GEORGIA INSTITUTE OF TECHNOLOGY

OFFICE OF CONTRACT ADMINISTRATION

PROJECT ADMINISTRATION DATA SHEET

[X] ORIGINAL

[ ] REVISION NO. ____

Project No. E-21-660

DATE: 4/15/81

Project Director: Erik I. Verriest

School/Lab Electrical Engineering

Sponsor: National Science Foundation, Washington, D. C. 20550

Type Agreement: Grant No. ECS-8105509

Award Period: From 7/1/81 To 12/31/83* (Performance) - - - - - - - (Reports)

Sponsor Amount: $45,163

Contracted through:

Cost Sharing: $6,768 (E-21-364)

GTRI/XXX

Title: "Research Initiation: On Generalized Balanced Realizations and Applications to Model Reduction"

ADMINISTRATIVE DATA

OCA CONTACT Leamon R. Scott, Ext. 4820

1) Sponsor Technical Contact: Abraham H. Haddad, NSF Program Official; National Science Foundation; Washington, D. C. 20550 Phone: 202-357-9618

2) Sponsor Admin./Contractual Contact: Lee A. DeHerrera, NSF Grants Official; National Science Foundation; Washington, D. C. 20550 Phone: 202-357-9602

Reports: See Deliverable Schedule

Security Classification: N/A

Defense Priority Rating: N/A

RESTRICTIONS

See Attached NSF Supplemental Information Sheet for Additional Requirements.

Travel: Foreign travel must have prior approval - Contact OCA in each case. Domestic travel requires sponsor approval where total will exceed greater of $500 or 125% of approved proposal budget category.

Equipment: Title vests with Georgia Institute of Technology

COMMENTS: *Includes the usual six (6) month unfunded flexibility period.

COPIES TO:

Administrative Coordinator
Research Property Management
Accounting Office
Procurement Office

Research Security Services
Reports Coordinator (OCA)
Legal Services (OCA)
Library, Technical Reports

EES Information Office (2)
Project File (OCA)
Other: ____

GEORGIA INSTITUTE OF TECHNOLOGY OFFICE OF CONTRACT ADMINISTRATION

52-869/07

SPONSORED PROJECT TERMINATION/CLOSEOUT SHEET

Date 8/31/84

Project No. E-21-660 School/~~Lab~~ EE

Includes Subproject No.(s)

Project Director(s) Erik I. Verriest GTRI ~~XXXX~~

Sponsor National Science Foundation, Washington, D.C. 20550

Title "On Generalized Balanced Realizations and Applications to Model Reduction"

Effective Completion Date: 12/31/83 (Performance) 3/31/84 (Reports)

Grant/Contract Closeout Actions Remaining:

- [ ] None
- [X] Final ~~Invoice or Final Fiscal Report~~ FCTR
- [ ] Closing Documents
- [X] Final Report of Inventions if positive
- [X] Govt. Property Inventory & Related Certificate if positive
- [ ] Classified Material Certificate
- [ ] Other

Continues Project No. Continued by Project No.

COPIES TO:

Project Director
Research Administrative Network
Research Property Management
Accounting
Procurement/EES Supply Services
Research Security Services
Reports Coordinator (OCA)
Legal Services

Library
GTRI
Research Communications (2)
Project File
Other I. Newton

Form OCA 60:1028

E-21-660

GEORGIA INSTITUTE OF TECHNOLOGY
SCHOOL OF ELECTRICAL ENGINEERING
ATLANTA, GEORGIA 30332

ELEPHONE: (404) 894-2949

September 24, 1982

Dr. Abraham H. Haddad
Engineering, System Theory and Application
National Science Foundation
Washington, DC 20550

Subject: Annual Progress Report for NSF Grant No. ECS-8105509, "On Generalized Balanced Realizations and Applications to Model Reduction" (covering 1 July 1981 to 31 December 1983).

Dear Dr. Haddad:

Significant progress has been made during the first year of the above research grant.

An investigation of the special properties of the "classical" balanced realizations (i.e. for time invariant stable systems) has lead to a simplified characterization of these realizations. An exhaustive analysis of the second order case gave some insight in the relation between order reduction based on dominance and order reduction via the balanced realization technique.

Further results have been obtained in the extension of the balanced realization concept to time varying systems. These results are reported in:

E. I. Verriest and T. Kailath: "On Generalized Balanced Realizations," to be published in the June 1983 issue of the IEEE Trans. on Automatic Control.

In order to practically test the usefulness of the proposed model reduction for time varying systems, a software package is being developed, as part of a graduate project.

An important new direction of the research is the application of the balancing concept to the closed-loop system. This results in what we called the LQG-balanced realizations and is based on the underlying Riccati equations rather than the (open loop) Lyapunov equations. This is a crucial advance, however many problems remain to be investigated (e.g. stability of the reduced order model). These results were reported in:

E. I. Verriest: "Suboptimal LQG-Design via Balanced Realizations," Proceedings of the 20th IEEE Conference on Decision and Control, pp. 686-687, San Diego, CA, December 1981.

Also for the time invariant case an LQG-reduction program is being developed.

Important new insight in the fidelity of discrete applications of continuous systems was gained by using concepts of balanced realizations. Applications to the design of digital filters were obtained as well. In particular it was shown that with the "usual" discretization procedures input and output properties were not conserved. Our method yields a fidelity up to second order in the stepsize. These results are reported in:

E. I. Verriest: "Reachability-Observability and Discretization," accepted for the 21st IEEE Conference on Decision and Control," Orlando, FL, December 1982. This paper is also under review for publication in the Transactions on Automatic Control.

E. I. Verriest: "Digital Filter Design based on a High Fidelity Discretization Procedure," submitted to the 1983 IEEE, ICASSP Conference, Boston, MA, April, 1983.

A study of the applicability of model reduction (via balanced realizations) for infinite dimensional systems has begun. So far only a simple parabolic system has been considered. The method starts from a large (but finite) dimensional approximation. The finite element method and the expansion in orthogonal functions are compared.

We are continuing our efforts on several fronts: distributed systems (parabolic, elliptic and hyperbolic), delay differential systems, the relation to model reduction methods based on singular perturbation, applications to stochastic modeling, and finally model reduction of nonlinear systems, using concepts of differential geometry to obtain a local linear state space.

If additional information is needed, please contact us and we will supply it to you. Your support is greatly appreciated. Thanking you,

Sincerely,

Erik I. Verriest

EIV:krd

enclosures: copies of all above cited papers.

TA6 - 10:45

SUBOPTIMAL LQG-DESIGN VIA BALANCED REALIZATIONS

Erik I. Verriest

School of Electrical Engineering
Georgia Institute of Technology
Atlanta, GA. 30332

Abstract

A new low sensitivity realization for the quadratic regulator is derived. The main idea is the generalization of the "balanced realizations" of Moore via the separation principle. The results lead directly to a new approximation method for the LQG problem and reduced order regulator synthesis.

Introduction

The solution to the LQG-problem is well known to fall apart into a deterministic controller and a stochastic observer synthesis. The internal structure of the combined estimator-controller is immaterial however. Hence, this freedom can be exploited to obtain a realization which minimizes the sensitivity, taking the overall stochastic dynamics and the performance index into account. Skelton [2] suggested a weighting with respect to the "component costs." In stochastic context such is undesirable since it may lead to a certain imbalancedness. If the uncertainty associated with a dynamical element, with low cost contribution, is high, then the actual cost contribution in a sample process can be quite different from the estimated (low) cost. This motivates a balancing with respect to the optimal-deterministic controller and the stochastic observer via the separation principle.

To fix the ideas, consider the stochastic system

$$\begin{aligned} \dot{x} &= Fx + Gu + w \\ y &= Hx + v \end{aligned} \quad (1)$$

and $Ex_0 = 0$, $Ex_0x_0' = P_0$, where w and v are independent, zero mean gaussian with covariances Q and R. Let the goal be the minimization of the performance index

$$J = \frac{1}{2} E \int_{t_0}^{t_f} (x'Ax + u'Bu)dt + \frac{1}{2} Ex'(t_f)S_f\, x(t_f) \quad (2)$$

The deterministic optimal closed loop system has the dynamics:

$$\dot{x} = (F - GC)x \quad (3)$$

$$C = B^{-1}G'S \quad (4)$$

$$\dot{S} = -S(F-GC) - (F-GC)'S - A - C'BC;\ S(t_f) = S_f \quad (5)$$

The solution $S(t;\ t_f,\ S_f)$ of (5) has an interpretation as a weighting matrix for the minimum "cost-to-go" from the state x(t) at time t. Note that for $S_f = 0$, S is exactly the observability gramian of the closed loop system with fictitious output

$$y = Lx,\quad L' = [A^{1/2} \quad C'B^{1/2}] \quad (6)$$

The performance index is then the output energy of this system. Similarly, the filter error dynamics are given by

$$\dot{\tilde{x}} = (F-KH)\tilde{x} + M\varepsilon\ ;\quad \tilde{x} = x - \hat{x} \quad (7)$$

$$M = [Q^{1/2} \quad K'R^{1/2}] \quad (8)$$

where

$$K = PH'R^{-1}\ ,\quad P = E(\tilde{x}\tilde{x}') \quad (9)$$

$$\dot{P} = (F-KH)P + P(F-KH)' + Q + KRK'\ ;\ P(t_0) = P_0 \quad (10)$$

and ε is a noise vector of unit variance. Again, $P(t;\ t_0,\ P_0)$ is a measure for the uncertainty in the closed loop system (7), and characterizes the "disturbability" by the noise. Our goal is now to find a new coordinate system in which the components contributing a lot to the performance index, at the same time have the largest uncertainty about them. Hence it follows that one should balance with respect to the solutions P and S of the Riccati equations, rather than the open loop gramians [1].

The LQG-Balanced Realization

In [5], a state-space transformation was derived to obtain the LQG-balanced realization for the (steady state) regulator. This can easily be generalized to the non-stationary case. We summarize.

Theorem 1: Given the system (1)*, which is observable and reachable in (t_0, t_f), and the performance index (ε) then there exists a similarity transformation such that in the new coordinates $\bar{P}$ and $\bar{S}$ are equal and diagonal. (= Π)

We shall refer to Π as the "canonical riccatian." Its elements are the eigenvalues of PS (computed in

*All parameters are real analytic functions of time.

any realization). Since balancing is primarily based on the singular value decomposition, the numerical properties of it are shared, hence the insensitivity of this realization with respect to the design parameters. The different types of balancing [3,4] can then be taken advantage of for robust LQG-design. In Fixed-Interval-Balancing (FIB) the interval (t_0, t_f) is fixed. This leads to robust terminal controller design.

Theorem 2: If $(F,G,H;Q,R;A,B)$ is FIB-LQG-balanced over $[t_0,t_f]$ with canonical riccatian Π, then the cost functional [2] evaluates to

$$J = \frac{1}{2}\sum_i \pi_i(t_0)\, Ex_i(t_0)^2 + \int_{t_0}^{t_f}\sum_i \pi_i\left[\left(\frac{A+Q}{2}\right)_{ii} + \pi_i F_{ii}\right]dt \tag{11}$$

In Infinite-Interval-Balancing (IIB), the cost rate is minimized ([2] is infinite). The technique is suitable for regulator design as shown in [5]. Theorem 2 can be modified accordingly, the optimal costrate (for stationary systems) being the integrand in (11). Sliding-Interval-Balancing (SIB) can be used to design robust suboptimal controllers (for stabilizing purposes) [6]. At time $t_0 = t$, one considers the performance index (2) with the receding horizon $t_f = t + T$ (T fixed), and computes the instantaneous gain at t from $S(t, t+T, S_f)$. Its dual Kalman filter is similarly computed from the approaching horizon estimation, with gain computed from $P(t, t-T, P_0)$. Balancing is in the sense that $\Pi(t) = \bar{S}(t, t+T, S_f) = \bar{P}(t, t-T, P_0)$ P_0 and S_f are design parameters. Remark also that for stationary systems SIB(T) leads to a stationary balanced realization.

LQG-MODEL APPROXIMATION

Once the system is brought in LQG-balanced form, the relative importance of each "axis" of the coordinate system is displayed. A large value on the diagonal of Π indicates that both the cost contribution and the uncertainty are high for the corresponding state component. Hence, partition the canonical riccation into a part containing large and a part containing small elements (according to some norm) and consider the consistent partition on the system

$$x' = [x_1' \quad x_2']$$

$$\bar{F} = \begin{bmatrix} F_1 & F_{12} \\ F_{21} & F_2 \end{bmatrix}, \quad \bar{G} = \begin{bmatrix} G_1 \\ G_2 \end{bmatrix}, \quad \bar{H} = [H_1 \quad H_2]$$
$$\bar{A} = \begin{bmatrix} A_1 & A_{12} \\ A_{12} & A_2 \end{bmatrix}, \quad Q = \begin{bmatrix} Q_1 & Q_{12} \\ Q_{12} & Q_2 \end{bmatrix}, \quad \Pi = \begin{bmatrix} \Pi_1 & \\ & \Pi_2 \end{bmatrix} \tag{12}$$

As motivated by Skelton [2], we can use this information for (i) Simplification of the original model by leaving out the subspaces for which the components are fairly well certain and at the same time do not contribute much to the overall performance index; (ii) The design of reduced order regulators.

For FIB and IIB the following results hold.

Theorem 3: If the LQG-balanced system is partitioned as in (11), then the reduced order model $\{F_1,G_1,H_1,A_1,B,Q_1,R\}$ is also LQG-balanced with the canonical riccation Π_1.

Corollary: The optimal regulator for the reduced order system of theorem 3, is given by the subsystem of the consistently partitioned optimal regulator for the full system.

The proofs follow simply from the partitioning in the riccati equations. In particular

$$\dot{\Pi}_1 = \Pi_1 F_1' + F_1\Pi_1 - \Pi_1 H_1' R^{-1} H_1 \Pi_1 + Q_1 \tag{13}$$

$$\dot{\Pi}_1 = -\Pi_1 F_1 - F_1'\Pi_1 - \Pi_1 G_1 B^{-1} G_1'\Pi_1 + A_1 \tag{14}$$

As a consequence the closed loop subsystem is also asymptotically stable. The corresponding properties for SIB are not so direct since (13) and (14) fail to hold.

SUBOPTIMAL CONTROL

The optimal control for the subsystem in the previous section can be used as a suboptimal reduced order control for the full order system. The overall system, in terms of the $(2n_1 + n_2)$ dimensional state vector $X' = [\tilde{x}_1', \hat{x}_1', x_2']$ satisfies

$$\dot{X} = \begin{bmatrix} F_1 - K_1H_1 & 0 & F_{12}-K_1H_2 \\ K_1H_1 & F_1 - G_1C_1 & K_1H_2 \\ F_{21} & F_{21}-G_2C_1 & F_2 \end{bmatrix} X + \begin{bmatrix} I & 0 & -K_1 \\ 0 & 0 & K_1 \\ 0 & I & 0 \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \\ v \end{bmatrix} \tag{15}$$

Let P be the variance of X partitioned accordingly as

$$P = \begin{bmatrix} \Omega & M & N \\ M' & \Sigma & T \\ N' & T' & \Lambda \end{bmatrix} \tag{16}$$

It satisfies a Lyapunov equation, derived from (15) which must be solved to evaluate the performance (2). In [5] the costrate for stable stationary systems (15) is evaluated.

REFERENCES

[1] B. C. Moore, "Singular Value Analysis of Linear Systems. Pt. II: Controllability, Observability and Model Reduction," Systems Cont. Report No. 7802, Dept. of EE, U. of Toronto, Toronto, 1978.
[2] R. E. Skelton, "Component Cost Analysis of Linear Systems," 19th IEEE Conf. on Decision & Control, Albuquerque, N.M., Dec. 80.
[3] E. I. Verriest and T. Kailath, "On Generalized Balanced Realizations," Proc. of 19th IEEE CDC-Conference, Albuquerque, N.M., Dec. 80, pp. 504-505.
[4] E. I. Verriest, "On Balanced Realizations for Time Variant Linear Systems," PhD Dissertation, Dept. of EE, Stanford Univ., 1980.
[5] E. I. Verriest, "Low Sensitivity Design and Optimal Order Reduction for the LQG-Problem," 18th Symp. on Circuits & Systems, Jn. 81, Albuquerque, N.M.
[6] W. H. Kwon, "Linear Quadratic Gaussian Problems with Moving Finite Horizons." Preprint.

DIGITAL FILTER DESIGN BASED ON A HIGH FIDELITY DISCRETIZATION PROCEDURE

E.I. Verriest

Assistant Professor
Electrical Engineering
Georgia Institute of Technology
Atlanta, GA 30332
(404) 894-2949

ABSTRACT

This paper is concerned with the design of digital filters. The design uses a new high fidelity discrete approximation developed recently by the author [1],[2], of a given continuous time system. The new fidelity criterion is a measure for the "consideration" of the input-to-state and state-to-output mapping properties in the discretization. Furthermore the structure of the balanced relatizations [4] is exploited in order to lower the sensitivity.

Some design applications for discrete systems start from an intrinsic continuous time system. For instance a commonly used technique in the design of digital filters is based on the discretization of an equivalent continuous transfer function. Another example is the digital control design of an analog system or plant modeled by a set of dynamical equations in continuous time. Here the system model is first discretized and the estimation and control algorithms are then designed according to this discrete-time description [5].

Usually, the direct form (canonical) realizations are used because the state update requires at most $n-1$ multiplications for an n-th order filter. It was shown in [3] that these realizations do not have the best properties with regards to the effects of finite word length. These effects are minimized in the "balanced realization" [4] for which the state update requires at most n^2+n multiplications. However it can be shown that these realizations have a much higher fidelity towards the original continuous time specification. Hence a larger stepsize (Δ_b say) can be allowed, than one would require of the canonical form realization with the same fidelity. (Let this stepsize be Δ_c.) This tradeoff can lead to an increase in stepsize (sampling period) and hence a smaller number of multiplications per time unit than for the direct form realizations with the same fidelity follows.

High fidelity is achieved if the eigenvalues of the reachability and observability gramians of the discrete approximations are "close" to the corresponding eigenvalues of the continuous-system reachability and observability gramians. Hence the original transfer function is first realized in balanced form, next a stepsize Δ is chosen which is compatible with the expected signal spectrum and the discretization described in [1] is performed.

The procedure is tested in computer simulations and compared with the conventional methods [5].

REFERENCES

[1] E.I. Verriest, "Reachability, Observability and Discretization," 21st IEEE Conference on Decision and Control, Orlando, FL, December 1982.

[2] E.I. Verriest, "Reachability, Observability and Discretization," submitted to *IEEE Trans. AC,* 1982.

[3] C.T. Mullis and R.A. Roberts, "Synthesis of Minimum Roundoff Noise Fixed Point Digital Filters," *IEEE Trans. CAS,* Vol. CAS-23, No. 9, pp. 551-562, 1976.

[4] B.C. Moore, "Principal Component Analysis in Linear Systems: Controllability, Observability and Model Reduction," *IEEE Trans. AC,* Vol. AC-26, No. 1, Feb. 1981, pp. 17-32.

[5] G.F. Franklin and J.D. Powell, *Digital Control of Dynamic Systems,* Addison-Wesley, 1980.

On Generalized Balanced Realizations*

E. I. Verriest† and T. Kailath‡

Abstract

The class of analytic time varying linear systems is considered. Different balanced realizations of such systems are defined, and their existence and properties are analyzed. The results are then used to derive reduced order approximations (also for unstable systems). A method is suggested to determine the order of a "good" approximation.

* This work was supported in part by the National Science Foundation under Grant ECS78-10003 and Grant ECS-81-05509, and by the Air Force Office of Scientific Research, Air Force Systems Command under Contract AF49-620-79-C-0058.

† Formerly at Stanford University, now with the School of Electrical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

‡ Information Systems Laboratory, Stanford University, Stanford, California 94305.

I. INTRODUCTION

In recent years, a new "canonical" realization for finite dimensional linear time-invariant systems (A, B, C) was introduced by B. C. Moore [2]; these realizations exhibit certain symmetries (balance) between the input and the output maps of the realization, and are called "balanced realizations." The input and output maps of a system are characterized in a direct quantitative way by the controllability and reachability gramians, and the observability and constructibility gramians respectively. Deterministic as well as stochastic interpretations can be given to these gramians. This gramian formalism lends itself to include time-variant systems as well, as was first shown by Silverman et al. [3], for the special class of uniform realizations as defined in [16].

