IDENTIFICATION OF NONLINEAR SYSTEMS THROUGH QUASI-WHITE TEST SIGNALS

Thesis by

Vasilis Zissis Marmarelis

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1976

(Submitted October 29, 1975)

ACKNOWLEDGMENTS

I am grateful to Professors Gilbert D. McCann, my advisor, and Thomas K. Caughey for their guidance, stimulating discussions, support and concern that made this work a rewarding intellectual experience rather than an educational burden.

I am grateful to Dr. Panos Z. Marmarelis, my brother, whose insight, inspiration and encouragement provided the early stimulation for my involvement in this work.

I want to thank D. Knutsen and S. Tappel for greatly appreciated assistance in dealing with the computational aspects of my work.

I would like to express my thanks and appreciation to R. Fargason and G. Lyon for valuable help with the experiments, and to D. Aranovich and B. Ellert for technical assistance.

I want to express my deep appreciation to S. Booth and I. Stork for their excellent work in preparing this manuscript, and to E. Johnson for general secretarial assistance.

Finally, I thank the California Institute of Technology for supporting me financially throughout my graduate years, and for providing me with an excellent research environment and facilities for my work. I want to extend my thanks to the National Science Foundation for supporting through grants the biggest part of this research effort.

ABSTRACT

The use of quasi-white test signals, i.e. physically realizable signals that approximate the statistical properties of ideal white noise, in nonlinear system identification through the crosscorrelation technique is comprehensively studied. Important theoretical aspects of the subject are illustrated (e.g. the mathematical mechanisms of kernel estimation through crosscorrelation, the role of the several orthogonal functional series, the meaning of the corresponding kernels, the accuracy of the obtained truncated models etc.), and useful tools for the actual application of the method are developed (e.g. analytical expressions for the kernel estimation errors, optimum test procedure etc.).

In addition to the widely known and used band-limited gaussian white noise and pseudorandom signals based on m-sequences, a new family of quasi-white test signals is introduced and its properties are thoroughly studied. The various advantages and disadvantages of these three families of quasi-white signals are discussed independently as well as in a comparative perspective. The accuracy of the several estimated models is found to be comparable for all these families of quasi-white signals, with small differences pertaining to the specific system under study or random factors. The theoretical study is followed and confirmed by actual applications on computer simulated and physiological systems.

The newly introduced family is simplifying, clarifying and

unifying the concept of the quasi-white signal in connection with its use in the crosscorrelation technique.

Some special purpose tests are also presented, with one of them (the "general nonlinearity test") possessing the potentiality of a totally different identification method aiming at the Volterra kernels of the system.

TABLE OF CONTENTS

Table of Contents (cont.)

Table of Contents (cont.)

CHAPTER I

INTRODUCTION

The perception of the surrounding phenomena as the observable expressions of a complex invisible machinery, which conceptually takes the form of a network of interconnected and interdependent variables, is fairly common and popular among physical scientists. The mechanistic view of the physical world and the causal chain of physical phenomena evolvement seem also to be the common stand of physical sciences. In this regard, the system formalization of a physical relation comes as a natural and suitable approach, provided, of course, that the scientific means exist to tackle the problem under this formulation. This is the task of system science and the general frame in which the present dissertation places itself.

The system conceptualization of the surrounding world has a philosophical quality that can hardly be bypassed. For this reason, the first two sections attempt to give a brief, concise account of the philosophical ideas involved. We expect these philosophical inferences to be more or less argumental, but we also consider this philosophical framing an introductory necessity of the subject.

1.1 The Philosophical Frame of System Conceptualization

The knowledge that the human kind possesses is the whole of its organized and expressible experience.

Experience is the natural result of the interaction between the human nervous system and the surrounding physical world.

Man experiences the surrounding physical world through his

senses and elaborates on the obtained experience creating intellectual states within his brain, which correspond to pieces of this experience. Further, the human brain exhibits the natural ability of constructing relational intellectual structures from these intellectual states, which depict the perceived experience at some area and level of conception.

This is a self-changing function, in which the operator is the operand as well. The reason is that every new intellectual construction can be subsequently used to create the next intellectual structure. This process is naturally leading towards an intellectual depiction of the perceived physical experience, which serves best the natural purpose of existence of the human intellect. We cannot state categorically what this natural purpose is; however, we know that it provokes the growth of these intellectual structures. In that respect, this is an evolutionary process, where the driving motor is a natural tendency of the human brain to increase and improve its intellectual structures.

This process seems to be self-sustained in a monotonic course (within the limits posed by natural constraints of the carrying physical space) and exhibits the same basic characteristics of the evolution of species through natural selection. At the present stage of being, it seems that the evolution of the human intellectual capacity (that is, the whole of the intellectual structures in his brain) is a more rapidly self-accelerated process than the evolution of species; nevertheless, this may be a temporary situation, reinforced by the contemporary parameters of the carrying environment . A striking realization here, is that whatever the conception of the evolution of human intelligence,

this conception is inevitably by itself still one more expression of the present stage of this same process.

Being inescapably locked within the limits of our physical perceptual ability, we can do nothing more than play with the intellectual images of our subjective perception of the surrounding world. These intellectual images are the concepts, corresponding to the various pieces of experience that the human brain happened to elaborate upon, through the long journey of the human intellectual evolution.

