

ON THE GENERALIZATION OF AR PROCESSES TO RIEMANNIAN MANIFOLDS

João Xavier*

Instituto Superior Técnico
Instituto Sistemas e Robótica
Av. Rovisco Pais, 1049-001
Lisbon, Portugal
email: jxavier@isr.ist.utl.pt

Jonathan H. Manton

Department of Information Engineering
Research School of Information Sciences and Engineering
Australian National University
Canberra ACT 0200, Australia
email: j.manton@ieee.org

ABSTRACT

The autoregressive (AR) process is fundamental to linear signal processing and is commonly used to model the behaviour of an object evolving on Euclidean space. In real life, there are myriad examples of objects evolving not on flat spaces but on curved spaces such as the surface of a sphere. For instance, wind-direction studies in meteorology and the estimation of relative rotations of tectonic plates based on observations on the Earth's surface deal with spherical data, while subspace tracking in signal processing is actually inference on the Grassmann manifold. This paper considers how to extend the AR process to one evolving on a curved space, or in a general, a manifold. Doing so is non-trivial, and in fact, several different extensions are proposed, along with their advantages and disadvantages. Algorithms for estimating the parameters of these generalized AR processes are derived.

1. INTRODUCTION

The autoregressive (AR) process is a popular model for a stochastic source on Euclidean space. However, in many engineering scenarios the observations lie on a curved space (spheres, orthogonal matrices, etc) or are more naturally interpreted in a quotient space (projective spaces, Grassman manifolds, etc). The field of directional statistics is filled with such examples: examining wind directions corresponds to studying time-series on the unit-circle, spherical data arises when examining relative rotations of Earth's tectonic plates, etc. See [1, 2, 3, 4] and the references therein for more examples. It is thus of interest to extend the AR modelling tool to such spaces. A common feature on the aforementioned observation spaces is the fact that they can be structured as Riemannian manifolds. In this paper, we discuss the extension of the AR process to Riemannian manifolds. We restrict our attention to the discrete-time case. Although some particular continuous-time stochastic models have been proposed for manifolds, e.g., Lévy processes in Lie groups [5], the specific AR discrete-time case has received little attention (also, there are some striking differences between continuous-time and discrete-time stochastic processes on manifolds: see [4] for an example on the unit-circle).

2. REVIEW OF AR PROCESSES ON THE UNIT-CIRCLE

In order to put in perspective our generalization of AR models to Riemannian manifolds, we start by reviewing the three main approaches

*This work was done while the first author visited the Research School of Information Sciences and Engineering (RSISE) at ANU, Australia

taken in constructing AR processes on the unit-circle $S^1 = \{z \in \mathbb{R}^2 : \|z\| = 1\}$. See [1, 2] for a more in-depth discussion. To the best of our knowledge, the unit-circle is the manifold for which most AR models have been proposed. There are essentially three types of approaches: projected processes, wrapped processes and linked processes.

Projected processes: this approach induces a stochastic process on S^1 by radially projecting onto the unit-circle a process $\{(x_k, y_k) : k = 1, 2, \dots\}$ evolving on the plane \mathbb{R}^2 . The processes $\{x_k\}$ and $\{y_k\}$ may be coloured. As a special case, if $\{x_k\}$ and $\{y_k\}$ represent independent zero-mean Gaussian processes with the same variance, then each $z_k = (x_k, y_k) / \|(x_k, y_k)\|$ is uniformly distributed on S^1 ;

Wrapped processes: let $\{x_k\}$ be an AR process of order p on the real line \mathbb{R} . By wrapping \mathbb{R} onto the unit-circle as $x \mapsto e^{i2\pi x}$ (here, we are identifying S^1 with the subset of complex numbers with unit-modulus) we construct the so-called *circular wrapped autoregressive process* of order p , see [3];

Linked processes: a link function is a bijective, odd monotone map from \mathbb{R} onto $(-\pi, \pi)$. Popular choices: the inverse tan and the scaled probit links given by $g(x) = 2 \tan^{-1} x$ and $g(x) = 2\pi(\Phi(x) - 0.5)$, respectively (Φ is the distribution function of the standard normal). Similar to the wrapping map, link functions can transfer processes $\{x_k\}$ on the real line to ones on the unit-circle as $\{z_k = (\cos(g(x_k)), \sin(g(x_k)))\}$. If $\{x_k\}$ denotes an AR process of order p then $\{z_k\}$ is a so-called *circular linked autoregressive process* of order p on the unit-circle. A mean direction μ can be assigned to $\{z_k\}$ through multiplication with the phasor $e^{i\mu}$. Within this model, realizations of z_k near $e^{i(\mu+\pi)}$ are scarce, because only realizations of x_k with large magnitude are mapped there.