The applicability of the stationary balanced realizations to robust design problems was successfully shown by Mullis and Roberts [9] in the context of digital filtering (although the realizations were not in the "final" form discussed herein and in the following references). Moore [1], [2] and then Pernebo and Silverman [11] and Kung [23] showed their use in model reduction problems. Kung relates them further to some identification methods. This success motivated one study of balanced realizations for an important class of time-variant systems.

The main contributions of this paper lie in the existence proofs of balanced realizations for the class of analytic systems [4] (which can be used to approximate almost all systems). Moreover, the conditions can be given in terms of the system parameters, which is a major advantage over other known results in the time variant case ([3], [24], [25]). We remark that the class of analytic systems is neither included in, nor does it include, the class of uniform realizations. The model reduction techniques are thereby also extended to these systems. Further, where previously the reduction was restricted to stable realizations, we show the feasability of reducing unstable systems. We also discuss a method to approximate the fast or the slow behavior of a system.

The rest of the paper is organized as follows. In section 2, the connection of various gramians to the input and output maps of a system is briefly reviewed, and their transformation properties and subsequent uses as "balancing" pairs are discussed. In section 3, a pointwise (w.r.t. time) balancing transformation is shown to exist as a Lyapunov transformation. Sections 4, 5, and 6 describe some special balanced realizations of particular interest: the so-called fixed-interval, infinite-interval, and sliding-interval balanced realizations. Each has its own domain of applicability. Applications of balanced realizations to model order reduction are discussed in section 7; in fact one of the major extensions is the availability of reduced models for unstable systems. This paper is an extended version of the conference paper [21] and is based on the dissertation [4].

2. PRELIMINARY CONCEPTS

We consider the time varying system

$$\begin{aligned} \dot{x}(t) &= A(t)\, x(t) + B(t)\, u(t) \;; \quad x(t_0) = x_0 \\ y(t) &= C(t)\, x(t) \end{aligned} \tag{2.1}$$

and define the following Gramians ($\Phi(\cdot,\cdot)$ being the transition matrix of $A(\cdot)$) :

$$C\,[t_0,t_f] = \int_{t_0}^{t_f} \Phi(t_f,t)\, B(t)\, B'(t)\, \Phi'(t_f,t)dt \tag{2.2}$$

$$\bar{C}\,[t_0,t_f] = \int_{t_0}^{t_f} \Phi(t_0,t)\, B(t)\, B'(t)\, \Phi'(t_0,t)dt \tag{2.3}$$

$$O\,[t_0,t_f] = \int_{t_0}^{t_f} \Phi'(t,t_0)\, C'(t)\, C(t)\, \Phi(t,t_0)dt \tag{2.4}$$

$$\bar{O}\,[t_0,t_f] = \int_{t_0}^{t_f} \Phi'(t,t_f)\, C'(t)\, C(t)\, \Phi(t,t_f)dt \tag{2.5}$$

They are known, respectively, as the reachability, controllability, observability and constructability gramians, see e.g. [12]. This nomenclature is inherited from the diverse input-state and state-output maps, as the gramians are directly related to these maps. We illustrate this below for the reachability and the observability maps.

Suppose a system is in the zero state at time t_0 and an input function $u(t)$ belonging to a certain class of permissible functions is applied. In our case we shall focus on inputs with finite energy in the interval $[t_0,t_f]$ where $t_f > t_0$. This class of input functions forms the Hilbert space $L_2^m\,[t_0,t_f]$ of square integrable m-component vector functions. The set $X_r(t_f)$ of reachable states at time t_f , is then

$$X_r(t_f) = \{ \int_{t_0}^{t_f} \Phi(t_f,t)B(t)u(t)dt \mid u(\cdot)\varepsilon\, L_2^m\,[t_0,t_f]\}$$

Clearly $X_r(t_f)$ is a subset of $\mathbb{R}^n$. The reachability map L is

$$L: \; L_2^m\,[t_0,\, t_f] \rightarrow \mathbb{R}^n$$

$$L(u(\cdot)) = \int_{t_0}^{t_f} \Phi(t_f,t)\; B(t)\; u(t)\; dt \tag{2.6}$$

The inner product of $v(\cdot)$ and $w(\cdot)$ in $L_2^m\,[t_0,t_f]$ is defined as

$$\langle v(\cdot),\, w(\cdot)\rangle = \int_{t_0}^{t_f} v(t)'\; w(t)\; dt$$

while the inner product of x and y in $\mathbb{R}^n$ is as usual

$$\langle x,y\rangle = x'y$$

It is well known that the minimum norm solution to $L(u(\cdot)) = x$ is given by $u = L^*(z)$ where z is any solution of $LL^*(z) = x$ (see [22] pp. 161-163). The operator LL^* maps $\mathbb{R}^n$ into $\mathbb{R}^n$, and is exactly the above defined reachability gramian (2.2). If the system is completely reachably in $[t_0,t_f]$, or equivalently if L is surjective, then $LL^* = C[t_0,t_f]$ is invertible, and the required minimum "energy" is the L_2-norm

$$||u|| = ||x||_{(LL^*)^{-1}} = (x'(LL^*)^{-1}x)^{1/2} \tag{2.7}$$

The connection of $\overline{C}$ to the controllability map is analogous.

Reachability and controllability of an event (x,t) can be viewed as symmetric concepts with respect to t. While reachability of (x,t) involves driving the state from zero to z using a "past" input (i.e. $u(\theta)$, $\theta \le t$), controllability of the event (x,t) involves driving the state to zero, using a "future" input (i.e. $u(\theta)$, $\theta \ge t$). For more details, see [20]. Similarly observability of an event (x,t) is the ability to detect the "occurrence" of an event by scrutinizing the output after t, while the notion of constructability deals with the ability to construct or simulate the state of the system on line (i.e. towards the future).

The observability map

$$\begin{aligned} M:\quad & \mathbb{R}^n \to L_2^P\,[t_0,t_f] \\ & M(x) = C(t)\,\Phi(t,t_0)x \end{aligned} \tag{2.8}$$

and its gramian are related through

$$\mathcal{O}[t_0,t_f] = M^*M$$

where again M^* is the adjoint transformation of M.

The significance stems from the dual problem of finding the state $x \in \mathbb{R}^n$ which minimizes $||y - Mx||$ for $y(\cdot) \in L_2^P\,[t_0,t_f]$. If the solution is completely observable, the solution is ([22] p. 160).

$$x = (M^*M)^{-1}\,M^*y$$

If a "signal plus noise" model

$$y = Mx_0 + v$$

is given for y, where $v(\cdot)$ is a zero mean, unit covariance white noise, then the observation error y-Mx has covariance $(M^*M)^{-1}$. Again there is an analogous relation between the constructability gramian and map.

In several problems, in particular the optimization problems, the "adjoint system" of $(A(\cdot), B(\cdot), C(\cdot))$ arises. This is the realization $(-A'(\cdot), C'(\cdot), B'(\cdot))$, defined in reverse time ([12] section 9.3).

$$\begin{cases} \dfrac{d\lambda(t)}{dt} = -A'(t)\,\lambda(t) + C'(t)\,\mu(t) \quad ; \quad \lambda(t_f) = \lambda_f \\ u(t) \;\; = B'(t)\,\lambda(t) \qquad\qquad\qquad ; \quad t_0 \le t \le t_f \end{cases} \tag{2.9}$$

Because this backward evolution is prohibitive in an actual simulation, the "modified adjoint system" is introduced. It is defined by

$$\begin{cases} \dfrac{d\tilde{\lambda}(t)}{dt} = A'(t_f - t)\,\tilde{\lambda}(t) + C'(t_f - t)\,\tilde{\mu}(t) \quad ; \quad \tilde{\lambda}(0) = \tilde{\lambda}_f \\ \tilde{y}(t) \;\; = B'(t_f - t)\,\tilde{\lambda}(t) \qquad\qquad\qquad 0 \le t \le t_f - t_0 \end{cases} \tag{2.10}$$

It is a simple matter [4] to show that the gramians of the adjoint system and the modified adjoint system bear the following relationship. To obtain symmetry of the arguments for the modified adjoints, the initialization $t_0 = 0$ is taken.

Adjoint Relationship

$$\begin{tikzcd} C[t_0,t_f] \arrow[dr, leftrightarrow] & \bar{C}[t_0,t_f] \arrow[dl, leftrightarrow] \\ O[t_0,t_f] & \bar{O}[t_0,t_f] \end{tikzcd}$$

Modified Adjoint Relationship

$$\begin{tikzcd} C[0,t] \arrow[d, leftrightarrow] & \bar{C}[0,t] \arrow[d, leftrightarrow] \\ O[0,t] & \bar{O}[0,t] \end{tikzcd}$$

If another basis is chosen in $\mathbb{R}^n$, the state space representation of the system will vary accordingly. Let $T(\cdot)$ be a differentiable nonsingular state transformation (also called "algebraic" [3])

$$\hat{x}(t) = T(t)\,x(t)$$

It is well known that the corresponding new representation $(\hat{A}(\cdot), \hat{B}(\cdot), \hat{C}(\cdot))$ satisfies

$$\hat{A}(\cdot) = (\dot{T}(\cdot) + T(\cdot)\, A(\cdot))\, T^{-1}(\cdot)$$

$$\hat{B}(\cdot) = T(\cdot)\, B(\cdot)$$

$$\hat{C}(\cdot) = C(\cdot)\, T^{-1}(\cdot)$$

The transition matrix is in the new coordinates

$$\hat{\Phi}(t,\tau) = T(t)\, \Phi(t,\tau)\, T^{-1}(\tau)$$

It follows that the effect on the gramians is as a congruence:

$$\hat{\mathcal{C}}[t_0,t_f] = T(t_f)\mathcal{C}[t_0,t_f]T^T(t_f) \tag{2.11}$$

$$\hat{\mathcal{O}}[t_0,t_f] = T^{-T}(t_0)\mathcal{O}[t_0,t_f]T^{-1}(t_0) \tag{2.12}$$

$$\hat{\overline{\mathcal{C}}}[t_0,t_f] = T(t_0)\overline{\mathcal{C}}[t_0,t_f]T^T(t_o) \tag{2.13}$$

$$\hat{\overline{\mathcal{O}}}[t_0,t_f] = T^{-T}(t_f)\overline{\mathcal{O}}[t_0,t_f]T^{-1}(t_f) \tag{2.14}$$

from which it is easily checked that any of the following products (the order of the factors in the product can be reversed) transforms as a similarity.

$$\begin{array}{ll} \mathcal{O}[t,t_1]\ \mathcal{C}[t_0,t] & t_0 < t < t_1 \\ \overline{\mathcal{O}}[t_0,t]\ \mathcal{C}[t_1,t] & t_0,\ t_1 < t \\ \mathcal{O}[t,t_0]\ \overline{\mathcal{C}}[t,t_1] & t < t_0,\ t_1 \\ \overline{\mathcal{O}}[t_0,t]\ \overline{\mathcal{C}}[t,t_1] & t_0 < t < t_1 \end{array} \tag{2.15}$$

It is not essential that the definition intervals of the gramians have equal length. Hence the eigenvalues of the above products will be invariant with respect to a similarity transformation.

By "balancing" a realization we mean that we "symmetrize" a certain input property (controllability or reachability over some interval) with a certain output property (observability or constructability over some interval) through a suitable choice of basis. This can be made precise by requiring that in the new representation the two gramians of consideration are made equal and diagonal. Loosely speaking we seek the coordinate basis for which the states requiring a high input energy (2.7) (either to "control" or to "reach") are also the states that are highly uncertain (large covariance), either considered backward (observation) or forward (construction) in time.

A choice of a particular input gramian with a particular output gramian is called a "balancing pair" if their arguments match in the sense that their product is one of those in table (2.15). Hence an input gramian and an output gramian form a balancing pair if and only if the eigenvalues of their product is invariant with respect to a coordinate transformation. Which balancing pair is chosen depends on the particular problem we have in mind.

Suppose now that we have a regulator problem for a system in the interval $[t_0,t_f]$. Let the state at time $t \in (t_0,t_f)$ be x. From the physical interpretation, the energy needed to control the state x at t to zero at t_f is proportional to $x'\overline{C}[t,t_f]^{-1}x$. In this problem, the weighting matrix $\overline{C}[t,t_f]$ rather than the reachability matrix C, should be looked at for a quantitative study of the "regulating behavior." What is a good choice for the corresponding output map in the regulator problem? In other words, in pursuing our goal of getting an "input-output symmetrized" realization (we will use the term "balanced" after the notion has been made precise), against which output gramian do we weight the controllability gramian? The observability gramian $\mathcal{O}[t,t_f]$ is a sensible answer. If the output is to be regulated, $x'\mathcal{O}x$, a measure of the energy at the output of the free running system, should be made small for "hard-to-control" states (e.g., x corresponding to small eigenvalues of $\overline{C}[t,t_f]$). Conversely if we allocate "cheap controls" to the states that would yield big output deviations (as measured by the total energy) when free running, then the relative importance of the internal variables will be more transparent. This indicates that the pair $\mathcal{O}[t_0,t_f]$, $\overline{C}[t_0,t_f]$ is a good choice for balancing input and output properties. Furthermore, we have the added bonus that the spectrum of their product does <u>not</u> depend on the chosen coordinate system. Duality suggests that the pair $\overline{\mathcal{O}}[t_0,t_f]$, $C[t_0,t_f]$ will play a major role in the problems of trajectory control, terminal controllers, etc., where the issue is reachability of the states at t_f and also the constructability at t_f (Kalman filter).

The rest of this paper deals uniquely with the C-$\mathcal{O}$ type balancing. Other results are readily transposed.

3. Existence of Balanced Realizations for Analytic Systems

Consider the fixed intervals (t_0,t_1) and (t_1,t_f) and the gramians $\mathcal{C}[t_0,t_1]$ and $\mathcal{O}[t_1,t_f]$. It is well known that $\mathcal{C}$ and $\mathcal{O}$ can be made equal and diagonal with the aid of a suitably chosen matrix T [2] if both are nonsingular (i.e., if the system $(A(\cdot), B(\cdot), C(\cdot))$ is reachable in (t_0,t_1) and observable in (t_1,t_f)). The time variation of the system does not play a role since we are only concerned with fixed intervals and we only need to find a T such that

$$T\mathcal{C}T' = T^{-T}\mathcal{O}T^{-1}$$

For reasons that will become clear later, such a T is called a "balancing transformation." In fact, if $\mathcal{C}$ and/or $\mathcal{O}$ are singular, an invertible transformation and an integer $\rho < n$ exists [4] such that

$$\text{i)} \quad \mathcal{C}[t_0,t_1] = \begin{bmatrix} \Lambda & \\ & \tilde{\mathcal{C}} \end{bmatrix}$$

$$\text{ii)} \quad \mathcal{O}[t_1,t_f] = \begin{bmatrix} \Lambda & \\ & \tilde{\mathcal{O}} \end{bmatrix}, \quad \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_\rho) \text{ with } \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_\rho > 0$$

where

iii) $\tilde{\mathcal{C}}\tilde{\mathcal{O}}$ is nilpotent.

ρ is then the dimension of the observable and reachable (minimal) system. Λ, the part of the gramians that is restricted to readable and observable subspace at time t_1 is called the "Canonical Gramian."

Suppose that the two intervals (t_0,t_1) and (t_1,t_f) are parametrized by α (taking values in an ordered set A) such that

$$t_0(\alpha) < t_1(\alpha) < t_f(\alpha) \quad , \forall \alpha \in A$$

Then we can find the family of balancing transformations

$\{T(t_1(\alpha); t_0(\alpha); t_f(\alpha)) | \alpha \in A\}$ or $\{T(\alpha) | \alpha \in A\}$ for short.

Letting A be some interval I in $\mathbb{R}$, and with suitable reparametrization such that $t_1(t) = t$, we can define the pointwise balancing transformations $\{T(t) | t \in I\}$. Restricting $t_0(\cdot)$ and $t_1(\cdot)$ to the class of differentiable functions, what are the properties of the induced mapping

$$\mathcal{T} : I \to \mathbb{R}^{n\mathrm{x}n} : \mathcal{T}(t) = T(t) \tag{3.1}$$

from I to the space of transformations of $\mathbb{R}^n$?

We shall show that under certain conditions, differentiability and hence continuity follow, and that $\mathrm{Im}\mathcal{T} \subseteq GL(n,\mathbb{R})$ (i.e., $T(t)$ is invertible for all $t \in I$). This then implies the existence of $T(t)$ as a Lyapunov transformation. Moreover, it can then be shown that this pointwise defined Lyapunov transformation works as the time varying balancing transformation.

Definition 1 [19]: Two time-variant systems are called topologically equivalent if one can be transformed in the other by a Lyapunov transformation.

It follows from the continuity of $t_0(t)$ and $t_f(t)$ on a compact interval I, and $t_0(t) < t < t_f(t)$ $\forall t \in I$ that the induced interval

$$\hat{I} = [\min_I t_0(t), \max_I t_f(t)]$$ is compact.

Theorem 1: If a compact interval I is such that the system $(A(\cdot), B(\cdot), C(\cdot))$ is analytic, completely reachable and observable in $\hat{I}$, then it is topologically equivalent in I, to a realization $(\hat{A}(\cdot), \hat{B}(\cdot), \hat{C}(\cdot))$, which is balanced with respect to $C[t_0(t), t]$ and

$O[t, t_f(t)]$ provided that $t_0(\cdot)$ and $t_f(\cdot)$ are continuous and differentiable on I, and at least one is analytic on I.

Before proceeding with the proof of this theorem, we indicate some classes of balanced realizations which are of particular interest.

[illegible]

[illegible]

[illegible]

[illegible] abbreviated $FIB(t_0,t_f)$.

iii) The infinite interval balanced realization, IIB, if it exists can be regarded as a limit case of i) for $T \to \infty$, or a limit of ii) if [illegible]

[illegible] in the context of uniform realizations. It has the serious limitation that only stable systems can be reduced. $FIB(t_0,t_f)$ seems to be more natural in finite time control problems, and both $FIB(t_0,t_f)$ and SIB(T) are possible for unstable systems, filling in an important gap existing with previous balancing methods. For IIB, a modification of theorem 1 which is valid over compact intervals, is needed. This is elaborated in section 5. First we show some preliminary results.

Lemma 1: If $A(t)$ is real analytic in an open interval I, then the associated transition matrix $\Phi(t,\tau)$ is analytic in $I \times I$.

Proof: [5] p. 44

It follows from lemma 1 that if $A(t)$ is analytic for all t greater than some t_0, and if the system $\dot{x}(t) = A(t)\, x(t)$ is stable, the transition matrix is a bounded analytic matrix function in $(t_0,\infty) x (t_0,\infty)$.

Lemma 2: If $A(\cdot)$, $B(\cdot)$ and $C(\cdot)$ are real analytic in an interval I containing t_0 and t_f, then $C[t_0,t_f]$ and $\mathcal{O}[t_0,t_f]$ are real-analytic functions of t_0 and t_f.

Proof: By lemma 1, $\Phi(t_f,\tau)$ is analytic in $I \times I$, and therefore $\Phi(t,\tau)$ can be extended to a holomorphic function $\Phi(t,z)$ in a domain $I \times \Omega$, where $\Omega \subseteq \mathbb{C}'$ encloses the interval I of the real line. Similarly $B(t)$ can be extended to $B(z)$ holomorphic in $\Omega' \subseteq \mathbb{C}$ enclosing I. Letting $D = \Omega \cap \Omega'$, for $\bar{t} \in I$, the function $\Phi(\bar{t},z)\, B(z)\, B'(z)\, \Phi'(\bar{t},z)$ is holomorphic in D. Since I is a connected path in D, the indefinite integral

$$M(\bar{t},t_0;t) \triangleq \int_{t_0}^{t} \Phi(\bar{t},z)\, B(z)\, B'(z)\, \Phi'(\bar{t},z)\, dz$$

is holomorphic $\forall t \in I$, and $C[t_0,t_f] = \Phi(t_f,\bar{t})\, M(\bar{t},t_0;t_f) \Phi'(t_f,\bar{t})$ is analytic with respect to t_f in I by the previous and lemma 1. The other statements are proven analogously.