The concepts are intellectual images of the perceived information through our senses from the surrounding physical world. From the natural process of forming these concepts, we can distinguish two principal categories. The first category includes those concepts, which are self-existent as intellectual images of a perceived entity; while, the second category includes those concepts, which attain an intellectual content with respect to two or more perceived entities. For example, watching a brown shoe we first form the concept "shoe", then we want to elaborate more on this piece of perceived visual information and we form the concept "brown", then we make the intellectual conjunction that "the shoe is brown". The concept "is" does not attain any conceptual content by itself. But it does provide an intellectual link between the concepts: "shoe" and "brown". Therefore, it does have an intellectual content with respect to these two concepts. Of course, there is a great variety of ways to analyze this piece of visual experience but, in any case, the brain will follow a similar pattern of formalization with concepts corresponding either to entities or relations. These are

respectively the two categories of concepts.

We understand that this subject has a philosophical quality, in the sense that truth cannot be rigorously proved but can only be portrayed as such from a composition of indications that usually hold strong in the eyes of the beholder. Any formalization of this sort, is a practicable way that a brain follows in order to handle the perceived information. We can never claim that it reliably or uniquely maps an objectively and undeniably true physical order.

In the following, we only use the assertion that concepts exist as intellectual images of pieces of experience, and that there are two categories of concepts,the entities and the relations. The entities are intellectually self-sustained; while the relations refer to two or more entities.

It must be emphasized that this formalization corresponds to a certain state of human intelligence and consequently it has an inherent character of change. As the human intellectual capacity evolves, it is expected that a specific intellectual arrangement of entities and relations will be extended in number and complexity.

For one more reason, the human brain possesses the natural ability of creating concepts, which do not correspond directly to pieces of physical experience, but they are the products of an intellectual procedure (abstraction) based upon the already existing concepts and structures within the brain. This results in a powerful intellectual capacity to conceptualize an entities-relations arrangement in a variety of ways by introducing abstract concepts.

On the basis of this abstractive capability, a single concept (entity or relation) may as well be intellectually analyzed in an interconnected complex of entities and relations. We call this property of the human intellect, "conceptual extensibility", and it must be emphasized that it is of fundamental importance in the system conceptualization.

1.2 General Foundation of System Theory

At a given point of the evolution of human intelligence, there is a concrete way in which the human experience is formally organized and expressed. This is the human knowledge at this point of time.

The formal expression of this knowledge is done through the language. The basic structural (syntactical) elements of the language are the nouns, the adjectives and the verbs. In this regard, the nouns and adjectives correspond to the intellectual entities, that we discussed in the previous section, while the verbs correspond to the intellectual relations.

The regular sentence has traditionally been the principal object of study in the language syntax. A regular sentence is the fundamental syntactical arrangement of the form: subject-verb-object. The subject and the object are members of the same grammatical groups. They are nouns or adjectives.

In a similar way, the fundamental conceptual group of the form entity-relation-entity has traditionally been (either implicitly or explicitly) one of the principal objects of the study of physical experience.

Science, as the whole of the human methodical effort to study physical experience, is greatly concerned with problems of that conceptual form.

At this point, we have to distinguish between two categories of relations. In order to do that, we first have to discuss the intellectual content of "change" and "dependence".

Change is a relational concept referring to two intellectual entities, which correspond to different pieces of experience, which exhibit a great amount of common characteristics in such a way, that the brain retains the common parts under the same concept; while creating another concept to relate intellectually these two different pieces of experience. For example, two different positions of a chair are two intellectual entities corresponding to two different pieces of visual experience. However, they both exhibit a great amount of common visual characteristics, which makes the brain (for reasons of its own natural economy, construction, organization and evolution) correspond the same concept (chair) to the common experience parts; while forming a concept (position) to relate (change) these two different pieces of experience. Clearly, this new concept is a changeable concept, while the first one is unchangeable. It must be emphasized again that this is simply a conceptual formalization and not an objective physical order.

The concept of dependence is formed from the perception of the relation between two changeable concepts.

Having the concept of dependence, the two categories of relations can be easily defined: The first category includes the descriptive relations. These are defined as relations that do not have the character

of dependence. The second category includes the functional (this is not the mathematical notion of "functional") relations. These are defined as relations that have the character of dependence. Clearly, the functional relations can relate only changeable concepts.

The study of the functional relations has traditionally been a great challenge of the human scientific thought. A basic question here is to what extent functional relations exist among existing concepts. Recalling the intellectual extensibility and the perceived continuity of the physical world, it is reasonable to accept a principle of universal interdependence, in the sense that there are relational links among all existing concepts, on the basis of the intellectual extensibility. Of course, the degree of this interdependence can vary very widely from one pair of interdependent concepts to another. In this regard, the totality of concepts (entities and relations) form a compact universal network, which corresponds to a given stage of human intelligence and intellectual extensibility.

In this formalization there are two basic notions involved: the structural (expressed by a specific arrangement of entities and relations) and the operational(expressed by the specific kind of functional relations). Both of these notions compose a conceptual lattice of functional inter-relations, which constitutes the arena of the system formalization.

In this context, system theory develops as the scientific tool for studying the functional relations among concepts.

