IFAC

INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL

WARSZAWA 1969

Identification

State Identification and Process Observability

Fourth Congress of the International Federation of Automatic Control
Warszawa 16–21 June 1969

TECHNICAL SESSION 12

NOT

Organized by
Naczelna Organizacja Techniczna w Polsce

INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL

Identification

State Identification and Process Observability

TECHNICAL SESSION No 12

FOURTH CONGRESS OF THE INTERNATIONAL
FEDERATION OF AUTOMATIC CONTROL
WARSZAWA 16 – 21 JUNE 1969

Organized by
Naczelna Organizacja Techniczna w Polsce

Contents

Paper No			Page
12.1	SU	- L.A.Rastrigin, V.S.Trahtenberg - Multidimensional Linear Extrapolation in the Optimal Design and Control	3
12.2	USA	- R.E.Bailey, L.J.Habegger - Minimum Variance Estimation of Parameters and States in Nuclear Power Systems	19
12.3	USA	- G.A.Phillipson, S.K.Mitter - State Identification of a Class of Linear Distributed Systems	34
12.4	J	- Akira Sano, Mitsuru Terao - Measurement Optimization in Batch Process Optimal Control	57
12.5	J	- Takashi Sekiguchi - Observability of Linear Dynamic Measuring System and Some Applications	75
12.6	IND	- Shri K.K. Bandyopadhyay, S. Dasgupta - Identification of the Parameters of a Tidal Channel from Simulation and other Approaches	95

Wydawnictwa Czasopism Technicznych NOT - Polska

Zakład Poligraficzny WCT NOT. Zam. 61/69.

МНОГОМЕРНАЯ ЭКСТРАПОЛЯЦИЯ В ЗАДАЧАХ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ И ПРОЕКТИРОВАНИЯ.

Л.А.РАСТРИГИН, В.С.ТРАХТЕНБЕРГ

Институт электроники и вычислительной техники Латв.ССР
Рига СССР

Одной из основных проблем современной автоматики является создание самонастраивающихся систем автоматического управления, адаптирующихся к изменяющимся условиям так, чтобы поддерживать экстремум некоторого заданного функционала (критерия качества). Задачу приспособления системы к изменяющимся условиям можно сформулировать следующим образом.

Пусть функция

$$J = J(X, A) \tag{1}$$

выражает зависимость критерия качества системы от ее управляемых параметров, характеризуемых вектором $X = (x_1, \dots, x_n)$, и внешних условий или ситуации $A = (a_1, \dots, a_m)$, в которой функционирует система. (Предполагается, что ситуация полностью описывается m числами $a_1, \dots, a_m$). В общем случае зависимость (1) может быть не формализована.

Пусть S - множество всех возможных ситуаций A, а U - множество допустимых управлений X. Тогда задача оптимального приспособления системы будет решена, если найден алгоритм или оператор преобразования $S \to U$, который каждому вектору A множества S возможных состояний системы ставит в соответствие вектор $X^* \in U$ оптимальных параметров системы, экстремизирующих ее критерий качества (1), например, $J[X^*(A), A] = \min$. Таким образом определение зависимости $X^* = X^*(A)$ решает задачу адаптации (предполагается, что задача идентификации ситуации A уже решена).

Наиболее универсальным методом приспособления системы является поиск [1, 2], который позволяет экстремизировать заданный критерий (1) путем сбора информации о поведении объекта с целью отыскания его оптимальной реакции X^* для новых условий A.

Однако, если зависимость (1) не формализована, поиск необходимо проводить в реальном масштабе времени непосредственно на объекте, что значительно снижает эффективность и оперативность приспособления. Очень часто поиск на объекте запрещен или существенно ограничен из технологических соображений. В этих случаях задачу приспособления системы целесообразно решать с применением беспоисковых методов. В качестве одного из таких методов адаптации системы к новой ситуации может быть использовано обучение на результатах предыдущего опыта.

Пусть для ряда различных ситуаций A_i $(i = 1, \ldots, k)$ каким-то определенным образом получены соответствующие им оптимальные параметры:

$$\left.\begin{array}{l} A_1 \rightarrow X_1^* \\ \cdots\cdots \\ A_k \rightarrow X_k^* \end{array}\right\} \qquad (2)$$

Эту совокупность соответствий будем называть обучающей последовательностью. Очевидно, что (2) можно рассматривать как результаты наблюдения неизвестной функциональной зависимости

$$X^* = X^*(A) \tag{3}$$

Тогда задача определения оптимальных параметров системы в некоторой новой ситуации A_{k+1} сводится прежде всего к восстановлению указанной зависимости (3) по указанным k наблюдениям, а затем к последующему определению значения X^*_{k+1} в точке A_{k+1} : $X^*_{k+1} = X^*(A_{k+1})$. В такой постановке задачу можно решить, например, с помощью метода потенциальных функций [3], методов стохастической аппроксимации [4], или родственной указанным методам итерационной процедуры [5].

Однако эти методы обеспечивают удовлетворительное восстановление неизвестной зависимости (3) только при достаточно большом числе наблюдений. Кроме того, эффективность здесь существенно зависит от выбора системы функций, линейная комбинация которых аппроксимирует неизвестную функциональную зависимость. Это означает, что для функционирования указанных процедур необходимо иметь значительную априорную информацию о поведении функции (3). Однако в практических ситуациях, особенно в задачах оптимального проектирования, сведения о поведении указанной функции крайне скудны и не могут лечь в основу выбора системы функций, по которым раскладывается (3). Более того, число наблюдений k практически всегда очень мало, что так же затрудняет указанный выбор.

Ниже предлагается и анализируется метод, условно названный методом многомерной линейной экстраполяции, с помощью которого оказывается возможным находить достаточно обоснованные оценки для оптимальных параметров системы в новой ситуации на базе весьма ограниченного предыдущего опыта. В частности, число наблюдений k может быть меньше размерности пространства ситуаций:

$$k < m \tag{4}$$

Метод многомерной линейной экстраполяции. Оценки для оптимальных параметров системы X^* в новой ситуации согласно предлагаемому методу находятся следующим образом [6]. Через векторы, входящие в состав обучающей последовательности соответственно в пространстве ситуаций S и пространстве оптимальных решений U проводятся гиперплоскости S' и U'. Очевидно, любой элемент $A' \in S'$ в предположении о линейности пространства ситуаций можно представить в виде линейной комбинации:

$$A' = A_1 + \sum_{i=1}^{k-1} \lambda_i (A_{i+1} - A_1) \tag{5}$$

где λ_i являются координатами A' в базисе, построенном на элементах обучающей последовательности. Если $A \in S'$ то в качестве оценки вектора оптимальных параметров системы в этой ситуации предлагается принять вектор X, определяемый выражением:

$$X = X_1^* + \sum_{i=1}^{k-1} \lambda_i (X_{i+1}^* - X_1^*) \tag{6}$$

т.е. между гиперплоскостями S' и U' образуется линейная связь. Таким образом, искомая функциональная зависимость (3) линеаризуется на подпространствах S' и U', что позволяет для любой ситуации $A' \in S'$ находить оценку вектора оптимальных параметров системы путем линейной экстраполяции.

Если новая ситуация A не принадлежит гиперплоскости S', то ее естественно отождествлять с ближайшей к ней в определенном смысле [redacted] ситуацией, лежащей в S'

Для этого в пространстве ситуаций вводится метрика, позволяющая каждой паре ситуаций поставить в соответствие число ρ, определяющее меру их близости, например, расстояние между ними:

$$\rho(A, A') = |A - A'|^2 \qquad (7)$$

В этом случае ситуация $A \in S$ будет отождествляться со своей ортогональной проекцией A' на гиперплоскость S'. Отождествление $A \equiv A'$ по сути дела означает, что свойства подпространства S' распространяются, или экстраполируются на окружающую и близкую к ней область пространства S

Алгоритм экстраполяции можно представить в виде

$$A \equiv A' \rightarrow X \equiv X^* \qquad (8)$$

Из этого алгоритма непосредственно следуют условия его применимости. Так, указанный метод целесообразно осуществлять лишь в том случае, если выполняются следующие условия:

1. Область S изменения параметров ситуации достаточно мала, так, что для поддержания качества системы в

желаемых пределах достаточно корректировать оптимальные параметры системы (без изменения ее структуры).

2. Имеется опыт оптимального приспособления системы к некоторым ситуациям из области S , на основании которого строится обучающая последовательность.

3. Неизвестную функциональную зависимость (3) возможно с достаточной точностью линеаризовать в области S .

Последнее условие при $k \leqslant m$ вовсе снимается, т.к. при отсутствии априорных сведений о функции (3)в этом случае число показов k не достаточно для однозначного определения линейного преобразования $S \rightarrow U$. Если же длина обучающей последовательности больше размерности пространства ситуаций, то последнее условие применимости метода ослабляется: следует предполагать,что неизвестную функциональную зависимость (3) можно с достаточной точностью линеаризовать в окрестности $q < k$ точек ($q \leqslant m$) из области S . В этом случае оценка для X^* определяется не по всем располагаемым наблюдениям, а лишь по ближайшим q (в смысле введенной метрики) к исследуемой ситуации.

Выбор оптимального представления функции близости и оптимального значения q , может производиться следующим путем. Выбросим i -ый элемент обучающей последовательности и экстраполируем решение по A_i . Полученный вектор $X^*_{iэ}$ вообще говоря не совпадет с имеющимся X_i^* . Невязка

$$\Delta = \sum_{i=1}^{k} \left| X_i^* - X_{iэ}^* \right| \qquad (9)$$

определяет эффективность выбора функции близости и оптимальности q. Поэтому оптимальный выбор функции близости и должен минимизировать эту невязку.

Теоретическое исследование влияния соотношения длины k обучающей последовательности и размерности m пространства ситуаций на величину среднеинтегральных потерь δ_F качества системы по области ее возможных состояний было проведено для случаев линейной и нелинейной зависимости (3). Оказалось, что в обоих случаях потери монотонно уменьшаются по мере приближения k к m. При $k > m$ линейная функция восстанавливается, естественно, безошибочно, и потери равны нулю. В случае нелинейной функции в этом случае осуществляется локально-линейное приближение.

Для сравнения были определены среднеинтегральные потери δ_F^* качества системы при восстановлении тех же функций по "ассоциативному" методу, согласно которому новая ситуация отождествляется с ближайшей, входящей в состав обучающей последовательности. Как видно из рис.1, экстраполяционный метод восстановления линейной зависимости (3) приводит к существенно меньшим интегральным потерям качества системы. Методы равноценны только в том случае, когда информация о восстанавливаемой функции ничтожно мала ($k \leq m$, т.е. $N = \frac{m}{k} \to \infty$). Аналогичная картина наблюдается в случае нелинейной зависимости (3). Однако здесь относительное преимущество предложенного экстраполяционного метода уменьшается, поскольку линеаризация приводит к дополнительным потерям качества системы.

Оценки для оптимальных параметров системы, найденные путем экстраполяции, являются случайными в силу случайности предыдущего опыта. Приведенное в работе [7] исследование статистических свойств этих оценок по множеству обучающих последовательностей показало, что дисперсность оценок уменьшается по мере увеличения длины обучающей последовательности.

Некоторые применения метода.

Были рассмотрены некоторые практические задачи по применению предложенного метода многомерной линейной экстраполяции.

I. Задача оптимального проектирования системы, как известно [8, 9], связана с трудоемкой процедурой отыскания оптимальных параметров, при которых система наилучшим образом удовлетворяет экономическим, техническим и временным требованиям. Неизбежные изменения отдельных исходных условий, характеристик, ограничений и т.п., возникающие в процессе проектирования системы, приводят к необходимости многократного повторения указанных трудоемких рсчетов с целью соответствующей корректировки оптимальных параметров системы. Сложность этих расчетов и, следовательно, время проектирования системы можно значительно сократить, если воспользоваться предложенным экстраполяционным методом. Для этого вначале производится определенное исходное количество корректировок параметров системы путем применения известных прямых методов многопараметрической оптимизации. Эти расчеты и образуют обучающую последовательность. Последующие корректировки осуществляются экстраполяционным методом на основании этого опыта. Следует подчеркнуть, что обучающую последовательность можно формировать только из результатов оптимизации системы фиксированной структуры. (См. I-ое условие применимости метода). Вектор ситуации должен полностью определять совокупность условий, для которых получено оптимальное решение.

Проверку последнего условия применимости экстраполяционного метода нетрудно установить по величине минимальной невязки (9), которая отражает эффективность метода, для данной задачи.

2. Задача о предсказании поведения функции, заданной в m точках интервала наблюдения, сводится к подбору аналитического выражения, которое в определенном смысле наилучшим образом аппроксимирует наблюдаемую функцию. Пусть, например, из предыдущего опыта известно, что функции, подобные наблюдаемой, наилучшим образом аппроксимируются полиномами степени n, причем

$$n \geq m \tag{10}$$

Как известно, задача о проведении полинома степени n через $m \leq n$ точек допускает бесчисленное множество решений. Необходимо выбрать из всевозможных решений единственное, которое наименее противоречит результатам предыдущего опыта. Для этого, согласно [8] формируется следующая обучающая последовательность. Компонентами вектора ситуации A_j являются j -е значения ранее наблюдавшейся функции в отмеченных точках интервала наблюдения; компонентами вектора оптимальных параметров X_j^* служат значения коэффициентов полинома, наилучшим образом аппроксимирующего j -ю функцию. Теперь, воспользовавшись методом многомерной линейной экстраполяции, можно найти оценку для коэффициентов искомого полинома и решить задачу предсказания поведения наблюдаемой функции.

При этом необходимо соблюдать следующее условие: функции, входящие в состав обучающей последовательности и вновь наблюдаемые, должны быть близкими в том смысле, что они описывают аналогичные процессы, протекающие в аналогичных условиях.

3. Задача идентификации динамического объекта заключается в построении математической модели, изоморфной объекту по поведению. Структура и параметры такой модели подбираются из условия, чтобы при одинаковых сигналах на входах объекта и модели достигалось наименьшее в определенном смысле рассоглаоованию их выходных сигналов. Если структура модели определена, то задача оптимальной настройки сводится к нахождению вектора оптимальных параметров, экстремизирующих показатель рассогласования выходных сигналов.

Однако, в реальных условиях свойства объекта не остаются постоянными во времени. Возникает необходимость постоянной или периодической корректировки оптимальных параметров модели. Показано [II] , что в случае квазистационарного объекта задачу адаптации модели удобно решать с использованием метода многомерной линейной экстраполяции. Априорные сведения об объекте на этапе обучения экстраполятора можно получить с помощью поисковых методов самонастройки модели [IO, II] .

Применительно к задаче идентификации объекта обучающая последовательность должна устанавливать соответствие между некоторой характеристикой, определяющей его текущее состояние и параметрами математической модели объекта в этом состоянии. В процессе нормальной эксплуатации характеристикой состояния объекта могут служить или непосредственно записи его входного и выходного сигналов, или их статистические свойства. Известно [10], что наиболее информативной характеристикой является взаимнокорреляционная функция выходного и входного сигналов объекта. В отдельных случаях изменение состояния объекта допускается наблюдать по изменению автокорреляционной функции его выходного сигнала. Использование любой из упомянутых функций в качестве ситуации требует предварительной их параметризации. Учитывая исключительную трудоемкость общепринятого способа параметризации корреляционных функций [10], в работе [11] принят простейший способ представления функции в виде вектора: его компонентами являются непосредственно значения функции, например, $K_{yx}(\tau_i)$, соответствующие определенным m моментам независимого переменного τ_i $(i=1,\dots,m)$.

Экспериментальное исследование процесса корректировки оптимальных параметров модели экстраполяционным методом было проведено на ЦВМ для класса объектов, имеющих монотонную переходную характеристику. Полученные результаты при статистических оценках взаимнокорреляционных функций подтвердили возможность предложенного подхода к решению задачи оперативного приспособления модели. Наименьшая погрешность достигается при наблюдении взаимнокорреляционной функции объекта.

ВЫВОДЫ:

I. Предложен и исследован метод беспоисковой адаптации сложной системы к изменениям многомерной ситуации.

2. Метод основан на экстраполяции предыдущего поведения, накопленного в процессе оптимальной работы объекта при аналогичных условиях. Характерной особенностью метода является то, что он позволяет принимать достаточно обоснованные решения при весьма ограниченном объеме априорных сведений.

3. Экспериментальное исследование метода в задачах оптимального проектирования, предсказания, идентификации и управления показало эффективность развитого подхода к решению задачи адаптации.

ЛИТЕРАТУРА

1. Д.ДЖ.УАЙЛД. Методы поиска экстремума. Изд-во "Наука", М., 1967.
2. Л.А.РАСТРИГИН. Случайный поиск в задачах оптимизации многопараметрических систем. Изд-во "Зинатне", Рига, 1965.
3. М.А.АЙЗЕРМАН, Э.М.БРАВЕРМАН, Л.И.РОЗОНОЭР. Метод потенциальных функций в задачах о восстановлении характеристики функционального преобразователя по случайно наблюдаемым точкам. Автоматика и телемеханика, № 12, 1964.
4. Я.З.ЦЫПКИН. Адаптация и обучение в автоматических системах. М., "Наука", 1968.
5. Н.С.РАЙБМАН, В.М.ЧАДЕЕВ. Адаптивные модели в системах управления. Изд-во "Советское радио", М., 1966.
6. Л.А.РАСТРИГИН, В.С.ТРАХТЕНБЕРГ. Применение экстраполяции при оптимальном проектировании сложных систем. В кн.: Методы статистической оптимизации. Изд-во "Зинатне", Рига, 1968.
7. Л.А.РАСТРИГИН, В.С.ТРАХТЕНБЕРГ. Оценка числовых характеристик функции методом многомерной экстраполяции. В кн.: Проблемы статистической оптимизации. Изд-во "Зинатне", Рига, 1968.
8. В.С.ТРАХТЕНБЕРГ. Некоторые проблемы, связанные с решением экстремальных задач оптимального проектирования поисковы-

ми методами. Автоматика и вычислительная техника, № IO, Рига, 1965.

9. А.Л.ГАЙДУКОВ. Применение случайного поиска при оптимальном проектировании. В кн.: Прикладные задачи технической кибернетики. Изд-во "Советское радио", М., 1966.

IO. Я.А.ГЕЛЬФАНДБЕЙН. Методы кибернетической диагностики ди-намических систем. Изд-во "Зинатне", Рига, 1967.

II. В.С.ТРАХТЕНБЕРГ. Поисковая и беспоисковая самонастройка модели динамического объекта. Автоматика и вычислительная техника, № 2, 1968.

Авторы:

Растригин Леонард Андреевич - доктор технических наук, профессор.

Трахтенберг Валерий Симонович - кандидат технических наук.

Институт электроники и вычислительной техники АН Латвийской ССР.

НАДПИСЬ К РИСУНКУ

доклада Л.А.Растригина и В.С.Трахтенберга "Многомерная экстраполяция в задачах оптимального управления и проектирования".

Рис. I. Соотношение ошибок восстановления линейной функции экстраполяционным и ассоциативным методами.

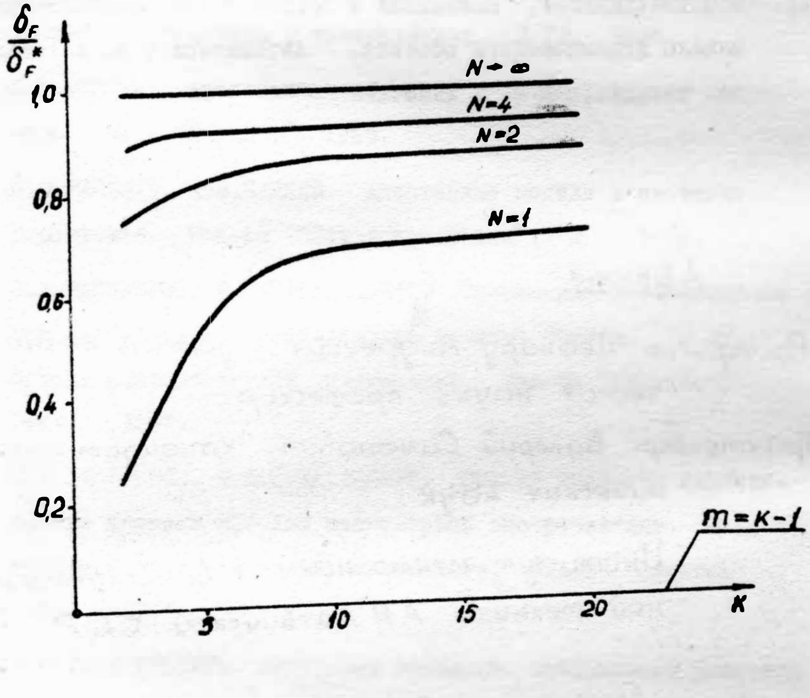


Рис. 1

MINIMUM VARIANCE ESTIMATION OF PARAMETERS AND STATES IN NUCLEAR POWER SYSTEMS

L. J. Habegger
Argonne National Laboratory
Argonne, Illinois, U.S.A.

and

R. E. Bailey
Purdue University
Lafayette, Indiana, U.S.A.

INTRODUCTION

Experimental identification of dynamic characteristics of nuclear reactor systems by current methods, such as 1/e period measurements or transfer function measurements via control rod oscillations or random noise inputs, are limited by one or more of the following:

(1) a capability of giving information about only a few isolated parameters,

(2) a restriction to stationary linear systems,

(3) a need for long experimental running time, and

(4) a need for highly specialized equipment for producing a particular system input.

In this paper, a minimum variance sequential estimation procedure - suitable for linear and nonlinear systems - is shown to be capable of bypassing these limitations. This procedure estimates a system's dynamic parameters using measurements of the system output from an arbitrary known input. Results are presented from the actual experimental application of the procedure to three reactor systems - the EBWR (Experimental Boiling Water Reactor) and EBR-II (Experimental Breeder Reactor - II) power reactors and the PUR-I (Purdue University Reactor - I) research reactor.

The work reported here was largely motivated by a need for an experimental procedure which is short enough to allow continuous or periodic determination of changes which frequently occur in the dynamic characteristics of nuclear reactor systems. The dynamic characteristics of a power reactor system may change, for example, as a result of changes in coolant flow rate, temperature, power level, pressure, control rod configuration, isotope buildup, or fuel configuration.

Applicability to nonlinear systems was also an important consideration in the development of a parameter estimation procedure since reactor dynamics are basically nonlinear.[1]

The most widely known and used method for experimentally determining reactor characteristics involves the excitation of the reactor at various frequencies with an oscillating control rod to obtain the system's transfer function.[2] These experiments are quite lengthy and thus not suited to frequent redetermination of the system characteristics. Reactor dynamics measurements have more recently been accomplished through the analysis of output noise resulting from either externally produced input noise or naturally occurring internal noise. This method significantly reduces experimental running time. However, because of a lack of information on the nature of the internal noise in power reactors, applications to this type of system usually require elaborate equipment for externally producing input noise. Both the noise analysis and rod oscillator experiments are in general restricted to stationary linear systems. Experimental procedures such as 1/e period measurements identify isolated parameters but give only limited information on the system response to an arbitrary input.

The minimum variance estimation procedure used in this paper is an extension of Kalman's solution to the linear filtering and prediction problem.[3] Aerospace applications of this procedure have also been proposed.[4,5]

NOTATIONAL CONVENTIONS

$\underline{a}$: Underlined lower case letters are column vectors.

$\underline{M}$: Underlined upper case letters are matrices.

$\underline{a}^*, \underline{M}^*$: The superscript (*) represents the transpose operation.

$E[v]$: Mean value of a random variable v.

δ_{ij}: Kronecker delta function with zero value for $i \neq j$ and unit value for $i = j$.

$\delta(t)$: Dirac delta function defined by

$$\int_{t_1 \le 0}^{t_2 \ge 0} f(t)\delta(t)dt = f(0) .$$

$\dfrac{\partial \underline{f}(\underline{x})}{\partial \underline{x}}$: Jacobi matrix with $(ij)^{th}$ element $\dfrac{\partial f_i(\underline{x})}{\partial x_j}$.

PROBLEM FORMULATION AND METHOD OF SOLUTION

The problem to be solved here is the estimation of the parameters $\underline{a}$ of the system which is described by the state equations

$$\dot{\underline{x}} = \underline{f}(\underline{x}(t), \underline{a}, t) + \underline{r}(t) \qquad (1)$$

The estimates are to be based on the discrete measurements

$$\underline{y}_i = \underline{M}_i \underline{x}(t_i) + \underline{w}_i, \; i = 1,2,\cdots,\xi \; . \tag{2}$$

Here $\underline{r}(t)$ and $\underline{w}_i$ are continuous and discrete random variables, respectively, with

$$E[\underline{r}(t)] = \underline{0}, \quad E[w_i] = \underline{0}$$
$$E[\underline{r}(t)\underline{r}^*(\tau)] = \underline{R}(t)\delta(t-\tau), \; E[\underline{w}_i\underline{w}_j^*] = \underline{W}_i \delta_{ij} \tag{3}$$
$$E[\underline{r}(t)\underline{w}_i^*] = \underline{0}$$

The system can be redefined by considering Eqn. (1) to be augmented with the equations

$$\dot{\underline{a}} = \underline{0} + \underline{r}'(t) \tag{4}$$

and using the new definition

$$\begin{bmatrix} \underline{x} \\ \underline{a} \end{bmatrix} \rightarrow \underline{x} \tag{5}$$

plus corresponding redefinitions for $\underline{f}(\cdot)$, $\underline{r}(t)$, and $\underline{M}_i$. The approach used in the procedure presented is to estimate the entire augmented state vector $\underline{x}$ in (5), which includes the parameters $\underline{a}$. Although the emphasis in this paper is on the estimation of the system parameters, the estimates of the system states, which are also given in this approach, could be of value in a feedback control system.

The algorithm for minimum variance sequential estimation of the augmented state vector $\underline{x}(t_i)$ is given below. The derivation of this procedure appears elsewhere[4,5] and will thus not be given here. In the algorithm, the estimate at time t_i , based on the measurements $\underline{y}_i$, i=1,2,$\cdots$,k, is is denoted by $\underline{x}_{i/k}$.

Given (1) the dynamic system in Eqn. (1) (augmented with Eqn. (4)), (2) the measurements in Eq. (2), and (3) the initial values $\underline{x}_{o/o}$ and $\underline{J}_{o/o}$, estimates of the states $\underline{x}(t_i)$ are given by the recursive relations

$$\underline{x}_{i/i} = \underline{x}_{i/i-1} + \underline{B}_i \left[\underline{y}_i - \underline{M}_i \underline{x}_{i/i-1}\right] \tag{6}$$

In Eqn. (6), $\underline{x}_{i/i-1}$ is the solution at time t_i of

$$\dot{\underline{x}}_a = \underline{f}(\underline{x}_a(t),t), \; \underline{x}_a(t_{i-1}) = \underline{x}_{i-1/i-1} \tag{7}$$

and

$$\underline{B}_i = \underline{J}_{i/i-1} \underline{M}_i^* (\underline{M}_i \underline{J}_{i/i-1} \underline{M}_i^* + \underline{W}_i)^{-1} \tag{8}$$
$$\underline{J}_{i/i-1} = \underline{\Phi}_{i/i-1} \underline{J}_{i-1/i-1} \underline{\Phi}^*_{i/i-1} + \underline{R}_i \tag{9}$$

In Eqn. (9) $\underline{\Phi}_{i/i-1}$ is the solution at time t_i of the matrix equation

$$\dot{\underline{\Phi}}(t,t_{i-1}) = \frac{\partial \underline{f}(\underline{x}_a(t),t)}{\partial \underline{x}_a(t)} \underline{\Phi}(t,t_{i-1}), \quad \underline{\Phi}(t_{i-1},t_{i-1}) = \underline{I} \tag{10}$$

and

$$\underline{R}_i = \int_{t_{i-1}}^{t_i} \underline{\Phi}(t_i,\tau)\underline{R}(\tau)\underline{\Phi}^*(t_i,\tau)\, d\tau \, . \tag{11}$$

Also,

$$\underline{J}_{i/i} = \underline{J}_{i/i-1} - \underline{B}_i \underline{M}_i \underline{J}_{i/i-1} \tag{12}$$

This estimation procedure has the characteristic that if $\underline{x}_{i-1/i-1}$ is unbiased and has a variance given by $\underline{J}_{i-1/i-1}$, and if, furthermore, $\underline{x}(t_i)$ can be approximated by the solution, at time t_i, of the linear relation

$$\dot{\underline{x}} \approx \underline{f}(\underline{x}_a(t),t) + \frac{\partial \underline{f}(\underline{x}_a(t),t)}{\partial \underline{x}_a(t)} (\underline{x}(t) - \underline{x}_a(t)) + \underline{r}(t), \tag{13}$$

then $\underline{x}_{i/i}$ given by Eqn. (6) will have the smallest variance of all estimates which are a linear combination of the measurements $\underline{y}_j$, $j=1,2,\cdots,i$. Under these conditions, $\underline{x}_{i/i}$ will be unbiased and will have a variance $\underline{J}_{i/i}$ as given by Eqn. (12).

The features which make the above procedure attractive for estimating reactor parameters are:

1. The experimental and computational time required is short. Thus, frequent reestimation of the system's parameters becomes practical. Since the procedure is sequential, the computations can theoretically be completed simultaneously with the transient measurements, however, the speed of the computer often prohibits this from occurring. Also, if the initial estimates, $\underline{x}_{o/o}$, are poor, the data may have to be used more than once to obtain satisfactory estimates.
2. The technique is applicable to nonlinear systems and systems with time varying coefficients.
3. The type of system input to be used is not specified. Thus the need for specialized equipment is reduced.
4. If during the transient more than one system state, or linear combinations of states, are measured simultaneously, these measurements can all be used in the estimation. The accuracy of the estimates generally increases with the order of the measurement vector, $\underline{y}_i$.
5. The effects of both measurement noise and system input noise are considered in the estimation procedure.

The major restriction in the use of the procedure is that the user must provide the following *a priori* information:

(1) the initial estimate $\underline{x}_{o/o}$

(2) the form of the model in Eqn. (1) describing the system dynamics, and

(3) the second order statistics $\underline{J}_{o/o}$, $\underline{R}(t)$, and $\underline{W}_i$.

EXPERIMENTAL APPLICATIONS

Experimental Boiling Water Reactor (EBWR)

The EBWR at Argonne National Laboratory during the period 1956 - 1962 served as a prototype nuclear power plant using a natural circulation, pressurized, boiling water reactor in which the steam produced was used directly to operate a turbine generator.[6] From 1964 to 1967 the EBWR was again placed into operation for the primary purpose of plutonium irradiation in the Plutonium Recycle Program.[7] The experimental data used in the following analysis of EBWR characteristics was obtained during this latter period of operation. Figure 1 shows, in block diagram form, the simplified theoretical model used for the identification of the EBWR dynamics. In this model, the plant feedback reactivity, $\delta k_f(t)$, is assumed to result only from void formation in the core. The power voids represent a change in the void formation rate due to a change in the heat formation rate. The flashing voids are voids which form as the result of pressure changes. The boundary voids represent the effect of variations in the height of the boundary between the boiling and non-boiling regions. A similar model has previously been used for the analysis of EBWR transfer function measurements.[2]

The zero-power reactor kinetics were assumed to be of the well known form[8]

$$\dot{N} = \left[\frac{\delta k_c(t) + \delta k_f(t)}{\ell} - \beta/\ell\right] N + \sum_{i=1}^{3} \lambda_i C_i \qquad (14)$$

$$\dot{C}_i = \frac{\beta_i}{\ell} N - \lambda_i C_i \ , \ i = 1, 2, 3 \qquad (15)$$

where

$N(t)$:	Reactor power level
$C_i(t)$:	Concentration of the i^{th} delayed neutron precursor
β:	Fraction of neutrons produced which are delayed
β_i/β:	Fraction of delayed neutrons from the i^{th} precursor group
λ_i:	Decay constant for the i^{th} precursor group
$\delta k_c(t)$:	Control rod reactivity
$\delta k_f(t)$:	Plant feedback reactivity

ℓ: Prompt neutron lifetime

The power output in a boiling water reactor contains non-negligible random fluctuations which are mainly the result of random boiling void formation. These random fluctuations were modeled by assuming the addition of a random input, r(t), on the right hand side of Eqn. (14).

In the EBWR experiments, the transient used in the estimation was obtained by dropping a control rod. This input, as a function of the measured time dependent rod position, h(t), was modeled by

$$\delta k_c(t) = \Delta k\ p(t) \tag{16}$$
$$= k_c(t) - k_c(t_o)$$

where

$$k_c(t) = \Delta k \left[\frac{h(t)}{H} - \frac{1}{2\pi} \sin \left(2\pi \frac{h(t)}{H}\right) \right] \tag{17}$$

In Eqn. (16), Δk is the reactivity worth of a fully inserted rod and is an unknown parameter to be estimated. The active core height is H and t_o is the time at which the rod was dropped. The form of Eqn. (17) is derived from perturbation theory for a thin control rod in an unreflected, homogeneous reactor.[9]

Through the use of this model, the transient measurements, and the a priori information $\underline{x}_{o/o}$, $\underline{J}_{o/o}$, $\underline{W}_1$, and $\underline{R}(t)$, the minimum variance sequential estimation of states and parameters in the EBWR system is possible. However, as will often be the case in experimental applications, a limited imprecision exists in the specification of the required a priori information. Therefore, prior to the experimental application, the effects on the estimation resulting from the use of various assumptions for this information were investigated using computer simulated measurements of EBWR power transients.

The results of this study, which are given in Reference 10, indicated that satisfactory parameter estimates can still be obtained if the a priori information used contains inaccuracies of the limited magnitude normally occurring in the experimental applications.

In the actual experimental application to the EBWR, the parameter estimates were based on simultaneous measurements of both pressure and power during rod drop transients. The sampling interval during the transients was 0.06 seconds. The parameter estimation was repeated at six different power levels, ranging from 15.1 to 69.3 megawatts (thermal) to determine the parameters as a function of power. The parameters estimated were Δk, T_h, T_{PV}, T_r, τ, σ, $(K_{\delta k}V)$, $(K_{\delta k}d)$, $(K_{\delta k}b)$, and (BK) (see Fig. 1).

Because of the large number of parameters estimated and, at several power levels, a fairly large discrepancy between the data and the transients predicted by the initial estimates (see for example Fig. 2), there was a possibility of obtaining unsatisfactory estimates during the simultaneous estimation of all ten of the unknown parameters. This difficulty was avoided by taking advantage of several observations concerning the system transient. First, the major drop in power occurs during the short period (~0.25 seconds) in which the control rod is dropping at the start of the transient. The magnitude of this initial power change is highly dependent on the control rod worth parameter, Δk, but is only slightly dependent on the remaining unknown parameters which relate to the relatively slow feedback effects. Therefore the parameter Δk can be estimated independently by using only this first portion of the transient. A second observation is that the four unknown feedback parameters T_h, T_{PV}, σ, and $(K_{\delta k}V)$ play the major role in determining the system response after the initial power change and can therefore be estimated independently using this region of the measured response. Although this second observation was originally based on empirical results using different parameters in the model, a similar conclusion can be obtained by considering the physical processes of the system. In a boiling water reactor, the principal feedback effect is from power voids and, from Fig. 1, this is related to the parameters T_h, T_{PV}, and $(K_{\delta k}V)$. The pressure change is directly related to the parameter σ which determines the steaming rate. Using these two observations, a first coarse estimate can be made of the parameters Δk, T_h, T_{PV}, σ, and $(K_{\delta k}V)$, followed by a more refined simultaneous estimate of all ten unknown parameters. Following this procedure, the parameter estimates in Table 1 were obtained.