$\square$

Remark that under the additional assumption of boundedness of $B(\cdot)$ and stability of $A(\cdot)$ it follows that $C[t_0,t_f]$ is bounded for bounded t_0 and t_f since the class of bounded holomorphic functions in D: HB[D] forms a subclass of H[D]([6] p. 75).

Proposition 1:

Let $t_0(t)$ and $t_f(t)$ be differentiable functions of t in I, such that for all t in I, $t_0(t) < t_f(t)$. If $A(t)$, $B(t)$ and $C(t)$ are real analytic (matrix) functions in the interval $\hat{I} = [\min\limits_{I} t_0(t), \max\limits_{I} t_f(t)]$, then the elements of the singular value decomposition of

$C(t) = C[t_0(t),t]$ and $\mathcal{O}(t) = \mathcal{O}\,[t,t_f(t)]$

are continuous and differentiable in I.

Proof: From lemma 2, $C[t_0,t]$ is analytic in both t_0 and t on the rectangle $\hat{I} \times I$. By analytic continuation the domain of definition of $C[t_0,\cdot\,]$ can be extended to a holomorphic function $C[t_0,z]$ in the domain $\hat{I} \times D$ where D is a simply connected complex domain, containing I. The holomorphic family $C[t_0,z]$ is symmetric (i.e., $C[t_0,z]$ is symmetric for $z \in I$), hence its eigenvalues $\lambda_i(t_0,z)$ and eigenprojections $P_i(t_0,z)$ are holomorphic on the real axis and the eigennilpotents $D_i(t_0,z)$ vanish identically([7], p.120). Hence there exist a diagonal matrix $\Lambda_c(t_0,z)$ and a matrix $X_c(t_0,z)$ with normalized columns, both holomorphic on I such that

$$C[t_0,z]\; X_c(t_0,z) = X_c(t_0,z)\; \Lambda_c(t_0,z) \tag{3.2}$$

Hence, on I, both $X_c(t_0,z)$ and $\Lambda_c(t_0,z)$ are differentiable.
Since $X_c(t_0,z)$ is orthogonal on I, this shows that $C[t_0,t]$ has a singular value decomposition

$$C[t_o,t] = X_c(t_o,t)\; \Lambda_c(t_0,t)\; X_c'(t_0,t) \tag{3.3}$$

which elements are differentiable in the second argument. Proving differentiability in the first argument proceeds similarly.

Let now $\mathcal{C}[t_0(t),t] = \mathcal{C}(t)$ have a singular value decomposition

$$\mathcal{C}(t) = U(t)\,\Lambda_c(t)\,U'(t) \tag{3.4}$$

Since $t_0 = t_0(t)$, both $U(t) = X_c(t_0(t),t)$ and $\Lambda_c(t) = \Lambda_c(t_0(t),t)$ are continuous since they are continuous functions of a continuous function. Moreover, differentiability of $U(t)$ follows from the assumptions since

$$\frac{d}{dt}U(t) = \frac{\partial}{\partial t_0}U(t_0,t)\frac{dt_0}{dt} + \frac{\partial}{\partial t}U(t) \tag{3.5}$$

and similarly for the differentiability of $\Lambda_c(t)$. Differentiability of the singular value decomposition of $\mathcal{O}(t)$ is completely analogous. $\square$

<u>Proof of theorem 1</u>: There are two steps; by proposition 1, the singular value decomposition (3.4) has differentiable elements for all t in I. Complete and total controllability are equivalent concepts for analytic systems [8], hence $\mathcal{C}(t)$ (and thus $\Lambda_c(t)$) have full rank in I. In fact $\mathcal{C}(t)$ is bounded away from zero in I, for suppose that for some sequence t_k in I, there exists a nonzero $q \in \mathbb{R}^k$ for which

$$\lim_{t_k \to t^*} q'\,\mathcal{C}(t_k)\,q = 0$$

then by continuity of $\mathcal{C}(t)$ in I, the singularity of $\mathcal{C}(t^*)$ follows, while since I is compact, $t^* \in I$. This contradicts the assumption of complete reachability. Hence $\Lambda_c^{-1}(t)$ is bounded on I.

The transformation $T_1(t) = \Lambda_c^{-1/2}(t)\,U'(t)$ is then continuous, differentiable and has a bounded inverse on compact I since

$$\dot{T}_1(t) = -\frac{1}{2}\dot{\Lambda}_c(t)\,\Lambda^{-3/2}(t)\,U'(t) + \Lambda_c^{-1/2}(t)\,\dot{U}'(t) \tag{3.6}$$

and

$$T_1(t)^{-1} = U(t)\,\Lambda_c^{1/2}(t) \tag{3.7}$$

Hence $T_1(t)$ is a Lyapunov transformation, in fact, it is analytic if $t_0(\cdot)$ is. Then the transformed system matrices $A_1(t)$, $B_1(t)$ and $C_1(t)$ are again analytic in $\mathcal{I}$. Moreover the new gramians are

$$\mathcal{C}_1(t) = I \tag{3.8}$$

$$\mathcal{O}_1(t) = \Lambda_c^{1/2}(t)\, U'(t)\, \mathcal{O}(t)\, U(t)\, \Lambda_c^{1/2}(t) \tag{3.9}$$

for which reason we shall refer to $T_1(t)$ as the "pre-input normalizing transformation." In the second step proposition 1 is again invoked to show that $\mathcal{O}_1(t)$ has a singular value decomposition

$$\mathcal{O}_1(t) = W(t)\, \Lambda^2(t) W'(t) \tag{3.10}$$

where $W(t)$ and $\Lambda(t)$ are continuous and differentiable. Complete observability of the system implies then further the nonsingularity and boundedness of $\mathcal{O}_1(t)$ and $\Lambda(t)$ in $\mathcal{I}$. Hence the transformation

$$T_2(t) = \Lambda^{1/2}(t)\, W'(t) \tag{3.11}$$

is a Lyapunov transformation in $\mathcal{I}$. The balancing transformation for the original realization

$$T(t) = \Lambda^{1/2}(t)\, W'(t)\, \Lambda_c^{-1/2}(t)\, U'(t) \tag{3.12}$$

is consequently a Lyapunov transformation. The system $(A(t), B(t), C(t))$ is then topologically equivalent on $\mathcal{I}$ to the balanced realization

$$(\hat{A}, \hat{B}, \hat{C}) = (\dot{T}T^{-1} + TAT^{-1}, TB, CT^{-1}) \tag{3.13}$$

with canonical gramian $\Lambda(t)$. If $t_f(\cdot)$ is analytic rather than $t_0(\cdot)$, a "pre-input normalizing transformation" should be constructed first.

□

Final remark: For the class of analytic systems, complete reachability is equivalent to the existence of a $t \in \mathcal{I}$ such that the time-varying controllability matrix [8]

$$C_n(t) = [B(t), (A(t) - p)\, B(t), \ldots (A(t) - p)^{n-1}\, B(t)] \tag{3.14a}$$

has full rank. Similarly, complete observability is equivalent to the existence of a $t \in I$, such that

$$\mathcal{O}_n(t) = [C'(t), (A'(t) + p)\, C'(t), \ldots (A'(t) + p)^{n-1}\, C'(t)]' \quad (3.14b)$$

has full rank. $(p = \frac{d}{dt})$

So the conditions of theorem 1 avoid the need for the (often unavailable) transition matrix to establish the existence of a topologically equivalent balanced realization.

All system matrices encountered in the rest of the paper are assumed to be analytic.

Remark that in [3] and subsequently in [24] and [25] unlike here, the existence of a topologically equivalent balanced realization cannot be given directly in terms of the system parameters. For uniform realizations the algebraic transformation is constructed and checked a posteriori for Lyapunovness.

4. Properties of FIB Realizations:

We first characterize a FIB(t_0,t_f) realization and derive some of its properties. SIB(T) realizations will only be briefly discussed, while in the IIB-case, we need additional conditions to guarantee the existence of a topologically equivalent balanced realizations.

Theorem 2: The n-th order realization $(A(\cdot), B(\cdot), C(\cdot))$ is FIB (t_0,t_f) iff there exist n nonnegative and differentiable functions $\{\lambda_i(\cdot)$: $i = i, \ldots, n\}$ such that

$$\Lambda(t_0) = \Lambda(t_f) = 0 \tag{4.1}$$

$$\left.\begin{aligned} A'\Lambda + \Lambda A + C'C &= \dot{\Lambda} \qquad (4.2a) \\ A\Lambda + \Lambda A' + BB' &= -\dot{\Lambda} \qquad (4.2b) \end{aligned}\right\} \quad t \in (t_0,t_f)$$

where $\Lambda(t) = \mathrm{diag}\,\{\lambda_i(t) : i = 1, \ldots, n\}$

Proof: Sufficienty follows from the uniqueness of the solution of a linear ordinary differential equation.
Necessity is direct since the observability gramian of any system satisfies $\mathcal{O}(t_f) = 0$ and (4.2a), and the reachability gramian $\mathcal{C}(t_0) = 0$ and (4.2b). By definition of balancedness they are equal.

$\square$

This connection between the elements of a balanced realization can be exploited to improve the accuracy in an actual computation. For stability analysis, the following corollary is important.

<u>Corollary 1</u>: The matrix $A(t)$ of a (not necessarily stable) $FIB(t_0,t_f)$ realization is negative semidefinite for all t in (t_0,t_f).

<u>Proof</u>: From (5.2a) and (5.2b) we get (A_s = symmetric part of A).

$$A_s(t)\ \Lambda(t) + \Lambda(t)\ A_s(t) = -(BB' + C'C)_t \tag{4.3}$$

Since $\Lambda(t) > 0$ in (t_0,t_f) it follows that

$$A_s(t) = -\int_0^\infty e^{-\Lambda(t)\tau}\ (BB' + C'C)_t\ e^{-\Lambda(t)\tau}\ d\tau \leq 0 \tag{4.4}$$

Thus for all t in (t_0,t_f) $A(t)$ is negative semi-definite since its symmetric part is. $\square$

It can be shown [4] that actually $A(t)$ is negative definite at t iff the pair $\{\Lambda(t), [B(t), C'(t)]\}$ "frozen" at t is controllable. This result is very unusual since it relates (pointwise) properties of a time varying system to the properties of a time-invariant one, constructed from a "frozen" time-varying system. A sufficient condition for pointwise positive definiteness of $A(t)$ can then be derived from this result as the condition that at time t all λ_i are disjoint and none is stationary.

Another result generalizes the symmetry properties reported in [2] for scalar time invariant systems (derived for IIB).

<u>Corollary 2</u>: A single input-single output $FIB(t_0,t_f)$ can be completely characterized by the 2n functions

$\{\lambda_i(t),\ \sigma_i\beta_i(t)\ ;\ i = 1, \ldots, n\}$ where $\lambda_i(t)$ and $\beta_i(t) \geq 0$ and $\Sigma = \text{diag}\ \{\sigma_i\}$ is a signature matrix.

Proof:

$$\text{Set } \beta_i(t) = \sqrt{\frac{b_i^{\,2}(t) + c_i^{\,2}(t)}{2}} \tag{4.5}$$

It follows then from theorem 3 that

$$b_i = \varepsilon_i \cdot \sqrt{\beta_i^{\,2} - \dot{\lambda}_i} \tag{4.6}$$

$$c_i = \varepsilon_i\sigma_i \cdot \sqrt{\beta_i^{\,2} + \dot{\lambda}_i} \tag{4.7}$$

where the ε_i are arbitrary signs and σ_i = sign β_i .

After some algebra, equations (5.2) yield

$$2a_{ii}\lambda_i = -\beta_i^{\,2} \tag{4.8}$$

$$(\lambda_i^{\,2} - \lambda_j^{\,2})a_{ji} = \left(\sigma_i\sigma_j\lambda_j\sqrt{(\beta_i^{\,2} + \dot{\lambda}_i)(\beta_j^{\,2} + \dot{\lambda}_j)} - \lambda_i\sqrt{(\beta_i^{\,2} - \dot{\lambda}_i)(\beta_j^{\,2} - \dot{\lambda}_j)}\right)\varepsilon_i\varepsilon_j \qquad i \neq j \tag{4.9}$$

A sign change of b_i (ε_i) corresponds to the (elementary) similarity transformation of changing the sign of the i^{th} state component. The overall transfer is not affected, hence without loss of generality one can set all ε_i equal to plus one. We then have the parametrization of the balanced realization, assuming that all λ_i are non zero and disjoint.

$$b_i = \sqrt{\beta_i^2 - \dot{\lambda}_i}\ , \quad c_i = \sigma_i\sqrt{\beta_i^2 + \dot{\lambda}_i}\ , \quad a_{ii} = -\frac{\beta_i^2}{2\lambda_i}$$

$$a_{ji} = \sigma_i \frac{\sigma_j\lambda_j\sqrt{(\beta_i^2 + \dot{\lambda}_i)(\beta_j^2 + \dot{\lambda}_j)} - \sigma_i\lambda_i\sqrt{(\beta_i^2 - \dot{\lambda}_i)(\beta_j^2 - \dot{\lambda}_j)}}{\lambda_i^2 - \lambda_j^2} \qquad (4.10)$$

$\square$

FIB-realizations are useful in problems where fixed initial and final time are important (as in certain regulation problems to the nominal trajectories in rendezvous, etc.) Nonstationarity of the system typically arises from the linearization of a nonlinear system about a nominal trajectory.

The importance and applicability of balanced realizations over semi-infinite intervals was first illustrated by Moore [2] and in the context of digital filters by Roberts and Mullis [9] for stationary systems. They have been further studied by Sastry [10] and Pernebo [11] for stationary systems, and by Silverman et al [3], [24], [25] for uniform systems. The existence conditions in the latter work are less elegant since they rely on an a posteriori check for "Lyapunovness" of the balancing transformation. On the other hand the class of uniform realizations has or conserves remarkable properties [19].

Many of the previous results will carry over if we formally let $t_0 \to -\infty$ and $t_f \to +\infty$. Unfortunately, the existence conditions established in section 3 only deal with systems defined over compact intervals. Additional conditions are needed to guarantee the existence of IIB-realizations.

The main difficulty in establishing the results for IIB is the requirement of boundedness of T, $\dot{T}$ and T^{-1} as $t \to \pm\infty$. However in many applications only a finite timespan is of interest. (e.g. terminal controller problems)

Of course, $A(\cdot)$, $B(\cdot)$, and $C(\cdot)$ still have to be defined for all $t \in \mathbb{R}$ in order that the infinite-interval gramians are well-defined. What we mean here is that t in $\mathcal{C}(-\infty,t] \triangleq \mathcal{C}_\infty(t)$ and $\mathcal{O}[t,+\infty) \triangleq \mathcal{O}_\infty(t)$ only varies in a compact interval. This eliminates the problems of unboundedness at infinity, and on compact sets, boundedness and continuity are equivalent. Then we can take full advantage of the existence conditions developed in Section 3, where compact intervals were considered. Note that from a practical viewpoint, we are always "limited" to some finite time interval, and thus all we really have to consider is boundedness for finite time (i.e., continuity). For these reasons a broader class of transformations than the (strict-) Lyapunov transformations can be allowed. It is therefore useful to define a notion of "almost Lyapunov" or quasi-Lyapunov transformation.

Definition 2: An algebraic transformation $T(t)$ is quasi-Lyapunov iff $T(t)$, $T^{-1}(t)$ and $\dot{T}(t)$ are continuous and bounded on compact intervals.

Theorem 3: If $A(t)$, $B(t)$ and $C(t)$ are analytic matrices such that $C_\infty(t)$ and $O_\infty(t)$ exist (i.e., are finite) and are nonsingular for all (finite) t, then $(A(\cdot), B(\cdot), C(\cdot))$ can be IIB balanced by a quasi-Lyapunov transformation.

The proof requires only a slight modification to the proof of theorem 1. Under the assumptions, it follows that $C_\infty(t)$ and $O_\infty(t)$ are analytic for all (finite) t. Hence they are continuous and differentiable, and obviously invertible. The extension of proposition 1 is then direct; and from here the proof of theorem 1 carries over verbatim for I: any compact interval in $(-\infty,\infty)$. $\square$

A sufficient condition for the existence of $C_\infty(t)$ and $O_\infty(t)$ is the square integrability of $||\Phi(t,\tau)\ B(\tau)||$ and $||\Phi'(\tau,t)C'(\tau)||$, which is in turn implied by the analyticity of $(A(\cdot), B(\cdot), C(\cdot))$ in any compact interval in $(-\infty,\infty)$ with the asymptotic stability of $A(t)$. Remark that the existence of $C_\infty(t)$ and $O_\infty(t)$ implies the existence of $C_\infty(t')$ for all $t' \leq t$ and $O_\infty(t'')$ for all $t'' \geq t$. The invertibility of $C_\infty(t)$ and $O_\infty(t)$, if they exist, is equivalent to complete reachability and observability of $(A(\cdot), B(\cdot), C(\cdot))$ in $(-\infty,+\infty)$. The above result shows in particular that the complete reachability for all $(-\infty,t)$, and the existence of $C_\infty(T)$ implies that the pre-input normalizing transformation $T_1(t) = \Lambda_c^{-1/2}(t)U(t)$ is quasi-Lyapunov in any compact interval enclosed in $(-\infty,T]$. In general, it will not be a Lyapunov transformation since T_1, $\dot{T}_1$ and T_1^{-1} might not be bounded as $t \to \pm\infty$. Finally, note that a quasi-Lyapunov transformation preserves finite-time stability and is therefore significant in simulations.

If (strict)-Lyapunovness of the balancing transformation is required, a stronger version of theorem 3 is needed. Boundedness of T_1 and T_1^{-1} is implied by the boundedness of C^{-1} and C respectively since

$$||T_1(t)|| = ||\Lambda_c^{-1/2}(t)|| = ||\Lambda_c(t)^{-1}||^{1/2} = ||C^{-1}(t)||^{1/2}$$

$$||T_1(t)^{-1}|| = ||\Lambda_c^{1/2}(t)|| = ||\Lambda_c(t)||^{1/2} = ||C(t)||^{1/2}$$

It is therefore natural to add a certain "uniformity" condition on $C_\infty(\cdot)$ (and similarly on $\mathcal{O}_\infty(\cdot)$).

Definition 3:

i) $(A(\cdot), B(\cdot))$ is "boundedly completely reachable" (b.c.r.) iff there exist $0 < \alpha_m < \alpha_M < \infty$ such that

$$0 < \alpha_m I \leq C_\infty(t) \leq \alpha_M I < \infty \,, \quad \forall\, t \in \mathbb{R} \tag{5.1}$$

ii) $(A(\cdot), C(\cdot))$ is "boundedly completely observable" (b.c.O) iff there exist $0 < \beta_m < \beta_M < \infty$ such that

$$0 < \beta_m I \leq \mathcal{O}_\infty(t) \leq \beta_M I < \infty \,, \forall\, t \in \mathbb{R} \tag{5.2}$$

A minimal stationary system is clearly boundedly completely reachable and boundedly completely observable. Note that this definition does not imply and is not implied by uniform reachability and observability. [13].
The transformation $T_1(t) = \Lambda_c^{-1/2}(t)\, U(t)$ is guaranteed to be a bounded algebraic transformation (with bounded inverse) if $(A(\cdot), B(\cdot))$ is b.c.r. With an additional condition on the behavior of $C_\infty(t)$ at infinity, $T_1(t)$ also has a bounded derivative.

We say that $M(t)$ has disjoint eigenvalues as $t \to \infty$, iff for any two eigenvalues $\lambda_i(t)$ and $\lambda_j(t)$ of $M(t)$ $i \neq j$, there exists a $\tau_{ij} < \infty$ and $\varepsilon_{ij} > 0$ such that

$$|\lambda_i(t) - \lambda_j(t)| > \varepsilon_{ij} \qquad \forall\, t > \tau_{ij} \tag{5.3}$$

This implies for finite dimensional systems the existence of $\tau < \infty$ and $\varepsilon > 0$ such that for all $i \neq j$

$$|\lambda_i(t) - \lambda_j(t)| > \varepsilon \qquad \forall\, t > \tau \tag{5.4}$$

<u>Theorem 4</u>: If $A(t)$ and $B(t)$ are bounded analytic matrices such that $(A(\cdot), B(\cdot))$ is a boundedly completely reachable pair and if their reachability gramian $C_\infty(t)$ has disjoint eigenvalues as t approaches infinity, then the pre-input normalizing transformation $T_1(t)$ is a Lyapunov transformation.

