Group Action Induced Distances on Spaces of High-Dimensional Linear Stochastic Processes

Bijan Afsari and René Vidal

Center for Imaging Science, Johns Hopkins University, Baltimore MD 21218, USA
bijan,rvidal@cis.jhu.edu

Abstract. This paper studies the geometrization of spaces of stochastic processes. Our main motivation is the problem of pattern recognition in high-dimensional time-series data (e.g., video sequence classification and clustering). First, we review some existing approaches to defining distances on spaces of stochastic processes. Next, we focus on the space of processes generated by (stochastic) linear dynamical systems (LDSs) of fixed size and order (this space is a natural choice for the pattern recognition problem). When the LDSs are represented in state-space form, the space of LDSs can be considered as the base space of a principal fiber bundle. We use this fact to introduce a large class of easy-to-compute group action-induced distances on the space of LDSs and hence on the corresponding space of stochastic processes. We call such a distance an alignment distance. One of our aims is to demonstrate the usefulness of control-theoretic tools in problems related to stochastic processes.

Keywords: Stochastic processes, pattern recognition, linear dynamical systems, extrinsic and intrinsic geometries, principal fiber bundle.

1 Introduction and motivation

Pattern recognition (e.g., classification and clustering) of time series data is important in many real world data analysis problems. Early applications include the analysis of one-dimensional data such as speech and seismic signals (see, e.g., [18] for a review). More recently, applications in the analysis of video data (e.g., activity recognition [1]), robotic surgery data (e.g., surgical skill assessment [5]), or biomedical data (e.g., analysis of multichannel EEG signals) have motivated the development of statistical techniques for the analysis of high-dimensional (or vectorial) time-series data. There are different grand strategies in dealing with this problem [18]. One is a parametric approach based on *modeling* the observed time series and then performing statistical analysis and inference on a *space of models* (instead of the space of the observed *raw* data). In many real-world instances (e.g., video sequences [1,5,21]), one could model the observed high-dimensional time series with low-order Linear Dynamical Systems (LDSs). In such instances the mentioned strategy could prove beneficial, e.g., in terms of implementation (due to significant compression achieved in high dimensions), statistical inference, and synthesis of time series. These motivations lead us to state the following prototype problem.

Problem 1 (Statistical analysis on spaces of LDSs). Let $\{\mathbf{y}^i\}_{i=1}^N$ be a collection of p-dimensional time series indexed by time t. Assume that each time series $\mathbf{y}^i = \{\mathbf{y}_t^i\}_{t=1}^\infty$ can be approximately modeled by an (stochastic) LDS M_i of output-input size (p, m) and order n[1] realized as

$$\mathbf{x}_t^i = A_i \mathbf{x}_{t-1}^i + B_i \boldsymbol{v}_t, \tag{1}$$
$$\mathbf{y}_t^i = C_i \mathbf{x}_t^i + D_i \boldsymbol{v}_t, \quad (A_i, B_i, C_i, D_i) \in \widetilde{\mathcal{SL}}_{m,n,p} = \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times m} \times \mathbb{R}^{p\times n} \times \mathbb{R}^{p\times m}$$

where $\boldsymbol{v}_t$ is a common stimulus (e.g., white Gaussian noise with identity covariance) and where the realization $R_i = (A_i, B_i, C_i, D_i)$ is learnt and assumed to be known. The problem is to: (1) Choose an appropriate space $\mathcal{S}$ of LDSs containing the learnt models $\{M_i\}_{i=1}^N$, (2) geometrize $\mathcal{S}$, i.e., equip it with an appropriate geometry (e.g., define a distance on $\mathcal{S}$), (3) develop tools (e.g., probability distributions, averages or means, variance, PCA) to perform statistical analysis (e.g., classification and clustering) in a computationally efficient manner.