Although no true values for the parameters are available for evaluating the parameter estimates, these estimates display the power dependency expected from the model. The estimate changes also reflect a dependency on the changes in other system variables such as boric acid concentration in the coolant, coolant flow rate, and fuel depletion. The parameter estimates are also in the general range of values given by rod oscillator tests. The rod oscillator estimates, however, cannot be used to absolutely verify the estimates in Table 1 since the oscillator results are based on a linearized model which required much smaller experimental system state variations.

For many applications, such as to adaptive control, any parameter estimates are satisfactory that give good agreement between predicted and experimental trajectories. For both power and pressure, the estimates in Table 1 resulted in such a desirable agreement. Examples of the theoretical and experimental trajectories are given in Figures 3 and 4. These figures also show the trajectories obtained using the initial parameter estimates.

The evaluation of the parameter estimates from the viewpoint of their physical meaning is limited by the accuracy of the simplified system model shown in Fig. 1. One shortcoming of this model is the assumption that the transfer function in each process in the feedback is independent of the states of the system. Another approximation in the model is the assumption of space-independent dynamics. The result of these and other assumptions is that the parameter estimates may vary from the values associated with their intended physical definition so that the inexact model will still approximate the system response.

Purdue University Reactor - I (PUR-I)

In the estimation of the EBWR parameters it was assumed that the parameters in the zero-power reactor kinetics, Eqns. (14) and (15), were those given by rod-oscillator tests at low power ($\delta k_f(t) \approx 0$). However, these parameters could also have been obtained at this low power using the minimum variance estimation procedure. This capability was experimentally verified using a transient power measurement from the Purdue University low power, light water moderated, research reactor.[10]

Experimental Breeder Reactor - II (EBR-II)

The EBR-II, located at the National Reactor Testing Station in Idaho, is a first generation liquid sodium cooled fast breeder reactor with the primary purpose of providing information on complete breeder reactor fuel cycles. This reactor also serves as a prototype for larger power plants of this type. The EBR-II has a design power level of 62 megawatts (thermal).[11]

The minimum variance estimation procedure was applied to this system using measurements of a power transient provided to the authors by the National Reactor Testing Station. This transient was the result of dropping a control rod at a power level of 45 megawatts (thermal). The system reactivity input was represented by

$$\delta k_c(t) = \Delta k \; p(t) \tag{19}$$

where Δk is the control rod worth, as in Eqn. (17), and p(t), the "shape function," was obtained from a previous experiment.[10]

The system model in this application was assumed to consist of the reactor kinetics in Eqns. (14) and (15) with the feedback reactivity

$\delta k_f(t)$, given by

$$\delta k_f = D\delta k_f + U \frac{N-N_o}{N_o} \qquad (20)$$

In a fast reactor such as the EBR-II, the principal feedback effect is due to various core temperature changes. Thus the parameters D and U in Eqn. (20) can be roughly related to the time constant and gain, respectively, of a weighted average core temperature change in response to a power change.

Theoretically, the minimum variance estimation procedure may now be used directly to obtain estimates of parameters in Eqns. (14), (15), (19), and (20), but computational problems require a further manipulation of these equations. In a fast reactor such as the EBR-II, the neutron lifetime, ℓ , is very small. As a result, the individual terms on the right hand side of Eqn. (14) are very large, although their sum, $\dot{N}$, is relatively small. Consequently, the numerical integration time step must be made prohibitively small to prevent errors in the calculation of $\dot{N}$.

To avoid this computational difficulty, the following approximation was made:

From Eqn. (14),

$$(\delta k_c(t) + \delta k_f(t) - \beta)\, X_1 + \sum_{i=1}^{m} \beta_i X_{i+1} = \ell \dot{X}_1 \approx 0 \qquad (21)$$

since ℓ is small. In Eqn. (21)

$$X_1 \equiv \frac{N-N_o}{N_o}\,, \quad X_{i+1} \equiv \frac{\ell\lambda_i}{\beta_i} C_i - 1 .$$

The approximate algebraic relation in Eqn. (21) could now be used directly to eliminate X_1 . But since X_1 , the reactor power, is the measured variable, the variable $\delta k_f(t)$ will be eliminated. This is accomplished by differentiating Eqn. (21) with respect to time and then using Eqn. (20) and (21) to eliminate $\delta k_f(t)$ and $\delta\dot{k}_f(t)$ in the resulting equations. The final set of equations is

$$\dot{X}_1 = \frac{(X_1 + 1)}{\sum_{i=1}^{m} \beta_i (X_{i+1} + 1)} \left\{ \sum_{i=1}^{3} \beta_i \lambda_i (X_1 - X_{i+1}) + (X_1 + 1) \left[\delta\dot{k}_c(t) - D(\delta k_c(t) - \beta) + X_1 U \right] \right\} - (X_1 + 1) D \qquad (22)$$

$$\dot{X}_{i+1} = \lambda_i (X_1 - X_{i+1})\,, \quad i = 1, 2, 3 . \qquad (23)$$

This approximate form gave a transient which differed from the original by less than 0.5% at each point while decreasing the required number of integration steps by a factor of 250.

Although the EBWR model which was used is of higher order, the use of the EBR-II model in Eqns. (22) and (23) provides an additional test of the estimation procedure because of the highly nonlinear nature of Eqn. (22). The capability of the procedure to estimate parameters in this model was demonstrated using computer simulated noise free data. Using this data, the parameters Δk, D, and U, which initially had a 20% average error, were simultaneously estimated to an accuracy of 6 significant figures.

In the application to the EBR-II experiments, the parameters Δk, D, and U were estimated by the minimum variance estimation procedure to be

$$\Delta k = 0.0003034, \quad D = -8.755, \quad U = -0.008377.$$

The measured transient and the transient using the initial and final parameter estimates are shown in Fig. 4. From this figure, it is seen that the estimation procedure was again successful in obtaining estimates which give a very good fit to the data.

CONCLUSIONS

Experimental applications to the boiling water, zero-power, and fast breeder nuclear reactor types have demonstrated that a minimum variance sequential estimation procedure can be a powerful and versatile tool for the identification of reactor system dynamics. A short experimental running time, applicability to nonlinear systems, and small requirements for experimental equipment, make the procedure attractive as a replacement or supplement for conventional reactor identification procedures. The simultaneous estimation of the system states by the procedure, in addition to the parameter estimation, introduces the possibility of applications to adaptive feedback control systems.

Although the emphasis in this paper is on the application to nuclear reactor systems, the procedure is applicable to any system described by ordinary differential equations with a forward-biased independent variable. Use with systems described by partial differential equations is feasible if the system equations can be approximated by ordinary differential equations through modal or nodal expansions.

In addition to the experimental applications, the minimum variance estimation procedure may be of value in the optimization of a system design. For example, the procedure may be used to estimate the design parameters which give the "best fit" of the predicted system response to a known optimal response.

REFERENCES

1. Chernick, J., "A Review of Nonlinear Reactor Dynamics Problems," BNL-774 (1962).
2. DeShong, J. A., Jr., Lipinski, W. C., "Analyses of Experimental Feedback Transfer Functions for a Natural Circulation Boiling Water Reactor," ANL-5850 (1958).
3. Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems," Trans. ASME Journal of Basic Engr., Vol. 82, p. 35 (1960).
4. Kopp, R., Orford, R. J., "Linear Regression to System Identification for Adaptive Systems," Journal of AIAA, Vol. 1, No. 10, p. 2300 (1962).
5. Lear, W. M., "On the Use of Ultrastable Oscillators and a Kalman Filter to Calibrate the Earth's Gravitational Field," Ph.D. Thesis, Purdue University (1965).
6. "The Experimental Boiling Water Reactor," ANL-5607 (1957).
7. Toppel, B. J., "EBWR Plutonium Recycle Experiment," ANL-7010, p. 69 (1965).
8. Weaver, L. E., System Analysis of Nuclear Reactor Dynamics, Rowman and Littlefield, Inc. New York, N. Y., p. 51 (1963).
9. Weinberg, A. M., Wigner, E. P., The Physical Theory of Neutron Chain Reactors, University of Chicago Press, Chicago, Ill., p. 548 (1959).
10. Habegger, L. J., "The Use of a Kalman Filter for the Estimation of Nuclear Reactor Parameters," Ph.D. Thesis, Purdue University (1968).
11. Loftness, R. L., Nuclear Power Plants, Van Nostrand Co., Inc., Princeton, N. J. (1964).

Table 1. Minimum Variance Estimates of EBWR Parameters

Power, Mw(th)	15.1	28.5	38.5	49.5	57.0	69.3
Δk	0.0142	0.0163	0.0103	0.0080	0.0087	0.0080
T_h	9.00	10.52	6.00	8.62	5.24	5.00
T_{PV}	0.0164	0.0150	0.0207	0.00945	0.00692	0.00503
T_r	13.0	13.0	11.0	11.9	76.9	74.1
τ	0.099	0.099	0.099	0.097	0.111	0.100
σ	20.5	19.4	25.5	45.6	43.8	68.6
$K_{\delta k}V$	0.0124	0.0276	0.0144	0.0290	0.0316	0.0321
$K_{\delta k}d$	0.000108	0.000107	0.000102	0.000084	0.000115	0.00011
$K_{\delta k}b$	0.000412	0.000403	0.000397	0.000477	0.000606	0.00061
BK	.237	.238	.192	.282	.769	1.043

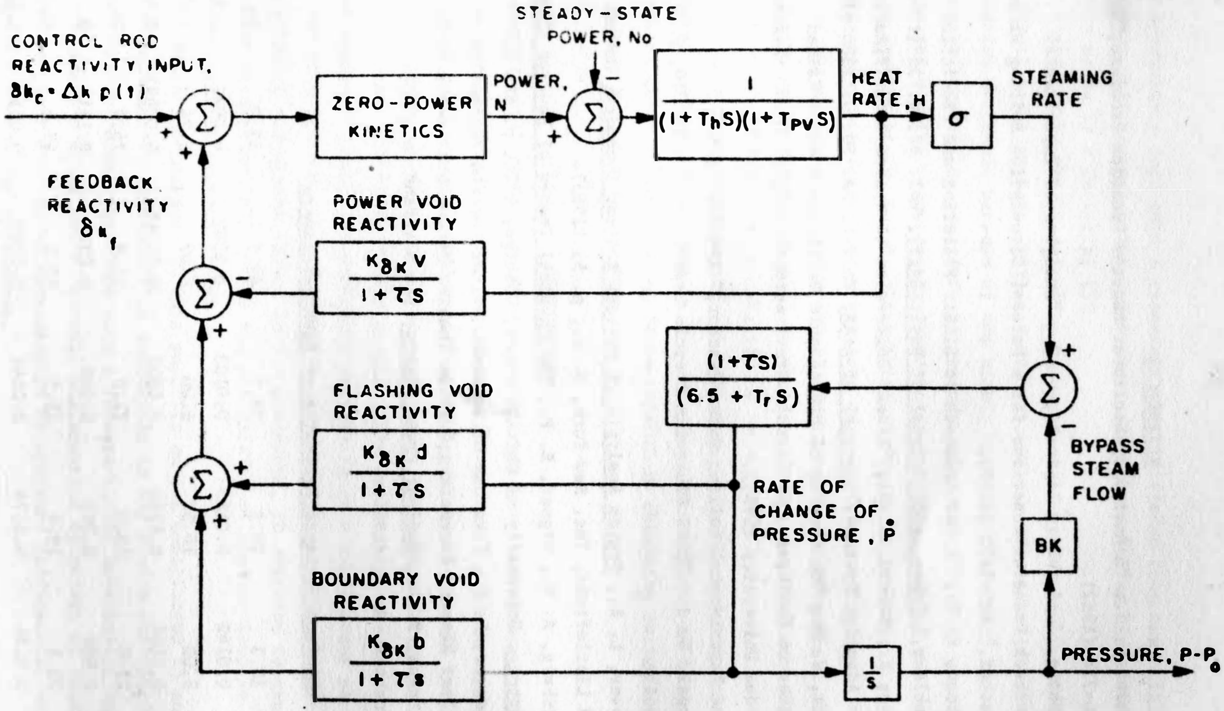


Figure 1 . Block Diagram of EBWR Dynamics Model.

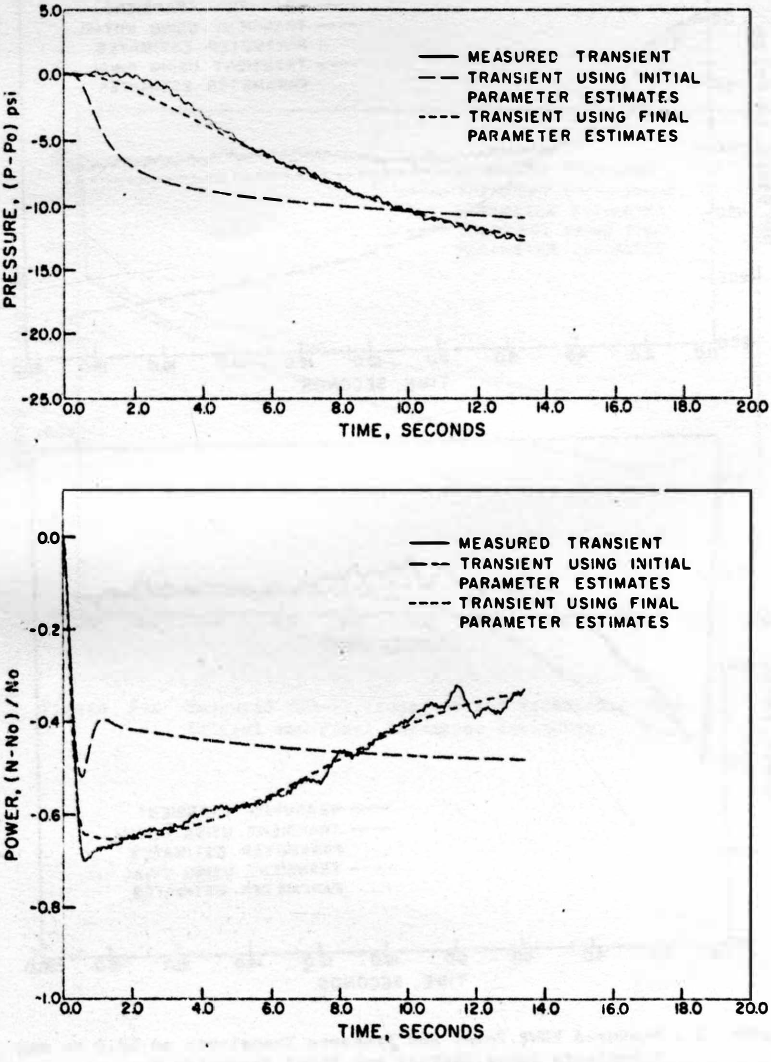


Figure 2 . Measured EBWR Power and Pressure Transients at 28.5 Mw and Transients Using Initial and Final Parameter Estimates.

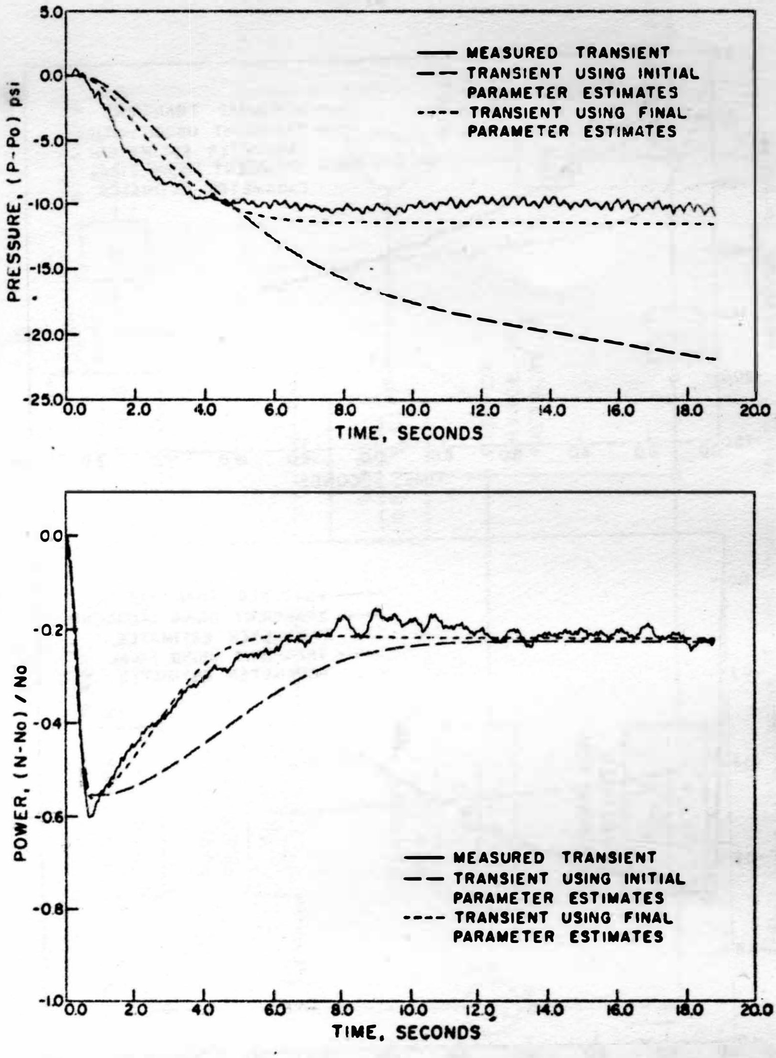


Figure 3 . Measured EBWR Power and Pressure Transients at 57.0 Mw and Transients Using Initial and Final Parameter Estimates.

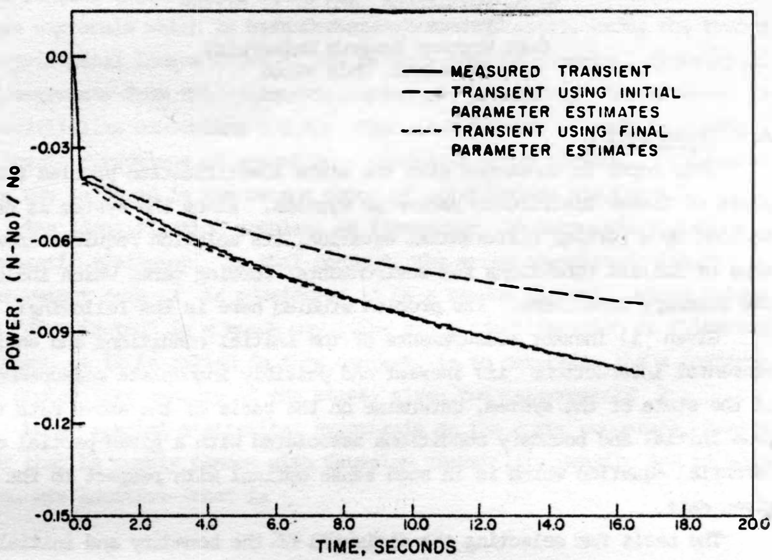


Figure 4 . Measured EBR-II Transient and Transients Using Initial and Final Parameter Estimates.

STATE IDENTIFICATION OF A CLASS OF LINEAR DISTRIBUTED SYSTEMS

by

G. A. Phillipson* and S. K. Mitter
Systems Research Center
Case Western Reserve University
Cleveland, Ohio 44106

A. Introduction

This paper is concerned with the state identification problem for a class of linear distributed parameter systems. Since the system is described by a partial differential equation, its solution requires knowledge of initial conditions and environmental forcing terms which include the boundary conditions. The problem studied here is the following:

Given i) inexact measurements of the initial conditions and environmental interactions ii) inexact and possibly incomplete measurements of the state of the system, determine on the basis of the above data the true initial and boundary conditions associated with a given partial differential equation which is in some sense optimal with respect to the given data.

The basis for selecting the estimates of the boundary and initial conditions associated with a given partial differential equation, that is, the criterion of optimality, is that of "least squares". To be more precise, we mean the following:

Given:

(1) The measurement data, which we denote here by Z , and

(2) An (arbitrary) solution of the partial differential equation, denoted here by $Y(\underline{v})$, where $\underline{v}$ is an arbitrary estimate of the true initial state and boundary conditions, then

Obtain:

(1) $\underline{v}$ which extremizes the error functional

$$J(\underline{v}) = ||Z - Y(\underline{v})||^2 ,$$

---

*Now at Shell Development Company, Houston, Texas.

where $||\cdot||^2$ is some appropriate squared metric.

The identification problem, treated here is thus a variational problem—that of characterizing extremals to a given functional, constrained by a partial differential equation. We obtain a characterization of these extremals which is both necessary and sufficient, using the theory of variational inequalities.[1] Two methods for the numerical recovery of the extremals from this characterization are presented. One of these is a Ricatti-like decoupling and the other is a "direct method" involving conjugate directions of search on a quadratic error surface. Collateral work may be found in the recent paper of Balakrishnan and Lions.[2]

The identification problem, as introduced, is customarily given a stochastic treatment. In that context, the error associated with the measurement data Z is considered to be a random variable, whose values are "distributed" in a known way. The state identification or filtering problem, as it is called in this context, is to determine the a posteriori probability density of the state, given the measurements Z.

Under special statistical hypothesis on the error processes, namely that they be purely random with Gaussian probability density and in addition, are additive—that is

$$Z = Y(\underline{u}) + E$$

where $\underline{u}$ is the true "state of nature" and E is the error process, then if the system state evolution process is also linear, the a posteriori density of the states is also Gaussian. It can be shown[3] that the filtered estimate (given in terms of the sufficient statistics of the Gaussian distribution of the states, the mean and variance) coincides with the "least-squares" estimate. Thus, under these special hypotheses the variational and stochastic approaches yield identical results.

We remark that the variational problems arising in distributed optimal control are amenable to the solution techniques suggested in the sequel. In particular, optimal boundary controllers are recovered efficiently by the "direct method" already mentioned.

B. Definitions and Mathematical Preliminaries

Let Ω be a simply connected, bounded open set in R^r. Points of Ω are denoted by $x = (x_1\ x_2\ \ldots\ x_r)$. Γ is the boundary of Ω. Let t denote time, $t \in (0,T]$. Define the sets:

$$\Sigma = \Gamma \times (0, T] \quad ; \quad Q = \Omega \times (0, T]$$

We adopt a notational convention regarding the functions f :

$f(x,t)$ is a point in R^1 , $(x,t) \in Q$

$f(\cdot,t)$ is an element of a Hilbert space $K(\Omega)$

$f(\cdot,\cdot)$ is an element of the Hilbert space $L^2(0,T; K)$,

where $L^2(0,T; K)$ is the space of functions (equivalence classes) which are square integrable with values in K . When appropriate, we shall consider derivatives of f to be taken in the distribution sense, that is, given a function $\phi(\cdot) \in C^1(\Omega)$ with compact support in Ω , then for $f(\cdot) \in K(\Omega)$, the mapping

$$\frac{\partial f(x)}{\partial x_i} : \phi(x) \to - \int_\Omega f \frac{\partial \phi}{\partial x} dx \; ; \; i = 1, 2, \ldots r$$

is called the distribution derivative of the function f . Higher order derivatives are taken in an analogous way.

Define the second order elliptic operator $A[\cdot]$:

$$A[\psi] = - \sum_{i,j=1}^{r} \frac{\partial}{\partial x_i} \left[a_{ij}(x,t) \frac{\partial \psi(x,t)}{\partial x_j} \right] + a_o(x,t)\psi$$

where $a_{ij}(x,t)$, $(i,j = 1, 2 \ldots r)$ are bounded, measurable and exhibit the coercive property:

$$\sum_{i,j=1}^{r} a_{ij}(x,t)\, \xi_i\, \xi_j \geq \alpha(\xi_1^2 + \ldots + \xi_r^2) \quad \text{for all} \quad \xi_i \in R^1 \; , \; (i = 1, 2 \ldots r)$$

$$a_o(x,t) \geq \alpha$$

$$\alpha > 0 \; , \quad (x,t) \in Q \; .$$

By $A_s[\cdot]$ we mean the operator $A[\cdot]$ with an additional symmetry condition:

$$a_{ij}(x,t) = a_{ji}(x,t) \quad (i,j = 1, 2 \ldots r) \; , \text{ for all } (x,t) \in Q \; .$$

The case where the coeficients $a_{ij}(x,t) = a_{ij}(x)$, $a_o(x,t) = a_o(x)$ leads to the classical Sturm Liouville operator, denoted by A_{st} .

We shall be concerned with the properties of solutions to the Sturm Liouville problem

$$A_{st}[w] - \lambda \rho w = 0 \tag{1}$$

with any one of the boundary conditions

$$\text{(I)}\quad w(s) = 0 \qquad s \in \Gamma$$

$$\text{(II)}\quad \frac{\partial w(s)}{\partial \nu} = 0 \qquad s \in \Gamma$$

$$\text{(III)}\quad \frac{\partial w(s)}{\partial \nu} + \beta(s)\, w(s) = 0 \qquad s \in \Gamma\ ;\quad \beta(s) > 0 \text{ for all } s \in \Gamma\ .$$

Solutions to (1) with any one of (I) (II) (III) are complete in $L^2(\Omega)$. Of special interest are the solutions to

$$\left.\begin{aligned} &\frac{\partial^2 w}{\partial x^2} + \lambda w = 0 \qquad x \in (0,1) \\ &w(0) = 0 \\ &w(1) = 0 \end{aligned}\right\} \qquad (1)^*$$

namely,

$$\{\sqrt{2} \sin \sqrt{\lambda_i}\, x\}_{i=1,2\dots} \quad ; \quad \lambda_i = (i\pi)^2$$

C. The Distributed Systems

We consider in detail identification problems associated with the distributed system whose evolution equation is linear, parabolic, with inhomogeneous boundary conditions of the Dirichlet type:

$$\left.\begin{aligned} &\frac{\partial y(x,t)}{\partial t} + A[y(x,t)] = f(x,t) && (x,t) \in Q \\ &y(s,t) = u_1(s,t) && (s,t) \in \Sigma \\ &y(x,o) = u_2(x) && x \in \Omega \end{aligned}\right\} \qquad (2)$$

Hypothesis on $f(x,t)$ and $u_1(s,t)$, $u_2(x)$ are:

$$\left.\begin{aligned} &f(\cdot,\cdot) \in L^2(Q) \\ &u_1(\cdot,\cdot) \in L^2(\Sigma) \\ &u_2(\cdot) \in L^2(\Omega) \end{aligned}\right\} \qquad (3)$$

For the system (2) with hypothesis (3), we have the following Lemma:

Lemma 1 (Lions-Magenes)

There exists one and only one solution to (2) with (3) such that $y(\cdot,\cdot) \in L^2(Q)$. In addition,

$$\frac{\partial y}{\partial x}(\cdot,\cdot) \notin L^2(Q)\ .$$

Remark All our results hold in the case where the boundary conditions on (2) are Neumann or "Mixed". Moreover, systems whose evolution equation is of second order hyperbolic type, namely:

$$\frac{\partial^2 y(x,t)}{\partial t^2} + A_s[y(x,t)] = f(x,t)$$

with any of the three boundary conditions fall within the jurisdiction of our results.[3]

D. Mathematical Statement of the Identification Problem

Given: (i) System evolution process-equation (2).

(ii) Input measurements:

$$\underline{z} = \begin{bmatrix} z_1(s,t) \\ \\ z_2(x) \end{bmatrix} = \begin{bmatrix} u_1^*(s,t) + K_1 N_1(t) \\ \\ u_2^*(x) + K_2 N_2(t) \end{bmatrix} \tag{5}$$

where $\underline{u}^* = [u_1^*(s,t) \vdots u_2^*(x)]^T$ is the true "state of nature", and K_1 and K_2 are constants.

(iii) Output measurements:

(a) $z(x,t) = y(x,t;\underline{u}^*) + K_o N_o(t)$ (6)

(b) $z(x^i,t) = y(x^i,t;\underline{u}^*) + K_o^i N_o^i(t)$; $i = 1,2..\nu$ (7)

where $x^i \in \Omega$, $N_o(t)$, $N_o^i(t)$, $N_1(t)$ and $N_2(t)$ are random error processes and K_o is a constant.

Identification Problem: Obtain $\underline{u}$, a "refined estimate" of $\underline{u}^*$, based on the data contained in the input and output measurements. The "refined estimate" is defined as that $\underline{u}$ in an admissible set of functions V which extremizes a certain quadratic error functional $J(\underline{v})$. That is, choose $\underline{u}$ such that

$$J(\underline{u}) = \inf_{\underline{v} \in V} J(\underline{v}) \; ; \quad V = L^2(\Sigma) x L^2(\Omega) \; .$$

It is possible to consider a large variety of error functionals $J(\underline{v})$. This variety is induced by the type of measurement data available ((ii) and (iii)). A careless choice of functional $J(\underline{v})$ can lead to erroneous results. We postpone a discussion of "well set" functionals to Section E. Two specific error functionals considered in this study are induced by the two output measurements (iiia) and (iiib). They are:

$$\text{(a)}\quad J(\underline{v}) = \int_Q [y(x,t;\underline{v}) - z(x,t)]^2 dx\,dt + \int_\Sigma [v_1(s,t) - z_1(s,t)]^2 ds\,dt$$

$$+ \int_\Omega [v_2(x) - z_2(x)]^2 dx \quad . \qquad (8)$$

$$\text{(b)}\quad J(\underline{v}) = \int_0^T \sum_{i=1}^{\nu} [y(x^i,t;\underline{v}) - z(x^i,t)]^2 dt + \int_\Sigma [v_1(s,t) - z_1(s,t)]^2 ds\,dt$$

$$+ \int_\Omega [v_2(x) - z_2(x)]^2 dx \quad . \qquad (9)$$

Remarks The output measurement process (iiia) is physically unrealistic as it is not possible to measure the entire spatial profile. For the same reason, so is the input measurement process $z_2(x)$. The latter case can be rationalized however, by asserting that $z_2(x)$ is obtained by computing an initial steady state profile which is in error. Although not considered in this paper, it is possible to treat other measurement processes (provided they are appropriately formulated) by using the methods of this paper.

For notational convenience, we shall consider (in detail) the identification problem associated with (8) and report formally the results for (9).

E. Characterization of Extremals

The characterization of extremals to $J(\underline{v})$ is afforded by the results of Lions and Stampacchia.[1] We first introduce the appropriate framework.

Let $a(\underline{v},\underline{w})$ be a coercive continuous bilinear form, $\underline{v},\underline{w} \in V = L^2(\Sigma) \times L^2(\Omega)$

$\ell(\underline{v})$ be a continuous linear form.

Then, if

$$J(\underline{v}) = a(\underline{v},\underline{v}) - 2\ell(\underline{v}) + c \qquad (10)$$

we have the following theorem:

Theorem 1 (Lions-Stampacchia):[1] There exists one and only one $\underline{u} \in V$ such that

$$J(\underline{u}) \le J(\underline{v}) \quad \text{for all} \quad \underline{v} \in V$$

and it is characterized by

$$a(\underline{u},\underline{v}) - \ell(\underline{v}) = 0 \quad \text{for all} \quad \underline{v} \in V \qquad (11)$$

Theorem 1 is an appropriate "maximum principle" for the purposes of solving the given identification problem. It is necessary to check whether $J(\underline{v})$ given by (8) (or (9)) has the representation (10). Using (8), we can define

$$a(\underline{v},\underline{v}) = \int_Q [y(x,t;\underline{v}) - y(x,t;\underline{o})]^2 dx\, dt + \int_\Sigma v_1(s,t)^2 ds\, dt + \int_\Omega v_2(x)^2 dx \tag{12}$$

$$\ell(\underline{v}) = -\{\int_Q [y(x,t;\underline{v}) - y(x,t;\underline{o})][y(x,t;\underline{o}) - z(x,t)]dx\, dt - \int_\Sigma v_1(s,t)\, z_1(s,t) ds\, dt - \int_\Omega v_2(x)\, z_2(x) dx\} \tag{13}$$

$$c = \int_Q [y(x,t;\underline{o}) - z(x,t)]^2 dx\, dt + \int_\Sigma z_1(s,t)^2 ds\, dt + \int_\Omega z_2(x)^2 dx \tag{14}$$

Then it is clear that $J(\underline{v})$, given by (8), can be written:

$$J(\underline{v}) = a(\underline{v},\underline{v}) - 2\ell(\underline{v}) + c$$

with $a(\underline{v},\underline{v})$, $\ell(\underline{v})$ and c given by (12), (13) and (14), respectively. Moreover, the hypothesis on $a(\underline{v},\underline{v})$, $\ell(\underline{v})$ and $\underline{c}$ are satisfied. Hence, by Theorem 1, the refined estimate $\underline{u}$, which minimizes $J(\underline{v})$, is uniquely characterized by:

$$\int_Q [y(x,t;\underline{u}) - z(x,t)][y(x,t;\underline{v}) - y(x,t;\underline{o})]dx\, dt + \int_\Sigma [u_1(s,t) - z_1(s,t)][v_1(s,t)]ds\, dt + \int_\Omega [u_2(x) - z_2(x)]\, v_2(x) dx = 0 . \tag{15}$$

Equation (15) is not of immediate utility. However, by defining a system adjoint to (2), (15) can be manipulated to yield a more workable result. Thus, define $p(x,t)$, the adjoint variable to $y(x,t)$, which evolves according to:

$$\left.\begin{aligned} -\frac{\partial p(x,t)}{\partial t} + A[p(x,t)] &= y(x,t;\underline{u}) - z(x,t) && x,t \in Q \\ p(s,t) &= 0 && s,t \in \Sigma \\ p(x,T) &= 0 && x \in \Omega \end{aligned}\right\} \quad (16)$$

It can be shown[3] that (15) is equivalent to:

$$\left.\begin{aligned} -\frac{\partial p(s,t)}{\partial \nu^*} + u_1(s,t) - z_1(s,t) &= 0 && s,t \in \Sigma \\ p(x,o) + u_2(x) - z_2(x) &= 0 && x \in \Omega \end{aligned}\right\} \quad (17)$$

Thus the simultaneous solution of (2), (16) and (17) defines the defined estimate $\underline{u}$ and yelds the refined estimate of the state, $y(x,t;\underline{u})$.

Remark The extremal to the functional $J(\underline{v})$ given by (9) is given by solving (2) and (17) simultaneously with an equation for $p(x,t)$ given by

$$\left.\begin{aligned} -\frac{\partial p(x,t)}{\partial t} + A[p(x,t)] &= \sum_{i=1}^{\nu} y(x,t;\underline{u}) - z(x,t)]\, \delta(x - x^i) \quad (i=1,2..\nu) \\ p(s,t) &= 0 \\ p(x,T) &= 0 \end{aligned}\right\} \quad (18)$$

It can be shown,[4] that (16) and (18) have solutions such that $\frac{\partial p}{\partial \nu^*}(\cdot,\cdot) \in L^2(\Sigma)$, so that (17) makes sense.

