International Journal of Theoretical Physics, Vol. 30, No. 12, 1991

Unified Theory of Model Reduction via Gleason Measures

Erik I. Verriest[1] **and David Finkelstein**[2]

Received May 31, 1991

Previous work on the stochastic realization and approximation problem has cast this problem in the framework of the *RV*-coefficient, a measure of correlation recently introduced in the multivariate statistical literature. This allowed the introduction of a common measure for the "goodness of fit" for the different realization algorithms. This paper explores the deeper geometrical and logical foundation for this common measure in a unified theory for the data-driven and the exact covariance approaches.

1. INTRODUCTION

1.1. Scope of the Paper

In the theory of identification, signal processing, and digital filtering, a problem of fundamental importance is that of finding a finite-dimensional Markovian representation of a stochastic process from the covariance information. This problem is known as the stochastic realization problem, and has received a great deal of attention. Whenever a finite set of real data is gathered, all processing is done over finite sets, and an underlying probabilistic description is absent in most cases. As a result, covariances must be estimated by sample covariances, and a "degradation" of the theoretical realization solutions results. A more direct, data-driven approach is needed. Moreover, for many applications, the Markovian representation or state space model may be too complex, due to its high dimensionality, thus barring efficient computational management. This motivates the quest for approximate lower-order models, and the need for common measures to evaluate

[1]School of Electrical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0250.
[2]School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332-0430.

and compare different approaches. In this paper the geometry of the stochastic realization problem, both exact and approximate, is investigated, and it is linked to some notions in the theoretical foundations of quantum mechanics. More precisely, measures on the subspaces of a Hilbert space are introduced, which relate to the density matrices in the quantum mechanical context.

In Section 2, the stochastic realization problem is briefly reviewed, and cast as a problem of finding "good" subspaces. The Gleason measures are introduced in Section 3, and form the main basis for the further development. Section 4 shows that they lead to natural criteria for the evaluation of different subspaces, conditioned on some prior. As a particular application, the different stochastic realization algorithms can be unified (and evaluated).

1.2. Historical Background

Akaike (1975) has developed a stochastic realization theory based on the information interface between the past and the future of a time series and the concepts of canonical correlation analysis. The theory is further refined by Faurre (1976). The two canonical vectors are shown to be related to the states of the forward and backward innovations representation of the process. Moreover, the canonical correlation analysis provides a rational basis for obtaining reduced order models. Baram (1981) extended the results to the nonstationary case. A similar algorithm for obtaining the stochastic realization and the reduced order models, called canonical realization algorithm (CRA), was introduced by Desai and Pal (1982). They obtain forward–backward dual models with state covariances which are equal and diagonal. (These diagonal elements are the canonical correlations.) They are the stochastic counterpart of the deterministic balanced realizations introduced by Moore (1981) and extended to the time-varying case by Verriest and Kailath (1983). Finkelstein and Finkelstein (1983) studied the prediction of the output of an arbitrary automaton from its input through the Galois lattice of the transition relation.

Arun and Kung (1986) proposed the Karhunen–Loeve method (KLM) as a basis for the stochastic realization. KLM is equivalent to a principal component analysis of instrumental variables.

Ramos and Verriest (1984) unified the theory by showing that both CRA and KLM, given the exact covariances, are special cases of a more general optimization problem, using the RV-coefficient introduced by Escoufier (1973). It was shown that this common statistical measure of information provides a rationale for drawing inferences about the performance of the algorithms. Verriest (1985) explored the connection of the exact covariance and real data case further by relating the RV-coefficient to certain operators in a tensor product space $G \otimes H$, where G and H are separable Hilbert spaces. Here G is a base-space, R^p say, and H is, respectively, $L_2^p(\Omega, B, M)$ and R^N. The idea is to associate for a given element in the tensor product space (considered as a prior for the problem) an operator in H. The set of such operators has the structure of a Hilbert space under the inner product inherited from the inner product in H itself. The usual induction

$$\text{inner product} \rightarrow \text{norm} \rightarrow \text{distance}$$

leads to rigorous definition of similarity (or correlation) of subspaces $G_1 \otimes H$ and $G_2 \otimes H$.

