On the General Theory of Control Systems

R. E. KALMAN

Introduction

In no small measure, the great technological progress in automatic control and communication systems during the past two decades has depended on advances and refinements in the mathematical study of such systems. Conversely, the growth of technology brought forth many new problems (such as those related to using digital computers in control, etc.) to challenge the ingenuity and competence of research workers concerned with theoretical questions.

Despite the appearance and effective resolution of many new problems, our understanding of fundamental aspects of control has remained superficial. The only basic advance so far appears to be the *theory of information* created by Shannon[1]. The chief significance of his work in our present interpretation is the discovery of general 'laws' underlying the process of information transmission, which are quite independent of the particular models being considered or even the methods used for the description and analysis of these models. These results could be compared with the 'laws' of physics, with the crucial difference that the 'laws' governing man-made objects cannot be discovered by straightforward experimentation but only by a purely abstract analysis guided by intuition gained in observing present-day examples of technology and economic organization. We may thus classify Shannon's result as belonging to the *pure theory* of communication and control, while everything else can be labelled as the *applied theory*; this terminology reflects the well-known distinctions between pure and applied physics or mathematics. For reasons pointed out above, in its methodology the pure theory of communication and control closely resembles mathematics, rather than physics; however, it is not a branch of mathematics because at present we cannot (yet?) disregard questions of physical realizability in the study of mathematical models.

This paper initiates study of the pure theory of control, imitating the spirit of Shannon's investigations but otherwise using entirely different techniques. Our ultimate objective is to answer questions of the following type: What kind and how much information is needed to achieve a desired type of control? What intrinsic properties characterize a given unalterable plant as far as control is concerned?

At present only superficial answers are available to these questions, and even then only in special cases.

Initial results presented in this Note are far from the degree of generality of Shannon's work. By contrast, however, only *constructive* methods are employed here, giving some hope of being able to avoid the well-known difficulty of Shannon's theory: methods of proof which are impractical for actually constructing practical solutions. In fact, this paper arose from the need for a better understanding of some recently discovered computation methods of control-system synthesis[2-5]. Another by-product of the paper is a new computation method for the solution of the classical Wiener filtering problem[7].

The organization of the paper is as follows:

In Section 3 we introduce the models for which a fairly complete theory is available: dynamic systems with a finite dimensional state space and linear transition functions (i.e. systems obeying linear differential or difference equations). The class of random processes considered consists of such dynamic systems excited by an uncorrelated gaussian random process. Other assumptions, such as stationarity, discretization, single input/single output, etc., are made only to facilitate the presentation and will be absent in detailed future accounts of the theory.

In Section 4 we define the concept of *controllability* and show that this is the 'natural' generalization of the so-called 'deadbeat' control scheme discovered by Oldenbourg and Sartorius[21] and later rederived independently by Tsypkin[22] and the author[17].

We then show in Section 5 that the general problem of optimal regulation is solvable if and only if the plant is completely controllable.

In Section 6 we introduce the concept of *observability* and solve the problem of reconstructing unmeasurable state variables from the measurable ones in the minimum possible length of time.

We formalize the similarities between controllability and observability in Section 7 by means of the *Principle of Duality* and show that the Wiener filtering problem is the natural dual of the problem of optimal regulation.

Section 8 is a brief discussion of possible generalizations and currently unsolved problems of the pure theory of control.

Notation and Terminology

The reader is assumed to be familiar with elements of linear algebra, as discussed, for instance, by Halmos[8].

Consider an n-dimensional real vector space X. A *basis* in X is a set of vectors $\mathbf{a}_1 \ldots, \mathbf{a}_n$ in X such that any vector $\mathbf{x}$ in X can be written uniquely as

$$\mathbf{x} = x_1\mathbf{a}_1 + \ldots + x_n\mathbf{a}_n \qquad (1)$$

the x_i being real numbers, the *components* or *coordinates* of $\mathbf{x}$. Vectors will be denoted throughout by small bold-face letters.

The set $X^\star$ of all real-valued linear functions $\mathbf{x}^\star$ (= *covectors*) on X, with the 'natural' definition of addition and scalar multiplication, is an n-dimensional vector space. The value of a covector $\mathbf{y}^\star$ at any vector $\mathbf{x}$ is denoted by $[\mathbf{y}^\star, \mathbf{x}]$. We call this the *inner product* of $\mathbf{y}^\star$ by $\mathbf{x}$. The vector space $X^\star$ has a natural basis $\mathbf{a}^\star_1 \ldots, \mathbf{a}^\star_n$ associated with a given basis in X; it is defined by the requirement that

$$[\mathbf{a}^\star_i, \mathbf{a}_j] = \delta_{ij} \qquad (2)$$

Using the 'orthogonality relation' 2, we may write 1 in the form

$$\mathbf{x} = \sum_{i=1}^{n} [\mathbf{a}^\star_i, \mathbf{x}]\mathbf{a}_i \qquad (3)$$

which will be used frequently.

For purposes of numerical computation, a vector may be considered a matrix with one column and a covector a matrix

481

with one row. Then the inner product is expressed by the product of these two matrices.

By means of the dual basis, one can set up an isomorphism between X and $X^\star$: to each element $\mathbf{x}$ in X there corresponds one and only one element $\mathbf{x}^\star$ in $X^\star$ such that $\mathbf{x}^\star$ has the same coordinates with respect to $\mathbf{a}^\star_1 \ldots, \mathbf{a}^\star_n$ as $\mathbf{x}$ with respect to $\mathbf{a}_1 \ldots, \mathbf{a}_n$. Note, however, that this correspondence is not a 'natural' one since it depends on the particular basis chosen.

This having been said, it is clear that $[\mathbf{x}^\star, \mathbf{x}]^{\frac{1}{2}}$ is defined for every $\mathbf{x}$ in X—this is the *euclidean norm* of $\mathbf{x}$ (with respect to a fixed basis in X).

Linear transformations of X into itself or their matrix representations will be denoted by bold-face capitals. A special case is a *positive-definite* [*positive semi-definite*] transformation or matrix $\mathbf{Q}$, which is defined by requiring that the quadratic form $[\mathbf{x}^\star, \mathbf{Qx}]$ be positive [non-negative] for every $\mathbf{x} \neq 0$. It is easily seen that if $\mathbf{Q}$ is positive-definite $[\mathbf{x}^\star, \mathbf{Qx}]^{\frac{1}{2}}$ is a norm which is called the *generalized euclidean norm* of $\mathbf{x}$ (again defined with respect to a fixed basis in X); to remember this fact, we introduce the special notation

$$[\mathbf{x}^\star, \mathbf{Qx}] = \|\mathbf{x}\|^2_\mathbf{Q} \tag{4}$$

it being understood that $\mathbf{Q}$ is positive-definite or semi-definite.

Dynamic Models

In any mathematical study of control, it is necessary to restrict attention to a certain class of *models* which represent the dynamic phenomena to be studied. The purpose of this section is to introduce the requisite terminology on models.

We call the physical object to be controlled the *plant*. In practice, this may be (i) an automobile, (ii) an airplane, (iii) a chemical reactor. To control the plant, we must be able to change certain physical quantities, called *inputs*. These may be (i) the accelerator or steering wheel of an automobile, (ii) the control surfaces of an airplane, (iii) the material or heat input to a chemical reactor. Another requirement for control is that one must be able to measure the behaviour of the plant; the physical variables of the plant which can be measured directly are called *outputs*. Examples are (i) the speed (measured by a speedometer) or position (measured by the human eye) of a car; (ii) the altitude, speed, or acceleration of an airplane; (iii) the temperature, colour, quantity, etc., of chemicals formed by a reaction. The following terminology is standard in describing *dynamic systems*:

The *state* of a dynamic system is the smallest collection of numbers which must be specified at time $t = t_0$ in order to be able to predict the behaviour of the system for any time $t \geq t_0$. In other words, the state is the minimal 'record' of the past history needed to predict the future behaviour. The future states $\mathbf{x}(t)$ of the system which was in state $\mathbf{x}(t_0)$ at time t_0 are then given by the *transition function* or *motion*

$$\mathbf{x}(t) = \boldsymbol{\phi}(t; \mathbf{x}(t_0), t_0) \quad (t \geq t_0) \tag{5}$$

For this definition to be consistent, we stipulate

$$\boldsymbol{\phi}(t_0; \mathbf{x}(t_0), t_0) = \mathbf{x}(t_0) \tag{6}$$

$$\boldsymbol{\phi}(t_2; \boldsymbol{\phi}(t_1; \mathbf{x}(t_0), t_0), t_1) = \boldsymbol{\phi}(t_2; \mathbf{x}(t_0), t_0), t_2 \geq t_1 \geq t_0 \tag{7}$$

the second condition assuring that $\boldsymbol{\phi}$ is unique.

By the Principle of Causality, any dynamic system may be described from the 'state' point of view. Moreover, it is clear that any output of the system must be a function of the state. The transition function $\boldsymbol{\phi}$ depends on the inputs to the system which take place after $t = t_0$, but as a rule it is not necessary to indicate this dependence explicitly.