1.3 The System Formulation of a Problem

The study of a functional relation between two (or more) changeable concepts requires a formal and convenient way of describing the changing characteristics of these concepts. This is done by employing the concept of "measure" as the outcome of relating, in a special way called "measurement", the changeable concept with a proper unchangeable one.

The measurement as an intellectual function of the human brain is connected to a natural quantitative sense that follows the ability of the brain to compare pieces of perceived information and arrange them in a certain order. In this way, we attain the very fundamental sense of time and space, and a variety of other quantitative (changeable) concepts. The measurement is an attempt to make a rigorous and formal description of one dimension of a quantitative concept.

A very important realization is that every piece of experience refers to certain ordinates of space and time. The perception of change, on the other hand, requires two different pieces of experience. Therefore, a perceived change is associated with two sets of space and time ordinates. Because of this basic fact, a measure can be followed by an associated space or time ordinate. A quantitative description of a changeable concept which is given in terms of measures followed by the associated time ordinates will be called a signal.

The signals are the observed quantities in the study of phenomena which obey the cause-and-effect scheme. The conceptual description of such a phenomenon is a functional relation. Thus, a

functional relation appears in this case as an operator performing the transformation of some signals to others. This operator is called a system.

In that sense, a system is the conductor of changes among signals that represent measures of changeable concepts in time. The direction of this conduction must be uniquely determined in each case. Of course, there may be a bidirectional interdependence between two concepts, but in that case, two systems must be employed (one for each direction). A system is an operator, operating upon some signals, which we call input signals, to produce other signals, which we call output signals.

A system being conceptually a relation can also be dependent on another concept. In this case, we are talking about a "parametric" system, while the system which exerts this parametric influence is called "regulatory".

The activity of a regulatory system can be thought of as an additional input to the system; however, the distinction occurs because of the fact that the activity of a regulatory system is not directly related to an output of the system and, therefore, it is not participant in a causal scheme of change conduction. The effect of the regulatory activity takes place within the system itself and, therefore, the input-output system formulation does not take hold in this case. Of course, the entire arrangement of entities and relations in a certain situation is conceptual and, therefore, arbitrary on behalf of the human subject who studies the phenomenon. Nevertheless, the conceptualization of parametric and

regulatory systems seems accordant with our experience about the dynamics of the surrounding world and, consequently, it will be of probable convenience in some cases.

In Fig. 1.3.1, we show diagramatically a portion of a conceptual network. The signals corresponding to the changeable concepts are shown as circles. The functional relations are shown as parallelograms accompanied by arrows indicating the direction of the change conduction. The regulatory relation is shown as a rhombus. More specifically:

C_i: is the input signal for system r_{ij} and the output signal for system r_{ji}.

C_j: is the output signal for system r_{ij} and the input signal for system r_{ji}.

C_k: is the regulatory signal, i.e. the input of the regulatory system p_{kij}.

r_{ij}: is the rightward system between C_i and C_j, and it is a parametric system.

r_{ji}: is the leftward system between C_i and C_j, and it is not a parametric system.

p_{kij}: is the regulatory system of the parametric system r_{ij}.

Recalling the principle of universal interdependence, we realize that the group in Fig. 1.3.1, is just a piece from the universal compact network. Nevertheless, the study of physical relations in practice requires the isolation of small segments of the universal network. This can be done easily (and practically reliably), because the interdependence among the concepts of interest and the rest of the universal network can

