

https://dspace.org/

The optimal control of infinite dimensial linear systems

Jones, Elwyn Lloyd

2015-04-21
http://hdl.handle.net/10539/17472

Downloaded from a DSpace repository. https://dspace.org/

THE OPTIMAL CONTROL
OF INFINITE DIMENSIONAL LINEAR SYSTEMS

BY E. L. JONES

Submitted by the author in part fulfillment of the requirements for the degree of M.Sc.(Eng.) at the University of the Witwatersrand, Johannesburg.

June 1976

I, Elwyn Lloyd Jones, hereby declare that this dissertation is my own work and that it has not previously been submitted to any University or Institute of Learning for any purpose whatsoever.

11th June 1976.

CONTENTS

II. Preface

IV. Acknowledgement

1. Introduction

7. Early attempts at the formulation and solution of a Linear Optimal Control Problem

11. The Linear Quadratic Problem and the Kalman-Bucy Solution

15. An alternative formulation - the Limited State Linear Regulator Problem

17. Conclusion

19. Key References

APPENDIX A : On the Solution of Wiener-Hopf type Equations

APPENDIX B : On the solution of the Matrix Valued Algebraic Riccati Equation

APPENDIX C : On the Solution of the Operator Valued Algebraic Riccati Equation

APPENDIX D : On the Linear Quadratic Problem with Complexity Constraints

Preface:

The work to be covered by the ensuing dissertation was largely motivated by two problems encountered by the author while working in practice.

Firstly at the H.R.U* in Stellenbosch, Cape. Here, the unit was concerned with the physical modelling of coastal regions, with the expressed purpose of determining the geomorphological effects of man-made structures such as harbours or marinas. A weir was constructed "off shore" of the model, and its height was controlled by means of a small thyristor controlled D.C. machine. The problem was to control the D.C. machine in such a way that the resultant "tidal effect" at some measurement spot in the model corresponded as closely as possible to the observed tide at the corresponding measurement station in the real world.

The second problem was encountered at number ten ammonium nitrate plant at A.E. & C.I.**, Modderfontein. After doing a reliability-, and an approximate dynamic-analysis of the plant, the author came to the conclusion that the only effective means of detecting a dangerous situation, for alarm and shut-down purposes, was through reactor temperature. However, for structural reasons, the sensors (therme-couples) were heavily clad in metal. As a result, the rise time was uncomfortably long. The problem was to design a filter to optimally compensate for the sensor sluggishness.

A third problem had been encountered by the author in his B.Sc.(Eng.) thesis on electronically steered micro-wave antennae. The problem was to control the relative excitation of two sources off the focus of a parabolic reflector, in such a way that the far-field radiation pattern suitably scanned the horizon.

*Hydraulic Research Unit of the Mechanical Engineering Research Institute of the Council for Scientific and Industrial Research of South Africa.

**African Explosives and Chemical Industries.

These systems have one thing in common, and that is that they cannot be described by means of ordinary differential equations. In Engineering parlance, they are referred to as Distributed Parameter Systems, however in this dissertation they will be referred to by the slightly more general name of Infinite Dimensional Systems. Further, attention will be restricted to linear systems. The objective then is to explore the principles and methods of determining the optimum controller (or filter) for Infinite Dimensional Linear Systems; three methods of approaching this problem have been isolated.

The first of these methods is due to Wiener[1], and was developed during the war effort to improve gunnery control, unfortunately it is not generally taught at our universities and those text-books that do mention it appear to under-emphasize its elegance, and generality. The reader is referred to Appendix A for some of the author's suggestions on using this technique.

The second approach is due to Kalman[3], and was originally devised for use on Finite Dimensional Linear Systems encountered in the space programme. The author has explored the possibility of extending these techniques to Infinite Dimensional Linear Systems. A review of the theory appears in Appendix C , while an efficient algorithm devised by the author appears in Appendix B.

The third approach takes into account some of the practical difficulties that the two previous approaches ignore. The original work was started by Levine and Athans[9], and it has been extended by the author in Appendix D.

Recommendations are made for the Engineer who would like to design the optimum compensator for any Linear System; and suggestions are made for further practical and theoretical research in this field.

LB.26/5/76

Acknowledgement:

I would like to thank Professor David Jacobson my un-official mentor, Dr. N.C.Powers who helped me with the mathematics, Mr. Mike Dewe, members of the Electrical Engineering Department, Dr. van der Merwe, and Professor John Flower who encouraged me in the crucial final stages of the project. I would also like to extend my sincere thanks to Mrs. Berger for typing the manuscript.

Introduction.

A linear electric circuit made up of resistors, inductors, capacitors and idealized active elements may be described by differential equations and algebraic equations of the form:-

$$v(t) = R\, i(t) \qquad v_o(t) = \mu\, V_1(t)$$

$$v(t) = L \frac{d}{dt} i(t) \qquad i_o(t) = \beta\, i_1(t)$$

$$i(t) = C \frac{d}{dt} v(t) \qquad i(t) = g_m\, v(t)$$

where the v's are voltages, and the i's are currents, these equations being linked by the network topology to yield the embracive state space representation:-

$$\dot{x} = Ax + Bu \qquad t > t_o$$

$$y = Cx \qquad x(t_o) = x_o$$

where x,y and u are n,m and r dimensional real vector valued functions of time respectively, and A, B, and C are matrices, each of consistent dimension. The elements of x, y and u will be the voltage or currents in different parts of the circuit; the entries in x being referred to as the state variables, those in u as the input, or control variables, and y, the output variables. In all real systems the entries in A, B and C will all be finite; furthermore, x_0 is included to specify the initial state of the system:- the problem of describing the behaviour of this system is thus well-posed. For the record, the above representation is said to be formulated in an n-dimensional Euclidean space, E^n.

If we look at the behaviour of the above circuit under very high frequency conditions, we may find strange effects, such as parasitic oscillations, or loss of attenuation. On closer examination, these effects may be found to be due to circuit layout, rather than circuit topology; the explanation

being that the system can no longer be described by simple Kirchhoff laws, but rather, account must be taken of the spatial aspects of circuit layout, this is done by using Maxwell's Equation:-

$$\nabla \times E = -\dot{B}$$

$$\nabla \cdot D = \rho$$

$$\nabla \times H = J + \dot{D}$$

$$\nabla \cdot B = 0$$

$$D = \varepsilon E$$

$$B = \mu H$$

Together with ohm's law:-

$$J = \sigma E$$

where all the variables are functions of space and time; but not the parameters, ε, μ and ρ which, in a time invariant homogenous system are real valued functions of space only. The above equations may be reduced to:-

$$\dot{H} = -\mu^{-1} \nabla \times E$$

$$\dot{E} = \varepsilon^{-1} \nabla \times H - \varepsilon^{-1} \sigma E$$

which clearly is of the form:-

$$\dot{U}(t) = AU(t)$$

where U is a vector function of time containing as elements, the spatial distributions of E and H; and A is some linear operator comprised of various scalar operators, and differential operators on these distributions.

In the analysis of a closed region, the effect of input variables, and output variables, may be taken into account by means of operators B and C such that:-

$$\dot{U}(t) = AU(t) + Bv(t)$$
$$y(t) = CU(t)$$

where $v(t)$ and $y(t)$ may or may not be spatially dependent. Of course, the proper solution of this problem requires initial conditions U_0, and boundary conditions where applicable. Here the operator A is defined on the space $C^2[R^3]$; that means that the elements of $U(t)$ can only be real valued twice continuously differentiable bounded functions of x.

If the domain of definition is infinite, one must only make sure that the state functions decrease sufficiently fast for the total energy in the system to remain finite. Thus $C^2[R^3]$ may be considered as a subspace of the Lebesque integrable space L_2, with norm:-

$$\|U(t)\|^2 = \tfrac{1}{2}\mu \int_\Omega |H|^2 dv + \tfrac{1}{2}\varepsilon \int_\Omega |E|^2 dv$$

See P.K.C. Wang[5] for further discussion of this type of problem.

Another space in which a system may be formulated, is obtained by extending n to infinity in E^n, and restricting attention to those systems for which total energy remains a meaningful concept, or more precisely, are square summable; the resultant space is called ℓ_2. Such systems probably never occur naturally, however, they have many of the properties of L_2 (distributed parameter systems), while retaining much of the simplicity of E^n (Finite dimensional systems). They may occur when using finite element models of distributed parameter systems.

A theory formulated in abstract Hilbert space, H , would cover all the cases of E^n , L_2 , and ℓ_2 ; such a theory would cover most engineering systems, and would have the advantage of being the most general framework for which many of the well known system theoretic properties still hold. For instance, a system in H , may be described by a single limit as n tends to infinity of a finite element model, where each element has been

systematically labelled : 1, 2, ... , n. Also the eigenvalues (if finite eigenvalues exist) occur as points in the complex s-plane, and not as lines, or dense areas. Finally, it still makes sense to talk of a quadratic cost function, controllability, and observability. Furthermore, a very nice theory exists for the solution of certain problems in H.

Given the system:-

$$\dot{U}(t) = AU(t) + Bv(t)$$
$$y(t) = CU(t) \qquad t > t_0 \; U(t_0) = U_0$$

where $U(t)$ is an element of some Hilbert space, A is a bound linear operator on H, and B is a linear operator mapping the control space into H. Then the solution is given by the Kernel representation:-

$$U(t) = T(t)\,U_0 + \int_{t_0}^{t} T(t-s)\,Bv(s)ds$$

$$y(t) = CU(t) \qquad t > t_0.$$

Where $T(t)$ is the solution of the autonomous system ($v(t)$ equal naught for all $t > t_0$), and is given by:-

$$T(t) = e^{A(t-t_0)} \qquad t > t_0$$

Here the bound set $\{T(t) : t > t_0\}$ is referred to as a Strongly Continuous Semigroup of Operators, and A is referred to as the Infinitesimal Generator of the Semigroup. Furthermore, there exist finite constants δ, $M(\delta)$ such that :-

$$\|T(t)\| < Me^{\delta t}$$

The reader is referred to Ladas and Lakshmikantham[18] for a fuller treatment of this subject.

So, let us compare the above technique, with the classical techniques of taking the Fourier, or Laplace transforms of $h(t)$ directly.

Fourier Transform.

A necessary, and sufficient condition for a function such as $h(t)$ to be Fourier Transformable, is that it is an element of L_1 (Lebesque integrable), and is of bounded variation (The derivative is bounded almost everywhere). The inverse of the Fourier Transform then constructs $h(t)$ at all but a countable number of discontinuous points, this is known as the Dirichlet condition. Certain design and analysis techniques exist in the frequency domain, mainly involving the theorems of Bode; however, they cannot be applied to unstable systems, and difficulties may arise with non-minimum phase systems, so a more powerful method had to be developed.

Laplace Transform.

Sufficient conditions for a function to be Laplace Transformable, are that it be causal (zero for t less than naught), Lebesque integrable over every finite interval, and that there exists a constant, δ, such that:-

$$\lim_{t\to\infty} e^{-\delta t} \int_0^t h(s)ds = 0$$

Thus such extraordinary functions as a dirac semi-comb can easily be handled by Laplace Transform techniques whereas the other methods find it impossible to handle something as simple as a delayed unit step. Once the Laplace Transform has been found, several techniques such as the initial value theorem, and the final value theorem may be used for system analysis. Another is the very powerful Nyquist stability criterion which may likewise still be used for infinite dimensional systems. Although the Nyquist criterion is traditionally used for stability it may easily be used to test for poles or zero's in any specified domain of the complex s-plane. We thus have a fair collection of basic analytic tools, but,

unfortunately no useful design tools for distributed parameter systems per se, or ever for large dimensional systems.

Early attempts at the formulation and solution of a Linear Optimal Control Problem.

Stimulated largely by the war effort, a linear optimal control problem was formulated, and solved by Norbert Wiener[1]. His work was conducted at about the same time as that of Kolmogoroff, in Russia; however, it was only completely declassified, and made publically available in 1949, when Wiener published: "Extrapolation, Interpolation, and Smoothing of Stationary Time Series."

His work can be recognised to be essentially in the frequency domain; being free of any reference to initial conditions (the noise and signal power density spectrums playing the equivalent rôle), and notably, being free of any reference to system dimension, thus infinite dimensional systems may be handled with equal facility, provided the input, and output vectors are finite.