The plants to be studied in this paper are dynamic systems subject to two essential (very restrictive but standard) assumptions:

The state space X *of the plant is an* n-*dimensional vector space* (8)

The transition function $\boldsymbol{\phi}$ *depends linearly on the initial state* $\mathbf{x}(t_0)$ *and the inputs after* $t = t_0$ (9)

Less abstractly, these assumptions mean that the plant is governed by a linear vector differential equation (*continuous-time dynamic system*):

$$d\mathbf{x}/dt = \mathbf{F}(t)\mathbf{x} + \mathbf{D}(t)\mathbf{u}(t)$$
$$\mathbf{y}(t) = \mathbf{B}(t)\mathbf{x}(t) \tag{10}$$

or a vector difference equation (*discrete time dynamic system*):

$$\mathbf{x}(t_{k+1}) = \boldsymbol{\Phi}(t_k)(\mathbf{x}(t_k) + \mathbf{A}(t_k)\mathbf{u}(t_k))$$
$$\mathbf{y}(t_k) = \mathbf{B}(t_k)\mathbf{x}(t_k) \tag{11}$$

where $\mathbf{u}(t)$ resp. $\mathbf{u}(t_k)$ is an n-dimensional vector $(m \leq n)$ denoting the *inputs* to the plant; $\mathbf{y}(t)$ resp. $\mathbf{y}(t_k)$ is a p-dimensional vector $(p \leq n)$ denoting the *outputs* of the plant; the coordinates $x_i(t)$ of the state (with respect to some fixed basis) are called *state variables*.

Although our theory extends to classes of plants of type 10 or 11, in the interest of simplifying conceptual and mathematical technicalities, we make the following, inessential, further assumptions:

The plant is stationary. (In other words, $\mathbf{F}(t)$, $\mathbf{D}(t)$, $\boldsymbol{\Phi}(t_k)$ and $\mathbf{A}(t_k)$ are constant matrices) (12)

The plant has a single input. (In other words, $m = 1$) (13)

The plant has a single output. (In other words, $p = 1$) (14)

With these assumptions, the equations of the plant become:

$$d\mathbf{x}/dt = \mathbf{F}\mathbf{x} + \mathbf{d}u_1(t)$$
$$y_1(t) = [\mathbf{b}^\star, \mathbf{x}(t)] \tag{15}$$

or

$$\mathbf{x}((k+1)T) = \boldsymbol{\Phi}[\mathbf{x}(kT) + \mathbf{a}u_1(kT)] \quad (T > 0)$$
$$y_1(kT) = [\mathbf{b}^\star, \mathbf{x}(kT)] \tag{16}$$

It will be useful, though inessential, to assume also that

$$\boldsymbol{\Phi} \text{ in } 11 \text{ is non-singular} \tag{17}$$

It is a triviality to reduce 15 to 16 by means of the operation of *sampling* or *discretization*. This step corresponds to finding the z-transform[9] of equations 15; however, the 'state' point of view and the use of matrix notation obviates the need for transform considerations. The principal assumption is that $u_1(t)$ is a piece-wise constant function over the intervals $\ldots$, $(-T, 0)$, $(0, T) \ldots$, where T is positive but otherwise arbitrary. Using this, the explicit solution of equation 15 is[2-4, 10, 11]:

$$\mathbf{x}((k+1)T) = (\exp T\mathbf{F})\mathbf{x}(kT) + \int_{kT}^{(k+1)T} [\exp (T-\tau)\mathbf{F}]\mathbf{d}u_1(kT) \, d\tau$$

482

492

Therefore if

$$\Phi = \exp T\mathbf{F}$$

$$\mathbf{a} = \left(\int_0^T \exp \tau\mathbf{F} \, d\tau \right) \mathbf{d}$$

equation 15 reduces to 16–17, since $\exp T\mathbf{F}$ is non-singular for any T. The calculations required to find Φ and $\mathbf{a}$ can be conveniently performed either by Laplace transform techniques or by evaluating rapidly converging power series[11]. We refer in particular to references 2, 3 and 11 for pedagogical material showing how the concepts of this section are related to conventional control engineering terminology.

The vectors $\mathbf{d}$ respectively $\mathbf{a}$ and $\mathbf{b}$ specify how the plant is *constrained* with respect to inputs or outputs. If every state can be affected directly by the input, the plant is *not constrained with respect to input*; if every state variable $x_i(t)$ can be measured, the plant is *not constrained with respect to output*. A plant which has no input is said to be *free*.

To aid the intuition, it will be convenient to make use of block diagram notation as is customary in control engineering. See *Figure 1*, which represents equation 16. The sole difference
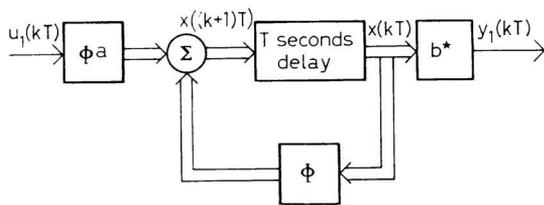

Figure 1

between the present notation and the customary one is that the 'blocks' represent, in general, matrix multiplication as indicated by the fat arrows in *Figure 1*.

To control a plant of type 15 or 16, it is necessary to express the input $u_1(t)$ or $u_1(kT)$ as a function of the state and of all present and future desired states $\mathbf{x}^d(t)$. (It is immediate from the definition of the state that u_1 need not depend on any other functions of time, but of course will depend on the constants Φ, $\mathbf{a}$ characterizing the plant.) If the state cannot be measured directly, i.e. if 14 is assumed, then u_1 will depend on all present and future desired states and all present and past measured values of the output. We call some particular dependence of this sort the *control law of the plant*.

The concept of 'desired state' is a generalization of what is usually called in the control engineering literature 'command input' or 'desired output'. Again, in the interests of simplicity, (for a more detailed treatment, see references 5 and 6) we assume that

The desired state of the plant is identically zero for all t   (18)

In other words, the object of control will be to transfer any initial state to the *equilibrium state* $\mathbf{0}$. (This terminology is due to the fact that if $\mathbf{x}(t_0) = \mathbf{0}$ and $u_1(t) = 0$ for all $t \geq t_0$, then $\phi(t; \mathbf{0}, t_0) = \mathbf{0}$ for all $t \geq t_0$, i.e. the plant remains at $\mathbf{0}$ for all $t \geq t_0$.)

Under these hypotheses, the control law will be expressed in the continuous time or discrete time case by

$$u_1(t) = \chi(\mathbf{x}(t), t) \quad \text{or} \quad u_1(kT) = \chi(\mathbf{x}(kT), kT)$$

if the plant is not constrained at the output, and by

$$u_1(t) = \chi(y_1(s), s \leq t; t)$$

or

$$u_1(kT) = \chi(y_1(jT), j \leq k; kT)$$

if assumption 14 holds. In these relations, χ is an arbitrary function; if χ is linear in $\mathbf{x}$ or y_1, we say the *control law is linear*; if χ does not depend explicitly on t or kT, the *control law is stationary*.

It will be convenient to refer to $u_1(t)$ or $u_1(kT)$ as the *control signal*; the values kT of time are called *sampling points*.

Controllability

In this section, we shall assume (except in 29) that the plant in question is always *linear, stationary, single input*.

Our first objective is to find an intrinsic characterization of the manner in which a plant is constrained in regard to control. This leads to:

A state $\mathbf{x}$ *of a plant is said to be 'controllable' if there exists a control signal* $u_1(t)$ *defined over a finite interval* $0 \leq t \leq t_1$ *such that* $\phi(t_1; \mathbf{x}, 0) = \mathbf{0}$. *In general, the time* t_1 *will depend on* $\mathbf{x}$. *If every state is controllable, the plant is said to be 'completely controllable'*   (19)

This concept originated as follows. In a discussion of a now well known paper by Bergen and Ragazzini[12], it was pointed out in 1954 that it is possible to design a sampled-data controller for any single input/single output linear stationary plant in such a way that the error in response to a step input is identically zero after a finite length of time. (This is nowadays usually called a *dead-beat system*.) While this observation is certainly correct in the 'usual' cases encountered in practice, no rigorous proof was possible at that time. In 1957 essential improvements were obtained in the original arguments by means of the 'state' method; an engineering exposition of this (without complete proofs) has appeared[2]. Now we witness an illustration of the dictum that 'a good theorem eventually becomes a definition'.

A necessary and sufficient condition for complete controllability in the discrete time case is the following[2, 5, 6, 13]:

a discrete time plant is completely controllable (i) *if and* (ii) *only if the vectors* $\mathbf{a}, \Phi^{-1}\mathbf{a} \ldots, \Phi^{-n+1}\mathbf{a}$ *are linearly independent*   (20)

Before proving this theorem, we introduce some terminology. For every positive integer i we denote $\Phi^{-i+1}\mathbf{a}$ by $\mathbf{e}_i$. It is easy to show by induction, using the fact that Φ is non-singular[13]:

The set $\Gamma(q)$ *of all initial states which can be transferred to the origin in at most* q *steps by the application of appropriate control signals is given by*   (21)

$$\Gamma(q) = \left\{ \mathbf{x}; \mathbf{x} = \sum_{i=1}^{q} \xi_i \mathbf{e}_i \right\};$$

the ξ_i *being arbitrary real numbers.*

We need also the easily proven fact:

If $\mathbf{e}_{k+1}$ *is linearly dependent on* $\mathbf{e}_1 \ldots, \mathbf{e}_k$, *then so is* $\mathbf{e}_m$ *for every* $m > k$. *Hence, in particular, if* $\Gamma(k) = \Gamma(k+1)$, *then* $\Gamma(k) = \Gamma(m)$ *for all* $m \geq k$.   (22)

Proof of 20. (i) If the set $\mathbf{e}_1 \ldots, \mathbf{e}_n$ is linearly independent, it is a basis for X. Hence every state can be transferred to the

483

origin in at most $q = n$ steps, and so the plant is completely controllable.

(ii) Conversely, suppose that only $k < n$ of the vectors $e_1 \ldots, e_n$ are linearly independent but the plant is completely controllable. By 22, it follows that actually the *first* k of these vectors are linearly independent and therefore $\Gamma(k) = \Gamma(m)$ for all $m \geqslant k$. Hence $\Gamma(k)$ is the set of all controllable states. But dim $\Gamma(k) = k < n$ so that $\Gamma(k) \neq X$ which contradicts the fact that the plant is completely controllable.