The proof is deferred to the appendix.

Sufficient conditions for the existence of an IIB realization, topologically equivalent to the given realization, are then:

<u>Theorem 5</u>: If $A(t)$, $B(t)$, and $C(t)$ are bounded analytic matrices on $(-\infty, \infty)$ such that the realization $(A(\cdot), B(\cdot), C(\cdot))$ is boundedly completely reachable and controllable, and if further either

i) C_∞ and $C_\infty \mathcal{O}_\infty$ have disjoint eigenvalues for $t \to +\infty$ and $t \to -\infty$ respectively, or

ii) C_∞ and $C_\infty \mathcal{O}_\infty$ have disjoint eigenvalues for $t \to -\infty$ and $t \to +\infty$ respectively,

then a topologically equivalent balanced realization exists, valid over $(-\infty, \infty)$.

Proof: From theorem 4, it follows that the matrix

$$\bar{O}_\infty(t) = T_1^{-T}(t)\; O_\infty(t) T_1^{-1}(t)$$

is also Lyapunov. Its eigenvalues are those of $C_\infty(t)O_\infty(t)$, and hence are disjoint as $t \to -\infty$. Let $\bar{O}_\infty(t)$ have the singular value decomposition

$$\bar{O}_\infty(t) = W(t)\;\Lambda^2(t)\;W'(t)$$

then from lemma's A1 and A2 (or the dual of theorem 4) it follows that $\dot{W}(t)$ and $\dot{\Lambda}(t)$ are bounded, and that

$$T_2(t) = \Lambda^{1/2}(t)\;W(t)$$

is Lyapunov. If instead the eigenvalue condition on O_∞ is known to hold, then we can "dualize" by first considering the transformation that brings the observability gramian to the identity matrix.

$\square$

In the rest of this section we shall assume that the IIB-realization exists.

Theorem 6: If $(A(\cdot), B(\cdot), C(\cdot))$ is IIB with canonical gramian $\Lambda(\cdot)$, then

$$\begin{aligned} A'\Lambda + \Lambda A + C'C &= \dot{\Lambda} \\ A\Lambda + \Lambda A' + BB' &= -\dot{\Lambda} \end{aligned} \tag{5.5}$$

The proof is trivial (cf. theorem 2). Note that the conditions (4.1) do not carry over, and the sufficiency-equivalent of theorem 2 is lost. The corollaries to theorem 2 still hold where IIB is possible, in particular the negative semidefiniteness of $A(t)$, because only the necessity part is needed. An important special case is formed by the stationary realizations; while FIB$[t_0,t_f]$ leads to a time varying realization, the IIB-realization, if it exists, is also stationary, and the relations (4.10) imply the symmetry relations in the scalar case

$$\begin{aligned} b &= \Sigma c' \\ A' &= \Sigma A \Sigma \end{aligned} \tag{5.6}$$

In fact, the transferfunction h(s) can be written as (for disjoint λ_i)

$$h(s) = b'(s\Sigma - BEB)^{-1} b, \tag{5.7}$$

where $B = \text{diag}\,\{b_i\}$ and $E_{ij} = -(\lambda_i\sigma_i + \lambda_j\sigma_j)^{-1}$ (5.8)

<u>Theorem 7</u>: The n^{th} order balanced realization of a stationary single-input-single-output minimal system of order n cannot have a zero element in its $\hat{b}$ and/or $\hat{c}$ vector, nor on the principal diagonal of $\hat{A}$, provided that the corresponding element of its canonical gramian is distinct from all the others.

<u>Proof</u>: From theorem 6:

$$a_{ii}(t) = -\frac{b_i^2}{2\lambda_i} = -\frac{c_i^2}{2\lambda_i}$$

hence $b_i = 0 \Leftrightarrow a_{ii} = 0 \Leftrightarrow c_i = 0$ (5.9)

Assume now (5.9)

It follows then from (4.10) that for all $j \neq i$, $a_{ij} = a_{ji} = 0$, and thus that the i^{th} state component is unobservable and unreachable, contradicting the existence of the IIB realization.

6. Sliding Interval Balanced Realizations

Both the fixed interval balanced realization(FIB) and the infinite interval balanced realization analyzed in Sections 4 and 5 suffer from some serious problems. In the first case the reachability gramian is zero at the starting time, and gradually builds up, while the observability gramian decreases and is zero at the final time. Those points are therefore singular points for the balancing transformation. This reflects in the elements of the "balanced realization" being unbounded as t closes in to t_0 or t_f.

For infinite interval balancing, the issue is stability, or convergence of the integrals defining the gramians. This motivates us to consider the gramians over an interval with fixed finite, nonzero length. If T is the length of the interval, then the gramians leading to "Sliding Interval Balancing (SIB(T)) are

$$\mathcal{C}[t - T,t] \triangleq \mathcal{C}_T(t)$$

$$\mathcal{O}[t,t + T] \triangleq \mathcal{O}_T(t)$$

This ties in with the theory of uniform balanced realizations [3],[24], and [25].

Theorem 8: If $(A(\cdot), B(\cdot), C(\cdot))$ is analytic and completely reachable and observable in a compact $[t_0,t_f]$, then the SIB(T) realization exists for all $0 < T < (t_f-t_0)/2$ and $t_0 + T < t < t_f - T$.

Proof: If the realization is completely reachable and observable in the interval (t_0,t_f), by the analyticity assumption it is then reachable and observable in every subinterval of (t_0,t_f). Thus in particular $\mathcal{C}_T(t)$ and $\mathcal{O}_T(t)$ are nonsingular for all t and T. By theorem 1 the balancing transformation exists.

□

A sliding-interval-balanced realization of a stationary system is also stationary. This "structure-conservation" property also holds for periodic systems. SIB(T) realizations and their associated canonical gramian of a periodic system are itself periodic.

Unlike $FIB(t_0,t_f)$ and IIB, the Lyapunov-type differential equations do not lead to a simple criterion for "balancedness" of a realization. The canonical gramian for the SIB(T) realization satisfies

$$\dot{\Lambda}(t) = A(t)\Lambda(t) + \Lambda(t)A'(t) + B(t)B'(t) - \Phi(t,t-T)B(t)B'(t)\Phi'(t,t-T) \tag{6.1}$$

$$\dot{\Lambda}(t) = -A'(t)\Lambda(t) - \Lambda(t)A(t) - C'(t)C(t) + \Phi'(t + T,t)C'(t)C(t)\Phi(t + T,t) \tag{6.2}$$

The need for the transition matrix of A(t) obscures the usefulness of these equations. However they infer that the symmetry relations

$$A' = \Sigma A \Sigma \tag{6.3}$$

$$b = \Sigma c' \tag{6.4}$$

where Σ is some signature matrix, still hold for SIB(T) realizations of a scalar stationary system, as is readily verified.

7. Applications to Model Order Reduction

The canonical gramian is a direct measure for the relative importance of each "dimension". This makes the balanced realization theory very useful for model reduction. This fact was initially recognized by B. C. Moore [2], for the case of infinite interval balancing of stationary systems, both in the continuous and discrete case.

The reduced order model is given by a consistent partitioning of the full order balanced realization, i.e., if

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad C = [\; C_1 \quad C_2 \;] \tag{7.1}$$

consistent with

$$\Lambda = \begin{bmatrix} \Lambda_1 & \\ & \Lambda_2 \end{bmatrix}, \; \Lambda_1 > \Lambda_2$$

then (A_{11}, B_1, C_1) is a reduced order subsystem. Corresponding to the different types of balancing, we get different reduced order models for a given system. The following theorems describe their properties. A_s indicates the symmetric part of A.

Theorem 9 :

If $(A(\cdot), B(\cdot), C(\cdot))$ is FIB(t_0,t_f) and partitioned as in (8.1), then the subsystem $(A_{11}(\cdot), B_1(\cdot), C_1(\cdot))$

i) has a transition matrix which is contractive, and strictly contractive if $(A_{11}(t))_s$ is not singular for all t.

ii) is automatically balanced.

Proof: By equation (4.4), $A_s(t)$ is negative semi-definite for all t in (t_0,t_f), whence also

$$[\hat{A}_s(t)]_{11} \leq 0 \quad . \tag{7.2}$$

Wazewski's inequality ([15] pp. 117-119) implies for the solution of

$$\dot{x}_1(t) = \hat{A}(t)_{11}\, x_1(t) \quad , \quad x_1(t_0) = x_0$$

$$||x_2(t;x_0,t_0)|| \leq ||x_0|| \exp\left(\frac{1}{2} \int_{t_0}^{t} \lambda_{max}(\tau) d\tau\right) \leq ||x_0||, \; t \geq t_0 \tag{7.3}$$

where $\lambda_{max}(\tau)$ is the largest eigenvalue of $[\hat{A}_s(t)]_{11}$. If $(A_{11}(t))_s$ is not singular for all t, then by analyticity, $\lambda_{max}(\tau)$ in (7.3) is zero on a set with measure zero, and negative elsewhere. So there is a $0 < K < 1$, in general, depending on t_0 such that

$$||x(t)|| \leq ||x_0||\, K$$

The inheritance of balancedness to the subsystem follows at once from the sufficiency part of theorem 2.

□

The important fact learned from Theorem 9 is that the state of all free-running principal subsystems (and therefore the balanced realization itself) of a $FIB(t_0,t_f)$ realization is bounded by the initial state.

Corollary: The subsystem (A_{11},B_1,C_1) of an $FIB(t_0,t_f)$ realization (A,B,C) is bounded input-bounded state stable in (t_0,t_f).

Proof: From theorem 9 :

$$||x(t)|| \le ||\Phi_{A_{11}}(t,t_0)\, x_0 + \int_{t_0}^{t} \Phi_{A_{11}}(t,\tau)\, B_1(\tau)\, u(\tau)d\tau||$$

$$\le ||x_0|| + ||\int_{t_0}^{t_f} \Phi_{A_{11}}(t,\tau)B'(\tau)u(\tau)d\tau||$$

$$\le ||x_0|| + ||\int_{t_0}^{t} \Phi_{A_{11}}(t,\tau)B_1(\tau)B_1'(\tau)\Phi'_{A_{11}}(t,\tau)d\tau||^{1/2}$$

$$\cdot ||\int_{t_0}^{t_f} u'(\tau)u(\tau)d\tau||^{1/2}$$

$$\le ||x_0|| + ||\Lambda_1(t)||^{1/2} \cdot ||u||_2$$

using Schwarz's inequality. $\Lambda_1(t)$, as a submatrix of $\Lambda(t)$ is bounded. $\square$

For IIB the situation is more delicate since the conditions of theorem 6 are no longer sufficient. The equivalent to part ii) of theorem 9 will need additional conditions. Part i) carries over integrally.

Theorem 10:

If $(A(\cdot),B(\cdot),C(\cdot))$ is IIB and partitioned as in (8.1), then the subsystem $(A_{11}(\cdot),\ B_1(\cdot),\ C_1(\cdot))$

i) is internally stable, i.e., the solution to the homogeneous system remains bounded as $t \to \infty$. Its transition matrix is strictly contractive if $(A_{11}(t))_s$ is not singular for all t.

ii) is IIB balanced with canonical gramian $\Lambda_1(t)_1$ if $\lim \Lambda(t)$ exists at $\pm\infty$, and

a) $A_{11}(t)$ and $A_{11}'(-t)$ are asymptotically stable

b) $$C_1(-\infty) = \lim_{t\to-\infty} \int_{-\infty}^{t} \Phi_{A_{11}}(t,\tau)B_1(\tau)B_1'(\tau)\Phi_{A_{11}}'(t,\tau)d\tau$$

exists

c) $$O_1(+\infty) = \lim_{t\to+\infty} \int_{t}^{\infty} \Phi_{A_{11}}'(\tau,t)C_1'(\tau)C_1(\tau)\Phi_{A_{11}}(\tau,t)d\tau \text{ exists}$$

d) For some t_* : $A_{11}(t_*)$ and $-A_{11}(t_*)$ have no common eigenvalue.

Remark that strict contractivity does not imply asymptotic stability, hence the stronger assumption a) in part ii). A sufficient condition for asymptotic stability is [15].

$$\int_{t_0}^{\infty} \lambda_{max}(\tau)d\tau = -\infty \quad , \quad \forall\, t_0$$

<u>Proof of part ii)</u>: If $C_1(t)$ is the reachability gramian of the system (A_{11},B_1,C_1), then $C_1(t)$ satisfies

$$\dot{C}_1(t) = A_{11}(t)\, C_1(t) + C_1(t)\, A_{11}(t)' + B_1(t)\, B_1(t)'$$

By theorem 6, and (7.1), the difference $C_1(t) - \Lambda_1(t)$ satisfies

$$\dot{X}(t) = A_{11}(t)\, X(t) + X(t)\, A_{11}'(t) \quad . \tag{7.4}$$

The homogeneous equation (7.4) has the general solution

$$X(t) = \Phi_{A_{11}}(t,0)\, M\Phi_{A_{11}}'(t,0) + X_C \tag{7.5}$$

where M is a (symmetric) matrix of integration constants, and X_C is a constant (symmetric) solution (if any) of the algebraic equation

$$A_{11}(t)\, X_C + X_C A_{11}'(t) = 0 \quad . \tag{7.6}$$

But, this equation has only the null solution iff $A_{11}(t)$ and $-A_{11}(t)$ have no common eigenvalues. Invoking continuity, condition d) sufficies to ensure that $X_c = 0$, and thus

$$C_1(t) = \Lambda_1(t) + \Phi_{A_{11}}(t,0)\, M \Phi_{A_{11}}'(t,0) \tag{7.7}$$

Similarly, for some symmetric N

$$O_1(t) = \Lambda_1(t) + \Phi_{A_{11}}'(0,t)\, N \Phi_{A_{11}}(0,t) \tag{7.8}$$

The asymptotic stability of $A_{11}'(-t)$ implies that

$$\lim_{t \to -\infty} ||\Phi_{A_{11}}(0,t)|| \to 0$$

and thus the unboundedness of $\Phi_{A_{11}}(t,0)\, M \Phi_{A_{11}}'(t,0)$ as $t \to -\infty$, for any nonzero M. Boundedness of $\Lambda_1(t)$ as $t \to -\infty$ is implied since it is contained in the bounded $\Lambda(t)$. Condition b) and equation (7.7) imply then that M can only be zero and thus $C_1(t) = \Lambda_1(t) \quad , \ \forall\, t$

The part $O_1(t) = \Lambda_1(t)$ is analogous, using c) and asymptotic stability of A(t).

$\square$

Remark that asymptotic stability does not imply the conditions b) and c), even if B and C are bounded as the counterexample $A(t) = -2t$, and $B(t) = 1$ readily illustrates.

Theorem 11:

If (A,B,C) is IIB, partitioned as in (7.1), then the following implications hold:

i) (A_{11},C_1) completely observable $\rightarrow A_{11}(t)$ is asymptotically stable

ii) (A_{11},B_1) completely reachable $\rightarrow A_{11}'(-t)$ is asymptotically stable

<u>Proof</u>: i) Consider the function $V(x)= x'\Lambda_1 x$. Along trajectories of the subsystem we find

$$\dot{V}(x) = x'(A_{11}'\Lambda_1 + \dot{\Lambda}_1 + \Lambda_1 A_{11})x$$

$$= -x' C_1' C_1 x$$

$$= -x_0' \Phi_{A_{11}}'(t,0) C_1(t)' C_1(t) \Phi_{A_{11}}(t,0) x_0$$

By the complete observability, the columns of $C_1(t)\Phi_{A_{11}}(t,0)$ are linear independent. Thus $\dot{V}(x) \leq 0$ and not identically zero along trajectories. Hence $V(x)$ is a Lyapunov function.

The proof of ii) is similar.

Statement i) is adopted from [3], where it is stated in the context of uniform realizations, and hence the stronger requirement of uniform complete observability was used, to yield the stronger result of uniform asymptotic stability. This uniform condition implies c) in theorem 10. Also the existence of $\Lambda(t)$ at infinity is implied if (A,B,C) is a uniform balanced realization [3]. Boundedness of $B_1(t)$ and uniform asymptotic stability of $A_{11}(t)$ (which is again equivalent to BIBO stability for uniform systems [16]) implies condition b). Hence for balanced uniform realizations (BUR), we can modify theorem 1^ to:

<u>Theorem 10"</u>:

If (A(·), B(·), C(·)) is a <u>uniform</u> IIB realization, partitioned as in (8.1), and if the subsystem $(A_{11}(\cdot), B_1(\cdot), C_1(\cdot))$ is uniformly completely observable, then it is

i) asymptotically stable

ii) IIB if also

a) $(A_{11}(\cdot), B_1(\cdot), C_1(\cdot))$ BIBO-stable

b) $A_{11}(t)$ and $-A_{11}(t)$ have no common eigenvalue for some t

Remarks:

i) If (A,B,C) is a time invariant BUR, the conditions a) and b) above are implied by the asymptotic stability, and we retrieve the original result by B. C. Moore for stationary systems ([2], Theorem 6).

ii) Theorem 10' can also be "dualized"

iii) Shokoohi et al. [24] and [25] focus on the stability of approximations of uniformly balanced asymptotically stable systems.

Using our SIB(T) method for variable T, we are able to obtain reduced order models corresponding to the slow (large T) or fast (small T) response of the system. This technique for small T can be linked to singular perturbation methods [4].

For practical applications, a criterion is suggested to decide on the order of a reduced model. To this end , the "real dimension" of a system (with respect to some type of balancing) is defined as

$$\rho = 2^{H_\Lambda} \tag{7.9}$$

where H_Λ is the entropy (logarithms to the base 2) associated with the corresponding canonical gramian (i.e., the entropy of the distribution $\{\frac{\lambda_i}{tr\Lambda}, i = 1, \ldots, n\}$ where $\Lambda = \text{diag}\ (\lambda_1 \ldots \lambda_n)$). The reduced order should be larger than ρ. In the example below, the real dimension with respect to SIB(T), is close to one for T very small and T very large. In the first case, the reduced order system mimics the slow (unstable) behavior, while for T large, the stable fast response is modeled.

In the feedback design, this added freedom helps in matching the approximate model to the specific design problem more adequately.

8. Conclusions

The concept of balancing a given realization is extneded to analytic time-varying systems. It is further generalized to obtain different types of balancing. In particular we introduced the new concepts of Fixed Interval Balancing and Sliding Interval Balancing, and established for analytic systems

i) existence conditions,

ii) properties of the various balanced realizations, and

iii) applications to model reduction, and their properties.

The ideas of balancing a realization can be extended to the LQG-design problem, as already shown by the author in [17] and by the independent work of Jonckheere et al. [18].

APPENDIX: Proof of Theorem 4

Several preliminary results towards the proof of the existence theorem of IIB realizations are first established.

