \in V_{k,m}$. The unobserved states
We can write

\[ p(Z_s|X_s) = \text{etr}(Z_s'X_s\Omega_s)/0 F_1(1/2 m; 1/4 \Omega_s^2), \]

\[ p(X_s|X_{s-1}, Y_{s-1}) = \text{etr}(X_s'Y_{s-1}X_{s-1}\Phi_s)/0 F_1(1/2 m; 1/4 \Phi_s^2), \]

and

\[ p(Y_s|Y_{s-1}) = \text{etr}(Y_s'Y_{s-1}\Psi_s)/0 F_1(1/2 m; 1/4 \Psi_s^2) \quad \text{for} \quad s = 1, 2, \ldots, t. \]  

(2.19)

We assume that \((X_0, Y_0)\) are given.

We put

\[ Z_s = \{Z_0, Z_1, \ldots, Z_s\} \quad \text{with} \quad Z_0 = \emptyset, \]

\[ X_s = \{X_0, X_1, \ldots, X_s\} \quad \text{and} \quad Y_s = \{Y_0, Y_1, \ldots, Y_s\}. \]  

(2.20)

Given \(Z_s\), we estimate \(X_s\) and \(Y_s\) by the posterior modes.

We have the posterior density

\[ p(X_s, Y_s|Z_s) \propto p(X_s, Y_s, Z_s). \]

We can write

\[ p(X_s, Y_s, Z_s) = p(Z_s|X_{s-1}, Y_{s-1}) \cdot p(X_{s-1}, Y_{s-1}) \]

\[ = p(Z_s|X_{s-1}, Y_{s-1}) \cdot p(X_{s-1}, Y_{s-1}|X_{s-1}, Y_{s-1}, Z_{s-1}) p(X_{s-1}, Y_{s-1}, Z_{s-1}), \]

where

\[ p(X_{s-1}, Y_{s-1}|X_{s-1}, Y_{s-1}, Z_{s-1}) = p(X_{s-1}, Y_{s-1}|X_{s-1}, Y_{s-1}, Z_{s-1}) p(Y_{s-1}|X_{s-1}, Y_{s-1}, Z_{s-1}). \]

(2.21)

Hence, from our model (2.19), we obtain

\[ p(X_s, Y_s|Z_s) \propto \prod_{s=1}^t p(Z_s|X_s) p(X_s|X_{s-1}, Y_{s-1}) p(Y_s|Y_{s-1}) \]

\[ \propto \prod_{s=1}^t \left[ \text{etr}(Z_s'X_s\Omega_s) \text{etr}(X_s'Y_{s-1}X_{s-1}\Phi_s) \text{etr}(Y_s'Y_{s-1}\Psi_s) \right] \]

and hence

\[ L(X_s, Y_s|Z_s) = \log p(X_s, Y_s|Z_s) \]

\[ \propto \sum_{s=1}^t \left[ \text{tr}(Z_s'X_s\Omega_s + X_s'Y_{s-1}X_{s-1}\Phi_s) + \text{tr}(Y_s'Y_{s-1}\Psi_s) \right]. \]  

(2.22)
We are concerned with the following procedure for estimating $X_s$ and $Y_s$ by the posterior modes, which is similar to the one suggested in Section 2.1 but a little more involved.

(i) At the stage $s = 1$, with $X_0^* = X_0$ and $Y_0^* = Y_0$ given, we have

$$L(X_1, Y_1|Z_1) \propto \text{tr}[X_1'(Z_1\Omega_1 + Y_0^*X_0^*\Phi_1)] + \text{tr}(Y_1'Y_0^*\Psi_1)$$
$$= \text{tr}[X_1'G_{X,1}(X_0^*, Y_0^*)] + \text{tr}(Y_1'G_{Y,1}(Y_0^*)) \quad \text{say},$$

(2.23)

whose modes are given by

$$\hat{X}_1 = H[G_{X,1}(X_0^*, Y_0^*)] \quad \text{and} \quad \hat{Y}_1 = H[G_{Y,1}(Y_0^*)].$$

(2.24)