The state-space representation (1) is quite general and with n large enough it can approximate a large class of processes. In Problem 1, obviously, the first two steps (which are the focus of this paper) have significant impacts on the third one. One has different choices for the space $\mathcal{S}$ as well as geometries on that space. The gamut ranges from an *infinite dimensional linear* space to a *finite dimensional (non-Euclidean) manifold*, and the geometry can be either *intrinsic* or *extrinsic*. By an intrinsic geometry we mean one in which a shortest path between two points in a space stays in the space, and by an extrinsic geometry we mean one where the distance between the two points is measured in an *ambient* space. Our approach is somewhere in between: to design an *easy-to-compute* extrinsic distance, while keeping the ambient space *not* too large.

This paper is organized as follows: In Section 2, we review some existing approaches in geometrization of spaces of stochastic processes. In Section 3, we focus on processes generated by LDSs of fixed order, and in Section 4, we study smooth fiber bundle structures over spaces of LDSs generating such processes. Finally, in Section 5, we introduce our class of group action induced distances namely the *alignment* distances. The paper is concluded in Section 6. Due to space limitation, proofs will appear elsewhere and the reader is referred to standard texts on differential geometry or control theory for some basic definitions.

2 A Review of Existing Approaches

This review, in particular, since the subject appears in a range of disciplines is non-exhaustive. Unless otherwise stated, by a *process* we mean a discrete-time (wide sense) stationary zero mean Gaussian stochastic process with no deterministic component. In view of Problem 1, our main interest is in the finite dimensional spaces of LDSs of fixed order and the processes they generate. However, since such a space can be embedded in the larger infinite dimensional space of "virtually all processes," first we consider the latter.

[1] Typically in video analysis: $p \approx 1000 - 10000$, $m, n \approx 10$ (see e.g., [1,5,21]).

Geometrizing the space of power spectral densities. A process can be *identified* with its *covariance sequence* or equivalently the Fourier (or z) transform of its covariance sequence, namely the *power spectral density*. We denote the space of all $p \times p$ power spectral density (PSD) matrices by $\mathcal{P}_p$. This is an *infinite dimensional* cone which also has a convex *linear* structure coming from matrix addition and multiplication by nonnegative reals. The subset of $\mathcal{P}_p$ comprised of PSD matrices which are of rank p at every frequency $\omega \in [0, 2\pi]$ is denoted by $\mathcal{P}_p^+$. Most of the literature is devoted to geometrizations of $\mathcal{P}_1^+$.

In the case of $\mathcal{P}_1^+$ the famous Itakura-Saito distance between PSDs has been used in practice, at least since the 1970's (see [18] for references). The high-dimensional version of the Itakura-Saito distance has also been known since the 1980's [13] but less used in practice. The Itakura-Saito distance is essentially induced from the Kullback-Leibler divergence in the time domain between (infinite dimensional) probability densities of two processes.[2] Therefore, it is not a true distance. Amari's information geometry-based approach [3, Ch. 5] allows to geometrize $\mathcal{P}_1^+$ in various ways and yields different distances (e.g., the Itakura-Saito distance). Furthermore, in this framework one can define geodesics between two processes under various Riemannian or non-Riemannian *connections*. Recently, in [11] a Riemannian framework for geometrization of $\mathcal{P}_p^+$ for $p \geq 1$ has been proposed. In general, such approaches are not suited for large p (due to computational costs and the full-rankness requirement).

The space $\mathcal{P}_p$ (or even $\mathcal{P}_p^+$) is *too large*. For, it includes, e.g., ARMA processes of arbitrary large orders, and it is not clear, e.g., how an *average* of some ARMA models or processes of equal order might turn out (it is reasonable to require the average to be of the same order). The ideal is a workable Riemannian framework which respects all nonlinearities and allows to define geodesics, averages, etc.