3. A NEW AR MODEL FOR RIEMANNIAN MANIFOLDS

When considering the construction of AR processes on arbitrary Riemannian manifolds, it is natural to start by trying to extend the three approaches delineated in section 2 for the unit-circle. However, we argue that none is really suited for our purpose, which consists in constructing AR processes on a Riemannian manifold with intrinsic-only tools. We will make a new proposal for an AR model for arbitrary Riemannian manifolds – its particularization for the unit-circle is discussed in section 5. Our objections to the aforementioned three approaches are as follows.

Projected processes: this approach relies on projecting an AR process evolving on an ambient Euclidean space onto our Riemannian manifold M . Clearly this construction is not intrinsic to M : we need first to represent M as a submanifold of some \mathbb{R}^n . Although such a smooth embedding can always be found at least theoretically

by Whitney's theorem [6] (even an isometric one exists by Nash's theorem) the main drawback here is that all the subsequent analysis will depend on the particular choice of the embedding we adopted;

Linked processes: essentially, the same argument used in the previous paragraph applies here. Indeed, this approach requires again the presence of an extrinsic object to M , the link map, and all the features of the resulting AR process will depend strongly on the particular link adopted (we leave aside the discussion about the easiness of finding such a link for a generic manifold);

Wrapped processes: this approach is an intrinsic one and can readily be extended to Riemannian manifolds, at least those which are connected and geodesically complete. Indeed, for this class of manifolds the Riemannian exponential map at $p \in M$, denoted $\text{Exp}_p : T_p M \rightarrow M$, which acts from the tangent space to M at p is an onto map (this is Hopf-Rinow's theorem [6]). Thus, we can create a standard AR process on the linear tangent space $T_p M$ of a particular point $p \in M$ and wrap it onto M through Riemannian exponentiation. Note that, for the particular case of $M = S^1$, we do recover the circular wrapped autoregressive process discussed earlier (by letting $p = (1, 0)$ and assuming, as usual, that S^1 is viewed as a Riemannian submanifold of \mathbb{R}^2). Our main discomfords regarding this approach are: (i) a particular point p is granted a privileged role among all the others in the manifold M ; (ii) the manifold M (and any additional structure that it might possess, e.g., a Lie group structure) is essentially ignored in the construction of the AR process: the latter evolves on the flat tangent space $T_p M$ and M enters the picture only as the range space of the Exp_p map.

New AR model. Let (M, g) be a Riemannian manifold. For simplicity, we shall assume that M is connected and the Riemannian metric g makes M geodesically complete (e.g., compactness of M is a sufficient condition for this to occur). Our inspiration for the new AR model originates in a possible interpretation of an AR model on \mathbb{R}^n . Let $\{x_k\}$ be an AR process of order p evolving on \mathbb{R}^n . We have

$$x_k = A_1 x_{k-1} + \dots + A_p x_{k-p} + \text{noise} \quad (1)$$

where each A_i denotes an $n \times n$ matrix. Defining the increments $\Delta_k = x_k - x_{k-1}$ we have

$$x_k = x_{k-1} + \Delta_k \quad (2)$$

where the sequence of increments is governed by the recursive iteration

$$\Delta_k = A_1 \Delta_{k-1} + \dots + A_p \Delta_{k-p} + \text{noise}. \quad (3)$$