(ii) We shall state the procedure for the general stages $s = 2, 3, \ldots, t - 1$, with $\hat{X}_{s-2}^*$ and $\hat{Y}_{s-2}^*$ having been already estimated and

$$\hat{X}_{s-1} = H[G_{X,s-1}(\hat{X}_{s-2}^*, \hat{Y}_{s-2}^*)] \quad \text{and} \quad \hat{Y}_{s-1} = H[G_{Y,s-1}(\hat{Y}_{s-2}^*)]$$

(2.25)

given at the end of the stage $s - 1$, where $G_{X,s-1}(\cdot, \cdot)$ and $G_{Y,s-1}(\cdot)$ will be defined below. Using the symbol $\propto$ similar to (2.9) and (2.10), we have

$$L(X_s, Y_s|Z_s) \propto \text{tr}[X'_s\{Z_s\Omega_s + Y_{s-1}X_{s-1}\Phi_s\}] + \text{tr}(Y'_sY_{s-1}\Psi_s)$$
$$= \text{tr}[X'_sG_{X,s}(X_{s-1}, Y_{s-1})] + \text{tr}(Y'_sG_{Y,s}(Y_{s-1}))$$
$$+ \text{tr}[X'_sG_{X,s-1}(\hat{X}_{s-2}^* + Y_{s-1}X_{s-1}\Phi_s)] + \text{tr}(Y'_sG_{Y,s-1}(\hat{Y}_{s-2}^*)) \quad \text{say}$$
$$= \text{tr}[X'_s\{G_{X,s-1}(\hat{X}_{s-2}^*, \hat{Y}_{s-2}^*) + Y_{s-1}X_{s-1}\Phi_s\}] + \text{tr}(X'_sZ_s\Omega_s)$$
$$= \text{tr}[X'_s\{F_{X,s-1}(\hat{X}_{s-2}^*, \hat{Y}_{s-2}^*) + Y_{s-1}X_{s-1}\}] + \text{tr}(X'_sZ_s\Omega_s) \quad \text{say}.$$

(2.26)

With $(\hat{X}_{s-1}, \hat{Y}_{s-1})$ given by (2.25), we estimate $(X_s, Y_s)$ from (2.26) as

$$\hat{X}_{s(1)} = H[G_{X,s}(\hat{X}_{s-1}, \hat{Y}_{s-1})] \quad \text{and} \quad \hat{Y}_{s(1)} = H[G_{Y,s}(\hat{Y}_{s-1})]$$

(2.28)

and, with $(\hat{X}_{s-2}^*, \hat{Y}_{s-2}^*, \hat{Y}_{s-1}, \hat{X}_{s(1)}, \hat{Y}_{s(1)})$ given at this moment, we estimate $(X_{s-1}, Y_{s-1})$ from (2.27) as

$$\hat{X}_{s-1(1)} = H[F_{X,s-1}(\hat{X}_{s-2}^*, \hat{Y}_{s-2}^*, \hat{Y}_{s-1}, \hat{X}_{s(1)})]$$

and

$$\hat{Y}_{s-1(1)} = H[F_{Y,s-1}(\hat{Y}_{s-2}^*, \hat{Y}_{s(1)})].$$

(2.29)

Repeating the process yields the estimates $(\hat{X}_{s(2)}, \hat{Y}_{s(2)})$ of $(X_s, Y_s)$ given by (2.28) with $(\hat{X}_{s-1(1)}, \hat{Y}_{s-1(1)})$ replacing $(\hat{X}_{s-1}, \hat{Y}_{s-1})$, and then the estimates $(\hat{X}_{s-1(2)}, \hat{Y}_{s-1(2)})$ of $(X_{s-1}, Y_{s-1})$ given by (2.29) with $(\hat{Y}_{s-1(1)}, \hat{X}_{s(2)}, \hat{Y}_{s(2)})$ replacing $(\hat{Y}_{s-1}, \hat{X}_{s(1)}, \hat{Y}_{s(1)})$, and we proceed similarly. For some $N$ large enough to ensure the convergence of the estimates $(X_{s-1(j)}, Y_{s-1(j)})$, we set the final estimates of $(X_{s-1}, Y_{s-1})$ as

$$\hat{X}_{s-1} = \hat{X}_{s-1(N)}, \quad \hat{Y}_{s-1} = \hat{Y}_{s-1(N)}$$

(2.30)

and put

$$\hat{X}_s = H[G_{X,s}(\hat{X}_{s-1}^*, \hat{Y}_{s-1}^*)] \quad \text{and} \quad \hat{Y}_s = H[G_{Y,s}(\hat{Y}_{s-1}^*)].$$

