from which it follows that $r=0$. As a consequence, $\det[\lambda I - M(s)]$ has degree n_0 in λ, which means that $n_0=m$, and hence $k_{n_0}=k_m=0$. This contradicts the conclusion that $k_{n_0}>k_0$, which means that the case cannot occur.

3) $n_0 \geq 1$, and $u_{0k}>0$ for one value of $k \in \{1,2,\cdots,n_0\}$. Corresponding to this u_{0k} we obtain an nth-order Butterworth configuration, yielding n closed-loop poles at once. Consequently $q=0$ and $r=0$. Since $r=0$, $\det[\lambda I - M(s)]$ is of degree n_0 in λ, so that $n_0=m$, which in turn implies $k_{n_0}=k_m=0$. Since u_{0k} can be nonzero for one value of k only (otherwise several nth-order patterns of closed-loop poles would result, which is impossible), we evidently have $\det[\lambda I - M(s)]=-a\lambda^{m-1}+\lambda^m$, with $a>0$. Unless $m=1$, this implies $\det[M(s)]=0$, which is contrary to assumption. The case $m=1$, in which a single nth-order Butterworth pattern is obtained, corresponds to the single-input case where $H^T(-s)QH(s)=c/\phi(s)\phi(-s)$, with c a constant.

Summarizing, we have demonstrated that if $n_0=0$, one or several Butterworth patterns are obtained. The case $n_0=1$, which results in a single nth-order Butterworth pattern, only occurs when in the single-input case $H^T(-s)QH(s)=c/\phi(s)\phi(-s)$, with c a constant. The case $n_0>1$ does not occur.

References

[1] S. S. L. Chang, *Synthesis of Optimum Control Systems*. New York: McGraw-Hill, 1961.
[2] R. E. Kalman, "When is a linear control system optimal?" *Trans. ASME* (J. Basic Eng.), ser. D, vol. 86, pp. 51–60, 1964.
[3] J. S. Tyler and F. B. Tuteur, "The use of a quadratic performance index to design multivariable control systems," *IEEE Trans. Automat. Contr.*, vol. AC-11, pp. 84–92, Jan. 1966.
[4] H. Kwakernaak and R. Sivan, "Asymptotic pole locations of time-invariant linear optimal regulators," in *Proc. 3rd Annu. Southeastern Symp. System Theory*, Atlanta, GA, Apr. 5–6, 1971.
[5] —, *Linear Optimal Control Systems*. New York: Wiley-Interscience, 1972.
[6] W. M. Wonham, *Linear Multivariable Control: A Geometric Approach* (lecture notes in economics and mathematical systems), vol. 101. Berlin: Springer, 1974.
[7] G. Rosenau, "Höhere Wurzelortskurven bei Mehrgrössensystemen," *IFAC Symp. Multivariable Systems, Preprints*, Oct. 7–8, 1968.
[8] L. Weinberg, *Network Analysis and Synthesis*. New York: McGraw-Hill, 1962.
[9] F. R. Gantmacher, *The Theory of Matrices*, vol. I. New York: Chelsea, 1959.

On Optimal and Suboptimal Actuator Selection Strategies

Y. VANBEVEREN AND M. R. GEVERS, MEMBER, IEEE

***Abstract*—This short paper studies a particular class of optimization problems dealing with the selection, at each instant of time, of one out of many actuators in order to obtain a determined result. A cost is associated with each actuator. The cost function is the integral of a weighted combination of the achieved accuracy on the state of the system and the control energy. The control energy term depends upon both the selected actuator and the magnitude of the applied control. The problem is to design an optimal actuator selection strategy. The analysis is limited to the class of linear deterministic systems with measurable states. A discrete approach is considered. The analytic solution to this optimization problem is given first. When the number of actuators and the number of stages in the time interval become large the optimal analytic solution requires a considerable combinatorial work; a suboptimal algorithm is then proposed to alleviate this defect.**

I. Introduction

The problem of selecting, at each instant of time, one out of many available actuators is presently untreated in the literature. There are, however, applications in which several different or incompatible actions can be applied on a process. Classes of examples are: problems with a bottleneck (such as hierarchical systems in which a single line is to transmit different effects having the same potentialities to the various subsystems), or problems with different zones for the control (e.g., a gearbox). In this last example the problem is both to select the best gear and to determine the pressure on the accelerator.