We remarked in Section D that it was possible to construct functionals $J(\underline{v})$ which were not "well set". By well set, we mean that a representation for $J(\underline{v})$ given by (10) is possible. As an example of a non well set problem, consider

$$J(\underline{v}) = \int_{\Omega} [y(x,T;\underline{v}) - z(x,T)]^2 dx + \int_{\Sigma} [v_1(s,t) - z_1(s,t)]^2 ds\, dt + \int_{\Omega} [v_2(x) - z_2(x)]^2 dx \quad (19)$$

As before, we can define

$$a(\underline{v},\underline{v}) = \int_\Omega [y(x,T;\underline{v}) - y(x,T;\underline{o})]^2 dx + \int_\Sigma v_1(s,t)^2 ds\,dt$$

$$+ \int_\Omega v_2(x)^2 dx \tag{20}$$

Now, $a(\underline{v},\underline{v})$ is not continuous[3] and the representation fails. It is possible to construct several such ill-posed problems.[2,3] Appropriate reconstruction can, however, relieve these difficulties.[2,3]

F. Recovery of the Extremals

As we announced in Section A, two methods for the recovery of extremals from the characterization given by (2), (16) or (18) and (17) can be proposed. Consider first a Ricatti-like Decoupling.

F.1 Ricatti-like Decoupling

We note that (2), (16) or (18) and (17) constitutes a two point (time) boundary-value problem. That is, the "initial" conditions on $y(x,t)$ and $p(x,t)$ are split. It is possible to determine an equation for $y(x,T)$, with which the system of equations (2), (16) or (18) and (17) can be solved (in principle) as an initial value problem. However, $y(x,T)$ --that is, $y(x,T;\underline{u})$ is the refined state estimate at the terminal time T, which is fixed, but arbitrary. Thus we shall consider the identification problem to be solved once having obtained an equation for $y(x,T;\underline{u})$. We give the result as a theorem:

Theorem 2 Given the system of Equations (2), (16) and (17), then if $P(x,\xi,t)$ satisfies

(a) $$\frac{\partial P(x,\xi,t)}{\partial t} - A_\xi[P(x,\xi,t)] - A_x[P(x,\xi,t)]^* - \delta(\xi - x)$$

$$+ \int_{\Gamma_s} \frac{\partial P(x,s,t)}{\partial \nu^*} \frac{\partial P(s,\xi,t)}{\partial \nu^*} ds = 0\ , \quad (x,\xi,t) \in \Omega\times\Omega\times(0,T] \tag{21}$$

$$P(x,s,t) = P(s,\xi,t) = 0 \tag{22}$$

$$P(x,\xi,o) = \delta(x - \xi) \tag{23}$$

(b) $P(\cdot,\cdot,t) \in H^2(\Omega\times\Omega)$, $\frac{\partial P}{\partial t}(\cdot,\cdot,t) \in L^2(\Omega\times\Omega)$, $t \in (0,T]$,

($H^2(\Omega\times\Omega)$ is the second Sobolev space) then:

(i) There exists one and only one $\hat{y}(\cdot,\cdot) \in L^2(Q)$ such that

$$y(x,T;\underline{u}) = \hat{y}(x,T) \tag{24}$$

where $\hat{y}(x,t)$ is the unique solution of the following linear integral equation of the second kind

$$\int_\Omega P(x,\xi,t)\{ \frac{\partial \hat{y}(\xi,t)}{\partial t} + A[\hat{y}(\xi,t)] - f(\xi,t)\}d\xi = z(x,t) - \hat{y}(x,t) \tag{25}$$

The conditions satisfied by $\hat{y}(x,t)$ on the closure of Q are:

$$\left.\begin{aligned} \hat{y}(s,t) &= z_1(s,t) \qquad (s,t) \in \Sigma \\ \hat{y}(x,o) &= z_2(x) \qquad x \in \Omega \end{aligned}\right\} \tag{26}$$

For a proof of this theorem, see Phillipson.[3] The numerical solution of equations (21) through (26) is not trivial. However an approximate solution is possible, using an eigenvalue expansion.[3] With the definitions

$$P(x,\xi,t) \overset{\text{l.i.m}}{=} P_m(x,\xi,t) \; ; \; P_m(x,\xi,t) = \sum_{i,j=1}^{m} P_{ij}(t)\, w_i(x)\, w_j(\xi)$$

$$\hat{y}(x,t) \overset{\text{l.i.m}}{=} \hat{y}_m(x,t) \; ; \; \hat{y}_m(x,t) = \sum_{i=1}^{m} \hat{y}_i(t)\, w_i(x)$$

where $w_i(x)$ and $w_j(x)$ satisfy (1), it can be shown[3] that (21) through (26) yield the familiar "lumped" results:

$$\frac{d}{dt}(P^{-1}(t)) + P^{-1}(t)A + AP^{-1}(t) + P^{-1}(t)\, IP^{-1}(t) - WW^T = 0 \tag{27}$$

$$P^{-1}(0) = I \tag{28}$$

$$\frac{d\hat{y}(t)}{dt} + A\hat{y}(t) - f(t) + z_1(t) = P^{-1}(t)[z(t) - \hat{y}(t)] \tag{29}$$

$$\hat{y}(0) = z_2 \tag{30}$$

where

$P(t) = \{P_{ij}(t)\}_{i,j=1,2..m}$

$\hat{y}(t) = \{\hat{y}_i(t)\}_{i=1,2..m}$

$A = \text{diag}\{\lambda_i\}_{i=1,2..m}$

$$WW^T = \int_\Gamma \frac{\partial w}{\partial \nu^*} \frac{\partial w^T}{\partial \nu^*} ds \; ; \; \frac{\partial w}{\partial \nu^*} = \left\{ \frac{\partial w_i}{\partial \nu^*} \right\}_{i=1,2..m}$$

$$z_1(t) = \{z_{1i}(t)\}_{i=1,2..m} \; ; \; z_{1i} = \int_\Gamma z_1(s,t) \frac{\partial w_i}{\partial \nu^*} ds$$

$$z_2 = \{z_{2i}\}_{i=1,2..m} \; ; \; z_{2i} = \int_\Omega z_2(x)\, w_i(x)\, dx$$

$$f(t) = \{f_i(t)\}_{i=1,2..m} \; ; \; f_i(t) = \int_\Omega f(x,t)\, w_i(x)\, dx \; .$$

In Section G, we report results using the suggested decoupling and subsequent approximation, for a simulated example. There, we also give results pertaining to the "discrete measurement" case induced by the functional (9).

F.2 A Direct Variational Method

To recapitulate, the problem is to select $\underline{u} \in V$ such that

$$J(\underline{u}) = \operatorname*{Inf}_{\underline{v} \in V} J(\underline{v}) \; . \tag{31}$$

As we have seen, there is a unique $\underline{u} \in V$ with the property (31) and it is characterized by

$$a(\underline{u},\underline{v}) - \ell(\underline{v}) = 0 \tag{32}$$

Equation (32) is the derivative of the functional $J(\underline{v})$ evaluated at $\underline{u}$. In terms of the gradient $\underline{G}(\underline{u})$, (32) is equivalent to:

$$(\underline{G}(\underline{u}),\underline{v})_V = 0$$

with $G(\underline{u})$ given by (17). The direct method for determining $\underline{u}$ involves searching on the quadratic surface $J(\underline{v})$ along directions $\underline{s}^k(\underline{G})$ which lead eventually to $\underline{u}$. That is,

$$\underline{u}^{k+1} = \underline{u}^k + \alpha^k \underline{s}^k(\underline{G}) \tag{33}$$

Because of the demonstrated efficiency of conjugate directions of search, we shall employ them here. The algorithm is as follows:

(i) Select $\underline{u}^o \in V$ (Initial guess)

(ii) Evaluate $\underline{G}(\underline{u}^o)$ via (17). If $G(\underline{u}^o) = 0$, then by (32), $\underline{u}^o$ is the solution. If $\underline{G}(\underline{u}^o) \neq 0$, then for the (i+1)st iteration, (i=0,1... ,) proceed as follows:

(iii) $\underline{u}^{i+1} = \underline{u}^i + \alpha^i \underline{s}^i$; $\underline{s}^o = -\underline{G}(\underline{u}^o)$.

$$\underline{s}^{i+1} = -\underline{G}(\underline{u}^{i+1}) + \beta^i \underline{s}^i$$

$$\beta^i = \frac{(\underline{G}(\underline{u}^{i+1}), \underline{G}(\underline{u}^{i+1}))_V}{(\underline{G}(\underline{u}^i), \underline{G}(\underline{u}^i))_V}$$

in addition, α^i is chosen so that

$$J(\underline{u}^{i+1}) = \operatorname*{Inf}_{\gamma^i \in R^1} J(\underline{u}^i + \gamma^i \underline{s}^i)$$

It is possible to obtain an explicit expression for α^i:

$$\alpha^i = -\frac{a(\underline{s}^i,\underline{u}^i) - \ell(\underline{s}^i)}{a(\underline{s}^i,\underline{s}^i)} = \frac{(\underline{G}(\underline{u}^i),\underline{G}(\underline{u}^i))_V}{a(\underline{s}^i,\underline{s}^i)} .$$

We review some properties of the algorithm in the following theorems:

<u>Theorem 3</u> If $\underline{G}(\underline{u}^i) \neq 0$, $J(\underline{u}^{i+1}) < J(\underline{u}^i)$

Corollary: The sequence of real numbers $J(\underline{u}^i)$ is monotone decreasing and has a limit in the extended reals:

$$\lim_{i \to \infty} J(\underline{u}^i) = J_\infty = \operatorname*{Inf}_{\underline{v} \in V} J(\underline{v}) .$$

<u>Theorem 4</u> The sequence $\{\underline{u}^i\}$ converges weakly to a unique $\underline{u} \in V$ and the limit $\underline{u}$ has the property that

$$J(\underline{u}) = \operatorname*{Inf}_{\underline{v} \in V} J(\underline{v})$$

that is,

$$\lim_{i \to \infty} \underline{u}^i \xrightarrow{\text{weakly}} \underline{u} \in V \ \text{(unique)}$$

$$\text{and } J(\underline{u}) = J_\infty .$$

We remark that at each iteration, the evaluation of $\underline{G}(\underline{u}^i)$ involves the numerical solution of (2) forwards in time, then (16) backwards in time, which are the (numerically) stable directions of solution.

As before, the numerical solution requires an approximation of the solutions to a set of partial differential equations. Again, we use the eigenfunctions of (1) to achieve this approximation. We discuss the results in the next section.

G. Numerical Results

The system chosen was the following:

$$\frac{\partial y(x,t)}{\partial t} - \frac{\partial^2 y(x,t)}{\partial x^2} = 212 \qquad (x,t) \in (0,1)x(0,T] \qquad (34)$$

$$y(o,t) = u_1(o,t) \qquad t \in (0,T]$$
$$y(1,t) = u_1(1,t) \qquad t \in (0,T]$$
$$y(x,o) = u_2(x) \qquad x \in (0,1)$$

Input Measurements

$$z_1(o,t) = u_1^*(o,t) + k_1 N_1(t) \; ; \; u_1^*(o,t) = 70 + 10 \sin 2\pi t$$
$$z_1(1,t) = u_1^*(1,t) + k_2 N_2(t) \; ; \; u_2^*(1,t) = 54.5$$
$$z_2(x) = u_2^*(x) + k_3 \; ; \; u_2^* = 70\, e^{-0.25x}$$

$k_1 = 4.2$, $k_2 = 0$, $k_3 = 2.8$

$N_i(t)$ purely random function with amplitude ± 1.0.

Output Measurements

(a) $z(x,t) = y(x,t;\underline{u}^*) + k_o N_o(t)$

(b) $z(x^i,t) = y(x^i,t;\underline{u}^*) + k_o^i N_o^i(t)$; $(x^i = 0.2, 0.4, 0.6, 0.8)$.

$k_o^i = 8.0$, N_o, N_o^i are random telegraph signals with amplitude ± 1.0.

The eigenfunctions appropriate to the suggested approximations are those given by $(1)^*$.

The results obtained for our example using the method of Section F.1 are shown in Figures 1, 2 and 3. An eight term expansion was adopted, and several selected variables are shown. It should be stated that the integration step size necessary to obtain a numerically stable solution for the $\{P_{ij}(t)\}$ was small (0.001) and this resulted in a large computational effort. The total time for solution was of the order of three minutes. On the other hand, using the direct method of Section F.2, three iterations, (sufficient to recover $\underline{u}$ such that $G(\underline{u}) \approx 0$) were accomplished in only 75 seconds. Some selected results are shown in

Figures 4, 5, 6, 7, and 8. Note that $J(\underline{v})$ is minimized rapidly (Figure 9) and observe that $J(\underline{v})_{\text{measurement (a)}} < J(\underline{v})_{\text{measurement (b)}}$.

We remark that $G(\underline{u}^i)$ given by (17) was approximated by $G_m(\underline{u}^i)$:

$$G_m(\underline{u}^i) = - \sum_{i=1}^{4} p_i(t) \frac{\partial w_i(s)}{\partial \nu^*} + u_1^i(s,t) - z_1(s,t)$$

$$\sum_{i=1}^{4} p_i(o)\, w_i(x) + u_2^i(x) - z_2(x) \quad .$$

We observed that for the example chosen, the last three terms in the summation were identically zero, that is,

$$G_m(\underline{u}) = G(\underline{u})$$

However in general, it is not clear in what sense $G_m(\underline{u}) \to G(\underline{u})$, and we are attempting to establish an appropriate result.

H. Summary and Conclusions

A special variational phrasing of a distributed identification problem resulted in a framework in which solutions were characterized using the theory of variational inequalities. Numerical techniques were suggested for recovering extremals to the variational problem, one of which, the direct method, yielded promising results. This direct method is also applicable to the problem of determining optimal boundary controls for certain distributed optimal control problems.

References

1. Lions, J.L., Stampacchia, G., "Variational Inequalities", to appear in the Comm. Pure and Applied Math. Vol. XX, No.3, August 1967, 493-519.
2. Balakrishnan, A.V., Lions, J.L., "State Estimation for Infinite-Dimensional Systems", J. Comp. and System Sciences, Vol. 1, (1967) pp. 391-403.
3. Phillipson, G.A., "The State Identification of a Class of Distributed Systems", Ph.D. Thesis, Case Western Reserve University, June 1967.
4. Lions, J.L., Magenes, E., "Remarques sur les Problemes aux Limites pour Operateurs Paraboliques", C.R. Acad. Sc., Paris, Vol. 25, (1960) pp. 2118-2120.

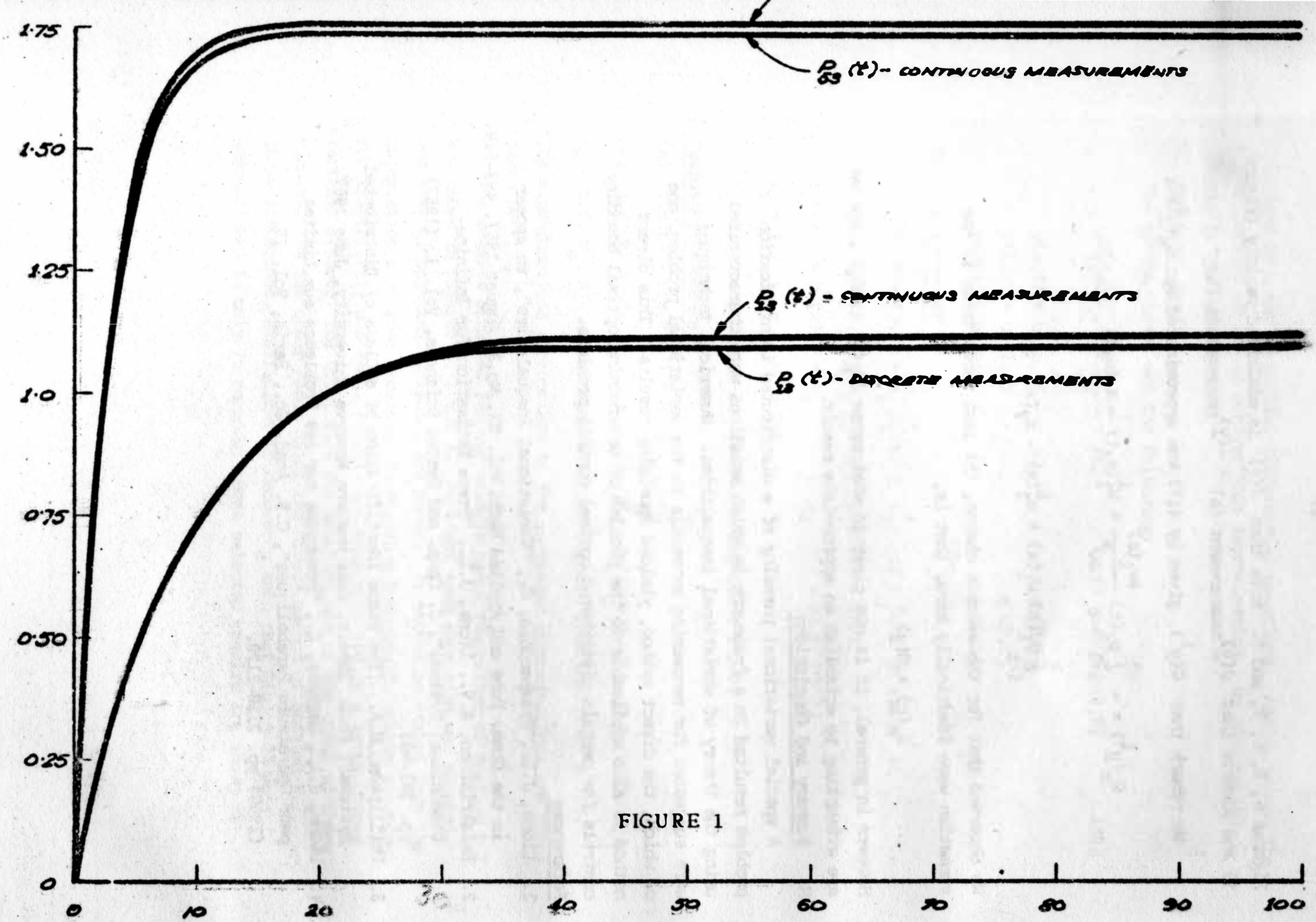
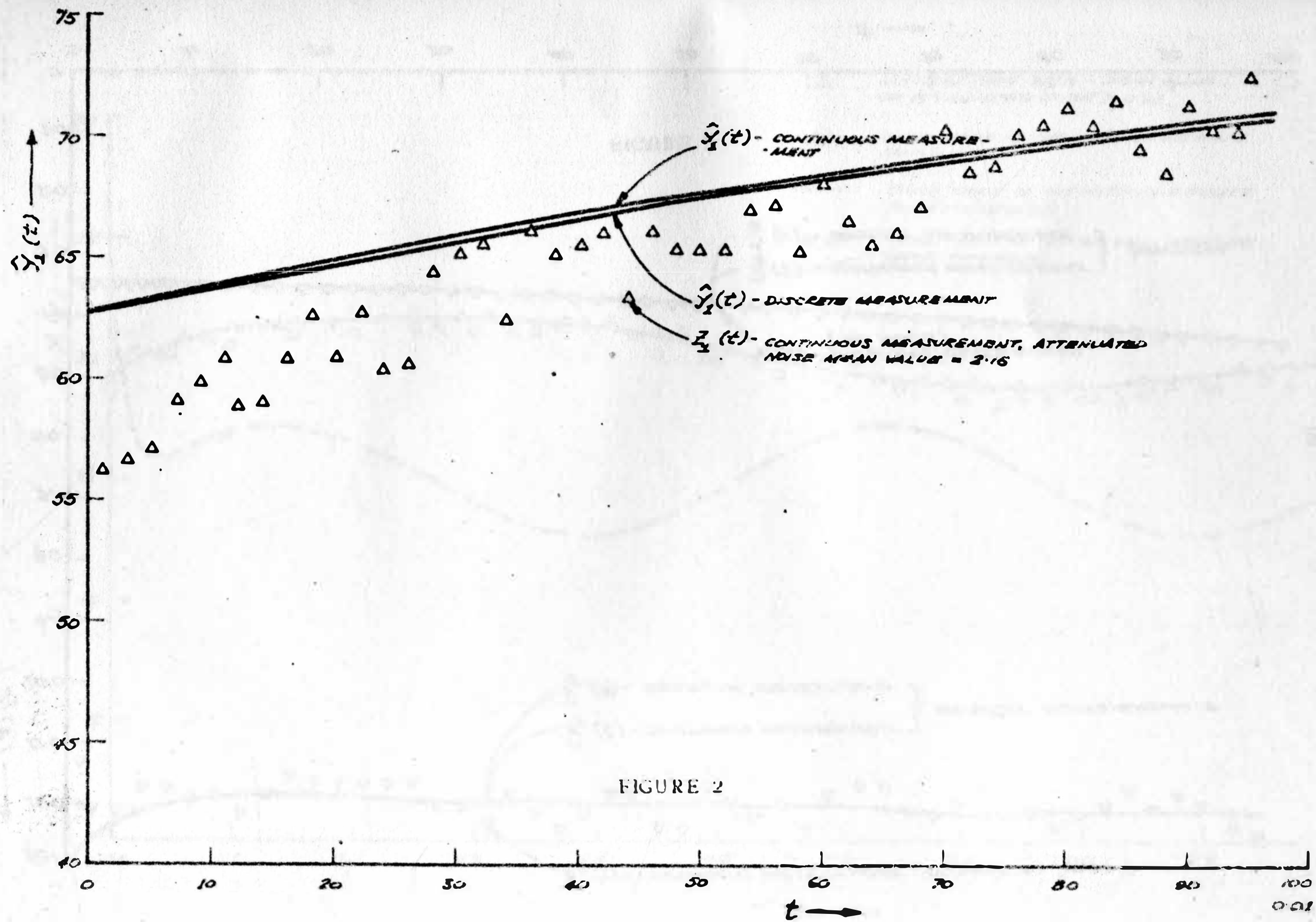

(t)- CONTINOUS MEASUREMENTS
(t) - CONTINUOUS MEASUREMENTS
(t)- DISCRETE MEASUREMENTS
1·75
1·50
1·25
1·0
0·75
0·50
0·25
0
10
20
40
50
60
70
80
90
100

FIGURE 1



$\hat{y}_1(t)$
75
70
65
60
55
50
45
40
0
10
20
30
40
50
60
70
80
90
100
0·01
t
$\hat{y}_1(t)$ - CONTINUOUS MEASURE- -MENT
$\hat{y}_1(t)$ - DISCRETE MEASUREMENT
$\bar{z}_1(t)$ - CONTINUOUS MEASUREMENT, ATTENUATED NOISE MEAN VALUE = 2·16

FIGURE 2

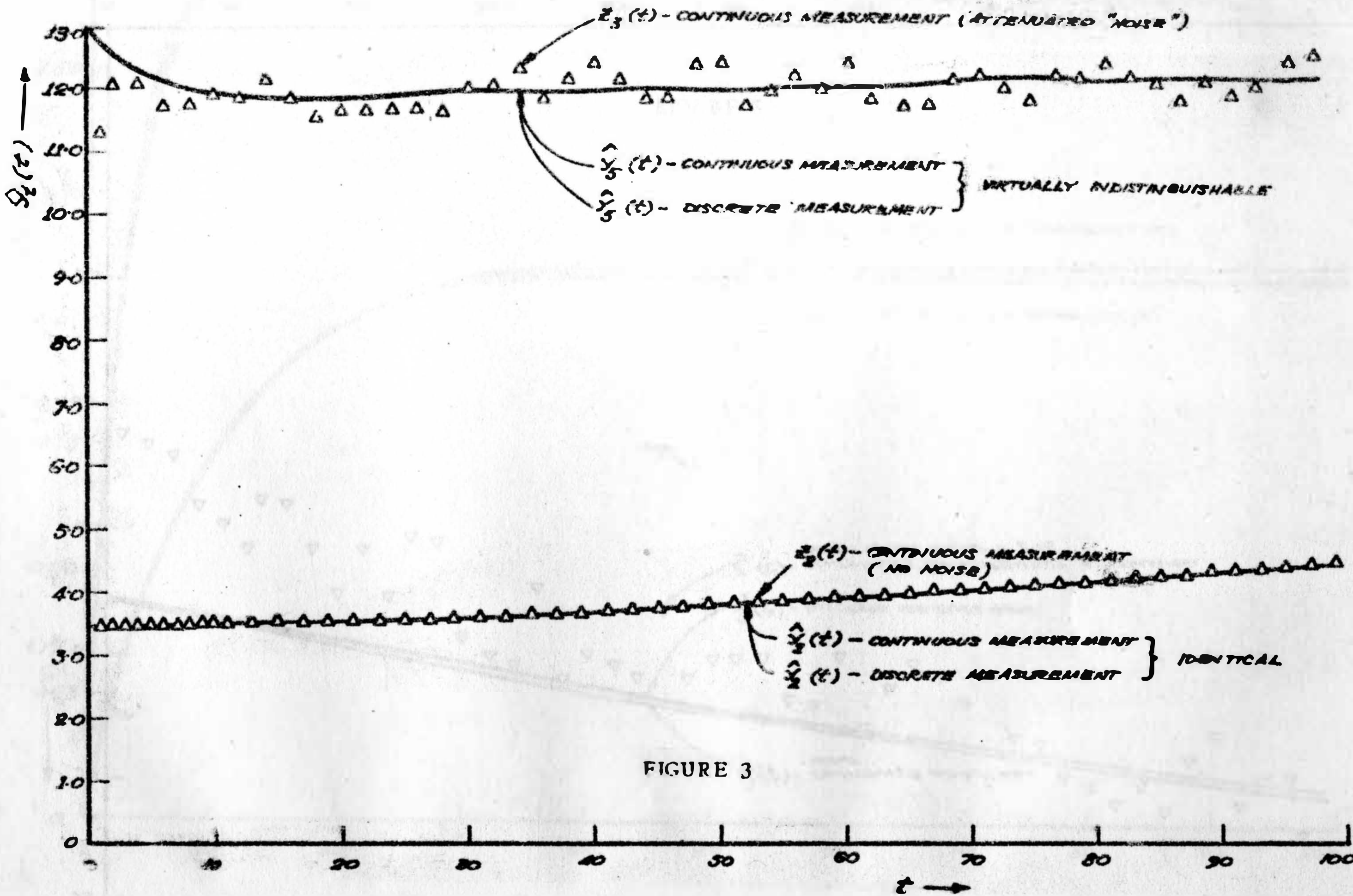

$z_3(t)$ - CONTINUOUS MEASUREMENT (ATTENUATED "NOISE")
$\hat{y}_5(t)$ - CONTINUOUS MEASUREMENT
$\hat{y}_5(t)$ - DISCRETE MEASUREMENT
VIRTUALLY INDISTINGUISHABLE
$z_2(t)$ - CONTINUOUS MEASUREMENT (NO NOISE)
$\hat{y}_2(t)$ - CONTINUOUS MEASUREMENT
$\hat{y}_2(t)$ - DISCRETE MEASUREMENT
IDENTICAL
t
13.0
12.0
11.0
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0
10
20
30
40
50
60
70
80
90
100

FIGURE 3

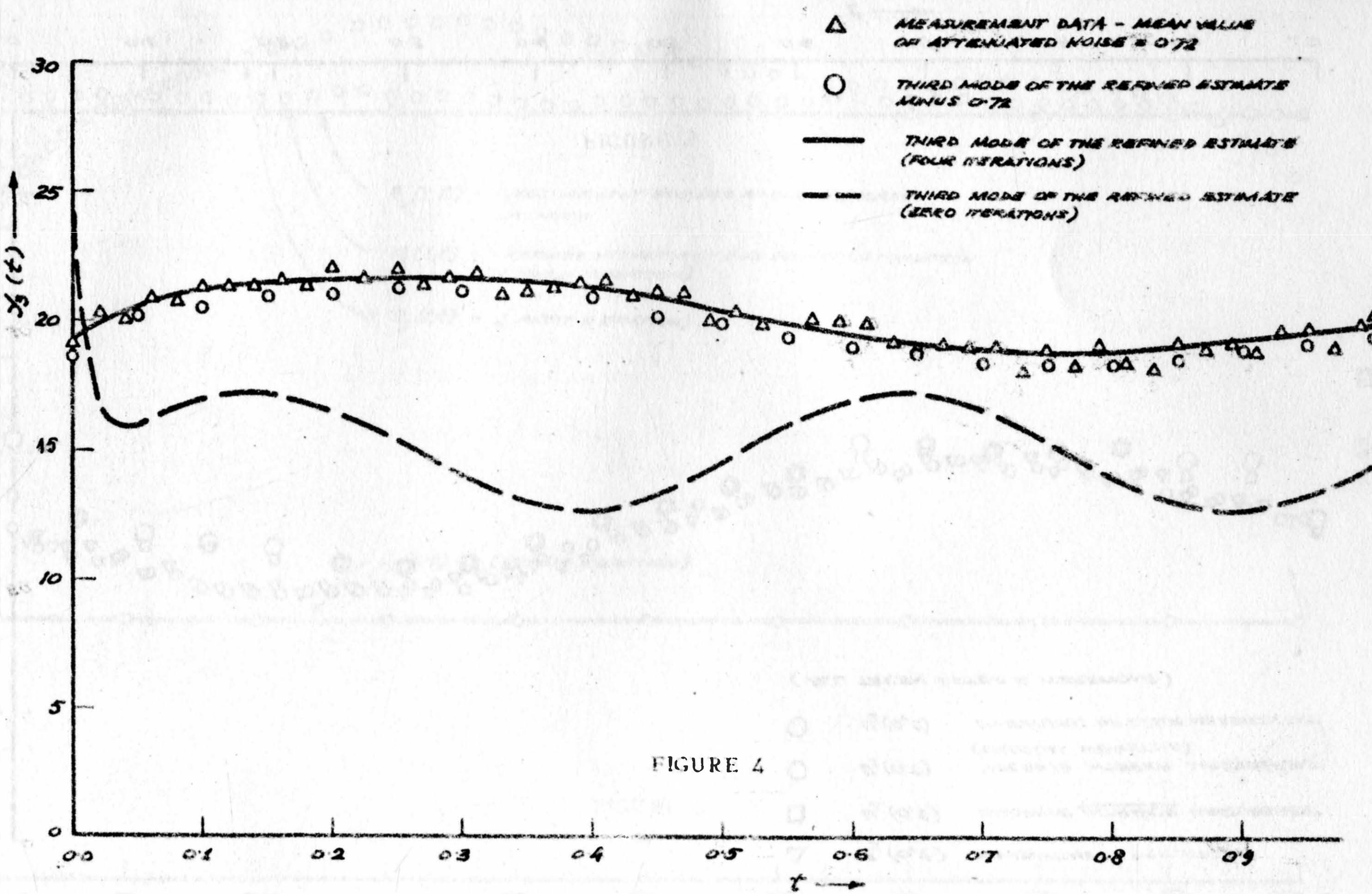
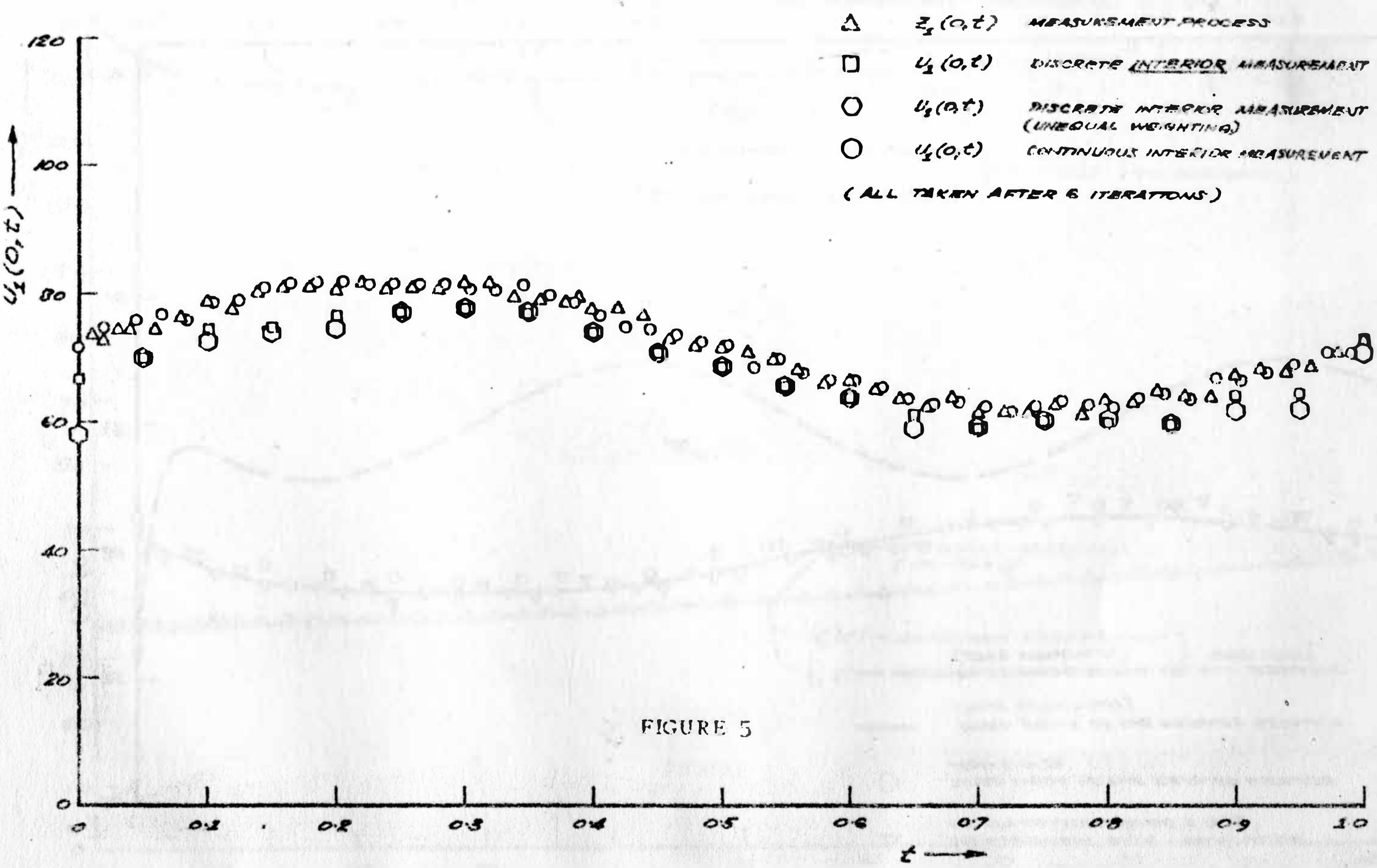
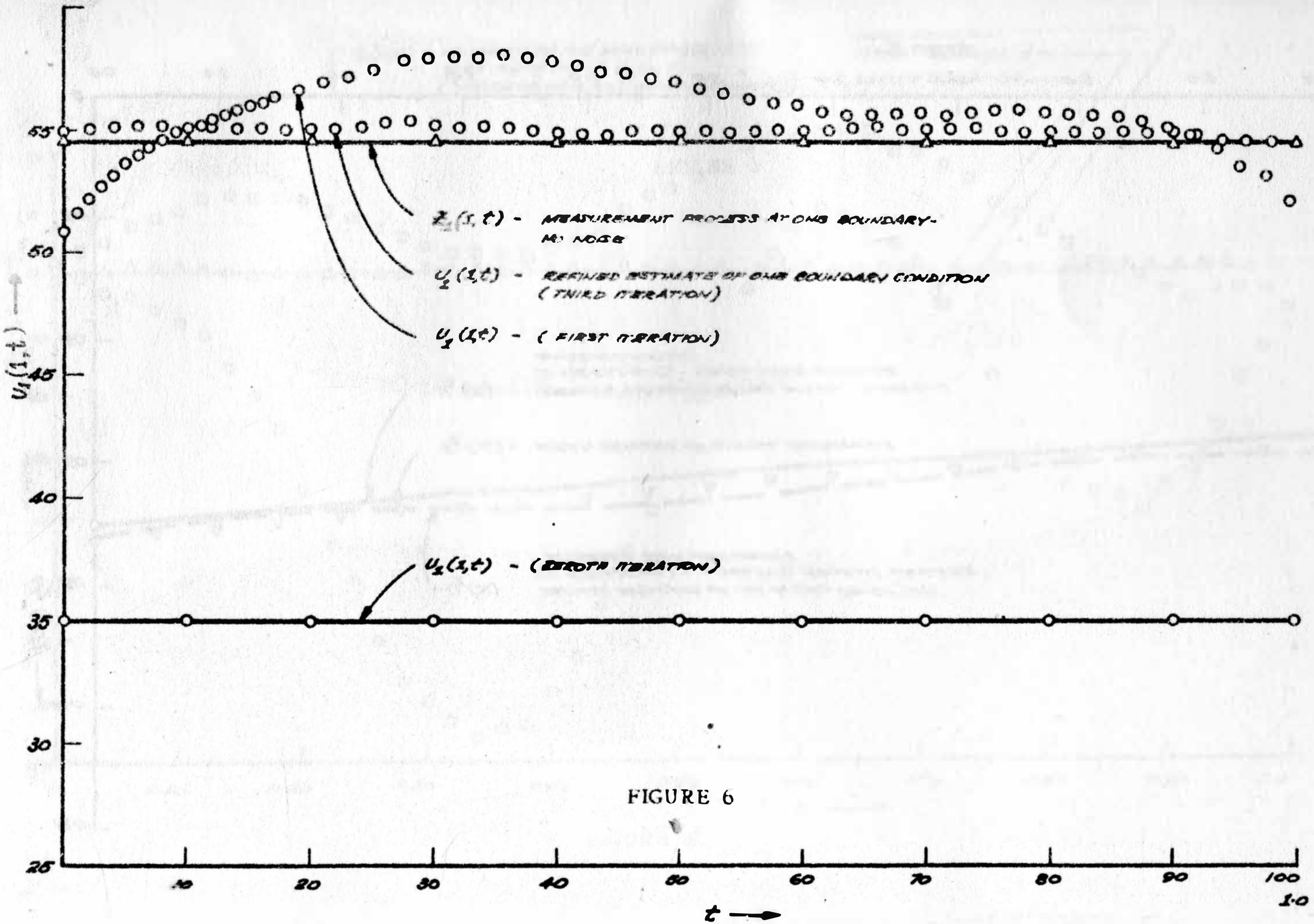