Clearly $\Gamma(k)$ in the preceding paragraph is the linear space generated by the infinite set of vectors $(a, \Phi^{-1}a \ldots)$; from this, it is obvious that $\Gamma(k)$ is invariant under the transformation Φ^{-1}, i.e. that $\Phi^{-1}\Gamma(k) = \Gamma(k)$ or $\Phi\Gamma(k) = \Gamma(k)$. Hence 20 can also be expressed by saying

> *A discrete time plant is completely controllable if and only if the only Φ-invariant subspace of* X *containing* a *is* X *itself.* (23)

In an entirely similar fashion, we can also obtain conditions for controllability in the continuous time case, or, more precisely, directly from the differential equations 15:

> *In the case of a continuous time plant the following propositions are equivalent:* (i) *the plant is completely controllable;* (ii) *the vectors* d, Fd ..., $F^{n-1}d$ *are linearly independent;* (iii) *the only* F-*invariant subspace of* X *containing* d *is* X *itself.* (24)

This condition has also been used, independently and in a different context, by Pontryagin[15]. However, in his case the required independence appears as a purely technical mathematical requirement, unaccompanied by the intuitive justification contained in definition 19 of controllability.

From 23 and 24 it follows immediately:

> *A discrete time [continuous time] plant is completely controllable for SOME* a [*SOME* d] *if and only if in the Jordan canonical form of Φ [F] no two blocks are associated with the same eigenvalue.* (25)

Proof of 25. This is an immediate consequence of the definition of the Jordan canonical form, for which see Coddington and Levinson, reference 10, Chapter 3, Section 1. A *sufficient* condition for 25 is that all eigenvalues be distinct.

> *Example.* The following plant is not completely controllable for any d. Take two transfer functions in parallel, each being $1/(s+1)$. Then $-F$ = unit matrix, and $\Gamma(1) = \Gamma(2)$ is the one-dimensional linear space generated by the vector d.
>
> Now assume that the output y_1 of the plant is the sum $x_1 + x_2$ of the outputs of the two transfer functions. The total transfer function is then $(d_1 + d_2)(s+1)/(s+1)^2$. In the usual semi-rigorous engineering treatment, one would naturally say that this is the 'same' as the transfer function $(d_1 + d_2)/(s+1)$, i.e. the factor $(s+1)$ can be cancelled. What is really involved here is replacing x_1, x_2 by y_1; control can then be exerted on y_1 but surely not on $y_2 = d_2x_1 - d_1x_2$. Thus if cancellation is possible, the effect of the control signal on some of the state variables is lost. (26)

Combining this observation with 25, we may state, in engineering language:

> *A single input/single output plant (continuous time or discrete time) is completely controllable if and only if the input excites all natural frequencies of the plant; in other words, if no cancellation of poles is possible in the transfer function.* (27)

The very elegance of signal-flow-graph and other similar methods for obtaining transfer functions of, say, networks, often obscures the cancellation of terms. In cases of doubt, conditions 20 or 24 must be employed.

It may happen in particular that the forbidden cancellation takes place in passing from the continuous time to the discrete time case. In general, this can happen only if the plant contains complex poles and then only in the following two ways. Consider the (continuous time) transfer function

$$(a) \qquad G(s) = b^2s/(s^2 + 2as + b^2)$$

The Z transform (discrete time transfer function) is given by[9]:

$$Z[G(s)] = [e^{aT}z(e^{aT}z - \cos bT)]/[e^{2aT}z^2 - 2(e^{aT}\cos bT)z + 1]$$

Unless $bT \neq r\pi$ (r = integer), the factor $(e^{aT}z - \cos r\pi)$ appears in both the numerator and the denominator.

The other type of forbidden cancellation is the following. Consider the continuous time transfer function

$$(b) \qquad G(s) = 3/(s^2 + 1)(s^2 + 4)$$

The corresponding discrete time transfer function is

$$Z[G(s)] = (z \sin T)/(z^2 - 2z \cos T + 1) - (z \sin 2T)/2(z^2 - 2z \cos 2T + 1)$$

Now if the sampling period is $T = 2\pi/3$, we find that

$$Z[G(s)] = 3\sqrt{2}\, z/4(z^2 - \sqrt{2}\, z + 1)$$

and if $T = 2\pi$, then $Z[G(s)] = 0$, so that with these sampling periods the plant is surely not completely controllable.

We therefore conclude (a direct, rigorous proof is not difficult):

> *A plant which is completely controllable in the absence of sampling remains completely controllable after the introduction of sampling if and only if, for every eigenvalue $\lambda(F)$ of F Re λ_i (F) = Re λ_j (F) implies* (28)
>
> Im $[\lambda_i(F) - \lambda_j(F)] \neq 2r\pi/T$ (r = positive integer)

This result provides the rigorous basis for the developments in reference 2. The characterization of the set of all controllable states may be regarded as a fundamental problem in the pure theory of control. While in the linear case the problem is now completely solved[14], only fragmentary results are available in the stochastic or non-linear cases. One well known fact is[13,15,16]:

> *Consider a plant which is linear and completely controllable. Introduce ideal saturation, i.e. require that $|u_1(t)| \leqslant 1$. Let X be the direct sum of two Φ-invariant [F invariant] subspaces $Y + Z$, such that the eigenvalues of Φ restricted to Y are less than or equal to one in absolute value [the eigenvalues of F restricted to Y have non-positive real parts].*
>
> *The set of all controllable states under saturation is the Cartesian product $Y \times W$, where W is a convex, compact subset of Z containing the origin.* (29)

The principal consequence of complete controllability is the following:

484

I. FUNDAMENTAL THEOREM FOR LINEAR CONTROL SYSTEMS

I. FUNDAMENTAL THEOREM FOR LINEAR CONTROL SYSTEMS—Consider a discrete time, completely controllable plant. Every state is transferred to the origin in a minimal number ($\leqslant$ n) of sampling periods if and only if the control law is given by:

$$u_1(kT) = -[e^\star_1, x(kT)] \qquad (30)$$

Here $e^\star_1$ is the first vector in the dual basis $e^\star_1 \ldots, e^\star_n$ of the basis $e_1 \ldots, e_n$. $\qquad$ (31)

A 'vector' block diagram of the closed loop system is shown in *Figure 2*. We see that the control law 30 satisfies the principle of feedback, since the control signal is computed from the measurements of the output, i.e. of the state. The feedback nature of the control law is not an *a priori* assumption but is forced on us by the requirement that every state be transferred to the origin in minimum time. This control law may not be
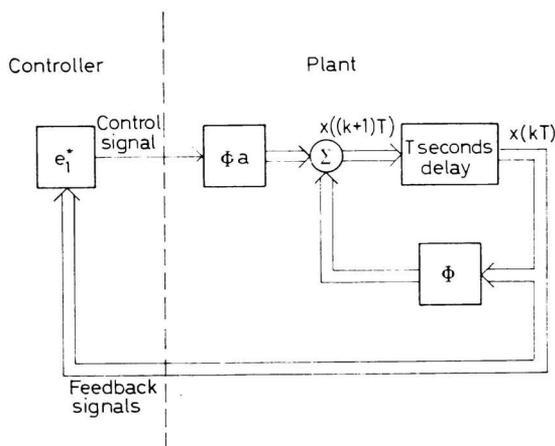

Figure 2

physically realizable, however, since constraints may prevent the instantaneous measurement of all state variables.

If a linear continuous time plant is discretized by sampling, the choice of the sampling period is arbitrary. Hence we see from 31 that:

In a continuous time, completely controllable plant any initial state may be returned to the origin in an arbitrarily short length of time τ by taking $T = \tau/n$, sampling, and using the control law 30. $\qquad$ (32)

It is also possible to achieve this without discretizing[14], but then the control law becomes non-stationary (time-varying feedback coefficients), which usually makes physical realization impractical.

Proof of Theorem I. At any time kT, any vector in $\Gamma(m)$ (where $1 \leqslant m \leqslant n$) can be represented in the form:

$$x(kT) = \sum_{i=1}^{m} \xi_i e_i = \sum_{i=1}^{m} [e^\star_i, x(kT)]e_i$$

Hence, using the control law 30,

$$x((k+1)T) = \Phi \sum_{i=2}^{m} [e^\star_i, x(kT)]e_i$$

$$= \sum_{i=1}^{m-1} [e^\star_{i+1}, x(kT)]e_i \quad \text{since } \Phi e_{i+1} = e_i$$

$$= \sum_{i=1}^{m-1} [(e^\star_i, x(k+1)T)]e_i$$

Suppose now that $\xi_m \neq 0$, i.e. that $x(kT)$ is in $\Gamma(m)$ but not in $\Gamma(m-1)$. From the preceding calculation, we see that 30 is the *only* control law assuring that $x((k+1)T)$ will be in $\Gamma(m-1)$, so that everything is proved.

A short reflection on this proof shows that 31 could have been phrased also in the following fashion:

A plant is completely controllable if and only if there exists a sampled-data controller which has 'dead-beat' response. $\qquad$ (33)

In various discussions of the dead-beat scheme which have appeared in the literature, usually without acknowledgement of the original contributions[21, 22, 17] (for example in reference 18, pp. 195-198), the possibility of dead-beat response is confused with extraneous matters, such as requiring the plant to be stable in the absence of control.

By relaxing the 'minimality' condition in 31, and demanding only that every state be transferred to the origin in at most $p \geqslant n$ steps, the control law can be made to satisfy other desiderata, for instance, minimizing the energy required for control[5].