Some aspects of the dual problem on the optimal selection of sensors have been solved by Athans [1], Herring and Melsa [2], and Bensoussan [3].

Athans [1] has considered the determination of optimal costly measurement strategies in the case of finite-dimensional systems. At each instant during a time interval, one out of a finite number of sensors must be selected to minimize a payoff that depends on two terms: the accumulated observation cost and the prediction accuracy at final time. The accumulated prediction error cost is not considered.

Herring and Melsa [2] have generalized these results to allow the selection at each instant of time of the best combination of a finite number of sensors. The payoff depends on the observation cost as before, but also on the accuracy of prediction at each instant of the time interval considered.

Bensoussan [3] has extended Athans' results (but with different methods) to infinite-dimensional spaces in order to optimize the location of sensors in a distributed parameter system. He uses the same payoff as Athans. Aidarous, Gevers, and Installé have derived a numerically implementable algorithm for the optimal allocation of sensors [7] and actuators [8] in a distributed parameter system.

In this short paper the problem of designing an optimal actuator selection strategy is solved using the optimality principle. The cost function is not the dual of any of the measurement strategy problems mentioned above, since it includes an instantaneous cost depending upon both the chosen actuator and the control energy. The problem is stated in Section II, and the N-stage optimization problem is solved in Section III. Two criteria are presented for the *a priori* elimination of certain "bad" sequences. For the remaining sequences the solution depends on the initial state. For a long time interval (N large) or a large amount of actuators, the computational effort required to find the optimal actuator policy can become prohibitive. Therefore, a suboptimal algorithm has been developed that drastically reduces the computation time. This "forward-backward" algorithm is presented in Section IV. All the simulations performed so far show that the "forward-backward" algorithm is near optimal; some numerical results are given in Section V.

II. Problem Statement

Consider a time-invariant linear dynamic system

$$X(i+1)=AX(i)+BU(i) \tag{1}$$

where X is an $n\times 1$ state vector and U is an $mq\times 1$ control vector. A and B are $n\times n$ and $n\times mq$ matrices. B will be represented as follows:

$$B=[b_1\ b_2\ \cdots\ b_m]$$

where b_j is an $n\times q$ matrix corresponding to the jth actuator. m actuators are available, but only one actuator can be used at any given time. Therefore,

$$U(i)=\begin{bmatrix} u_1(i) \\ u_2(i) \\ \\ \\ u_m(i) \end{bmatrix} \in U, \quad U \triangleq \left\{ \begin{bmatrix} u \\ 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ u \\ 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix}, \cdots, \begin{bmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ u \end{bmatrix} \right\}, \quad u \in R^q. \tag{2}$$

Hence, if the jth actuator is chosen at time i, $u_j(i)$ may take any real value, while $u_k(i)=0$, $k\neq j$.

It is assumed that each pair $[A,b_j]$ is completely controllable, and that the state $X(i)$ is exactly measurable. The cost function to be minimized for a N-stage problem is

$$J_N=\sum_{i=0}^{N-1}[X'(i+1)QX(i+1)+U'(i)RU(i)] \tag{3}$$

Manuscript received March 13, 1975; revised January 14, 1976. Paper recommended by D. L. Kleinman, Chairman of the IEEE S-CS Optimal Systems Committee.

Y. Vanbeveren was with Universite Catholique de Louvain, Louvain-la-Neuve, Belgium. He is now with the Centre d'Automatique, Fontainebleau, France.

M. R. Gevers is with the Universite Catholique de Louvain, Louvain-la-Neuve, Belgium.

where Q is a positive definite symmetric matrix and R is a positive definite block-diagonal matrix; $R=\text{diag}\{r_1,\cdots,r_m\}$, with r_j a $q\times q$ symmetric positive definite matrix.

At each instant of time, the optimal regulator must therefore decide what actuator should be used, and what value should be given to the control. Intuitively one can expect the following results.

Certain actuators or sequences of actuators may be rejected once and for all because whatever the initial state may be, they increase the payoff more than others.

For the remaining p actuators or sequences of actuators the regulator divides the state space into p zones: $S_1,S_2,\cdots,S_p$. If the initial state belongs to the zone S_j, then the sequence S_j will be chosen.