Viewing \mathbb{R}^n as a Riemannian manifold (the usual flat Riemannian metric is assumed) we can interpret (2) as

$$x_k = \text{Exp}_{x_{k-1}}(\Delta_k). \quad (4)$$

That is, x_k is found by travelling along the geodesic which departs from x_{k-1} in the direction of the tangent vector $\Delta_k \in T_{x_{k-1}} \mathbb{R}^n \simeq \mathbb{R}^n$ during a unit time interval (this geodesic is $\text{Exp}_{x_{k-1}}(t\Delta_k) = x_{k-1} + t\Delta_k$ for $t \geq 0$). Moreover, equation (3) can be interpreted as the stochastic law which generates the "fresh" increment Δ_k from the previous p ones $\Delta_{k-1}, \dots, \Delta_{k-p}$. Each increment lives in a distinct tangent space: $\Delta_{k-i} = \text{Log}_{x_{k-i-1}}(x_{k-i}) = x_{k-i} - x_{k-i-1} \in T_{x_{k-i-1}} \mathbb{R}^n$ for $i = 0, 1, \dots, p$. Here and for future reference, we use the notation Log_p to denote the Riemannian log map at $p \in M$. Since \mathbb{R}^n is a flat space (curvature tensor vanishes identically everywhere) any given two tangent spaces $T_a \mathbb{R}^n$ and $T_b \mathbb{R}^n$ are naturally identified by parallel-transportation $P_{a,b} :$

$T_a \mathbb{R}^n \rightarrow T_b \mathbb{R}^n$, e.g., along the geodesic which runs from a to b . Under the canonical isomorphism $T_p \mathbb{R}^n \simeq \mathbb{R}^n$ we have $P_{a,b} = \text{id}_{\mathbb{R}^n}$. Thus, equation (3) can be equivalently rewritten as

$$\begin{aligned} \Delta_k &= A_1 P_{x_{k-2}, x_{k-1}} \left(\text{Log}_{x_{k-2}}(x_{k-1}) \right) + \dots + \\ &A_p P_{x_{k-p-1}, x_{k-1}} \left(\text{Log}_{x_{k-p-1}}(x_{k-p}) \right) + w_k, \end{aligned} \quad (5)$$

where w_k denotes a random vector in $T_{x_{k-1}} \mathbb{R}^n$ – this implements the additive noise term in (3). In (5), we view each A_i as the matrix representation of an endomorphism (linear map) $A_i : T_{x_{k-1}} \mathbb{R}^n \rightarrow T_{x_{k-1}} \mathbb{R}^n$. The i th map processes the increment $\text{Log}_{x_{k-i-1}}(x_{k-i})$ (more exactly, its parallel-transported version at x_{k-1}).

We take (4) and (5) as the description of our AR model for generic Riemannian manifolds M . More precisely, we construct an AR process $\{x_k\}$ of order p on M as

$$\Delta_k = \sum_{i=1}^p A_i P_{x_{k-i-1}, x_{k-1}} \left(\text{Log}_{x_{k-i-1}}(x_{k-i}) \right) + W_k \quad (6)$$

$$x_k = \text{Exp}_{x_{k-1}}(\Delta_k). \quad (7)$$

Some remarks follow.

Remark 1: In (6) the symbol W_k denotes a random tangent vector in $T_{x_{k-1}} M$ (may be correlated with parallel-transported versions of W_{k-1}, W_{k-2}, \dots);

Remark 2: In general, the Riemannian log map Log_p is only defined in a neighborhood of p . However, in our case, we are considering M to be connected and geodesically complete. As a consequence, for any given $p \in M$, our manifold can be written as the disjoint union $M = V_p \cup \text{Cut}(p)$ where V_p is an open subset and $\text{Cut}(p)$ is the cut locus of p [7]. In fact, V_p is the diffeomorphic image under Exp_p of an open subset $U_p \subset T_p M$ which is star-shaped with respect to the origin of $T_p M$. Thus, Log_p is well-defined on V_p . Note that $\text{Exp}_p(\text{Log}_p(q)) = q$ for all $q \in V_p$. However, it is crucial to note that it is not necessarily true that $\text{Log}_p(\text{Exp}_p(X_p)) = X_p$ for any tangent vector $X_p \in T_p M$ satisfying $\text{Exp}_p(X_p) \in V_p$ (the latter equality holds only when $X_p \in U_p$). The subset $\text{Cut}(p)$ has measure zero with respect to the canonical Riemannian measure, say μ_M , carried by M [7]. Thus, if the noise model induces a probability measure which is continuous with respect to μ_M then, almost surely, for two consecutive points we have $x_{k+1} \notin \text{Cut}(x_k)$ which means that $\text{Log}_{x_{k-i-1}}(x_{k-i})$ is well-defined;