In order to justify the title 'fundamental theorem' for 31, we now show that this result implies and considerably improves the conventional Wiener–Hall theory of optimization of control systems in the absence of noise.

Optimization of Regulators

In this section, the plant will always be *linear, stationary, single input*, and *discrete time*.

To optimize a regulating system, it is necessary to introduce the concept of a *performance index*. This is usually taken as the integrated (or, in the case of discrete time systems, summed) *error* along the motions of the system. It is inessential but convenient to define the error as a positive definite quadratic form $\mu^k \|x(kT)\|^2_Q$ (see Section 2), where Q is a positive definite matrix and μ is a positive number. Then the performance index is a function of the initial state x given by

$$V(x) = \sum_{k=1}^{\infty} \mu^k \|\phi(kT; x, 0)\|^2_Q \qquad (34)$$

We can now define

OPTIMAL REGULATOR PROBLEM—Find a control such that 34 is minimized for every initial state x $\qquad$ (35)

At first sight, it is doubtful whether such a problem is meaningful at all, since $V(x)$ might turn out to be infinite, at least for some initial states. It is therefore remarkable that we have

II. GENERAL EXISTENCE AND UNIQUENESS THEOREM—If Q in 34 is positive definite, the solution of the optimal regulator problem exists for any $\mu > 0$ (i) if and (ii) only if the plant is completely controllable. Moreover, the resulting control system has the properties: (iii) its control law is unique, stationary, and linear; (iv) if $\mu \geqslant 1$ the closed loop system is asymptotically stable and $V°(x)$ is one of its Liapunov functions ($V°(x)$ is defined by 34, where ϕ is the transition function of the optimized closed loop system). $\qquad$ (36)

Proof. Let $V_N(x)$ denote the value of $V(x)$ when the summation indicated in 34 is between the limits $(1, N)$. Clearly then

$$V_{N+1}(x) = \mu\{V_N(\phi(T; x, 0)) + \|x\|^2_Q\} \qquad (37)$$

485

Let $V^\circ_N(\mathbf{x})$ be the value of $V_N(\mathbf{x})$ minimized with respect to the possible control signals at the sampling points $0 \ldots (N-1)T$. The *existence* of V°_N is proved by induction, see below.

It is now convenient to invoke the intuitively obvious

PRINCIPLE OF OPTIMALITY (Bellman[19])—*In order that the control signals defined at the sampling points* $0 \ldots , (N+1)T$ *be optimal. it is necessary, whatever the state resulting from the first control signal, that the control signals defined at the sampling points* $T \ldots , (N+1)T$ *be optimal.* (38)

This principle leads to the step-by-step determination of the optimal control law and of V°.

The plant being assumed stationary, the optimal control law does not depend on the time at which the initial state $\mathbf{x}$ occurs; so that if $V^\circ_N(\mathbf{x})$ is optimal for the interval $(0, NT)$, then $V^\circ_N(\boldsymbol{\phi}(T; \mathbf{x}, 0))$ is optimal for the interval $[T, (N+1)T]$. Thus from 37 we get, by the Principle of Optimality:

$$V^\circ_{N+1}(\mathbf{x}) = \min_{u_1(0)} \mu\{V^\circ_N(\boldsymbol{\phi}(T; \mathbf{x}, 0)) + \|\mathbf{x}\|^2_Q\} \quad (39)$$

Using this key recursive formula and the linear nature of $\boldsymbol{\phi}$, it is now easy to show by induction[3,4,6,] that $V^\circ_N(\mathbf{x})$ is given by a quadratic form $\|\mathbf{x}\|^2_{P(N)}$; moreover, if $\mathbf{Q}$ in 34 is positive definite, so is also $\mathbf{P}(N)$. The *existence* of V°_N is a trivial consequence of these calculations.

The essential remaining part of the proof is to show that the sequence $\mathbf{P}(0), \mathbf{P}(1) \ldots$ converges. To do this we topologize the space of all positive-definite matrices in the usual way[8], by defining the norm $\|\mathbf{P}\|$ of a positive-definite matrix by

$$\|\mathbf{P}\|^2 = \text{trace } \mathbf{P}^2 \quad (40)$$

(As is well known[8], $\|\mathbf{P}\|^2$ is the sum of the squared eigenvalues of $\mathbf{P}$.) It remains to prove the convergence of the sequence of numbers

$$\|\mathbf{P}(0)\|, \|\mathbf{P}(1)\| \ldots \quad (41)$$

(i) Suppose the plant is completely controllable. Then there is a control law, given by 31, such that motions of the closed loop system satisfy $\boldsymbol{\phi}(kT, \mathbf{x}, 0) = \mathbf{0}$ for all $\mathbf{x}$ and all $k \geqslant n$. For this particular control law, 34 is the sum of a finite number of terms and so V exists; V is obviously a positive-definite quadratic form which we denote by $V^1(\mathbf{x}) = \|\mathbf{x}\|^2_{P^1}$. Now $V^\circ_N(\mathbf{x}) \leqslant V^1(\mathbf{x})$ for all $\mathbf{x}$ and all N, since otherwise V°_N would not be optimal; this implies at once that $\|\mathbf{P}(N)\| \leqslant \|\mathbf{P}^1\| < \infty$, so that the sequence 41 is *bounded*.

Also, by 34 and the fact that $\mathbf{Q}$ is positive-definite, $\mathbf{P}(N+1) - \mathbf{P}(N)$ is positive definite for all N, so that $\|\mathbf{P}(N)\| \leqslant \|\mathbf{P}(N+1)\|$ which shows that the sequence 41 is *non-decreasing*.

A non-decreasing and bounded sequence converges; we write $\lim_{N \to \infty} \mathbf{P}(N) = \rho^0$. By optimality, the matrix $\mathbf{P}^0$ satisfying $\|\mathbf{P}^0\| = \rho^0$ is unique. Hence $\lim_{N \to \infty} \mathbf{P}(N) = \mathbf{P}^\circ$. A trivial argument shows that actually

$$\lim_{N \to \infty} V^\circ_N(\mathbf{x}) = \|\mathbf{x}\|^2_{P^0} = V^\circ(\mathbf{x});$$

in other words, the passage to the limit in N and the minimization can be interchanged. This proves (i).

(ii) If the plant is not completely controllable, we can write X as the direct sum $Y + \Gamma(n)$ where $\dim Y \neq 0$ since $\Gamma(n) \neq X$. To this direct-sum decomposition of X corresponds the partition of the matrix

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_{YY} & \mathbf{0} \\ \boldsymbol{\Phi}_{\Gamma Y} & \boldsymbol{\Phi}_{\Gamma\Gamma} \end{bmatrix} \quad (42)$$

where $\boldsymbol{\Phi}_{Y\Gamma} = \mathbf{0}$ because $\Gamma(n)$ is $\boldsymbol{\Phi}$-invariant. Now let

$$\nu = \inf_{\mathbf{y}} \|\boldsymbol{\Phi}_{YY}\mathbf{y}\|/\|\mathbf{y}\|;$$

$\nu > 0$ since otherwise $\boldsymbol{\Phi}_{YY}$ would be singular, implying by 42 that $\boldsymbol{\Phi}$ is singular, which contradicts the assumptions in Section 3.

Let $\mathbf{Y}$ be the projection of X on Y. Clearly, if $\mathbf{y}$ is in Y,

$$\nu^k\|\mathbf{y}\| \leqslant \|\boldsymbol{\Phi}_{YY}^k\mathbf{y}\| \leqslant \|\mathbf{Y}\boldsymbol{\phi}(kT; \mathbf{y}, 0)\| \leqslant \|\boldsymbol{\phi}(kT, \mathbf{y}, 0)\|;$$

and if $\mu\nu > 1$, the sum 34 will surely diverge for every $\mathbf{x}$ whose projection on Y is non-zero, which proves (ii).

(iii) From (i), we see at once that

$$V^\circ(\mathbf{x}) = \min_{u_1(t)} \mu\{V^\circ(\boldsymbol{\phi}(T; \mathbf{x}, 0)) + \|\mathbf{x}\|^2_Q\} \quad (43)$$

which shows that the optimal control law is stationary. A direct calculation based on 43, as in references 3, 4, and 6 shows that the control law is linear; this is, of course, a consequence of the fact that the error criterion $\|\mathbf{x}\|^2_Q$ is a quadratic form. This proves (iii).

(iv) We note that the difference in $V^\circ(\mathbf{x})$ along any motion of the optimized system is given by

$$V^\circ(\mathbf{x}) = T^{-1}\{V^\circ(\boldsymbol{\phi}(T, \mathbf{x}, 0)) - V^\circ(\mathbf{x})\}$$
$$= -T^{-1}\{(\mu - 1)V^\circ(\boldsymbol{\phi}(T; \mathbf{x}, 0)) + \mu\|\mathbf{x}\|^2_Q\},$$

which is surely negative-definite if $\mu \geqslant 1$ and $\mathbf{Q}$ is positive definite. Hence the usual stability theory of Liapunov is applicable[20], which completes the proof of 36.

We again have a feedback as in *Figure* 2; but of course now the linear functional defining the control law is not $\mathbf{e}^\star_1$ but is obtained after calculations based on 39[3,4,6].

This result can be easily generalized in numerous ways:

(*a*) In *terminal control* the upper limit of summation in 34 may be finite. In this case the control law is always non-stationary. Stability still follows in most cases[6].

(*b*) The performance criterion can be generalized

$$V(\mathbf{x}) = \sum_{k=1}^{\infty} \mu^k[\|\boldsymbol{\phi}(kT; \mathbf{x}, 0)\|^2_Q + \|\mathbf{u}(kT)\|^2_R] \quad (44)$$

to include the cost of the 'control energy'[3,5,6].