III. N-Stage Problem

The payoff to be minimized is now given by (3). We shall first compute the global cost for a predetermined sequence. The fixed actuator case, in which no selection is to be made, is classically solved using Bellman's optimality principle. In the present case, the formulation is not much different. We shall call $V_{N-i}[j_i]$ the optimal cost resulting from the optimization of the last $N-i$ stages of an N-stage problem with a given sequence $[j_i]$[1] of chosen actuators. Therefore, by the classical theory of optimal control

$$V_{N-i}[j_i]=X'(i)M_i[j_i]X(i) \tag{4}$$

where

$$\begin{cases} M_i[j_i]=A'W_{i+1}[j_{i+1}]A-A'W_{i+1}[j_{i+1}]b_{j_i} \\ \quad x\left(b'_{j_i}W_{i+1}[j_{i+1}]b_{j_i}+r_{j_i}\right)^{-1}b'_{j_i}W_{i+1}[j_{i+1}]A, \\ \quad i=0,\cdots,N-1 & \text{(5a)} \\ W_i[j_i]=M_i[j_i]+Q & \text{(5b)} \\ W_N=Q. & \text{(5c)} \end{cases}$$

In the same way

$$u_i[j_i]=-K_{j_i}[j_i]X(i) \tag{5d}$$

with

$$K_{j_i}[j_i]=\left(b'_{j_i}W_{i+1}b_{j_i}+r_{j_i}\right)^{-1}b'_{j_i}W_{i+1}A. \tag{5e}$$

Clearly the optimal choice of an actuator at the ith stage depends on the future behavior of the system, and so does the regulator gain. For a given sequence $[j_0]$ the optimal global cost can be written as

$$V_N[j_0]=X'(0)M_0[j_0]X(0). \tag{6}$$

For the N-stage problem, the optimal strategy consists in selecting a sequence of N actuators. In most cases the optimal sequence will be a function of the initial state $X(0)$. But it may happen that certain sequences or subsequences can be eliminated *a priori*, because they are dominated by others that are less expensive, whatever the initial state may be.

From (6) it follows immediately that in a N-stage problem the sequence $(k_0,\cdots,k_{N-1})$ can be rejected *a priori* [i.e., for all $X(0)$] if there exists a sequence $(j_0,\cdots,j_{N-1})$ such that

$$M_0(k_0,\cdots,k_{N-1})\geqslant M_0(j_0,\cdots,j_{N-1}).$$

But what can we say in an N-stage problem about an r-subsequence that can be rejected *a priori* in an r-stage problem? The following rules provide an answer to this question.

Lemma: In a one-stage optimization problem the kth actuator can be eliminated *a priori* if there exists $j\neq k$ such that

$$b_k r_k^{-1}b'_k\leqslant b_j r_j^{-1}b'_j. \tag{7}$$

[1]*Notation:* In this N-stage problem $[j_i]$ refers to the sequence $(j_i,j_{i+1},\cdots,j_{N-1})$ of chosen actuators.

Proof: For $N=1$, and using the matrix inversion lemma, (6) can be rewritten as

$$V_1(j)=X'(0)A'\left[Q^{-1}+b_j r_j^{-1}b'_j\right]^{-1}AX(0).$$

Therefore the kth actuator can be rejected *a priori* if, for some $j\neq k$,

$$\left[Q^{-1}+b_k r_k^{-1}b'_k\right]^{-1}\geqslant\left[Q^{-1}+b_j r_j^{-1}b'_j\right]^{-1}.$$

This is equivalent with (7). ■

Roughly speaking the term $b_k r_k^{-1}b'_k$ is the ratio of the "power of the kth actuator" to the cost resulting from its use. Intuitively condition (7) means that if the cost of the kth actuator is large and its influence on the state of the system is weak, then this actuator can be suppressed. This criterion should be compared with the dual condition obtained by Bensoussan [3] in the measurement context, in which the quantity $b_k r_k^{-1}b'_k$ is replaced by what Athans [1] interprets as a "signal-to-noise" ratio related to the kth measurement device.

Proposition 1: Let the sequence $(k_0,\cdots,k_{r-1})$ be eliminated *a priori* in an r-stage problem, $1\leqslant r\leqslant N$. Then any sequence terminating by the subsequence $(k_0,\cdots,k_{r-1})$ can be eliminated *a priori* in an N-stage problem.