Lemma A1: Let $X(t)$ be a symmetric positive definite Lyapunov transformation, with singular value decomposition $X(t) = U(t)\,\Lambda^2(t)\,U'(t)$, then both $\dot{\Lambda}(t)$ and $A(t)\,\Lambda(t) - \Lambda(t)\,A(t)$ are bounded on $(-\infty,\infty)$ where:

$$A = U'\dot{U} \quad . \tag{A.1}$$

Proof:

Let $S(t)$ be defined as the unique positive definite symmetric square root of $X(t)$: (take $\Lambda(t) > 0$)

$$S(t) = U(t)\,\Lambda(t)\,U'(t) \quad . \tag{A.2}$$

It follows from a lemma by Silverman ([19], Corollary 1) that $S(t)$ is also a Lyapunov transformation, and thus that $\dot{S}$ is bounded. But

$$\dot{S} = \dot{U}\Lambda U' + U\dot{\Lambda}U' + U\Lambda\dot{U}' \quad . \tag{A.3}$$

Pre- and post-multiplication of (A.3) by U' and U respectively yields:

$$U'\dot{S}U = U'\dot{U}\Lambda + \dot{\Lambda} + \Lambda\dot{U}'U \quad . \tag{A.4}$$

From the orthogonality of U, we obtain

$$\dot{U}'U = -U'\dot{U} \quad . \tag{A.5}$$

Substituting equation (A.5) in equation (A.4) gives

$$U'\dot{S}U = U'\dot{U}\Lambda + \dot{\Lambda} - \Lambda U'\dot{U} \tag{A.6}$$

or, by virtue of definition (A.1)

$$U'\dot{S}U = A\Lambda + \dot{\Lambda} - \Lambda A \quad . \tag{A.7}$$

At the same time (A.5) indicates that A is an antisymmetric matrix. The left-hand side of equation (A.7) is bounded because S is Lyapunov. Thus the elements of the matrix on the right-hand side must be bounded. The diagonal elements are

$$a_{ii}\lambda_i + \dot{\lambda}_i - \lambda_i a_{ii} = \dot{\lambda}_i = (\dot{\Lambda})_i \tag{A.8a}$$

while the off-diagonal elements reduce to

$$a_{ij}\lambda_j - \lambda_j a_{ij} = (A\Lambda - \Lambda A)_{ij} \quad , \quad i \neq j \tag{A.8b}$$

The statement of the lemma thus follows.

□

Lemma A2: Let X(t) be a symmetric positive definite Lyapunov transformation with singular value decomposition $X(t) = U(t)\Lambda^2(t)\,U'(t)$, then $\dot{U}(t)$ is bounded on $(-\infty,\infty)$ if the eigenvalues of X(t) are uniformly disjoint.

Proof:

Since X(t) is Lyapunov, its inverse is bounded for all t, thus all eigenvalues are strictly bounded away from zero. Under the additional assumption of uniform disjointness of elements of Λ^2, there exists an $\varepsilon > 0$ such that

$$|\lambda_i^2 - \lambda_j^2| \geq \varepsilon_{ij} \tag{A.9}$$

thus for some $\varepsilon > 0$

$$|\lambda_i - \lambda_j| \geq \frac{\varepsilon_{ij}}{\lambda_i + \lambda_j} \geq \varepsilon > 0 \qquad \forall t \;, \; \forall i \neq j \tag{A.10}$$

By lemma A1, the elements $a_{ij}(\lambda_i - \lambda_j)$ are bounded (by M say), and thus

$$|a_{ij}| \le \frac{M}{|\lambda_i - \lambda_j|} \le \frac{M}{\varepsilon} < \infty$$

where a_{ij} is the (ij)-element of the matrix A defined in equation (A.1). Note that the diagonal elements are zero. Thus A is bounded and

$$||\dot{U}|| = ||UA|| = ||A|| < \infty \quad .$$

□

Lemma A2 can now be applied to the reachability gramian of a boundedly completely reachable, analytic pair (A(t),B(t)), proving theorem 4.

Proof of theorem 4:

From the discussion preceeding the statement of the theorem it suffices to show that U(t) and $\Lambda_c(t)$ have bounded derivatives at infinity.

From the boundedness of A(t) and B(t), it follows that

$$\begin{aligned} ||\dot{C}_\infty(t)|| &\le ||A(t)C_\infty(t) + C_\infty(t)A'(t) + B(t)B'(t)|| \\ &\le 2||A(t)||\ ||C_\infty(t)|| + ||B(t)||^2 \end{aligned} \tag{A.11}$$

and thus boundedness of $\dot{C}_\infty(t)$ in $(-\infty,\infty)$ is ensured by boundedly completely reachability. Thus $C_\infty(t)$ is a Lyapunov transformation and from Lemmas A1 and A2 (note that "disjointness" at $+\infty$ suffices, since boundedness on compact intervals follows from the analyticity assumptions.) we obtain that $\dot{\Lambda}_c$ and $\dot{U}$ are bounded on $(-\infty,\infty)$. Finally it follows that

$$\dot{T}_1(t) = -\frac{1}{2}\Lambda_c^{-3/2}\dot{\Lambda}_c U' - \Lambda_c^{-1/2}\dot{U}' \qquad (A.12)$$

is bounded as t approaches plus or minus infinity. Since $T_1(t)$ was already shown to be a bounded algebraic transformation on $\mathbb{R}$, (A.12) proves that $T_1(t)$ is indeed a Lyapunov transformation.

$\square$

References

[1] B. C. Moore, "Principal Component Analysis in Linear Systems: Controllability, Observability, and Model Reduction," IEEE Trans. on Automatic Control, Vol. AC-26, No. 1, Feb. 1981, pp. 17-32.

[2] B. C. Moore, "Singular Value Analysis of Linear Systems, Parts I, II," Dept. Electr. Eng., Univ. of Toronto, Ont. Syst. Contr. Rep 7801 and 7802, July 1978.

[3] L. M. Silverman, S. Shokoohi, P. Van Dooren, "Balanced Realization and Approximation of Time-Variable Systems," 4th International Symp. on Math. Theory of Networks & Systems, Delft, 1979.

[4] E. I. Verriest, "On Balanced Realizations for Time Variant Linear Systems," Ph.D. Dissertation, Dept. of Electrical Engineering, Stanford University, Stanford, CA. 1980.

[5] S. Lefschetz, "Differential Equations: Geometric Theory." Dover 1977.

[6] E. Hille, "Analytic Function Theory," Vol. I. Chelsea, 1976.

[7] T. Kato, Perturbation Theory for Linear Operators, Springer-Verlag, 1976.

[8] L. M. Silverman, and H. E. Meadows, "Controllability and Observability in Time-Variable Linear Systems," J. SIAM Control, 5(1):1967.

[9] C. T. Mullis, and R. A. Roberts, "Symthesis of Minimum Roundoff Noise Fixed Point Digital Filters," IEEE Trans. CS, CAS-23(9):551-562, 1976.

[10] S. S. Sastry, "Near Uncontrollability and Near Unobservability of Linear Systems," M.S. Thesis, Univ. of Calif., Berkely, Feb. 1979.

[11] L. Pernebo, and L. M. Silverman, "Balanced Systems and Model Reduction," Proc. 18th Conference on Decision and Control, Fort Lauderdale, FL., Dec., 1979.

[12] T. Kailath, Linear Systems, Prentice Hall, Englewood Cliffs, New Jersey, 1980.

[13] R. E. Kalman, "Contribution to the Theory of Optimal Control," Bol. Soc. Mat. Mex., pp. 102-119, 1960.

[14] E. I. Verriest, "Bounds on the Pole Location for Reduced Order Balanced Systems," in preparation.

[15] J. L. Willems , Stability Theory of Dynamical Systems, John Wiley & Sons, N.Y., 1970.

[16] L. M. Silverman, and B. D. O. Anderson, "Controllability, Observability and Stability of Linear Systems, J. SIAM on Control, 6:121-130, 1968.

[17] E. I. Verriest, "Low Sensitivity Design and Optimal Order Reduction for the LQG-Problem," Proc. 24th Midwest Symp. on Circuits & Systems, Albuquerque, N.M., June 1981, pp. 365-369.

[18] E. A. Jonckheere and L. M. Silverman, "A New Set of Invariants for Linear Systems - Application to Approximation," Int. Symp. on Math. Theory of Networks and Systems, Santa Monica, CA., August 1981.

[19] L. M. Silverman, "Synthesis of Impulse Response Matrices by Internally Stable and Passive Realizations," IEEE Trans. on Circuit Theory, CT-15(3):238-245, Sept. 1968.

[20] R. E. Kalman, P. L. Falb, and M. A. Arbib, Topics in Mathematical System Theory, McGraw-Hill, 1969.

[21] E. I. Verriest, and T. Kailath, "On Generalized Balanced Realizations," Proc. of the 19th Conference on Decision and Control, Albuquerque, N.M., Dec. 1981, pp. 504-505.

[22] D. G. Luenberger, "Optimization by Vector Space Methods," J. Wiley & Sons, 1968.

[23] S. Y. Kung, "A New Identification and Model Reduction Algorithm via Singular Value Decompositions," Proc. 12th Asilomar Conf. Circuits, Syst., Comp., Pacific Grove, CA., pp. 705-714, Nov. 1978.

[24] S. Shokoohi, L. Silverman, and P. Van Dooren, "Stability of Balanced Time-Variable System Approximations," Proc. 1980 IEEE Conference on Decision and Control, Albuquerque, N.M.

[25] S. Shokoohi, L. Silverman, and P. Van Dooren, "Stable Approximation of Time Variable Systems," Intl. Fed. of Automatic Control, Kyoto, Japan, Aug. 1981.

REACHABILITY, OBSERVABILITY AND DISCRETIZATION

Erik I. Verriest
School of Electrical Engineering
Georgia Institute of Technology
Atlanta, GA 30332

21st CDC

This work was supported by the National Science Foundation under Grant ECS-81-05509.

Abstract:

It is shown that a commonly used discretization method leads to a certain infidelity regarding the reachability and observability properties of the original versus the discretized version. In a digital implementation this infidelity is amplified due to round off errors. In particular, the reduced order modeling is shown to be very sensitive. A new discretization method of high fidelity is proposed. As an application the commutativity of reduced order modeling and discretization is studied for balanced realizations and a criterion for determining the stepsize corresponding to a certain degree of fidelity is presented.

1. Introduction and background material.

In many cases a mechanical, electrical or chemical process, plant or system is modeled by a set of dynamic equations in continuous time, especially in those instances where the physical insight underlies the modeling. In general one may have a time varying nonlinear set of differential equations. Linearization about a steady state nominal operating point, or a nominal trajectory leads then to a set of linear differential equations which are time-varying in general.

In this paper we shall only consider process models described by time invariant linear differential equations

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx \end{aligned} \tag{1.1}$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, $y(t) \in \mathbb{R}^p$ are respectively the state vector, input vector and output vector, and

$$A \in \mathbb{R}^{nxn}, \qquad B \in \mathbb{R}^{nxm}, \qquad C \in \mathbb{R}^{pxn}$$

Without loss of generality we may assume that the realization (1.1) is minimal. Further, the assumption that A is stable is made.

For the purpose of estimation and/or control via a digital computer, one needs to convert the given continuous time system model (1.1) into a set of difference equations, say

$$\begin{aligned} z_{k+1} &= \Phi z_k + \Gamma u_k \\ y_k &= Hz_k + Du_k \end{aligned} \tag{1.2}$$

Consequently, this discrete model is then used in the control-law design. Many discretization schemes are known, depending on the sample-and-hold procedure used. Further, a whole class of equivalent state space realizations correspond to any given input-output behavior. Finally the stepsize in the discretization may be variable.

In this paper we shall motivate what we mean by a "best" discretization scheme, and actually show its performance in the context of control, estimation, digital filtering and model reduction.

Rather than directly requiring a match between input and output of the original and the discretized system, we shall look at their internal properties. For this purpose, the gramian formalism is thought to be a good approach.

Suppose that for the description (1.1) the initial state is zero. It is then well known from optimal control theory that in order to reach the state x_0 as $t \to \infty$, the input needs an energy of at least

$$||u||_{L_2}^2 \geq x_0' \, C^{-1} \, x_0 \tag{1.3}$$

where C is the reachability gramian of the system (1.1).

$$C = \int_0^\infty e^{At} BB' e^{A't} dt \tag{1.4}$$

Equality holds in (1.3) iff $u(t) = u_0(t)$ the optimal control minimizing the L_2-norm

$$||u||_{L_2} = \left(\int_0^\infty u'u dt \right)^{1/2}$$

while meeting the constraint $x(\infty) = x_0$. For the discretized model (1.2) one can find similarly that, in order to reach z^o, the input-"energy" is bounded by

$$||u||_{L_2}^2 \geq z^{o\prime} C_d^{-1} \, z^o \tag{1.5}$$

where now

$$C_d = \sum_{k=0}^{\infty} \Phi^k \Gamma\Gamma' \Phi'^k \tag{1.6}$$

and equality holds in (1.5) iff u_k is the optimal control minimizing the l_2-norm

$$||u_k||_{\ell_2} = \left(\sum_{k=0}^{\infty} u_k' \, u_k\right)^{1/2}$$

and meeting the constraint $z_\infty = z^0$. Hence, in both cases the inverse of the reachability gramian is a weighting matrix for the computation of the minimal reachability cost or energy. Alternatively, a stochastic interpretation can be given to the gramians. If either system is perturbed by white noise with unit covariance, then the state covariance after all transients died out is equal to the reachability gramian. Thus C and C_d are a quantitative measure for the disturbability of the states for the continuous and discretized system respectively. A good discretization scheme should neither introduce nor destroy reachability or disturbability, in order to keep the same evaluation of the control cost. Hence one should look for a discretization scheme for which C_d is close to C in some sense.

Analogously, one can also look at the properties of the state-to-output mapping. For a stable system with initial state x_0 and no forcing function, the total output energy, as measured by the L_2 norm is

$$||y||_{L_2} = x_0' \mathcal{O} x_0 \tag{1.7}$$

where

$$\mathcal{O} = \int_0^{\infty} e^{A't} C C e^{At} dt \tag{1.8}$$

is the observability gramian of (1.1).

For the discretized system (1.2) one finds

$$||y||_{l_2} = z^{o\prime} O_d z^o \tag{1.9}$$

with the observability gramian

$$O_d = \sum_{k=0}^{\infty} \Phi'^k H'H\Phi^k \tag{1.10}$$

A stochastic interpretation of the observability gramians can also be given. Let for the undriven system, the output be corrupted by white noise, then the information (as measured by the inverse of the covariance matrix) conveyed in the output process about the initial state is respectively O and O_d for the continuous system (1.1) and the discretized system (1.2). Thus the observability gramians are a quantitative measure for the observability properties of the system. Again a good discretization scheme should neither create nor destroy observability, and O_d should be made as close to O as possible.

The role played by the gramians is further quintessential in the model reduction techniques based on the balanced realizations ([1], [2]). This technique actually also works for timevarying systems ([3],[4]). Here the input-to-state and state-to-output properties are "balanced" by a suitably chosen state-space transformation, so that disturbability of the state in a certain direction is made euqal to the observability of the state component in that direction. Thus for a balanced realization, one has equality of the reachability and observability gramian (which are then further made diagonal). In this way near-redundancy of state components can be detected as those directions x in the state space for which

$$\frac{x'Cx}{||x||^2} = \frac{x'Ox}{||x||^2}$$

is small compared to other directions. If $C = O = \Lambda = \text{diag}\ \{\lambda_i\}$, then the elements λ_i of the canonical gramian are interpreted as measures of the importance of the i-th dimension in the state space.

Thus in particular, when the original system (1.1) is given in balanced form, the discretized version (1.2) should conserve this property, if a discretization method with "high fidelity" in the reachability and observability properties is used. If then a reduced model for (1.1) is chosen, it automatically results in the corresponding reduced discrete model, for a suitably chosen stepsize.

In the context of digital filtering, it is further well known [5] that the balanced realizations play an important role in the minimization of finite word length effects (probability of overflow and roundoff noise). In fact it is shown in [5] that the balanced realization or principle axis realization minimizes these effects, although to the expense of requiring a larger number of multiplications. Hence, if for some discretization method the gramians are "conserved", then automatically "balancedness" will be conserved, and thus the parasitic noise effects due to the finite word length in the discrete model are minimal. Moreover, the following scheme is made commutative in some sense:

$$
\begin{tikzcd}
(A,B,C) \arrow[r, "\text{model}", "\text{reduction}"'] \arrow[d, "\text{discretization}"'] & (A_r,B_r,C_r) \arrow[d] \\
(\Phi,\Gamma,H) \arrow[r] & (\Phi_r,\Gamma_r,H_r)
\end{tikzcd}
$$

In what follows we shall also need the Lyapunov equations satisfied by the gramians (for strictly stable A and Φ). In the continuous time case C

and $\mathcal{O}$ solve respectively

$$AC + CA' + BB' = 0 \tag{1.11}$$

$$A'\mathcal{O} + \mathcal{O}A + C'C = 0 \tag{1.12}$$

while the discrete gramians C_d and $\mathcal{O}_d$ satisfy

$$\Phi C_d \Phi' + \Gamma\Gamma' = C_d \tag{1.13}$$

$$\Phi' \mathcal{O}_d \Phi + H'H = \mathcal{O}_d \tag{1.14}$$

The rest of the paper is organized as follows: In section 2, we define a measure for the fidelity of any discretization scheme, and show that a commonly used method has a low fidelity. A first improvement by scaling is suggested. A better high fidelity scheme is worked out in section 3, and it is shown that the fidelity is up to second order in Δ, the stepsize. Finally application to the model reduction techniques are discussed in section 4.

2. A Fidelity Measure for the Discretization Scheme.

To fix the ideas, let C and O be the reachability and observability gramians for the original continuous time model, and C_d and O_d those for any discretized version of the form (1.2). Motivated by our discussion in the previous section we define the discretization fidelities.

Definition 1: i) The fidelities of the discretization with respect to the reachability and observability properties are respectively

$$f_c = \exp\left\{ -\sqrt{\frac{\sum_i |\lambda_i(C) - \lambda_i(C_d)|^2}{\sum_i \lambda_i^2(C)}} \right\} \tag{2.1}$$

$$f_o = \exp\left\{ -\sqrt{\frac{\sum_i |\lambda_i(O) - \lambda_i(O_d)^2}{\sum_i \lambda_i^2(O)}} \right\} \tag{2.2}$$

ii) The overall discretization fidelity is

$$f = f_o f_c \tag{2.3}$$

Obviously the fidelities are between one and zero, and can be lower bounded by

Theorem 1: The fidelities are bounded by

$$\begin{aligned} f_c &\geq \exp\left\{ - \frac{||C - C_d||_F}{||C||_F} \right\} \\ f_o &\quad \exp\left\{ - \frac{||O - O_d||_F}{||O||_F} \right\} \end{aligned} \tag{2.4}$$

where $||\cdot||_F$ is the Frobenius (Schur or Hilbert-Schmidt) norm. If the

Frobenius norm of the difference $C - C_d$ and $O - O_d$ is sufficiently small, the following lower bounds are useful.

$$f_c \geq 1 - \frac{\|C - C_d\|_F}{\|C\|_F}$$
$$f_o \geq 1 - \frac{\|O - O_d\|_F}{\|O\|_F} \qquad (2.5)$$

Proof: (2.5) follows directly from (2.4) since for all x, the inequality $e^x \geq 1 - x$ holds. To show (2.4), the Wielandt-Hoffman inequality is involved

$$\sum_i |\lambda_i(O) - \lambda_i(O_d)|^2 \leq \|O - O_d\|_F^2$$

and the fact that for any Hermitian matrix X

$$\|X\|_F^2 = \sum_i |\lambda_i(X)|^2$$ ∎

A commonly used method for discrete representation of a continuous time linear time invariant system (1.1) is given by the zero order hold ([6], p. 135).

$$x_{k+1} = \Phi x_k + \Gamma u_k$$
$$y_k = C x_k \qquad (2.4)$$

where

$$x_k = x(k\Delta), \qquad y_k = y(k\Delta)$$
$$u_k = u(k\Delta), \qquad \Delta = \text{stepsize}$$

and

$$\Phi = e^{A\Delta} , \tag{2.5}$$

$$\Gamma = \int_0^\Delta e^{A\tau} d\tau \ B \tag{2.6}$$

For improved numerical properties, one can modify the above using the matrix exponential

$$\Phi = e^{A\Delta} = I + A\Delta + \frac{A^2\Delta^2}{2!} + \ldots$$

$$= I + A\Delta\psi \tag{2.7}$$

where

$$\psi = I + \frac{A\Delta}{2!} + \frac{A^2\Delta^2}{3!} + \ldots \tag{2.8}$$

and is usually evaluated by a series in the form

$$\psi_N \triangleq I + \frac{A\Delta}{2} (I + \frac{A\Delta}{3} (\ldots \frac{A\Delta}{N-1} (I + \frac{A\Delta}{N})) \ldots) \tag{2.9}$$

$$\psi_N \rightarrow \psi \text{ as } N \rightarrow \infty$$

The read-in matrix (2.6) becomes, using term by term integration

$$\Gamma = \psi \, \Delta B \tag{2.10}$$

The discrete time system (2.4) has reachability and observability gramians respectively given by

$$C_\Delta = \sum_{k=0}^{\infty} (I + A\Delta\psi)^k \psi\Delta BB'\Delta \psi'(I + \psi'\Delta A')^k \tag{2.11}$$

$$O_\Delta = \sum_{k=0}^{\infty} (I + \psi' \Delta A')^k C'C (I + A\Delta\psi)^k \tag{2.12}$$

and satisfying respectively the discrete time Lyapunov equations

$$(I + A\Delta\psi) C_\Delta (I + \psi'\Delta A') + \Delta^2 \psi BB'\psi' = C_\Delta \tag{2.13}$$

$$(I + \psi'\Delta A') O_\Delta (I + A\Delta\psi) + C'C = O_\Delta \tag{2.14}$$

or

$$(A\psi C_\Delta + C_\Delta \psi'A') + \Delta(\psi BB'\psi' + A\psi C_\Delta\psi'A') = 0 \tag{2.13'}$$

$$\Delta(\psi'A'O_\Delta + O_\Delta A\psi) + \Delta^2\psi'A'O_\Delta A\psi + C'C = 0 \tag{2.14'}$$