Geometrizing the spaces of models. Any distance on $\mathcal{P}_p$ (or $\mathcal{P}_p^+$) induces a distance, e.g., on a subspace corresponding to AR or ARMA models of a fixed order. This is an example of an extrinsic distance induced from an *infinite dimensional ambient* space to a *finite dimensional subspace*. In general, this framework is not ideal and we might try to, e.g., define an *intrinsic* distance on the finite dimensional subspace. In fact, Amari's original paper [2] lays down a framework for this approach, but lacks actual computations. For the one-dimensional case in [22], based on Amari's approach, distances between models in the space of ARMA models of fixed order are derived. For high order models or in high dimensions, such calculations are, in general, computationally difficult [22].

Alternative approaches also have been pursued. For example, Piccolo's [20] approach is based on first calculating the (possibly) infinite order AR equivalent of an invertible ARIMA model and then defining a distance between models via embedding the AR(∞) models in the Hilbert space of ℓ^2 sequences. In principle, this approach is extendible to $\mathcal{P}_p^+$ (albeit computationally difficult). In [19] a conceptually similar approach is pursued by defining a distance between two

[2] Notice that defining distances between probability densities in the time domain is a more general approach than PSD-based approaches and can be employed in the case of nonstationary as well as non-Gaussian processes.

processes via a weighted ℓ^2 distance between the sequences of the cepstrum coefficients of the two processes. Interestingly, in the case of ARMA models this specific distance can be expressed in closed form in terms of the poles and zeros of the models. In [7], this approach is given a simple interpretation in terms of state-space parameters (in particular subspace angles between extended observability matrices). In [6], this approach is shown to be a special case of the family of Binet-Cauchy kernels introduced in [24]. In [4], the space of AR processes of order p is geometrized using the geometry of Toeplitz matrices (via reflection coefficient parameterizations). Extensions of all these methods to $\mathcal{P}_p$ for $p > 1$ do not seem obvious.

More relevant to us are [16] and [9], where (*intrinsic*) state-space based Riemannian distances between LDSs of fixed size and fixed order have been studied. Such approaches ideally suit Problem 1, but they are computationally demanding. More recently, in [1] we introduced group action induced distances on a specific class of LDS spaces of *fixed size* and *order*. Here, we give further theoretical foundation for that approach.

3 Processes generated by finite order LDSs

Consider an LDS, M, of the form (1) with *a* realization $R = (A, B, C, D) \in \widetilde{\mathcal{SL}}_{m,n,p}$. In the sequel, for various reasons, we will restrict ourselves to increasingly smaller submanifolds of $\widetilde{\mathcal{SL}}_{m,n,p}$ which will be denoted by additional superscripts. Recall that the $p \times m$ matrix transfer function is $T(z) = D + C(I_n - z^{-1}A)^{-1}B$, where $z \in \mathbb{C}$ and I_n is the n-dimensional identity matrix. The output PSD matrix (in the z-domain) is the $p \times p$ matrix function $S(z) = T(z)T^\top(z^{-1})$ ($^\top$ is matrix transpose). The PSD is a rational matrix function of z whose rank (a.k.a. *normal rank*) is constant almost everywhere in $\mathbb{C}$. Stationarity of the output process is guaranteed if M is asymptotically stable, or equivalently if all eigenvalues of A are of modulus smaller than one. We denote the submanifold of such realizations by $\widetilde{\mathcal{SL}}^{\mathrm{a}}_{m,n,p} \subset \widetilde{\mathcal{SL}}_{m,n,p}$.