MEASUREMENT DATA - MEAN VALUE OF ATTENUATED NOISE = 0·72
THIRD MODE OF THE REFINED ESTIMATE MINUS 0·72
THIRD MODE OF THE REFINED ESTIMATE (FOUR ITERATIONS)
THIRD MODE OF THE REFINED ESTIMATE (ZERO ITERATIONS)
$y_3(t)$
t
30
25
20
15
10
5
0
0·0
0·1
0·2
0·3
0·4
0·5
0·6
0·7
0·8
0·9

FIGURE 4



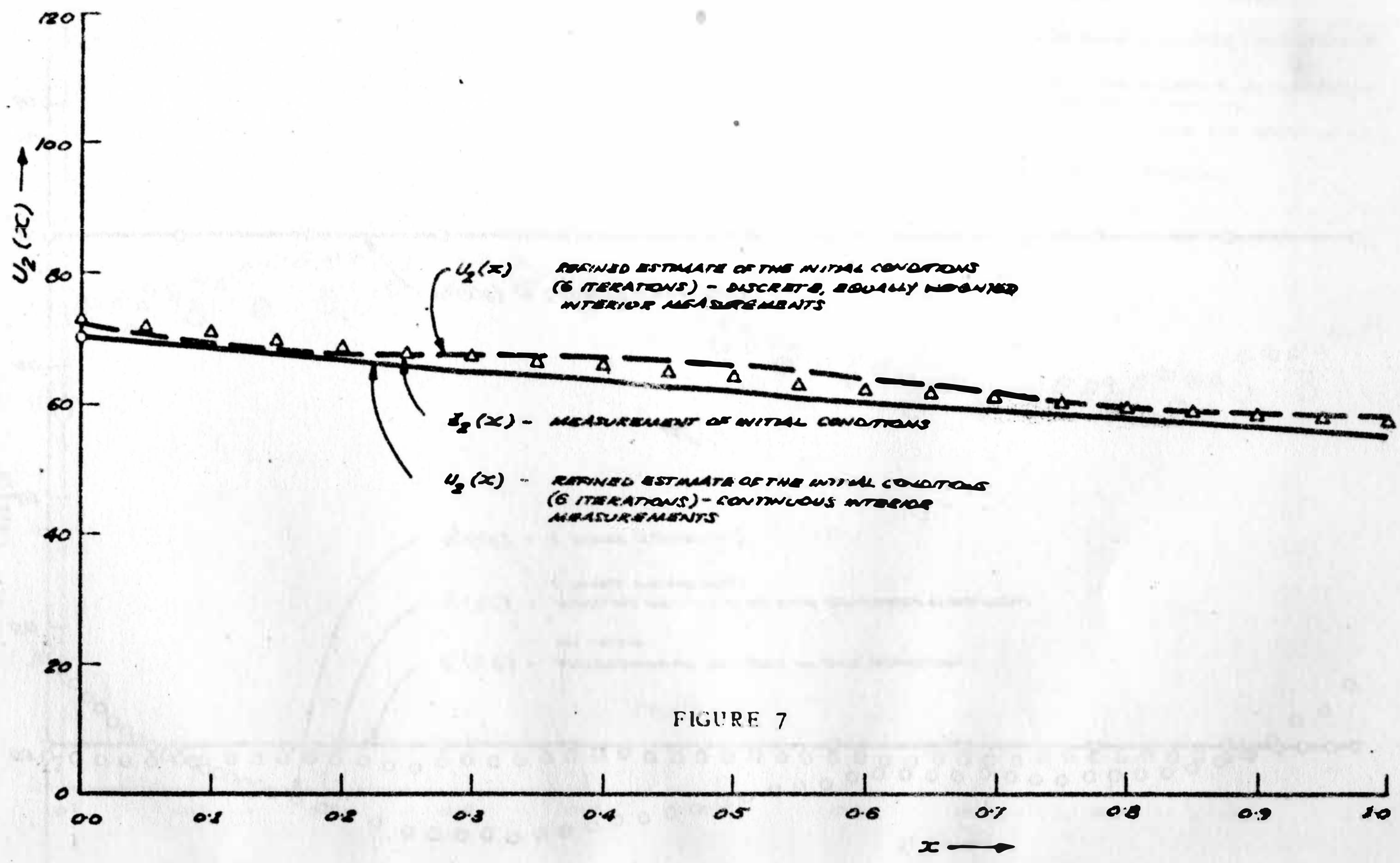
$z_1(0,t)$ MEASUREMENT PROCESS
$u_1(0,t)$ DISCRETE INTERIOR MEASUREMENT
$u_1(0,t)$ DISCRETE INTERIOR MEASUREMENT (UNEQUAL WEIGHTING)
$u_1(0,t)$ CONTINUOUS INTERIOR MEASUREMENT
(ALL TAKEN AFTER 6 ITERATIONS)
$u_1(0,t)$
120
100
80
60
40
20
0
0·1
0·2
0·3
0·4
0·5
0·6
0·7
0·8
0·9
1·0
t

FIGURE 5



Z1(1, t) - MEASUREMENT PROCESS AT ONE BOUNDARY-
NO NOISE
U1(1,t) - REFINED ESTIMATE OF ONE BOUNDARY CONDITION
(THIRD ITERATION)
U1(1,t) - (FIRST ITERATION)
U1(1,t) - (ZEROTH ITERATION)
U1(1, t) →
55
50
45
40
35
30
25
10
20
30
40
50
60
70
80
90
100
1.0
t →

FIGURE 6



120
100
80
60
40
20
0
$U_2(x)$ →
0·0
0·1
0·2
0·3
0·4
0·5
0·6
0·7
0·8
0·9
1·0
x →
$U_2(x)$ REFINED ESTIMATE OF THE INITIAL CONDITIONS (6 ITERATIONS) - DISCRETE, EQUALLY WEIGHTED INTERIOR MEASUREMENTS
$Z_2(x)$ - MEASUREMENT OF INITIAL CONDITIONS
$U_2(x)$ - REFINED ESTIMATE OF THE INITIAL CONDITIONS (6 ITERATIONS) - CONTINUOUS INTERIOR MEASUREMENTS

FIGURE 7

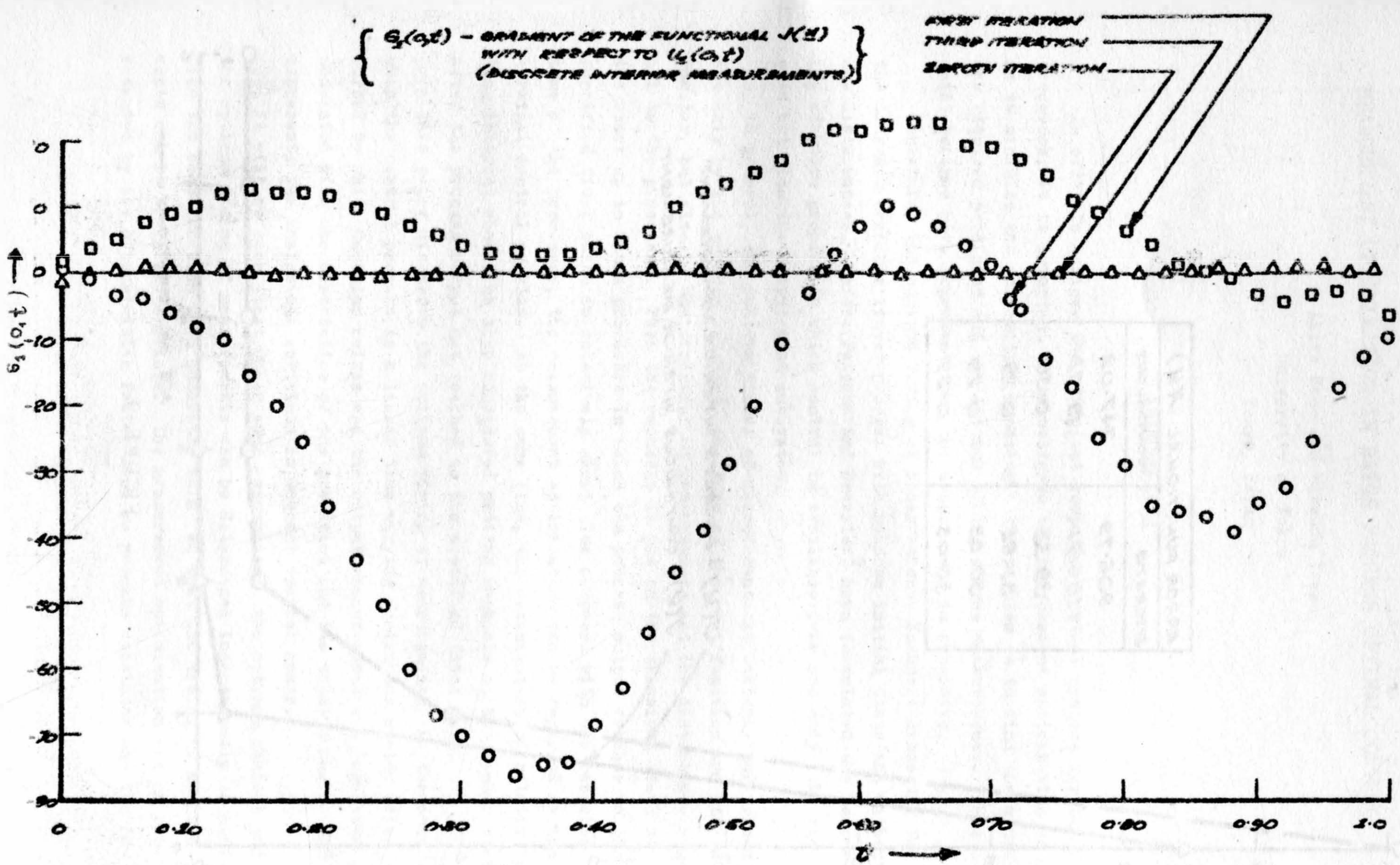
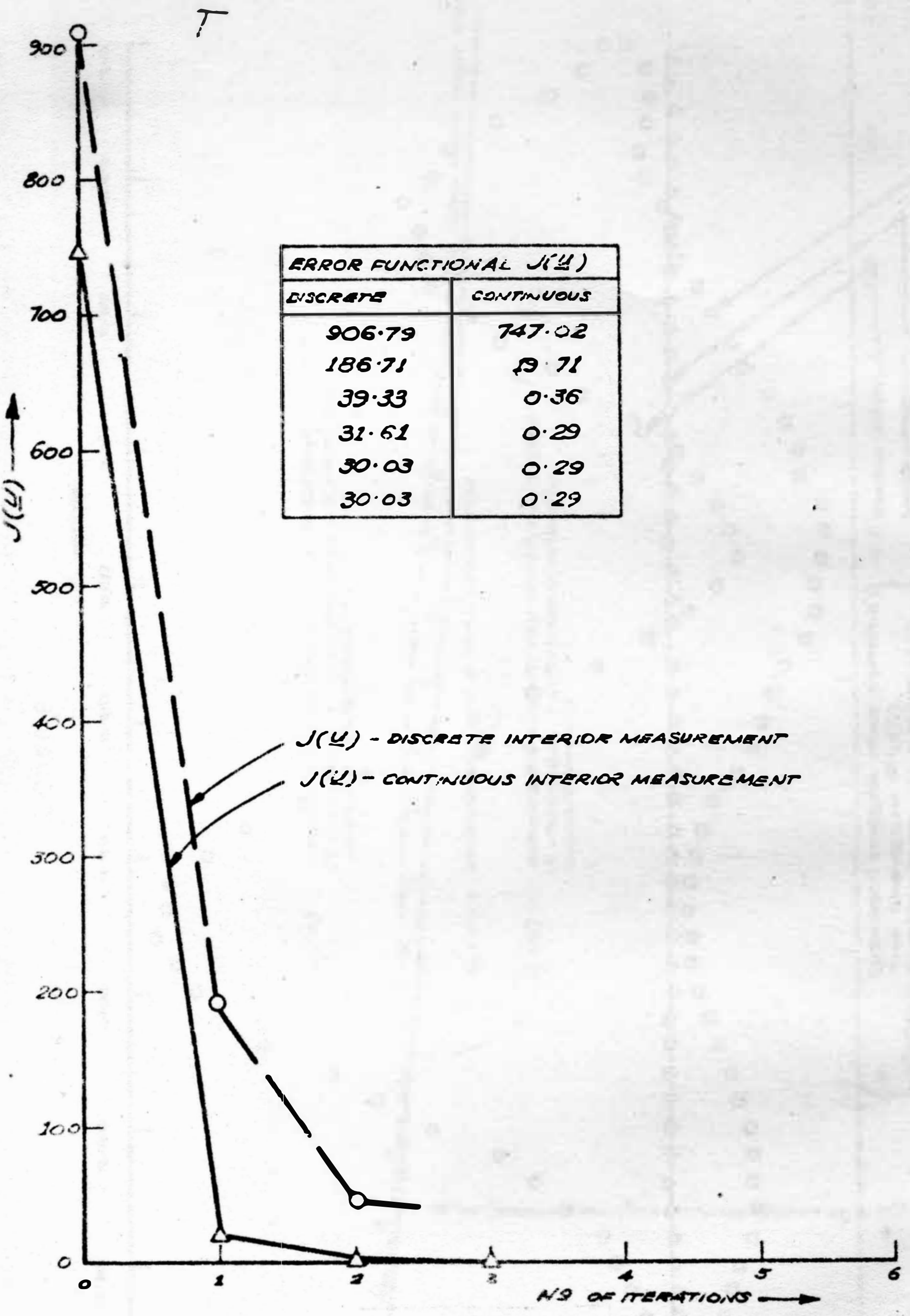

$G_3(0,t)$ - GRADIENT OF THE FUNCTIONAL $J(U)$ WITH RESPECT TO $U_2(0,t)$ (DISCRETE INTERIOR MEASUREMENTS)
FIRST ITERATION
THIRD ITERATION
EIGHTH ITERATION
$G_3(0,t)$
t

FIGURE 8



ERROR FUNCTIONAL J(U)	
DISCRETE	CONTINUOUS
906·79	747·02
186·71	19·71
39·33	0·36
31·61	0·29
30·03	0·29
30·03	0·29

FIGURE 9

MEASUREMENT OPTIMIZATION IN BATCH PROCESS OPTIMAL CONTROL

Akira Sano and Mitsuru Terao
University of Tokyo
Tokyo, Japan

INTRODUCTION

In order to achieve the efficient computing control systems for batch processes, it is important to investigate the measurement optimization so as to minimize the information handling. The problem is to find the minimum required precision and the optimum timing of the measurements. This point becomes more important when the measurement is expensive.

Concerning the optimum timing of observations, Kushner[1] discussed for the discrete unstable linear process with unknown initial state and without any disturbances. Meier, Peschon and Dressler[2] have formulated the measuring adaptive problem in which control is available over not only the process but also the measurement subsystem.

In general, the optimum timing of measurements is strongly influenced not only by the process constant, the form of the cost function and its parameter, but also by the precision of measurements and the disturbance signal to the processes. When the emphasis is put on the transient state of the start up in batch processes in which the process output reaches to the specified final value in relatively short time compared with the response time of the processes, the measurement optimization can be regarded as the terminal control problem. At the same time, the optimization of relatively long interval should be also considered putting emphasis on the steady state after the process output has reached to the specified final value.

This paper discusses the optimum timing of measurements in the case when the control process is a linear time-varying continuous system disturbed by white Gaussian noises and the output measurements are impulsive modulated with the restriction of the precision and the total number, and represents the analytical solutions in several special cases.

It will be seen that under some assumptions the optimum control and the optimum timing of measurements can be determined independently to minimize the cost function with quadratic form so that the timing of measurements can be specified a priori. The measurement optimization can be represented by the deterministic facter such as process constants and the parameters of cost functions as well as by the statistical factors which are the ratio of the variance of measurement error and the initial uncertainty to the variance of external disturbances.

The two types of cost function are considered. The one consists of quadratic in the control and terminal error, and the other consists of quadratic in the control and the steady state error. The steady state optimization is especially required in the direct digital control. The minimum required information quantity of measurements can be reduced if we allow some degradation of the performance (cost function). A computational result presented in the final section shows that there is a compromise between the number and the precision of measurements for various near optimum performances.

PROBLEM STATEMENT

The dynamics of batch processes is generally described by the following linear time-varying noise perturbed differential equation

$$\dot{x}(t)=A(t)x(t)+B(t)u(t)+v(t) \tag{1}$$

where $x(t)$ is an n-state vector, $u(t)$ is an r-control vector and $v(t)$ is n-white Gaussian noise vector with zero mean and covariance

$$Ev(t)v'(t')=V(t)\delta(t-t'). \tag{2}$$

The state of the process is estimated from a set of noise corrupted impulsive measurements given by

$$y(t)=[M(t)x(t)+w(t)]\,\delta(t-t_i), \qquad i=1,2,\cdots,k \tag{3}$$

where $y(t)$ is an m-measurement vector and $w(t)$ is the white Gaussian measurement error with zero mean and covariance

$$Ew(t)w'(t')=W(t)\delta(t-t'). \tag{4}$$

The uncertainty of the initial state is normally distributed with mean value x_0 and covariance matrix P_0. For simplicity, the random variables $x(0)$, $v(t)$ and $w(t)$ are independent of each others.

The number and the precision of possible measurements are usually limited because of physical or economical constraints so that it is important to find the optimum timing of the k measurement pulses which minimizes the cost function with the following quadratic form

$$J=\frac{1}{2}E\left[x'(T)Fx(T)+\int_0^T\{x'(t)Q(t)x(t)+u'(t)R(t)u(t)\}dt\right] \tag{5}$$

where F and $Q(t)$ are $n\times n$ nonnegative definite matrices and $R(t)$ is an $r\times r$ positive definite matrix.

EVALUATION OF THE OPTIMUM COST PERFORMANCE

The minimum value of the cost function can be evaluated by the use of the linear estimation theory and stochastic optimization technique.

The problem of minimizing the cost function (5) subject to (1), (2), (3) and (4) may be decoupled to the problem of determining the minimum variance linear estimate of $x(t)$ and the problem of stochastic optimum control $u^o(t)$, using the conditional expected value $\hat{x}(t)$ with knowledge of a set of measurements[3,4] where

$$\hat{x}(t)=E[x(t)\,|\,y(t_1),y(t_2),\cdots,y(t_i);\ t_i\le t]. \tag{6}$$

Therefore, the minimum cost of (5) can be derived by the choice of the optimum timing of measurements $\{t_i\}$ and the optimum control $u^o(t)$ as

$$J^o=\min_{\{t_i\}}\cdot\min_{u(t)} J=\min_{\{t_i\}}[J_1+\min_{u(t)} J_2] \tag{7}$$

where

$$J_1=\frac{1}{2}E[(x(T)-\hat{x}(T))'F(x(T)-\hat{x}(T))]+\frac{1}{2}E[\int_0^T(x(t)-\hat{x}(t))'Q(t)(x(t)-\hat{x}(t))dt], \tag{8}$$

$$J_2=\frac{1}{2}E[\hat{x}'(T)F\hat{x}(T)+\int_0^T\{\hat{x}'(t)Q(t)\hat{x}(t)+u'(t)R(t)u(t)\}dt]. \tag{9}$$

The J_1 of (8), that is the estimation error, is rewritten by denoting covariance matrix of estimation error with $P(t)=E(x(t)-\hat{x}(t))(x(t)-\hat{x}(t))'$ as

$$J_1=\frac{1}{2}tr[FP(T)]+\frac{1}{2}\int_0^T tr[Q(t)P(t)]dt. \tag{10}$$

The conditional expected value $\hat{x}(t)$ and its covariance $P(t)$ are derived by the following equations using Kalman's linear estimation theory.[5,6,7] If there is no measurement in an interval $[t_{i-1}^+,\ t_i^-]$ where t_{i-1}^+ and t_i^- denote the instant after the observation at t_{i-1} and before the observation at t_i respectively, then

$$\dot{\hat{x}}(t)=A(t)\hat{x}(t)+B(t)u(t),\qquad t_{i-1}^+\le t\le t_i^-, \tag{11}$$

$$\dot{P}(t)=A(t)P(t)+P(t)A'(t)+V(t),\qquad t_{i-1}^+\le t\le t_i^- \tag{12}$$

and if there is one measurement and correction about $\hat{x}(t)$ and $P(t)$ is performed, then

$$\hat{x}(t_i^+)=\hat{x}(t_i^-)+K(t_i)[y(t_i)-\hat{y}(t_i^-)]\qquad t=t_i \tag{13}$$

where $\hat{y}(t_i^-)=M(t_i)\hat{x}(t_i^-)$, and

$$P(t_i^+)=P(t_i^-)-K(t_i)M(t_i)P(t_i^-), \qquad t=t_i \tag{14}$$

where

$$K(t_i)=P(t_i^-)M'(t_i)[M(t_i)P(t_i^-)M'(t_i)+W(t_i)]^{-1}. \tag{15}$$

Combining (11) with (13) for $\hat{x}(t)$ and (12) with (14) for $P(t)$

$$\dot{\hat{x}}(t)=A(t)\hat{x}(t)+B(t)u(t)+\sum_{i=1}^{t_i\le t} z(t_i)\delta(t-t_i), \tag{16}$$

$$\dot{P}(t)=A(t)P(t)+P(t)A'(t)-\sum_{i=1}^{t_i\le t} K(t_i)M(t)P(t_i^-)\delta(t-t_i)+V(t) \tag{17}$$

where the third term in (16) is easily shown to be a white random variable with zero mean and covariance

$$Ez(t_i)z'(t_j)=Z(t_i)\delta_{ij}=P(t_i^-)M'(t_i)K'(t_i)\delta_{ij}. \tag{18}$$

Solving for $P(t)$ using the initial condition $P(0)=P_0$, we obtain the estimation error term (8).

On the other hand, we can obtain the optimum control $u^o(t)$ and the optimum performance $J_2^o(\hat{x}(0), 0)$, regarding (16) as the noise perturbed control process equation with the quadratic performance index (9). The dynamic programming approach in the stochastic optimal control[8,4] leads the Hamilton Jacobi's equation

$$-\frac{\partial J_2^o(\hat{x}(t),t)}{\partial t} = \min_{u(t)} \Big[\frac{1}{2}\sum_{i=1}^{t_i\le t}\sum_{j,l}^{n} z_{jl}(t_i)\frac{\partial^2 J_2^o}{\partial \hat{x}_j \partial \hat{x}_l} + \{A(t)\hat{x}(t)+B(t)u(t)\}'\frac{\partial J_2^o}{\partial \hat{x}}$$
$$+\frac{1}{2}\{\hat{x}'(t)Q(t)\hat{x}(t)+u'(t)R(t)u(t)\}\Big] \tag{19}$$

where $z_{jl}(t)$ is the j,l-element of the covariance matrix $Z(t)$ given by (18). Thus the optimum control is determined from (19) as

$$u^o(t)=-R^{-1}(t)B'(t)\frac{\partial J_2^o}{\partial \hat{x}(t)}. \tag{20}$$

The approach assuming the solution of the form

$$J_2^o(\hat{x}(t),t)=\frac{1}{2}\hat{x}'(t)S(t)\hat{x}(t)+\frac{1}{2}U(t) \tag{21}$$

yields the following differential equations for $S(t)$ and $U(t)$. Substituting (20) and (21) into (19), we have

$$\dot{S}(t)=-S(t)A(t)-A'(t)S(t)+S(t)B(t)R^{-1}(t)B'(t)S(t)-Q(t), \tag{22}$$

$$\dot{U}(t)=-\sum_{i=1}^{t_i\le t} tr[S(t)Z(t_i)]\delta(t-t_i) \tag{23}$$

with boundary condition $S(T)=F$ and $U(T)=0$.

Therefore, the optimum performance J_2^o is given by

$$J_2^o(\hat{x}(0),0)=\frac{1}{2}\hat{x}_0'S(0)\hat{x}_0+\frac{1}{2}\sum_{i=1}^{k} tr[K(t_i)M'(t)P(t_i^-)S(t)]\delta(t-t_i). \tag{24}$$

As a result, the original optimum performance (7) is reduced to the following, using (10) and (24) and omitting the term which does not affect

$\{t_i\}$,

$$J^o=\min_{\{t_i\}}\left[\frac{1}{2}tr[FP(T)]+\frac{1}{2}\int_0^T tr[Q(t)P(t)]\,dt+\frac{1}{2}\sum_{i=1}^{k}tr[K(t_i)M'(t)P(t_i^-)S(t_i)]\right]. \tag{25}$$

The first and the second terms denote the cost due to the state uncertainty caused by external disturbances and measurement error, and the third term denotes the cost of control corrections based on the measurements. The following sections illustrate several examples in a scalar case.

SOME ILLUSTRATIONS FOR TERMINAL CONTROL PROBLEM

As an example, the optimum timing of measurements is derived for a scalar linear system with a cost function that is quadratic in the control and terminal error. Let $A(t)=a$, $B(t)=M(t)=F=1$, $Q(t)=0$, $R(t)=r$, $V(t)=V$ and $W(t)=W$ in (1)$\sim$(5), then the minimum cost function of (25) in scalar cases is reduced to

$$J^o=\min_{\{t_i\}}\left[\frac{1}{2}P(T)+\frac{1}{2}\sum_{i=1}^{k}\frac{S(t)P^2(t_i^-)}{P(t_i^-)+W}\delta(t-t_i)\right] \tag{26}$$

where $S(t)$ is the solution of the Riccati's differential equation (22) given by

$$S(t)=\frac{2ar}{1+(2ar-1)e^{-2a(T-t)}}. \tag{27}$$

1. The case with no disturbances

The case in which there is only an initial uncertainty is discussed. Normalizing J^o of (26) by the variance of the initial uncertainty P_0 and manipulating $P(t_i^-)$ using (17), the optimum performance is given as

$$J^o/P_0=\min_{\{t_i\}}\left[\frac{\lambda e^{2aT}}{2(\sum_{j=1}^{k}e^{2at_j}+\lambda)}+\sum_{i=1}^{k}\frac{ar\,e^{4at_i}}{\{1+(2ar-1)e^{-2a(T-t)}\}(\sum_{j=1}^{i-1}e^{2at_j}+\lambda)(\sum_{j=1}^{i}e^{2at_j}+\lambda)}\right] \tag{28}$$

where $\lambda=W/P_0$ is the ratio of the variance of measurement error to the initial uncertainty. The cost performance (28) is to be minimized with respect to t_i, $i=1,2,\ldots,k$. If $k=1$, the case of a single measurement, the optimum timing may be easily derived as follows. If $a<0$ (stable process having time constant of $1/a$), the both first and second terms in (28) for $k=1$ increase as the t_1 increases, and so the optimum t_1 is always zero. If $a>0$ (unstable process having a pole in right half plane) the optimum t_1 is obtained by differentiating (28) for $k=1$ with respect to t_1. As a result,

$$t_1=0, \qquad a\le 0 \tag{29a}$$

$$t_1=T-\frac{1}{2a}\ln\left[\sqrt{2ar(1+e^{2aT}/\lambda)}+1\right], \quad 0\le t_1\le T, \qquad a>0. \tag{29b}$$

In Fig. 1, the optimum timing is shown against the process parameter a for various values of λ and r in the case of T=1. It is interesting to know that although the optimum timing for $a<0$ is always the initial time, the timing for $a>0$ becomes later and approaches to T/2 for large a, as indicated by Kushner[1]. Fig. 1 shows that the optimum timing becomes later as the precision of measurement and the weighting coefficient r decrease and it seems to be consistent with our intuition.

2. <u>The case with disturbances</u>

It is conceivable that the control processes are always perturbed by external disturbances. We discuss the following case when the measurements are quite precise or the disturbance is quite large compared with the initial uncertainty and the measurement error.

If k=1, then the optimum performance of (26) normalized by the variance of disturbances is given calculating $P(t_1^-)$ of (17) subject to disturbances and assuming that the measurement is quite precise, as

$$J^0/V = \min_{t_1}\left[\frac{-e^{2at_1}+e^{2aT}}{4ae^{2at_1}} + \frac{r\{(2a\eta+1)e^{2at_1}-1\}}{2\{1+(2ar-1)e^{-2a(T-t_1)}\}}\right] \qquad (30)$$

where $\eta = P_0/V$. Therefore the optimum timing of one measurement pulse may be uniquely determined differentiating (30) as

$$t_1 = T - \frac{1}{2a}\ln\left[\mp\sqrt{2ar\{2ar-1+(2a\eta+1)e^{2aT}\}} - (2ar-1)\right], \qquad a \lessgtr 0 \qquad (31)$$

The optimum timing obtained analytically by (31) is shown in Fig. 2 against the process parameter a for various weighting coefficient r and initial uncertainty η. It is seen that for large positive a (very unstable precess) the optimum timing approaches to the middle of the control interval and for large negative a (stable process having short time constant) it approaches to the terminal time T asymptotically. For an ordinary stable process, if the initial uncertainty is negligible, i.e., $\eta=0$, the timing is to be near the terminal time of the control interval as represented by the solid lines in Fig. 2, however, if the initial uncertainty is present, i.e., $\eta=1$, the timing becomes earlier as represented by the dashed lines in Fig.2. If the initial uncertainty η is quite large, the measurement is to be made at the initial instant of the interval.

When the measurement has arbitrary precision it is not so simple to derive the optimum timing analytically since there is not always only one local optimum. So we obtain the optimum t_1 for the case numerically. Fig.3 shows the optimum timing against process constant for several measurement grades of precision normalized by the variance of disturbance.

STEADY STATE OPTIMIZATION

In this section the cost function is assumed to consist of quadratic in the control and the steady state error. If the control interval $[0,T]$ is to be large enough compared with the process parameter a, the steady state optimization can be achieved. Let $A(t)=a$, $B(t)=M(t)=Q(t)=1$, $F=0$, $R(t)=r$, $V(t)=V$ and $W(t)=W$, then the optimu performance (25) is given by

$$J^o = \min_{\{t_i\}} \left[\frac{1}{2}\int_0^T P(t)dt + \frac{1}{2}\sum_{i=1}^{k} \frac{S(t)P^2(t_i^-)}{P(t_i^-)+W} \delta(t-t_i) \right] \tag{32}$$

where $S(t)$ is the sollution of the Riccati's differential equation,

$$S(t)=\frac{1-e^{-2\beta(T-t)}}{\beta-a+(\beta+a)e^{-2\beta(T-t)}}, \qquad \beta=\sqrt{\frac{1}{r}+a^2}. \tag{33}$$

If the control optimization interval T is considered to be large enough, then (33) becomes

$$S \doteq r(\beta+a) \tag{34}$$

1. The case with no disturbances

We usually requiire more than one measurements to achieve the steady state optimization. However, for some a nalytical investigation, it may be convenient that we consider the case of $k=1$. We can also obtain the optimum timing for arbitrary k by using the simple hill climbing method since there is only one local optimum in this case.

Normalizing J^o in (36) by P_o we obtain the following for $k=1$,

$$J^o/P_o = \min_{t_1} \left[\frac{e^{4at_1}-e^{2at_1}+\lambda(e^{2aT}-1)}{4a(e^{2at_1}+\lambda)} + \frac{r(\beta+a)e^{4at_1}}{2(e^{2at_1}+\lambda)} \right] \tag{35}$$

where $\lambda=W/P_o$. The optimum timing of t_1 may be obtained by differentiating (35) with repect to t_1 and the result is

$$t_1=0, \qquad a \leq 0, \tag{36a}$$

$$t_1=\frac{1}{2a}\ln\left[\sqrt{\lambda^2+\frac{\lambda e^{2aT}}{2ar(\beta+a)+1}}-\lambda\right], \qquad a>0. \tag{36b}$$

Fig. 4 shows the solution given by (36) against the process parameter $a>0$ for various λ and r in the case of $T=5$. When the penalty is paid for the steady state error more than for the control corrections, that is, the weighting coefficient r decreases, the optimum t_1 approaches

$$t_1=\frac{1}{2a}\ln\left[\sqrt{\lambda^2+\lambda e^{2aT}}-\lambda\right], \qquad a>0. \tag{37}$$

Fig. 4 shows that the optimum timing does not depend on r so much, but on the measurement precision λ. The more precise the measurement is (the smaller λ), the earlier the measurement should be made.