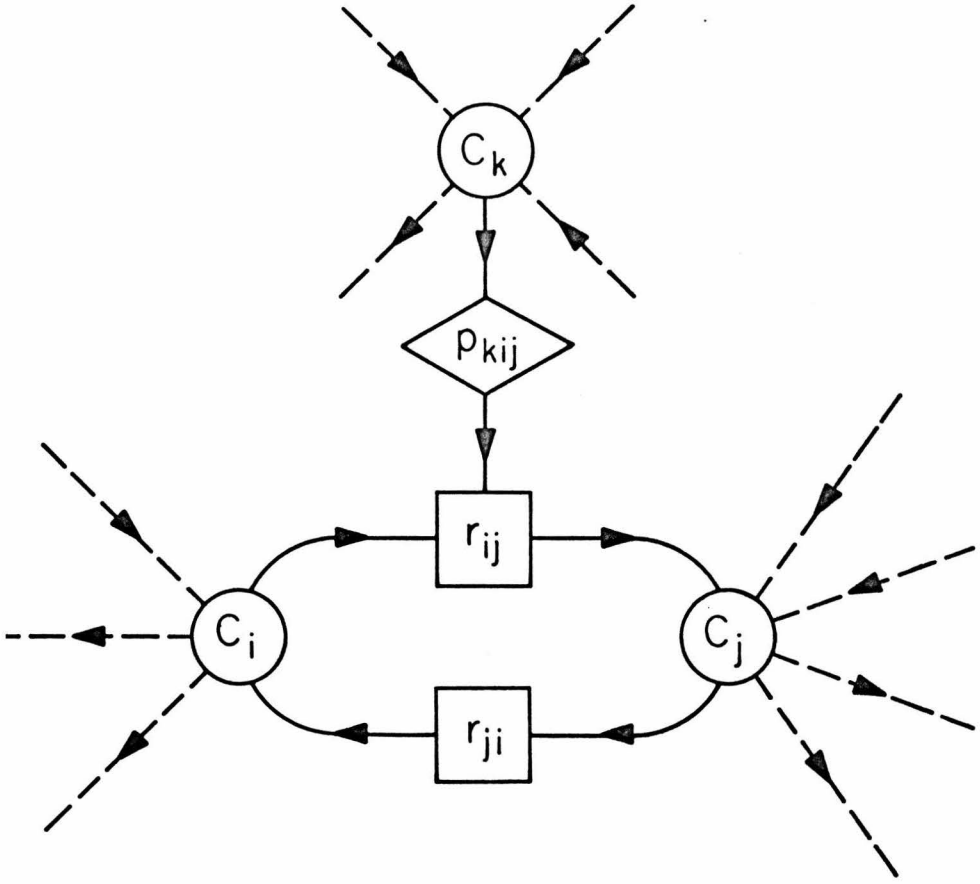


Fig. 1.3.1: Portion of an entity-relation conceptual network

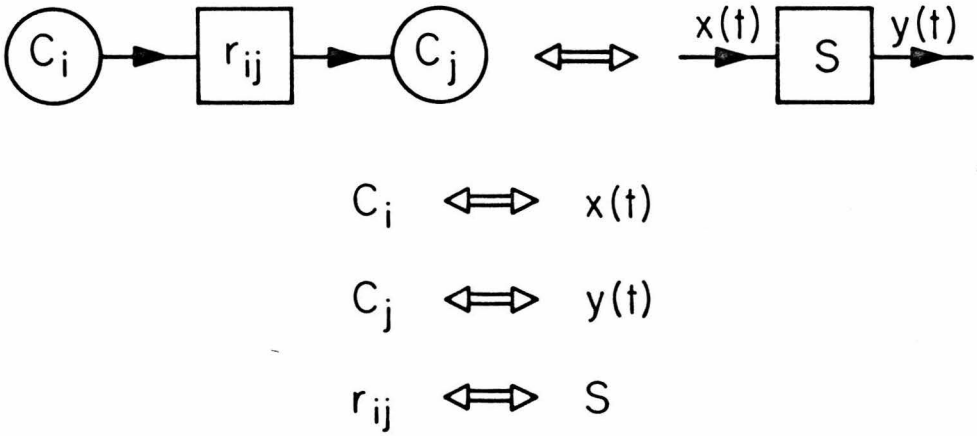


Fig. 1.3.2: Correspondence between a conceptual relation and an one-input/one-output system

be practically eliminated by designing the proper experiment and observational equipment. This is possible because the significant interdependences of a concept are usually limited in number. However, if there is remaining observable interdependence with uncontrollable or unknown factors, we consider it as external noise present in our observations. Proper methods have been designed to analyze the data in a case like that.

In conclusion, the system formulation and solution of a problem follows the following basic steps:

(1) Determine the concepts of interest in your study and form the entity-relation network of them. Also, determine the concepts which have significant interdependence with the concepts of your network, and try to neutralize them as much as the practical considerations allow.

(2) Perform any kind of experiment and analysis of the obtained data, which will give you an understanding of the existing functional relations.

(3) Study the behavior of the network from the aspects of interest, on the basis of the achieved understanding of the existing functional relations.

Therefore, the system formulation and solution of a problem comprises three basic steps. We call these steps: System analysis, identification and synthesis; in direct correspondence to the steps described above. It must be noted that the system analysis usually extends into part of the second step as well. In any case, the analysis is based

upon all the available information about the system and, in most of the cases, it takes the form of a hypothesis concerning the system structure (and basic functional features), which is to be tested through the identification and synthesis steps.

For example, if the system under study is the mass-spring well-known mechanical system, then:

(a) The concepts involved in the functional relation under study are the position of the mass and the applied force.

(b) The input signal is the forcing function and the output signal is the position function (both functions of time).

(c) The analysis step comprises the determination of the concepts involved (as above) and the specification of a second-order linear differential equation with constant coefficients as a reasonable formal description of the functional relation under study.

(d) The identification step comprises the determination (or estimation) of the unknown coefficients of the differential equation.

(e) The synthesis step comprises the study of the system response (mass position function) to a given forcing function, on the basis of the estimated differential equation (model), and in comparison to the actual experimental results. It also comprises the study of any interesting aspects of the system behavior (like stability, natural modes etc.), and any other desirable utilization of the

achieved knowledge about this functional relation. We must note, that this is a very simple and well-studied phenomenon, and for this reason the analysis step occupies a dominant role. In other cases, less simple or studied than this, the identification step may have the dominant role.

The study of complex systems (networks) is naturally the ultimate target of system science. Having in mind that the scientific course is one of building up from the simpler, well-studied cases to the more and more complex with small, studious, systematic steps, we believe that the first fundamentally important step of system science is the thorough study of the elementary network, which comprises only the functional relation between two concepts (i. e. one input, one output system). (Fig. 1.3.2)

The study of this fundamental case, with main emphasis on the identification problem, is the principal concern of the present dissertation.

1.4 An Outline of the Present Dissertation

The principal aim of the present dissertation is to provide a concise and accurate account of the use of quasi-white test signals in nonlinear system identification through crosscorrelation methods.

In chapter 2, we review the system identification problem, and particularly in connection with the Volterra-Wiener approach to that problem. Several important aspects of the Wiener white noise method are discussed.