For small Δ, we consider the zeroth approximation of ψ (i.e. we only take the first term in the series $\psi \approx I$). Then we get for (2.13)

$$AC_\Delta + C_\Delta A' + \Delta BB' + \Delta A C_\Delta A' = 0$$

Expanding C_Δ in a series, we get

$$C_\Delta = C^{(0)} + \Delta C^{(1)} + \ldots \tag{2.15}$$

and

$$0 = A(C^{(0)} + \Delta C^{(1)}) + (C^{(0)} + \Delta C^{(1)})A' + \Delta BB' + \Delta A(C^{(0)} + \Delta C^{(1)})A'$$

or

$$AC^{(0)} + C^{(0)}A' = 0 \tag{2.17}$$

$$AC^{(1)} + C^{(1)}A' + BB' + AC^{(0)}A' = 0 \tag{2.18}$$

Since A is stable, (2.17) has only the zero solution, and then the first order term $C^{(1)}$ reduces to the reachability gramian of the continuous system, since the solution to (11) is unique.
Thus, up to first order in Δ we have

$$C_\Delta \simeq \Delta C \tag{2.19}$$

An analogous expansion of O_Δ

$$O_\Delta = O^{(0)} + \Delta O^{(1)} + \dots \tag{2.20}$$

yields

$$\Delta[A'(O^{(0)} + \Delta O^{(1)}) + (O^{(0)} + \Delta O^{(1)})A] + \Delta^2 A'(O^{(0)} + \Delta O^{(1)})A + C'C = 0$$

from which we obtain the inconsistency $C'C = 0$, unless we let $O^{(0)}$ be of the order of Δ^{-1}. In this case

$$A'O^{(0)} + O^{(0)}A + C'C = 0 \tag{2.21}$$

$$A'O^{(1)} + O^{(1)}A + A'O^{(0)}A = 0 \tag{2.22}$$

Again, by unity

$$O^{(0)} = O$$

and thus the first approximation to the discrete observability gramian is

$$O_\Delta = \frac{1}{\Delta} O \tag{2.23}$$

We find in (2.19) and (2.23) a discrepancy in the balance of the gramians in the continuous versus the discrete case, due to the appearance of the scalar factors Δ and $\frac{1}{\Delta}$ in the first term of he approximation. This can be worked away easily by rescaling the state variables. Indeed if we let $\bar{x}_k = \frac{1}{\sqrt{\Delta}} x_k$, then the discrete system

$$\begin{aligned} \bar{x}_{k+1} &= \Phi \bar{x}_k + \bar{\Gamma} u_k \\ y_k &= \bar{C} \bar{x}_k \end{aligned} \tag{2.24}$$

with $\bar{\Gamma} = \Gamma/\sqrt{\Delta}$ and $\bar{C} = \sqrt{\Delta}\, C$, has gramians matching those of the corresponding continuous system in the first approximation.

<u>Definition 2</u>: A discretization procedure will be called an <u>equilibrium discretization</u> if the gramians $\mathcal{C}_d$ and $\mathcal{O}_d$ reduce to $\mathcal{C}$ and $\mathcal{O}$ respectively as Δ approaches zero.

For the above developed equilibrium discretization, we can show after some algebra using matching expansions that up to third order:

$$\begin{aligned} \mathcal{C}_\Delta &= \mathcal{C} - \frac{\Delta^2}{12} A\mathcal{C}A' \\ \mathcal{O}_\Delta &= \mathcal{O} + \frac{\Delta}{2} C'C - \frac{\Delta^2}{12} (A'C'C + C'CA) \end{aligned} \tag{2.25}$$

Hence, for this particular discretization method, the reachability properties and the observability properties do not have an equal fidelity. Indeed, using theorem 1, we obtain (keeping the first terms only)

$$f_c \geq 1 - \Delta^2 a_c$$

$$f_o \geq 1 - \Delta a_o$$

where

$$a_c = \frac{1}{12} \|A\|_F^2$$

$$a_o = \|A\|_F^2$$

and we used the fact that $\|XY\|_F \leq \|X\|_F \|Y\|_F$ and the Liapunov equation (1.12). The overall discretization method has therefore a fidelity which deviates linearly from one.

3. A high fidelity zero order hold discretization scheme.

The discrepancy in the fidelity that was shown to exist in the usual zero order hold discretization method can be traced back to the non-symmetrical role of the input and the output operations. Indeed in (2.4), a weighting of the input $u(n\Delta)$ is chosen corresponding to the effect the continuous time system would have had with a constant input during an interval of length Δ. The output equation is however a "one shot" computation: the instantaneous output at the sampling interval is computed. Alternatively, we can compute an averaged output during the interval of length Δ, and hold this average output rather than the instantaneous output, as computed by (2.4).

Thus, again sampling $u(t)$ with period Δ, and using a zero order hold, we obtain as before $(0 \le \tau \le \Delta)$.

$$x(n\Delta + \tau) = e^{A\tau}x(n\Delta) + \int_0^{\tau} e^{A\theta}d\theta Bu(n\Delta)$$

$$= e^{A\tau}x(n\Delta) + A^{-1}(e^{A\tau} - I)Bu(n\Delta) \qquad (3.1)$$

Here we assumed that the continuous time system is strictly stable, so that all its poles are in the open left half plane. A is therefore nonsingular. The average value of the state in $[n\Delta, (n+1)\Delta)$ is then

$$\bar{x}_n = \frac{1}{\Delta}\int_0^{\Delta} x(n\Delta + \tau)d\tau$$

$$= \frac{1}{\Delta}\int_0^{\Delta} e^{A\tau}d\tau\; x(n\Delta) + \frac{1}{\Delta}A^{-1}\int_0^{\Delta}(e^{A\tau} - I)\, d\tau Bu(n\Delta)$$

$$= \frac{A^{-1}}{\Delta}(e^{A\Delta} - I)\; x(n\Delta) + A^{-1}\left[\frac{A^{-1}(e^{A\Delta} - I)}{\Delta} - I\right]Bu(n\Delta) \qquad (3.2)$$

The corresponding average output of the continuous time system is

$$\bar{y}_n = C\bar{x}_n$$

$$= \frac{CA^{-1}}{\Delta}(e^{A\Delta} - I)x(n\Delta) + CA^{-1}[\frac{A^{-1}(e^{A\Delta} - I)}{\Delta} - I]Bu(n\Delta) \tag{3.3}$$

As discussed in the previous section, we can properly renormalize the state equations by letting the discrete time state z_n correspond to

$$z_n = \frac{1}{\sqrt{\Delta}}\ x(n\Delta) \tag{3.4}$$

The equations (3.3) and (3.1) where we set τ equal to Δ are then rewritten as

$$\begin{aligned} z_{n+1} &= \bar{F}\, z_n + \bar{G}\, u_n \\ \bar{y}_n &= \bar{H}\, z_n + \bar{D}\, u_n \end{aligned} \tag{3.5}$$

where

$$\bar{F} = e^{A\Delta} \tag{3.6}$$

$$\bar{G} = \frac{A^{-1}(e^{A\Delta} - I)B}{\sqrt{\Delta}} \tag{3.7}$$

$$\bar{H} = \frac{C(e^{A\Delta} - I)A^{-1}}{\sqrt{\Delta}} \tag{3.8}$$

$$\bar{D} = CA^{-1}\ [\ \frac{A^{-1}(e^{A\Delta} - I)}{\sqrt{\Delta}} - I\ ]\ B \tag{3.9}$$

or, letting again as in the previous section

$$\Psi = I + \frac{A\Delta}{2!} + \frac{A^2\Delta^2}{3!} + \ldots \tag{3.10}$$

one obtains:

$$\bar{F} = I + A\Psi\Delta \tag{3.11}$$

$$\bar{G} = \sqrt{\Delta}\ \Psi B \tag{3.12}$$

$$\bar{H} = \sqrt{\Delta}\ C\Psi \tag{3.13}$$

and

$$\begin{aligned}\bar{D} &= CA^{-1}(\Psi - I)B \\ &= \Delta C\bar{\Psi}B\end{aligned} \tag{3.14}$$

where

$$\bar{\Psi} = \frac{I}{2!} + \frac{A\Delta}{3!} + \frac{A^2\Delta^2}{4!} + \ldots \tag{3.15}$$

can be accurately approximated by

$$\bar{\Psi}_N = \frac{1}{2}\left[\ I + \frac{A\Delta}{3}\left(I + \frac{A\Delta}{4}\left(\ldots\frac{A\Delta}{N}\left(I + \frac{A\Delta}{N+1}\right)\right)\ldots\right)\right] \tag{3.16}$$

It is now obvious from equations (3.12) and (3.13) that $\bar{G}$ and $\bar{H}$ play symmetrical roles with respect to the original continuous time system. This is even more pronounced if one considers the Lyapunov equations satisfied by the reachability and the observability gramians respectively. Note that these gramians exist since, if A has all its eigenvalues in the open left half plane, then all the eigenvalues of $\bar{F}$ are strictly inside the unit circle. Letting C_Δ and $\mathcal{O}_\Delta$ be the reachability and observability gramians for the discretized system, we get (noting that A and Ψ commute)

$$(I + A\Psi\Delta)C_\Delta(I + A\Psi\Delta)' + \Delta\Psi BB'\Psi' = C_\Delta \tag{3.17}$$

$$(I + A\Psi\Delta)'\mathcal{O}_\Delta(I + A\Psi\Delta) + \Delta\Psi'C'C\Psi = \mathcal{O}_\Delta \tag{3.18}$$

or after some algebra and observing that Ψ is invertible (since $\bar{F}$ has no eigenvalues on the unit circle)

$$AC_\Delta \Psi^{-T} + \Psi^{-1} C_\Delta A' + BB' + \Delta A C_\Delta A' = 0 \tag{3.19}$$

$$A' O_\Delta \Psi^{-1} + \Psi^{-T} O_\Delta A + C'C + \Delta A' O_\Delta A = 0 \tag{3.20}$$

It is apparent that C_Δ and O_Δ will be perturbations of C and O respectively, hence the technique of matching expansions in Δ will be employed to obtain exact formulas for the perturbations. The result is stated as a theorem, the proof is deferred to the appendix.

First we show that Ψ^{-1} can be expanded in a converging sequence in Δ.

Lemma 1: The inverse of Ψ as defined in (3.10) can be expanded in the form

$$\Psi^{-1} = \sum_{i=0}^{\infty} \alpha_i A^i \Delta^i \tag{3.21}$$

where the α_i are recursively defined by

$$\frac{\alpha_0}{(m+1)!} + \frac{\alpha_1}{m!} + \frac{\alpha_2}{(m-1)!} + \dots + \frac{\alpha_{m-1}}{2!} + \frac{\alpha_m}{1!} = 0 \tag{3.22}$$

$$\alpha_0 = 1$$

proof: Direct multiplication of (3.10) and (3.21) yields

$$\Psi\Psi^{-1} = \sum_{i=0}^{\infty} \frac{A^i \Delta^i}{(i+1)!} \sum_{j=0}^{\infty} \alpha_j A^j \Delta^j$$

$$= \sum_{m=0}^{\infty} A^m \Delta^m \sum_{i=0}^{m} \frac{\alpha_i}{(m+1-i)!}$$

from which (3.22) follows.

$\square$

Lemma 2: The explicit solution of the equations (3.22) is

$$\alpha_m = \frac{1}{m!} \left(\frac{d}{dx} \right)^m \left. \frac{x}{e^x - 1} \right|_{x=0} \tag{3.23}$$

The sequence $\{\alpha_m\}$ converges to zero.

proof: Appendix.

$\square$

It follows then from lemma 2 that the series (3.21) is absolutely convergent, at least for

$$\Delta < ||A||^{-1} \, 2\pi \tag{3.24}$$

The first coefficients in the series are: $\alpha_0 = 1$, $\alpha_1 = -\frac{1}{2}$, $\alpha_2 = \frac{1}{12}$, $\alpha_3 = 0$, $\alpha_4 = -\frac{1}{720}$, $\alpha_5 = 0$. This suggests that the actual region of convergence for (3.21) may be much larger than that given in (3.24), but we have been unable to show this. We show also in the appendix that the coefficients α_i can be computed alternatively as the coefficients of the powers of x in the telescoping series

$$1 - \frac{1}{2}\,\theta(x) + \frac{1}{3}\,\theta^2(x) - \frac{1}{4}\,\theta^3(x) + \ldots$$

where

$$\theta(x) = \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$$

Now, we can state the

Theorem 2: The reachability and observability gramians $\mathcal{C}_\Delta$ and $\mathcal{O}_\Delta$ of the discretized system (3.5) can be expanded in a norm-converging series if $\Delta < 2\pi||A||^{-1}$ as:

$$\mathcal{C}_\Delta = \sum_{i=0}^{\infty} \Delta^i \mathcal{C}_\Delta^{(i)} \tag{3.25}$$

$$\mathcal{O}_\Delta = \sum_{i=0}^{\infty} \Delta^i \mathcal{O}_\Delta^{(i)} \tag{3.26}$$

where $\mathcal{C}_\Delta^{(0)}$ and $\mathcal{O}_\Delta^{(0)}$ are respectively $\mathcal{C}$ and $\mathcal{O}$, the gramians of the original continuous time system, and for $k \neq 0$, the perturbation terms are computed recursively from:

$$A\mathcal{C}_\Delta^{(k)} + \mathcal{C}_\Delta^{(k)}A' + A \sum_{i=2}^{k} \alpha_i (\mathcal{C}_\Delta^{(k-i)}A'^{i-1} + A^{i-1}\mathcal{C}_\Delta^{(k-1)})A' = 0 \tag{3.27}$$

$$A'\mathcal{O}_\Delta^{(k)} + \mathcal{O}_\Delta^{(k)}A + A' \underset{i=2}{\overset{k}{\Delta}} \alpha_i (\mathcal{O}_\Delta^{(k-1)}A^{i-1} + A'^{i-1}\mathcal{O}_\Delta^{(k-1)})A = 0 \tag{3.28}$$

where the coefficients α_i are as in lemma 2, and the last summation vanishes if k is 1.

proof: Appendix

$\square$

Application of theorem 1 yields

$$A\mathcal{C}_\Delta^{(1)} + \mathcal{C}_\Delta^{(1)}A' = 0$$

which has the unique solution $\mathcal{C}_\Delta^{(1)} = 0$ since A is strictly stable. The second order perturbation term is the solution to

$$A\mathcal{C}_\Delta^{(2)} + \mathcal{C}_\Delta^{(2)}A' + \frac{1}{12} A(\mathcal{C}A' + A\mathcal{C})A' = 0$$

which can be reduced using (1.11) to

$$A\mathcal{C}_\Delta^{(2)} + \mathcal{C}_\Delta^{(2)}A' - \frac{1}{12} ABB'A' = 0$$

which has as solution

$$\mathcal{C}_\Delta^{(2)} = -\frac{1}{12} \int_0^\infty e^{At}ABB'A'e^{A't}dt = -\frac{1}{12} A\mathcal{C}A'$$

The third term is zero since (2.27) yields

$$AC_\Delta^{(3)} + C_\Delta^{(3)}A' + A[\alpha_2(C_\Delta^{(1)}A' + AC_\Delta^{(1)}) + \alpha_3(C_\Delta^{(0)}A'^2 + A^2C_\Delta^{(0)})]A' = 0$$

while $\alpha_3 = 0$ and $C_\Delta^{(1)} = 0$

The fourth term solves, noting again that $\alpha_3 = 0$.

$$AC_\Delta^{(4)} + C_\Delta^{(4)}A' + A[\alpha_2(C_\Delta^{(2)}A' + AC_\Delta^{(2)}) + \alpha_4(C_\Delta^{(0)}A'^3 + A^3C_\Delta^{(0)})]A' = 0$$

The term in between the brackets can be reduced to

$$-\alpha_2^2(ACA'^2 + A^2CA') + \alpha_4(CA'^3 + A^3C)$$

$$= (\alpha_2^2 + \alpha_4)ABB'A' - \alpha_4A^2BB' - \alpha_4BB'A'^2$$

The solution is then

$$C_\Delta^{(4)} = \int_0^\infty e^{At}A[(\alpha_2^2 + \alpha_4)ABB'A' - \alpha_4A^2BB' - \alpha_4BB'A'^2]A'e^{A't}dt$$

$$= (\alpha_2^2 + \alpha_4)A^2CA'^2 - \alpha_4A^3CA' - \alpha_4ACA'^3$$

One can continue the process ad libidum. Up to the 4^{th} order we get thus:

$$C_\Delta = C - \frac{\Delta^2}{12}ACA' + \frac{\Delta^4}{720}(4A^2CA'^2 - A^3CA' - ACA'^3) + O(\Delta^5) \qquad (3.29)$$

Observing that equation (3.20) is the "dual" of (3.19) where A, Ψ, B, C_Δ are respectively replaced by A', Ψ', C', $\mathcal{O}_\Delta$ we obtain directly

$$\mathcal{O}_\Delta = \mathcal{O} - \frac{\Delta^2}{12}A'\mathcal{O}A + \frac{\Delta^4}{720}(4A'^2\mathcal{O}A^2 - A'^3\mathcal{O}A - A'\mathcal{O}A^3) + O(\Delta^5) \qquad (3.30)$$

The proposed discretization scheme is thus such that the reachability and observability properties of the original continuous time system - as displayed by the respective gramians - are inherited up to second order in Δ. Beyond this the perturbation terms display further a very nice symmetry.

The eigenvalues of the reachability gramian can be interpreted as the reciprocal of the minimal required energy associated with displacing the state in certain eigendirections [7]. Hence, a comparison of the eigenvalues of C and C_Δ will give an indication of the fidelity with regard to the reachability properties in the discretization. Similarly, the eigenvalues of the observability gramian are associated with the information conveyed by the observed output about the components of the initial state in certain eigendirections [4].

In what follows we shall assume that Δ is sufficiently small so that all terms of order 4 and above are negligible in (3.29) and (3.30). Thus, we let

$$C_\Delta = C - \varepsilon ACA'$$

$$O_\Delta = O - \varepsilon A'CA$$

where we replaced $\frac{\Delta^2}{12}$ by ε.

If the eigenvalues of C are

$$\lambda_1^c \geq \lambda_2^c \geq \ldots \geq \lambda_n^c$$

and those of C_Δ

$$\lambda_1^\varepsilon \geq \lambda_2^\varepsilon \geq \ldots \geq \lambda_n^\varepsilon$$

then it follows from the minimax theorem [8] that

$$\lambda_i^c - \varepsilon\lambda_{max}(ACA') \leq \lambda_i^\varepsilon \leq \lambda_i^c - \varepsilon\lambda_{min}(ACA') \qquad (3.31)$$

If the eigenvalues of ACA' are not known, then we can use the more conservative spectral norm

$$||ACA'|| \geq \lambda_{max}(ACA')$$

and obtain

$$\lambda_i^{c} - \varepsilon||ACA'|| \leq \lambda_i^{\varepsilon} \leq \lambda_i^{c} \qquad (3.31)$$

Thus all the eigenvalues of the discrete gramian are smaller than those of the continuous gramian by at most

$$\frac{\Delta^2}{12}||ACA'|| \leq \frac{\Delta^2}{12}||A||^2||C|| = \frac{\Delta^2}{12}\lambda_1^{c}||A||^2$$

Similarly, the eigenvalues $\{\bar{\lambda}_i^{\varepsilon}\}$ of the discrete observability gramian are bounded by

$$\lambda_i^{0} - \varepsilon||A'OA|| \leq \bar{\lambda}_i^{\varepsilon} \leq \lambda_i^{0} \qquad (3.32)$$

where

$$\lambda_1^{0} \geq \lambda_2^{0} \geq \ldots \geq \lambda_n^{0} > 0$$

are the ordered eigenvalues of O.