Remark 3: In (6) we use the notation $P_{p,q} : T_p M \rightarrow T_q M$ to denote parallel-transportation along the geodesic which runs from p to q (this geodesic is unique for $q \notin \text{Cut}(p)$ which we suppose occurs with probability one – see previous remark);

Remark 4: In (6) each symbol A_i denotes a smooth section of the tensor bundle $T_1^1(M)$. In equivalent terms, each A_i is a smooth field of endomorphisms on M : A_i assigns a linear map $T_p M \rightarrow T_p M$ for any given $p \in M$. We let $A_i X_p$ be the linear map $A_i|_p : T_p M \rightarrow T_p M$ acting on $X_p \in T_p M$, e.g., in (6) all these linear maps are operating on $T_{x_{k-1}} M$. By smoothness of each field A_i we mean that the vector field $p \mapsto A_i X_p \in T_p M$ is smooth, for any given smooth vector field X on M . Note that the fields A_1, A_2, \dots, A_p generalize the matrices A_1, A_2, \dots, A_p appearing in AR processes on \mathbb{R}^n , see (1).

4. TWO ALTERNATIVE EXTENSIONS

The appearance of $\text{Cut}(p)$ in remark 2 of section 3 is not aesthetically pleasing. Since $\text{Cut}(p)$ can only cause problems if the AR pro-

cess takes too big a step, the way to remove this problem is to consider instead the infinitesimal generator of the AR process. Specifically, a discrete time AR process on a manifold can be obtained by first defining a linear stochastic differential equation on the manifold, such as an Ornstein-Uhlenbeck process, and then sampling it at unit intervals, or equivalently, integrating the differential equation over unit intervals to form a discrete time process. Needless to say, in general there will not be a closed form expression for the resulting discrete time process.

Yet another interpretation of a *stable* discrete time AR process on \mathbb{R}^n is the following. The origin in \mathbb{R}^n plays a distinguished role; if the noise excitation goes to zero, then a stable AR process will approach the origin. Moreover, the AR can be rewritten in the form that the next point is equal to the weighted centre of mass of the previous p points together with the origin, followed by a perturbation caused by noise. By replacing centre of mass by Karcher mean, this idea extends to an arbitrary Riemannian manifold. Note though that for manifolds with positive curvature, if the points of the AR process are too far apart, the Karcher mean may not be unique and some mechanism for handling this needs to be introduced.

5. AR MODEL FOR THE UNIT-CIRCLE

We specialize our AR model in (6) and (7) to the unit-circle S^1 , viewed hereafter as a Riemannian submanifold of \mathbb{R}^2 . To obtain compact expressions, we shall identify $\mathbb{R}^2 \simeq \mathbb{C}$ and $S^1 \simeq \{z \in \mathbb{C} : |z| = 1\}$. Thus, the tangent space to S^1 at z corresponds to $T_z S^1 \simeq \text{span}\{iz\}$ where $i = \sqrt{-1}$. Within this identification the Riemannian metric is $\langle \delta, \omega \rangle = \text{Re}\{\delta \bar{\omega}\}$ for $\delta, \omega \in T_z S^1$ (\bar{z} represents the complex conjugate of z). For $z \in S^1$ we have $\text{Cut}(z) = \{-z\}$ and $\text{Log}_z(w) = z \log(z^{-1}w)$ whenever $w \notin \text{Cut}(z)$ (here, \log denotes the complex logarithm associated with the branch $(-\pi, \pi)$). Moreover, $\text{Exp}_z(\delta) = z \exp(z^{-1}\delta)$ for any $\delta \in T_z S^1$ (here, \exp denotes the complex exponential). Finally, parallel-transportation $P_{z,w} : T_z S^1 \rightarrow T_w S^1$ is given by $P_{z,w}(\delta) = w z^{-1} \delta$. To specify an AR process of order p we have to define p smooth fields of endomorphisms a_1, a_2, \dots, a_p . The i th field a_i smoothly assigns an endomorphism $T_z S^1 \rightarrow T_z S^1$ to each tangent space. Since each $T_z S^1$ is one-dimensional, each endomorphism $T_z S^1 \rightarrow T_z S^1$ acts as multiplication by a real number, say, $a_i(z) \in \mathbb{R}$. The field of endomorphisms thus created is smooth if the function $z \mapsto a_i(z)$ is smooth on S^1 . All said, the AR model of order p on the unit-circle takes the form