(2.31)
We put

\[ \varphi_{s-2}(\hat{X}_{t-2}, \hat{Y}_{t-2}) \] \quad \text{and} \quad \varphi_{s-1}(\hat{X}_{t-1}, \hat{Y}_{t-1}) \]

given at the end of the stage \( t - 1 \), we carry out the procedure stated in (ii) with \( s = t \). We obtain, for some \( N_t \), the final estimates of \((X_{t-1}, Y_{t-1})\) as

\[ \hat{X}_{t-1} = \hat{X}_{t-1}(N_t) = H[F_{X,t-1}(\hat{X}_{t-2}, \hat{Y}_{t-2}, \hat{Y}_{t-1}(N_{t-1}), \hat{X}_{t}(N_t))] \]

and

\[ \hat{Y}_{t-1} = \hat{Y}_{t-1}(N_t) = H[F_{Y,t-1}(\hat{Y}_{t-2}, \hat{Y}_{t}(N_t))] \] \quad (2.32)

and we write the final estimates of \((X_t, Y_t)\) as

\[ \hat{X}_t = H[G_{X,t}(\hat{X}_{t-1}, \hat{Y}_{t-1})] \quad \text{and} \quad \hat{Y}_t = H[G_{Y,t}(\hat{Y}_{t-1})]. \] \quad (2.33)

Thus, we established the estimates

\[ \hat{X}_t = \{\hat{X}_1^*, \hat{X}_2^*, \ldots, \hat{X}_t^*\} \quad \text{and} \quad \hat{Y}_t = \{\hat{Y}_1^*, \hat{Y}_2^*, \ldots, \hat{Y}_t^*\}. \]

Remark 2.4. The same statement concerning the conditional distribution of \( Z_s \) given \( X_s \) as in Remark 2.1 can be expanded for our model treated in Section 2.2.

Remark 2.5. The other types of state space model involving multiple sequences of unobservable state matrices may be constructed and treated in similar ways.

3. State space models on \( P_{k,m-k} \)

Let us consider a simple state space model on the manifold \( P_{k,m-k} \). We are given time series observations \( \{Q_1, Q_2, \ldots, Q_t\} \) on \( P_{k,m-k} \) related to a sequence of unobservable state matrices \( \{P_0, P_1, \ldots, P_t\} \) on \( P_{k,m-k} \). The observations \( Q_s \) given \( P_s \) are independently distributed as matrix Langevin \( L(P)(m, k; \alpha_s P_s) \), with the mode \( P_s \in P_{k,m-k} \) when \( \alpha_s > 0 \). The unobserved states \( P_s \) given \( P_{s-1} \) are independently distributed as matrix Langevin \( L(P)(m, k; \beta_s P_{s-1}) \), with the mode \( P_{s-1} \) when \( \beta_s > 0 \). Here, we assume that the coefficients \( \alpha_s \) and \( \beta_s \) are known real constants. These conditional densities are

\[ p(Q_s|P_s) = \text{etr}(\alpha_s P_s Q_s)/F_1(\frac{1}{2}k; \frac{1}{2}m; \alpha_s I_k) \]

and

\[ p(P_s|P_{s-1}) = \text{etr}(\beta_s P_{s-1} P_s)/F_1(\frac{1}{2}k; \frac{1}{2}m; \beta_s I_k) \quad \text{for} \quad s = 1, 2, \ldots, t. \] \quad (3.1)