Embedding stochastic processes in LDS spaces. Two (stochastic) LDSs are indistinguishable if their output PSDs are equal. Using this equivalence on the entire set of LDSs is not useful, because it induces a complicated many-to-one correspondence between the LDSs and the subspace of stochastic processes they generate. However, if we restrict ourselves to the subspace of *minimum phase* LDSs the situation improves. The LDS in (1) is minimum phase if its transfer function has constant rank m *everywhere* outside the unit circle in the complex plane $\mathbb{C}$ (including at infinity). Denote the subspace of minimum phase realizations by $\widetilde{\mathcal{SL}}^{\mathrm{a,mp}}_{m,n,p} \subset \widetilde{\mathcal{SL}}^{\mathrm{a}}_{m,n,p}$. This is clearly an open submanifold of $\widetilde{\mathcal{SL}}^{\mathrm{a}}_{m,n,p}$. In $\widetilde{\mathcal{SL}}^{\mathrm{a,mp}}_{m,n,p}$, the spectral factorization of the output PSD is unique up to an orthogonal matrix [23]: let $T_1(z)$ and $T_2(z)$ have realizations in $\widetilde{\mathcal{SL}}^{\mathrm{a,mp}}_{m,n,p}$ and let $T_1(z)T_1^\top(z^{-1}) = T_2(z)T_2^\top(z^{-1})$, then $T_1(z) = T_2(z)\Theta$ for a unique $\Theta \in O(m)$, where $O(m)$ is the Lie group of $m \times m$ orthogonal matrices.

Equivalent classes of realizations under group actions. A fundamental fact is that there are *symmetries* or *invariances* due to certain Lie *group actions* in the model (1). Let $GL(n)$ denote the Lie group of $n \times n$ non-singular (real) matrices. We say that the Lie group $GL(n) \times O(m)$ *acts* on the realization space $\widetilde{\mathcal{SL}}_{m,n,p}$ (or its subspaces) via the action $\bullet$ defined as

$$(P, \Theta) \bullet (A, B, C, D) = (P^{-1}AP, P^{-1}B\Theta, CP, D\Theta). \tag{2}$$

One can easily verify that under this action the output covariance sequence remains invariant. In general, the converse is not true. That is, two output covariance sequences might be equal while their corresponding realizations are not related via $\bullet$ (due to non-minimum phase and the action not being *free* [17], see below). For the converse to hold we need to impose further *rank* conditions.

Observability and controllability. Recall that *controllability* and *observability* matrices of order k are defined as $\mathcal{C}_k = [B, AB, \ldots, A^{k-1}B]$ and $\mathcal{O}_k = [C^\top, (CA)^\top, \ldots, (CA^{k-1})^\top]^\top$, respectively. A realization is called *controllable* (resp. *observable*) if $\mathcal{C}_k$ (resp. $\mathcal{O}_k$) is of rank n for $k = n$. We denote the subspace of controllable (resp. observable) realizations by $\widetilde{\mathcal{SL}}^{\mathrm{co}}_{m,n,p}$ (resp. $\widetilde{\mathcal{SL}}^{\mathrm{ob}}_{m,n,p}$). The space $\widetilde{\mathcal{SL}}^{\min}_{m,n,p} = \widetilde{\mathcal{SL}}^{\mathrm{co}}_{m,n,p} \cap \widetilde{\mathcal{SL}}^{\mathrm{ob}}_{m,n,p}$ is called the space of *minimal* realizations. An important fact is that we cannot reduce the order (i.e., the size of A) of a minimal realization without changing its input-output behavior.

Tall and full rank LDSs. Another (less studied) rank condition is when C is of rank n (here $p \geq n$ is required). Denote by $\widetilde{\mathcal{SL}}^{\mathrm{tC}}_{m,n,p} \subset \widetilde{\mathcal{SL}}^{\mathrm{ob}}_{m,n,p}$ the subspace of such realizations and call a corresponding LDS *tall and full-rank*. Such LDSs are closely related to *generalized linear dynamic factor models* for (very) high-dimensional time series [8] and also appear in video sequence modeling [1,5,21]. It is easy to verify that all the above realization spaces are smooth open submanifolds of $\widetilde{\mathcal{SL}}_{m,n,p}$. Their corresponding stable or minimum-phase submanifolds (e.g., $\widetilde{\mathcal{SL}}^{\mathrm{a,mp,co}}_{m,n,p}$) are defined in an obvious way.

The following proposition forms the basis of our approach to defining distances between processes: any distance on the space of LDSs with realizations in the above submanifolds (with rank condition) can be used to define a distance on the space of processes generated by those LDSs.