Wiener used an integral-square error (I.S.E.) performance index, involving only the final output vector, this restriction hardly distracts from its usefulness. He represented the system by its impulse response, commonly designated $h(t)$, and he used the statistics $Cy(t)$ and $C\,uy(t)$, where $Cy(t)$ is the auto-correlation of the system output, and $C\,uy(t)$ is the cross-correlation between system input and system output. He was primarily interested in filtering, i.e. finding $u(t)$, given $y(t)$, where the system is noisy, i.e.:-

$$\dot{x}(t) = A\,x(t) + B\,u(t) + D\,v(t)$$
$$y(t) = C\,x(t) + w(t)$$

where $v(t)$ and $w(t)$ are zero-mean noisy inputs. His answer, the Wiener filter, is a system that accepts $y(t)$ as an input, and produces a minimum variance estimate of $u(t)$ as an output; however, with some

manipulation, this may be applied to control problems. The answer is in the form of $K(t)$, the impulse response of the optimum Wiener filter, it is given by the solution of the Wiener-Hopf integral equation of the first kind, viz.

$$C_{uy}(t) = \int_0^\infty K^T(\tau)\, C_y(t-\tau)d\tau \qquad t > 0$$

This equation is not only necessary, but is also a sufficient condition for optimality; however, it is by no means easy to solve, in particular it must only be satisfied by $K(\tau)$ for non-negative τ, for in order to preserve causality, and hence realizability it is necessary that $K(\tau)$ be identically zero for negative τ. Together with this, we have that $C_y(\tau)$ and $C_{uy}(\tau)$ are two-sided functions. The usual method of solution is done by factorizing which may be summerized as follows:-

Find two functions ψ_1 and ψ_2 such that:-

$$\psi_1(t) = 0 \qquad t < 0 \quad \text{causal part.}$$

$$\psi_2(t) = 0 \qquad t > 0 \quad \text{anti-causal part.}$$

$$C_y(t) = \int_{-\infty}^{\infty} \psi_1(t-\tau)\psi_2^T(\tau)d\tau$$

This can be recognised as an essential feature of the Bilateral Laplace transform. Next, find a function $a(\cdot)$ such that:-

$$C_{uy}(t) = \int_{-\infty}^{\infty} a(t-\tau)\psi_2^T(\tau)d\tau$$

It is now an easy matter to modify the Wiener-Hopf equation, to get the sufficient condition for optimality:-

$$a(t) = \int_0^\infty K^T(\tau)\psi_1(t-\tau)d\tau \qquad t > 0$$

Now because, $K(\tau)$ and $\psi_1(\tau)$ are both one-sided, causal functions, the convolution may be solved by using ordinary Laplace transforms.

Nowadays the problem is usually argued entirely in the frequency domain, using spectral densities instead of correlations, and solved by using Bilateral Laplace Transform theory; Hoever, such an approach adds nothing, and the method outlined above is more lucid.

The problem of factorizing has remained a stumbling block for many years, and it is perhaps for this reason that the Wiener filter has fallen out of favour; however, perhaps it is time for it to be re-visited, especially in the light of the availability of low-cost high-density memories and shift registers, enabling the impulse response of the compensator to be synthesized directly. The major research effort since the publication of Wiener's book, has been on finding a tractable method of solution, and on extending the results to time-varying systems. The author has also considered the problem independently, his findings being included in appendix A for reference. Although the results are not competitively practicable, they do serve to illustrate the nature of the problem.

The first line of attack is the sequency approach. It assumes a sample data controller; the impulse response of a stable single input single output system may thus be described by a summable series in z_1, where the output has been averaged over one sample period. Now if the state transition operator is nilpotent, or if the system is output degenerate (that is, there exists a T, such that if $t > T$, then $y(t) = 0$, for all x_0) then the problem becomes finite dimensional, thus enabling numerical solutions to be found.

The second approach tried by the author is the gradient method, it recognizes that the solution of the Wiener-Hopf equation is an infinite dimensional problem, and tries to solve it as such, it is of necessity a rather theoretical solution and one that can only be found if the answer has certain nice properties.

The Wiener filter is essentially an infinite dimensional filter. If the Wiener-Hopf equation could be solved in general, and if the resulting filter could be constructed, we would have a powerful design tool that could be applied to all linear systems with equal ease. In the next section it will be shown how the first of these problems has been surmounted for finite dimensional systems, and the possibility of extending the results to infinite dimensional systems will be investigated.

The Linear Quadratic Problem and the Kalman-Bucy Solution.

The major break-through in linear optimal control theory occured with a number of publications by R.E. Kalman, in the early 1960's. His research was largely stimulated by, and his results immensely useful to the "space race" of the 1960's. The finite-time linear Quadratic Problem for finite dimensional systems is now famous in control, and filtering (or estimation) theory; so is the resultant Riccati Differential Equation, named as such for its similarity to the equations studied by Count Riccati over two hundred years previously.

By way of review, one is concerned with minimizing a performance index of the form:-

$$J = \int_{t_o}^{t_f} \left(\frac{1}{2} x^T C^T C x + \frac{1}{2} u^T R u \right) dy + x(t_f)^T Q_f x(t_f)$$

subject to $\dot{x} = Ax + Bu \qquad t > 0$

$$x(t_o) = x_o$$

where the system is defined in a finite dimensional Euclidean space, R is positive (Legendre-Clebsch necessary condition for Optimizability) and invertible. Here t_o (and x_o) and t_f (and hence Q_f) are fixed, however, A, B, C, and R may be time-varying; and $u(\cdot)$ must be chosen from the class of measurable functions. The solution is both necessary and sufficient and is given by:-

$$u(t) = -R^{-1}(t)\, B^T(t)\, S(t)\, x(t)$$

where $S(t)$ is the solution of the Riccati Differential Equation:-

$$-\dot{S}(t) = C^T C + S(t)A + A^T S(t) - S(t)\, BR^{-1}B^T\, S(t)$$

$$S(t_f) = Q_f$$

Note here, that the solution $S(t)$ is independent of initial conditions x_o, this fact will take on greater significance in later work. The infinite time (time invariant) case is derived from the above when A, B, C and R are constant for all time, by extending t_f to infinity (t_o is usually taken without loss of generality to be zero) and arbitrarily reducing Q_f to zero. The solution is then given by:-

$$u(t) = -R^{-1} B^T S \quad x(t)$$

where S is the steady state solution of the Riccati Differential Equation, if it exists, i.e.:-

$$S = \lim_{t \to \infty} S(t)$$

where

$$\dot{S}(t) = C^TC + S(t)A + A^TS(t) - S(t)BR^{-1}B^TS(t)$$

$$S(0) = 0$$

The existence of this limit can be guaranteed if (A,B) is stabilizable, and (A,C) is detectable; detectability being the dual of stabilizability, in the sense without time reversal. Furthermore S is unique and positive, and satisfies the matrix equation:

$$C^TC + S\,A + A^T\,S - S\,BR^{-1}\,B^T\,S = 0$$

The beauty of this solution is that it gives an explicit method for calculating the optimal control, namely integrating the Riccati Equation, and this has been done a great deal in practice. Now, as we let the Euclidean model approach a real system in a Hilbert space such as ℓ_2, two difficulties arise, the first being a practical difficulty fondly known as "the curse of dimensionality"; and the second being that nagging doubt, that in spite of what may be intuitively obvious, we have as yet no rigorous proof that the solution of the optimal control problem in a finite dimensional space approaches the solution of the actual problem

in the infinite dimensional space. A closer look at this will prove fruitful.

The first of these difficulties was somewhat lessened, by finding more efficient algorithms for solving the Matrix Riccati Equation : Potter reduced the problem to an Eigenvalue problem involving an $2n \times 2n$ matrix; Kleinman introduced an itterative technique, claiming quadratic convergence, but requiring an initial stabilizing estimate of the solution. The author has tried to find a more efficient method, which has been included in Appendix B for reference. It may best be used for systems with only real roots, in which case it reduces the problem to one of finding the Eigenvalues of an $n \times n$ real symmetric matrix, which is distinctly easier than for a general matrix.

Supposing these computational difficulties can be overcome, the theoretical problem remains.

Now the Riccati Operator Valued Equations, as an operator on L_2 has been studied by Butkovskii[5] in the U.S.S.R. and P.K.C. Wang[6] in the U.S.A. ; and as an operator on general Hilbert space, by for instance, David Russel[7]. The extensions follow through with very little difficulty, when A, B, C^TC, R^{-1} are bounded as in the finite dimensional case. A major lack, is that sufficient conditions for convergence to the infinite time solution appear not to have been formulated; optimizability just being assumed.

The key question regarding convergence in dimension, as tackled by David Russel has been reviewed by the Author and is included in appendix C. Mr. Russel's paper has been found to be somewhat lacking; firstly the assumption that A , and B are bound precludes such phenomena as transport lag and boundary control, it is on this that he builds the proof that

J_K^o converges to J^o. Secondly, he proposes convergence "from outside", in other words, using the control that gives us J_K^o, does not even imply that J will be finite. Thirdly he only proves strong convergence in x_o, and finally since $u_K^o(o) = -R^{-1}B_K P_K x_o$ only converges strongly in x_o, he has effectively only calculated the optimal control $u^o(\cdot)$ in the vicinity of the origin. The author has overcome the last of Russel's problems by restricting x_o to a compact subspace, this can be implied by the idea of "spatially band-limited" in L_2 for instance, however general sufficient conditions for x_o to always be an element of some compact subspace, still needs to be resolved.

In conclusion it should be noted that in general, any finite dimensional approximation to an infinite dimensional optimal control problem yields a solution that is no longer independent of the initial condition, x_o.

An alternative formulation - the limited state linear regulator problem.

The Wiener Filter leads to an infinite-dimensional filter, the Kalman solution leads to a distributed measurement problem, or an infinite-dimensional Kalman Filter; an alternative approach is to recognise that these ideals can never be realized in practice and allowance be made from the start. A further justification, is that very often sensor type and location must often be decided "a priori", often before the transfer function is known; furthermore, component tolerances, cost and reliability usually limit the dimension of any real filter. The above considerations lead to what is known as a problem with complexity constraints, which leads to a parameter optimizing problem.

Gupta and Hasdorff[8] have briefly mentioned this problem as a variation of the Wiener filter in their popular text-book. A more tractable solution has been found by Levine and Athans[9] though they have only found necessary conditions for optimality; the author of this dissertation has virtually completed the solution of the problem by finding sufficient conditions, both for the existence of a solution,and for the solution to be a weak local minimizer. It being noted that the conditions for a global minimizer being not so easily found. The details of this work are given in appendix D, a more applications orientated version of this paper has been presented by the author at the S.A.C.A.C. Symposium on Control Theory.[11] Here again, the author has experimented with Newtonian type algorithms to find the optimum, but has generally found them less effective than solving the Riccati Equation, when this is possible.

It should be noted that a certain knowledge of the noise statistics, or moments of the initial conditions is required before the problem can be solved; thus again, the solution is not independent of x_0. On reflection, one observes that as the measurements progress, a closer estimate of

current state could be found, which may lead to a different optimum solution; this suggests that a non-linear regulator may well lead to better results.

Conclusion.

The Kalman solution to the Linear Quadratic problem has excelled when applied to medium order system, but much theoretical work still needs to be done on the Operator Valued Linear Qaudratic Problem. At the present state of the art, it is strongly recommended, that for large dimensional systems, the methods of Wiener should be used.

Another of the problems of distributed parameter systems, is the synthesis problem. It may be overcome by using a constrained optimization technique such as the Limited State approach, or the Sequency approach, but much room exists for further innovation.

In general a dynamic optimization problem has three interlaced parts, the forward part, the backward part, and the instantaneous optimization part. The finite dimensional Linear Quadratic Problem is unique, in that the backward part, and the forward part are entirely separated. In all three infinite dimensional cases considered under the major chapter headings, this was found to be not generally true, but rather, that the solution of the backward part, required some *a priori* knowledge of the forward part. It is thus not always possible to solve the Linear Quadratic Problem, unless one has some information on the probable, or possible initial states, or disturbances; but, it is only when the infinitesimal generator and the control operators are bounded, and when the solution may be synthesised accurately, that the results hold for the entire state space.

The following four areas are suggested for further research:-

1. The problem of synthesising distributed parameter filters can always be tackled from the innovative point of view for particular examples.
2. The gradient solution of the Wiener filter, outlined briefly in

Conclusion.

The Kalman solution to the Linear Quadratic problem has excelled when applied to medium order system, but much theoretical work still needs to be done on the Operator Valued Linear Qaudratic Problem. At the present state of the art, it is strongly recommended, that for large dimensional systems, the methods of Wiener should be used.