2. THE STOCHASTIC REALIZATION PROBLEM

2.1. Problem Formulation

Given the covariance sequence $\Lambda(k)$ of a rational, stationary, zero-mean, discrete-time vector sequence $\{y_k\}$, the stochastic realization problem consists in finding a Markovian representation of the form

$$x_{k+1} = Fx_k + w_k \tag{2.1}$$

$$\bar{y}_k = Hx_k + v_k \tag{2.2}$$

where $\{w_k\}$ and $\{v_k\}$ are white Gaussian noises with

$$\forall k, l \,|\, E\begin{bmatrix} w_k \\ v_k \end{bmatrix} [w_l' \quad v_l'] = \begin{bmatrix} Q & S \\ S' & R \end{bmatrix} \delta_{kl} \tag{2.3}$$

such that $E(\bar{y}_{k+n}\bar{y}_k') = \Lambda(n)$. Here δ_{kl} is the Kronecker delta.

The solution to this problem is described by Faurre (1976). Given the covariance sequence, one forms the (infinite) Hankel matrix

$$\hat{H} = \begin{bmatrix} \Lambda(1) & \Lambda(2) & \cdots \\ \Lambda(2) & \Lambda(3) & \\ \vdots & & \end{bmatrix} \tag{2.4}$$

The time sequence is rational if and only if this Hankel matrix has finite rank (say n). It follows then from the deterministic realization theory that the order of any minimal Markovian representation of $\{y_k\}$ is precisely n, and a triple (F, G, H) can be constructed such that

$$\Lambda(k) = HF^kH^1 + \Lambda_0\delta_{k0}, \qquad k \geq 0$$

$$\Lambda(k) = \Lambda'(-k), \qquad k \leq 0 \tag{2.5}$$

where in order to have a Markovian representation, the following needs to be satisfied:

$$P - FPF' = Q \tag{2.6}$$

$$G - FPH' = S \tag{2.7}$$

$$\Lambda_0 - HPH' = R \tag{2.8}$$

$$\begin{bmatrix} Q & S \\ S' & R \end{bmatrix} \geq 0, \qquad P > 0 \tag{2.9}$$

Here P is interpreted as the state covariance matrix

$$P = E(x_k x_k') \tag{2.10}$$

The triple (F, G, H) together with Λ_0 do not uniquely specify the covariances P, Q, S, and R. However, P completely specifies Q, S, and R, and therefore characterizes the Markovian representation. Furthermore, note that any minimal realization of the covariance sequence is unique, modulo a similarity transformation.

2.2. Information Interface Between Past and Future

Assume that the stochastic time series $\{y_k\}$ is Gaussian (with zero mean). The relevant random variables are then in the Hilbert space $L_2(\Omega, \beta, P)$ and conditional expectations can be interpreted as orthogonal projections onto subspaces $L_2(\Omega, \mathscr{F}_k^{\{y_k\}}, P)$.

For the time series $\{y_k\}$ define the infinite vectors

$$y_k^+ = \begin{bmatrix} y_k \\ y_{k+1} \\ \vdots \end{bmatrix}, \qquad \text{the future}$$

$$y_k^- = \begin{bmatrix} y_{k-1} \\ y_{k-2} \\ \vdots \end{bmatrix}, \qquad \text{the past} \tag{2.11}$$

and define the semi-infinite covariance matrices

$$\hat{H} = E\{Y_k^+ (Y_k^-)'\}, \qquad R^+ = E\{Y_k^+ (Y_k^+)'\}, \\ R^- = E\{Y_k^- (Y_k^-)'\} \tag{2.12}$$

Within this representation, the forward and backward predictor subspaces are

$$X_k = \mathrm{Span}(Y_k^+ \mid Y_k^-), \qquad Z_{k-1} = \mathrm{Span}(Y_k^- \mid Y_k^+) \tag{2.13}$$

$(A \mid B)$ denotes the projection of span (A) onto the Hilbert space spanned by the components of B. These two spaces form the information interface between R^+ and R^-. Either one can be used to define a minimal Markovian representation (forward or backward). The canonical correlations lead to a natural distance measure between the past and the future, which for the Gaussian case is exactly the Kullback–Leibler information.

Alternatively, the past can be treated as the instrumental variables for predicting the future. This is the principal component approach taken by Arun and Kung (1986). Retaining the components of the past that have a significant contribution to the predictive efficiency of the future is their motivation.

Despite the nice intuitive appeal of a canonical correlation analysis, their critique on the method centers on two issues: (1) variables may be strongly correlated, and yet not extract significant portions of the variance; (2) robustness may be at stake if one deals with sample covariances rather than the exact ones.