(*c*) Several control variables may be considered[4,6,14].

(*d*) Non-stationary systems can be treated analogously[6,14]. This includes also the case where $\mathbf{Q}(kT)$ is some more complicated function of time than $\mu^k\mathbf{Q}$. Non-stationarity greatly increases the computations required to find the optimal control law[6].

(*e*) Non-quadratic error criteria may be considered[14]. This leads to nothing new theoretically, but the optimal control law will be almost always non-linear and its explicit calculation may be completely impractical.

(*f*) The requirement that $\mathbf{Q}$ be positive-definite may be weakened.

We cannot forego an important remark concerning (*f*) above[20]. Suppose that $\|\mathbf{x}\|^2_Q = [\mathbf{q}^\star, \mathbf{x}]^2$. If $[\mathbf{q}^\star, \boldsymbol{\Phi}\mathbf{a}] \neq 0$, then the optimal control minimizing $V(\mathbf{x})$ defined by 34 is obviously

$$u_1(t) = -[\mathbf{q}^\star, \boldsymbol{\Phi}\mathbf{x}]/[\mathbf{q}^\star, \boldsymbol{\Phi}\mathbf{a}] \quad (45)$$

since then $[\mathbf{q}^\star, \boldsymbol{\phi}(kT; \mathbf{x}, 0)] = 0$ for all $k > 0$.

Formula 45 is the essence of the well-known paper of Bergen and Ragazzini[12]. As pointed out already by these authors, the *optimal* control law 45 does not necessarily result in a *stable* closed loop system. In fact, if $[\mathbf{q}^\star, \boldsymbol{\Phi}(T)\mathbf{a}(T)]$ is zero for some

486

$T = T_0$, then for values of T very close but not equal to T_0 the control law 45 will require very high loop gain so that we would intuitively expect to have instability. This can be readily shown by numerical examples. It is much less easy to find similar cases in continuous time control.

It is not yet known what *necessary* and sufficient conditions on **Q** assure the conclusions of Theorem 36.

Observability

Imitating the methods of Section 4, we now seek an intrinsic characterization of plants constrained at the output. In this section and the next, the plant will always be *linear, stationary, discrete time, free*, and *single output*.

Let $X^\star$ be the dual vector space of the state space X, i.e. the space of all linear functions on X. An element $\mathbf{z}^\star$ or $\mathbf{x}^\star$ of $X^\star$ is called a *costate*.

> *A costate* $\mathbf{z}^\star$ *of a plant is said to be 'observable' if its exact value* $[\mathbf{z}^\star, \mathbf{x}]$ *at any state* $\mathbf{x}$ *at time 0 can be determined from measurements of the output signal* $y_1(t) = [\mathbf{b}^\star, \boldsymbol{\phi}(t; \mathbf{x}, 0)]$ *over the finite interval* $0 \geqslant t \geqslant t_2$. *The time* t_2 *will depend in general on* $\mathbf{z}^\star$. *If every costate is observable, we say that the plant is 'completely observable'.* (46)

If a plant is completely observable, then its state variables with regard to any basis $\mathbf{a}_1 \ldots, \mathbf{a}_n$ can be determined in a finite length of time; the state variable x_1 with respect to this basis is $[\mathbf{a}^\star_i, \mathbf{x}]$, (where $\mathbf{a}^\star_i$ is the ith member of the dual basis), which is observable by assumption. However, we cannot label states *per se* as observable because there is no natural correspondence (i.e. independent of the particular basis used) between states and costates. If $\mathbf{a}^\star_j$ is not an observable costate, then the state variable x_j cannot be determined; in other words, this aspect of the plant's behaviour cannot be inferred from the measurements.

In analogy with Section 4, we have:

> *A plant is completely observable (i) if and (ii) only if the set of covectors* $(\boldsymbol{\Phi}^\star)^{-1} \mathbf{b}^\star \ldots, (\boldsymbol{\Phi}^\star)^{-n}\mathbf{b}^\star$ *is linearly independent.* (47)

For every positive integer i, let $\mathbf{f}^\star_i$ denote the covector $(\boldsymbol{\Phi}^\star)^{-i}\mathbf{b}^\star$. If a costate $\mathbf{z}^\star$ can be determined from the past q observations of the output, we can write:

$$[\mathbf{z}^\star, \mathbf{x}] = \sum_{i=1}^{q} \xi^\star_i [\mathbf{b}^\star, \boldsymbol{\phi}(-iT; \mathbf{x}, 0)]$$

Since the plant is free,

$$= \sum_{i=1}^{q} \xi^\star_i [\mathbf{b}^\star, \boldsymbol{\Phi}^{-i}\mathbf{x}]$$

$$= \sum_{i=1}^{q} \xi^\star_i [\mathbf{f}^\star_i, \mathbf{x}]$$

Since this must hold for every $\mathbf{x}$,

$$\mathbf{z}^\star = \sum_{i=1}^{q} \xi^\star_i \mathbf{f}^\star$$

From this, it follows by easy induction

> *The set* $\Gamma^\star(q)$ *of costates whose exact value at any state can be determined by operating on at most the last* q *measurements of the output is given by*

$$\Gamma^\star(q) = \left\{ \mathbf{z}^\star; \mathbf{z}^\star = \sum_{i=1}^{q} \xi^\star_i \mathbf{f}^\star_i \right\} \quad (48)$$

> *the* $\xi^\star_i$ *being arbitrary real numbers.*

Moreover, it follows easily:

> *If* $\mathbf{f}^\star_{k-1}$ *is linearly dependent on* $\mathbf{f}^\star_1 \ldots, \mathbf{f}^\star_k$ *then so is* $\mathbf{f}^\star_m$ *for all* m > k. *Hence if* $\Gamma^\star(k) = \Gamma^\star(k+1)$, *then* $\Gamma^\star(k) = \Gamma^\star(m)$ *for all* k $\geqslant$ m. (49)

The proof of 47 now follows exactly as the proof of 20.

The form and method of proof of Theorem 47 immediately suggests the problem: Does there exist some analogue of Theorem 1? The suspected analogy can be phrased in numerous ways. It turns out, however, that all such analogies, save one, lead to unnatural answers.

The correct analogy is the answer to the following question: If $\mathbf{z}^\star$ is in $\Gamma^\star(q)$, does there exist a 'scheme' which assures that the value of $\mathbf{z}^\star$ at any state $\mathbf{x}(t)$ becomes (and remains) known exactly after at most q measurements of the output of the plant?

In order to answer this question, it is necessary first to give an explicit prescription for computing the actual *value* of $[\mathbf{z}^\star, \mathbf{x}(t)]$ at any time t from the knowledge of the past measurements of the output of the plant. It does not seem possible to *deduce* such a scheme at the present state of the theory; however, the following *assumptions* provide a very natural framework for future developments.

Let $\bar{\mathbf{x}}(t)$ denote the estimate of the state of the plant at time t. This quantity will be determined by agreeing to use $[\mathbf{z}^\star, \bar{\mathbf{x}}(t)]$ as the estimate of the actual value of the costate $\mathbf{z}^\star$ at time t. We call $\tilde{\mathbf{x}}(t) = \mathbf{x}(t) - \bar{\mathbf{x}}(t)$ the *estimation error*; this is motivated by noting that, by linearity, the estimation error in the value of the costate is $[\mathbf{z}^\star, \mathbf{x}(t)] - [\mathbf{z}^\star, \bar{\mathbf{x}}(t)] = [\mathbf{z}^\star, \tilde{\mathbf{x}}(t)]$. With these conventions and considering the discrete time case only, the answer to the problem posed in the preceding paragraph amounts to specifying how $\bar{\mathbf{x}}((k+1)T)$ should be computed from past measurements $y_1(0) \ldots, y_1(kT)$ of the output of the plant.

Now we have an analogue of 30:

> *III. FUNDAMENTAL THEOREM FOR LINEAR OBSERVATION SYSTEMS*—Consider a discrete time, completely observable plant. The fastest observation scheme (in the sense that the exact value of every costate $\mathbf{z}^\star$ at any state $\mathbf{x}(t)$ is found after the smallest number ($\leqslant$ n) of measurements) is uniquely determined by the relations (50)

$$\bar{\mathbf{x}}((k+1)T) = \boldsymbol{\Phi}\{\bar{\mathbf{x}}(kT) - \mathbf{b}.[\mathbf{b}^\star, \tilde{\mathbf{x}}(kT)]\} \quad (51)$$

> where

$$\mathbf{b} = \boldsymbol{\Phi}^{-1}\mathbf{f}_1, \quad \mathbf{b}^\star = \boldsymbol{\Phi}^\star\mathbf{f}^\star_1$$

> $\mathbf{f}_1$ being the first element of the dual basis of $\mathbf{f}^\star_1 \ldots, \mathbf{f}^\star_n$.

The optimal observation scheme is implemented by means of a feedback system, as shown in *Figure* 3. The 'dynamics' in the open loop is the exact copy of the dynamics of the system whose state we wish to observe. The scalar error signal

$$[\mathbf{b}^\star, \tilde{\mathbf{x}}(kT)] = [\mathbf{b}^\star, \mathbf{x}(kT)] - [\mathbf{b}^\star, \bar{\mathbf{x}}(kT)] \quad (52)$$

can be measured directly. The scheme is physically realizable (for instance, by an analogue or digital computer) since there is one whole sampling period available for the computation of the next estimate.