Another of the problems of distributed parameter systems, is the synthesis problem. It may be overcome by using a constrained optimization technique such as the Limited State approach, or the Sequency approach, but much room exists for further innovation.

In general a dynamic optimization problem has three interlaced parts, the forward part, the backward part, and the instantaneous optimization part. The finite dimensional Linear Quadratic Problem is unique, in that the backward part, and the forward part are entirely separated. In all three infinite dimensional cases considered under the major chapter headings, this was found to be not generally true, but rather, t olution of the backward part, required some *a priori* knowledge of the forward part. It is thus not always possible to solve the Linear Quadratic Problem, unless one has some information on the probable, or possible initial states, or disturbances; but, it is only when the infinitesimal generator and the control operators are bounded, and when the solution may be synthesised accurately, that the results hold for the entire state space.

The following four areas are suggested for further research:-

1. The problem of synthesising distributed parameter filters can always be tackled from the innovative point of view for particular examples.
2. The gradient solution of the Wiener filter, outlined briefly in

appendix A, should be investigated from the Engineering point of view, particularly with the intention of implementing as an adaptive regulator.

3. The possibility of using memoryless non-linear output-input controllers, for the control of large dimensional linear systems coule be investigated from both the practical and theoretical points of view.

4. The Operator Valued Linear Quadratic Problem still needs to be refined; in particular, sufficient conditions for the existence of a solution are needed, as are techniques for finding the solution.

KEY REFERENCES, IN CHRONOLOGICAL ORDER

1. Norbert Wiener, "Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications" John Wiley & Sons, 1942 (declassified 1949).
2. E. Hille & R.S. Phillips, "Functional Analysis and semi-groups" Colloq. Amer. Math. Soc., Vol. 31, 1957.
3. R.E. Kalman, "Contributions to the Theory of Optimal Control" Bol. Soc. Mat. Mexicana, Vol 5, 1960, pp 102-119.
4. A.G. Butkovskii & A. Ya. Lerner, "Optimal Control of Systems with Distributed Parameters" Avtomatika i Telemekhanika, Vol 21, 1960. Trans: Automation and Remote Control, Vol 21, 1960, pp 472-477.
5. P.K.C. Wang, "Control of Distributed Parameter Systems" Advances in Control Theory, Vol 1, 1964, pp 75-172.
6. A.V. Balakrishnan, "Optimal Control Problems in Banach Spaces" J. SIAM Control, Ser A, Vol 3, 1965, pp 152-180.
7. D.L. Lukes & D.L. Russell, "The Quadratic Criterion for Distributed Systems" SIAM J. Control, Vol 7, 1969, pp 101-121.
8. S.C. Gupta & L. Hasdorff, "Fundamentals of Automatic Control" John Wiley & Sons, 1970.
9. W.S. Levine, T.L. Johnson & M. Athans, "Optimal Limited State Variable Feedback Controllers for Linear Systems" IEEE Trans. Automatic Control, Vol AC-16, 1971, pp 785-792.
10. G.E. Ladas & V. Lakshmikantham, "Differential Equations in Abstract Spaces" Mathematics in Science and Engineering, Vol 85. 1972.
11. E.L. Jones, "Linear Quadratic Problems with Complexity Constraints" S.A.C.A.C. Symposium on Control Theory with the cooperation of South African Mathematical Society, 1975.

Appendix A

On the Solution of Wiener-Hopf type Equations.

Page 1: The Gradient Method for the Solution of the Wiener-Hopf Equation.

Page 4: A Sequency Approach to Optimal Estimation.
- After a lecture given by the Author to final year Control Design students, at the University of the Witwatersrand, 1975.

THE GRADIENT METHOD FOR THE SOLUTION OF THE WIENER-HOPF EQUATION

PROBLEM FORMULATION [1]

Suppose : there exists a Co-variance stationary time series $u(t) \in R^r$, that has a continuous, Fourier Transformable auto-correlation:-

$$C_u(t) = \lim_{T \to \infty} \frac{1}{T} \int_{-T}^{T} u(t + \tau) u^T(\tau)\, d\tau \qquad (1$$

and, that $u(t)$ serves as an input to some unspecified linear system, possibly with other zero-mean noisy inputs; and that the output is $y(t) \in R^m$, which has a continuous, Fourier Transformable, auto-correlation:-

$$C_y(t) = \lim_{T \to \infty} \frac{1}{T} \int_{-T}^{T} y(t + \tau) y^T(\tau)\, d\tau \qquad (2$$

and, that between input and output, there exists a continuous, Fourier Transformable, cross-correlation:-

$$C_{uy}(t) = \lim_{T \to \infty} \frac{1}{T} \int_{-T}^{T} u(t + \tau) y^T(\tau)\, d\tau$$

$$= \lim_{T \to \infty} \frac{1}{T} \int_{-T}^{T} u(\tau) y^T(t - \tau)\, d\tau \qquad (3$$

then, determine the optimum linear causal filter $K(t)$, such that the estimate:-

$$\hat{u}(t) = \int_0^{\infty} K^T(\tau) y(t - \tau)\, d\tau \qquad (4$$

is as close to $u(t)$ as possible, in the sense that:-

$$J(K) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} |u(t) - \hat{u}(t)|^2\, dt \qquad (5$$

is a minimum.

SOLUTION AND PROPOSED ALGORITHM

Expand (5) , and substitute into it, (1) , (2), (3) and (4) :-

$$J(K) = \frac{1}{2}\,\mathrm{tr}\, C_u(0) - \mathrm{tr}\int_0^\infty C_{uy}(t)K(t)\,dt + \frac{1}{2}\,\mathrm{tr}\int_0^\infty\int_0^\infty K^T(s)C_y(s-t)K(t)\,ds\,dt \qquad (6$$

Now investigate the change in $J(K)$ for a weak variation in K ; that is a variation that may be represented by $\varepsilon K'$, where K' is bound, and ε is a small positive number:-

$$J(K - \varepsilon K') - J(K) = -\varepsilon\,\mathrm{tr}\int_0^\infty C_{uy}(t)K'(t)\,dt + \varepsilon\,\mathrm{tr}\int_0^\infty\int_0^\infty K^T(s)C_y(s-t)K'(t)\,ds\,dt + \frac{1}{2}\varepsilon^2\,\mathrm{tr}\int_0^\infty\int_0^\infty K'^T(s)C_y(s-t)K'(t)\,ds\,dt \qquad (7$$

Now the last term is positive (sufficient condition), and tends to zero faster than the other two, so for sufficiently small ε , we may write:-

$$J(K - \varepsilon K') - J(K) = \varepsilon\,\mathrm{tr}\int_0^\infty\left[\int_0^\infty K^T(s)C_y(s-t)\,ds - C_{uy}(t)\right]K'(t)\,dt + \tfrac{1}{2}\,O(\varepsilon^2) \qquad (8$$

If we now choose:-

$$K'(t) = -\left[\int_0^\infty K^T(s)C_y(s-t)\,ds - C_{uy}(t)\right]^T \qquad (9$$

and if $K'(t)$ is bound, we can choose ε small enough, such that $J(K - \varepsilon K') \leq J(K)$, with equality only being achieved at a desirable solution. Also since $J(K)$ is bound below, the conditions of Polak's model have been satisfied, and we can propose the following algorithm:-

SOLUTION AND PROPOSED ALGORITHM

Expand (5) , and substitute into it, (1) , (2), (3) and (4) :-

$$J(K) = \frac{1}{2}\,\mathrm{tr}\; C_u(0) - \mathrm{tr}\int_0^\infty C_{uy}(t)K(t)\,dt + \frac{1}{2}\,\mathrm{tr}\int_0^\infty\!\!\int_0^\infty K^T(s)C_y(s-t)K(t)\,ds\,dt \qquad (6$$

Now investigate the change in $J(K)$ for a weak variation in K ; that is a variation that may be represented by $\varepsilon K'$, where K' is bound, and ε is a small positive number:-

$$J(K - \varepsilon K') - J(K) = -\varepsilon\,\mathrm{tr}\int_0^\infty C_{uy}(t)K'(t)\,dt + \varepsilon\,\mathrm{tr}\int_0^\infty\!\!\int_0^\infty K^T(s)C_y(s-t)K'(t)\,ds\,dt + \frac{1}{2}\varepsilon^2\,\mathrm{tr}\int_0^\infty\!\!\int_0^\infty K'^T(s)C_y(s-t)K'(t)\,ds\,dt \qquad (7$$

Now the last term is positive (sufficient condition), and tends to zero faster than the other two, so for sufficiently small ε , we may write:-

$$J(K - \varepsilon K') - J(K) = \varepsilon\,\mathrm{tr}\int_0^\infty\left[\int_0^\infty K^T(s)C_y(s-t)\,ds - C_{uy}(t)\right]K'(t)\,dt + \frac{1}{2}O(\varepsilon^2) \qquad (8$$

If we now choose:-

$$K'(t) = -\left[\int_0^\infty K^T(s)C_y(s-t)\,ds - C_{uy}(t)\right]^T \qquad (9$$

and if $K'(t)$ is bound, we can choose ε small enough, such that $J(K - \varepsilon K') < J(K)$, with equality only being achieved at a desirable solution. Also since $J(K)$ is bound below, the conditions of Polak's model[3] have been satisfied, and we can propose the following algorithm:-

Given : $\varepsilon_0 > 0$.

Step 0 : Compute a $K_0(\cdot)$, and set $i = 0$.

Step 1 : Set $\varepsilon = \varepsilon_0$.

Step 2 : Compute $K'(\cdot)$ using eqn. (9).

Step 3 : Set $Z(\cdot) = K_i(\cdot) - \varepsilon K'(\cdot)$.

Step 4 : If $J(Z) - J(K_i) \leq -\varepsilon$, set $K_{i+1}(\cdot) = Z(\cdot)$, set $i = i + 1$, and go to 1 ; else go to 5.

Step 5 : If $K'(\cdot) = 0$, set $K_{i+1}(\cdot) = K_i(\cdot)$ and stop; else, set $\varepsilon = \varepsilon/2$ and go to 2.

Now the space on which the impulse response $K(\cdot)$ has been defined, has not been well defined, apart from being some normed linear space. However, if it is compact the above algorithm will converge to an optimal solution.

CONCLUSION

The restriction that it be compact does have its drawbacks, for consider the trivial problem, where $y(t) = u(t)$, and

$$C_u(t) = C_{uy}(t) = C_y(t) = C(t)$$

we are required to solve :-

$$C(t) = \int_0^\infty K^T(s)C(t - s)\, ds$$

now this has the immediate solution $K^T(s) = I\,\delta(s)$ where $\delta(s)$ is the dirac delta function, and is not bound for all s , thus the above algorithm can not be used to find this solution.

If we expect the answer to be stable, of exponential order, and sufficiently smooth, then we can expect the above algorithm to work without any difficulty.

A SEQUENCY APPROACH TO OPTIMAL ESTIMATION

INTRODUCTION [3]

Let a discrete time series be represented by the sequence of finite terms:-

$$u^* = u(1) , u(2) , u(3) , \ldots \tag{1}$$

If this sequence is zero after the r'th element, it may be considered as an r-dimensional vector, designated u , by writing the entries in column fashion. Let this sequence pass through a linear system with a pulse response given by:

$$h^* = h(1) , h(2) , \ldots , h(n) \tag{2}$$

The response at the i'th time interval will be given by:-

$$y(i) = \sum_{j=1}^{n} h(j)\,u(i - j) \tag{3}$$

$$u(k) = 0 , \quad k < 1 \quad \text{or} \quad k > r$$

The sequence y^* will thus terminate after $m = r + n - 1$ terms. The convolution operator depicted in (3) , can thus be written as an m by r matrix , H.

$$H = \begin{bmatrix} h(1) & 0 & \ldots & 0 \\ h(2) & h(1) & & \vdots \\ \vdots & \vdots & & \vdots \\ h(n) & \vdots & & h(1) \\ 0 & h(n) & & \vdots \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & h(n) \end{bmatrix} \tag{4}$$

(r columns)

The vector y , may thus be written:-

$$y = H u \tag{5}$$

PROBLEM FORMULATION 'ND SOLUTION

Let us now try and determine the pulse response K^* , of the optimum compemsator, such that:-

$$\hat{\eta} = K y \tag{6}$$

is as close to η , the response of the ideal system , I , to the time series u^*:-

$$\eta = I u \tag{7}$$

in the sense that J is a minimum, where J is given by:-

$$J = \frac{1}{2} \sum_{i=1}^{\ell} q(i)(\eta(i) - \hat{\eta}(i))^2 \tag{8}$$

Here $\ell = m + p^{-1}$, where p is the sequence length of the pulse response of the compensator to be determined; and $q(i)$ is positive for all i.