Proposition 1. *Let $\tilde{\Sigma}_{m,n,p}$ be $\widetilde{\mathcal{SL}}^{a,mp,co}_{m,n,p}$, $\widetilde{\mathcal{SL}}^{a,mp,ob}_{m,n,p}$, $\widetilde{\mathcal{SL}}^{a,mp,min}_{m,n,p}$, or $\widetilde{\mathcal{SL}}^{a,mp,tC}_{m,n,p}$. Then realizations $R_1, R_2 \in \tilde{\Sigma}_{m,n,p}$ generate the same PSD matrix if and only if there is a unique $(P, \Theta) \in GL(n) \times O(m)$ such that $(P, \Theta) \bullet R_1 = R_2$.*

4 Principal fiber bundle structures over spaces of LDSs

An LDS M is an equivalent class of realizations related by the action $\bullet$. Hence M sits naturally in a *quotient* space, namely $\widetilde{\mathcal{SL}}_{m,n,p}/(GL(n) \times O(m))$. However, this space is not smooth or even Hausdorff. Recall that if a Lie group G acts on a manifold *smoothly*, *properly*, and *freely* then the quotient space has the structure of a *smooth manifold* [17]. Smoothness of $\bullet$ is obvious. In general, the action of a *non-compact* group such as $GL(n)\times O(m)$ is *not* proper. However, one can verify that the rank conditions we imposed in Proposition 1 are enough to make $\bullet$ both a proper and free action on the realization submanifolds. The resulting quotient manifolds are denoted by dropping the tilde superscript $\sim$, e.g., $\mathcal{SL}^{\mathrm{a,mp,min}}_{m,n,p}$. The next theorem, which is an extension of existing results, e.g., in [9] shows that, in fact, we have a principal fiber bundle structure.

Theorem 1. *Let $\tilde{\Sigma}_{m,n,p}$ be as in Proposition 1 and $\Sigma_{m,n,p}=\tilde{\Sigma}_{m,n,p}/(GL(n)\times O(m))$ be the corresponding quotient LDS space. The realization-system pair $(\tilde{\Sigma}_{m,n,p},\Sigma_{m,n,p})$ has the structure of a smooth principal fiber bundle with structure group $GL(n)\times O(m)$. In the case of $\mathcal{SL}^{a,mp,tC}_{m,n,p}$ the bundle is trivial (i.e., diffeomorphic to a product), otherwise it is trivial only when $m=1$ or $n=1$.*

This theorem implies that the minimality condition is a complicated nonlinear constraint, in the sense that it makes the bundle twisted and nontrivial.

Reduction of structure group to the orthogonal group. Next, we recall the notion of reducing a bundle with non-compact structure group to one with a compact structure group. This will be useful in our geometrization approach in the next section. Interestingly, bundle reduction also appears in *statistical analysis of shapes* under the name of *standardization* [14]. The basic fact is that any principal fiber G-bundle $(\tilde{\Sigma},\Sigma)$ can be reduced to an H-subbundle $\widetilde{\mathcal{O}\Sigma}\subset\tilde{\Sigma}$, where H is the maximal compact subgroup of G [15]. This reduction means that Σ is diffeomorphic to $=\widetilde{\mathcal{O}\Sigma}/H$ (i.e., no topological information is lost by going to the subbundle and the subgroup). Therefore, in our cases of interest we can reduce a $GL(n)\times O(m)$-bundle to an $O(n)\times O(m)$-subbundle. We call such a subbundle a *standardized* realization space or (sub)bundle. One can perform reduction to various standardized subbundles and there is no canonical reduction. However, in each application one can choose an *interesting* one. A reduction is in spirit similar to the Gram-Schmidt orthonormalization. As an example, consider $R=(A,B,C,D)\in\widetilde{\mathcal{SL}}^{\mathrm{a,mp,tC}}_{m,n,p}$, and let $C=UP$ be an orthonormalization of C, where $U^\top U=I_n$ and $P\in GL(n)$. Now the new realization $\hat{R}=(P^{-1},I_m)\bullet R$ belongs to the $O(n)$-subbundle $\widetilde{\mathcal{OSL}}^{\mathrm{a,mp,tC}}_{m,n,p}=\{R\in\widetilde{\mathcal{SL}}^{\mathrm{a,mp,tC}}_{m,n,p}|C^\top C=I_n\}$. Other forms of bundle reduction, e.g., in the case of the nontrivial bundle $\widetilde{\mathcal{SL}}^{\mathrm{a,mp,min}}_{m,n,p}$ are possible. In particular, via a process known as *realization balancing* (see [10]), we can construct a large family of standardized subbundles. Due to space limitations, the details will appear elsewhere.