Proof: The whole cost to go from stage 0 to stage N may be broken up into the cost to go from stage 0 to $N-r+1$ plus the cost to go from stage $N-r+1$ to N. Proposition 1 is then stated using the principle of optimality [4]. Indeed if a sequence is eliminated for all possible initial states in an r-stage optimization problem, it is *a fortiori* eliminated in the r last stages of an N-stage problem. ■

Proposition 2: In an N-stage problem one can eliminate *a priori* any sequence containing an actuator that would be eliminated in a one-stage problem. (i.e., if the kth actuator is eliminated *a priori* in a one-stage problem, then any sequence containing k can be *a priori* eliminated in an N-stage problem, whatever the position of k in this sequence).

Proof: Assume that the "bad" actuator k appears at the $(N-r+1)$th position of an N-stage sequence, and that the last r actuators selected in this N-sequence are $(k,j_1,j_2,\cdots,j_{r-1})$. We shall show that all N-sequences ended by $(k,j_1,j_2,\cdots,j_{r-1})$ can be eliminated. We first show that this r-sequence can be rejected *a priori* in an r-stage problem. By hypothesis, there exists a $j\neq k$ such that (7) is true. Therefore,

$$\left[\left(M_1(j_1,\cdots,j_{r-1})+Q\right)^{-1}+b_k r_k^{-1}b'_k\right]^{-1}\geqslant\left[\left(M_1(j_1,\cdots,j_{r-1})+Q\right)^{-1}+b_j r_j^{-1}b'_j\right]^{-1}.$$

Equivalently,

$$V_r(k,j_1,\cdots,j_{r-1})\geqslant V_r(j,j_1,\cdots,j_{r-1}).$$

Therefore, $(k,j_1,\cdots,j_{r-1})$ can be eliminated in an r-stage problem. By Proposition 1 it can be eliminated in the last r stages of an N-stage problem, which concludes the proof. ■

Assume that after *a priori* elimination of certain sequences or subsequences, p admissible sequences remain, with $p\leqslant m^N$. The selection of an optimal sequence of actuators among these p admissible sequences is now a funtion of the initial state $X(0)$. The state space is divided into subspaces; the sequence $[k_0]=(k_0,\cdots,k_{N-1})$ is optimal if $X(0)$ belongs to the subspace S_{k_0} defined as

$$S_{k_0}=\left\{X(0)\,|\,X'(0)M_0[k_0]X(0)\leqslant X'(0)M_0[j_0]X(0),\,[j_0]\neq[k_0]\right\}.$$

The separation surfaces are pieces of hyperplanes.

The selection of an optimal sequence for the state $X(0)$, therefore, requires the solution of a combinatorial problem, that can be solved by the "decision tree" method, since the number of admissible sequences is finite. In the worst case, the solution of an N-stage optimal control problem leads to the examination of m^N different sequences. For a large number of actuators or a long time interval, the computation time could become prohibitive. It was, therefore, necessary to develop a suboptimal but much faster algorithm.

IV. The "Forward-Backward" Algorithm

The control problem is now considered as a discrete dynamic programming problem. The N-stage problem consists in finding the sequence $\{U(0), U(1),\cdots,U(n-1)\}$ that minimizes the payoff (3). Equivalently, it is required to find the functional solutions of the following set of equations:

$$\begin{cases} V(X(i),i) = \min\limits_{U(i)} \{F(X(i+1),U(i)) + V(X(i+1),i+1)|X(i)\}, \\ \quad i=0,\cdots,N-1 \qquad (8) \\ V(X(N),N)=0 \qquad (9) \end{cases}$$

subject to the constraint (1) with a given $X(0)$, and with

$$F(X(i+1),U(i)) \triangleq X'(i+1)QX(i+1)+U'(i)RU(i).$$

Let us recall that two types of information are necessary to construct $U(i)$, the selected actuator, say j_i, and the value of the gain, say K_{j_i}; the pair $\{j_i,K_{j_i}\}$ completely determines $U(i)$ for a given $X(i)$.

To solve (8) and (9) recursively, we apply an iterative "forward-backward" algorithm that is closely related to Bellman's "approximation in policy space" procedure [5].