The performance index J, may thus be written:-

$$J = \frac{1}{2} (\eta - \hat{\eta})^T Q(\eta - \hat{\eta})$$

$$\text{where } \eta = I u$$

$$\hat{\eta} = K H u \tag{9}$$

Instead of writing $\hat{\eta}$ in the above form , we observe that since convolution is a commutative operator, we can define a matrix U , such that:-

$$U = \overbrace{\begin{bmatrix} u(1) & 0 & \cdots & 0 \\ u(2) & u(1) & & \vdots \\ \vdots & \vdots & & \vdots \\ u(r) & \vdots & & u(1) \\ 0 & u(r) & & \vdots \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & u(r) \end{bmatrix}}^{p+n-1} \qquad (10$$

Thus:- $$\hat{\eta} = U H k \qquad (11$$

This equation will be accepted without modification if $r = p$, otherwise, H must be redefined to be $p + n - 1$ by p. Expanding the performance index, we thus get:-

$$J = \frac{1}{2} u^T I^T Q I u - u^T I^T Q U H k + \frac{1}{2} k^T H^T U^T Q U H k \qquad (12$$

minimising w.r.t. k gives

$$k = (H^T U^T Q U H)^{-1} H^T U^T Q I u \qquad (13$$

where the inverse exists, since Q is positive definite, and U and H are of full rank. Furthermore, it is comparatively easy to calculate, because it is only p-by-p, and is symmetric.

DISCUSSION

In the case of Q being the identity matrix (time invariant performance index), equation (13) may be written in the form:-

$$\sum C_y(i - j)k(i) - C_{uy}(j) = 0 \qquad (14$$

where $C_y(i - j)$ is the i,j'th element of the symmetric auto-correlation matrix:-

$$H^T U^T U H$$

and $C_{uy}(j)$ is the j'th element of the cross-correlation vector:-

$$H^T U^T I u$$

Thus equation (14) can be seen to be the discrete form of the Wiener-Hopf equation. The reason for it being so easy to solve in this case, is that the problem has been forcibly kept finite dimensional . The problem of realizability has been overcome by the commutation of equation (13) , which forbids entries in k of negative index, while otherwise preserving the system structure.

REFERENCES CITED

1. Norbert Wiener, "Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications", John Wiley & Sons, 1950.

2. E. Polak, "Computational Methods in Optimization, A Unified Approach", Mathematics in Science and Engineering, Vol 77, 1971.

3. M.Cuénod & A. Durling, "A Discrete-Time Approach for System Analysis", Academic Press, 1969.

Appendix B

On the Solution of the Matrix Valued Algebraic Riccati Equation

- After a paper published by the Author in the I.E.E.E. Transactions on Automatic Control.

A Reformulation of the Algebraic Riccati Equation Problem.

by E.L. Jones.

Abstract:- *The Algebraic Riccati Equation Problem is reformulated , so as to yield a simple solution when the system has only real roots , as may occur when using a spatially quantized distributed parameter model. A restriction is also placed on the choice of the synthetic output matrix , C.*

Introduction.

It is well known that the problem :-

$$\text{minimize } J = \int_{t_o}^{\infty} (\tfrac{1}{2} x^T C^T C x + \tfrac{1}{2} u^T R u)\, dt \qquad (1$$

$$\text{subject to :- } \dot{x} = Ax + Bu \; ; \qquad x(t_o) = x_o \qquad (2$$

where u and x are real valued vector functions of time of dimension m and n respectively , and A,B,C, and R are real constant matrices , such that R is positive definite ;

and (A,B) is stabilizable

(A,C) is detectable.

requires the solution of the Algebraic Riccati Equation [1] , [2].

$$C^T C + SA + A^T S - SBR^{-1} B^T S = 0 \qquad (3$$

The author is with the Department of Electrical Engineering , University of the Witwatersrand , Johannesburg , Jan Smuts Avenue , 2001 , Johannesburg , South Africa. Telephone 724-1311.

/2...

Various methods have been proposed to solve (3) with minimal computational effort , the most notable being by Kleinman [3] and Potter [4]. An alternative procedure is to be presented that is based on A commuting with C , and yields the most pleasing results when A is diagonalizable.

The Procedure.

Let the problem be formulated in a suitable co-ordinate system , choose C such that it is invertible , and commutes with the matrix A [5]. Then change the basis of the space :-

$$y = Cx \tag{4}$$

After substituting , the problem becomes :-

$$\text{minimize } J = \int_{t_o}^{\infty} (\tfrac{1}{2} y^T y + \tfrac{1}{2} u^T Ru)\, dt \tag{5}$$

$$\text{subject to } \dot{y} = CAC^{-1}y + CBu$$

$$= Ay + CBu \; ; \qquad y(t_o) = Cx_o \tag{6}$$

which admits as a solution

$$u^o = -R^{-1} B^T C^T S y$$

where S is the unique positive definite solution of :-

$$I + SA + A^T S - SCBR^{-1}B^T C^T S = 0 \tag{7}$$

We know that S is unique, and positive definite , because (A,I) is completely observable for all A. Now S is positive definite , so its inverse exists , multiply (7) on both sides by S^{-1} , giving :-

$$S^{-2} + AS^{-1} + S^{-1}A^T - CBR^{-1}B^T C^T = 0 \tag{8}$$

which may be re-written :-

/3...

$$(S^{-1} + A)(S^{-1} + A)^T = CBR^{-1}B^TC^T + AA^T \quad (9$$

The notation may be abbreviated be defining the following two matrices :-

$$Q = CBR^{-1}B^TC^T + AA^T \quad (10$$

$$P = S^{-1} + A \quad (11$$

and the problem becomes to find P such that :-

$$PP^T = Q \quad (12.1$$

$$P - P^T = A - A^T \quad (12.2$$

$$P - A > 0 \quad (12.3$$

Now, if A is symmetric, the solution of these three equations is particularly simple. First, because Q is symmetric it may always be diagonalized, so that : -

$$Q = HDH^T \quad (13$$

where D is strictly diagonal. Then the positive definite square root, P is given by :-

$$P = HD^{\frac{1}{2}}H^T \quad (14$$

which satisfies (12.3), since $P^2 - A^2 \geq 0$, and (A,I) is completely observable.

And the solution to (7) is then given by :-

$$S = (P - A)^{-1}$$

So that the control that minimizes (1) is :-

$$u^o = -R^{-1}B^TC^TSCx$$

/4...

Conclusion

The class of problems, for which A is symmetric (has only real roots) is non-trivial, as it often arises in first-order finite element models [7]. Furthermore, Wonham and Johnson [6] have shown that only n parameters are necessary to determine the optimal control, and as Gantmacher [5] has shown that commuting matrices have at least n independent parameters, it is reasonable to expect that restricting C to the class of matrices that commutes with A does not inhibit the problem, but rather reduces the number of decisions demanded of the designer. Note that the case of a singular C may easily be handled as a perturbation provided (A,CB) remains stabilizable, and (A,C) remains detectable. However if the number field is extended, to cope with complex roots, the optimal control becomes complex ; whereas if A is not symmetricised the equation (12.2) becomes non-zero, and not so easy to solve. The actual diagonalizat n of the symmetric matrix, such as in equation (13), may be achieved by means of the Jacobi-von Neumann Algorithm. [8]

/5...

References

[1] R. W. Brockett , "Finite Dimensional Linear Systems." John Wiley & Sons, pp 147-152 , 1970.

[2] W.M. Wonham , "On a matrix Riccati equation of stochastic control." SIAM J. Control , vol.6 , pp 681-698 , 1968.

[3] D.L. Kleinman , "On an iterative technique for Riccati equation computation." IEEE Trans. Automatic Control , vol.AC-13 , pp 114-115 , Feb. 1968.

[4] J.E. Potter , "Matrix quadratic solutions." J.SIAM , Appl. Math., vol.14 , pp 496-501 , 1966.

[5] F.R. Gantmacher , "The theory of Matrices." Chelsea Publishing Co., vol.1 , pp 220-224 , 1960.

[6] W.M. Wonham & C.D. Johnson , "Optimal Bang-Bang Control with quadratic performance index." Trans. ASME J. Basic Eng., series D , vol.86 , No. 1 , pp 107-111 , March , 1964.

[7] A.G. Butkovskii , "Some approximate methods for solving problems of optimal Control of Distributed Parameter Systems." Avtomatika i Telemekhanika , vol.22 , No.12 , pp 1565-1575 , December , 1961. Trans: Automation and Remote Control , vol.22 , pp 1429-1438.

[8] A. Ralston & H.S. Wilt (ed.) , "Mathematical methods for digital computers." John Wiley & Sons , vol.1 , pp 84-91 , 1960.

Appendix C

On the Solution of the Operator Valued Algebraic Riccati Equation.

- After the work of D.L. Russel, and reviewed by the Author.

Introduction.

Consider a System (A,B) described by a linear differential equation of the form:-

$$\dot{x}(t) = Ax(t) + Bu(t) \qquad t \geq t_o$$

$$x(t_o) = x_o$$

Where x is defined in a Hilbert space H_1 with norm $\|\cdot\|_1$ induced by the inner product $\langle \cdot,\cdot \rangle_1$. Similarly u will be defined in a Hilbert space H_2 with norm $\|\cdot\|_2$ induced by the inner product $\langle \cdot,\cdot \rangle_2$. The norm of A, a linear operator on H_1, will be defined by:-

$$\|A\| = \sup_{\|x\|_1 \leq 1} \|Ax\|_1 \qquad x \in H_1$$

and will be assumed to be finite; the norm of B, a linear operator from H_2 into H_1, will be defined by:-

$$\|B\| = \sup_{\|x\|_2 \leq 1} \|Bu\|_1 \qquad u \in H_2$$

and will be assumed to be bounded.

The closed loop Linear Quadratic Problem (L.Q.P) is the problem of choosing $u = F(x,t)$, such that $u(t)$ is measurable, and such that the behaviour of the System minimizes a performance index of the form:-

$$J = \int_{t_o}^{t_f} [\langle x(t), Wx(t) \rangle_1 + \langle u(t), R\,u(t) \rangle_2]\, dt$$

Where W is a bounded self-adjoint operator on H_1 in the same sense that A is bounded, and R is a one-to-one self-adjoint operator of H_2 onto H_2 with bounded inverse in the sense that

$$\|R^{-1}\| = \sup_{\|u\|_2 \leq 1} \|R^{-1}u\|_2 \qquad u \in H_2$$

is bounded.

This problem has been solved by Kalman[1] in 1960, where H_1 and H_2 are finite dimensional Euclidean spaces. He has shown, that F is in fact a linear operator of H_2 into H_1, and:

$$F = -R^{-1}B^T S(t)$$

where $S(t)$ is the solution of the Riccati[2] equation:-

$$-\dot{S}(t) = W + S(t)A + A^T S(t) - S(t)BR^{-1}B^T S(t)$$

$$S(t_f) = 0$$

The infinite time problem in Euclidean space ($t_o = 0$, $t_f = \infty$) has also been tackled, an expositary treatment being given in: R.W. Brockett[3] "Finite Dimensional Linear Systems". In this case:

$$F = -R^{-1}B^T S$$

where

$$S = \lim_{t \to \infty} S(t)$$

and $S(t)$ is given by the solution of:-

$$\dot{S}(t) = W + S(t)A + A^T S(t) - S(t)BR^{-1}B^T S(t)$$

It can be appreciated that the above limit only exists if some rather important assumptions on A,B and W are satisfied. Furthermore, it can be shown that $S > 0$, and that S satisfies:-

$$W + SA + A^T S - SBR^{-1}B^T S = 0$$

Wonham[4] has shown in "On a Matrix Riccati Equation of Stochastic Control", that under various assumptions, the above equation has a unique positive-definite solution. This has led to various efficient algorithms for calculating S, for instance Potter,[5] Kleinman,[6] and myself.[7]

Returning now to the problem of the space in which the problem has been defined, it has been solved in infinite dimensional spaces by Butkovskii[8] in Russia, and Wang[9] in the U.S.A. The solution of the infinite time problem in general Hilbert space, will be assumed to be rigorously proves, by Russel[10]; the problem remaining, that it may be exceedingly difficult to find numerical solutions to the operator-valued Riccati equation.