Proof of 50. Let the measurement of the output of the plant start at $t = 0$. It is evidently sufficient to show that 51 is the only scheme which guarantees the following for all $k = 1, 2 \ldots$:

$$[\mathbf{z}^\star, \bar{\mathbf{x}}(kT)] = [\mathbf{z}^\star, \mathbf{x}(kT)] \quad \text{for all } \mathbf{z}^\star \text{ in } \Gamma^\star(k)$$
$$[\mathbf{f}^\star_1, \bar{\mathbf{x}}(kT)] = 0 \quad \text{for all } i > k \quad (53)$$
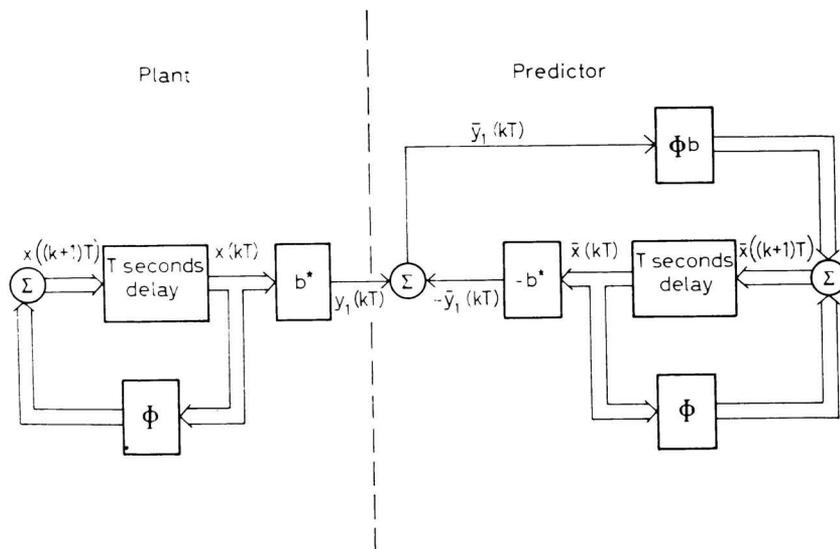
487

Figure 3

We proceed by induction. The first estimate is

$$\bar{x}(T) = \Phi b[b^{\star}, x(0)] = f_1[f^{\star}_1, x(T)]$$

and 53 follows trivially.

Now assume that for $k = q$

$$[z^{\star}, \bar{x}(qT)] = [z^{\star}, x(qT)]$$

for every costate $z^{\star}$ in $\Gamma^{\star}(q)$.

$$[w^{\star}, \Phi\bar{x}(qT)] = [w^{\star}, x((q+1)T)]$$

where

$$w^{\star} = (\Phi^{\star})^{-1}z^{\star}$$

Now if $z^{\star}$ is in $\Gamma^{\star}(q)$, then $w^{\star}$ is a costate in $\Gamma^{\star}(q+1)$ such that its first coordinate vanishes: $\xi^{\star}_1 = [w^{\star}, f_1] = 0$. Therefore $w^{\star}$ is orthogonal to the second term on the right-hand side of 51 and we have

$$[w^{\star}, \bar{x}((q+1)T)] = [w^{\star}, x((q+1)T)] \qquad (54)$$

By the same reasoning, if $[f^{\star}_1, \bar{x}(qT)] = 0$ for all $i > q$, then $[f^{\star}_1, \bar{x}((q+1)T)] = [f^{\star}_1, \Phi\bar{x}(qT)] = [f^{\star}_{i+1}, x(qT)] = 0$ for all $i > q+1$.

It remains to prove that relation 54 holds for any covector in $\Gamma^{\star}(1)$.

By direct calculation,

$$[f^{\star}_1, \bar{x}((q+1)T)] = [f^{\star}_1, \Phi\bar{x}(qT) + f_1[f^{\star}_1, x((q+1)T) - \Phi\bar{x}(qT)]]$$

By linearity,

$$= [f^{\star}_1, \Phi\bar{x}(qT)] + [f^{\star}_1, x((q+1)T) - \Phi\bar{x}(qT)]$$

$$= [f^{\star}_1, \bar{x}((q+1)T)]$$

which proves the induction. The fact that the proof goes through only when the estimation scheme is given by 51 is evident from the calculations. Q.E.D.

An immediate consequence of this result is the following, perhaps even more natural, analogue of 30:

FASTEST CONTROL OF PLANT WITH OUT-PUT CONSTRAINT—Consider a completely observable plant which is constrained at its input.

Given any unknown initial state x_0, every state variable $\phi_1(kT; x_0, 0)$ is reduced to zero in the minimum number ($\leqslant$ n) of sampling periods if and only if the closed loop system is governed by (55)

$$x((k+1)T) = \Phi\{x(kT) - b.[b^{\star}, x(kT)]\} \qquad (56)$$

Replacing e_1 by b and $e^{\star}_1$ by $b^{\star}$, this is just the system in *Figure* 2. However, the initial constraints on the plant which led to this system were, of course, different.

To prove 55, we observe by subtracting 51 from 15 that the error $\tilde{x}(kT)$ in the optimal observation scheme is also governed by a free dynamic system of type 56:

$$\tilde{x}((k+1)T) = \Phi\{\tilde{x}(kT) - b.[b^{\star}, \tilde{x}(kT)]\} \qquad (57)$$

Identifying x with $\tilde{x}$, the desired conclusion follows immediately.

The Principle of Duality

The analogies already apparent between controllability and observability can be expressed cogently by what we shall call the *Principle of Duality*.

Consider a discrete time, n-input, single output plant. The most general form of the transition equations of *any* feedback system with linear (but not necessarily stationary) control law built around this plant is

$$x((k+1)T) = \Phi\{x(kT) + a((k+\tfrac{1}{2})T)[b^{\star}, x(kT)]\} \qquad (58)$$

where $a((k+\tfrac{1}{2})T)$ is an arbitrary vector; the notation is intended to indicate that its value is fixed for the sampling interval $kT \leqslant t \leqslant (k+1)T$.

Let $\phi(kT; x, 0)$ denote the motions of 58. Given an arbitrary state x and an arbitrary costate $z^{\star}$, we define by induction for every $k = 0, 1 \ldots$ the unique function $\phi^{\star}$ satisfying

$$[z^{\star}, \phi(kT; x, 0)] = [\phi^{\star}(0; z^{\star}, kT), x] \qquad (59)$$

It is easily seen that $\phi^{\star}$ represents the motions of the following dynamic system defined in $X^{\star}$:

$$x^{\star}((k-1)T) = \Phi^{\star}(x^{\star}(kT) + (\Phi^{\star})^{-1}b^{\star}[x^{\star}(kT), \Phi a((k-\tfrac{1}{2})T)]) \qquad (60)$$

488

Since $\mathbf{b}^\star$ is fixed and $\mathbf{a}((k+\frac{1}{2})T)$ is arbitrary, we see that 60 defines in $X^\star$ a single input, n-output plant which we call the *dual plant*. We have arrived at the

PRINCIPLE OF DUALITY—Considering the class of feed-back systems with linear control law, the dual plant defined in $X^\star$ is obtained by the following steps: (i) Replace $\boldsymbol{\phi}$ by its dual $\boldsymbol{\phi}^\star$; (ii) interchange input and output constraints, (iii) reverse the direction of time. (61)

See *Figure 4*.

A number of remarks are in order to place this idea in proper perspective:

(*A*) The dual or *adjoint* of a free linear dynamic system is well known in the theory of linear differential equations[10]. The adjoint of the free system 10 is

$$d\mathbf{x}^\star/dt = -\mathbf{F}^\star(t)\mathbf{x}^\star$$

(*B*) In the present context, the duality principle was discovered by the author[7] in the course of an investigation of the Wiener–Kolmogorov filtering and prediction problem. The
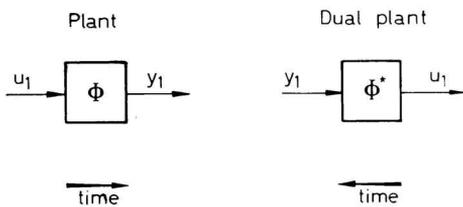

Figure 4

important point is that by the duality principle a plant constrained at the output can be converted into a plant constrained at the input so that the theories of optimal regulation and Wiener filtering can be compared directly. See below.

Two immediate consequences of the Principle of Duality are the following:

A plant is completely controllable if, and only if, its dual is completely observable, and conversely. (62)

To prove 62, it suffices to observe that if $\mathbf{a}_1 \ldots, \mathbf{a}_n$ is any basis in X and $u_1{}^i(0) \ldots, u_1{}^i(kT) \ldots$ is a sequence of control signals which transfers $\mathbf{a}_i$ to the origin in finite time, then the control signal defined by

$$u_1(kT) = \sum_{i=1}^{n} u_1{}^i(kT) x_i(kT)$$

will transfer every state to the origin in finite time. Hence in questions of controllability and observability it suffices to consider linear feedback systems.

Similarly,

the solution of the minimal time observation problem is identical with the solution of the minimal time control problem for the dual plant. (63)

We examine now the implications of 'dualization' of results of Section 5. Concerning the dual plant, consider the minimization of

$$V^\star(\mathbf{z}^\star) = \sum_{k=1}^{\infty} \phi^\star(kT;\mathbf{z}^\star,0)\,{}^2_{\mathbf{Q}^\star}$$ (64)

where $\mathbf{Q}^\star$ is positive-definite.