We put

\[ \mathcal{Q}_s = \{Q_0, Q_1, \ldots, Q_s\} \quad \text{with} \quad Q_0 = \emptyset \quad \text{and} \quad \mathcal{P}_s = \{P_0, P_1, \ldots, P_s\} \quad \text{assuming that} \quad P_0 \quad \text{is given} \] \quad (3.2)

and we estimate \( P_s \) given \( Q_s \) by the posterior modes, following the previous discussion in Section 2. We obtain the posterior log-density

\[ L(P_s|Q_s) = \log p(P_s|Q_s) \propto \sum_{s=1}^{t} \text{tr}(\alpha_s P_s Q_s + \beta_s P_{s-1} P_s). \] \quad (3.3)
(i) At the stage $s = 1$, with $P_0^* = P_0$ given, we have
\[
L(P_1|Q_1) \propto \text{tr}[P_1(z_1Q_1 + \beta_1P_0^*)] = \text{tr}[P_1G_1(P_0^*)] \quad \text{say.} \tag{3.4}
\]
We let the spectral decomposition of $G = G_1(P_0^*)$ be
\[
G = HDH' \quad \text{for} \quad H = (H_1 H_2) \in O(m) \quad \text{with} \quad H_1 \text{ being } m \times k \quad \text{and}
\]
\[
D = \text{diag}({\lambda}_1, {\lambda}_2, \ldots, {\lambda}_m) \quad \text{with} \quad {\lambda}_1 > {\lambda}_2 > \cdots > {\lambda}_m \quad \text{almost everywhere} \tag{3.5}
\]
and put
\[
M[G] = H_1 H_1' \in P_{k,m-k}. \tag{3.6}
\]
The mode of the log-density (3.4) is given by
\[
\hat{P}_1 = M[G_1(P_0^*)]. \tag{3.7}
\]

(ii) In general, our iterative procedure for obtaining the estimates $\hat{P}_t^* = (\hat{P}_1^*, \hat{P}_2^*, \ldots, \hat{P}_t^*)$ takes similar steps with the $P_s$ to those (2.11)–(2.16) stated in Section 2.1, where we replace $G_s(X_{s-1})$ defined in (2.9) and $F_s-1(X_{s-1}, X_s)$ defined in (2.10) by
\[
G_s(P_{s-1}) = \alpha_s Q_s + \beta_s P_{s-1} \tag{3.8}
\]
and
\[
F_{s-1}(P_{s-2}, P_s) = G_{s-1}(P_{s-2}) + \beta_s P_s, \tag{3.9}
\]
respectively, and replace $H[\cdot]$ (2.7) by $M[\cdot]$ (3.6). The detailed discussion for the rest of the iterative procedure is omitted.

4. Concluding remarks

The existing literature of state space models focused the discussion on the Euclidean spaces. In this paper, we develop some state space models on the Stiefel and Grassmann manifolds, specifying the conditional matrix Langevin distributions for the noise processes. The matrix Langevin distributions, which are parametrized by modal orientations and concentrations, are most useful distributions for statistical analysis on these manifolds and play important roles analogous to those of the normal distributions on the Euclidean spaces; see Chikuse [5]. We investigate the estimation of unobservable states, i.e., the modal orientations and/or orthogonal regression matrices for the matrix Langevin noise processes, via Bayes methods using the posterior modes. This may be related to the estimation of unobservable states, the means, for the normal noise processes of the Kalman filter on the Euclidean spaces, being given via Bayes methods using the posterior means (see e.g., [22]). The concentration parameters are assumed to be known throughout the paper. Chikuse’s [5] sampling methods may be applied to the estimation of the concentrations and the research will be left to a future paper.

The state space models developed in this paper may arise in practical applications of orientational statistical analysis in those fields described in Introduction. Some examples of orientational statistics and statistical analysis on the manifolds are given in Chikuse [5]. The model discussed in Section 2.2, in consideration of orientational regressions for the noise processes of states, would broaden the range of applications of state space models on the manifolds.
Acknowledgments

The author is grateful to the referees for helpful comments and suggestions.

References

Control Conference, University of Michigan, Ann Arbor, MI, 1968, The American Society of Mechanical Engineers,
New York, pp. 87–105.
[7] Y. Chikuse, G.S. Watson, Large sample asymptotic theory of tests for uniformity on the Grassmann manifold, J.
1270–1285.
(1964) 475–501.
[18] P.E. Jupp, K.V. Mardia, Maximum likelihood estimators for the matrix von Mises–Fisher and Bingham distributions,
1032–1041.
1983.