It is common engineering practice to model a distributed parameter system, as a lumped parameter model; interpreted mathematically, this is equivalent to projecting the infinite dimensional Hilbert space onto a finite dimensional sub-space. The question now arises, can we solve the Riccati equation as an operator on the finite-dimensional sub-space, and expect this approximation to be reasonably close to the desired solution as an operator on the whole space. The rest of this paper will attempt to answer this question; however, before proceeding, it is thought as well to put some qualitative factors into perspective.

The greatest emphasis should be on keeping H_1 as general as possible, in fact, in some applications even a Hilbert space may be too restrictive. H_2 on the other hand, is selected more general than the Euclidean space E^3. Next A , and B are usually determined external to one's self, and should be admitted to a rather general class; boundedness being the severest restriction allowed, but with the hope that it may later be removed. Finally W and R are usually determined rather arbitrarily, if not , semi-subjectively; hence assumptions regarding W , and R will generally not be considered restrictive, but will often help by reducing the class of all possible W or R.

Assumptions.

The assumptions concerning A , B , and possibly W ; will now be clearly defined.

"Completely Controllable".

A system (A,B) is completely controllable, if for any x_o (W.o.l.g = 0) and for any x_f , there exists a $u(\cdot)$, an element of the class of admissable controls, and a t_f , such that:

$$\begin{aligned} x(t_f) &= x_f \\ \dot{x}(t) &= Ax(t) + Bu \qquad t_o \leq t \leq t_f \\ x(t_o) &= x_o \end{aligned}$$

"Stabilizable".

A system (A,B) is stabilizable, if there exists an F (a linear mapping from H_1 into H_2) , such that for every x_o , and every t_o:-

$$x(t) \to 0 \quad \text{as} \quad t \to \infty$$

where $$\dot{x}(t) = (A - BF)\, x(t) \qquad t \geq t_o$$

$$x(t_o) = x_o$$

"Optimizable".

A system (A,B) is optimizable relative to W if there exists an admissable control u , and a constant, M_o , such that

$$\int_o^\infty | < x, Wx >_1 + < u, R\, u >_2 | \, dt \leq M_o \|x_o\|^2$$

subject to $\dot{x}(t) = Ax(t) + Bu(t) \qquad t \geq 0$

$$x(0) = x_o \qquad \forall x_o \in H_1$$

In the finite dimensional case complete controllability implies stabilizability, which in turn implies optimizability; the optimalizability assumption will be adopted in this paper.
In addition, the following structure of projections will be assumed.

Assumption on E_k

There is a sequence $\{E_k\}$ of projections on H_1 such that:-

(i) $\lim_{k\to\infty} E_k x = x$; $\lim_{k\to\infty} E_k^* x = x$ $\qquad x \in H_1$

(ii) $E_k A = A E_k$

This naturally leads to the following definitions:-

Definitions.

(i) $M_k = E_k(H_2)$

(ii) $x_k = E_k x$ $\qquad x \in H_1$

(iii) $B_k = E_k B$

(iv) $A_k = E_k A E_k$

(v) $W_k = E_k^* W E_k$

Further assumptions will make the proof more easy:-

Sufficient Assumptions.

(iii) $E_k E_\ell = E_\ell E_k = E_\ell$ $\qquad k > \ell$

(iv) $E_k^* W = W E_k$

These assumptions and definitions have the following immediate implications:-

(a) $A_k = E_k A = A E_k$

i.e. M_k reduces A.

(b) $W_k = E_k^* W = W E_k$

(c) $\|E_k\| \leq 1 \qquad \forall k$

Russel's earlier theorem will now be summarized.

Theorem. (Operator Valued L.Q.P.)

If (1) P(t) is the solution of the Riccati equation:-

$$\dot{P}(t) = W + A^*P(t) + P(t)A - P(t)BR^{-1}B^*P(t)$$

$$P(0) = 0 \qquad t > 0$$

(2) The system (A,B) is Optimizable relative to W.

Then

(1) P(t) is monotonic increasing with t, and has a limit:-

$$\lim_{t\to\infty} P(t) = P$$

(2) P is a positive self-adjoint operator that satisfies the quadratic equation:-

$$W + PA + A^*P - PBR^{-1} P = 0$$

(3) If $u^o = R^{-1}B^*P\,x$, then J reaches its minimum with u^o i.e.

$$J(u^o) \leq J(u) \qquad \text{for all admissable } u(\cdot)$$

(4) The minimum value of $J(\cdot)$ may be evaluated as:-

$$J(u^o) = \langle x_o , P\,x_o \rangle$$

Development of convergence in K.

The first two theorems show that optimazibility of a system implies optimizability of the sub-systems as generated by $\{E_k\}$.

Theorem 1.1

If (1) W is positive.

(2) $E_k^* W = W E_k$

Then

$$\langle x, W x \rangle \geq \langle x, W_k x \rangle \qquad \forall x \in H_1$$

Proof.

If W is positive, then for all $x \in H_1$

$$\langle (I - E_k)x , W(I - E_k)x \rangle_1 \geq 0$$

$$\therefore \quad \langle x, (I - E_k)^* W (I - E_k)x \rangle_1 \geq 0$$

$$\therefore \quad \langle x, (W - E_k^* W - W E_k + E_k^* W E_k)x \rangle_1 \geq 0$$

$$\therefore \quad \langle x, (W - E_k^* W E_k - E_k^* W E_k + E_k^* W E_k)x \rangle_1 \geq 0$$

$$\therefore \quad \langle x, (W - E_k^* W E_k)x \rangle_1 \geq 0$$

$$\therefore \quad \langle x, (Wx - W_k x) \rangle_1 \geq 0$$

$$\therefore \quad \langle x, W x \rangle_1 - \langle x, W_k x \rangle_1 \geq 0$$

$$\therefore \quad \langle x, W x \rangle_1 \geq \langle x, W_k x \rangle_1$$

□

Theorem 1.2

If $\langle x, Wx \rangle \geq \langle x, W_k x \rangle$ and if the syatem satisfies the optimizability assumption, then all the sub systems satisfy the optimizability assumption.

Proof.

If the system satisfies the optimizability assumption, then $\exists u, M$ such that:-

$$M\|x_o\|^2 \geq \int_0^\infty [\langle x, Wx\rangle_1 + \langle u, Ru\rangle_2]\,dt$$

where $\dot{x}(t) = Ax(t) + Bu(t) \qquad t \geq 0$

$$x(0) = x_o \qquad \forall x_o \in H_1$$

Now since $\langle x, W_k x\rangle \leq \langle x, Wx\rangle$, this gives:-

$$M\|x_o\|^2 \geq \int_0^\infty [\langle x, W_k x\rangle_1 + \langle u, Ru\rangle_2]\,dt$$

where $\dot{x}(t) = Ax(t) + Bu(t) \qquad t \geq 0$

$$x(0) = x_o$$

Now, by the definition of W_k , the idempotency of E_k , and the observation that

$$\dot{x}_k(t) = E_k Ax(t) + E_k Bu(t) = A_k x_k(t) + B_k u(t)$$

we get:

$$M\|x_k(0)\|^2 \geq \int_0^\infty [\langle x_k, W_k x_k\rangle_1 + \langle u, Ru\rangle_2]\,dt$$

where $\dot{x}_k(t) = A_k x_k(t) + B_k u(t) \qquad t \geq 0$

$$x_k(0) = E_k x_o$$

This may be interpreted as, $\exists\, u$, M such that the sub-system is optimizable. □

First strong convergence for finite t will be proved, it will depend on the following theorem due to Russel[11], which will be assumed:-

Theorem. (Russel)

Let $X(\mu,t)$ be the solution of

$$\dot{X} = F(X,\mu,t)$$

where $X(\mu,0) = X_o(\mu)$

where $\mu \in \Omega$ a compact space

$X \in B$ a Banach space (of Bound linear transformations)

and $t \in \mathbf{R}$ the real line.

also $F(X,\mu,t)$ is a polynomial in X, with coefficients strongly continuous functions of $\mu \in \Omega$ and $t \in \mathbf{R}$.

If $X_o(\mu)$ is strongly continuous in $\mu \in \Omega$, then $\exists T$ such that $X(\mu,t)$ is strongly continuous in the set $\{\mu \in \Omega, t \in [0,T]\}$.

Furthermore $X(\mu,t)$ is strongly continuous in the set $\{\mu \in \Omega, t > 0\}$ if and only if there exists a non-negative (increasing) function $g(t)$ such that:-

$$\|X(\mu,t)\| \leq g(\tau) \qquad \mu \in \Omega$$
$$0 \leq t < \tau < \infty$$

Theorem 2.

If $P_k(t)$ is the solution of the Riccati Equation:-

$$\dot{P}_k(t) = W_k + A^*_k P_k(t) + P_k(t) A_k - P_k(t) B_k R^{-1} B^*_k P_k(t)$$

$$P_k(0) = 0$$

where A_k, B_k, and W_k have been defined above, and $E_k x \to x$ strongly in x; and if the system (A,B) satisfies the optimizability assumption, then:-

$$P_k(t)x \to P(t)x$$

strongly in the set $\{x \in H_1, t > 0\}$

Proof.

It is only necessary to show that the conditions of Russel's earlier theorem are fulfilled. First note that k is a positive integer, append to this set the symbol ∞, and define $P_\infty(t) = P(t)$, etc. Now define a neighbourhood system on these, the extended positive integers; Ω.

(i) N is a neighbourhood of a finite integer n, if N is any subset of Ω which includes n.

(ii) N is a neighbourhood of ∞ if N is a subset of Ω that includes ∞ and all but finite many of n.

Ω thus forms what is called a compact topological space. Next note that $P_k(t)$ is a self-adjoint operator on H_1, with norm

$$\| P_k(t) \| = \sup_{\|x\|_1 \leq 1} \langle x, P_k(t)x \rangle$$

Now since every Cauchy sequence of self-adjoint operators, converges to a self-adjoint operator, the $P_k(t)$ form a real Banach space $\mathcal{B}$. Furthermore, since $P_k(t)$ increases monotomically to P_k, and all the $\langle x, P_k x \rangle$ are bound above by $\langle x, Px \rangle$, by the optimizability assumption; we have that $\mathcal{B}$ is a Banach space of bound linear operators. Finally note that t is real, and hence that our problem is formulated in the correct spaces.

Next note that:-

$$F(P_k(t),k,t) = W_k + A_k^* P_k(t) + P_k(t)A_k - P_k(t)B_k R^{-1} B_k^* P_k(t)$$

i.e. that $F(P_k(t), k,t)$ is a polynomial in $P_k(t)$ with co-efficients independent of - (and hence strongly continuous in-) t. Now $E_k x \to x$ strongly, so:-

$$W_k \; x \to W \; x$$

$$A_k \; x \to A \; x$$

$$B_k R^{-1} B^*_k \quad x \to B R^{-1} B^* \; x$$

These show that the co-efficients of F are strongly continuous functions of $k \in \Omega$, in the abstract sense in which Ω has been defined. Finally note that $P_k(o) = O$, and is independent both of k, and of t; and may thus be thought of as strongly continuous in k, and t.

Russel's theorem now assures us that $P_k(t)$ is strongly continuous in the set $\{k \in \Omega, t \in [O,T]\}$. To extend this set to include all positive t we call on the optimizability assumption again, to assure us that for all x_o;

$$\langle x_o , P_k(t) \; x_o \rangle \leq \langle x_o , P \; x_o \rangle$$

Thus $\| P_k(t) \|$ is bounded above by $\|P\|$, and $P_k(t)$ is a strongly continuous function in the set $\{k \in \Omega, t \geq O\}$. □

In the next group of theorems we shall prove the important result that $f(k) \to f(\infty)$; where $f(k) = \langle x_o , P_k x_o \rangle$ and $f(\infty) = \langle x_o , P x_o \rangle$. This result will be both significant from the theoretical, and the practical point of view. It guarantees continuity of $f(k)$ on Ω, even at $t = \infty$ thus completing the previous theorem; it also forms an essential premiss of Dini's theorem, to follow. Practically it means that for any initial state (x_o), a k-dimensional model may be found to give a performance $f(k)$ (measured in terms of the model) as close to optium $(f(\infty))$ as desired; however, the convergence may still be confounded by a different choice of x_o.

Theorem 3.1

If (1) W is positive

(2) $E_k^* W = W E_k$

(3) $E_k E_\ell = E_\ell E_k = E_\ell$ if $k > \ell$.