Given an initial approximation $\{U^0(0),U^0(1),\cdots,U^0(N-1)\}$ for the control policy, the state trajectory $\{X^0(1),\cdots,X^0(N)\}$ is computed using (1) and the given initial state $X(0)$. A new control sequence $U^1(i)$, $i=N-1, N-2,\cdots,0$ is then computed backwards by applying N times the following one-step minimization procedure.

$$\min_{U(i)} \{F(X(i+1),U(i)) + V^0(X^0(i+1),i+1)|X^0(i)\}, \quad i=N-1,\cdots,0 \qquad (10)$$

with

$$\begin{cases} V^0(X^0(N),N)=0 \qquad (11) \\ V^0(X^0(i),i)=[F(X(i+1),U^1(i)) + V^0(X^0(i+1),i+1)|X^0(i)]. \qquad (12) \end{cases}$$

The minimizing control sequence $\{U^1(0),\cdots,U^1(N-1)\}$ defines a new state trajectory $\{X^1(1),\cdots,X^1(N)\}$, and, by application of the backward minimization procedure, a new sequence $V^1(X^1(i),i)$, and so on. The forward-backward algorithm therefore requires two steps that may be repeated until no significant decrease in the cost function is obtained:

—a backward minimization step that consists in computing the $U^k(i-1)$ in terms of the $V^{k-1}(X^{k-1}(i),i)$.

—a forward reconstruction step that consists in computing the state trajectories $X^k(i)$ in terms of $U^k(i-1)$; these state trajectories are necessary to compute the new $V^k(X^k(i),i)$.

The practical implementation of the "forward-backward" algorithm in the actuator selection problem is as follows.

1) Initialization step: Starting from the given $X(0)$, a first sequence of pairs $\{j_i^0,K_{j_i}^0\}$, $i=0,\cdots,N-1$, is obtained by solving N times a one-stage optimization problem, as shown in Section III. This defines a first state trajectory $\{X^0(1),\cdots,X^0(N)\}$.

2) Backward step: The following step is performed N times in the backward sense starting with $W_N^1=Q$· given $X^0(i)$, j_i^1, and $K_{j_i}^1$ are computed using the one-stage minimization procedure of Section III, with Q replaced by $W_{i+1}^1[j_{i+1}^1]$; $M_i^1[j_i^1]$ and $W_i^1[j_i^1]$ are then computed from $W_{i+1}^1[j_{i+1}^1]$ and $K_{j_i}^1$ using (5). Notice that this requires only one iteration of the Riccati equation (5).

3) Forward step: Using the pairs $\{j_i^1,K_{j_i}^1\}$ a new state trajectory $\{X^1(1),\cdots,X^1(N)\}$ is computed by solving (1) in the forward sense.

Steps 2) and 3) are repeated until the pairs $\{j_i^k,K_{j_i}^k\}$ converge to a stable sequence $\{j_i^*,K_{j_i}^*\}$.

Numerical computations have shown that this algorithm converges very fast, and that the number of iterations necessary for the sequence to converge to the stable sequence $\{j_i^*,K_{j_i}^*\}$ is always smaller than N. The convergence can be accelerated by using the results of Proposition 2. Lew [6] has shown that a bounded perturbation of discrete functional equations of the form (8)–(9) yields a bounded change in its solution. This is, as Lew points out, a stability property; approximation methods can therefore be expected to result in bounded errors. It is reasonable to expect that the sequences $V^k(X(0),0)$ will be closer and closer to the optimal cost function, but this can not, in general, be proved. The algorithm is therefore suboptimal. Notice that if the cost function converges close to the optimal cost, it does not imply that the selected actuator strategy converges to the optimal sequence. However, in most cases this is unimportant, since the only objective is to minimize the cost.

If the optimal actuator sequence is reached, then, by construction, the optimal regulator gains are also obtained, and the strategy is optimal.