5 Extrinsic quotient geometry and the alignment distance

Equipping a principal fiber bundle with a group invariant Riemannian metric and constructing an *intrinsic* distance on the base space (via a process known as Riemannian submersion) is well known. However, in our applications, this approach (e.g., as introduced in [16]) is computationally prohibitive, since it involves solving a high-dimensional nonlinear geodesic ODE. Instead, we propose to use the large class of *extrinsic* unitary invariant distances on *a* standardized realization subbundle to build distances on *the* LDS base space. The main benefits are that such distances are abundant, the ambient space is *not* too large (e.g., not infinite dimensional), and calculating the distance in the base space boils down to a simple optimization problem (albeit non-convex). Specifically, let $\tilde{d}_{\widetilde{\mathcal{O}\Sigma}_{m,n,p}}$ be a unitary invariant distance on a standardized realization subbundle $\widetilde{\mathcal{O}\Sigma}_{m,n,p}$ with the base $\Sigma_{m,n,p}$ (as in Theorem 1). One example of such a distance is

$$\tilde{d}^2_{\widetilde{\mathcal{O}\Sigma}_{m,n,p}}(R_1,R_2)=\lambda_A\|A_1-A_2\|_F^2+\lambda_B\|B_1-B_2\|_F^2+\lambda_C\|C_1-C_2\|_F^2+\lambda_D\|D_1-D_2\|_F^2, \tag{3}$$

where $\lambda_A, \lambda_B, \lambda_C, \lambda_D > 0$ are constants and $\|\cdot\|_F$ is the matrix Frobenius norm. A group action induced distance (called the *alignment* distance) between two LDSs $M_1, M_2 \in \Sigma_{m,n,p}$ with realizations $R_1, R_2 \in \widetilde{\mathcal{O}\Sigma}_{m,n,p}$ is found by solving the *realization alignment* problem

$$d^2_{\Sigma_{m,n,p}}(M_1, M_2) = \min_{(Q,\Theta)\in O(n)\times O(m)} \tilde{d}^2_{\widetilde{\mathcal{O}\Sigma}_{m,n,p}}\big((Q,\Theta)\bullet R_1, R_2\big). \tag{4}$$

This distance has been used for efficient video sequence classification and clustering (e.g., via defining *averages* or a k-means like algorithm) [1]. In [12] a fast algorithm is developed which (with little modification) can be used to compute this distance. The above alignment distance based on (3) is only an example. In any application, a large class of such distances can be constructed and among them a suitable one can be chosen or they can be combined in a machine learning framework (such distances may even correspond to different standardizations).

6 Conclusion

In this paper our focus was the geometrization of spaces of stochastic processes generated by LDSs of fixed size and order, for use in pattern recognition of high-dimensional time-series data (e.g., in the prototype Problem 1). We argued how existing approaches lack practicality. We then introduced a general and flexible geometrization framework, based on the quotient structure of the space of such LDSs, which leads to a large class of extrinsic distances. Our approach can be extended in various directions and that will be the subject of our future work.

Acknowledgements This work was supported by the Sloan Foundation and by grants ONR N00014-09-10084, NSF 0941362, NSF 0941463 and NSF 0931805.