Then, if $k > \ell$, we have :

$$\langle x, W_k x \rangle \geq \langle x, W_\ell x \rangle \qquad \forall x \in H_1$$

Proof. For all $x \in H_1$ and, if W is positive,

$$\langle (E_k - E_\ell)x , W(E_k - E_\ell)x \rangle_1 \geq 0$$

$$\therefore \quad \langle x, (E_k - E_\ell)^* W (E_k - E_\ell)x \rangle_1 \geq 0$$

$$\therefore \quad \langle x, (E_k^* W E_k - E_k^* W E_\ell - E_\ell^* W E_k + E_\ell^* W E_\ell)x \rangle_1 \geq 0$$

Now by assumption (2), this implies:-

$$\langle x, (E_k^* W E_k - W E_k E_\ell - W E_\ell E_k + E_\ell^* W E_\ell)x \rangle_1 \geq 0$$

And by assumption (3):-

$$\langle x, (E_k^* W E_k - W E_\ell - W E_\ell + E_\ell^* W E_\ell)x \rangle_1 \geq 0$$

Then by the idempotency of E_ℓ, and again using assumption (2), this implies:-

$$\langle x, (E_k^* W E_k - E_\ell^* W E_\ell - E_\ell^* W E_\ell + E_\ell^* W E_\ell)x \rangle_1 \geq 0$$

So we get:-

$$\langle x, (E_k^* W E_k - E_\ell^* W E_\ell)x \rangle_1 \geq 0$$

$$\therefore \quad \langle x, (W_k - W_\ell)x \rangle_1 \geq 0$$

$$\therefore \quad \langle x, (W_k x - W_\ell x) \rangle_1 \geq 0$$

$$\therefore \quad \langle x, W_k x \rangle_1 - \langle x, W_\ell x \rangle_1 \geq 0$$

Thus proving the result:-

$$\langle x, W_k x \rangle_1 \geq \langle x, W_\ell x \rangle_1 \qquad k > \ell \qquad \square$$

Theorem 3.2

If $\langle x, W_k x \rangle \geq \langle x, W_\ell x \rangle \qquad k > \ell$

and if $E_k E_\ell = E_\ell E_k = E_\ell \qquad k > \ell$

then $\langle x_o, P_k x_o \rangle \geq \langle x_o, P_\ell x_o \rangle \; \forall x_o \in H_1 \qquad k > \ell$

i.e. f(k) is monotonic increasing.

Proof. By the solution of the Operator Valued L.Q.P.:-

$$\langle x_o, P_k x_o \rangle_1 = \int_0^\infty [\langle x_k, W_k x_k \rangle_1 + \langle u_k^o, R\, u_k^o \rangle_2] dt$$

where

$$\dot{x}_k = A_k x_k + B_k u_k^o \qquad t \geq 0$$

$$x_k(0) = E_k x_o \qquad \forall x \in H_1$$

Now since $\langle x, W_k x \rangle_1 \geq \langle x, W_\ell x \rangle_1 \qquad \forall x \in H_1$, we have:-

$$\langle x_o, P_k x_o \rangle \geq \int_0^\infty [\langle x_k, W_\ell x_k \rangle_1 + \langle u_k^o, R\, u_k^o \rangle_2] dt$$

where $\dot{x}_k(t) = A_k x_k(t) + B_k u_k^o(t) \qquad t \geq 0$

$$x_k(0) = E_k x_o \qquad \forall x_o \in H_1$$

Now since $x_\ell = E_\ell x = E_\ell E_k x = E_\ell x_k$,

we can write the above expression as:-

$$\langle x_o, P_k x_o \rangle \geq \int_0^\infty [\langle x_\ell, W_\ell x_\ell \rangle_1 + \langle u_k^o, R\, u_k^o \rangle_2] dt$$

subject to $\dot{x}_\ell(t) = A_\ell x_\ell(t) + B_\ell u_k^o(t) \qquad t \geq 0$

$$x_\ell(0) = E_\ell x_o \qquad \forall x_o \in H_1$$

Referring back to the Operator Valued L.Q.P. we see that the control

$u_k^o(t)$ is sub-optimal for the system (A_ℓ, B_ℓ), i.e.

$$\int_0^\infty [\langle x_\ell\, W_\ell x_\ell \rangle_1 + \langle u_k^o\, ,\, R\, u_k^o \rangle_2]\, dt \geq \langle x_o, P_\ell x_o \rangle$$

subject to

$$\dot{x}_\ell(t) = A_\ell x_\ell(t) + B_\ell\, u_k^o(t) \qquad t > 0$$

$$x_\ell(0) = E_\ell x_o \qquad \forall\, x_o \in H_1$$

Combining, we get:-

$$\langle x_o, P_k\, x_o \rangle \geq \langle x_o,\, P_\ell x_o \rangle \qquad \forall\, x_o \in H_1,\ k > \ell$$

i.e. that $f(k) \triangleq \langle x_o, P_k x_o \rangle$ is monotonic increasing with k. □

Theorem 3.3

If $f(k)$ is monotonic increasing for all k, and the system satisfies the optimizability assumption; then $f(k)$ has a limit, and this limit is not greater than $\langle x_o, P\, x_o \rangle$.

Proof.

From the Operator Valued L.Q.P. if the system (A,B) satisfies the optimizability assumption, then there exists a P, such that:-

$$\min_{u(\cdot)} \int_0^\infty [\langle x,\, Wx \rangle_1 + \langle u,\, B\, u \rangle_2]\, dt = \langle x_o,\, P\, x_o \rangle$$

subject to

$$\dot{x}(t) = Ax(t) + Bu(t) \qquad t > 0$$

$$x(0) = x_o \qquad \forall\, x_o \in H_1\ ,$$

Now since $f(k)$ is monotonic increasing, we must have $f(k) \leq \langle x_o, P x_o \rangle$ for all finite k ; thus $f(k)$ has a limit as $k \to \infty$, and

$$\lim_{k\to\infty} f(k) \leq \langle x_o, P\, x_o \rangle \qquad \forall\, x_o \in H_1$$ □

Theorem 3.4

If (1) $P_k(t)$ increases monotonically with t, and $P_k(t) \to P_k$ $k \in \Omega$

(2) $P_k(t)x_o \to P(t)x_o$ strongly

(3) $\lim_{k\to\infty} f(k) \leq \langle x_o, P\, x_o \rangle$

Then $\lim_{k\to\infty} f(k) = \langle x_o, P\, x_o \rangle$

Proof.

Assume the conclusion to be false, then by premise (3):-

$$f(k) \to (x_o, P\, x_o) - \delta$$

where δ is some positive number; now observe:

$$\begin{aligned} & f(k) - f_t(k) + \delta \\ = \; & f(k) - (x_o, P\, x_o) + \delta + (x_o, P\, x_o) - (x_o, P(t)x_o) \\ & + (x_o, P(t)x_o) - f_t(k) \\ \leq \; & |f(k) - (x_o, P\, x_o) + \delta| + |(x_o, P\, x_o) - (x_o, P(t)x_o)| \\ & + |(x_o, P(t)x_o) - f_t(k)| \end{aligned}$$

where $f_t(k) \triangleq \langle x_o, P_k(t)x_o \rangle$ as usual.

Now by premise (1), it is possible to choose a T such that

$$|\langle x_o, P\, x_o \rangle - \langle x_o, P(T)x_o \rangle| \quad < \delta/3$$

like-wise, by premise (2), and the assumption, it is possible to choose a $K(T)$ such that:-

(i) $|\langle x_o, P(T)x_o \rangle - f_T(k)| < \delta/3 \qquad \forall k > K$

(ii) $|f(k) - \langle x_o, P\, x_o \rangle + \delta| < \delta/3 \qquad \forall k > K$

Thus:-

$$f(k) - f_t(k) + \delta < \delta$$

i.e. there exists at least one k, and (some x_o) such that:-

$$\langle x_o, P_k(T)x_o \rangle > \langle x_o, P_k\, x_o \rangle$$

but this contradicts the premise that $P_k(t)$ increases monotonically with t, to the limit P_k; so the assumption is false, and

$$\lim_{k\to\infty} f(k) = \langle x_o, P\, x_o \rangle \qquad \square$$

The following theorem due to Dini may be found, for example in M. Zamansky's [12] "Linear Algebra and Analysis", and will be assumed to be valid.

Theorem (Dini)

If a monotone sequence of real functions f_n, which are continuous on a compact space E, converges simply to a continuous function f, then it converges uniformly.

We shall now use Dini's theorem to prove the following:-

Theorem 4.1

If for all x_o, $P_k(t)$ is monotone increasing with t, and

(1) $P_k(t) \to P_k$ strongly in k, as $t \to \infty$

(2) $P_k(t)x_o \to P(t)x_o$

(3) $\langle x_o, P_k x_o \rangle \to \langle x_o, P\, x_o \rangle$

Then $\langle x_o, P_k(t)x_o \rangle \to \langle x_o, P_k\, x_o \rangle$ uniformly in k.

Proof.

Define $f_t(k) = (x_o, P_k(t)x_o$, and $f(k) = (x_o, P_k\, x_o)$, where k is defined on Ω, a compact (topological) space. Then premise (1) clearly implies $f_t(k) \to f(k)$ strongly on Ω, where $f_t(k)$ is monotone.

Premise (2) implies that for all (finite) t, $f_t(k)$ is continuous on Ω, and (3) implies that $f(k)$ is also continuous on Ω at $t = \infty$. The conditions of Dini's theorem have thus been satisfied, and we may conclude that $f_t(k) \to f(k)$ uniformly on Ω ; i.e. $(x_o, P_k(t)x_o) \to (x_o, P_k x_o)$ uniformly in k. □

We now strengthen this convergence in x_o.

Theorem 4.2

If (1) The system satisfies the optimizability assumption.

(2) $P_k(t)$ is monotonic in t.

(3) $(x_o, P_k(t)x_o) \to (x_o, P_k x_o)$ uniformly in k for all x_o.

Then $P_k(t)x_o \to P_k x_o$ uniformly in k

Proof.

$$\|P_k(t)x_o - P_k x_o\|^4$$

$$= \langle (P_k(t) - P_k)x_o \, , \, (P_k(t) - P_k)x_o \rangle^2$$

$$\leq \langle (P_k(t) - P_k)x_o \, , \, x_o \rangle \langle (P_k(t) - P_k)^2 x_o \, , \, (P_k(t) - P_k)x_o \rangle$$

$$\leq \{ \langle P_k(t)x_o \, , \, x_o \rangle - \langle P_k x_o \, , \, x_o \rangle \} \cdot \|P_k(t) - P_k\|^3 \cdot \|x_o\|^2$$

Now since $0 \leq P_\infty - P_k(t) \leq P_k \leq P$, the right-hand side converges to 0 uniformly in k, hence so does the left-hand side, so:-

$$P_k(t)x_o \to P_k x_o \text{ uniformly in } k$$ □

We may now deduce that $P_k x_o$ converges to $P x_o$ (without any involvement of t).

Theorem 5

If for each x_o ,

(1) $P_k(t)x_o \to P(t)x_o$ strongly in t

(2) $P_k(t)x_o \to P_k x_o$ uniformly in k

(3) $P(t)x_o \to P\,x_o$

Then $P_k x_o \to P\,x_o$

Proof.

First note:-

$$\|P_k x_o - P\,x_o\|$$

$$= \|P_k x_o - P_k(t)x_o + P_k(t)x_o - P(t)x_o + P(t)x_o - P\,x_o\|$$

$$\leq \|P_k x_o - P_k(t)x_o\| + \|P_k(t)x_o - P(t)x_o\| + \|P(t)x_o - P\,x_o\|$$

Now writing out the premises in full we see:-

(1) $(\delta)(x_o)(t)(\exists k)[\,k > K \Rightarrow \|P_k(t)x_o - P(t)x_o\| < \delta/3]$

(2) $(\delta)(x_o)(\exists T)(K)[\,t > T \Rightarrow \|P_k(t)x_o - P_k x_o\| < \delta/3]$

(3) $(\delta)(x_o)(\exists T)[\,t > T \Rightarrow \|P(t)x_o - P\,x_o\| < \delta/3]$

Combining, we get:-

$$(\delta)(x_o)(\exists T)(\exists K)[\,k > K \Rightarrow \|P_k x_o - P\,x_o\| < \delta]$$

Thus

$$P_k\,x_o \to P\,x_o$$

□

Finally we relate the convergence of P , to that of the control, u.