In chapter 3, we review the properties and the use of band-

limited gaussian white noise and pseudorandom signals (based on m-sequences) in nonlinear system identification.

In chapter 4, we introduce a new family of quasi-white signals (CSRS) and we discuss its properties and use in nonlinear system identification. A comparative discussion of the advantages and disadvantages of the several quasi-white signals is also given.

In chapter 5, we study and evaluate the several estimation errors that are committed in nonlinear system identification through the CSRS.

In chapter 6, we discuss the overall management of errors occurring during the identification process. Several suggestions are made as to how the effect of these errors can be minimized, and an optimum identification procedure is outlined.

In chapter 7, we present two special techniques referring to important aspects of the nonlinear system identification process. The first technique, in particular, appears to provide a potentiality for the development of a new complete identification method.

In chapter 8, we illustrate through computer simulated applications the use of CSRS in nonlinear system identification.

In chapter 9, we present the application of CSRS in the identification of a real physiological system.

In chapter 10, we present a comparative study of the several quasi-white signals through computer simulated and real system applications.

CHAPTER II

IDENTIFICATION OF A SYSTEM

2.1 The System Identification Problem

In the previous chapter, we introduced the system formalization as a conceptual construction of a network involving the significant functional relations among changeable concepts. A certain conceptual construction concerning a set of observable (directly or indirectly) phenomena is not unique. A variety of conceptual constructions can usually portray the same piece of physical experience.

From all these various constructions, some tend to be preferable to the others because of two basic reasons. One is that some of them allow the utilization of the existing scientific tools to analyze themselves, while others do not. The other is our natural tendency towards maximum efficiency (least effort for the same return), which leads to the maximum simplicity (according to some criteria) of the construction form. Thus, a certain construction form usually prevails and constitutes the system formulation of the observed phenomena. In any case, this conceptual construction, which takes the form of an entity-relation network, is created on the basis of all available information concerning the concepts involved.

After the entity-relation network is constructed, a proper formal description of it, is sought. The whole of formal descriptions of the functional relations and of the measure changeable entities, is contained in the mathematical science. Thus, the available tools to

analyze the entity-relation network are the existing mathematical tools.

A formal mathematical description of the entity-relation network is called a mathematical model. The formulation of a mathematical model describing the functional relations among some concepts of interest, is the object of "systems analysis". The specific form that a mathematical model will take in a given case is a matter open to creativity. The mathematical tools are finite in number and variety, however, the extent of creative combination of some of them into new forms is virtually infinite.

In any case, the constructed mathematical model involves a number of unknown parameters that have to be evaluated (estimated). The amount and the form of these parameters depend on the specific model formulation. Clearly, it is a great virtue of the model to have a few parameters in a simple form that allows their easy evaluation (estimation).

This task of parameter evaluation (estimation) is the "system identification problem". This problem, in connection with the Volterra-Wiener model formulation is the principal subject of the present study.

2.2 Methodology in System Identification

The methodology of system identification crucially depends on the system analysis stage, because the analysis process determines the amount and the form of the parameters to be identified.

Algebraic and differential equations have traditionally been the most popular forms of models to be used in the analysis stage. These forms are compact and fairly illustrative (in most of the cases) of the

modelized relations. However, they usually require a fair amount of knowledge concerning the entities and the relations involved. This knowledge is, in a lot of cases, available or can be obtained through observation and/or experimentation at various levels of structural complexity. However, there are even more cases where this kind of insight knowledge of the system structure is not available or cannot be obtained with the available means of data acquisition. This is the case where the system under study takes the form of a "black-box".

A general state-space differential equation formulation can be used to modelize a black-box; however, in that case, the identification process becomes usually very complicated and involved. A reasonable approach, in the case of a black-box, is to use an expansion-type open form mathematical relation, which alleviates the strict structural requirement of a closed form input-output mathematical relation, by simply relating the projections of the input-output signals upon some standard functional modes. This is the basic idea of the Volterra-Wiener approach which we will extensively discuss in the following.

In the Volterra formulation the input-output mathematical relation is a multiple-integral series expansion, where the n-th order term (corresponding to the n-th order standard functional mode) is an n-tuple integral describing the interaction among n infinitesimal pieces of the input signal's past. In this perspective, the study of any input-output relation (which is representable by the Volterra series) attains a common functional pattern, unifying and simplifying the analysis stage of a great number of systems.

For a long time, the scientific effort was confined to the study of systems, which can be represented quite reliably only by the first functional mode of the Volterra series. These are the so-called linear systems, for which the principle of superposition holds. Of course, the linearity of a system is, in some cases, a reasonable approximation to an actual nonlinearity, in which the first functional mode is strongly predominant. However, the linear systems obviously constitute a special and limited class of systems. The study of linear systems has been very extensive, thorough and successful, with great achievements on the synthesis stage (control theory etc.).

At the present stage of scientific research in system science, the most challenging subject is the study of the nonlinear systems. Evidently, the complexity and variety of nonlinear systems is overwhelming with respect to the linear ones, along with the difficulties and complications encountered in their study. Because of the great variety of nonlinear systems, we are promptly forced to specialize on classes of them and study them successively on a step-by-step procedure.