2. The case with external disturbances

The cost performance (32) subject to external disturbances is rewritten as

$$J^0/V = \min_{\{t_i\}} \left[\frac{1}{4a} \left\{ \sum_{i=0}^{k} \left(\varphi(t_i^+) + \frac{1}{2a}\right)\left(e^{2a(t_{i+1}-t_i)} - 1\right) - T \right\} + \frac{1}{2} \sum_{i=1}^{k} \frac{S\varphi^2(t_i^-)}{\varphi(t_i^-) + \xi} \right] \tag{38}$$

where $\varphi(t_i^+) = P(t_i^+)/V$, $\eta = P(0^+)/V$, and $\xi = W/V$.

For arbitrary η and ξ, it is too much elaborate to obtain the optimum $\{t_i\}$ analytically, so that we got numerical solutions by a digital computer. The results for k=1 are summarized in Fig.5 and for k=5 in Fig.6. For k=1, the optimum timing of the measurement is plotted for various intial uncertainty η and measurement precision ξ. When the controlled process is stable (a<0), the optimum timing of the measurement depends on the initial uncertainty rather sensitively and approaches the initial part of the control interval as η increases. On the other hand for the unstable process (a>0), the optimum timing approaches the middle of the interval for any η and ξ. In the steady state optimization, it is considered that there is not so much the effect of the initial uncertainly but of the external disturbances. In this case the optimum timing t_1 is obtained analytically as

$$t_1 = \frac{T}{2} - \frac{1}{4a} \ln\left[2ar(\beta + a) + 1\right] \tag{39}$$

and is approximately the middle of the interval. For arbitrary k it is recognized that k measurements are to be taken at almost uniform intervals as shown in Fig.6.

3. Compromise between the number and the precision of measurements

In the last section we confirmed that the timing of measurements in the steady state optimization subject to external disturbances has uniform intervals. In connection with the direct digital control for the process it is significant to reduce the total required information quantity of measurements as far as specified control performance is satisfied. The information quantity which connects with the number and the precision of measurements can be reduced remarkably if we allow some degradation of the control performance. Fig.7 and Fig.8 show the control performances for the specified number and precision of measurements in the case of T=5, for a=1 and a=-1 respectively. The performance of the near optimal systems is normalized by the ideal performance in which measurements are not constrained.

Fig.7 and Fig.8 give a compromise between the number k and the precision ξ of measurements for some specified performance. The J_{id} denotes the ideal performance which is obtained from (38) by $k \to \infty$ and $1/\xi \to 0$. It is noticeable that more information quantity should be required for optimizing an unstable process. If we specify the cost relation between the number and the precision, we can determine the optimum k and ξ by the use of the diagram shown in Fig.7 and Fig.8.

CONCLUSION

In connection with the computing control, the measurement optimizations have been discussed in the case where the control process is a linear continuous system disturbed by noises subject to the specified cost function which consists of the control cost and the terminal error or steady state error. For a stable scalar system the optimum timing of measurement is always the initial part of the control interval if the external disturbances are absent. In the presence of disturbances, however, the optimum timing strongly depends on the form of cost function. For a considerably unstable process, the measurements are desired to be made about at uniform intervals independent of the initial uncertainty, disturbances and the parameter of the cost function.

The steady state optimization which is especially required in the direct digital control has been discussed representing a compromise between the number and the precision of measurements for various desired near optimum performance.

These discussions have rather general nature and include not only the stable controlled process but the unstable one such as fermentation processes or nuclear reactors and may be applied to the measurement optimization for multivariable processes.

REFERENCES

1. H.J.Kushner, "On the optimal timing of observations for linear control system with unknown initial states," IEEE Trans. on AC, AC-9,pp.144-150, 1964.
2. L.Meier, J.Peschon and P.M.Dressler," Optimal control of measurement subsystems," IEEE Trans. on AC, AC-12, pp.528-536, 1967.
3. J.J.Florentin, "Partial observability and optimal control," J.Electronics and Control, 9, pp.65-80, 1960.
4. W.M.Wonham, "Stochastic problem in optimal control," IEEE Inter. Conv.

Rec., 11, pt.2, pp.114-124, 1963.

5. R.E.Kalman, "A new approach to linear filtering and prediction problems," Trans. ASME, J. of Basic Engineering, ser. D, 82, pp.35-45, 1960.
6. R.E.Kalman and R.S.Bucy, "New results in linear filtering and prediction theory," Trans. ASME, J. of Basic Engineering, ser.D, 83, pp.528-536, 1961.
7. S.S.L.Chang, "Optimum filtering and control of randomly sampled systems," IEEE Trans. on AC, AC-12, pp.537-546, 1967.
8. J.J.Florentin, "Optimal control of continuous time, Markov stochastic systems," J.Electronics and Control, 10, pp.473-488, 1961.

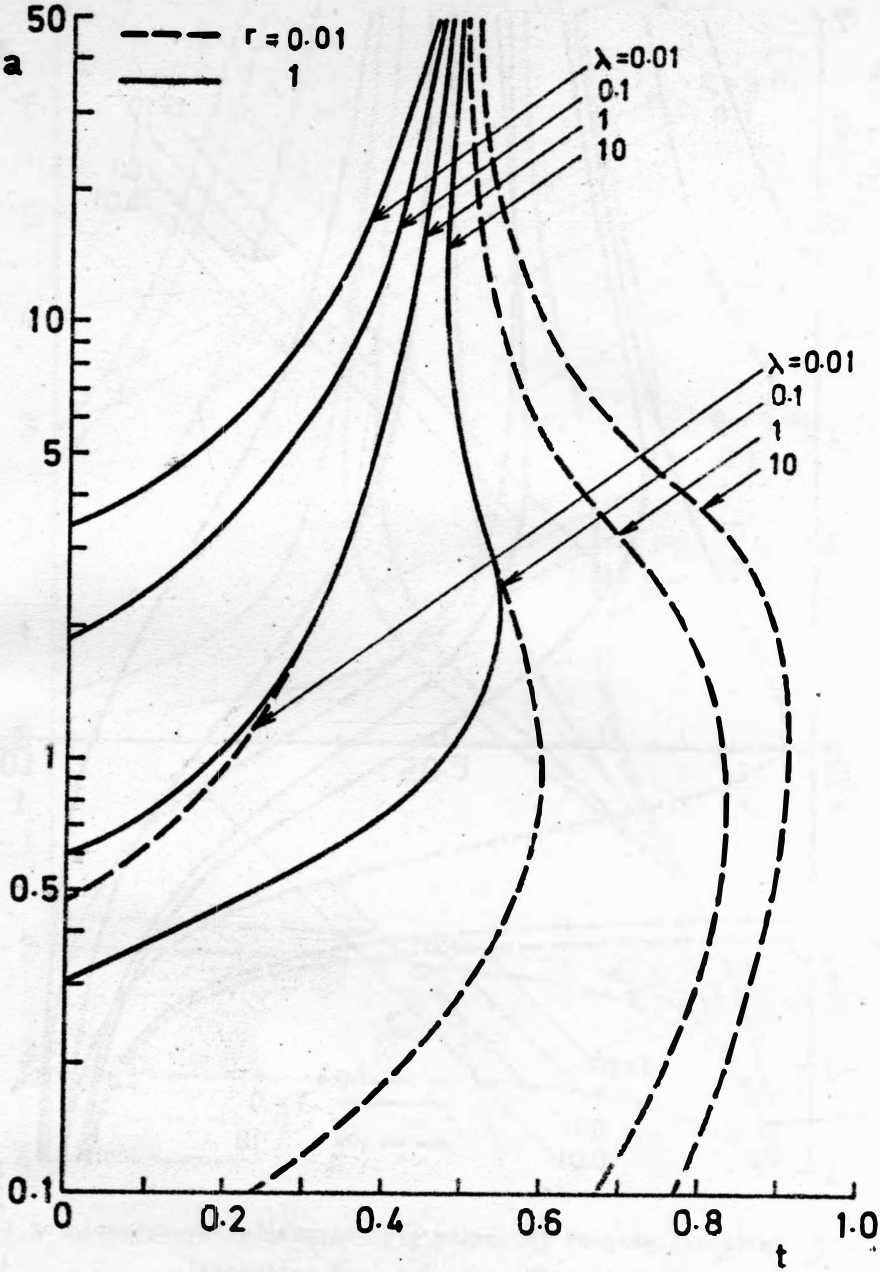


Fig.1 Optimum timing of the single measurement for the process with no disturbances: terminal optimization.

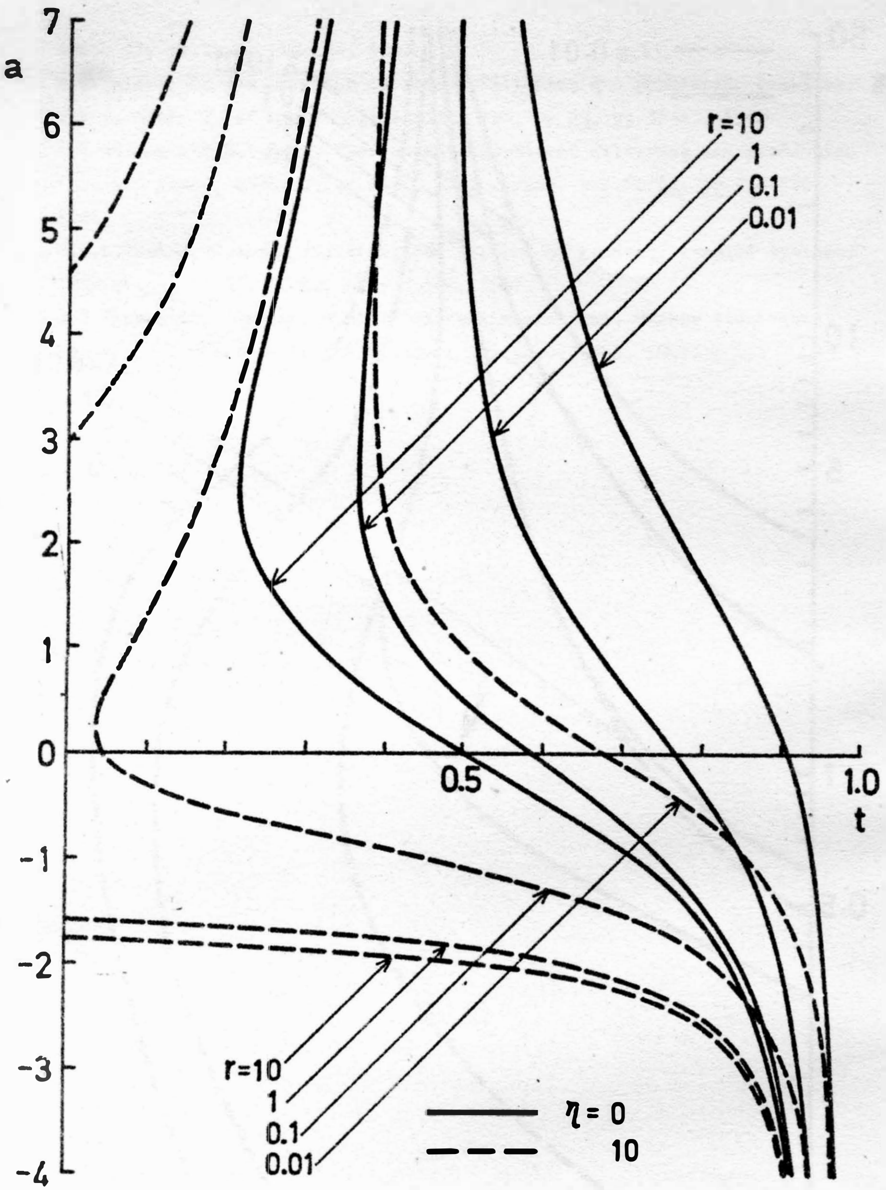


Fig.2 Optimum timing of the quite precise single measurement for the process with disturbances: terminal optimization.

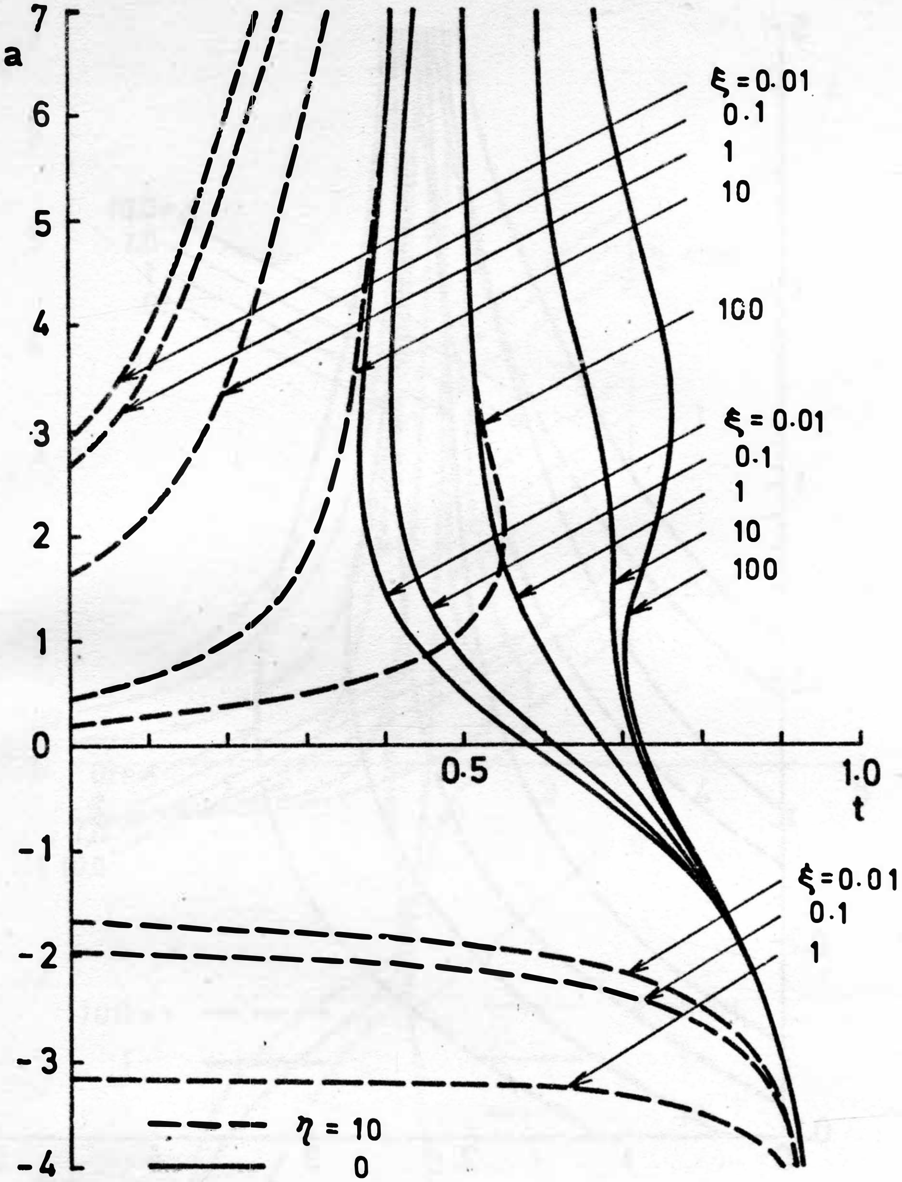


Fig.3 Optimum timing of the error corrupted single measurement for the process with disturbances: terminal optimization.

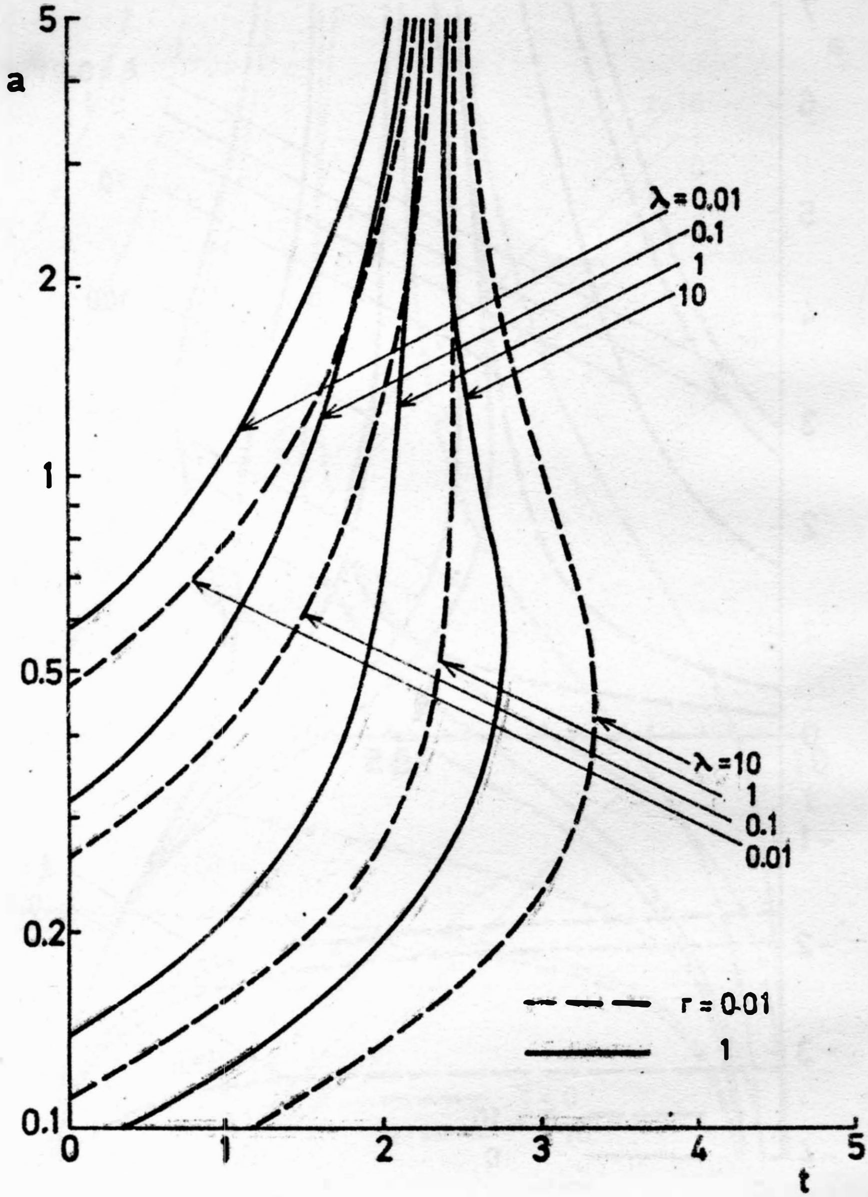


Fig.4 Optimum timing of the single measurement for the process with no disturbances: steady state optimization.

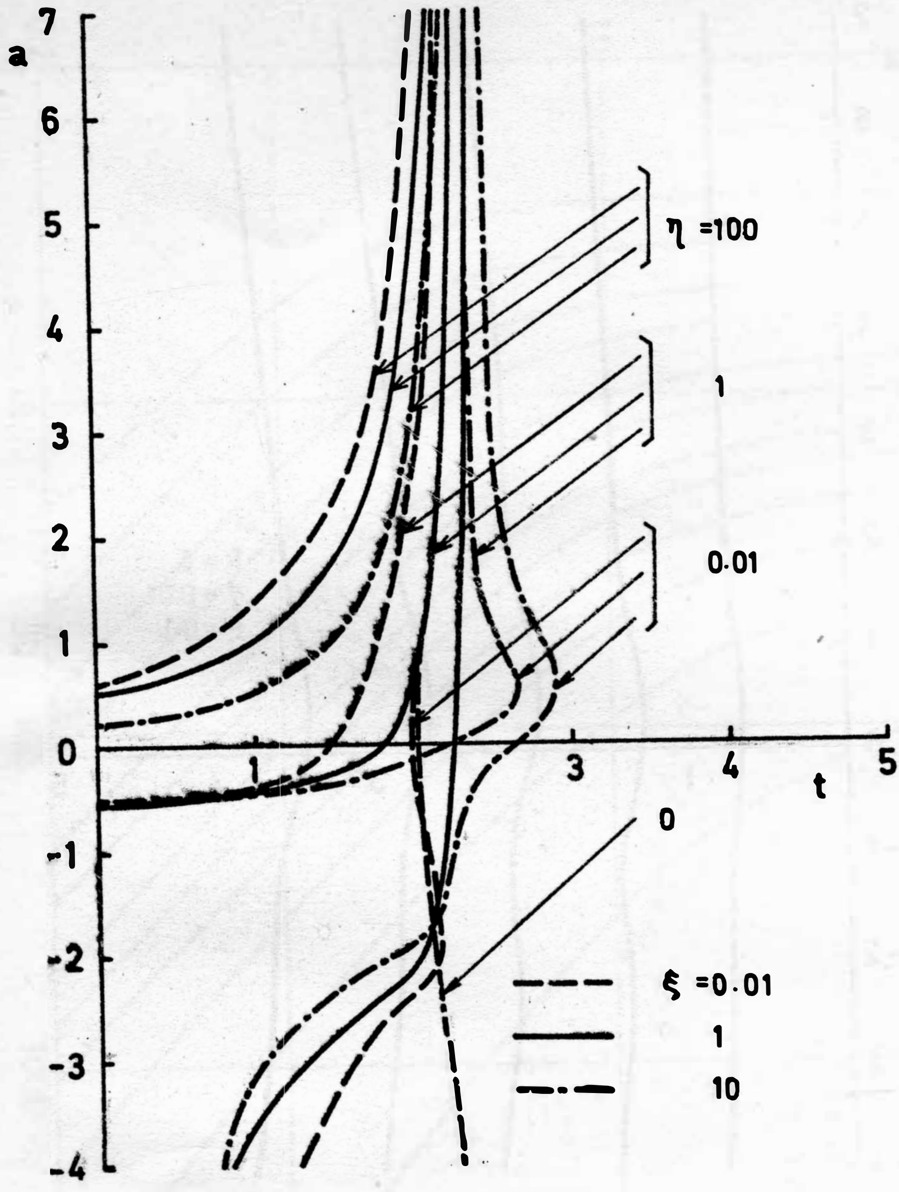


Fig.5 Optimum timing of the single measurement for the process with disturbances: steady state optimization.

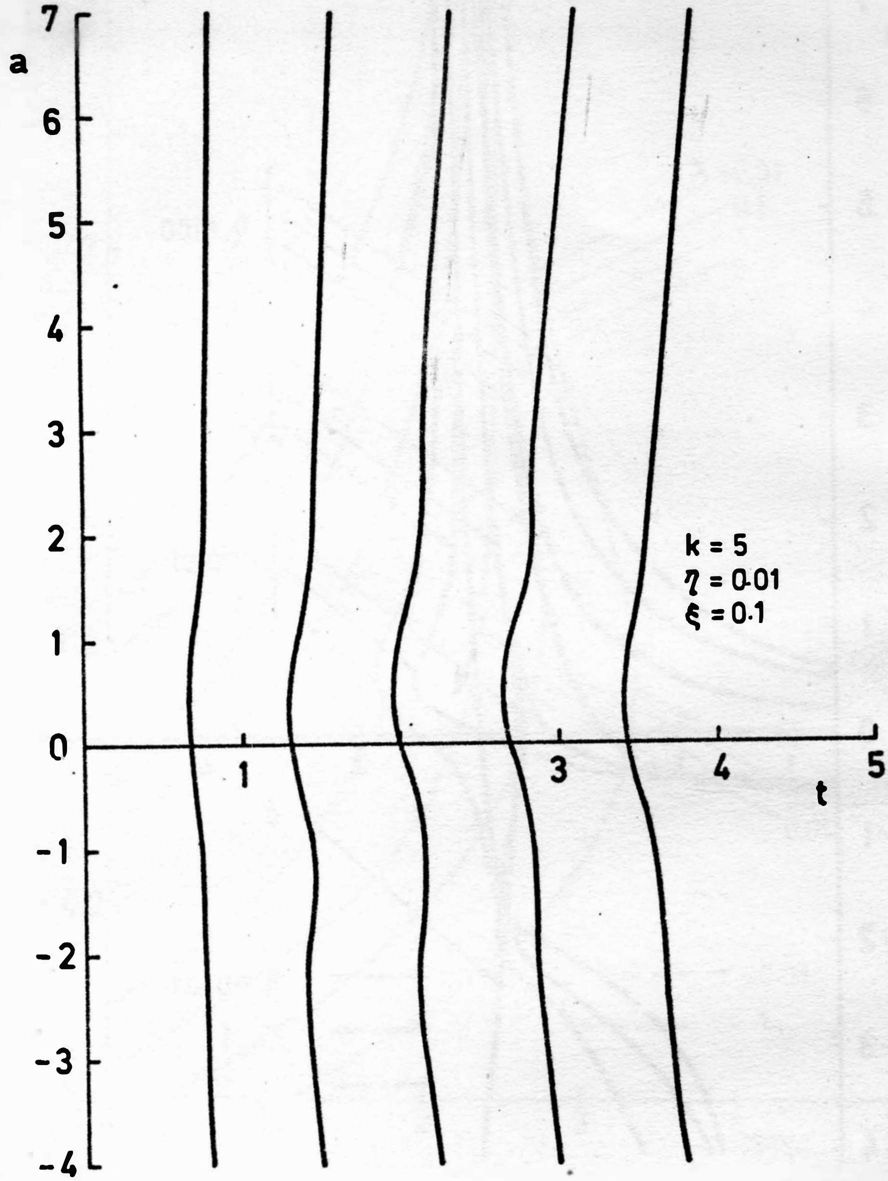


Fig.6 Optimum timing of five measurements for the process with disturbances: steady state optimization.

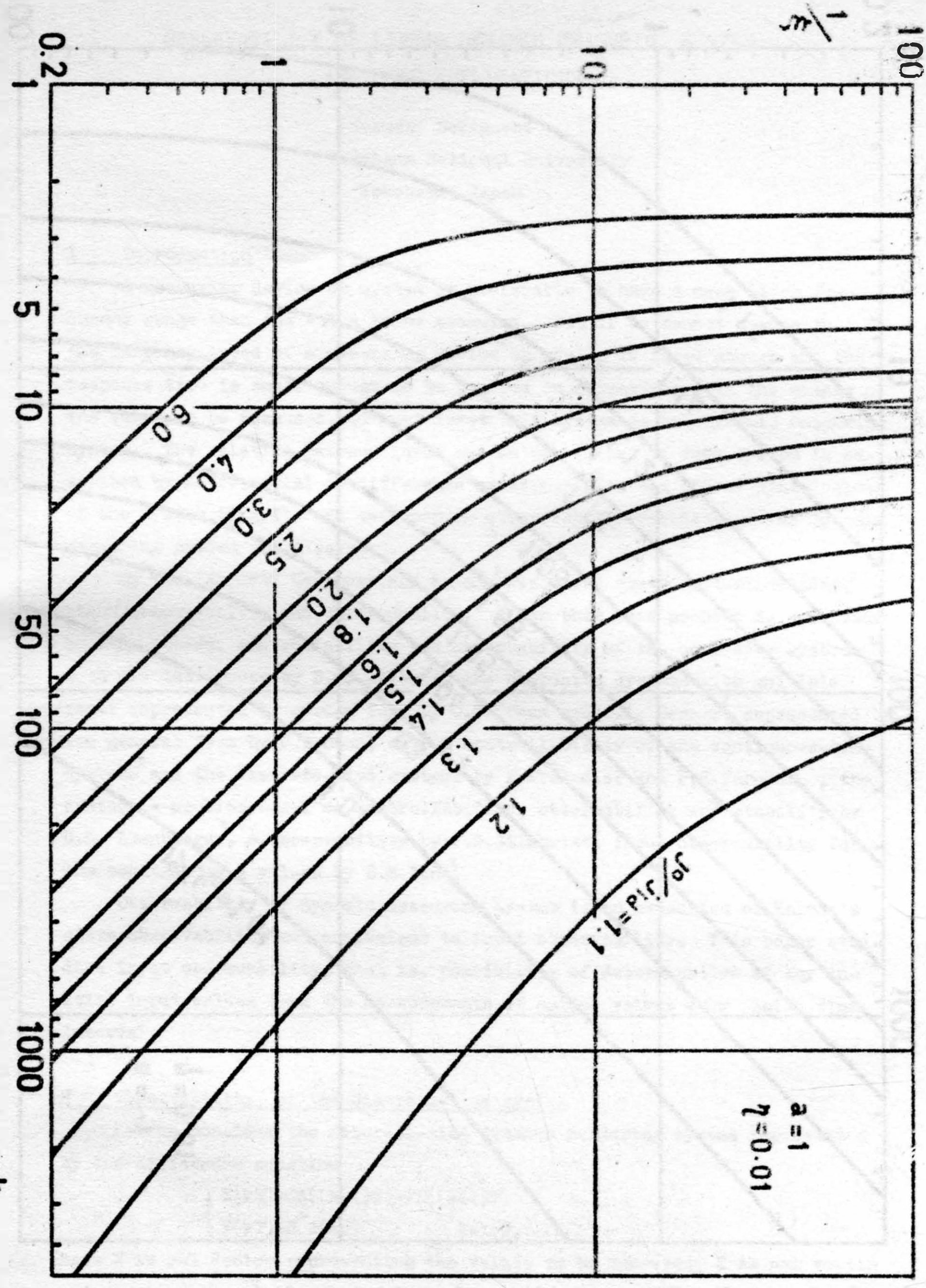


Fig.7 Compromise between number and precision of the measurements for an unstable process with disturbances.

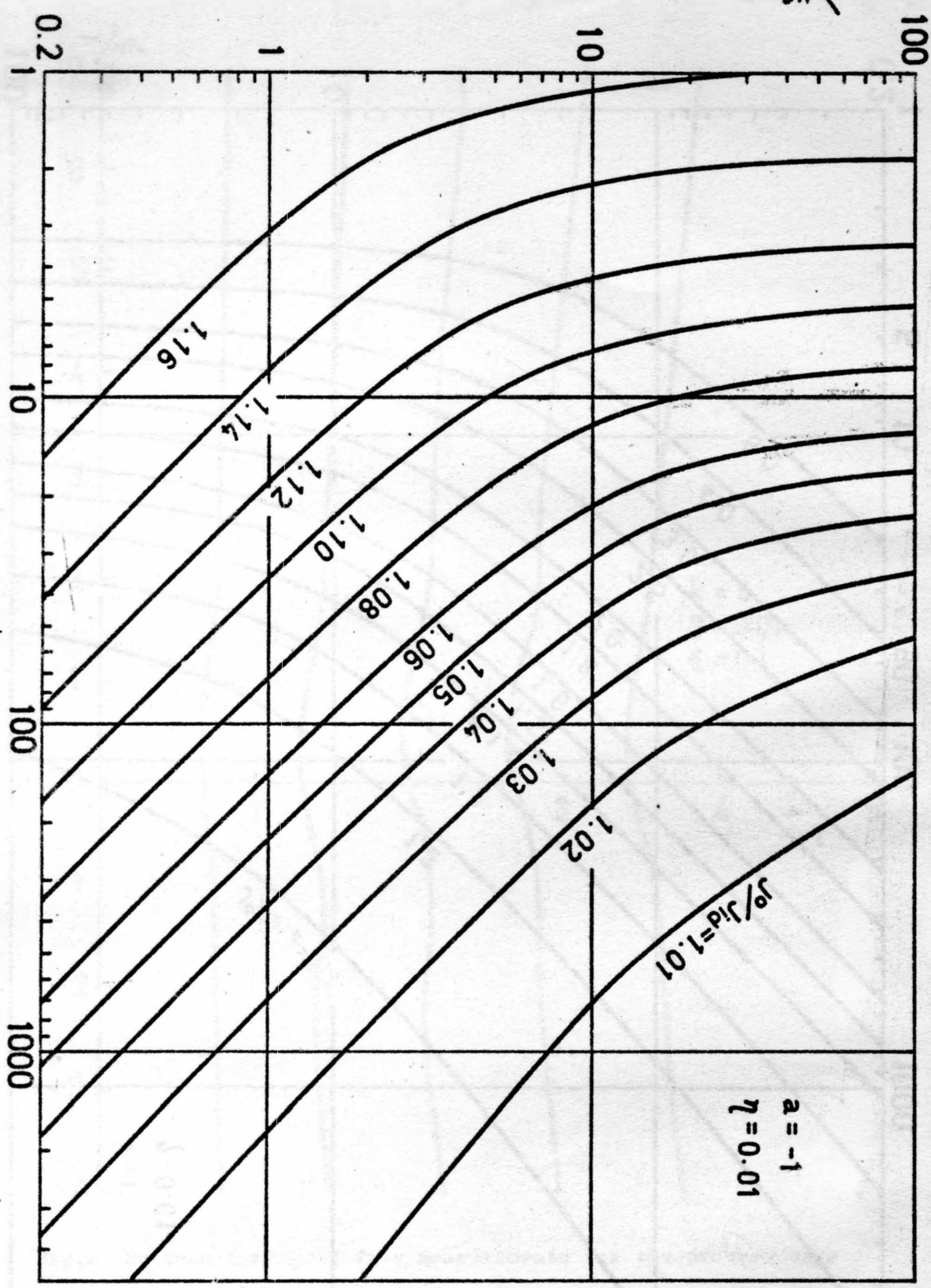


Fig.8 Compromise between number and precision of the measurements for a stable process with disturbances.

OBSERVABILITY OF LINEAR DYNAMIC MEASURING SYSTEM AND SOME APPLICATIONS

Takashi Sekiguchi
Yokohama National University
Yokohama, Japan

1 Introduction

A measuring device or system is preferable to have a much wider frequency range than the value to be measured. But if we cannot assume that the response speed of a measuring device or system is large enough and the response time is small enough to be ignored in comparison with the change of the value to be measured, we must treat this system as the dynamic measuring system. The relation between input and output values of this system is described by differential or difference equations with the aid of state values of the system (Fig.1). In this paper, observability problems of the dynamic measuring system are discussed.

On the 1st IFAC Congress, R.E.Kalman[1] discussed state controllability, state observability and their duality. After that this problem is discussed by many papers; controllability and observability of the composite systems with distinct roots by E.G. Gilbert[2], the composite systems with multiple roots represented by Jordan form by C.T. Chen and C.A. Desoer[3], represented by the general form by C.T.Chen[4], output controllability of the continuous-time systems[5] and the discrete-time systems[6] by E.Kreindler and P.E.Sarachik, the synthesis problem based on controllability, observability and stability by D.G. Luenberger[7], n-observability by J.D.Gilchrist[8], input observability for the constant mean values by B.E.Bona[9].