This is, of course, an abstract problem in $X^\star$. To see what this actually means in X, note first that any positive semi-definite quadratic form in $X^\star$ may be represented as

$$\|\mathbf{z}^\star\|^2_{\mathbf{Q}^\star} = \sum_{i=1}^{m} [\mathbf{z}^\star,\mathbf{q}_i]^2 \quad (m \leq n)$$ (65)

where the $\mathbf{q}_i$ are a set of orthogonal vectors

$$[\mathbf{q}^\star_i,\mathbf{q}_j] = \delta_{ij} \quad i,j = 1\ldots,m$$

and $m = n$ if and only if $\mathbf{Q}^\star$ is positive-definite. By 65,

$$V^\star(\mathbf{z}^\star) = \sum_{k=1}^{\infty} \sum_{i=1}^{m} [\phi^\star(kT;\mathbf{z}^\star,0),\mathbf{q}_i]^2$$

By 59 and time reversal,

$$= \sum_{k=1}^{\infty} \sum_{i=1}^{m} [\mathbf{z}^\star,\phi(0;\mathbf{q}_i,-kT)]^2$$

At this point, we make a crucial observation concerning the interpretation of the terms of the last equation.

Let $\mathbf{v}$ be a vector-valued random variable with zero mean and gaussian distribution. As is well known, such a random variable is uniquely specified by the knowledge of its covariance matrix $\mathbf{W} = cov\ \mathbf{v}$, which in turn is characterized by the following identity:

$$E[\mathbf{z}^\star,\mathbf{v}]^2 = \|\mathbf{z}^\star\|^2_{cov\ \mathbf{v}}$$ (66)

Since a covariance matrix is always positive semi-definite, we can use the representation 65. Thus

$$E[\mathbf{z}^\star,\mathbf{v}]^2 = \sum_{i=1}^{m} [\mathbf{z}^\star,\mathbf{w}_i]^2$$

(the $\mathbf{w}_i$ being eigenvectors of $cov\ \mathbf{v}$). The content of formula 66 can also be expressed as follows:

If $\mathbf{v}$ is a vector-valued gaussian random variable in X with zero mean, and covariance matrix $\mathbf{W}$, it can be expressed as the sum of independent random variables $\mathbf{v}_i$ (67)

$$\mathbf{v} = \mathbf{v}_1 + \ldots + \mathbf{v}_m \quad (m \leq n)$$

where $\mathbf{v}_i = \mathbf{w}_i.\alpha_i$, the α_i being scalar-valued, independent gaussian random variables with zero mean and unit variance.

Since the $\mathbf{v}_i$ are independent, their covariances add so that it suffices to prove that $cov \sum_i \mathbf{v}_i = \sum_i cov\ \mathbf{v}_i = cov\ \mathbf{v}$. But

$$E[\mathbf{x}^\star,\mathbf{v}] = \|\mathbf{x}^\star\|^2_{cov\ \mathbf{v}} = \sum_{i=1}^{m} [\mathbf{x}^\star,\mathbf{w}_i]^2$$

$$= \sum_{i=1}^{m} E[\mathbf{x}^\star,\mathbf{v}_i]^2 = \sum_{i=1}^{m} \|\mathbf{x}^\star\|^2_{cov\ \mathbf{v}_i} = \|\mathbf{x}^\star\|^2_{\sum_i cov\ \mathbf{v}_i}$$

and since this relation must hold for any costate $\mathbf{x}^\star$, the assertion 67 follows.

Finally, let $\mathbf{v}(kT)$ be a random process such that its values occurring at different instants of time are independent and at each instant $\mathbf{v}(kT)$ has the same covariance matrix $\mathbf{W}$.

Identifying the $\mathbf{q}_i$ with the $\mathbf{w}_i$ and using the assumptions of independence and zero mean

$$V^\star(\mathbf{z}^\star) = E\left[\mathbf{z}^\star,\sum_{k=-1}^{\infty} \phi(0;\mathbf{v}(kT),kT)\right]^2$$ (68)

Now consider, independently of the immediately preceding discussion, a *linear*, discrete time, single output dynamic system

489

subject to random perturbations

$$\mathbf{x}((k+1)T) = \mathbf{\Phi}\mathbf{x}(kT) + \mathbf{v}(kT)$$
$$y_1(kT) = [\mathbf{b}^\star, \mathbf{x}(kT)] \tag{69}$$

This system may be regarded as a model of a stochastic process, with $y_1(kT)$ being signal plus noise[7]. We wish to obtain an optimal estimate of every costate $\mathbf{z}^\star$ in such a way that $E[\mathbf{z}^\star, \tilde{\mathbf{x}}(kT)]^2$ is a minimum.

If we *assume* that the optimal estimation process is a linear dynamic system (this assumption in the gaussian case is proved in reference 7), then the estimate $\tilde{\mathbf{x}}(kT)$ will be computed, as in Section 6, by means of an 'artificial' plant which is not constrained at either the input or the output. Subtracting the transition equations of $\tilde{\mathbf{x}}$ from that of $\mathbf{x}$, we get for the estimation error the transition equation

$$\tilde{\mathbf{x}}((k+1)T) = \mathbf{\Phi}\{\tilde{\mathbf{x}}(kT) + \mathbf{a}((k+\tfrac{1}{2})T)[\mathbf{b}^\star, \tilde{\mathbf{x}}(kT)]\} + \mathbf{v}(kT) \tag{70}$$

Since the motions $\boldsymbol{\phi}$ on the right-hand side of 68 are those of a plant constrained at the output, $\boldsymbol{\phi}$ can be identified with motions of the *free* system 70. By linearity, it follows further that the summation on the right-hand side of 68 gives the present state of 70 due to the superposition of the effects of all the random disturbances in the past. In short,

$$V^\star(\mathbf{z}^\star) = E[\mathbf{z}^\star, \tilde{\mathbf{x}}(0)]^2$$
$$= \underset{\tilde{\mathbf{x}}(0)}{\text{Min}}\ E\{[\mathbf{z}^\star, \mathbf{x}(0) - \tilde{\mathbf{x}}(0)]^2 | y_1(-T), y_1(-2T) \ldots\} \tag{71}$$

i.e. $V^\star$ is the minimized conditional expectation of error in predicting the state by $\tilde{\mathbf{x}}(0)$, where $\tilde{\mathbf{x}}(0)$ is based on all past measurements of the sum y_1 of 'signal and noise'. The rigorous proof of the second equality is given in reference 7. Summing up,

> *the optimal regulator problem in* $\mathbf{X}^\star$ *is the Wiener filtering problem in* $\mathbf{X}$. (72)

In particular, the methods of Section 5 for determining the control law for the optimal regulator can be applied immediately by the Principle of Duality to obtain the defining equations of the optimal Wiener filter. The detailed formulae for doing this are worked out in reference 7.

This important discovery gives rise to many new problems as well as results, of which we cite only one.

> *The optimal prediction system* 70 *which solves the Wiener filtering problem is asymptotically stable if and only if the model* 69 *representing the COMBINED random processes of signal and noise is completely observable.* (73)

We must bear in mind that the model 69 is mathematical fiction: it is merely a representation of (presumably empirically obtained) auto- and cross-correlation functions of signal and noise. This representation is standard in the engineering theory of filtering and prediction. The hypothesis in theorem 73 can also be expressed as follows: *If the fictitious random excitation* $\mathbf{v}(kT)$ *is 'turned off', then the exact state of the system can be determined from a finite number of measurements of signal plus noise.*

To prove 73, we merely observe that if the (free) dual plant is stable in $X^\star$, the (free) plant itself is stable in X.

Generalizations and Open Problems

By putting together the results of Sections 4–6, we have the obvious conclusions:

IV. MAIN EXISTENCE THEOREM—Consider a discrete time single input/single output plant. The optimal regulator problem has a solution with properties as stated in 36 if and only if the plant is completely controllable and completely observable. (74)

This theorem may be regarded as the central result so far in the pure theory of control. It follows trivially from 36 and the definition of observability. For if we have observability, all state variables can be determined exactly in a finite number of steps, after which we have the situation covered by 36. Conversely, if some state variable is not observable, the performance index $V(\mathbf{x})$ will surely diverge for some μ sufficiently large if the unobservable coördinate of the initial state is not zero.

It is easy to relax the conditions of (i) 'discrete time', (ii) 'single input', (iii) 'single output' in Theorem 74.

The first generalization requires only assuming that the performance criterion is given by 44 with $\mathbf{R}$ positive-definite; for otherwise $V(\mathbf{x})$ can be made arbitrarily small for any fixed $\mathbf{x}$ by taking T small while at the same time $u_1(t)$ will be arbitrarily large.

The abstract definitions of controllability and observability need not be changed in the general case of multi-input, multi-output systems; of course the explicit conditions 20 and 47 have to be appropriately modified. The main new factor arising is that now the properties of complete controllability or observability are by no means obvious by inspection of the dynamic equations, particularly when these are obtained by linearization of non-linear equations. It is curious to note that the importance of these matters has remained undiscovered until now in the large, but superficial, literature of multiloop systems.

To what extent can the duality principle be divorced from the case of a linear control law? In the solution of the optimal regulator problem outlined in Section 5, no *a priori* assumption about linear control law is made, but this *follows* from using a quadratic performance index. Similarly in the direct solution of the Wiener filtering problem, no assumption is made that the control law is linear (or even that the optimal predictor is a dynamic system), but this follows from gaussianness as in reference 7. The Principle of Duality then connects these results; is it possible to do so *a priori*?

It can be shown that the properties of the optimal Wiener filters discussed in Section 7 and in reference 7 can be related to the information rate (in the sense of Shannon) conveyed by the measurements $y_1(kT)$ about the unknown state $\mathbf{x}(kT)$. By the duality principle, one would expect that the concept of information, too, has a dual. But what is it?

Going a step further, one should study the information requirements in controlling a plant subject to random disturbances. In particular, the 'quality of control' will depend on both the input and the output constraints. A measure of the latter is the information rate; a measure of the former may exist dually, as conjectured in the preceding paragraph. Can one then make precise quantitative statements about the performance of a control system from such quantitative measures of the input/output constraints? In other words, is it possible to regard controllability and observability as numeric quantities rather than just abstract properties?