Theorem 6.

$$u_k^o(0) \to u^o(0)$$

Proof.

$$u_k^o(0) = - R^{-1} B_k^* P_k x_o \qquad k \in \Omega$$

Now since B_k^* converges strongly to B^*, and P_k converges strongly to P, and since R^{-1}, B, and P are all bound we have that:-

$$\| R^{-1} B_k^* P_k x_o - R^{-1} B^* p\, x_o \| \to 0 \quad \text{as} \quad k \to \infty \qquad \square$$

Addendum.

Russel's[13] paper effectively proves that $u_k^o(t) \to u^o(t)$ strongly in x_o, and only at $t = 0$. If x_o can be assumed to be constrained to some compact subspace of H_1 (write $x_o \in E$), and if the optimal system can be assumed to be stable (Optimizability, as defined in the text, does not imply stabilizability); then uniform convergence can be proved, that is, that $u_k^o(t) \to u^o(t)$ for all t, and for all $x_o \in E$.

Proof.

Since A, B, R^{-1}, and P are all bound, so is the operator A', where

$$A' = A - BR^{-1}B^*P$$

Hence the operator $T(t)$ is also bound for every t, since the series of partial sums is:-

$$T(t) = \sum_{k=0}^{\infty} (A'^k/k!)t^k$$

is majorized by the exponential of the bound on $A't$. Now since the system is stable, $T(t)$ converges to zero, and is hence bound for all t. Now $x(t)$ is given by:-

$$x(t) = T(t)x_o$$

and since $T(t)$ is bound, and $x_o \in E$, a fundamental theorem of topology tells us that $x(t) \in E$. Now for all $x(t) \in E$, we have

(1) that $P_k x(t)$ is continuous in $x(t)$ since P_k is bounded, for all $k \in \Omega$.

(2) that $P_k x(t) \to Px(t)$ monotonically in k.

Thus, Dini's theorem tells us that $P_k \to P$ uniformly on E. □

References

1. R.E. Kalman, "Contributions to the theory of optimal control", Bol. Soc. Mat. Mexicana, Vol. 5 (1960) pp 102-119.

2. A relevant reference is believed to be the following:- D'Alembert, "Hist. de l'Acad. R. des Sci. de Berlin", XIX 1763) p242.

3. R.W. Brockett, "Finite Dimensional Linear Systems", John Wiley & Sons, (1970) pp 147-155.

4. W.M. Wonham, "On a Matrix Riccati Equation of Stochastic Control", S.I.A.M. J. Control, Vol. 6, No. 4, (1968) pp 681-697.

5. J.E. Potter, "Matrix quadratic solutions", J. S.I.A.M. Appl. Math. Vol. 14 (1966) pp 496-501.

6. D.L. Kleinman, "On an iterative technique for Riccati equation computation", I E E E Trans. Automatic Control, Vol. AC-13. (1968) pp 114-115.

7. E.L. Jones, "A Reformulation of the Algebraic Riccati Equation Problem", I E E E Trans. Automatic Control, Vol. AC-21, (1976) p 113.

8. A.G. Butkovskii and A. Ya. Lerner, "Optimal control of systems with distributed parameters", Avtomatika i Telemekhanika, Vol. 21, No. 6 (1960).

9. P.K.C. Wang, "Control of Distributed Parameter Systems", Advances in Control Systems, Vol. 1 (1964) pp 75-172.

10. D.L. Lukes and D.L. Russell, "The Quadratic Criterion for Distributed Systems" S.I.A.M. J. Control, Vol. 7 , No. 1, (1969) pp 101-121.

11. D.L. Russel, "Continuity in the Strong Topology of Operator-Valued Solutions of Non-linear Differential Equations with an Application to Optimal Control", S.I.A.M. J. Control, Vol.7, No. 1, (1969) pp 132-140.

12. M. Zamansky, "Linear Algebra and Analysis", Van Nostrand (1969) pp 293-294.

13. D.L. Russell, "Operator Solutions of Non-linear Equations in Optimal Control Problems", in "Non-linear Functional Analysis and Applications", ed. L.B. Rall.

Appendix D

On the Linear Quadratic Problem with Complexity Constraints.

- After a paper read by the Author at the S.A.C.A.C. Symposium on Control Theory at The University of the Witwatersrand.

Sufficient Conditions for the Limited State Linear Regulator Problem.

E.L. Jones

Abstract:- Sufficient conditions are found, for the existance of an optimal control, and for the control to be a local minimum of the infinite time linear multivariate dynamic system with quadratic cost function, and constant partial state feedback.

Introduction:

In general a dynamic system, may be of exceedingly high order, but with only a few state variables available for measurement. One approach to this type of optimal control problem is to re-construct the missing state variables, using a Kalman filter [1], or a Luenberger observer [2], but this again can lead to a dimensionality crisis in the state re-constructor. An alternative approach, largely attributed to Michael Athans and W.S. Levine [3] (though the general philosophy has been around a long time before) is first to determine the structure of the compensator, and then to treat the problem as a parameter optimizing problem. Levine and Athans have however only found necessary conditions for an optimal control (if it exists). In this paper sufficient conditions will be given for the existence of a local minimizer.

Notation.

$E[\cdot]$	Denotes Expectation of a Random Variable.
$\{\cdot\}$	Denotes the Range Space of an Operator.
$N\{\cdot\}$	Denotes the Null Space of an Operator.
$\langle \cdot , \cdot \rangle$	Denotes Inner Product in Hilbert Space.
$[\cdot \mid \cdot]$	Denotes Catenation (or Augmentation)
$(\cdot \otimes \cdot)$	Denotes Kronecker product.
$\{\phi\}$	Denotes the empty set.
$[\cdot]_s$	Subscript "s" denotes Symmetric Component.

Problem Formulation:

Consider the problem:-

$$\text{Minimize} \quad J = E\left[\tfrac{1}{2}\int_0^\infty (x^T C_0^T C_0 x + u^T R u)\,dt\right]$$

$$\text{where} \quad \dot{x} = Ax + Bu \quad ; \quad t \geq 0$$

$$y = Cx$$

$$u = -Fy$$

$$\text{and} \quad E\left[x(o)x^T(o)\right] = X_o = DD^T \tag{1}$$

where x,y, and u are real valued vector functions of time, with dimensions n,m and r respectively; and A,B,C,F,C_0,R,X_o and D are constant matrices consistant with the equations. There will be no loss of generality if R is assumed to be symmetric. F is the matrix of parameters to be varied, in order to minimize J.

Only if the integral exists, may the performance criterion be evaluated $\forall x(0)$, thus :-

$$J = \tfrac{1}{2} \quad E\left[x^T(o)\int_o^\infty e^{A_c{}^T t}\, Q\, e^{A_c t}\,dt\; x(o)\right]$$

where $A_c = A - BFC$

$$Q = C_0^T C_0 + C^T F^T R F C \tag{2}$$

$$\text{writing} \quad K = \int_o^\infty e^{A_c{}^T t}\, Q\, e^{A_c t}\,dt \tag{3}$$

we get

$$J = \tfrac{1}{2}\, tr\left[K\, D\, D^T\right] \tag{4}$$

it can be shown that K is given by the unique solution of:-

$$K A_c + A_c{}^T K + Q = O \tag{5}$$

where a sufficient condition for the uniqueness of K is that A_c be stable.

Alternatively, the performance criterion may be re-written:-

$$J = \tfrac{1}{2}\, tr\left[Q\int_o^\infty e^{A_c t}\, DD^T\, e^{A_c{}^T t}\,dt\right] \tag{6}$$

then writing

$$L = \int_0^\infty e^{A_c t} \, DD^T \, e^{A_c^T t} \, dt \tag{7}$$

we get

$$J = \tfrac{1}{2} \operatorname{tr} [Q \, L] \tag{8}$$

it can be shown that L is given by the unique solution of:-

$$L A_c^T + A_c L + DD^T = 0 \tag{9}$$

where a sufficient condition for the uniqueness of L is also that A_c be stable. [4]

The system (A,B,C) will be said to be stabilizable if and only if there exists a finite F such that A_c is stable.

The following remark will be useful: If (A,B) is stabilizable, and (A,C) is detectable, then either A is stable; or the matrix

$$W = C \int_0^\infty e^{At} \, BB^T \, e^{A^T t} \, dt \, C^T \tag{10}$$

is unbounded; but obviously not both. For if A is unstable, let ξ be an eigenvector corresponding to an eigenvalue λ of A, such that λ has non-negative real part. The space spanned by all possible ξ is called E_A^+, so we may say : $\xi \in E_A^+$. Now because (A,C) is detectable, we have that $C\xi \neq 0$. [5]

So there exists an $\eta \in R^m$, such that $C^T\eta$ has a component parallel to ξ. Use any such η to evaluate W :-

$$\eta^T W \eta > k \int_0^\infty e^{2 \operatorname{Re} \lambda t} \, \| \xi^T B \|^2 \, dt \tag{11}$$

where k is some positive scalar. Now in order for $\eta^T W \eta$ to be bounded, we must have $\xi^T B = 0$, i.e. :-

$$\xi^T A^i B = \lambda^i \xi^T B = 0 \qquad i = 0,1 \ldots , n-1 \qquad (12)$$

That is to say :-

$$\xi \in N\{\Gamma^T(A,B)\} \qquad (13)$$

where $\Gamma(A,B) = [B|AB| \ldots\ldots A^{n-1}B]$

but Wonham [5] has shown that the controllability subspace includes the unstable subspace, i.e. that :-

$$\{\Gamma(A,B)\} \supseteq E_A^+ \qquad (14)$$

which is equivalent to :-

$$N\{\Gamma^T(A,B)\} \cap E_A^+ = \{\phi\} \qquad (15)$$

thus contradicting our original assumption.

Existence of an Optimal Control.

As to the existence of an optimal F , called F* , we pose the following theorem.

Theorem 1.

If 1) (A,B,C) is stabilizable

2) (A,D_c) is stabilizable

3) (A,C_Q) is detectable .

4) R is positive definite

Then there exists an optimal F* with finite norm, that stabilizes (A,B,C) and minimizes J.

Proof. Let us refer to the class of F's with finite norm, that stabilizes (A,B,C) as χ , then if (A,B,C) is stabilizable, there exists an $F_i \in \chi$ such that $A - BF_iC$ is stable. Either this F_i is the optimal one in χ , thus satisfying the theorem; or there exists a better one, $F_{i+1} \in \chi$ such that $J(F_{i+1}) < J(F_i)$. We may proceed inductively, all that remains to be shown, is that the procedure has a limit, and that this limit is in χ. Now

clearly there is a minimal value of J, namely $J = 0$, and the sufficient condition for the existence of this minimum is given by $R > 0$, hence the procedure must have a limit F^*. Let us now assume that $A - BF_iC$ becomes unstable as $F_i \to F^*$, now clearly:

$$J \geq \tfrac{1}{2} \operatorname{tr}\left[C_0 \int_0^\infty e^{A_c t} DD^T e^{A_c t} \, dt \, C_0^T \right] \tag{16}$$

but since (A_c, D) is stabilizable, and (A_c, C_0) is detectable, this means that J becomes infinite, which contradicts the principle that J improves, so clearly $A_c - BF^*C$ is stable. Let us now suppose that $F_i \to \infty$ as $F_i \to F^*$, now clearly:

$$J \geq \tfrac{1}{2} \operatorname{tr}\left[F^T R \, F \, C \, L \, C^T \right] \tag{17}$$

and if $C \, L \, C^T$ is invertible for all F_i, this means that J becomes infinite, thus proving our theorem. If however, $C \, L \, C^T$ is singular for some i, then there exists at least one ΔF_i such that:-

$$\Delta F_i \, C \, N_i = 0 \tag{18}$$

where $N_i \, N_i^T = L_i$

Furthermore, any motion along any of the ΔF_i that satisfies (18) does not alter L_i, or $J(F_i)$, so we may construct a new sequence, S_i, by minimizing the norm (relative to all ΔF_i) at each step.

$$\begin{aligned} S_i &= F_i \quad ; \quad CL_iC^T \text{ invertible} \\ S_i &= \arg\min \left[\| F_i + \Delta F_i \| : \Delta F_i \, CN_i = 0 \right] ; CL_iC^T \text{ singular.} \end{aligned} \tag{19}$$

This new sequence has been chosen so that at each step,

$$S_i \, CN_i \neq 0$$

and

$$S_i \, C \, L_i \, C^T S_i^T > 0 \tag{20}$$

Let us now suppose that this new sequence becomes infinite as $F_i \to F^*$, then

$$J > \tfrac{1}{2} \operatorname{tr}\left[R \, S \, C \, L \, C^T S^T \right] \tag{21}$$

but since R is positive definite, and $S\,C\,L\,C^{T}S^{T}$ is non-singular, this means that J becomes infinite, which contradicts the principle that J improves, so $S_i \to S_o = F_m$ as $F_i \to F^*$ and there exists an F_m with finite norm, such that $J(F_m) = J(F^*)$, and $F_m \in \chi$.