Observability of dynamic measuring system is an extension of Kalman's state observability and equivalent to input observability. This paper studies input observability, that is, possibility of determination of any initial input values from the measurements of output values over finite time interval.

2 Observability of the discrete-time system

Let us consider the discrete-time dynamic measuring system represented by the difference equation

$$\begin{cases} Z(kT)=GZ((k-1)T)+FX((k-1)T) \\ Y(kT)=H^{*}Z(kT) \qquad k=1,2,\ldots\ldots \end{cases} \tag{1}$$

where X is rx1 vector representing the values to be measured, Z is nx1 vector representing the state values of the measuring system, Y is mx1 vector representing the state values of the measuring system, F is $n\times r$ matrix, G is $n\times n$ matrix and H is $n\times m$ matrix (* denotes conjugate transpose). The problem is to obtain the necessary and sufficient conditions for determining any r initial values from m values measurements. Then we can assume rank of F is r and rank of H is m. And let us assume rank of G is n. This system is shown on Fig.2. In case $Z(0)$ is unknown, it is first desirable to determine $Z(0)$. So we discuss two cases;in case 1 the determination of $X(0)$ where $Z(0)$ is known and in case 2 the determination of $Z(0)$.

2.1 In case initial state values (Z(0)) of the system (1) are known

Definition 1: The discrete-time dynamic measuring system is said to have the i-th observability if the minimum number of samplings to determine $X(0)$ is i.

This definition shows that $X(0)$ of the i-th observable system cannot be determined from $Y(T),\cdots, Y((i-1)T)$ but can be determined from $Y(T),\cdots, Y((i-1)T)$ and $Y(iT)$.

On each sampling we have m data, but also r unknown values. Then the following equation is necessary for observability of the system (1).

$$r\leq m\leq n \qquad (2)$$

Now we consider the i-th observable conditions of the system (1). We can find easily that the necessary and sufficient condition of the system (1) for the 1-st observability is $\mathrm{rank}[H^*F]=\ell_1=r$. If $\mathrm{rank}[H^*F]=\ell_1<r$, we must consider the next sampling data. From theorem I~III of Appendix I~IV, the necessary and sufficient condition for the 2-nd observability is

$$\mathrm{rank}\begin{bmatrix} H^*F & 0 \\ H^*GF & H^*F \end{bmatrix}=\ell_2=r+\ell_1$$

For example, let us consider the following system

$$G=\begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 2 & 3 & 1 \end{bmatrix} \qquad F=\begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad H=\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

Calculating the above equations, we have $\ell_1=1$ and $\ell_2=3=r+\ell_1$.

If $\ell_2<r+\ell_1$, we need more data observed, and the necessary and sufficient condition for the 3-rd observability is

$$\mathrm{rank}\begin{bmatrix} H^*F & 0 & 0 \\ H^*GF & H^*F & 0 \\ H^*G^2F & H^*GF & H^*F \end{bmatrix}=\ell_3=r+\ell_2$$

Thus we have the following theorem.

Theorem 1: The necessary and sufficient condition for the i-th observability of the system (1) for $X(0)$ is

$$\begin{cases} \ell_{k-1} < r + \ell_{k-2} \\ \ell_i = r + \ell_{i-1} \end{cases} \qquad k = 2,3,\cdots,i \quad \ell_0 = 0 \tag{3}$$

$$\text{where } \ell_k = \text{rank} \begin{bmatrix} H^*F & \cdots\cdots & 0 \\ \vdots & & \vdots \\ H^*G^{k-1}F & \cdots & H^*F \end{bmatrix} \tag{4}$$

On the k-th sampling, the first r columns of the matrix in equation (4) are added, and so we obtain the following corollary.

Corollary 1-1: The necessary condition for the i-th observability of the system (1) for X(0) is that rank of the following matrix is r

$$[F^*H \vdots F^*G^*H \vdots \cdots \vdots F^*(G^*)^{i-1} H]$$

We can also regard that the last m rows of the matrix in equation (4) are added on the k-th sampling, and so the following corollary.

Corollary 1-2: The necessary condition for the i-th observability of the system (1) for X(0) is that rank of the following matrix is r

$$[H^*F \vdots H^*GF \vdots \cdots \vdots H^*G^{i-1} F]$$

The above corollary 1-2 shows the necessity of the r new independent informations among m data of the i-th sampling.

On the other hand, the corollary 1-1 shows the necessity of independence of the r column vectors which are coefficients of $x_1(0),\cdots,x_r(0)$. The matrix in equation (4) can be regarded as the set of r×k coefficient column vectors, and so theorem 1 means that no one column vector among the first r coefficient column vectors can be represented by the linear combinations of the other r×i-1 coefficient column vectors. That is,

Corollary 1-3: The necessary and sufficient condition for the i-th observability of the system (1) for X(0) is for any γ_1(non-zero), $\gamma_2\cdots\gamma_i$

$$\begin{bmatrix} H^*F \\ \vdots \\ H^*G^{i-1}F \end{bmatrix}\gamma_1 + \begin{bmatrix} 0 \\ \vdots \\ H^*G^{i-2}F \end{bmatrix}\gamma_2 + \cdots + \begin{bmatrix} 0 \\ \vdots \\ H^*F \end{bmatrix}\gamma_i \neq 0 \tag{5}$$

where $\gamma_1^* = [\gamma_{11}^* \cdots \gamma_{1r}^*] \neq 0$, $\gamma_2^* = [\gamma_{21}^* \cdots \gamma_{2r}^*]$, $\cdots$, $\gamma_i^* = [\gamma_{i1}^* \cdots \gamma_{ir}^*]$.

If we cannot form the 1-st, the 2-nd,$\cdots$, and the i-th observability of the system (1), then from corollary 1-3 there are γ_1(non-zero), $\gamma_2\cdots,\gamma_i$ such as

$$\begin{bmatrix} H^*F \\ \vdots \\ \vdots \\ H^*G^{i-1}F \end{bmatrix}\gamma_1 + \cdots + \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ H^*F \end{bmatrix}\gamma_i = \begin{bmatrix} H^*F\gamma_1 \\ \vdots \\ \vdots \\ H^*(G^{i-1}F\gamma_1 + \quad + F\gamma_i) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix} \tag{6}$$

Let us put $f_1 = F\gamma_1, \cdots, f_i = G^{i-1}F\gamma_1 + \cdots + F\gamma_i$. Then $f_1,\cdots$, fi represent vectors in the n dimensional space. As the column vectors $H_1,\ldots$, Hm of H are independent, they span the m dimensional subspace in the n dimensional space. If we denote this m dimensional subspace by S(H), then equation (6) means that vectors $f_1,\cdots$, fi cross S(H) at right angle, i.e. $f_1 \perp S(H)$ & $f_2 \perp S(H)$ ..& $f_i \perp S(H)$. Vectors $f_1,\cdots$, fi lie in the n-m dimensional subspace which is complement of S(H) (Fig. 3). Now consider $f_{i+1}=Gf_i+F\gamma_{i+1}$. If $f_{i+1}\perp S(H)$ for some γ_{i+1}, then this system has not the i+1-th observability. Putting, here, i=n, then from Appendix V f_{n+1} can be represented by the linear combination of $f_1,\dots,f_n$. The vector f_{n+1} lies in the n-m dimensional complementary subspace and $f_{n+1}\perp S(H)$. That is, this system has not the n+1-th observability. Thus the theorem 2.

Theorem 2: If the system (1) does not have the n-th observability for X(0), then the system is non-observable, i.e. the order of observability for X(0) is at most n.

2.2 In case initial state values (Z(0)) of the system (1) are unknown

It is not always necessary to know Z(0) before determiming X(0), but it is sometimes desirable to determine Z(0) if possible. So the problem of this section is to obtain necessary and sufficient conditions for observability of the system (1) for Z(0) from Y(kT) not knowing X((k-1)T) (k=1,2···) precisely.

Definition 2: We say the system (1) is the i-th observable for Z(0), when i is the minimum number of samplings necessary to determine Z(0) without knowledge of X(kT).

In this case, if the order of observability for Z(0) is the same as that for X(0)

$$r < m \leq n \tag{7}$$

The minimum number of necessary samplings to determine Z(0) is dependent on wave forms of inputs, and it is attained when zero value inputs. Therefore sampling number must be equal to or larger than n/m.

Now, the coefficient matrix obtained from i samplings is

$$\begin{bmatrix} H^*G & H^*F & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ H^*G^i & H^*G^{i-1}F & \cdots & H^*F \end{bmatrix} \tag{8}$$

Let us denote rank of matrix (8) by l'_i, and we have the following theorem similar to theorem 1.

Theorem 3: The necessary and sufficient conditions for the i-th observability of the system (1) for Z(0) is

$$\begin{cases} l'_k < n+l_k & k=1,2,\dots,i-1 \\ l'_i = n+l_i \end{cases} \tag{9}$$

Corollary 3-1: The necessary condition for the i-th observability of the system (1) for Z(0) is that rank of following matrix is n

$$[H \vdots G^*H \vdots \cdots \vdots (G^*)^{i-1}H]$$

The above corollary 3-1 shows the necessity of independence of the first n column vectors in matrix (8), and theorem 3 is equal to the fact that none of the first n column vectors can be represented by the linear combinations of the other n+r×i-1 column vectors in matrix (8). thus the corollary similar to corollary 1-3 follows.

Corollary 3-2: The necessary and sufficient condition for the i-th observability of the system (1) for Z(0) is for any γ_0 (non-zero), $\gamma_1, \cdots, \gamma_i$

$$\begin{bmatrix} H^*G \\ \vdots \\ H^*G^i \end{bmatrix} \gamma_0 + \begin{bmatrix} H^*F \\ \vdots \\ H^*G^{i-1}F \end{bmatrix} \gamma_1 + \cdots + \begin{bmatrix} 0 \\ \vdots \\ H^*F \end{bmatrix} \gamma_i \neq 0 \qquad (10)$$

where $\gamma_0^* = [\gamma_{01}^* \cdots \gamma_{0\ell}^* \cdots \gamma_{0n}^*] \neq 0$, $\gamma_1^* = [\gamma_{11}^* \cdots \gamma_{1r}^*], \cdots, \gamma_i^* = [\gamma_{i1}^* \cdots \gamma_{ir}^*]$.

From Appendix VI and the same discussion as that in 2.1, we obtain the following theorem similar to theorem 2 except the order n-1.

Theorem 4: If the system (1) has some non-zero inputs and does not have the n-1 th observability for Z(0), then the system is non-observable, i.e. the order of observability for Z(0) is at most n-1.

3 Observability of the continuous-time system

Next we consider the continuous-time dynamic measuring system represented by the differential equation

$$\begin{cases} \dot{Z} = AZ + BX \\ Y = C^*Z \end{cases} \qquad (11)$$

where X is rx1 vector representing the values to be measured, Z is nx1 vector representing the state values of the measuring system, Y is mx1 vector representing the output values of the measuring system, B is nxr matrix, A is nxn matrix and C is nxm matrix. Here we assume the values to be measured X(t) are continuous and finite, and $X(t) \geq 0$ (or ≤ 0) without loss of generality (see Appendix VII). The problem is to determine the mean values X(to) ($0 \leq to \leq T$) of X(t) over small time interval $0 \leq t \leq T$ (Fig. 4).

3.1 In case initial state values (Z(0)) of the system (11) are known

From (11) we obtain

$$W(t) = Y(t) - C^* e^{At} Z(0) = C^* \int_0^t e^{A(t-\tau)} BX(\tau) d\tau$$

$$\therefore \; W(T) = C^* \int_0^T e^{A(T-\tau)} BX(\tau) d\tau$$

Since $e^{A(T-\tau)}$ is continuous function

$$W(T) = C^* e^{A(T-\xi_0)} B \int_0^T X(\tau) d\tau = C^* e^{A(T-\xi_0)} BTX(to) \qquad (0 \leq \xi_0 \leq T)$$

where X(to) is mean value, i.e.

$$X(to) = \frac{1}{T} \int_0^T X(\tau) d\tau$$

Let p denote the degree of minimal polynomial of A and $\alpha_\ell(t)$ scalar value of t^{ℓ}, and we have

$$W(T) = C^* \sum_{\ell=0}^{p-1} T\alpha_\ell(T-\xi_0) A^\ell BX(to) = \sum_{\ell=0}^{p-1} T\alpha_\ell(T-\xi_0)(B^*(A^*)^\ell C, X(to))$$

where (,) denotes inner product. Taking the inner product of $T\alpha_\ell(T-\xi_0)$ and a component of W(T), we obtain

$$\begin{bmatrix} (T\alpha_0, W) \\ \vdots \\ (T\alpha_{p-1}, W) \end{bmatrix} = \begin{bmatrix} (T\alpha_0, T\alpha_0) \cdots (T\alpha_0, T\alpha_{p-1}) \\ \vdots \qquad \vdots \\ (T\alpha_{p-1}, T\alpha_0) \cdots (T\alpha_{p-1}, T\alpha_{p-1}) \end{bmatrix} \begin{bmatrix} (B^*C, \; X(to)) \\ \vdots \\ (B^*(A^*)^{p-1}C, \; X(to)) \end{bmatrix}$$

Let us describe the above equation in brief

$$\left[(T\alpha_k,\ W)\right] = \left[(T\alpha_k,\ T\alpha_j)\right]\ \left[(B^*(A^*)^k C,\ X(to))\right]$$

Since the $\alpha_k(T-\xi_0)$ are linearly independent of each other, Gram matrix $\left[(T\alpha_k,\ T\alpha_j)\right]$ is non-singular, and we obtain the following

$$(B^*\ (A^*)^k C,\ X(to)) = \beta_k \qquad k=0,\cdots,p-1 \qquad \beta_k: m\times 1 \ \text{vector}$$

From the above equations, $x_1(0),\cdots,x_r(0)$ can be uniquely solved if and only if rank of the following matrix is r

$$\left[B^*C \,\vdots\, B^*A^*C \,\vdots\, \cdots\cdots \,\vdots\, B^*(A^*)^{p-1}C\right]$$

Rank of this matrix is the same as

$$\left[B^*C \,\vdots\, B^*A^*C \,\vdots\, \cdots\cdots \,\vdots\, B^*(A^*)^{p-1}C \,\vdots\, \cdots \,\vdots\, B^*(A^*)^{n-1}C\right]$$

Next we consider to use data over one more interval $T\leq t\leq 2T$.

From (10) we get

$$W(2T) = \sum_{k=0}^{p-1}\left\{ T\alpha_k(2T-\xi_0)(B^*(A^*)^k C, X(to)) + T\alpha_k(T-\xi_1)(B^*(A^*)^k C, X(T+t_1))\right\}$$

where $0\leq \xi_1,\ t_1\leq T$ and

$$X(T+t_1) = \frac{1}{T}\int_T^{2T} X(t)dt$$

If the $\alpha_k(2T-\xi_0)$ and the $\alpha_k(T-\xi_1)$ are linearly independent, we obtain

$$\begin{bmatrix}\left[(T\alpha_k(2T-\xi_0), W(2T))\right] \\ \left[(T\alpha_k(T-\xi_1), W(2T))\right]\end{bmatrix} = \Gamma \begin{bmatrix}\left[(B^*(A^*)^k C,\ X(to))\right] \\ \left[(B^*(A^*)^k C,\ X(T+t_1))\right]\end{bmatrix}$$

where
$$\Gamma = \begin{bmatrix}\left[(T\alpha_k(2T-\xi_0), T\alpha_j(2T-\xi_0))\right] & \left[(T\alpha_k(2T-\xi_0),\ T\alpha_j(T-\xi_1))\right] \\ \left[(T\alpha_k(T-\xi_1), T\alpha_j(2T-\xi_0))\right] & \left[(T\alpha_k(T-\xi_1), T\alpha_j(T-\xi_1))\right]\end{bmatrix}$$

Since the Gram matrix Γ is non-singular, we have

$$\begin{cases}(B^*(A^*)^k C,\ \ X(to)) = \beta_k & k=0,1,\cdots,p-1 \\ (B^*(A^*)^k C,\ \ X(T+t_1)) = \delta_k & \delta_k: m\times 1 \ \text{vector}\end{cases}$$

Therefore we can not get any more new informations about X(to). If $\alpha_i(T-\xi_1)$ can be represented by the linear combination of the others, i.e.

$$\alpha_i(T-\xi_1) = \sum C_{oj}\alpha_j(2T-\xi_0) + \sum_{j\neq i} C_{ij}\,\alpha_j(T-\xi_1)$$

where C_{oj}, C_{ij} are constants, we obtain

$$\begin{bmatrix}\left[(T\alpha_k(2T-\xi_0),\ W(2T))\right] \\ \left[(T\alpha_k(T-\xi_1),\ W(2T))\right]\end{bmatrix} = \Gamma_i \begin{bmatrix}\left[(B^*(A^*)^k C, X(to)) + C_{ok}(B^*(A^*)^i C,\ X(T+t_1))\right] \\ \left[(B^*(A^*)^k C, X(T+t_1)) + C_{ik}(B^*(A^*)^i C, X(T+t_1))\right]\end{bmatrix}$$

where Γ_i is the same as Γ without the terms $\alpha_i(T-\xi_1)$. Since the Gram matrix Γ_i is non-singular, we have

$$\begin{cases}(B^*(A^*)^k C, X(to)) + C_{ok}(B^*(A^*)^i C,\ X(T+t_1)) = \beta'_k & \beta'_k, \delta'_k: m\times 1 \ \text{vector} \\ (B^*(A^*)^k C, X(T+t_1)) + C_{ik}(B^*(A^*)^i C,\ X(T+t_1)) = \delta'_k & k=0,1,\cdots,p-1 \ \&\ k\neq i\end{cases}$$

In this case we can get no more new informations about X(to) as well. Thus the theorem.

<u>Theorem 5:</u> The necessary and sufficient condition for observability of the system (11) for X(to) is that the following matrix has rank r

$$\left[B^*C \,\vdots\, B^*A^*C \,\vdots\, \cdots \,\vdots\, B^*(A^*)^{n-1}C\right]$$

<u>3.2 In case initial state values (Z(0)) of the system (11) are unknown</u>

The problem is the same as 2.2. From (10) we get

$$Y(T)=C^*e^{AT}Z(0)+C^*e^{A(T-\xi_0)}BTX(t_0)=\sum_{\ell=0}^{p-1}\{\alpha_\ell(T)C^*A^{\ell}Z(0)+T\alpha_\ell(T-\xi_0)C^*A^{\ell}BX(t_0)\}$$

Taking the inner product, we obtain

$$\begin{bmatrix}[(\alpha_\ell(T),Y(T))]\\ [(T\alpha_\ell(T-\xi_0),Y(T))]\end{bmatrix}=\begin{bmatrix}[(\alpha_\ell(T),\alpha_j(T))] & [(\alpha_\ell(T),T\alpha_j(T-\xi_0))]\\ [(T\alpha_\ell(T-\xi_0),\alpha_j(T))] & [(T\alpha_\ell(T-\xi_0),T\alpha_j(T-\xi_0))]\end{bmatrix}\begin{bmatrix}[((A^*)^{\ell}C,Z(0))]\\ [(B^*(A^*)^{\ell}C,X(t_0))]\end{bmatrix}$$

Since the $\alpha_\ell(T)$ and the $T\alpha_\ell(T-\xi_0)$ are all linearly independent, we obtain

$$\begin{cases}((A^*)^{\ell}C,\ Z(0))=\varepsilon_\ell & \varepsilon_\ell: m\times 1 \text{ vector}\\ (B^*(A^*)^{\ell}C,\ X(t_0))=\rho_\ell & \ell=0,1,\cdots,p-1\end{cases}$$

Even if we use data over one more interval $T\leq t\leq 2T$, we can not get any more informations about $Z(0)$. So the following theorem.

Theorem 6: The necessary and sufficient condition for observability of the system (11) for $Z(0)$ is that rank of the following matrix is n

$$[C\ \vdots\ A^*C\ \vdots\cdots\vdots\ (A^*)^{n-1}C]$$

4 Observability of the time-varying system

Here we consider only the discrete-time system described by the following difference equation

$$\begin{cases}Z(kT)=G((k-1)T)Z((k-1)T)+F((k-1)T)X((k-1)T)\\ Y(kT)=H^*(kT)\ Z(kT) \qquad k=1,2,\cdots\end{cases}\qquad (12)$$

From i samplings, we obtain the following coefficient matrix

$$\begin{bmatrix}H^*(T)\Phi(T,0) & H^*(T)\Phi(T,T)F(0) & 0 & \cdots 0\\ \vdots & \vdots & \vdots & \vdots\\ H^*(iT)\Phi(iT,0) & H^*(iT)\Phi(iT,T)F(0) & H^*(iT)\Phi(iT,2T)F(T) & \cdots H^*(iT)\Phi(iT,iT)F((i-1)T)\end{bmatrix}$$

where Φ denotes the transition matrix. Putting as the following

$$\begin{cases}P_o^*\ (iT,0)=[\Phi^*(T,0)\ H(T)\ \vdots\cdots\vdots\ \Phi^*(iT,0)\ H(iT)]\\ P_i^*\ (iT,0)=[F^*(0)\Phi^*(T,T)H(T)\ \vdots\cdots\vdots\ F^*(0)\Phi^*(iT,T)\ H(iT)]\end{cases}$$

we get the following corollaries similar to corollary 1-1 and 3-1.

Corollary 1-4: The necessary condition for the i-th observability of the system (12) for $X(0)$ is that rank of $P_i^*\ (iT,0)$ is r, i.e. the following Gram matrix is non-singular

$$\Gamma_i=P_i^*\ (iT,0)\ P_i(iT,0)$$

Corollary 3-3: The necessary condition for the i-th observability of the system (12) for $Z(0)$ is that rank of $P_o^*\ (iT,0)$ is n, i.e. the following Gram matrix is non-singular

$$\Gamma_o=P_o^*\ (iT,0)\ P_o(iT,0)$$

5 The application to an open loop system

The analysis of dynamic behaviour of a non-linear system to an arbitrary input is in general very difficult, and we can not understand it without measurement. The problem is to grasp the characteristics of the non-linear system from measurement through the linear dynamic measuring system (Fig 5). Let us consider an induction motor as an example of non-linear system. The transient response, especially transient torque response to a sudden change of operating condition driven by sinusoidal voltage has been studied until now from the point of design and operation.[12~17] But the speed control of an induction motor with a variable frequency inverter has been possible because of the recent semiconductor's advance.[18~21] The voltage wave form supplied from the semiconductor inverter is not sinusoidal, and the new problems have arisen. There are some studies on static characteristics[22~24], dynamic and transient behaviour[25~27] and analog simulation.[28,29] Here we consider the observability problem of dynamic or transient torque of the induction motor.

The motor-load-measuring device system is shown in Fig.6. Putting $\theta_1 - \theta_2 = z_1$, $\dot{\theta}_1 = z_2$, $\dot{\theta}_2 = z_3$, $T_M = x_1$, $T_L = x_2$, $K/J_1 = a_1$, $E_1/J_1 = b_1$, $1/J_1 = c_1$, $K/J_2 = a_2$, $E_2/J_2 = b_2$, $1/J_2 = c_2$, we obtain the following state equation of the dynamic measuring system

$$\frac{d}{dt}\begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & -1 \\ -a_1 & -b_1 & 0 \\ a_2 & 0 & -b_2 \end{bmatrix}\begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ c_1 & 0 \\ 0 & -c_2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Using the torque meter of strain gage and two tachometers, the observing equation of this system is

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \mu & 0 & 0 \\ 0 & \mu_1 & 0 \\ 0 & 0 & \mu_2 \end{bmatrix}\begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix}$$

Therefore $[B^*C \,\vdots\, B^*A^*C \,\vdots\, B^*(A^*)^2 C] =$

$$\left[\begin{array}{ccc|ccc|ccc} 0 & \nu_1 c_1 & 0 & \mu c_1 & -\nu_1 b_1 c_1 & 0 & -\mu b_1 c_1 & \nu_1(b_1^2 - a_1)c_1 & \nu_2 a_2 c_1 \\ 0 & 0 & -\nu_2 c_2 & \mu c_2 & 0 & \nu_2 b_2 c_2 & -\mu b_2 c_2 & -\nu_1 a_1 c_2 & \nu_2(b_2^2 - a_1)c_2 \end{array}\right] \quad (13)$$

This system has observability for X(to). From the above matrix, it is easily seen that this system does not lose observability without two arbitrary measuring devices among three. We know observability for Z(0), because

$$[C \,\vdots\, A^*C \,\vdots\, (A^*)^2 C] = \left[\begin{array}{ccc|ccc|ccc} \mu & 0 & 0 & 0 & -\nu_1 a_1 & \nu_2 a_2 & -\mu(a_1 + a_2) & \nu_1 a_1 b_1 & -\nu_2 a_2 b_2 \\ 0 & \nu_1 & 0 & \mu & -\nu_1 b_1 & 0 & -\mu b_1 & \nu_1(b_1^2 - a_1) & \nu_2 a_2 \\ 0 & 0 & \nu_2 & -\mu & 0 & -\nu_2 b_2 & \mu b_2 & \nu_1 a_1 & \nu_2(b_2^2 - a_2) \end{array}\right] \quad (14)$$

But if we choose state variables as $\theta_1 = z_1$, $\dot{\theta}_1 = z_2$, $\theta_2 = z_3$, $\dot{\theta}_2 = z_4$, this system does not have observability for Z(0) any more, without losing observability for X(to) with only one arbitrary measuring device when $b_1 \neq b_2$. This means that the way to choose state variable has an effect on observability for Z(0) but does not for X(to).

If we put $x_1 = 0$, then the problem becomes to measure the load or braking torque behaviour. Putting $c_2 = 0$ in equation (13), we obtain observability for X(to) with an arbitray measuring device. The condition $c_2 = 0$ does not change the matrix (14) at all. This means that the way to choose input variables does not have an effect on observability for Z(0).

$x_2 = 0$ means the motor is driven with inertia load. Concerning observability, this is the same as $x_1=0$. Put $J_2=\infty$, and we have the condition with load locked or the condition on starting. Therefore $\theta_2=0$. Fig.7 shows the transient torque behaviour of the three phase induction motor (0,2kw, 200v, 4P) with load locked. These have been observed with the torque meter at the moment of switching-on. This system has $J_1=6,5\times9,8\times10^{-5}$ [$N\cdot m\cdot s^2$], $K=26\times9,8$ [N·m/rad]. Let us neglect the friction loss, i.e. $E_1=0$. The natural angular frequency $\omega_0=\sqrt{K/J_1}=632$. When the motor torque changes slowly or has no harmonics of larger than ω_0, we need not to treat this system as dynamic measuring system. Otherwise $J_1\ddot{\theta}_1$ has an effect on the torque measurement. Fig.8 and Fig.9 show the magnitudes of the influences of $J_1\ddot{\theta}_1$ in comparison with $K\theta_1$, when the torque changes sinusoidally. Fig.10 shows that $J_1\ddot{\theta}_1$ is larger than $K\theta_1$ on sudden changes.

Next we discretize this system[11].

$$G=\begin{bmatrix}\cos\omega_0T & \frac{1}{\omega_0}\sin\omega_0T\\ -\frac{1}{\omega_0}\sin\omega_0T & \cos\omega_0T\end{bmatrix}\quad F=\frac{1}{\sqrt{J_1K}}\begin{bmatrix}\frac{1}{\omega_0}(1-\cos\omega_0T)\\ \sin\omega_0T\end{bmatrix}\quad H=\begin{bmatrix}\mu\\ 0\end{bmatrix}$$

where T is the sampling period. The coefficient matrix is

$$\begin{bmatrix}H^*F & 0\\ H^*GF & H^*F\end{bmatrix}=\begin{bmatrix}\frac{\mu}{K}(1-\cos\omega_0T) & 0\\ \frac{\mu}{K}(1-\cos\omega_0T)\cos\omega_0T+\sin^2\omega_0T & \frac{\mu}{K}(1-\cos\omega_0T)\end{bmatrix}$$

Choosing the sampling period as $T=2\pi j/\omega_0$, $j=1,2,\cdots$, we have not observability of this system. Otherwise we do have (Fig.11). This shows the close relation between observability and sampling period, and the similarity to the sampling theorem.

6 The application to a closed loop system

L.I. Rozonoer points out the close relationship between controllability, observability and the possibility of describing a model by mathematics[1]. It is very useful to construct a model in order to understand the complex mechanism of nature. The deeper gets our recognition into the nature, the more precise becomes our model, and conversely we need the more precise model in order to understand the deeper nature. The propriety of a mathematical model depends on the degree of coincidence between the results of calculation and data of measurement. It is, therefore, desirable for a new mathematical model to possess observability and to be examined by experiment. Here we consider a mathematical model of the human thermal system. Since a living system has many closed loops and can not be examined in the state of open loop, we must study it in the usual state of closed loop (Fig.12).

There proposed some mathematical models such as cylinder models of C.H. Wyndham and A.R. Atkins[30] or of E.H. Wissler[31], slab model of R.J.Crosbie, J.D. Hardy and E.Fessenden[32]. Here we discuss the model proposed by Y.Kawashima,H.Yamamoto and M.Masubuchi[33](Fig.13). This model consists of three parts, i.e. body shell, body core and nervecenter for the thermal regulation. The thermal regulations are done in this model by evaporation, shivering and blood circulation. From heat balance, we obtain the following basic relation

$$\begin{cases} V_1 c_b \gamma_b \dot{t}_1 = c_b \gamma_b R(t_2 - t_1) + A_1 h_1 (t_4 - t_1) \\ V_2 c_b \gamma_b \dot{t}_2 = c_b \gamma_b R(t_1 - t_2) + A_2 h_2 (t_3 - t_2) \\ V_3 c_3 \gamma_3 \dot{t}_3 = A_2 h_2 (t_2 - t_3) + A_3 h_3 (t_e - t_3) + A_4 h_4 (t_4 - t_3) - N \\ V_4 c_4 \gamma_4 \dot{t}_4 = A_1 h_1 (t_1 - t_4) + A_4 h_4 (t_3 - t_4) + Q \end{cases}$$

After the proper operation of dimensionless and linearization around a steady state, we get

$$A = \begin{bmatrix} -1-R_s & R_s & 0 & 1 \\ \frac{R_s}{T_2} & \frac{-R_s - a_2}{T_2} & \frac{a_2}{T_2} & 0 \\ 0 & \frac{a_2}{T_3} & \frac{-a_2-a_4-a_5}{T_3} & \frac{a_4}{T_3} \\ \frac{1}{T_4} & 0 & \frac{a_4}{T_4} & \frac{-1-a_4}{T_4} \end{bmatrix} \quad B = \begin{bmatrix} t_{3s} - t_{1s} & 0 & 0 & 0 \\ \frac{t_{1s} - t_{3s}}{T_2} & 0 & 0 & 0 \\ 0 & \frac{a_3}{T_3} & \frac{-1}{T_3} & 0 \\ 0 & 0 & 0 & \frac{1}{T_4} \end{bmatrix} \quad C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

This system does not have observability for X(to) because the second and the third rows of the matrix $[B^*C \,\vdots\, B^*A^*C \,\vdots\, B^*(A^*)^2 C \,\vdots\, B(A^*)^3 C]$ are linearly dependent. This means that we cannot distinguish between evaporation and environment temperature influences. Therefore we must adopt one of the following two methods for examining this model by experiment. The first is to choose the environment temperature as the basis and constant. This temperature is changed as suddenly as possible like step function, and after that it is held constant. The other temperatures must be measured from this constant one. The data of shivering and evaporation are compared with the results of the calculations of the model. The accuracy of this method is not so good because of the roughness of the measurement of shivering and evaporation. It is better to use the environment temperature for comparison. The second method is to choose it and shivering as the inputs to be compared. This experiment must be done in the cold environment where the effect of evaporation is negligible.

7 Conclusions

Observability of linear dynamic measuring system or input observability is very useful for both measurement of complex system's behaviour and examination of a new mathematical model.

On the theoretical point of view there is one point made clear in this paper, i.e. the difference between discrete-time and continuous-time system. Theorem 5 and 6 give necessary and sufficient conditions, but the similar matrices of corollary 1-1 and 3-1 give only necessary conditions. The necessary conditions of (2) and (7) for the discrete-time systems have none of the correspondings for the continuous-time systems. The duality between output controllability and input observability is proved for continuous-time systems but not for discrete-time systems.

Acknowledgment.

The author wishes to express his gratitude to professor M.Masubuchi and his assistants H.Kanoh and Y.Kawashima who gave him many suggestions, to professor K.Ogata whose notes and book[11] were very useful for this work, to professor S.Matsuoka and the members of Study Group for Living Regulation Systems who discussed the model of the human thermal system, to professor I.Hatakeyama and the members of the physiologists' group who discussed from the physiological point of view. He also wishes to express his thanks to his assistant A.Ohsato for devoting much time to this work.

Appendix

I Let us consider the equation

$$w_k = \sum_{j=1}^{r} \varphi_{kj}\, x_j\ , \quad k = 1, \cdots, \ell, \cdots, m \qquad \text{(I-1)}$$

where rank$[\varphi] = \ell$, $\ell \leq r \leq m$. We can assume without loss of generality

$$\Delta = \begin{vmatrix} \varphi_{11} & \cdots & \varphi_{1\ell} \\ \vdots & & \vdots \\ \varphi_{\ell 1} & \cdots & \varphi_{\ell\ell} \end{vmatrix} \neq 0 \qquad \text{(I-2)}$$

Putting $x_{\ell+1} = \alpha_1, \cdots, x_r = \alpha_{\ell'}$, $\varphi_{1\ell+1} = \varphi'_{11}, \cdots, \varphi_{mr} = \varphi'_{m\ell'}$, we obtain from (I-1)

$$\sum_{j=1}^{\ell} \varphi_{kj}\, x_j = w_k - \sum_{j=1}^{\ell'} \varphi'_{kj}\, \alpha_j \qquad \text{(I-3)}$$

Solving (I-3), we get $x_1, \cdots, x_\ell$ each of which is the linear combination of the w_k and the α_j.

Let us, now, introduce the u_j $(j = 1, \cdots, \ell)$ satisfying the following

$$\sum_{j=1}^{\ell} \varphi_{kj}\, x_j = \sum_{j=1}^{\ell} \beta_{kj}\, u_j \qquad \text{(I-4)}$$

where rank$[\beta] = \ell$. We can put, for example, $\beta_{kj} = \varphi_{kj}$. We consider this case.

II From (I-3), we get

$$x_1 = \frac{1}{\Delta}\Big[\sum_{j=1}^{\ell} w_j A_{j1} - \sum_{k=1}^{\ell'} \alpha_k \sum_{j=1}^{\ell} \varphi'_{jk} A_{j1}\Big] \qquad \text{(II-1)}$$

where A_{jk} is the cofactor of the coefficient matrix of (I-3) about the j-th row and the k-th column. x_1 and α_1 are combined if and only if $\sum_{j=1}^{\ell} \varphi'_{j1} A_{j1} \neq 0$

Thus the theorem.