The investigation of these and similar questions should provide powerful incentive and encouragement in the further development of the pure theory of control.

Conclusions

Guided by recent developments in the theory of optimization of control systems, the framework of a *pure theory of control*

has been outlined providing a rigorous foundation for some past and many future investigations. An important achievement of this new point of view is the Principle of Duality relating the control and filtering problems. Further, at the present status of development of the theory, it has been possible to solve abstract problems side-by-side with effective computational methods superior to those currently in vogue in the so-called applied theory of control.

References

1 SHANNON, C. E. A mathematical theory of communication. *Bell System tech. J.* 27 (1948) 379, 623

2 KALMAN, R. E. and BERTRAM, J. E. General synthesis procedure for computer control and single- and multi-loop linear systems. *Trans. Amer. Inst. elect. Engrs* 77 II (1958) 602

3 KALMAN, R. E. and KOEPCKE, R. W. Optimal synthesis of linear sampling control systems using generalized performance indexes. *Trans. Amer. Soc. mech. Engrs* 80 (1958) 1820

4 KALMAN, R. E. and KOEPCKE, R. W. The role of digital computers in the dynamic optimization of chemical reactions. *Proc. Western Joint Comp. Conf.*, 1959, pp. 107–116

5 BERTRAM, J. E. and SARACHIK, P. E. On optimal computer control. First International Congr. on Automatic Control, Moscow, 1960

6 KALMAN, R. E. *Dynamic Optimization of Linear Control Systems*, to be published

7 KALMAN, R. E. On a new approach to filtering and prediction problems. *J. Basic Engng (Trans. A.S.M.E.)*, 82D (1960) 35

8 HALMOS, P. R. *Finite Dimensional Vector Spaces*, 2nd ed. 1958. New York; Van Nostrand

9 CHENG, D. K. *Analysis of Linear Systems*. 1959. Reading, Mass.: Addison-Wesley

10 CODDINGTON, E. A. and LEVINSON, N. *Theory of Ordinary Differential Equations*. 1955. New York; McGraw-Hill

11 KALMAN, R. E. and BERTRAM, J. E. A unified approach to the theory of sampling systems. *J. Franklin Inst.* 267 (1959) 405

12 BERGEN, A. R. and RAGAZZINI, J. R. Sampled-data processing techniques for feedback control systems. *Trans. Amer. Inst. elect. Engrs* 73 II (1954) 236

13 KALMAN, R. E. Optimal nonlinear control of saturating systems by intermittent action. *Wescon Conv. Rec. Inst. Radio Engrs, N.Y.* 1, Pt. 4 (1957) 130

14 KALMAN, R. E. Contributions to the theory of optimal control. *Proc. Symp. on Ordinary Differential Equations*, Mexico City, 1959. To appear in *Bell. Soc. Mat. Mexicana*, 1961

15 PONTRYAGIN, L. S. Optimal control processes (in Russian). *Progr. math. Sci.* 14 (1959) 3

16 LaSALLE, J. P. The time-optimal control problem. *Contributions to nonlinear differential equations. Ann. Math. Stud.* (1960)

17 KALMAN, R. E. Discussion of reference 12

18 JURY, E. L. *Sampled-data Control Systems*. 1958. New York; Wiley

19 BELLMAN, R. E. *Dynamic Programming*. 1957. Princeton Univ. Press

20 KALMAN, R. E. and BERTRAM, J. E. Control system analysis and design via the 'second method' of Lyapunov. *J. Basic Engng (Trans. A.S.M.E.)*, 82D (1960) 371, 394

21 OLDENBOURG, R. and SARTORIUS, H. *Dynamics of automatic control systems*, 2nd ed. 1951. Munich; Oldenbourg

22 TSYPKIN, YA. Z. Theory of intermittent control systems (in Russian). *Automat. Telemech.*, Moscow, 11 (1950) 300

Summary

This note initiates the study of what might be called the 'pure' theory of control. Restricting attention to linear, stationary, single input/single output plants, the new concepts of controllability and observability are introduced and it follows that solutions of the usual problems of control theory exist if and only if the plant is completely controllable and completely observable.

The connection between observability and controllability is formalized by the Duality Principle, which shows that the Wiener filtering and prediction problem is a special case of the author's theory of optimization of deterministic control systems, as has been shown earlier by less direct methods.

Sommaire

Cette note introduit l'étude de ce qu'on pourrait appeler la théorie 'pure' de la commande automatique. En restreignant l'attention aux systèmes linéaires, stationnaires, à entrée et sortie uniques, on présente les nouveaux concepts de contrôlabilité et d'observabilité, et il en découle qu'il n'existe de solutions aux problèmes habituels de la théorie de la commande automatique, que si le système est complètement contrôlable et complètement observable.

La relation entre observabilité et contrôlabilité est formalisée par le Principe de Dualité, qui montre que le problème de filtrage et de prédiction de Wiener est un cas particulier de la théorie de l'auteur sur l'optimisation des systèmes de commande déterministes, comme il a été montré antérieurement par des méthodes moins directes.

Zusammenfassung

Der Beitrag leitet das Stadium einer sozusagen 'reinen' Theorie der Regelung ein. Unter Beschränkung auf lineare, stationäre Systeme mit nur einer Ausgangs- und einer Eingangsgröße werden die neuen Begriffe der Regelbarkeit und der Erfaßbarkeit eingeführt. Es folgt daraus, daß Lösungen für die üblichen Aufgaben der Regeltheorie dann und nur dann existieren, wenn die Regelstrecke vollständig regelbar und erfaßbar ist. Die Verknüpfung zwischen Erfaßbarkeit und Regelbarkeit ist durch das Dualitätsprinzip gegeben, aus dem hervorgeht, daß das Filter- und Vorhersageproblem nach Wiener ein Spezialfall der Theorie des Verfassers zur Optimierung deterministischer Regelungsanordnungen ist, wie sich auch schon früher nach weniger unmittelbaren Verfahren zeigte.

DISCUSSION

J. BRAY (*U.K.*)

I consider the paper by Mr. Kalman is useful and I should like to elucidate one aspect in which, in my view, his analogy touches the theory of information. We find that for the type of systems described in the paper, when the order of the system or the dimension of the vector of state is between the limits 40–100, the difficulty consists in working out experiments on the plant for the purpose of determining

491

the coefficients in equation 16. Selection of the modulus, obtaining of the coefficients, prediction, optimization—all these are problems which are internal computational problems, on which experiments can be carried out over a long period without disturbing the operation of the object. However, if the experiments are carried out on an expensive plant it is essential that the time during which the experiment is carried out should be utilized as fully as possible. It appears that the classical statistics of experiments does not provide sufficient indication on the experimental work in recurrent relations. A reliable, although not the most effective, method which we used was to make a random selection based on an experimental calculation in the steady-state regime. I believe, however, that this problem can be solved more satisfactorily.

Most of the elements which we used in describing systems of higher order are, in reality, non-linear to some extent, and I believe that Mr. Kalman should extend his ideas to give at least a topological description of such systems although then we would like to know what we are talking about.

L. I. ROZONOER (*U.S.S.R.*)

Mr. Kalman's paper is interesting. The conception of controllability and observability is very natural and useful, particularly in conjunction with the 'duality principle', the essence of which is reflected in the theorem to the effect that if a plant is controllable, a related plant is observable and *vice versa*.

However, I believe that it is so far impossible to speak, in conjunction with the work of Mr. Kalman, of a general theory or even of its basic principles. First, even according to the purely formal characteristics, the paper does not contain any ideas on formulating problems and on results in non-linear cases, without which there is obviously no general theory. However, I believe that the conception of 'duality', which is closely linked with the canonic conjugation in the sense of analytical mechanics, can be extended also to arbitrary systems, including non-linear ones. Secondly, and this is the main point, it will only be possible to talk of a 'pure' control theory when at least the basic outlines of the mathematical apparatus become clear, which will enable describing by means of precise and general terms the concepts which at present are intuitive and unclear. This relates particularly to the concepts of controllability and observability which, are precise only when applied to systems of a specific type, and which lose their clarity completely (although they remain intuitively conceivable) in the case of using control systems more generally.

R. E. KALMAN, *in reply*. The comments of Dr. Bray on the tasks which have to be studied deserve serious attention. An effective solution of these problems will represent progress in the field of automatic control.

It is a good idea to remind ourselves periodically that the accepted formulation of the control theory does not satisfy practical requirements. Obviously, Dr. Bray realizes this more clearly than many other people. However, I believe that even he agrees that verification of the bases of the control theory has some justification. The concepts of controllability and observability were introduced and emphasized particularly in view of the fact that our present knowledge of the theory of automatic control proved unsuitable in view of the requirements to be met for the problems of present-day investigation. It is possible that Dr. Bray did not consider this fact and it is also possible that new problems to which he referred can be solved by ordinary means. In fact, if it is possible to develop investigations of the control theory without the conception of controllability and observability, my paper will become, to some extent, an erroneous sophistication. However, present-day investigations lead to a different conclusion, namely: before dealing successfully with solving new problems it is necessary to have a wider justification for the accepted control theory.

The remarks of Dr. Rozonoer, that the paper is of importance only for linear systems, are, in the final analysis, correct. However, I should like to add that the paper does not pretend to represent the development of a *general* theory of control systems, as is underlined in the headings of Paragraph 4. In the very general approach in which the paper has been written, basically nothing is known on non-linear systems. However, as was mentioned in the paper, the conditions of observability and controllability have already played an important role in the investigations of L. S. Pontryagin and others relating to non-linear systems.