In the present study, the approach that will be followed is the Volterra-Wiener one; because of the advantages that it provides in the analysis stage, and its generality and unification characters. This does not mean, of course, that this is an advantageous approach for all the cases; neither that a more general and advantageous approach cannot be designed. It simply seems to us as the most appropriate, at present, method to analyze and identify a wide class of nonlinear systems, which appear, with the present means of information acquisition, as black-

boxes.

2.3 The Volterra-Wiener Approach of Nonlinear System Identification

Consider the fundamental unit of an one-input/one-output system (Fig. 1.3.2). The functional relation between the output y(t) and the input x(t) can be in general denoted with the mathematical notion of a "functional":

$$y(t) = F[\, x(t)] \qquad (2.3.1)$$

The functional F is a mathematical operator operating on the values of the function x(t) to produce the values of the function y(t). In that respect, the functional F is a general mathematical description of the system S. Obviously, the system analysis and identification are directed towards the study of this functional, i.e. towards a more explicit expression of the functional relation F, which allows the evaluation (estimation) of the parameters involved.

For a physical system S, the causality principle is the first instrument of the analysis process. Causality holds in all physical systems that do not exhibit autonomous ability to act in a way that is not completely dictated by the stimulus past. Under the causality principle, a system only reacts to external stimulation, in a way determined by the past (and present) values of the stimulus. The causality principle is the formal expression of the deterministic cause-and-effect philosophical viewpoint; a viewpoint, that is strongly supported by the observable mechanistic outlook of physical phenomena.

Therefore, for a physical system S, we accept that the response

$y(t)$ depends solely upon the past (and present) values of the stimulus $x(t)$:

$$y(t) = F[\, x(t');\ t' \leq t\,] \qquad (2.3.2)$$

If only the present stimulus value affects the response at the present time, the system is called static. Apparently, the analysis and identification of static systems is a much simpler task than the analysis and identification of dynamic systems. Thus, the great challenge is the study of dynamic systems. There are also cases, where we cannot determine an invariant functional F, which governs the input-output relation for all times t. These are the hereditary systems, in which the relation is changing on the basis of the system history.

Further, any explicit mathematical expression of the functional F has a certain structural form involving a set of parameters and a set of constants. That certain structural form and the set of constants is the object of the analysis process. The set of parameters Ω is the object of the identification process, in which we seek to determine (estimate) these parameters. Therefore, we can use the notation:

$$y(t) = F[\, \Omega;\ x(t'),\ t' \leq t\,] \qquad (2.3.3)$$

to demonstrate the existence of the parameter set within the functional expression. It must be noted that the parameter set Ω is independent from the stimulus $x(t)$, but it might be dependent (explicity) on time.

In the second step of analysis, we confine ourselves to a class of systems, which exhibit the following three basic characteristics:

(1) They are stationary: i.e. their functional characteristics (formally represented as parameters in the explicit mathe-

matical functional expression) do not change in time. In other words, the parameter set Ω (eqn. 2.3.3) is independent of time for a stationary system.

(2) They have finite memory: i.e., the effect of a stimulus with finite energy on the response decays to arbitrarily small degree in finite time. This justifies the search for transfer functional description of the system, which is independent from the initial conditions.

(3) They are analytic: i.e., their differential behavior of all orders is continuous within some domain of stimulus values.

In the case of a system of this class, the corresponding functional F can be expanded in a functional power series, as V. Volterra suggested [1]. This functional power series is a series of multiple integrals of the form:

$$y(t) = \sum_{n=0}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} k_n(\tau_1, \ldots, \tau_n) x(t-\tau_1) \ldots x(t-\tau_n) d\tau_1 \ldots d\tau_n$$

$$\triangleq \sum_{n=0}^{\infty} I_n(t)$$

and it is known as the Volterra series. The Volterra series can be thought of as the limiting case of Taylor series expansion of a function with multiple arguments, when the number of these arguments tends to infinity. These arguments, in the case of the Volterra series, are the values of the function x(t) (infinite in number), and every kernel $k_n(\tau_1, \ldots, \tau_n)$ relates to the n-th order derivative of the functional F.

It must be noted that the kernels of a physical system are symmetric functions of their arguments, and they attain zero values for any

negative argument value. This is how causality is manifested mathematically in the Volterra series expansion.

The stationarity of the system is manifested by the independence of all kernels from time.

The finite memory is a characteristic relating to the convergence of the functional series. It is known that if every term $I_n(t)$ of the functional series is absolutely bounded by a positive constant a_n for all values of t in the domain of interest and the several a_n form a uniformly convergent numerical series, then the functional series is uniformly convergent as well (Weierstrass). That is, if there is a uniformly convergent numerical series Σa_n, for which,

$$\left| I_n(t) \right| \leq a_n \qquad (2.3.5)$$

for any n and t; then the functional series $\Sigma In(t)$ is also uniformly convergent. Consequently, a sufficient condition for the uniform convergence of the Volterra series is:

$$\left| I_n(t) \right| = \left| \int_0^\infty \cdots \int_0^\infty k_n(\tau_1, \ldots, \tau_n) x(t-\tau_1) \cdots x(t-\tau_n) d\tau_1 \cdots d\tau_n \right| \leq$$
$$\int_0^\infty \cdots \int_0^\infty \left| k_n(\tau_1, \ldots, \tau_n) \right| \left| x(t-\tau_1) \right| \cdots \left| x(t-\tau_n) \right| d\tau_1 \cdots d\tau_n \leq a_n \qquad (2.3.6)$$

for all n and t (in the domain). Suppose now that the stimulus x(t) is also absolutely bounded:

$$\left| x(t) \right| \leq M \quad ; \text{(for all t)} \qquad (2.3.7)$$

then, the sufficient condition for uniform convergence becomes:

$$\int_0^\infty \cdots \int_0^\infty \left| k_n(\tau_1, \ldots, \tau_n) \right| d\tau_1 \cdots d\tau_n \leq a_n / M^n \qquad (2.3.8)$$