Theorem1: x_k $(k = 1, \cdots, \ell)$ and $x_{j+\ell}$ $(j = 1, \cdots, \ell')$ of (I-1) are combined on the transformation (I-4) if and only if

$$\sum_{i=1}^{\ell} \varphi'_{ij} A_{ik} \neq 0 \qquad \text{(II-2)}$$

III Among the column vectors of the coefficient matrix of (I-1) only the first ℓ column vectors are linearly independent. Therefore

$$\begin{bmatrix} \varphi'_{1j} \\ \vdots \\ \varphi'_{mj} \end{bmatrix} = C_{1j} \begin{bmatrix} \varphi_{11} \\ \vdots \\ \varphi_{m1} \end{bmatrix} + \cdots + C_{\ell j} \begin{bmatrix} \varphi_{1\ell} \\ \vdots \\ \varphi_{m\ell} \end{bmatrix} \quad (j = 1, \cdots, \ell') \qquad \text{(III-1)}$$

Seeing that

$$C_{kj} = \frac{\sum \varphi'_{ij} A_{ik}}{\sum \varphi_{ik} A_{ik}}$$

we have

$$x_k = \frac{1}{\Delta} \sum_{j=1}^{\ell} w_j A_{jk} - \sum_{j=1}^{\ell'} C_{kj}\, \alpha_j$$

Therefore $x_\ell + \sum_{j=1}^{\ell} C_{\ell j}\alpha_j = \frac{1}{\Delta}\sum_{j=1}^{\ell} w_j A_{j\ell}$

If we put $u_\ell = x_\ell + \sum_{j=1}^{\ell} C_{\ell j}\alpha_j$

then the transformation (I-4) is performed.

Thus the theorem.

<u>Theorem II:</u> If we put $u_\ell = x_\ell + \sum_{j=1}^{\ell} C_{\ell j}\alpha_j$

and
$$\beta = \begin{bmatrix} \varphi_{11} & \cdots & \varphi_{1\ell} \\ \vdots & & \vdots \\ \varphi_{m1} & \cdots & \varphi_{m\ell} \end{bmatrix}$$

under the conditions (I-2) and (II-1), then the transformation (I-4) can be performed.

IV Let us consider the system (1). From the above theorems we can put $H^*FX = \beta U$, where β is the set of independent column vectors of H^*F and U is vector whose element is $u_\ell = x_\ell + \sum_{j=1}^{\ell} C_{\ell j}\alpha_j$.

Then
$$\begin{bmatrix} W(T) \\ W(2T) \end{bmatrix} = \begin{bmatrix} H^*F & 0 \\ H^*GF & H^*F \end{bmatrix}\begin{bmatrix} X(0) \\ X(T) \end{bmatrix} = \begin{bmatrix} H^*F & 0 \\ H^*GF & \beta \end{bmatrix}\begin{bmatrix} X(0) \\ U(T) \end{bmatrix}$$

Since rank of $\begin{bmatrix} H^*F & 0 \\ H^*GF & H^*F \end{bmatrix}$ is equal to rank of $\begin{bmatrix} H^*F & 0 \\ H^*GF & \beta \end{bmatrix}$ and rank $\begin{bmatrix} HF & 0 \\ HGF & \beta \end{bmatrix} < r + \ell_1$ means the impossibility of determining X independently from U, we obtain the following.

<u>Theorem III:</u> The system (1) has the 2-nd observability for X(0) if and only if
$$\operatorname{rank}\begin{bmatrix} H^*F & 0 \\ H^*GF & H^*F \end{bmatrix} = r + \ell_1$$

V $G^n = K_1 G^{n-1} + \cdots + K_n I$ where $K_1, \cdots, K_n$ are constants

$$\therefore \quad f_{n+1} = Gf_n + F\gamma_{n+1} = G^{n-1}F(K_1\gamma_1 + \gamma_2) + \cdots + GF(K_{n-1}\gamma_1 + \gamma_n) + F(K_n\gamma_1 + \gamma_{n+1})$$

On the other hand the linear combination of $f_1, \cdots, f_n$ is
$$c_1 f_n + \cdots + c_n f_1 = G^{n-1}Fc_1\gamma_1 + \cdots + GF(c_1\gamma_{n-1} + \cdots + c_{n-1}\gamma_1) + F(c_1\gamma_n + \cdots + c_n\gamma_1)$$

Since rank $\begin{bmatrix} \gamma_1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \gamma_n & \gamma_{n-1} & \cdots & \gamma_1 \end{bmatrix} = n$, there are $c_1, \cdots, c_n$ such that
$$\begin{bmatrix} K_1\gamma_1 + \gamma_2 \\ \vdots \\ K_n\gamma_1 + \gamma_{n+1} \end{bmatrix} = \begin{bmatrix} \gamma_1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \gamma_n & \gamma_{n-1} & \cdots & \gamma_1 \end{bmatrix}\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

Therefore $f_{n+1} = c_1 f_n + \cdots + c_n f_1$

VI Since $G^n = K_1 G^{n-1} + \cdots + K_n I$ and $f_{n-1} = G^{n-1}\gamma_0 + \cdots + F\gamma_{n-1}$
$$f_n = G^{n-1}(K_1\gamma_0 + F\gamma_1) + \cdots + I(K_n\gamma_0 + F\gamma_n)$$

On the other hand the linear combination of $f_1, \cdots, f_{n-1}$ is expressed by
$$G^{n-1}\gamma_0 c_1 + \cdots + G(F\gamma_{n-2}c_1 + \cdots + \gamma_0 c_{n-1}) + I(F\gamma_{n-1}c_1 + \cdots + F\gamma_1 c_{n-1})$$

Since rank
$$\begin{bmatrix} \gamma_0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ F\gamma_{n-2} & \cdots & \gamma_0 \\ F\gamma_{n-1} & \cdots & F\gamma_1 \end{bmatrix} = n-1$$

there are $c_1, \cdots, c_{n-1}$ except the case $\gamma_1 = \cdots = \gamma_n = 0$ such that

$$\begin{bmatrix} K_1\gamma_0+F\gamma_1 \\ \vdots \\ K_{n-1}\gamma_0+F\gamma_{n-1} \\ K_n\gamma_0+F\gamma_n \end{bmatrix} = \begin{bmatrix} \gamma_0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ F\gamma_{n-2} & \cdots & \gamma_0 \\ F\gamma_{n-1} & \cdots & F\gamma_1 \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_{n-2} \\ c_{n-1} \end{bmatrix}$$

Therefore $f_n = c_1 f_{n-1} + \cdots + c_{n-1} f_1$

VII Since X(t) is continuous and finite, there is some M such that

$$X'(t) = X(t) + M \geq 0 \quad (\text{or} \leq 0)$$

Choosing such a proper M, we obtain

$$\begin{cases} Z'(t) = Z(t) + \int_0^t e^{A(t-\tau)} BM\,d\tau \\ Y'(t) = C^*Z'(t) = Y(t) + C^*\int_0^t e^{A(t-\tau)} BM\,d\tau \end{cases}$$

and

$$\begin{cases} Z' = AZ' + BX' \\ Y' = C^*Z' \end{cases}$$

References

1 Kalman, R.E. On the Generel Theory of Control Systems 1-stIFAC p481~492('60)
2 Gilbert,E.G. Controllability and Observability in Multivariable Control Systems J.SIAM.Controlvol.1 No.2 p128 ~ 151 ('63)
3 Chen,C.T.&C.A.Desoer Controllability and Observability of Composite Systems IEEE Trans. vol.AC-12 No.4 p402 ~ 409 ('67)
4 Chen,C.T. General Theory ofControllability and Observability of composite Systems JACC of U.S.A. p786 ~ 790 ('67)
5 Kreindler,E& P.E. Sarachik On the Concepts of Controllability and Observability of Linear Systems IEEE Trans. AC-9 No.2 p129 ~ 136 ('64)
6 Sarachink,P.E. & E.Kreindler Controllability and Observability of Linear Discrete-time Systems Inter.J, Control vol.1 No.5 p419 ~ 432 ('65)
7 Luenberger,D.G Observers for Multivariable Systems IEEE Trans. AC-11 No.2 p190 ~ 197 ('66)
8 Gilchrist,J.D. n-Observability for Linear Systems IEEE Trans. AC-11 No.3 p388 ~ 396 ('66)
9 Bona,B.E. Observability of Mean Values IEEE Trans. AC-12 No.4 p473 ('67)
10 Fujiwara,M. Differential and Integral Calculus (Book) ('34)
11 Ogata,K. State Space Analysis of Control Systems (Book) ('67)
12 Chidambara,M.R. & S. Ganapathy Transient Torques in 3-Phase Induction Motors during SwitchingOperation IEEE PAS April '62 p47-55
13 Smith,I.R. & S.Sriharan Transient Performance of the Induction Motor IEE Proc. Vol, 113 No.7 p1173 ~ 1181 ('62)
14 Kron,G. The Application of Tensors to the Analysis of Rotating Electrical Machinery G.E.Review vol.35 etc.
15 Lyon,W.V. Transient Analysis of Alternating Current Machinery (Book) ('54)
16 Miyairi,S. Energy Conversion Engineering (Book) ('65)
17 Takeuchi,J. Tensor Analysis of Electrical Machines (Book) ('63)
18 Matsumoto,M & K.Juri Speed of Induction Motor by Variable Frequency Inverter Tokyo Branch Congress of IEE of Japan ('63)
19 Gerecke,E. New Methods of Power Control with Thyristors 3-rd IFAC ('66)
20 Carli,A.D., M.Murgo &.A. Ruberti Speed Control of Induction Motor by Frequency Variation. 3-rd IFAC p 4C ('66)
21 Heumann.K. Variable Speed Control of Induction Motors 3-rd IFAC p 4D ('66)
22 Хасаев, О.И. Работа Асинхронного Двигателя От Преобразователя Частоты На Полупроводниковых Триодах ЭЛЕКТРИЧЕСТВО No.9 p29 ~ 36 ('61)
23 Комриков,К.П. Определение Рабочих Характеристик Асинхронного Двигателя При Частотном регулировании ЭЛЕКТРОМЕХАНИКА No.3 p277 ~ 284 ('62)
24 Jain,G.C. The Effect of V[illegible]ltage Wave-shape on the Performance of a 3-phase Induction Motors IEEE PAS-83 p561 ~ 566 ('64)
25 Racz, I Dynamic Behaviour of Inverter Controlled Induction Motors

3-rd IFAC p 48 ('66)

26 Jain,G.C. Analytical Study of the Step Starting and Step Running of a 3- Phase Induction Motor IEEE PAS-85 No.2 p93 ~ 105 ('66)

27 Takeuchi,J,A.MAEDA & S.Iida Characteristics of 3-Phase Induction Motor Controlled by 3-Phase Bridge-type SCR Inverter J.IEE of Japan vol. 88-2 No.953 p335 ~ 344 ('68)

28 Jordan,H.E. Analysis of Induction Machines in Dynamic Systems IEEE PAS-84 No.11 p1080 ~ 1088 ('65)

29 Krause, P.C. & C.H. Thomas Simulation of Symmetrical Induction Machinery IEEE PAS-84 No.11 p1038 ~ 1053 ('65)

30 Wyndham,C.H. & A.R.Atkins An Approach to the Solution of the Human Biothermal Problem with the Aid of an Analog Computer Proc. 3-rd Inter.Cong. on Med.Elect. ('60)

31 Wissler,E.H. A Mathematical Model of the Human Thermal System Bul. Math. Bio. vol. 26 p147 ~ 166 ('64)

32 Crosbie,R.J.,J.D.Hardy & E. Fessenden Electrical Analog Simulation of Temperature Regulation in Man IRE Trans. BME-8 No.4 p245 ~ 252 ('61)

33 Kawashima,Y., H.Yamamoto & M.Masubuchi A Model of Human Thermal Regulating System 7-th M.E. Conference of Japan ('68)

Captions for the figures

OBSERVABILITY OF LINAR DYNAMIC MEASURING SYSTEM

AND SOME APPLICATIONS

Takashi Sekiguchi

Yokohama National University

Yokohama, Japan

Fig.1: Dynamic measuring system.

Fig.2: The r inputs, m outputs and the n-th order discrete-time dynamic measuring system.

Fig.3: Non-observable vector and subspaces.

Fig.4: Real and equivalent values of an input and an output.

Fig.5: An open-loop problem.

Fig.6: Motor-load-measuring device system.

Fig.7: Transient torque behaviour measured by the torque meter.

Fig.8: $J_1\ddot{\theta}_1$ and $K\theta_1$ of Fig 7-(a).

Fig.9: $\frac{J_1(\ddot{\theta}_1)_{max}}{K(\theta_1)_{max}}$ to frequency characteristics (observed waves are smoothed in Fig 7).

Fig.10: Effet of $J_1\ddot{\theta}_1$ on sudden changes.

Fig.11: Relation between sampling period and natural frequency.

Fig.12: Measurement of the human thermal regulating system.

Fig.13: Kawashima-Yamamoto-Masubuchi's model for the human thermal regulating system.

Fig.1 Dynamic measuring system.

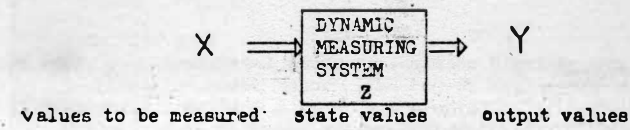


Fig.2 The r inputs, m outputs and the n-th order discrete-time dynamic measuring system.

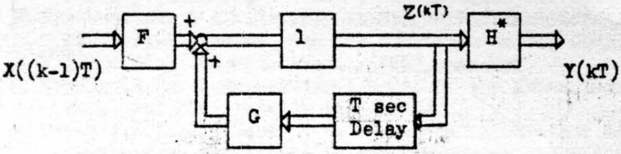


Fig.3 Non-observable vector and subspaces.

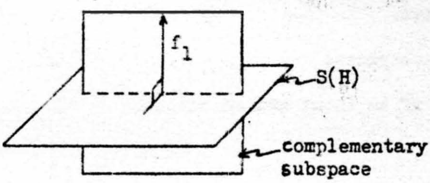


Fig.5 An open-loop problem.

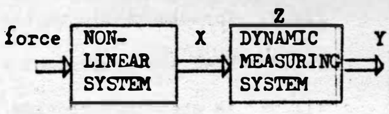


Fig.4 Real and equivalent values of the input and the output.

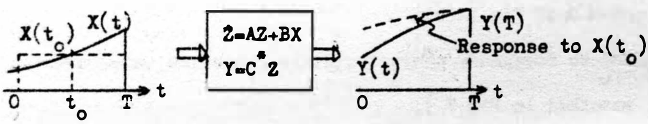
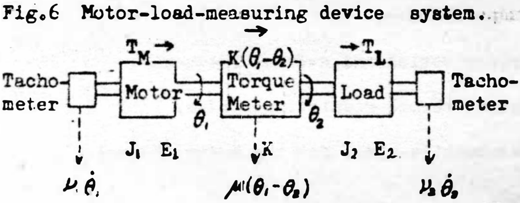


Fig.6 Motor-load-measuring device system.



J: moment of inertia

E: coefficient of friction loss

T_M: motor torque

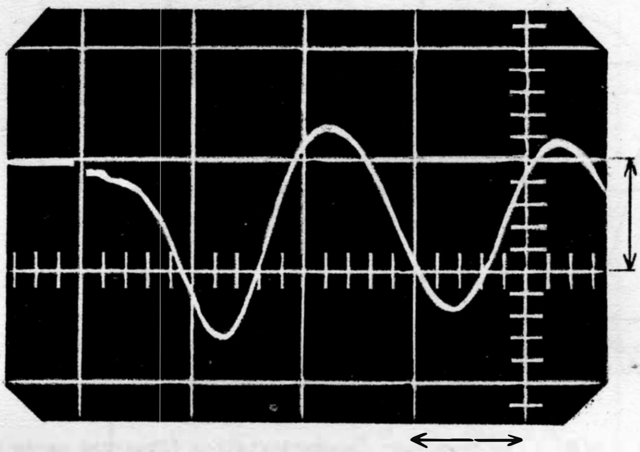
T_L: load torque

$K(\theta_1 - \theta_2)$: transmission torque

μ, μ_1, μ_2 : the coefficients of measurement

Fig.7 Transient torque behaviour measured by the torque meter.
Motor rated 0.2kw 200v 4P.
Torque meter rated 0.5x9.8 N-m.

(a) On switching-on of sine voltage (50Hz,130v).



(b) On switching-on of step-wise voltage (30Hz,320v peak to peak).

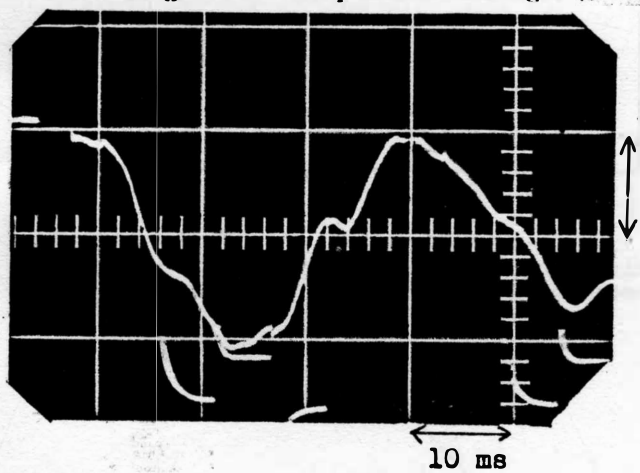
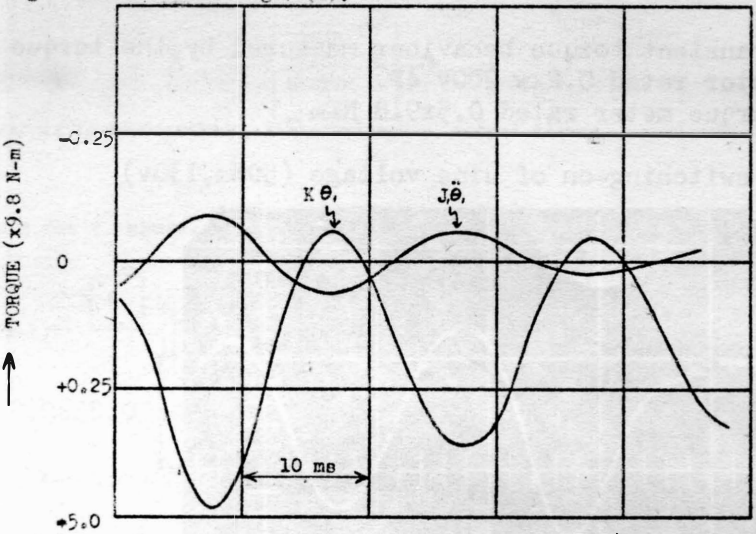
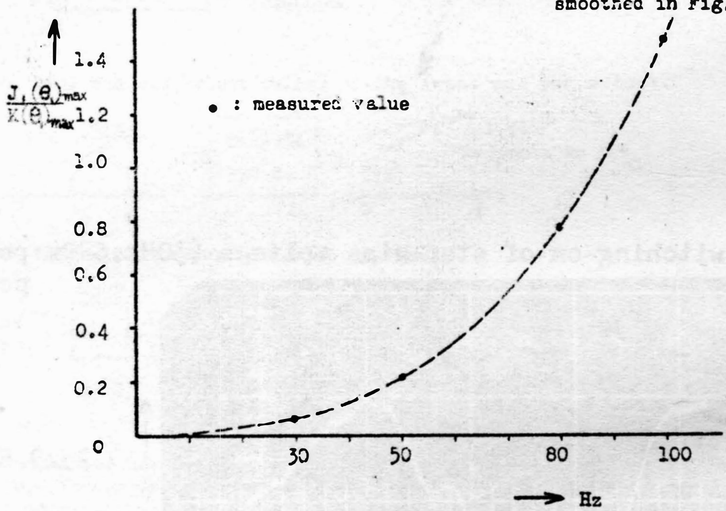


Fig.8 $J_1\ddot{\theta}_1$ and $K\theta_1$ of Fig.7-(a).



On switching-on of sine voltage (50 Hz,130V).
Motor rated 0.2kw 200v 1410rpm.

Fig.9 $J_1(\ddot{\theta}_1)_{max}/K\theta_{1max}$ to frequency characteristics (Observed waves are smoothed in Fig.7).



On switching-on of step-wise voltage (320V peak to peak).
Motor rated 0.2kw 200v 4P.

Fig.10 Effect of $J_1\ddot{\theta}_1$ on sudden changes.

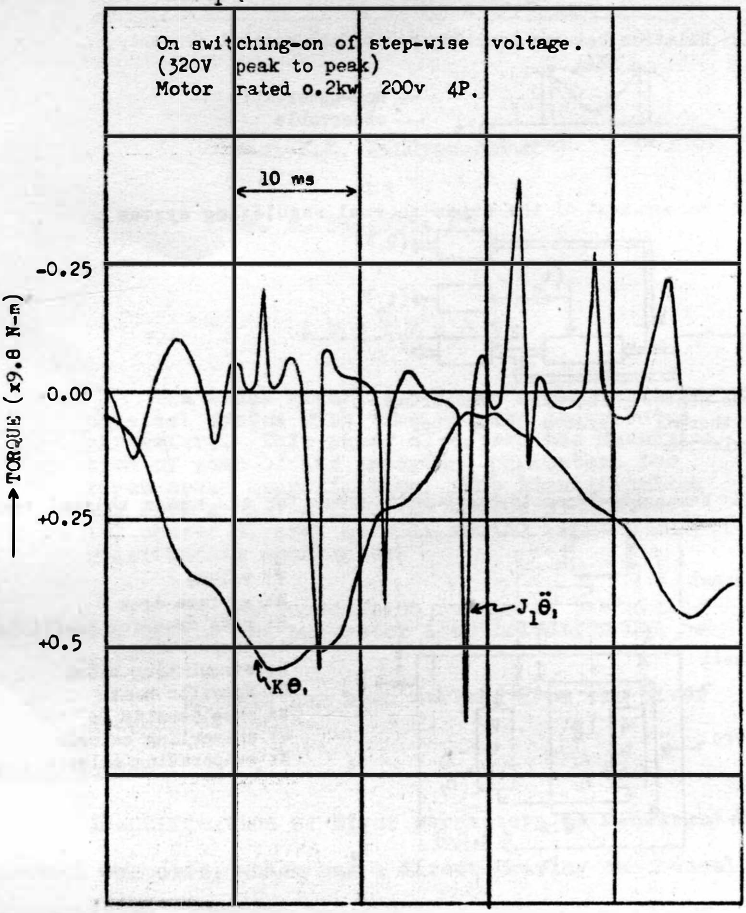


Fig.11 Relation between sampling period and natural freqency.

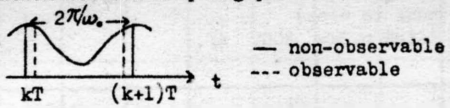


Fig.12 Measurement of the human thermal regulating system .

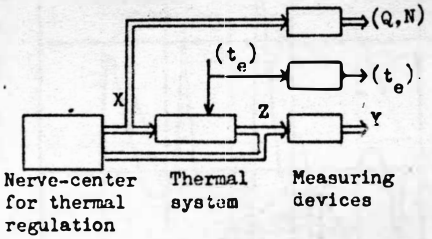
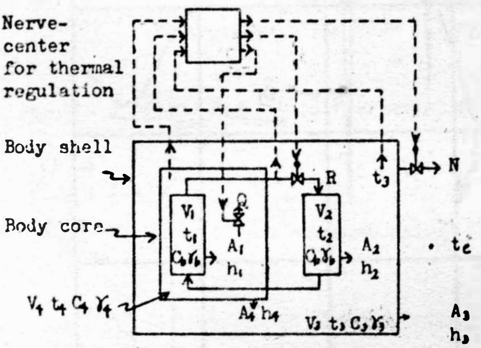


Fig.13 Kawashima-Yamamoto-Masubuchi's model for the human thermal regulating system.



V: volume
A: surface area
h: heat transfer coefficient
t: temperature
R: circulating blood
C: specific heat
γ: mass density
Q: shiverling calorie
N: evaporating calorie

IDENTIFICATION OF THE PARAMETERS OF A TIDAL CHANNEL FROM SIMULATION AND OTHER APPROACHES

- by -

Shri K.K. Bandyopadhyay *

and

Dr. S. Dasgupta **

Abstract

A tidal river is a large scale nonlinear physical system with time varying distributed parameters. This paper discusses the determination of some of its important parameters for first order approximation. The identification has been achieved from (a) analog simulation, (b) phasor diagram as well as (c) input - output relationship approaches.

An instrumentation scheme of the analog set up for the parameter identification has been given.

Example has been derived from the River Hooghly of India.

1. Introduction

Identification of River parameters is necessary for control purposes, which has a direct bearing on channel navigability.

A tidal river is a large scale nonlinear physical system with distributed parameters. Since the tidal influx changes from hour to hour, from day to day, from season to

* Senior Scientific Officer, Hydraulic Study Department, Calcutta Port Commissioners, Calcutta (India).

** Professor, Electrical Engineering Department, Jadavpur University, Calcutta (India).

season, these parameters are time varying as well. For exercising proper control, it is necessary to have some quick rough estimate of these parameters. There are methods in existence, which give estimates of these parameters, requiring unavoidable assumptions and lengthy iterations.

In the present paper new methods have been discussed for the determination of some important parameters based on measurement alone. In these methods, the estimation although is of first order approximation, it does not involve any assumption or iteration, whether simulated in an analog model or computed in a digital computer.

For the control of navigational channel ā priori knowledge of at least of the following parameters are necessary :-

(a) Phase and Range of the tide.

(b) System inertia (cross-sectional Area).

(c) System storage (storing width).

(d) Roughnesso of the bottom, presence of bends. bars, obstruction or training structures.

(e) The confluence and divergence of principal current streams.

(f) Shoaling etc.

To maintain and control the depths of the channel (i.e. the navigability), modification is exercised by dredging and training works.

Therefore, it is necessary to identify these parameters at least approximately from time to time.

The system parameters can be identified through (i) analog simulation where the differential equations that describe its dynamic behaviour, is the same as that in the prototype, (ii) the analysis of the phasor diagram of the system or (iii) the input - output relationship.

Example of the large scale physical system has been drawn from the tidal river Hooghly of India.

2.1. Identification from analog simulation

The simplified partial differential equations that describe the tidal motion in an estuary are :-

$$\left.\begin{aligned} &\frac{\partial Q}{\partial x} + b\,\frac{\partial h}{\partial t} = 0 \\ \text{and}\quad &\frac{\partial h}{\partial x} + \frac{1}{gbZ}\,\frac{\partial Q}{\partial t} + \frac{Q/Q}{C_c^2 b^2 Z^3} = 0 \end{aligned}\right\} \quad \ldots\ldots (1)$$

where
- Q : Discharge
- b : Width of the estuary
- Z : Average depth of the river section
- g : Accelaration due to gravity
- h : Water surface elevation
- C_c : Chezy's co-efficient
- x : Distance from the input end (source) along the channel
- t : Time

In tidal rivers, appreciable energy losses by dissipative forces, cause attenuation of the tide along the channel. These are losses, by friction along the bottom, form of sections, curvature and changing cross-sections in the channel. Seperate accounting makes the equations rather intractable. Therefore, all these losses are emerged into one equivalent resistance force assumed to be uniformly distributed along a reach. The estimation of this parameter again differs from schematisation to schematisation. The value of this co-efficient not only depends on the sand transport, salinity and suspended material in the flow but also on the depth of the channel.

Further the term $\frac{Q/Q/}{C_c^2 b^2 Z^3}$ renders the equation quasi-linear. Without much loss of accuracy the term may be represented linearly as :-

$$\frac{/Q/}{C_c^2 b^2 Z^3} = \frac{F_1}{gbZ}$$

$$\text{or} \quad F_1 = \frac{/Q/}{C_c^2 b Z^2} \; g.$$

Equation (1) can then be written as :-

$$\frac{\partial Q}{\partial x} + b\frac{\partial h}{\partial t} = 0$$

$$\text{and} \quad \frac{\partial h}{\partial x} + \frac{1}{gbZ}\frac{\partial Q}{\partial t} + \frac{F_1}{gbZ}Q = 0 \qquad \ldots\ldots (2)$$

where[1] $F_1 = (\frac{8}{3\pi})g \quad \frac{Q_m}{C_c^2 b Z^2}$ .... (2a)

and Q_m : Discharge amplitude.

2.2 The Tidal Estuary

In most estuaries, the width tapers off exponentialy, thus

$$b = b_o \varepsilon^{-\eta x} \qquad \ldots\ldots (3)$$

where

b_o : the width of the river at the input end

and η : a constant giving the taper rate.

Substituting eqn.(3) in eqn.(2) the general solution for sinusoidal input becomes :-

$$h_{(x,t)} = h_1 \epsilon^{(\frac{\eta}{2} - \alpha_1)x} \operatorname{Sin}(\omega t - \beta_1 x) + h_2 \epsilon^{(\frac{\eta}{2} + \alpha_1)x} \operatorname{Sin}(\omega t + \beta_1 x) \quad ..(4)$$

where

$$\alpha_1 + j\beta_1 = \sqrt{\frac{\eta^2}{4} + \gamma^2}$$

α_1 : the attenuation constant

β_1 : the phase constant

h_1 and h_2 are the two constants to be evaluated from the boundary conditions,

and

$$\gamma^2 = \left(\frac{F_1}{gb_o Z} + \frac{j\omega}{gb_e Z}\right) j\omega b_e$$

Clearly, the first part of the equation(4) represents, the incident wave while the second part, its reflected counterpart.

2.2.1 Analogous system

An analogous mathematical model may be observed in describing the propagation of low frequency electrical waves in exponentially tapered transmission line. The corresponding transmission line equations are[2] :-

$$\left.\begin{aligned} \frac{\partial i}{\partial x} + C\frac{\partial e}{\partial t} &= 0 \\ \text{and} \quad \frac{\partial e}{\partial x} + L\frac{\partial i}{\partial t} + Ri &= 0 \end{aligned}\right\} \quad .... (5)$$

where

$e_{(x,t)}$: the instantaneous voltage at any point.

$i_{(x,t)}$: the instantaneous current at any point.

C : the capacitance per unit length of line.

L : the inductance per unit length of line.

R : the resistance per unit length of line.

Comparing equations (2) and (5) one obtains :-

b, the river storage co-efficient is analogous to C, the capacitance.

$\frac{1}{gbZ}$, the river inertial co-efficient is analogous to L, the inductance.

$\frac{F_1}{gbZ}$, the river friction is analogous to R, the resistance.

h, the water surface elevation is analogous to e, the voltage.

Q, the discharge is analogous to i, the current.

Thus, considering the exponential taper :-

$$\left.\begin{aligned} b &= b_o\epsilon^{-\eta x} & &\triangleq C = C_o\epsilon^{-\eta x} \\ \frac{1}{gbZ} &= \frac{1}{gb_oZ}\epsilon^{\eta x} & &\triangleq L = L_o\epsilon^{\eta x} \\ \frac{F_1}{gbZ} &= \frac{F_1}{gb_oZ}\epsilon^{\eta x} & &\triangleq R = R_o\epsilon^{\eta x} \end{aligned}\right\} \quad \ldots. (6)$$

The corresponding solution for $e_{(x,t)}$ becomes :-

$$e_{(x,t)} = E_1\epsilon^{(\frac{\eta}{2} - \alpha)x}\,\text{Sin}(\omega t - \beta x) + E_2\epsilon^{(\frac{\eta}{2} + \alpha)x}\,\text{Sin}(\omega t + \beta x) \quad \ldots. (7)$$

$$\left.\begin{aligned} \text{where } \alpha + j\beta &= \sqrt{\frac{\eta^2}{4} + \gamma^2} \\ \text{and } \gamma^2 &= (R_o + j\omega L_o)\, j\omega C_o \end{aligned}\right\} \quad \ldots. (7a)$$

For long tidal channels, the reflected component is negligibly small[3].

Therefore we may write the corresponding electrical equation as

$$e_{(x,t)} = E_1 \epsilon^{(\frac{\eta}{2} - \alpha)x} \sin(\omega t - \beta x) \qquad \ldots. (8)$$

It is easy to derive the expression of current from the solution (8) and equation (5), with appropriate boundary conditions. Thus -

$$i_{(x,t)} = \frac{-C_o \omega E_1}{\sqrt{(\frac{\eta}{2} + \alpha)^2 + \beta^2}} \epsilon^{(-\frac{\eta}{2} + \alpha)x} \sin(\omega t - \beta x - \theta) \ldots (9)$$

where

$$\theta = \tan^{-1}\left\{\left(\frac{\frac{\eta}{2} + \alpha}{\beta}\right)\right\}$$

2.2.2 Determination of α

Measurement of amplitudes at points x_1 and x_2 on the transmission path from equation (8), is related by :-

$$l_n \frac{E_{(x_2)}}{E_{(x_1)}} = \left(\frac{\eta}{2} - \alpha\right)(x_2 - x_1) \qquad \ldots. (10)$$

Knowing $\frac{\eta}{2}$ from transmission channel (line) characteristics, α is determined.

2.2.3 Determination of F_1

From (7a) one obtains :-

$$\alpha = \sqrt{\frac{\omega^2 L_o C_o \left[\left(\frac{\eta^2}{4\omega^2 L_o C_o} - 1\right) + \sqrt{\left(\frac{\eta^2}{4\omega^2 L_o C_o} - 1\right)^2 + \frac{R_o^2}{\omega^2 L_o^2}}\right]}{2}} \qquad \ldots (11)$$

From the values of α, L_o, C_o and ω one obtains $\frac{R_o}{L_o}$ i.e. F_1.

2.2.4 Determination of β

Again from (7a) one obtains

$$\beta = \sqrt{\frac{\omega^2 L_o C_o \left[\left(1 - \frac{\eta^2}{4\omega^2 L_o C_o}\right) + \sqrt{\left(1 - \frac{\eta^2}{4\omega^2 L_o C_o}\right)^2 + \frac{R_o^2}{\omega^2 L_o^2}} \right]}{2}} \; ..$$

Knowing $\frac{R_o}{L_o}$, β is evaluated.

2.2.5 Determination of Q_m

From the values of α, β, η, ω and C_o, I_m is evaluated from eqn.(9). Again $I_m \doteq Q_m$ in its analog model scale.

2.2.6 Determination of Chezy's co-efficient

From the relationship (2a). The Chezy's friction co-efficient can be estimated for each section (reach) of the energy transmitting path.

3. Identification from the analysis of Phasor Diagram

One of the convenient methods of representation of sinusoidally time varying quantities is to represent them by Phasor[4]. For steady state analysis phasor diagram representation is a common practice in electrical engineering.

A large scale physical system transmitting energy, usually consisting of (i) Inertia, (ii) Storage and (iii) Friction parameters may also be identified separately from their respective phasor heads.

In the propagation of waves in a tidal channel :-

3.1 The inertial head phasor (in the absence of storage and friction).

The head loss due to inertia is given by :-

$$h_1 = \frac{1}{gA}\frac{dQ}{dt}$$

where

g : acceleration due to gravity.