$$z_k = z_{k-1} \exp \left(\left(\sum_{i=1}^p a_i(z_{k-1}) \log(z_{k-1}^{-1} z_{k-i}) \right) + w_k \right)$$

where $w_k \in T_{z_{k-1}} S^1$ denotes a random tangent vector.

6. AR MODEL FOR COMPACT CONNECTED LIE GROUPS

In this section, we discuss the application of the AR model (6)-(7) to the special case of compact connected Lie groups G equipped with a bi-invariant metric (the previous case S^1 can be captured in this framework, although we are not requiring here that G is Abelian). Note that because G is compact, such a bi-invariant metric exists [6]. For the sake of clarity, we consider only AR processes of order $p = 1$. We will see that when both the field of endomorphisms $A \equiv A_1$ and the random vector fields W_k are left-invariant (thus adapted to the underlying group structure), then the process $\{x_k\}$

proceeds by computing its increments in the Lie algebra of G , denoted hereafter by \mathfrak{g} . The Lie group structure of G is thus exploited within this construction and as a result the AR process definition can be described, to a large extent, only in linear-algebraic terms. Note the distinction between this result and the *wrapped process* construction: here, it is the increments which are generated in the Lie algebra \mathfrak{g} and not some AR process which is then projected onto G by exponentiation (as in constructing *wrapped processes*). Since elements in the Lie algebra (in any tangent space, for that matter) correspond naturally to small “displacements” within the manifold, it seems that this approach is more intuitive.

By saying that the field of endomorphisms $A = \{A|_g : T_g G \rightarrow T_g G\}$ is left-invariant we mean that the diagram in figure 1 commutes for any $g, h \in G$. Here, $L_g : G \rightarrow G$, $L_g(x) = g \cdot x$ denotes

$$\begin{array}{ccc} T_h G & \xrightarrow{A|_h} & T_h G \\ L_{g*} \downarrow & & \downarrow L_{g*} \\ T_{gh} G & \xrightarrow{A|_{gh}} & T_{gh} G \end{array}$$

Fig. 1. The field of endomorphisms A is left-invariant

left-translation in G by the group element g (the notation $g \cdot x$ denotes group multiplication of g by x) and L_{g*} stands for its differential map. This means that the field A is completely determined by sampling it at the Lie algebra $\mathfrak{g} \simeq T_e G$ (e denotes the identity element of G): $A|_g(X_g) = (L_{g*} \circ A|_e \circ L_{g^{-1}*})(X_g)$ for any $X_g \in T_g G$. To simplify notation, we write hereafter AX_e for $A|_e(X_e)$ for any $X_e \in \mathfrak{g}$. Also, if W_k denotes a random left-invariant vector field then we can write $W_k|_g = L_{g*}(\Omega_k)$ for all k and g , where $\{\Omega_k\}$ is a random sequence of vectors in \mathfrak{g} .

Lemma: Under the above assumptions, the model (6)-(7) boils down to

$$\Theta_k = A \text{Log}(x_{k-2}^{-1} \cdot x_{k-1}) + \Omega_k \quad (8)$$

$$x_k = x_{k-1} \cdot \text{Exp}(\Theta_k) \quad (9)$$

The proof is omitted due to paper length constraints. Here, the symbols Exp and Log denote the Riemannian exponential and log maps computed at the identity e (because the metric is bi-invariant Exp coincides with the Lie group \exp map which generates the one-parameter subgroups of G). All the action in (8) takes place in the (fixed) vector space \mathfrak{g} . The “old” increment $\text{Log}(x_{k-2}^{-1} x_{k-1})$ is processed through A to establish the next direction of movement, which is then disturbed additively by the random tangent vector Ω_k . In (9), group multiplication is invoked to move x_{k-1} according to the increment Θ_k . Note that (8)-(9) are formulated in terms of $T_e G \simeq \mathfrak{g}$ for mere convenience. The identity element $e \in G$ is not being granted a privileged role in the AR process construction: indeed, due to the left-invariance of both A and the random vector field W_k , any other point could equivalently be chosen as a “fiducial” point to express the process construction.