TABLE I

Strategies Considered	Global cost
1) Four successive "one-stage problems" are considered; the actuator selection is optimized at each stage.	60 081
2) The suboptimal "forward-backward" strategy is used.	39 723
3) The optimal strategy is used (in this case, $6^4=7776$ different sequences must be considered).	39 698

TABLE II

	3 Stages	4 Stages	12 Stages
Suboptimal algorithm	0.03 min	0.04 min	0.20 min
Optimal algorithm	0.13 min	0.90 min	?

V. Numerical Example

In this section, the suboptimal "forward-backward" algorithm is compared with the optimal solution and with an other strategy. The following system has been simulated:

$$X(i+1)=\begin{bmatrix} -1 & 1 & -1 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} X(i) + \begin{bmatrix} 1 & 1 & 0 & 3 & 1 & 4 \\ 0 & 2 & 1 & -2 & 1 & 0 \\ 1 & -1 & 5 & 2 & 6 & 1 \end{bmatrix} U(i),$$

i.e., there are six possible input vectors $b_1,\cdots b_6$. It is desired to design a four-stage control strategy with a scalar control, that minimizes the following payoff:

$$J=\sum_{i=1}^{4} (X'(i)QX(i)+U'(i-1)RU(i-1))$$

with $Q=\text{diag}\{3,5,2\}$ and $R=\text{diag}\{20,15,15,5,20,10\}$. The initial state of the system is $X(0)=(-1,+50,+60)$. Several strategies have been considered. (See Table I.)

The difference between the costs of the two last strategies is only 0.063 percent, which is negligible.

Table II compares the computation times necessary for a 3-stage, 4-stage, and 12-stage problem, respectively, with the optimal and the suboptimal algorithms, using an IBM 370/158 computer.

The time saving is appreciable when the suboptimal algorithm is used. It increases of course with the number of stages. From the viewpoint of precision, the comparison with other nonoptimal strategies shows that considerable savings can be achieved when using the suboptimal "forward-backward" algorithm as is evidenced by Table I. Several other numerical simulations have shown this to be true.

VI. Conclusion

The problem of the optimal selection of one out of many available actuators has been treated. Our study leads to optimal control strategies that can be obtained only by combinatorial methods. It has been shown how certain actuators or sequences of actuators can sometimes be rejected *a priori*. However, when the number of actuators or the length of the time interval become large, the "curse of dimensionality" makes the

use of combinatorial methods all but impossible. For such cases a suboptimal "forward-backward" algorithm has been proposed, which has been shown to be computationnally very attractive. It is much faster than other algorithms that have been proposed for the solution of similar optimal selection problems.

References

[1] M. Athans, "On the determination of optimal costly measurement strategies for linear stochastic systems," *Automatica*, vol. 18, pp. 397–412, July 1972.

[2] K. D. Herring and J. L. Melsa, "Optimum measurements for estimation," *IEEE Trans. Automat. Contr.*, vol AC-19, pp. 264–266, June 1974.

[3] A. Bensoussan, "Optimization of sensors' location in a distributed filtering problem," in *Proc. Int. Symp. Stability of Stochastic Dynamic Systems*, July 1971, pp. 62–84.

[4] J. S. Meditch, *Stochastic Optimal Linear Estimation and Control*. New York: McGraw-Hill, 1969.

[5] R. Bellman, *Adaptive Control Processes*. Princeton, NJ: Princeton Univ. Press, 1961.

[6] A. Lew, "A predictor-corrector method for dynamic programming," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 54–56, Feb. 1974.

[7] S. E. Aidarous, M. Gevers, and M. Installe, "Optimal sensors' allocation strategies for a class of stochastic distributed parameter systems," *Int. J. Contr.*, vol. 22, no. 2, pp. 197–213, 1975.

[8] ——, "Optimal pointwise discrete control and controllers' allocation strategies for stochastic distributed systems," *Int. J. Contr.*, to be published.

The Asymptotic Behavior of Constant-Coefficient Riccati Differential Equations

T. KAILATH, FELLOW, IEEE, AND L. LJUNG, MEMBER, IEEE

***Abstract*—A simple and self-contained proof is given of a general theorem on the convergence of a constant coefficient Riccati differential equation to a unique limiting value. In particular our result, which includes (strictly) previous results, does not require any analysis of the algebraic Riccati equation.**