The "discretized" eigenvalues are all smaller than the corresponding eigenvalues of the continuous observability gramian by at most

$$\frac{\Delta^2}{12}||A'OA|| \leq \frac{\Delta^2}{12}||A||^2||O|| = \frac{\Delta^2}{12}\lambda_1^{0}||A||^2$$

The fidelities (2.5) can now easily be lower bounded.

Indeed from theorem 1, we obtain

$$\begin{aligned} f_c &\geq \exp\left(-\frac{\Delta^2}{12}\frac{||ACA'||_F}{||C||_F}\right) \\ &\geq \exp\left(-\frac{\Delta^2}{12}||A||_F^2\right) \\ &\geq 1 - \frac{\Delta^2}{12}||A||_F^2 \end{aligned}$$

and similarly

$$f_0 \geq \exp\left(- \frac{\Delta^2}{12} ||A||_F^2\right)$$

$$\geq 1 - \frac{\Delta^2}{12} ||A||_F^2$$

where we used the fact that $||XY||_F \leq ||X||_F||Y||_F$. Another lower bound can be obtained by working directly from the definition, and using the maximal deviation for the eigenvalues:

$$\frac{\Sigma|\lambda_i(C) - \lambda_i(C_\Delta)|^2}{\Sigma\lambda_i(C)^2} \leq \frac{n\left(\frac{\Delta^2}{12} \lambda_1^0 \; ||A||^2\right)^2}{\lambda_1^{0\;2}} = n\left(\frac{\Delta^2}{12} ||A||^2\right)^2$$

whence

$$f_c \geq \exp\left(- \sqrt{n} \frac{\Delta^2}{12} ||A||^2\right)$$

$$\geq 1 - \frac{\Delta^2}{12} \sqrt{n} \; ||A||^2$$

and similarly

$$f_0 \geq 1 - \frac{\Delta^2}{12} \sqrt{n} \; ||A||^2$$

The results can be combined in the following.

<u>Theorem 3</u>

The fidelities of the discretization (3.5) are bounded by

$$f_0, f_c \geq 1 - \frac{\Delta^2}{12} \min(||A||_F^2 \; , \; \sqrt{n} \; ||A||^2) \tag{3.33}$$

The overall discretization fidelity is bounded by

$$f(\Delta) \geq 1 - \frac{\Delta^2}{6} \min(||A||_F^2, \sqrt{n}\ ||A||^2) \tag{3.34}$$

up to third order.

Comments:

i) The bounds in the theorem can be made tighter by taking the exponential, rather than the first terms in the series expansion. The bounds obtained are however very tractable.

ii) Computationally, $||A||_F$ is easier to obtain than $||A||^2$.

iii) The overall fidelity is quadratic in Δ, as opposed to the linear dependence in the usual discretization scheme of section 2.

4. Applications to Digital Filtering and Model Reduction

A commonly used technique in the design of digital filters is based on the discretization of an equivalent continuous transfer function. Usually, the direct form (canonical) realizations are used because the state update requires at most n+1 multiplications for an n-th order filter. It was shown in [5] that these realizations do not have the best properties with regards to the effects of finite wordlength. Those effects are minimized in the "balanced realizations" for which the state update requires at most $n^2 + n$ multiplications. In some cases, this augmentation of the number of multiplications can trade off with a larger fidelity in mimicing the continuous transfer function. This then may allow a decrease in step size (sampling period) and hence a smaller number of multiplications per time unit than for the direct form realizations for the same fidelity. The details will be presented in a forthcoming paper, but the remainder is based on the following.

If the continuous time system were given in balanced form, then by definition [7]

$$\mathcal{C} = \mathcal{O} = \Lambda \tag{4.1}$$

where Λ is diagonal and called the canonical gramian [4]. Hence the discretized realization has the gramians

$$\mathcal{C}_\Delta = \Lambda - \frac{\Delta^2}{12} A\Lambda A' + 0(\Delta^4) \quad (\Delta \to 0) \tag{4.2}$$

$$\mathcal{O}_\Delta = \Lambda - \frac{\Delta^2}{12} A'\Lambda A + 0(\Delta^4) \quad (\Delta \to 0) \tag{4.3}$$

In general, the matrices $A\Lambda A'$ and $A'\Lambda A$ are not diagonal, so the discretized realization is no longer balanced, although the imbalancing perturbation terms are only of second order. Suppose one "rebalances" the discretized system, then it is of interest to know how the canonical gramian of the discrete system relates to the original one.

<u>Theorem 3</u>: If (A,B,C) is a continuous time balanced realization with canonical gramian Λ, then the canonical elements $\lambda_i(\Delta)$ of the discretized system (3.5) are for sufficiently small step-size Δ

$$\lambda_i(\Delta) = \lambda_i - \frac{\Delta^2}{24} \sum_k (a_{ki}^2 + a_{ik}^2)\, \lambda_k + O(\Delta^4) \qquad (4.4)$$

<u>proof</u>: The canonical elements are the square roots of the eigenvalues of

$$\mathcal{C}_\Delta \mathcal{O}_\Delta = (\Lambda - \frac{\Delta^2}{12}\ A\Lambda A' + O(\Delta^4))(\Lambda - \frac{\Delta^2}{12}\ A'\Lambda A + O(\Delta^4))$$

$$= \Lambda^2 - \frac{\Delta^2}{12}(A\Lambda A'\Lambda + \Lambda A'\Lambda A) + O(\Delta^4)$$

This can be diagonalized by an orthogonal transformation T which is almost the identity. Hence we set

$$T = I + \Delta^2 \zeta \qquad (4.5)$$

for some skew symmetric ζ (because $TT' = I$). Now find ζ such that

$$(I + \Delta^2\zeta)(\Lambda^2 - \frac{\Delta^2}{12}(A\Lambda A'\Lambda + \Lambda A'\Lambda A) + O(\Delta^4))(I - \Delta^2\zeta)$$

$$= \Lambda^2 + \Delta^2(\zeta\Lambda^2 - \frac{1}{12}(A\Lambda A'\Lambda + \Lambda A'\Lambda A) - \Lambda^2\zeta) + O(\Delta^4) \qquad (4.6)$$

is diagonal (up to third order in Δ), or

$$(\zeta\Delta^2 - \frac{1}{12}(A\Lambda A'\Lambda + \Lambda A'\Lambda A) - \Lambda^2\zeta)_{ij} = 0 \qquad i \neq j$$

$$(\lambda_j^2 - \lambda_i^2)\zeta_{ij} - \frac{1}{12}\sum_k (a_{ik}\lambda_k a_{jk}\lambda_j + \lambda_i a_{ki}\lambda_k a_{kj}) = 0$$

This equation determines ζ_{ij} if $\lambda_j \neq \lambda_i$. However, we do not need ζ_{ij} explicitly for $i \neq j$. Remark that $\zeta_{ii} = 0$ since ζ is skew symmetric.

The diagonal elements of (4.6) are then

$$\lambda_i^2 - \frac{\Delta^2}{12}\lambda_i \sum_k \lambda_k(a_{ik}^2 + a_{ki}^2) + O(\Delta^4)$$

$$= \lambda_i^2 (1 - \frac{\Delta^2}{12}\sum_k \frac{\lambda_k}{\lambda_i}(a_{ik}^2 + a_{ki}^2) + O(\Delta^4))$$

For sufficiently small Δ, the square root is

$$\lambda_i (1 - \frac{\Delta^2}{24}\sum_k \frac{\lambda_k}{\lambda_i}(a_{ik}^2 + a_{ki}^2) + O(\Delta^4))$$

from which (4.4) follows. ∎

Related to this problem is also the combined model reduction and discretization. The reduction technique for the continuous as well as the discrete system is based on the balanced realizations [7], the main idea of which goes as follows.

Suppose $(\bar{A},\bar{B},\bar{C})$ is balanced with canonical gramian Λ, containing large elements in the first p diagonal entries and small ones in the other n-p entries. If one partitions the given realization in

$$\begin{bmatrix} \bar{A}_{11} & \bar{A}_{12} \\ \bar{A}_{21} & \bar{A}_{22} \end{bmatrix} \quad \begin{bmatrix} \bar{B}_1 \\ \bar{B}_2 \end{bmatrix} \quad [\bar{C}_1 \quad \bar{C}_2]$$

consistent with

$$\Lambda = \begin{bmatrix} \Lambda_1 & \\ & \Lambda_2 \end{bmatrix} \begin{matrix} \updownarrow p \\ \updownarrow n-p \end{matrix}$$

then a good approximation of this system is given by the stable balanced realization (A_{11}, B_1, C_1) with canonical gramian Λ_1.

Let now (A,B,C) be any n-th order (not necessarily balanced) realization with canonical gramian Λ. We investigate the commutation of discretization and reduction.

First we reduce the continuous system via the method of balanced realizations. Say that the result is $(\bar{A}_{11}, \bar{B}_1, \bar{C}_1)$ of order $p < n$. Using theorem 3, it follows then that the discretization of this model leads to a system with canonical elements.

$$\lambda_i^I(\Delta) = \lambda_i - \frac{\Delta^2}{24} \sum_{k=1}^{p} (\bar{a}_{ki}^2 + \bar{a}_{ik}^2)\lambda_k \qquad i = 1,\ldots,p \tag{4.7}$$

Alternatively, one can first discretize (A,B,C) and then reduce. Invoking (3.23) and (3.29), the canonical elements of the discretized system are the square roots of the eigenvalues of the product (up to third order in Δ)

$$\mathcal{C}_\Delta \mathcal{O}_\Delta = \mathcal{C}\mathcal{O} - \frac{\Delta^2}{12} (A\mathcal{C}A'\mathcal{O} + \mathcal{C}A'\mathcal{O}A)$$

This formula is not so useful to put bounds on the eigenvalues since the terms in the righthand side are not symmetric. In order to remedy this situation, we symmetrize by finding the transformation T such $TC_\Delta T' = I$, then the elements fo the canonical gramian are equal to the square root of the eigenvalues of the symmetric matrix $T^{-T}O_\Delta T^{-1}$.

Further, expanding T as $T = (I + \Delta^2\Theta)C^{-\frac{1}{2}}$ where $C^{\frac{1}{2}}$ is a square root of C (i.e. $C = C^{\frac{1}{2}}\, C^{T/2}$), we get

$$(I + \Delta^2\Theta)C^{-\frac{1}{2}}(C - \frac{\Delta^2}{12} ACA')C^{-T/2}(I + \Delta^2\Theta') = I$$

from which the first perturbation term is

$$\Delta^2[\Theta - \frac{1}{12} C^{-\frac{1}{2}} ACA'C^{-T/2} + \Theta'] = 0 .$$

Only the symmetric part of Θ is of importance.

$$\Theta_S = \frac{1}{24} C^{-\frac{1}{2}} ACA'C^{-T/2}$$

for the first perturbation, hence we can set the asymmetrical part zero and approximate T^{-1} as

$$T^{-1} = C^{\frac{1}{2}}(I + \Delta^2\Theta_s)^{-1} = C^{\frac{1}{2}}(I - \frac{\Delta^2}{24} C^{-\frac{1}{2}} ACA'C^{-T/2})$$

We get finally

$$T^{-1}OT^{-1} = (I - \frac{\Delta^2}{24} C^{-\frac{1}{2}} ACA'C^{-T/2})C^{T/2}$$

$$(O - \frac{\Delta^2}{12} A'OA)C^{\frac{1}{2}}(I - \frac{\Delta^2}{24} C^{-\frac{1}{2}} ACA'C^{-T/2})$$

$$= C^{T/2}OC^{\frac{1}{2}} - \frac{\Delta^2}{24} (C^{-\frac{1}{2}} ACA'OC^{\frac{1}{2}} + 2C^{T/2} A'OAC^{\frac{1}{2}} + C^{T/2} OACA'C^{-T/2})$$

A general formula for the eigenvalues $\lambda_i(\Delta)$ of the above matrix cannot be given, but we get the following bounds

$$|\lambda_i^2 - \lambda_i^2(\Delta)| \le \frac{\Delta^2}{24} \|C^{-\frac{1}{2}} ACA'OC^{\frac{1}{2}} + 2C^{T/2} A'OAC^{\frac{1}{2}} + C^{T/2} OACA'C^{-T/2}\|$$

$$\le \frac{\Delta^2}{24} \left\{ \| C^{-\frac{1}{2}} ACA'OC^{\frac{1}{2}} \| + 2 \|C^{T/2} A'OAC^{\frac{1}{2}}\| + \|C^{T/2} OACA'C^{-T/2}\| \right\}$$

$$\le \frac{\Delta^2}{24} \left\{ 4 \|ACA'O\| \right\}$$

$$\le \frac{\Delta^2}{6} \|A\|^2 \|C\| \|O\|$$

Hence, for sufficiently small Δ

$$\lambda_i - \frac{\Delta^2}{12} \frac{\|A\|^2 \|C\| \|O\|}{\lambda_i} \le \lambda_i(\Delta) \le \lambda_i + \frac{\Delta^2}{12} \frac{\|A\|^2 \|C\| \|O\|}{\lambda} \qquad (4.9)$$

The perturbation is again quadratic in Δ as expected, but can only be bounded by a λ_i dependent term, whereas we get an equality in (4.7). Letting $\frac{1}{12} \|A\|^2 \|C\| \|O\|$ equal α, then each $\lambda_i(\Delta)$ is inside the sector with top at λ_i and which widens inversely proportional to λ_i (Fig. 1). For both cases, it follows that if $\lambda_i > \lambda_j$, then not necessarily $\lambda_i(\Delta) > \lambda_j(\Delta)$ which would indicate that the important dimensions of the discrete system and the continuous system do not correspond.

Suppose now that the original system was reduced from order n to order p. Thus, we assume that

$$\Lambda = \begin{bmatrix} \Lambda_1 & \\ & \Lambda_2 \end{bmatrix} \begin{array}{l} \updownarrow p \\ \updownarrow n-p \end{array}$$

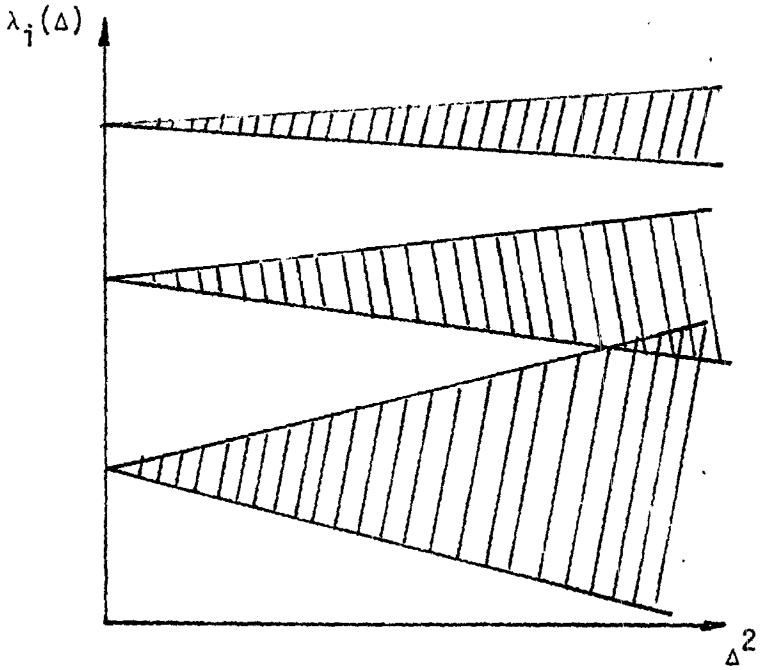


Figure 1. Bounds on the elements of the conoical gramian for the general reduced order realization.

where $\Lambda_1 > \Lambda_2$. Let λ_+ and λ_- be the least elements of Λ_1 and Λ_2 respectively. Then the p-th order optimal reduction of the discretized system will correspond to the discretization of the optimally reduced continuous system if

$$\lambda_+(\Delta) > \lambda_i(\Delta) \qquad i = p+1,\ldots,n \tag{4.10}$$

This is a condition on Δ.

Definition 3: The discretization and the model reduction are said to be compatible for steplength Δ if the elements of the set $\{\lambda_i(\Delta); i=1,\ldots,p\}$ are all larger than the elements of the set $\{\lambda_i(\Delta); i=P+1,\ldots,n\}$.

The bounds provided on the canonical elements $\lambda_i(\Delta)$ developed in (4.9) provide now a conservative scheme for computing the region of compatibility.

Theorem 4: A sufficient condition for the compatibility of the p-th order reduction and the discretization is that

$$\Delta \leq \sqrt{\frac{\lambda_+ - \lambda_i}{\lambda_+ - \lambda_i} \; \frac{\lambda_+\lambda_i}{\alpha}} \tag{4.11}$$

where the λ_i are the canonical elements of the (deleted) part Λ_2 and $\alpha = \frac{1}{12}$ $\|A\|^2 \; \|C\| \; \|O\|$.

proof: Using the bounds (4.9) we get

$$\lambda_+(\Delta) > \lambda_+ - \frac{\Delta^2\alpha}{\lambda_+}$$

and

$$\lambda_i(\Delta) < \lambda_i + \frac{\Delta^2\alpha}{\lambda_i} \qquad \text{for } i = p+1,\ldots,n.$$

For (4.10) to hold it is thus sufficient that

$$\lambda_+ - \frac{\Delta^2\alpha}{\lambda_+} \geq \lambda_i + \frac{\Delta^2\alpha}{\lambda_i}$$

or

$$\Delta^2 \leq \frac{\lambda_+ - \lambda_i}{\lambda_+ + \lambda_i} \; \frac{\lambda_+\lambda_i}{\alpha} \qquad i = p+1,\ldots,n \qquad \blacksquare$$

Remark that by bounding the righthand side of (4.11) we obtain the (more conservative) bound

$$\Delta \leq \sqrt{\frac{\lambda_+ - \lambda_-}{\lambda_+ + \lambda_-} \; \frac{\lambda_+\lambda_n}{\alpha}}$$

Remarks: The definition 3 is quite arbitrary, alternatively one could require that compatbility holds if also the ordering of the elements $\lambda_i(\Delta)$ and λ_i is kept for $i=1,\dots,p$, which is obviously a more restrictive condition.

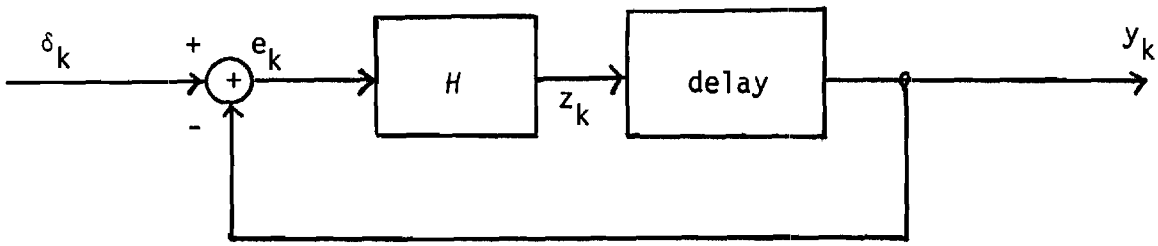
5. Conclusion.

A new ZOH-discretization scheme has been derived. It was shown that it "conserves" the reachability and the observability properties of the continuous system up to first order. The method can be applied for robust design of digital filters. In the application to model reduction an upper bound on the stepsize is given based on the order conservation of the canonical elements.

APPENDIX

A. Proof of lemma 2:

The sequence $\{\alpha_i\}$ can be generated as the return difference in the feedback system



where the system H has impulse response $h = \frac{1}{(k+1)!}$ with initialization $y_0 = 0$.