Ramos and Verriest (1984) and Ramos (1985) resolved the two methods by putting them in a common framework, optimizing Escoufier's (1973) *RV*-coefficient under different constraints. If for random vectors X and Y

$$\mathrm{cov}(X, Y) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

then

$$RV(X, Y) = \frac{T_r(\Sigma_{12}\Sigma_{21})}{[T_r(\Sigma_{11}^2)T_r(\Sigma_{22}^2)]^{1/2}} \geq 0 \tag{2.14}$$

This measure shares many of the properties of a correlation coefficient, but is not one itself. (It is the square of the correlation if X and Y are scalar.) It also allows the computation of a "figure of merit" for each algorithm in a consistent way. Our new Gleason approach gives a natural interpretation for (2.14) (and other) measures.

3. THE LATTICE OF SUBSPACES AND GLEASON MEASURES

Let H be a Hilbert space. The set of all closed subspaces of H has the structure of an orthocomplemented complete lattice, also called a logic. (The predicate calculus of a quantum mechanical system bears some similarity to the corresponding calculus of formal logic; one refers to this often as "quantum logic.") A one-to-one correspondence exists between the lattice of all closed subspaces of H and the lattice Proj H of all orthoprojectors on H.

In his study of the mathematical foundations of quantum mechanics, Mackey posed the problem of finding all positive measures on the closed subspaces of a Hilbert space. Such a measure must have the property that for any countable collection $\{S_i\}$ of mutually orthogonal closed subspaces the mapping is σ-additive, i.e.,

$$\sum_i \mu(S_i) = \mu\left(\sum_i S_i\right) \tag{3.1}$$

A measure satisfying the above property is for instance obtained by selecting a vector v in the Hilbert space H, and for each subspace A of H defining

$$\mu_v(A) = \|P^A(v)\|^2 \tag{3.2}$$

where P^A is the projection operation on A. Clearly, finite convex combinations of such measures also satisfy the conditions for such measures, and passing to the limit, any positive-semidefinite trace-class operator T also defines such a measure via

$$\mu(A) = \mathrm{Tr}(TP^A) \tag{3.3}$$

Gleason (1957) has shown that in a separable Hilbert space of dimension at least three, every measure on the closed subspaces can be represented as above, with T a positive-definite operator of trace class. Further extensions of Gleason measures which are vector- and operator-valued, have been surveyed by Jajte (1979).

Let H be a separable Hilbert space, and Proj H the lattice of all orthogonal projectors in H. Let also E be some Banach space, then:

Definition. A mapping ξ: Proj $H \to E$ is said to be an E-valued Gleason measure if:

(1) For any sequence of pairwise orthogonal projectors $P_1, P_2, \ldots$ from Proj H,

$$\sum_i \xi P_i = \xi\left(\sum_i P_i\right) \tag{3.4}$$

the series on the left-hand side being weakly convergent.

(2) $$\sup\{\|\xi P_i\| : P_i \in \mathrm{Proj}\ H\} < \infty \tag{3.5}$$

Jajte and Paszkiewicz (1978) have shown that every Gleason measure ξ on Proj H can be extended in a unique way to a continuous operator on $L(H)$, the algebra of all bounded linear operators in H.

An important class of Gleason measures taking values in a Hilbert space are the orthogonally scattered measures (OSG). Let H and K be two Hilbert spaces, with dimension at least three.

Definition. A Gleason measure ξ: Proj $H \to K$ is said to be an orthogonally scattered measure (OSG-measure) if for any orthogonal projectors P, Q in Proj H the following implication holds:

$$P \perp Q \Rightarrow \xi P \perp \xi Q \tag{3.6}$$

Note that automatically for all P, $\|\xi P\| \leq \|\xi I\|$ is implied, where I is the identity operator, corresponding with the (sub)space H.

Any OSG-measure defines a positive Gleason measure by

$$\mu P = \|\xi P\|^2, \qquad P \in \text{Proj } H \tag{3.7}$$

By Gleason's theorem, there exists then a nonnegative self-adjoint trace-class operator T such that

$$\mu P = \text{Tr } TP, \qquad P \in \text{Proj } H \tag{3.8}$$

The above can be interpreted as a "variance"; we similarly define a "covariance" by $\text{COV}\{P, Q\} = (\xi P, \xi Q)_K$. Then for any commuting projectors P, $Q \in \text{Proj } H$ one has

$$(\xi P, \xi Q)_K = \text{Tr } TPQ \tag{3.9}$$

This formula is not true in general for arbitrary projectors in a complex Hilbert space. However, if H and K are real Hilbert spaces, then it can be shown that

$$(\xi P, \xi Q)_K = \text{Tr } TPQ = \text{Tr } TQP, \qquad \text{all } P, Q \in \text{Proj } H \tag{3.10}$$

T is given by Gleason's theorem (3.9).