Therefore, the finite memory condition can be mathematically expressed as the absolute integrability of the kernels, in the way described by eqn. 2.3.8 , which guarantees the uniform convergence of the functional series.

In cases where the stimulus x(t) is a random process and possibly not absolutely bounded (e.g. the case of a gaussian random process), the expressions above can be written in terms of the expected values of the random quantities involved:

$$|E[I_n(t)]| = \left|\int_0^\infty \cdots \int_0^\infty k_n(\tau_1, \ldots, \tau_n)\, E[x(t-\tau_1)\ldots x(t-\tau_n)]\, d\tau_1 \ldots d\tau_n\right| \leq$$

$$\int_0^\infty \cdots \int_0^\infty |k_n(\tau_1, \ldots, \tau_n)|\, |\phi_n(t, \tau_1, \ldots, \tau_2)|\, d\tau_1 \ldots d\tau_n \leq a_n \qquad (2.3.9)$$

where, $\phi_n(t, \tau_1, \ldots, \tau_n)$ is the n-th order autocorrelation function of the random process x(t). If the process x(t) is stationary then ϕ_n is independent of time t. Consequently, in this case the uniform convergence condition is written for the absolute bounds of the autocorrelation functions (if they exist):

$$|\phi_n(t, \tau_1, \ldots, \tau_n)| \leq R_n \qquad \text{(for all } t, \tau_1, \ldots, \tau_n) \qquad (2.3.10)$$

$$\int_0^\infty \cdots \int_0^\infty |k_n(\tau_1, \ldots, \tau_n)|\, d\tau_1 \ldots d\tau_n \leq a_n / R_n \qquad (2.3.11)$$

Again, the uniform convergence condition takes the form of an absolute integrability condition for all kernels.

The analyticity of the system is manifested by the existence of the functional derivatives of all orders, which are directly related to the kernels k_n.

The set of kernels k_n constitutes the parameter set of the system, under this formulation. Therefore, the identification task in this case consists of determining (estimating) these kernels. To our great disappointment this is not an easy task. A general mathematical method to determine (estimate) these kernels is not yet available.

To facilitate the identification task, N. Wiener constructed a new functional series, on the basis of the Volterra series, in which the several functional terms are orthogonal to one another for a gaussian white noise stimulus. [2] The Wiener series is equivalent to the Volterra series but it spans more efficiently the function space (being orthogonal). The choice of the gaussian white noise (GWN) by Wiener was not accidental. Besides the fact that he was preoccupied with GWN from his studies on Brownian motion, Wiener was regarding GWN as a uniform blend of all possible signals and, therefore, he reasoned that the testing of the system with GWN would be an exhaustive one, providing a complete profile of the behavior of the system. On the other hand, GWN is a signal of extreme variability, since any two samples of it are statistically independent. Consequently, the passage of GWN through a system, which naturally acts as a correlator of successive input samples (if it has a nonzero memory), could reveal the pattern of the system behavior in terms of the correlation patterns among the output samples. These are some intuitive justifications of the use of GWN and probably there can be found more. The matter of the fact is that the GWN lived up to the expectations of its promoters, and its actual use in nonlinear system identification justified all these intuitive arguments.

The Wiener functional series is constructed with Volterra-type functionals, on the basis of a Gram-Schmidt-type orthogonalization procedure, in such a way that any two Wiener functionals are orthogonal for x(t) being GWN. The Wiener series has the form:

$$y(t) = \sum_{n=0}^{\infty} G_n [h_n(\tau_1, \ldots, \tau_n) ; x(t'), t' \leq t] \qquad (2.3.9)$$

$$= \sum_{n=0}^{\infty} G_n(t)$$

where, G_n is the n-th order Wiener functional, being characterized completely by the respective Wiener kernel $h_n(\tau_1, \ldots, \tau_n)$ and having the structural form:

$$G_{2n}(t) = \sum_{m=0}^{n} \frac{(-1)^m 2n!\, P^m}{2(n-m)!m!2^m} \int_0^{\infty} \cdots \int_0^{\infty} h_{2n}(\tau_1, \ldots, \tau_{2n-2m}, \sigma_1, \sigma_1, \ldots, \sigma_m, \sigma_m) x(t-\tau_1) \cdots x(t-\tau_{2n-2m}) d\tau_1 \cdots d\tau_{2n-2m}\, d\sigma_1 \cdots d\sigma_m \qquad (2.3.10)$$

$$G_{2n+1}(t) = \sum_{m=0}^{n} \frac{(-1)^m (2n+1)!\, P^m}{(2n-2m+1)!m!2^m} \int_0^{\infty} \cdots \int_0^{\infty} h_{2n+1}(\tau_1, \ldots, \tau_{2n-2m+1}, \sigma_1, \sigma_1, \ldots, \sigma_m, \sigma_m) x(t-\tau_1) \cdots x(t-\tau_{2n-2m+1}) d\tau_1 \cdots d\tau_{2n-2m+1}\, d\sigma_1 \cdots d\sigma_m \qquad (2.3.11)$$