Let us now review the necessary conditions for optimality:-

Theorem 2.

If F^* minimizes J , then

$$M\,L\,C^{T} = 0 \tag{22}$$

where $M = R\,F^*C - B^{T}K$

and where K and L are evaluated at F^*.

This is clearly the theorem due to Levine and Athans [3]. The importance of including it here, is that if it is read in conjunction with theorem (1), then we are given sufficient conditions for the existence of a solution to $M\,L\,C^{T} = 0$.

Sufficient Conditions for a Local Minimum.

In this section we shall look at differential variations in the cost, to second order; and establish sufficient conditions for its non-negativity. We shall save a lot in notational complexity if we adopt the following generalizations as definitions of the derivative. [6]

Definition. Let $f(\cdot)$ be a mapping from χ into γ , where χ and γ are Banach spaces. If x is an element of an open set in χ , and if :-

$$f(x+h) - f(x) = L_1(x,h) + w_1(x,h);\ h \in \chi \tag{23}$$

where $L_1(x,\cdot)$ is a linear operator from χ into γ , and

$$\frac{\|w_1(x,h)\|}{\|h\|} \to 0 \qquad \text{as} \quad \|h\| \to 0 \tag{24}$$

Then $L_1(x,\cdot)$ is called the Fréchet derivative of $f(\cdot)$ evaluated at x, and is unique, and is denoted by $f'(x)$. Now $f'(\cdot)$ itself is an element of $B(\chi,Y)$ where $B(\chi,Y)$ is the space of bounded linear operators from χ into Y, which is also a Banach space. With the derivative formulated in this way, we are able to repeat the process:-

Let $f'(\cdot)$ be a mapping from χ into $B(\chi,Y)$, where χ and $B(\chi,Y)$ are Banach spaces. If x is an element of an open set in χ, and if:-

$$f'(x+h) - f'(x) = L_2(x,h) + w_2(x,h) \qquad h \in \chi \tag{25}$$

where $L_2(x,\cdot)$ is a linear operator from χ into $B(\chi,Y)$, and

$$\frac{\|w_2(x,h)\|}{\|h\|} \to 0$$

$$\text{as} \quad \|h\| \to 0 \tag{26}$$

Then $L_2(x,\cdot)$ is called the Fréchet derivative of $f'(\cdot)$ evaluated at x, and is unique, and is denoted by $f''(x)$. Here $f''(\cdot)$ itself is an element of $B(\chi,B(\chi,Y))$ where $B(\chi,B(\chi,Y))$ is the space of bounded linear operators from χ into $B(\chi,Y)$, and the process may be repeated.

In our case χ is a Hilbert space, and Y is a subset of the real line, hence we may define the differentials:-

$$\delta^2 f = f''(x)h + O(h)^2$$

$$\text{and} \quad \delta f = \langle h, f'(x)\rangle + \tfrac{1}{2}\langle h, f''(x)h\rangle + O(h)^3 \tag{27}$$

Thus the sign of the differential variation in f may be determined from the definite-ness of the operator $f''(x)$, whenever $f'(x)$ is zero.

Here follows the main results, which is just an application of the above ideas to the problem at hand.

Theorem 3. If there exists a finite F^* that stabilzes (A,B,C) and satisfies the necessary conditions of optimality, and if the $mr \times mr$ symmetric matrix P is defined by:-

$$P = [4(M \otimes C)(A_c \otimes I + I \otimes A_c)^{-1}(B \otimes LC^T)]_s + R \otimes CLC^T \qquad (28)$$

Then $J(F^*)$ is a local minimum if P is positive definite.

Proof. We have:-

$$J'(F^*) = (RF^*C - B^T K)L\,C^T = 0 \qquad (29)$$

where K and L have been evaluated at F^*. Therefore:-

$$\begin{aligned} & J'(F^*+\Delta F) - J'(F^*) \\ = \; & (RF^*C - B^T K)\,\Delta\, L\, C^T \\ + \; & (R\,\Delta FC - B^T \Delta K)\, L\, C^T \\ + \; & (R\,\Delta FC - B^T \Delta K)\Delta L\, C^T \end{aligned} \qquad (30)$$

where

$$\begin{aligned} & \Delta\, L(A - BF^*C - B\Delta FC)^T \\ & + (A - BF^*C - B\,\Delta FC)\Delta\, L \\ & - L\,C^T\,\Delta F^T B^T - B\,\Delta FCL = 0 \end{aligned} \qquad (31)$$

and

$$\begin{aligned} & \Delta\, K(A - BF^*C - B\Delta FC) \\ & + (A - BF^*C - B\,\Delta F\,C)^T \Delta K \\ & + C^T\,\Delta F^T RFC + C^T F^T R\,\Delta FC \\ & - C^T\,\Delta F^T B^T K - KB\,\Delta FC \\ & - C^T\,\Delta F^T R\,\Delta FC = 0 \end{aligned} \qquad (32)$$

From continuity, and stability arguments, it may be argued that as F becomes small, we get:-

$$\delta^2 J = M\delta LC^T + (R\delta FC - B^T\delta K)\, L\, C^T + W_1(\delta F) \qquad (33)$$

where
$$\delta\, L\, A_c^T + A_c\,\delta\, L - LC^T\delta F^T B^T - B\delta\, F\, C\, L + W_2(\delta F) = 0 \qquad (34)$$

$$\delta K\, A_c + A_c^T\,\delta K + M^T\delta FC + C^T\delta F^T M + W_3(\delta F) = 0 \qquad (35)$$

where
$$M = RF^*C - B^T K$$

$$A_c = A - BF^*C$$

Nowit can be checked that if Λ_c is stable, and the mappings are continuous (bounded);

$$\frac{\|w_1(\delta F)\|}{\|\delta F\|} \to 0$$

$$\frac{\|w_2(\delta F\|}{\|\delta F\|} \to 0$$

$$\frac{\|w_3(\delta F)\|}{\|\delta F\|} \to 0$$

as $$\|\delta F\| \to 0 \qquad (36)$$

Since the necessary conditions for optimality are fulfilled, the variation in J is zero to first order, and we may write:-

$$\delta J = \tfrac{1}{2} < \delta F, M\delta LC^T + (R\delta FC - B^T\delta K)LC^T > + O(\delta F)^3 \qquad (37)$$

writing δF, δL, and δK as column vectors, and using the results in appendix 1, we get:-

$$\delta J = \tfrac{1}{2} < \delta F_v, (M \otimes C)\,\delta L_v + (R \otimes CLC^T)\,\delta F_v - (B^T \otimes CL)\,\delta K_v > + O(\delta F_v)^3 \qquad (38)$$

where
$$(A_c \otimes I + I \otimes A_c)\,\delta L_v - (B \otimes LC^T + LC^T \otimes B)\,\delta F_v = 0 \qquad (39)$$

$$(A_c^T \otimes I + I \otimes A_c^T)\,\delta K_v + (M^T \otimes C^T + C^T \otimes M^T)\,\delta F_v = 0 \qquad (40)$$

noting that the system is stable, and substituting for δL_v and δK_v, we get:-

$$\delta J = \tfrac{1}{2}\,\delta F_v^T\, P\, \delta F_v + O(\delta F_v)^3 \qquad (41)$$

where

$$\begin{aligned} P = {} & (M \otimes C)(A_c \otimes I + I \otimes A_c)^{-1}(B \otimes LC^T) \\ & + (M \otimes C)(A_c \otimes I + I \otimes A_c)^{-1}(LC^T \otimes B) \\ & + (B^T \otimes CL)(A_c^T \otimes I + I \otimes A_c^T)^{-1}(M^T \otimes C^T) \\ & + (B^T \otimes CL)(A_c^T \otimes I + I \otimes A_c^T)^{-1}(C^T \otimes M^T) \\ & + R \otimes CLC^T \end{aligned} \qquad (42)$$

Noting that the third term is just the transpose of the first term, and that:-

$$F_v^T (M \otimes C)(A_c \otimes I + 1 \otimes A_c)^{-1}(L\,C^T \otimes B)\,F_v$$
$$= F_v^T (C \otimes M)(A_c^T \otimes I + I \otimes A_c^T)^{-1}(B \otimes LC^T)\,F_v \qquad (43)$$

for all F_v ; the expression for P reduces to that given in the original theorem statement. And the rest of the theorem follows directly.

Now we know from the previous section that the second variation cannot be positive definite if CLC^T is singular, in this case , we have the corollary:-

Corollary: If there exists a finite F_m of minimum norm that stabilizes (A,B,C) and satisfies the necessary conditions of optimality, then $J(F_m)$ is a local minimum, in the stated class of F's , if $(I \otimes N^T C^T)\,P\,(I \otimes CN)$ is positive definite. Where the identity matrix is $r \times r$ and N is any factor of full rank, such that $NN^T = L$.

The proof of this corollary is left to the reader.

Conclusion:

Sufficient conditions have been found for the non-negativity of the second variation; also, it has been shown that it is not necessary for $C\,L\,C^T$ to be invertible in order that an optimal control may exist. However, it should be noted that the results depend heavily on the optimal control being a stabilizing control, but sufficient conditions are given for this too. The recognition that the second variation is a linear operator in matrix space, which may be re-written as a matrix operator in vector space (by using the Kronecker product), suggests that a second order algorithm may be used to find the solution.

APPENDIX 1.

If S represents the matrix,

$$\begin{vmatrix} S_{11} & S_{12} & \cdots & S_{1n} \\ S_{21} & S_{22} & \cdots & S_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ S_{m1} & S_{m2} & & S_{mn} \end{vmatrix}$$

then let S_v denote the column vector whose elements are , S_{11}, S_{12} ... , S_{1n}, S_{21}, S_{22}, ... , S_{2n}, ... ,S_{m1}, S_{m2}, ... , S_{mn}. Then the following pairs may be checked to be equivalent.

(i) $A^T S + S A + C = O$

$(A^T \otimes I + I \otimes A^T) S_v + C_v = O$

(ii) $Q = A S B$

$Q_v = (A \otimes B^T) S_v$

(iii) $Q = B^T S^T A^T$

$Q_v = (B^T \otimes A) S_v$

as well as the following expressions :-

(iv) $(A \otimes B)^T = (A^T \otimes B^T)$

(v) $S_v^T (A \otimes B)(C \otimes D) S_v = S_v^T (B \otimes A)(D \otimes C) S_v \quad \forall S_v$

References:

1. KALMAN R.E., and BUCY R.S., "New results in linear filtering and prediction theory," Trans. ASME, J. Basic Engrg., Vol. 83, pp 95-108, March 1961.

2. LUENBERGER D.G., "Observers for multivariable systems," I.E.E.E. Trans. Automatic Control, Vol. AC-11, pp 190-197, April 1966.

3. LEVINE W.S., and ATHANS MICHAEL, "On the Determination of the Optimal Constant Output Feedback Gains for Linear Multivariable Systems," I.E.E.E. Trans. Automatic Control, Vol.AC-15, pp 44-48, February 1970.

4. BROCKETT R.W., "Finite Dimensional Linear Systems," John Wiley and Sons, pp 61-62, 1970.

5. WONHAM W.M., "On a Matrix Riccati Equation of Stochastic Control," SIAM J. Control, Vol. 6, No. 4, 1968, pp 681-697.

6. LADAS G.E., and LAKSHMIKANTHAM V., "Differential Equations in Abstract Spaces," Academic Press, pp 12-20, 1972.

Author Jones E L
Name of thesis The Optimal Control of Infinite Dimensional Linear Systems 1976

PUBLISHER:
University of the Witwatersrand, Johannesburg
©2013

LEGAL NOTICES:

Copyright Notice: All materials on the University of the Witwatersrand, Johannesburg Library website are protected by South African copyright law and may not be distributed, transmitted, displayed, or otherwise published in any format, without the prior written permission of the copyright owner.

Disclaimer and Terms of Use: Provided that you maintain all copyright and other notices contained therein, you may download material (one machine readable copy and one print copy per page) for your personal and/or educational non-commercial use only.

The University of the Witwatersrand, Johannesburg, is not responsible for any errors or omissions and excludes any and all liability for any errors in or omissions from the information on the Library website.