4. APPLICATIONS TO REALIZATION THEORY

4.1. A Correlation Measure for Subspaces

As shown by Verriest (1986), the (exact) stochastic realization and the (real) signal modeling benefit from the use of the RV-coefficient. In the first case, the formalism is used in the comparison of random variables, while in the second it compares data matrices. We present here a geometric point of view, motivated by the observation that for the stochastic realization problem, the underlying space $L_2^P(\Omega, B, m)$ and in the real data case, the space $R^{p\times N}$ are isomorphic with the tensor product spaces, respectively,

$$L_2^P(\Omega, B, m) \sim R^p \otimes L_2(\Omega, B, m) \tag{4.1}$$

$$R^{p\times N} \sim R^p \otimes R^N \tag{4.2}$$

In general, let G and H be separable Hilbert spaces. Let $\{\phi_i\}$ be a complete orthonormal set (CONS) in G, and $\{\psi_i\}$ a CONS in H. Any vector x in the tensor product space $G\otimes H$ has a decomposition

$$x = \sum_i |x_i\rangle\langle\psi_i| \tag{4.3}$$

where $x_i \in G$. The vector x in the tensor product space will be referred to as a "prior." Introduce now the superposition of measures on Proj G induced by the prior:

$$\mu_x = \sum_i \mu_i \tag{4.4}$$

The μ_i are the Gleason measures corresponding to x_i. For all subspaces of G, it follows that

$$\mu_x(A) = \mathrm{Tr}\ T_x P^A \tag{4.5}$$

where

$$T_x = \sum_i |x_i\rangle\langle x_i| = xx' \tag{4.6}$$

is interpreted as a gramian or covariance operator.

The measure $\mu_x(A)$ gives a numerical value to the closeness of A to G, given the prior x.

The problem of finding the subspace of fixed dimension which "looks most like H from the point of view of x" is then solved by letting P^A be the projector on the eigenspace of T_x with the largest principal components. See Aragon and Couot (1976), who also stated several equivalent problems relating to the principal component analysis. Note that $\mu_x(H) = \mathrm{Tr}\ T_x$.

However, this measure does not lead to a useful definition of the correlation between subspaces. Indeed, consistent with the above "variance" μ_x we have the covariance [using (3.10)]

$$(\xi_x(A), \xi_x(B)) = \mathrm{Tr}\ T_x P^A P^B \tag{4.7}$$

But for $A \perp B$, we get $(\xi_x(A), \xi_x(B)) = 0$. There is no interface between A and B. This situation is displeasing, but can be resolved. The operator $T_x\colon G \to G$ is a characteristic for the given x in $G\otimes H$ (in fact, a "sufficient statistic"), and one can think of T (or μ) as conditioned by the vector $x \in G \otimes H$. In this sense, the extended projectors $\tilde{P}^B \in \text{Proj } G \otimes H$ defined by

$$\begin{aligned}\tilde{P}^B x &= \sum_i P^B |x_i\rangle \langle \phi_i| \\ &= \sum_i \xi_{x_i}(B) \langle \phi_i| \end{aligned} \tag{4.8}$$

yield a "coherent" addition of OSG measures, conditioned on x (i.e., a posterior measure). The posterior variance of $A \in \text{Proj } G$, given x, is then the operator from $G \to G$,

$$(\tilde{P}^A x)(\tilde{P}^A x)' = \sum_i P^A |x_i\rangle \langle x_i| P^A = P^A T_x P^A \tag{4.9}$$

and the covariance

$$(\tilde{P}^B x)(\tilde{P}^A x)' = \sum_i P^B |x_i\rangle \langle x_i| P^A = P^B T_x P^A \tag{4.10}$$

This is simply interpreted as the restriction to B of the mapping T_x restricted to the subspace A, and displays the coupling or interface between A and B given x. In order to attach a numerical value to this interface, any norm on the various restrictions $P^B T_x P^A$ can be chosen. The following natural definition follows.