References

1. B. Afsari, R. Chaudhry, A. Ravichandran, and R. Vidal. Group action induced distances for averaging and clustering linear dynamical systems with applications to the analysis of dynamic visual scenes. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.

2. S. I. Amari. Differential geometry of a parametric family of invertible linear systems-Riemannian metric, dual affine connections, and divergence. *Mathematical Systems Theory*, 20:53–82, 1987.
3. S. I. Amari and H. Nagaoka. *Methods of Information Geometry*, volume 191 of *Translations of Mathematical Monographs*. American Mathematical Society, 2000.
4. F. Barbaresco. Information geometry of covariance matrix: Cartan-Siegel homogeneous bounded domains, Mostow/Berger fibration and Frechet median. *Matrix Information Geometry*, pages 199–255, 2013.
5. B. Béjar, L. Zappella, and R. Vidal. Surgical gesture classification from video data. In *Medical Image Computing and Comp. Assisted Intervention*, pages 34–41, 2012.
6. R. Chaudhry and R. Vidal. Recognition of visual dynamical processes: Theory, kernels and experimental evaluation. Technical Report 09-01, Department of Computer Science, Johns Hopkins University, 2009.
7. K. D. Cock and B. D. Moor. Subspace angles and distances between ARMA models. *System and Control Letters*, 46(4):265–270, 2002.
8. M. Deistler, B. O. Anderson, A. Filler, C. Zinner, and W. Chen. Generalized linear dynamic factor models: An approach via singular autoregressions. *European Journal of Control*, 3:211–224, 2010.
9. B. Hanzon. *Identifiability, Recursive Identification and Spaces of Linear Dynamical Systems*, volume 63-64. Centrum voor Wiskunde en Informatica (CWI), 1989.
10. U. Helmke. Balanced realizations for linear systems: a variational approach. *SIAM Journal on Control and Optimization*, 31(1):1–15, January 1993.
11. X. Jiang, L. Ning, and T. Georgiou. Distances and Riemannian metrics for multivariate spectral densities. *IEEE Trans. on Aut. Control*, 57(7):1723–1735, 2012.
12. N. D. Jimenez, B. Afsari, and R. Vidal. Fast Jacobi-type algorithm for computing distances between linear dynamical systems. In *European Control Conference*, 2013.
13. D. Kazakos and P. Papantoni-Kazakos. Spectral distance measures between Gaussian processes. *IEEE Transactions on Automatic Control*, 25(5):950–959, 1980.
14. D. G. Kendall, D. Barden, T. K. Carne, and H. Le. *Shape and Shape Theory*. Wiley Series In Probability And Statistics. John Wiley & Sons, 1999.
15. S. Kobayashi and K. Nomizu. *Foundations of Differential Geometry Volume I*. Wiley Classics Library Edition. John Wiley & Sons, 1963.
16. P. S. Krishnaprasad and C. F. Martin. On families of systems and deformations. *International Journal of Control*, 38(5):1055–l079, 1983.
17. J. M. Lee. *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics. Springer, 2002.
18. T. W. Liao. Clustering time series data – a survey. *Pattern Recognition*, 38:1857–1874, 2005.
19. A. Martin. A metric for ARMA processes. *IEEE Trans. on Signal Processing*, 48(4):1164–1170, 2000.
20. D. Piccolo. A distance measure for classifying ARIMA models. *Journal of Time Series Analysis*, 11(2):153–164, 1990.
21. A. Ravichandran and R. Vidal. Video registration using dynamic textures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(1):158–171, 2011.
22. N. Ravishanker, E. L. Melnick, and C.-L. Tsai. Differential geometry of ARMA models. *Journal of Time Series Analysis*, 11(3):259–274, 1990.
23. Y. A. Rozanov. *Stationary random processes*. Holden-Day, 1967.
24. S. Vishwanathan, A. Smola, and R. Vidal. Binet-Cauchy kernels on dynamical systems and its application to the analysis of dynamic scenes. *International Journal of Computer Vision*, 73(1):95–119, 2007.