I. Introduction

The behavior as $t\to\infty$ of the solution $P(t)$ of the (filtering) Riccati equation with constant coefficients,

$$\frac{d}{dt}P(t)=FP(t)+P(t)F^T-P(t)H^THP(t)+GG^T, \qquad t\geqslant 0 \quad (1)$$

with initial condition

$$P(0)=\Pi \quad (2)$$

has been the object of considerable investigation since the appearance of the first results of Kalman and Bucy. The reason is the importance of such results in ensuring the numerical well behavior of the estimation and control problems in which the solution of the Riccati equation is used. Our theorem in this short paper strictly includes all previous results on this problem and is established in a direct and self-contained manner. Perhaps the best known earlier results are those of Wonham [1], from whose work it follows that if

$$(H,F) \text{ is detectable} \quad (3)$$

and

$$(F,G) \text{ is stabilizable,} \quad (4)$$

then

$$\text{for all } \Pi\geqslant 0 \quad (5)^1$$

$$P(t)\to\bar{P} \text{ as } t\to\infty \quad (6)$$

where $\bar{P}$ is the unique nonnegative definite solution of the so-called *algebraic Riccati equation* (ARE)

$$0=F\bar{P}+\bar{P}F^T-\bar{P}H^TH\bar{P}+GG^T. \quad (7)$$

Wonham actually had the stronger assumption than (3) that

$$(H,F) \text{ is observable,} \quad (8)$$

the weakening to (3) being due to Kucera [2]. Now while the condition (5) that $\Pi\geqslant 0$ is of course very reasonable since Π is the covariance of the initial error, due to numerical errors it may happen that Π is not necessarily nonnegative definite, and therefore it is of interest to investigate convergence for more general initial conditions. Here the best results seem to be those of Willems [3] (see also Rodriguez-Canabal [4] and Bucy and Rodriguez-Canabal [5]). Under the stronger assumptions (8) and

$$(F,G) \text{ is controllable} \quad (9)$$

Willems [3, theorem 8] showed that convergence took place for all

$$\Pi>\bar{P}_- \quad (10)$$

where (with the usual ordering relationship for symmetric matrices)

$$\bar{P}_- = \text{the infimum over all solutions to the ARE (7).} \quad (11a)$$

(Canabal [5], defines a quantity

P_- = the infimum of all matrices P such that

$$FP+PF^T-PH^THP+GG^T\geqslant 0. \quad (11b)$$

If P_- exists it follows [5] that it is equal to $\bar{P}_-$.)

In view of (10), (11) it was natural that Willems' proof relied heavily on a close study of the properties of the ARE, but actually so did the proofs of Wonham and Kucera for the case $\Pi\geqslant 0$. In this short paper we shall present a more direct proof of convergence of $P(\cdot)$ that does not first require a close study of the limiting solution $\bar{P}$. This is not only philosophically more satisfying, but as will become clear in this short paper, it also enables us to obtain somewhat more general results than can be obtained by the methods of [3], [4]. In particular we can handle situations where P_- may not exist (as, for example, if (H,F) is not observable). Incidentally, we note that our results also apply to somewhat more general Riccati equations than (1), where, in particular, GG^T is replaced by a possibly indefinite matrix Q. To simplify the presentation, however, discussion of these extensions is deferred to Section III. Furthermore, we believe that our proof is simpler than any known so far and uses some simple identities that further elaborate the properties of the Riccati equation (1).

Three Useful Identities

The first of these identities is just the statement of the fact that if we know a solution of the Riccati equation (RDE) for one initial condition, then under certain conditions solutions for other initial conditions can be expressed in terms of the first solution. This approach to studying a family of solutions to the RDE in terms of one "special solution" is apparently due to Sandor [6] and Reid [7], though of course it had long been widely used to study scalar Riccati equations. The precise result is the following.

Let $P(\cdot)$ and $P_1(\cdot)$ be solutions of (1) for initial conditions Π and Π_1.

Manuscript received February 10, 1975; revised January 23, 1976. Paper recommended by E. J. Davison, Chairman of the IEEE S-CS Computational Methods and Digital Systems Committee. This work was supported in part by the Air Force Office of Scientific Research, AF Systems Command, under Contract AF 44-620-69-C-0101 and in part by the Joint Services Electronics Program under Contract N-00014-67-A-0112-0044.

T. Kailath is with the Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305.

L. Ljung was with the Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA. He is now with the Department of Automatic Control, Lund Institute of Technology, Lund, Sweden.

[1] $A > B$ means that $A-B$ is nonnegative definite.