Definition. The "correlation" between subspaces A and B in Proj G is

$$\rho(A, B|x) = \frac{\|P^A T_x P^B\|}{(\|P^A T_x P^A\| \, \|P^B T_x P^B\|)^{1/2}} \tag{4.11}$$

In particular, the Frobenius norm becomes

$$\rho_F(A, B|x) = \frac{\text{Tr}(P^A T_x P^B T_x)}{[\text{Tr}(P^A T_x)^2 \, \text{Tr}(P^B T_x)^2]^{1/2}} \tag{4.12}$$

The spectral norm gives

$$\rho_2(A, B|x) = \frac{\lambda_{\max}(P^A T_x P^B T_x)}{\lambda_{\max}(P^A T_x) \lambda_{\max}(P^B T_x)} \tag{4.13}$$

For the special case of N observed p-vectors x_i, let $G = R^p$ and $H = R^N$, so that the data are organized in the data matrix X. Then $T_x = XX'$ is the sample covariance matrix S, and the correlation between the complementary subspaces span$(\phi_1, \ldots, \phi_k)$ and span$(\phi_{k+1}, \ldots, \phi_p)$ is (for an obvious partition of S)

$$\rho_F = \frac{\operatorname{Tr}(S_{12}S_{21})}{(\operatorname{Tr} S_{11}^2 \operatorname{Tr} S_{22}^2)^{1/2}} \tag{4.14}$$

which is the RV-coefficient. Note that then also

$$\rho_F(A, H|x) = \frac{\operatorname{Tr} S^2 P^A}{(\operatorname{Tr} S^2 \operatorname{Tr} S^2 P^A)^{1/2}} = \left(\frac{\operatorname{Tr} S^2 P^A}{\operatorname{Tr} S^2}\right)^{1/2} \tag{4.15}$$

which again yields the principal component analysis, since A maximizes $\rho_F(A, H|x)$ iff A maximizes $\mu_x(A)$.

Let L be a bounded operator in G. The induced transformation $\tilde{L}$ in $G \otimes H$ is

$$\begin{aligned} \tilde{L}x &= \tilde{L}\left(\sum_i |x_i\rangle \langle \phi_i|\right) \\ &= \sum_i |Lx_i\rangle \langle \phi_i| \end{aligned} \tag{4.16}$$

yielding the transformation rule for the Gleason measure. For all $A \in \operatorname{Proj} G$

$$\mu_{\tilde{L}x}(A) = \operatorname{Tr} T_{\tilde{L}x} P^A = \operatorname{Tr} LT_x L' P^A \tag{4.17}$$

Let K be the fixed subspace span$(\phi_1, \ldots, \phi_k)$ of G. Letting further

$$L = MP^K + NP^{K^\perp} \tag{4.18}$$

then the principal component analysis and canonical correlation analysis are respectively (in Frobenius norm) [$O(K)$ is the set of bounded operators in K].

$$\text{PCA:} \quad \max_{M \in O(K)} \frac{\operatorname{Tr}(MT_{12}T_{21}M')}{[\operatorname{Tr}(MT_{11}M')^2 \operatorname{Tr}(T_{22})^2]^{1/2}} \tag{4.19}$$

$$\text{CCA:} \quad \max_{\substack{M \in O(K) \\ N \in O(K^\perp)}} \frac{\operatorname{Tr}(MT_{12}NN'T_{21}M')}{[\operatorname{Tr}(MT_{11}M')^2 \operatorname{Tr}(NT_{22}N')^2]^{1/2}} \tag{4.20}$$

which are the formulas obtained by Robert and Escoufier (1976). If $G \otimes H = L_p^2(\Omega, B, m)$ or R^{pN}, then T is respectively the covariance Σ or sample covariance S.

4.2. Data-Driven Stochastic Realization Solution

Assuming that a data stream $\{y_k, |k| \leq N\}$ is observed, a data matrix Y can be formed by considering $(Y_{-N}, 0, \ldots, 0)'$,