Then indeed $y_{k+1} = z_k = -\sum_{i=0}^{k} h_i(y_{k-i} + \delta_{k-i})$ and $e_0 = 1$, $e_i = y_i \quad i > 0$

satisfies the recursion (3.22). In the transform domain we get

$$H(z) = \sum_{i=0}^{\infty} \frac{z^{-i}}{(i+2)!} = z^2(e^{\frac{1}{z}} - z^{-1} - 1)$$

and

$$E(z) = \frac{1}{1 + z^{-1}H(z)} = \frac{1}{z(e^{\frac{1}{z}} - 1)} \tag{A.1}$$

The inverse z-transform and hence the coefficients α_i can then be identified from the expansion of (A.1) in negative powers of z, or letting $x = z^{-1}$, we obtain

$$E(x^{-1}) = \frac{x}{e^x - 1}$$

The Taylor expansion of this about $x = 0$ yields

$$E(x^{-1}) = \sum_{i=0}^{\infty} \frac{1}{i!}\left[\left(\frac{d}{dx}\right)^i \frac{x}{e^x - 1}\right]_{x=0} x^i$$

Hence

$$\alpha_i = \frac{1}{i!}\left[\left(\frac{d}{dx}\right)^i \frac{x}{e^x - 1}\right]_{x=0}$$

If we also set $e^x - 1 = \theta$, then $x = \ln(\theta + 1)$ and we get

$$E(x^{-1}) = \hat{E}(\theta) = \frac{\ln(\theta + 1)}{\theta} = 1 - \frac{1}{2}\,\theta + \frac{1}{3}\,\theta^2 + \ldots$$

where also:

$$\theta = \frac{1}{z} + \frac{1}{2!}\,\frac{1}{z^2} + \frac{1}{3!}\,\frac{1}{z^3} + \cdots$$

giving also a solution of α_i as the coefficients of z^i in the above telescoping series. These coefficients are also related to the Bernoulli-numbers, which appear in the theory of the regular prime numbers [9], B_n defined by

$$\frac{x}{e^x - 1} = \sum_{n=0}^{\infty} \frac{B_n}{n!}\, x^n$$

Thus $\alpha_n = \frac{B_n}{n!}$

From the known properties of the Bernoulli numbers [10] we obtain for $n \geq 1$

$$\alpha_{2n+1} = 0 \qquad \text{(A.2)}$$

$$\frac{2}{(2\;)^{2n}}\,\frac{1}{1 - 2^{1-2n}} > (-1)^{n+1}\,\alpha_{2n} > \frac{2}{(2\pi)^{2n}} \qquad \text{(A.3)}$$

Thus clearly $\alpha_{2n} \to 0$ for $n \to \infty$

B. Proof of Theorem 1:

Consider the equation (3.19) where we substitute (3.21) and (3.25)

$$A \sum_{i=0} \Delta^i C_\Delta^{(i)} \sum_{j=0} \alpha_j \Delta^j A'^j + \sum_{i=0} \Delta^i \alpha_i A^i \sum_{j=0} \Delta^j C_\Delta^{(j)} A' + BB' +$$

$$+ \sum_{i=0} \Delta^{i+1} A C_\Delta^{(i)} A' = 0$$

$$A \sum_{k=0} \Delta^k \sum_{i=0}^{k} C_\Delta^{(i)} A'^{(k-1)} \alpha_{k-1} + \sum_{k=0} \Delta^k \sum_{i=0}^{k} \alpha_i A^i C_\Delta^{(k-1)} A' +$$

$$+ BB' + \sum_{k=1} A \Delta^k C_\Delta^{(k-1)} A' = 0$$

$$(AC_\Delta^{(0)} + C_\Delta^{(0)} A' + BB') + \sum_{k=1} \Delta^k \Bigg[A \sum_{i=0}^{k} C_\Delta^{(i)} A'^{(k-i)} \alpha_{k-i} +$$

$$+ \sum_{i=0}^{k} \alpha_{k-1} A^{k-1} C_\Delta^{(i)} A' + AC_\Delta^{(k-1)} A' \Bigg] = 0$$

hence

$$AC_\Delta^{(0)} + C_\Delta^{(0)} A' + BB' = 0 \tag{B.1}$$

and

$$\sum_{i=0}^{k} (AC_\Delta^{(i)}A'^{(k-i)} + A^{k-i}C_\Delta^{(i)}A')\alpha_{k-i} + AC_\Delta^{(k-1)}A' = 0$$

$$k = 1, \ldots$$

or equivalently,

$$AC_\Delta^{(k)} + C_\Delta^{(k)}A' + \sum_{i=2}^{k} \alpha_i(AC_\Delta^{(k-i)}A'^{i} + A^{i}C_\Delta^{(k-i)}A') = 0 \qquad \text{(B.2)}$$

The exact solutions of (B.1) and (B.2) are, noting that $\alpha_{2k+1} = 0$ for $k = 1, 2 \ldots$

$$C_\Delta^{(0)} = C$$

$$C_\Delta^{(2k)} = \int_0^\infty e^{At} \sum_{i=1}^{k} \alpha_{2i}(AC_\Delta^{(2(k-i))}A'^{2i} + A^{2i}C_\Delta^{(2(k-1))}A')e^{A't}dt$$

$$C_\Delta^{(2k-1)} = 0$$

A bound on $C_\Delta^{(2k)}$ is obtained as

$$||C_\Delta^{(2k)}|| \leq$$

$$\sum_{i=1}^{k} ||A||^2|\alpha_{2i}| \; ||\int_0^\infty e^{At}(C_\Delta^{(2(k-i))}A'^{2i-1} + A^{2i-1}C_\Delta^{(2(k-i))})e^{A't}dt||$$

The integral is bounded by

$$\frac{2||A||^{2i-i}||C_\Delta^{(2(k-i))}||}{2\sigma_0}$$

where $\sigma_0 = \min |\mathrm{Re}\lambda(A)|$

hence, we obtain recursively

$$||C_\Delta^{(2k)}|| \leq \frac{1}{\sigma_0} \sum_{i=1}^{k} |\alpha_{2i}| \ ||A||^{2i}||C_\Delta^{(2(k-i))}||$$

$$\leq \frac{1}{\sigma_0} \sum_{i=1}^{k} \frac{4}{(2\pi)^{2i}} ||A||^{2i}||C_\Delta^{(2(k-i))}||$$

By induction it is then easily shown that

$$||C_\Delta^{(2k)}|| \leq \frac{||A||^{2k}}{\sigma_0^{\ k}} \ \frac{(4+\sigma_0)^{k-1}}{(2\pi)^{2k}} \ 2||C||$$

and hence that the series (3.24) converges for at least

$$\Delta < \frac{2\pi}{||A||} \ \frac{\sigma_0}{4+\sigma_0} < \frac{2\pi}{||A||}$$

Acknowledgement: The author appreciates the help from Dr. Marc Clements in recognizing the relation of the α_i's to the Bernoulli numbers.

References

[1] B. C. Moore, "Principal Component Analysis in Linear Systems: Controllability, Observability, and Model Reduction," IEEE Trans. on Automatic Control, Vol. AC-26, No. 1, Feb. 1981, pp. 17-32.

[2] L. Pernebo, and L. M. Silverman, "Balanced Systems and Model Reduction," Proc. 18th Conference on Decision and Control, Fort Lauderdale, FL., Dec. 1979.

[3] L. M. Silverman, S. Shokoohi, P. Van Dooren, "Balanced Realization and Approximation of Time-Variable Systems," 4th International Symp. on Math. Theory of Networks & Systems, Delft, 1979.

[4] E. I. Verriest, "On Balanced Realizations for Time Variant Linear Systems," Ph.D. Dissertation, Dept. of Electrical Engineering, Stanford University, Stanford, CA. 1980.

[5] C. T. Mullis, and R. A. Roberts, "Synthesis of Minimum Roundoff Noise Fixed Point Digital Filters," IEEE Trans. CS, CAS-23(9):551-562, 1976.

[6] G. F. Franklin, and J. D. Powell, "Digital Control of Dynamic Systems," Addison-Wesley, 1980.

[7] B. C. Moore, "Singular Value Analysis of Linear Systems, Parts I, II," Dept. Electr. Eng., Univ. of Toronto, Ont. Syst. Contr. Rep 7801 and 7802, July 1978.

[8] B. N. Parlett, "The Symmetric Eigenvalue Problem," Prentice Hall, 1980.

[9] E. Kummer, Journal für Math., 40, 130-138, 1850.

[10] Handbook of Mathematical Functions, U.S. Department of Commerce, 1964, M. Abramowitz and I. A. Stegun, eds., p. 804.

GEORGIA INSTITUTE OF TECHNOLOGY
SCHOOL OF ELECTRICAL ENGINEERING
ATLANTA, GEORGIA 30332

TELEPHONE: (404) 894- 2949

August 8, 1984

Dr. Michael P. Polis
Division of ECSE
National Science Foundation
Washington, D.C. 20550

Dear Dr Polis:

Please find enclose my final report on grant ECS 81-05509, "On Generalized Balanced Realizations and Applications to Model Reduction".

Sincerely,

Erik I. Verriest

EIV/db

APPENDIX VII

NATIONAL SCIENCE FOUNDATION
Washington, D.C. 20550

FINAL PROJECT REPORT
NSF FORM 98A

PLEASE READ INSTRUCTIONS ON REVERSE BEFORE COMPLETING

PART I—PROJECT IDENTIFICATION INFORMATION

1. Institution and Address	2. NSF Program	3. NSF Award Number
Georgia Institute of Technology School of E.E., Atlanta, GA 30332	Research Initiation	ECS-8105509
	4. Award Period From 7/1/81 To 12/31/83	5. Cumulative Award Amount 45,163

6. Project Title

On Generalized Balanced Realizations and Applications to Model Reduction

PART II—SUMMARY OF COMPLETED PROJECT *(FOR PUBLIC USE)*

The main theme of the research was the detailed study of balanced realizations and its various extensions and generalizations. The main objectives of this study were state space modeling and model reduction. In modeling one important issue is the reduction of the sensitivity by an appropriate choice of the system parameters. In model reduction, the emphasis is to find an optimum way to approximate a given model by a lower dimensional one (i.e. with a smaller parameter set). Of course, optimality can only be specified with explicit reference to some performance (index). Our attention focussed on retaining the "energies" or "uncertainties" appearing in the model (i.e. second order moments) since they are directly linked to the eigenvalues of certain gramians, and hence mathematically tractable.

A comprehensive theory on the structure of balanced realizations for time-varying systems has been completed.

In the time invariant case the properties of multivariable balanced realizations was investigated, and a more direct balancing transformation (modulo an orthogonal matrix) have been obtained for the siso-case. Formulas for explicit solutions of Lyapunov equations were obtained.

All these notions have been applied to point out a discrepancy of a common discretization method, and an improved method was given. While this work was in progress, some results on the inverse "continuization" of a discrete process were obtained as well.

Further extensions to infinite dimensional systems and Volterra Systems were investigated. In the latter case, it was found that model reduction based on balancing techniques are not necessarily optimal with regard to the aforementioned energy or uncertainty properties.

Finally, a general theory on stochastic model reduction unifies the Desai-Pal and Arun-Kung techniques, resolving an ambiguity problem.

PART III—TECHNICAL INFORMATION *(FOR PROGRAM MANAGEMENT USES)*

1. ITEM *(Check appropriate blocks)*	NONE	ATTACHED	PREVIOUSLY FURNISHED	TO BE FURNISHED SEPARATELY TO PROGRAM: Check (√)	TO BE FURNISHED SEPARATELY TO PROGRAM: Approx. Date
a. Abstracts of Theses		Jose-Ramos			
b. Publication Citations		x			
c. Data on Scientific Collaborators					
d. Information on Inventions					
e. Technical Description of Project and Results					
f. Other *(specify)* Extended Report + papers		x			

2. Principal Investigator/Project Director Name *(Typed)*	3. Principal Investigator/Project Director Signature	4. Date
Erik I. Verriest		8/8/84

NSF Form 98A (3-83) Supersedes All Previous Editions

Form Approved OMB No. 3145-0058

Final Report for NSF Grant No. ECS-8105509, "On Generalized Balanced Realizations and Applications to Model Reduction", (covering July 1, 1981 to December 31, 1983).

Significant progress has been made during the first year of the above research grant.

An investigation of the special properties of the "classical" balanced realizations (i.e. for time invariant stable systems) has lead to a simplified characterization of these realizations. An exhaustive analysis of the second order case gave some insight in the relation between order reduction based on dominance and order reduction via the balanced realization technique.

Further results have been obtained in the extension of the balanced realization concept to time varying systems. These results are reported in:

E. I. Verriest and T. Kailath, "On Generalized Balanced Realizations", IEEE Trans. On Automatic Control, Vol. AC-28, No. 8, August 1983, pp. 833-844.

As we investigated time varying systems, as a side result - we obtained some results involving the stability of such systems. This evolved separately to a conference paper:

E. I. Verriest, "Stability of Linear Time-Varying Systems via Differential Inequalities", 22nd IEEE Conf. on Decision and Control, San Antonio, Texas, December 1983.

In order to practically test the usefulness of the proposed model reduction for time varying systems, a software package was developed, as part of a graduate project.

An important new direction of the research is the application of the balancing concept to the closed-loop system. This results in what we called the LQG-balanced realizations and is based on the underlying Riccati equations rather than the (open loop) Lyapunov equations. This is a crucial advance, however many problems remain open (e.g. stability of the reduced order model). These results were reported in:

E. I. Verriest, "Suboptimal LQG-Design via Balanced Realizations", Proceedings of the 20th IEEE Conference on Decision and Control, pp. 696-687, San Diego, California 1981.

Also for the time invariant case an LQG-reduction program is developed.

Important new insight in the fidelity of discrete applications of continuous systems was gained by using concepts of balanced realizations. Applications to the design of digital filters were obtained as well. In particular, it was shown that with the "usual" discretization procedures input and output properties were not conserved. Our method yields a fidelity up to second order in the stepsize. These results are reported in:

E. I. Verriest, " Reachability-Observability and Discretization", accepted for the 21st IEEE Conference on Decision and Control", Orlando, Florida, December 1982. This paper is also being reviewed for publication in the Transactions on Automatic Control.

The application to the design of digital filters is presented in:

E. I. Verriest, "Digital Filter Design based on a High Fidelity Discretization Procedure", 17th Annual Conference on Information Science and Systems, Johns Hopkins, March 1983.

During the 2nd year of the research, the work continued on several fronts:
i) Connecting to the work on discretization, we have looked at the inverse problem, namely that of modeling a discrete system by a process, continuous in time. It is shown that an n-th order discrete system can always be "continuized" by a minimal real system of order, possibly higher than n, but not exceeding 2n. This theory is then applied towards the equivalence of continuous and discrete Liapunov equations, hence again tying in directly with our main objective: the study of balanced realizations. It should also be noted that this procedure is further very significant in certain interpolation methods. The results are written up in:

E. I. Verriest, "The Matrix Logarithm and the Continuization of a Discrete Process", under review IEEE Transactions on Circuits and Systems.

ii) The structure of scalar time invariant balanced realizations (in particular the sign symmetry) is well known. The structural properties have for multi input - multi output balanced realizations has been analyzed. Unfortunately, not such nice properties as for the siso case exist. These results were discussed in an invited conference paper.

E. I. Verriest, "The Structure of Multivariable Balanced Realizations", International Symposium on Circuits and Systems, Newport Beach, California, May 1983.

The same paper also discusses a more direct way of obtaining the balancing transformation (modulo an orthogonal transformation) of a siso system, thus avoiding the dual singular value decomposition of the original approach of Moore. Also some explicit formulas for the Lyapunov equations in terms of reachability or observability matrices and what I have termed the "canonical gramian" of the characteristic polynominal are given. As indicated, this gramian is completely determined by the coefficients of the characteristic polynominal. A more detailed version of this conference paper will be written in the near future and submitted to the IEEE Transactions on Automatic Control.

iii) Another interesting question is the relation of model reduction based on balanced realizations and the classical method based on dominance. The results of this study are not yet quite at a satisfactory level, and some further work in this area is desired. We approached the problem from the inverse direction, i.e. given a balanced matrix A. Imbed this matrix into a larger one (by bordering) which is still balanced. The relation of the eigenvalues of the matrices are investigated and bounds are established. The results of this continued study will evolve in a more matrix-theory oriented paper: "The Eigenstructure of Balanced Matrices".

iv) In the area of stochastic modeling, we have tied together the stochastic balancing method proposed by Desai and Pal (canonical correlations) and the method of Arun and Kung based on the one-sided Karkunen Loeve expansion (principal component analysis). A dispute about which one is optimal is resolved by providing optimality of both methods, but with respect to different contraint sets. The unifying framework is the so-called RV-coefficient method by R. Escouffier and the redundancy analysis of Van Wollenberg. Preliminary results are reported in the invited paper:

J. Ramos and E. I. Verriest, "A Unifying Tool for Comparing Stochastic Realization Algorithms and Model Reduction Techniques", Proc. American Control Conference, San Diego, California, June 1984.

This unification fits further nicely in the more abstract framework of Gleason-measures on the logic of subspaces of a Hilbert space. This is under further investigation and a resulting publication will be forthcoming.

v) A study of the applicabiality of model reduction techniques via balanced realizations for infinite dimensional systems has begun. Only a simple parabolic system (the heat equation in a homogeneous medium) has been investigated. Two different approximations were used. A straight forward discretization, as used for the numerical solution of parabolic PDE's and the eigen function expansion. Both solutions converge to the exact operation solution as N, the dimensionality of the approximation, tends to infinity. The work slowed considerably as some unexpected difficulties arose, and is as yet not terminated.

vi) The germinal ideas on balancing in general and its use in model reduction spawned some new ideas towards the robust design of control systems.

The main idea is that while reducing a given deterministic model, part of the informaiton (i.e. some state components) is thrown away, thus resulting in some uncertainty. For this reason, any reduced model should have this uncertainty built into it: i.e., the reduced model of a deterministic system should be stochastic. The information loss should be conserved in the uncertainty in the stochastic model!

This must definitely be true in the finite approximation of infinite dimensional system. Similarly the effect of the nonlinearities of a mild nonlinear system could be modeled as a stochastic input, leading to more robust modeling on one hand, and to quantitative criteria for selecting weighting matrices and covariance matrices in the general LQG problem. A proposal to continue the research in this direction is forthcoming.

vii) Finally, an extension of the theory of balancing transformations to Volteria systems was proposed based on a tensor space formalism. Only some conservative bounds for the reduced orders could be given. This is inherent in the tensor space approach. This together with material on I/O approximation for certain nonlinear systems is presented in

E. I. Verriest, "Approximation and Order Reduction in Nonlinear Models using an RKHS-approach", 18th Annual Conference on Information Sciences and Systems, Princeton University, March 1984.

"A Generalized Stochastic Realization Theory with Applications to Multivariate Streamflow Modeling and Optimal Control of Water Resource Systems"

Jose Ramos, Graduate Student Civil Engineering
Dissertation Advisor: Srinivas G. Rao, co-advisor E. I. Verriest

Abstract (Preliminary)

The joint problem of multivariate modeling of streamflows and optimal control of water resource systems in a particular river basin is approached in this dissertation from a stochastic realization point of view.

The streamflow modeling problem is formulated as that of finding linear transformations (basis vectors) of the forward and backward predictor spaces of a vector stochastic process. These basis vectors will then be the states of a forward and backward Kalman filter, respectively. A new method of solution is presented, namely the RV-coefficient method which in turn yields two previously developed algorithms as specific cases. When the two Kalman filters have equal and diagonal state covariance matrices, the system is said to be in balanced form and a reduced-order model can be easily obtained. It is shown that the RV-coefficient method is a general tool for solving the stochastic realization problem in that when applied to unbalanced systems, a balancing transformation can be found.

The optimal control problem is solved as an LQG problem via the separation principle, having a forward Riccati equation for the Kalman filter and a backward Riccati equation for the optimal control law. Since this pair of equations is not in balanced form, the RV-coefficient method is applied in order to find the balancing transformation and hence the reduced-order model which can be viewed as some form of model aggregation. Applications of this methodology to a specific river basin are under way.