Re-interpretation as a process on the tangent bundle. Any Lie group G is parallelizable and therefore its tangent bundle TG is trivial: in particular, we have the identification $TG \simeq \mathfrak{g} \times G$ (this is obtained by fixing a basis for \mathfrak{g} and left-translating it through G). The latter Cartesian product acts as a system of coordinates for TG which means that any tangent vector $X_g \in T_g G \subset TG$ can be represented uniquely as (Θ, g) for some $\Theta \in \mathfrak{g}$. In this sense, we

can interpret (8) and (9) as describing a Markov process $\{(\Theta_k, x_k)\}$ evolving on the tangent bundle TG . We only observe the second component $\{x_k\}$ though.

Parameter estimation. We briefly comment on the maximum likelihood (ML) strategy for estimating the AR parameter A in (8), given an observed time-series $\{x_1, x_2, \dots, x_K\}$. We focus on the simplest case $G = S^1$: this gives the flavor of the difficulties encountered in the general setup. In this case, we have the identification $\mathfrak{g} \simeq \text{span}\{\iota\} \simeq \mathbb{R}$. The latter identification works as $\iota t \mapsto t$ for a real t (we recall that $\iota = \sqrt{-1}$). Thus, $A : \mathfrak{g} \rightarrow \mathfrak{g}$ is represented by a real number $A \simeq a \in \mathbb{R}$. The ML estimator is

$$\hat{a}_{\text{ML}} = \arg \max_a p(x_1, x_2, x_3, \dots, x_K; a).$$

It is more convenient to change variables from $(x_1, x_2, \dots, x_K) \in S^1 \times \dots \times S^1$ (K times) to $(x_1, x_2, \gamma_3, \dots, \gamma_K) \in S^1 \times S^1 \times (-\pi, \pi) \times \dots \times (-\pi, \pi)$ ($K - 2$ times) where

$$\gamma_k = \text{Im} \{ \log(x_{k-1}^{-1} x_k) \} \quad (10)$$

and $\text{Im}\{z\}$ retrieves the imaginary part of the complex number z . Recall that \log denotes the complex logarithm associated with the branch $(-\pi, \pi)$. Since \log returns its values in \mathfrak{g} , (10) is nothing but the aforementioned identification $\mathfrak{g} \simeq \mathbb{R}$ at play. Note that, in terms of these new variables, the model (8)-(9) corresponds to

$$\theta_k = a\gamma_{k-1} + \omega_k \quad (11)$$

$$\gamma_k = \log(\exp(\iota\theta_k)). \quad (12)$$

Hereafter, we examine the independent and identically distributed (iid) noise additive case, i.e., $p(\omega_1, \dots, \omega_K) = \prod_{k=1}^K p_\omega(\omega_k)$ where p_ω denotes the common density. To implement the ML estimator

$$\hat{a}_{\text{ML}} = \arg \max_a p(x_1, x_2, \gamma_3, \dots, \gamma_K; a)$$

we need to write the likelihood function $p(x_1, x_2, \gamma_3, \dots, \gamma_K; a)$. We suppose that the first two points x_1 and x_2 are randomly generated according to some density $p(x_1, x_2)$ which does not depend on a : this creates the first increment $\gamma_2 = \text{Im}\{\log(x_1^{-1} x_2)\}$. The AR process then evolves according to (11)-(12) for $k \geq 3$. Due to the Markov structure of (11)-(12) we have $p(x_1, \dots, \gamma_K; a) = p(x_1, x_2) \prod_{k=3}^K p(\gamma_k | \gamma_{k-1}; a)$. Now

$$\begin{aligned} p(\gamma_k | \gamma_{k-1}; a) &= \int p(\gamma_k | \theta_k, \gamma_{k-1}; a) p(\theta_k | \gamma_{k-1}; a) d\theta_k \\ &= \int p(\gamma_k | \theta_k; a) p_\omega(\theta_k - a\gamma_{k-1}) d\theta_k \quad (13) \end{aligned}$$

$$= \sum_{n_k=-\infty}^{+\infty} p_\omega(\gamma_k + 2\pi n_k - a\gamma_{k-1}). \quad (14)$$