$(Y_{-N+1}, Y_{-N}, 0, \ldots, 0)', \ldots, (Y_N, Y_{N-1}, \ldots, Y_{-N})', \ldots, (0, \ldots, 0, Y_N)'$ as consecutive samples of the vector in $G=R^{2N+1}$ (the "pure states"). In order to avoid the nasty end effects due to the substitution of zeros where data are missing, a linear superposition of these states, weighted by the sequence $\{q_j \geq 0;\ j=1, \ldots, 4N+1\}$, may be used. Let $Q=\text{diag}\{q_i\}$. The Gleason operator is $T_{(Y,q)}=YQY'$. The "past" is span$\{|\phi_{-N}\rangle \cdots |\phi_{-1}\rangle\}$, and the future span$\{|\phi_0\rangle \cdots |\phi_N\rangle\}$. A recursive realization algorithm, which optimally uses all the data (in real time), would necessarily involve the update of $T_{(Y,q)_N}$. (Strictly speaking, its dimension is increasing also, but the operator can be trivially extended to one acting in some "big" space.) This is obvious since the new datum creates a new "sample vector" so that the weights q_i will have to be recalculated for the linear superposition. Another difficulty lies in the fact that *all* samples that were obtained by shifts now have the new observation attached to it (i.e., where before a "zero" occurred). This is avoided if a windowed (hence, nonoptimal) recursion is used.

5. CONCLUSIONS

By determining well-motivated measures for the correlation between subspaces of a Hilbert space, based on the available prior information, it was possible to unify several tools from multivariate analysis, and introduce common measures for their evaluation. We have only discussed the principal component and the canonical correlation analyses. Discriminant analysis and a rational way for discarding variables can also be treated. Our inspiration for this work came from the desire to better motivate and explain the use of the *RV*-statistic in these problems. In particular, exact realization theory and its signal processing counterpart (i.e., the real data case) are unified. The fact that the data should come first looks natural from this viewpoint. Deterministic modeling, cluster analysis (in pattern recognition), and quantization of random fields are other applications of our abstract framework. The variation lies in the choice of the spaces G and H, and the constraints that are natural for the problem.

REFERENCES

Akaike, A. (1975). *SIAM Journal of Control*, **13**, 162–173.

Aragon, Y., and Couot, J. (1976). *Comptes Rendus A*, **283**, 867–869.

Arun, K. S., and Kung, S. Y. (1986). Generalized principal components analysis, and its application in approximate stochastic realization, in *Theory and Practice of Modeling Stochastic Processes*, U. B. Desai, ed., Kluwer, Hingham, Massachusetts, Chapter 4, pp. 75–104.

Baram, Y. (1981). *IEEE Transactions on Automatic Control*, **AC-26**, 1225–1231.

Desai, U. B., and Pal, D. (1982). A realization approach to stochastic model reduction and balanced stochastic realizations, in *Proceedings of the 21st Conference on Decision and Control*, pp. 1105–1111.

Escoufier, Y. (1973). *Biometrics*, **29**, 751–760.

Faurre, P. L. (1976). Stochastic realization algorithms, in *System Identification: Advances and Case Studies*, R. K. Mehra and D. G. Lainiotis, eds., Academic Press, New York.

Finkelstein, D., and Finkelstein, S. R. (1983). *International Journal of Theoretical Physics*, **22**, 753–779.

Gleason, A. M. (1957). *Journal of Mathematics and Mechanics*, **6**, 885–893.

Jajte, R. (1979). Gleason measures, in *Probabilistic Analysis and Related Topics*, Vol. 2, Bharucha-Reid, ed., Academic Press, New York.

Jajte, J., and Paszkiewicz, A. (1978). *Studia Mathematica*, **58**, 229–251.

Moore, B. C. (1981). *IEEE Transactions on Automatic Control*, **AC-26**, 17–32.

Ramos, J. A. (1985). A stochastic approach to streamflow modeling, Ph.D. Dissertation, School of Civil Engineering, Georgia Institute of Technology.

Ramos, J. A., and Verriest, E. I. (1984). A unifying tool for comparing stochastic realization algorithms and model reduction techniques, in *Proceedings 1984 ACC*, San Diego, California.

Robert, P., and Escoufier, Y. (1976). *Applied Statistics*, **25**, 257–265.

Verriest, E. I. (1986). Projection techniques for model reduction, in *Modeling, Identification, and Robust Control*, A. Lindquist and C. Byrnes, eds., North Holland, Elsevier, Amsterdam, the Netherlands, pp. 381–396.

Verriest, E. I. (1986). Model reduction via balancing, and connections with other methods, in *Theory and Practice of Modeling Stochastic Processes*, U. B. Desai, ed., Kluwer, Hingham, Massachusetts, Chapter 6, pp. 123–154.

Verriest, E. I., and Kailath, T. (1983). *IEEE Transactions on Automatic Control*, **AC-28**, 833–844.