A : area of cross-section of the channel.

t : time

Q : discharge

For a sinusoidal input

$$Q = Q_1 e^{j\omega t}$$

where Q_1 : discharge amplitude

$$h_1 = \frac{j\omega}{gA} Q \quad \text{..... (12)}$$

3.2 Storage head phasor (in the absence of inertia and friction)

Head loss due to storage $h_s = \frac{\int Q dt}{b}$

Where b: breadth of the channel

For a sinusoidal input

$$h_s = \frac{1}{j\omega b} Q \quad \text{.... (13)}$$

3.3. Friction head phasor (in the absence of and storage)

The head loss h_f due to friction (linearised) is given by

$$h_f = \frac{Q/Q/}{C_c^2 A^2 Z} \quad \text{.... (14)}$$

where C_c : Chezy's co-efficient

Z : Wetted average flow depth below Mean Tide Level.

A : bZ (the cross-sectional area)

It will be noticed from equations (12), (13) and (14) that :-

(a) the inertial head phasor leads the discharge phasor by 90°.

(b) the storage head phasor lags the discharge phasor by 90°,

and (c) the frictional head phasor is in phase with the discharge phasor.

When a tidal channel is subjected to a tidal input, the physical system reach by reach, for the first approximation be represented in the 'π'-mode formation, with one half of total channel storage per reach being lumped at each end of the section and total friction and inertial head being concentrated in series. From known gauge and discharge curves of each ends of a section, the complete phasor diagram could be constructed for identification of each parameters involved.

4. <u>Identification with the help of input-output relationship (four terminal network concept)</u>

Schönfeld's[5] idea of N-Gate system can be utilised for the analysis of a non-uniform long tidal channel.

The analysis of a non-uniform large physical system, can be carried out by dividing it into small sub-sections, each of which are individually homogenous. The input and output at the ends of each sub-sections are related by four constants generally known[6] as '' A,B,C,D constants''. Which identify the characteristics of the network. In this sort of identification the character of the intervening structure between input and output terminals is immaterial except in so far as it determines the input and output conditions.

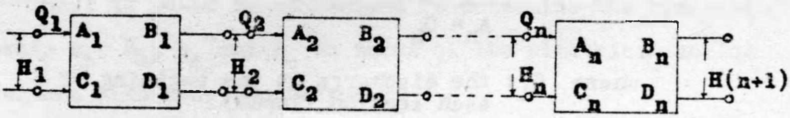


Fig.1

4.1 Estimation of "A,B,C,D, constants" in terms of input - output relationship.

For any section of a tidal reach if the elevation and discharge at the input and output ends be denoted by H_s, Q_s and H_R, Q_R respectively. They are related by[7].

$$H_S = A_h H_R + B_h Q_R \qquad \text{and} \qquad Q_S = C_h H_R + D_h Q_R \quad \ldots\ (15)$$

Where, after simplification and approximation one obtains :-

$$A_h = 1+\frac{1}{2}\left\{\frac{8}{3\pi}\,\frac{Q_1\left[1+\left(\frac{11}{4}\right)Z+5Z^2\right]l}{C_c^2A_o^2\ a_o} + j\,\frac{\omega l\left[1+\left(\frac{1}{2}\right)Z+\left(\frac{1}{2}\right)Z^2\right]}{gA_o}\right\} j\omega b_o l \quad \ldots\ (16)$$

$$B_h = -1+\frac{1}{4}\left\{\frac{8}{3\pi}\,\frac{Q_1\left[1+\left(\frac{11}{4}\right)Z+5Z^2\right]l}{C_c^2A_o^2\ a_o} + j\,\frac{\omega l\left[1+\left(\frac{1}{2}\right)Z+\left(\frac{1}{2}\right)Z^2\right]}{gA_o}\right\}(j\omega b_o l)\ x$$

$$\left\{\frac{8}{3\pi}\,\frac{Q_1\left[1+\left(\frac{11}{4}\right)Z+5Z^2\right]l}{C_c^2\ A^2_o\ a_o} + j\,\frac{\omega l\left[1+\left(\frac{1}{2}\right)Z+\left(\frac{1}{2}\right)Z^2\right]}{gA_o}\right\} \quad \ldots\ (1$$

and

$$C_h = -1+\frac{1}{4}\left\{\frac{8}{3\pi}\,\frac{Q_1\left[1+\left(\frac{11}{4}\right)Z+5Z^2\right]l}{C_c^2A_o^2\ a_o} + j\,\frac{\omega l\left[1+\left(\frac{1}{2}\right)Z+\left(\frac{1}{2}\right)Z^2\right]}{gA_o}\right\}(j\omega b_o l)(j\omega b_o l) \quad \ldots\ (18)$$

For symmetrical section

$$A_h = D_h$$

where Q_1 : the discharge at the begining of each section (input)

and $Z = \left(\frac{h_1}{a_o}\right)^2$

h_1 : tidal elevation at the begining of each section(input)

a_o : existing average depth of channel for that section

b_o : breadth of the flow channel

A_o : area of flow cross-section

C_c : Chezy's constant.

The equations become soi involved to account for the distortion of the wave form, the non linear friction being mainly responsible for it.

However for the first approximate values, these constants can be determined from theconsiderations of its electrical analog where[8]:-

$$A = D = \frac{E_s I_s + E_R I_R}{E_s I_R + E_R I_s} \quad \text{.... (19)}$$

$$B = \frac{E_s^2 - E_R^2}{E_s I_R + E_R I_s} \quad \text{.... (20)}$$

and

$$C = \frac{I_s^2 - I_R^2}{E_s I_R + E_R I_s} \quad \text{.... (21)}$$

Thus, it would be convenient to determine the hydraulic constants A_h, B_h, C_h and D_h in terms of the electrical analog counterpart with some limitations.

As already mentioned the system is divided into suitable sub-sections. Measurements are made at each junction of the consecutive sub-section. In other words, measurements concerning gauge heights H and Discharges Q (i.e. voltages E and currents I in its electrical transmission line analog) are made available. These give similar general circuit as per equations (19) (20) and (21) obtained after replacing, E by H and I by Q.

If an approximate lumped parameter equivalent is assumed in the 'π' mode, then

$$A = D = 1 + \frac{YZ}{2}$$

$$B = Z$$

$$C = Y\left(1 + \frac{YZ}{4}\right)$$

Figure 2 shows such an approximate network. The electrical analog and its hydraulic counterpart has been shown in figs. 5 and 6.

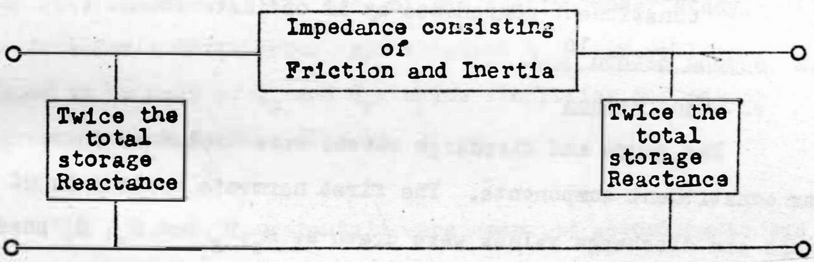


Fig.2

5. Procedure

The River Hooghly (of India) which is the navigable waterway for the Port of Calcutta (Fig.3), has been used as an example for all the three methods of parameter identification described above.

5.1 First Method

The survey data has been used for finding out, the taper rate ''η'' (Fig.4). It exhibits two distinct zones of taper rates 0.05/n.miles (Sec.A, Fig.4) between seaface (Saugor) and the first main branching off (Hooghly Point), whereas in the second zone this rate attains a value 0.0225/n.miles (Sec.B, Fig.4) and remains nearly constant till the end of the tidal compartment, with a transition zone in between.

The theory has been applied for an average tide (27.4.1955) under the following restraints :-

a) there is no upland discharge

b) the input tide is sinusoidal

c) there is no sudden change in depth or in width

d) the first harmonic component of the tidal elevation gauge curves for each station was used (The wave form was broken up into its constituent components by 12 ordinate scheme)[9].

5.2 Second Method[10]

A. Exact Method

The gauge and discharge curves were broken up into its constituent components. The first harmonic components of gauge and discharge values were drawn as H_s, H_r and Q_s, Q_r phasors respectively for eache ends of any section (Fig.7).

Referring to Fig.7 the phasor $Q_{cd} = H_s(\frac{j\omega bl}{2})$ for half the lumped storage of ''π'' - mode and is drawn being at 90° with H_s.

The subtraction of phasor Q_{cd} from Q_s brings out the phasor Q_{eq_1}, i.e. the effective discharge flowing in the reach, which encounters series friction as well as inertia.

From phasor tip A of the phasor H_s, a line AP is drawn parallel to Q_{eq_1} and from the line AP another perpendicular NB is drawn, so that it meets the phasor H_r at B.

The head loss AN in its proper scale, gives the friction drop (being in phase with Q_{eq_1}) and NB gives the inertial drop (being in quadrature with Q_{eq_1}) as already described. This procedure was repeated reach by reach (Fig.8) for identification and quick assessment of the channel friction. The method was applied for the same river and under the same restraints of the preceding method.

B. Approximate method

In tidal flow analysis, for all practical purposes, the discharge for a section is assumed to be mean of the discharges observed at the ends of the section under study. Therefore for a first order approximation Q_{eq_1} may well be assumed to be mean of Q_s and Q_r, which simplifies the phasor diagram approach further (Fig.9).

5.3 Third Method [8]

The ''A,B,C,D constants'' were computed according to the equation sets, (19), (20) and (21) from the first harmonic components of the gauge and discharge curves of the respective stations.

The river length being divided into corresponding four terminal network blocks (Fig.10).

6. **Results**

A. Table I shows, values of Chezy's constant as obtained by :-

a) Tapered transmission line analog

b) Exact phasor diagram approach

c) Approximate phasor diagram approach

The values have been compared with those obtained by the existing methods of :-

a) Iteration

and e) Phase measurement.

B. Table II shows, the identification of ''A,B,C,D constants'' by :-

a) Four terminal network concept

b) Hydraulic computation method.

They have been compared with the values obtained by :-

c) Tapered Transmission line concept

and d) River Geometry analysis.

7. Analog computation of ''A,B,C,D constants'' and Y,Z parameters in approximate ''π'' or ''T'' mode.

Fig.11 shows possible simple analog computation scheme to facilitate the determination of the '' A,B,C,D'' constants'' from the input data of gauge (tidal elevation) and discharge information of the River.

Analog computation scheme with the help of this power meter can measure Real or Imaginery components of a product of two complex quantities, it is necessary to introduce the complex conjugate of E_2 and E_4 so that complex products are performed. The synchros are set for the phase angles and the gain of the amplifiers are set for the respective magnitudes, then the value of $E_1 . \dot{E}_2 + E_3 . \dot{E}_4$ is obtained in the form a + jb measuring the Real power and the Imaginary power by properly switching S_2 for addition and then flipping switch S_1 back and forth for the real and imaginary components respectively. $E_1 E_2 - E_3 E_4$ is similarly measured by putting S_2 in proper position. $E_1^2 - E_2^2$ is obtained by setting the same values for E_1 and E_2 and for E_3 and E_4. A set of simple division then gives the values of "A,B,C,D constants" from equqtions (19) (20) and (21). It is also possible to determine the respective values of Y and Z for either the " π " or the "T" mode, by suitably selecting E_1, E_2, E_3 and E_4.

8. <u>Conclusion</u>

1) In the first method, the estimation of the greatest uncertain tidal channel parameter i.e. Chezy's co-efficient is much simpler compared to its existing counterpart.

The results are close to the standard method at least for the first order accuracy.

2) In the second method the representation gives a clearer picture of the system and hence clearer identification of the parameters.

For rapid schematization of a tidal river this method would be a powerful tool.

3. In the third method the identification of ''A,B,C,D constants'' - could be achieved in a quick and simple way even without the river survey data.

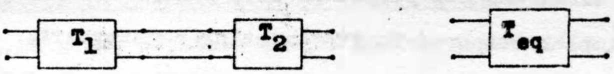
9. <u>Discussion</u>

Even when an exact identification is necessary, the initial identification by the above mentioned procedures should cut down the lengthy procedure considerably.

The tidal wave form becomes more and more distorted with the progress of the wave into upper tidal compartment. In that compartment the more realistic representation of the wave form would need more number of harmonic components. The first order approximate values derived with the help of fundamental therefore differs from the actual.

The transfer matrix for composite constants of an estuary with complex formations can easily be evaluated from the knowledge of the ''A,B,C,D constants'' of the individual sections. Typical formations are given below:-

(a) <u>When two channels are placed in tandem</u>



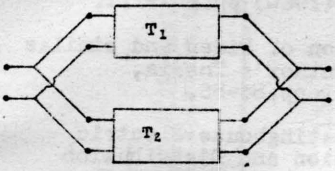
where $T_1 = \begin{vmatrix} A_1 & B_1 \\ C_1 & D_1 \end{vmatrix}$

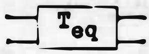
and $T_2 = \begin{vmatrix} A_2 & B_2 \\ C_2 & D_2 \end{vmatrix}$

Equivalent four terminal network corresponding to the system where -

$$T_{eq} = \begin{vmatrix} A_1A_2+B_1C_2 & A_1B_2+B_1D_2 \\ C_1A_2+D_1C_2 & C_1B_2+D_1D_2 \end{vmatrix}$$

(b) When two channels are placed in parallel :-





Equivalent four terminal network corresponding to the sytem

where $T_1 = \begin{vmatrix} A_1 & B_1 \\ C_1 & D_1 \end{vmatrix}$

and $T_2 = \begin{vmatrix} A_2 & B_2 \\ C_2 & D_2 \end{vmatrix}$

where

$$T_{eq} = \begin{vmatrix} \dfrac{A_1B_2 + B_1A_2}{B_1 + B_2} & \dfrac{B_1B_2}{B_1 + B_2} \\ C_1 + C_2 + \dfrac{(A_1 - A_2)(D_2 - D_1)}{B_1 + B_2} & \dfrac{B_1D_2 + D_1B_2}{B_1 + B_2} \end{vmatrix}$$

Thus, for tackling problems of tidal channels with branchings, this method of identification would be a good tool.

10. Acknowledgement

The authors are indebted to Prof. H.C.Guha, Vice-Chancellor, Jadavpur University and Sri B.B.Ghosh, Chairman, Calcutta Port Commissioners for providing all the facilities for carrying out the work. The authors owe a debt of gratitude to Dr. S.K. Bhattacharya, Chief Hydraulic Engineer, Calcutta Port Commissioners' office for his suggestion, discussion and encouragement during the progress of the work.

11. Bibliography

1 Dronkers, J.J. - Tidal computations in river and coastal waters. North Holland Publishing Co., Amsterdam(1961) pp.226.

2 Jhonson, W.C. - Transmission lines and Network Kogakusha Co.Ltd., Tokyo (1950) pp.8-9

3 Bandyopadhyay,K.K., N.C.Mazumdar and S.Dasgupta. - " A method of Estimation of Chezy's co-efficient in Tidal River" Journal of the Institution of Engineers (India), vol.48, No.6 (Part EL-3) Feb.1968, pp.406.

4 Kerchner, R.M. and G.F. Corcoran - "Alternating Current Circuits" John Wiley and Sons Inc., New York (1960) pp.107-111.

5 Schönfeld J.C., - Propagation of Tides and similar waves - Author's Thesis, Delft 1951 pp.84-85.

6 Central station Engineers of the Westinghouse Electric Corporation - Electrical Transmission and Distribution Reference Book - Oxford and IBH Publishing Company - Calcutta (1950)pp.266.

7 Dronkers, J.J., - Tidal Computation in rivers and coastal waters North Holland Publishing Co., Amsterdam (1964) pp.259-260, 278.

8 Bandyopadhyay,K.K., and S.Dasgupta - " Analysis of Tidal Channel with the help of Electrical four terminal Network concept" - Communicated for Publication in the Journal of the Institution of Engineer, India, 1968.

9 Scarborough, J.B., - " Numerical Mathematical Analysis " Oxford University Press (1962), London, pp.558-568.

10. Bandyopadhyay, K.K., - Phasor diagram representation of Tidal channel. Presented at the "12th Congress on Theoretical and Applied Mechanics", New Delhi, India, 1967.

*****

TABLE - I

Value of Chozy's constant (in ft.$^{\frac{1}{2}}$/sec.) as obtained by different methods.

Between Section.	Tapered Transmission line simulation method.	Interation method.	Phase measurement method.	Phasor diagram Analysis method	
				Exact	Approximate
Saugor and Gangra	176	144	193	163	150
Gangra and Balari	150	133	124	130	129
Balari and Diamond Harbour	133	135	111	95	100
Moyapur and Akra	196	117	82	92	101
Akra and Garden Reach	136	93	79	124	140
Garden Reach and Konnagar	101	112	84	87	81
Konnagar and Mulajore	108	112	87	82	76
Mulajore and Bansberia	97	90	77	111	110

TABLE - 11

Comparison of the values of A and B constants by different methods

Station	Four terminal network concept		By Hydraulic Computation		By Transmission Line concept		By river Geometry analysis (electrical equivalence concept)(determination of Z_A, Z_B of Fig.2)	
	A	B	A_h	B_h	A	B	A	B
F/man's Pt. to Moyapur	0.9876+j0.0293	3.67+j2.89	0.9695+j0.0464	3.93+j2.45	0.9707+j0.0304	-	0.9697+j0.02895	2.46+j2.57
Moyapur to Akra	0.9263+j0.0712	4.35+j4.90	0.9386+j0.1507	11.05+j3.53	0.9414+j0.0585	-	0.9346+j0.0605	4.21+j4.53
Akra to G.Reach	0.9919+j0.0139	3.25+j2.85	0.9869+j0.0336	6.61+j2.55	0.9889+j0.0151	-	0.9880+j0.0155	2.78+j2.16
G.Reach to Konnagar	0.9292+j0.1040	16.15+j9.65	0.9231+j0.1121	10.74+j6.35	0.9266+j0.1523	-	0.9192+j0.1586	13.79+j7.01
Konnagar to Mulajore	0.9777+j0.1284	13.35+j7.45	0.9354+j0.1104	15.33+j7.83	0.9386+j0.1286	-	0.9346+j0.1323	16.44+j8.12
Mulajore to Bansberia	0.9517+j0.0899	-	0.9502+j0.1566	30.37+j9.18	0.9490+j0.1558	-	0.9461+j0.1612	26.89+j9.02

MAP OF RIVER HOOGHLY

TIDAL GAUGE LOCATIONS

SCALE :- 1" = 16 MILES

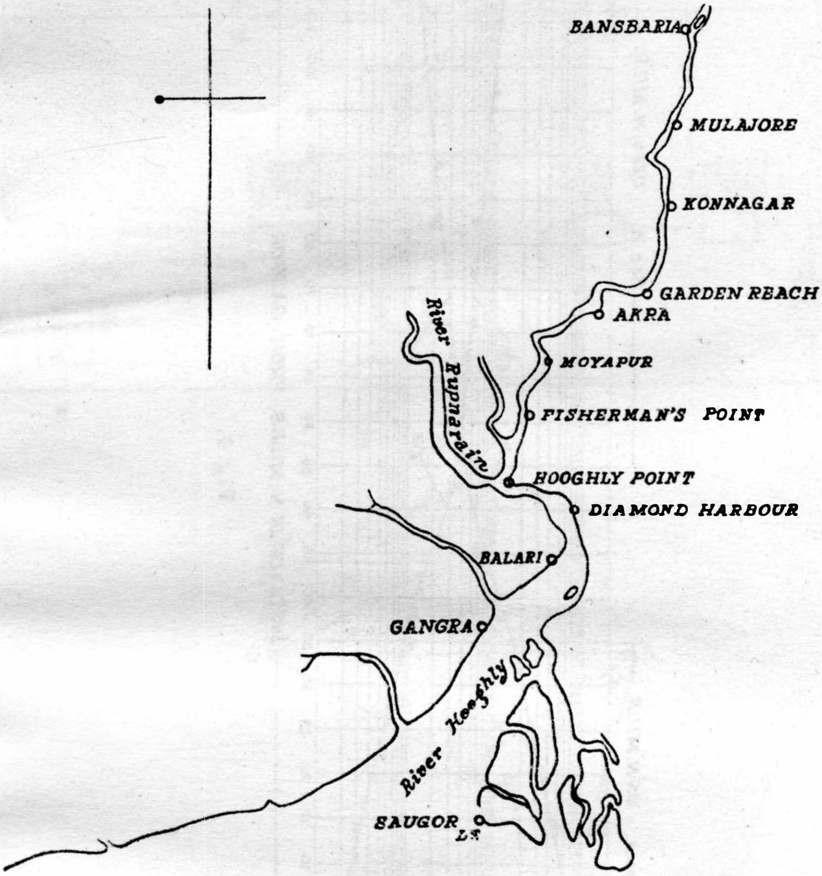


Fig -3.

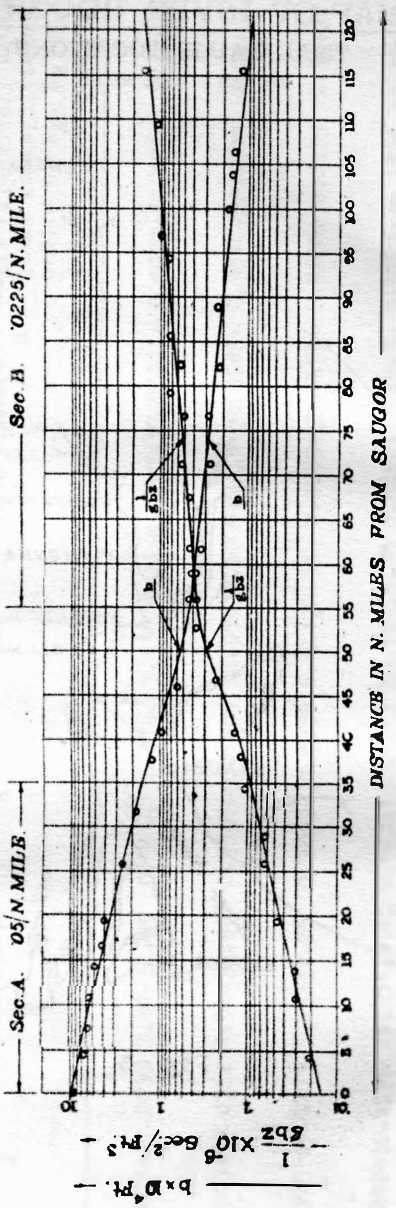

Sec. A. ·05/N. MILE
Sec. B. ·0225/N. MILE
DISTANCE IN N. MILES FROM SAUGOR
b
1/gbz

Fig. 4.

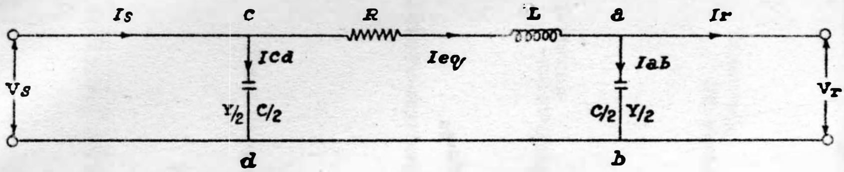


Fig-5.

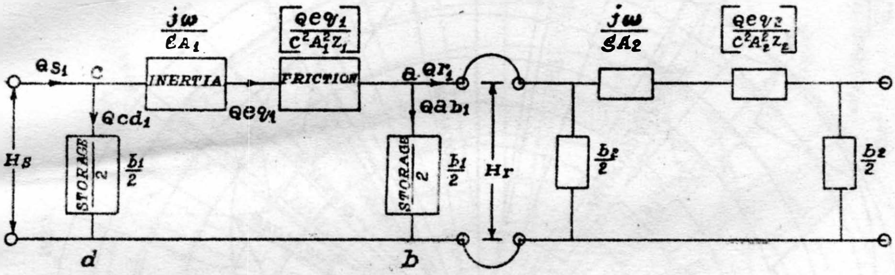


Fig-6.

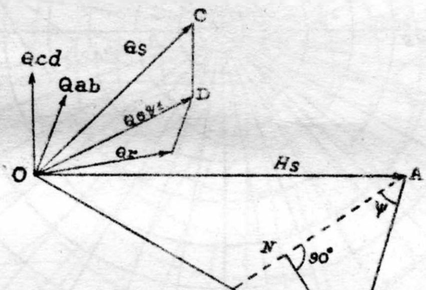
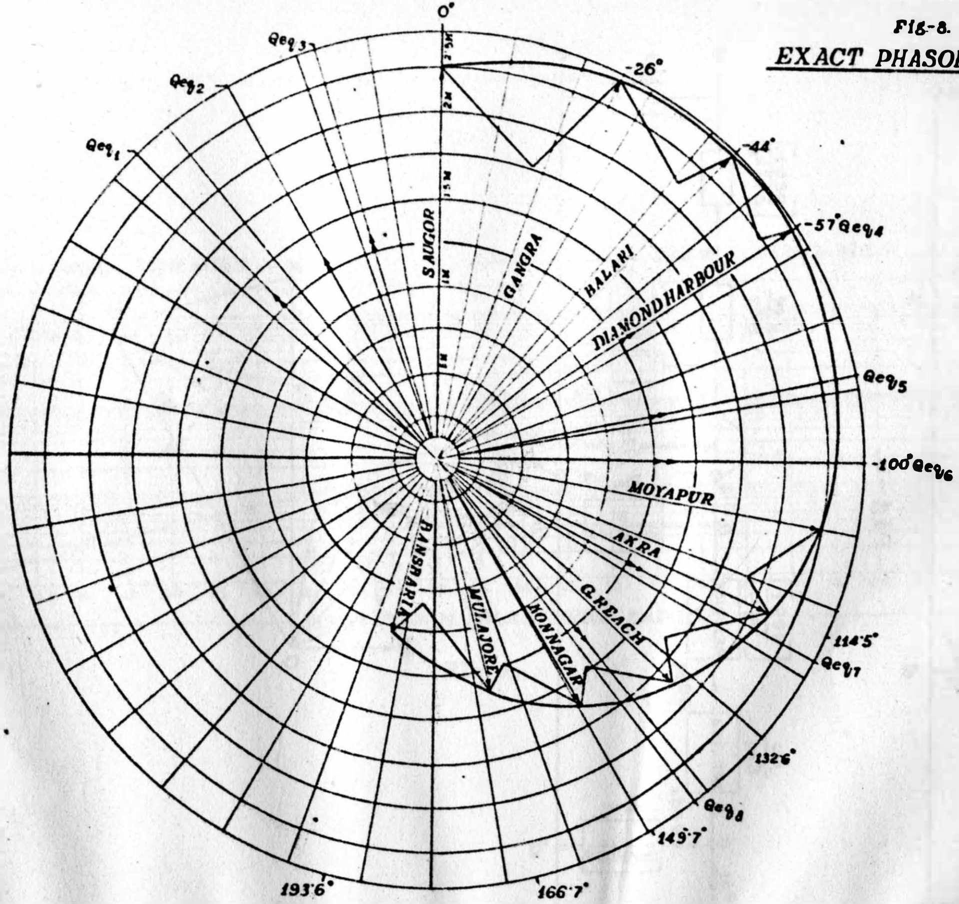


Fig-7.

Fig-8.

EXACT PHASOR DIAGRAM



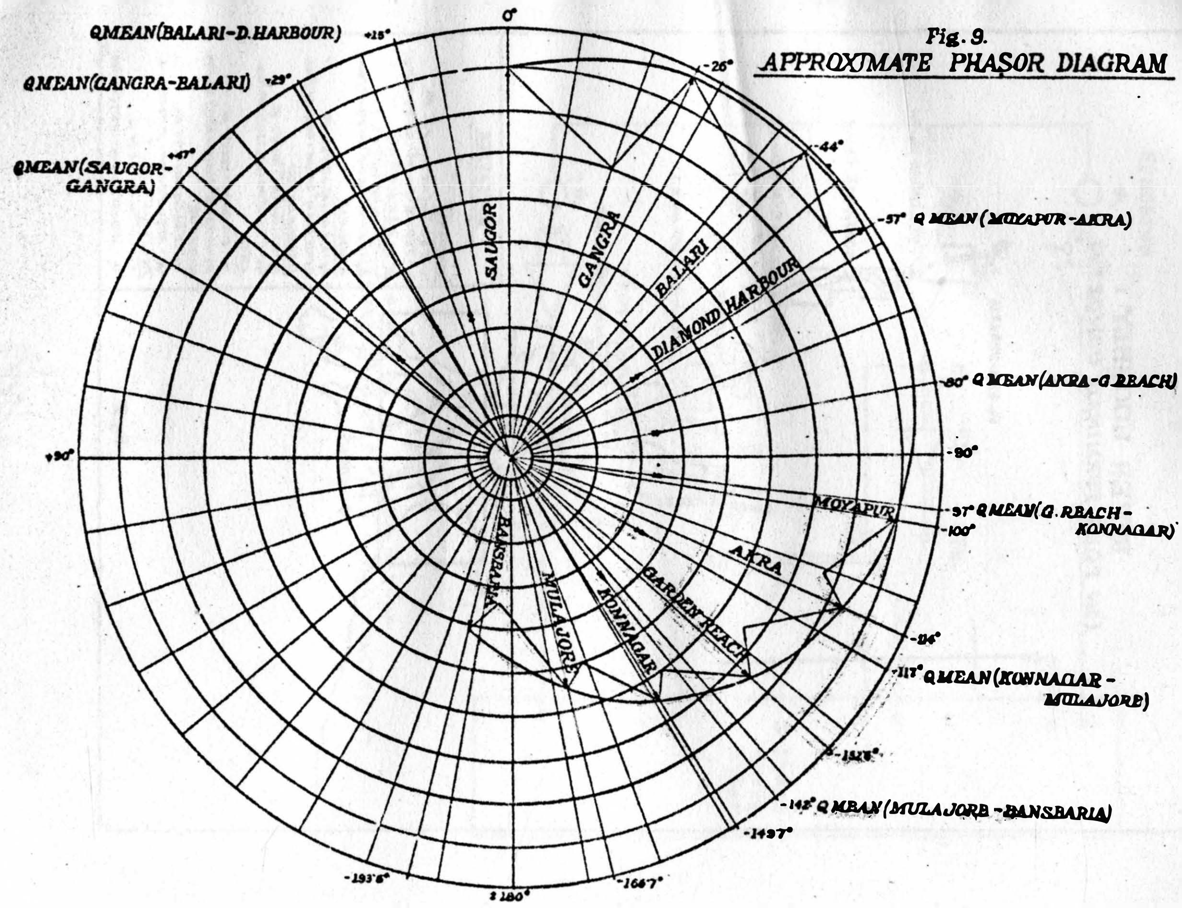


Fig. 9.
APPROXIMATE PHASOR DIAGRAM

RIVER HOOGHLY
(IN FOUR TERMINAL CONCEPT)

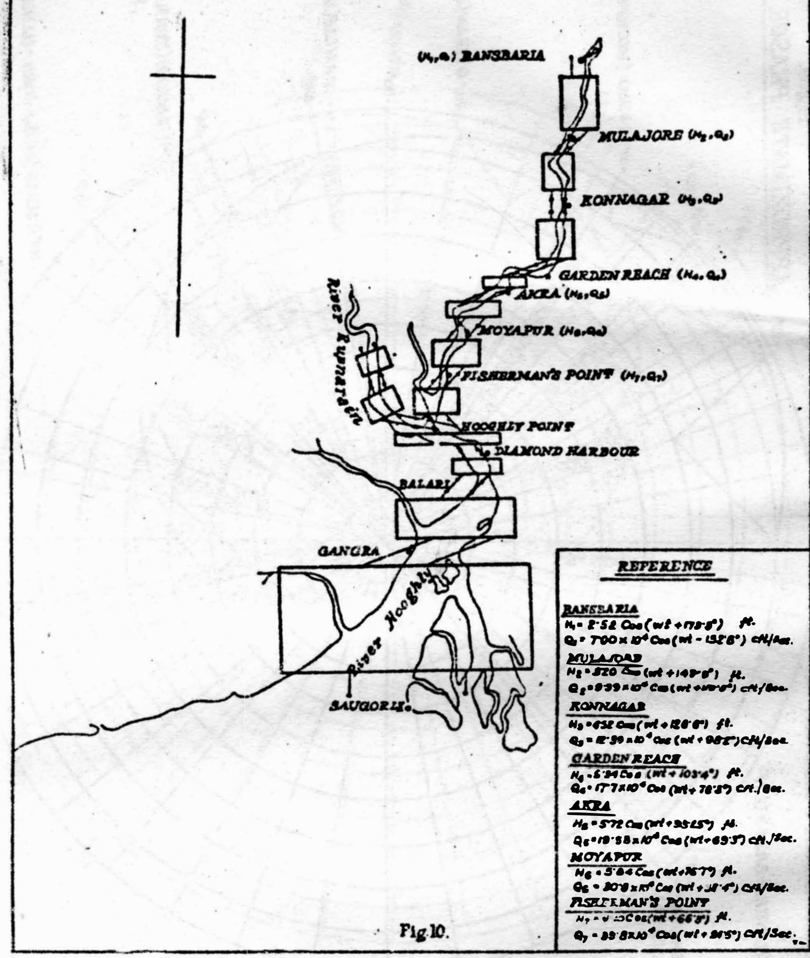


Fig.10.

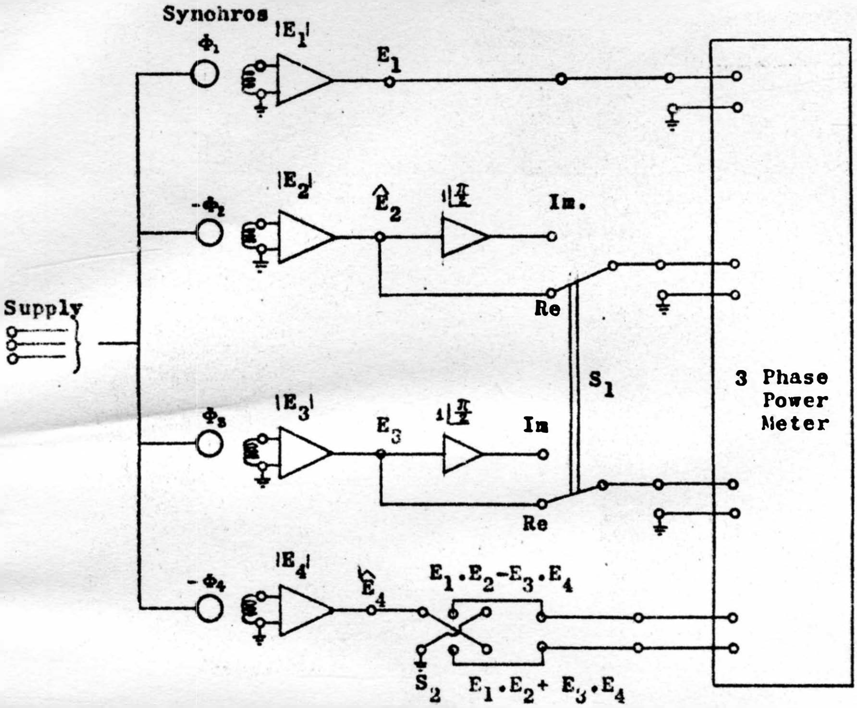


Fig.11
Schematic diagram of an Analog set-up for the determination of "A,B,C,D - constants"