Passage from (13) to (14) relies on (12) which establishes γ_k as a deterministic function of θ_k : we have $\gamma_k = \theta_k \bmod 2\pi$. We then use $p(\gamma_k | \theta_k; a) = \sum_{n_k} \delta(\gamma_k + 2\pi n_k - \theta_k)$ where δ is the Dirac delta function. This argument is admittedly informal but quite common and can be justified more rigorously by measure-theoretic principles, e.g., see equation (40) in [4]. The end result is

$$p(x_1, \dots, \gamma_K; a) \propto \prod_{k=3}^K \left(\sum_{n_k=-\infty}^{+\infty} p_\omega(\gamma_k + 2\pi n_k - a\gamma_{k-1}) \right) \quad (15)$$

where \propto means equality up to a multiplicative constant ($p(x_1, x_2)$ was dropped). We see that for practical implementation of the ML estimator, series truncation in (15) is almost unavoidable (the decay rate of the tails of p_ω should provide an educated guess to perform the truncation). Alternatively, one might consider an expectation-maximization (EM) approach for computing \hat{a}_{ML} : if the θ_k in (11) were observed, estimation of a corresponds essentially to a linear system identification problem. Detailed derivation of this approach is beyond the scope of this paper. The main point here is that equation (15) illustrates the typical format of the likelihood function for generic groups G . Indeed, the previous analysis extends to arbitrary G by defining $\Gamma_k = \text{Log}(x_{k-1}^{-1} x_k)$ and rewriting (8)-(9) as

$$\Theta_k = A\Gamma_{k-1} + \Omega_k \quad (16)$$

$$\Gamma_k = \text{Log}(\text{Exp}(\Theta_k)), \quad (17)$$

which specialize to (11)-(12) when $G = S^1$. The source of the complications is (17): the Θ_k are not observed and the map $\Theta_k \mapsto \Gamma_k$, written hereafter as \mathcal{F} , is many-to-one, as we already mentioned in section 3 (see remark 2). For $G = S^1$, we have $\mathcal{F}^{-1}(\gamma) = \{\gamma + 2\pi n : n \in \mathbb{Z}\}$. For other groups, the inverse set $\mathcal{F}^{-1}(\Gamma)$ must be derived and used in the equation corresponding to (13).

7. CONCLUSIONS AND EXTENSIONS

There is more than one way to define a process on a manifold such that, if the manifold happened to be \mathbb{R}^n , the process would correspond to an AR process. That is, there is not a single correct way to extend the concept of an AR process to a manifold. Rather, of great interest are processes on manifolds which can be used to model practical problems and which are amenable to analysis and parameter estimation. In this paper, three possible extensions of an AR process to a manifold were proposed, one of which was studied in detail. Applications of these processes to engineering problems will be considered in the future, as will a natural extension of the idea in section 6, namely, to the case when the manifold M is not itself a group but when a Lie group G acts on M transitively (in which case M is said to be a homogeneous space). One could envisage transferring an AR process on G to M via the group action.

8. REFERENCES

- [1] N. Fisher and A. Lee, "Time series analysis of circular data," *J. Royal Statist. Soc. B*, vol. 56, no. 2, pp. 327–339, 1994.
- [2] K. Mardia and P. Jupp, *Directional Statistics*, John Wiley & Sons, 2000.
- [3] J. Breckling, *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, Lecture Notes in Statistics, 61. Springer-Verlag, Berlin, 1989.
- [4] A. Willsky and J. Lo, "Estimation for rotational processes with one degree of freedom – part II: discrete-time processes," *IEEE Trans. on Automatic Control*, vol.AC-20, no.1, February 1975.
- [5] M. Liao, *Lévy processes in Lie groups*. Cambridge University Press, 2004.
- [6] W. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry*. 2nd edition, New York: Academic Press.
- [7] S. Gallot, D. Hulin, and J. Lafontaine. *Riemannian Geometry*, 2nd ed. Springer-Verlag.
- [8] B. Hall. *Lie groups, Lie algebras and representations: an elementary introduction*, Springer-Verlag.