The orthogonality of the Wiener functionals is manifested by the condition:

$$E[G_n(t) G_m(t)] = 0 \qquad \text{for any } m \neq n \qquad (2.3.12)$$

According to eqn. (2.3.10), the n-th order Wiener functional is a linear combination of multiple integrals of order lower or equal to n. The parameter P, which appears in eqn. (2.3.10) is the power level of the GWN x(t), defined from its autocorrelation function:

$$E[x(t)\, x(t-\tau)] = P \cdot \delta(\tau) \qquad (2.3.13)$$

We can make two interesting remarks on the structural form of the Wiener functionals:

(1) If we call "order" of each integral term the number of $x(t-\tau_i)$ factors that it contains, then every even order G-functional contains a "leading" integral term of the same order with the functional and several "following" integral terms of all the lower even orders. The analogous is true for the odd order G-functionals.

(2) The "following" integral terms are simply subtracting the expected value of the contribution of all possible formations of diagonal points of the kernel in the "leading" integral term.

The first four Wiener functionals are:

$$G_o(t) = h_o \tag{2.3.14}$$

$$G_1(t) = \int_o^\infty h_1(\tau_1)\, x(t-\tau_1)\, d\tau_1 \tag{2.3.15}$$

$$G_2(t) = \int_o^\infty\int h_2(\tau_1, \tau_2)\, x(t-\tau_1)\, x(t-\tau_2)\, d\tau_1 d\tau_2 - P\int_o^\infty h_2(\tau, \tau) d\tau \tag{2.3.16}$$

$$G_3(t) = \int_o^\infty\int\int h_3(\tau_1, \tau_2, \tau_3)\, x(t-\tau_1)\, x(t-\tau_2)\, x(t-\tau_3)\, d\tau_1 d\tau_2 d\tau_3$$

$$-3P\int_o^\infty\int h_3(\tau_1, \tau_1, \tau_2)\, x(t-\tau_2)\, d\tau_1 d\tau_2 \tag{2.3.17}$$

e.t.c.

Evidently, the Wiener series is equivalent to the Volterra series; since they are both constructed by linear expressions of the same mathematical objects (the multiple homogeneous integrals) and they are both expansions of the same function (the system response y(t)). Consequently, there is a unique analytical relation between the Volterra and the Wiener kernels of a system, as it is discussed in sec. 2.5, which represents the transformation of the functional expansion basis.

Notice that the Wiener series involves one more parameter, in addition to the system kernels: the power level P. This is because the Wiener series are constructed as to possess the special feature of orthogonality with respect to GWN, and the power level P determines the range of validity of this orthogonality in the function space, in much the same way that the domain of the independent variable determines the range of validity of the orthogonality of two functions. Why is orthogonality so insistently pursued? There are three main reasons for that; just as in the case of function expansions. The first reason is that an orthogonal basis spans the space within the range of its validity more efficiently. Therefore, the Wiener series is expected to have stronger convergence than the Volterra series, for an arbitrarily chosen stimulus signal; and consequently, it will provide expectedly (i.e. over a large number of signals)a better truncated model of the system. This does not exclude the possibility of a special case where the opposite is true. The second reason is that if the expansion basis is orthogonal then the truncated model can be extended to include higher order terms without affecting the already estimated (determined) terms. The third

reason is that the orthogonality enables us to estimate the system kernels in a relatively simple way (as it is discussed in the following); in the same way as the determination of the coordinates of a given vector is greatly simplified if the vector basis is orthogonal (diagonalization of the coefficient matrix).

This last advantage that orthogonality provides is especially important in the actual identification of nonlinear systems; since the determination (or estimation) of the Volterra kernels of a system is an impossible task in the general case of a nonlinear system and with the present analytical and computational means. Nonetheless, the employment of the orthogonality property of the G-functionals allows, as is shown below, the relatively simple estimation of the system Wiener kernels.

It must be noted that the GWN is not the only signal with respect to which the functional series can be orthogonalized. The orthogonalization can be achieved for any other signal that possesses the proper autocorrelation properties. For every such signal a corresponding set of kernels can be estimated (as discussed in chapters 3 and 4).

Surprisingly enough, the basic idea of utilizing the orthogonality of the G-functionals to directly estimate the Wiener kernels, did not occur to Wiener. Wiener suggested a fairly complicated and cumbersome method of expanding the kernels and the stimulus onto a basis of orthogonal functions, and then try to estimate the expansion coefficients. This method, though theoretically correct, did not give satisfactory results in actual applications; the main reason being the complexity, the

repetitious approximations and the computational length of the operations involved.

A few years later, Y. W. Lee and M. Schetzen suggested the utilization of the orthogonality of the G-functionals, in order to estimate the Wiener kernels of a system [3]. This is called the "crosscorrelation technique" and simply requires the computation of crosscorrelations between the response and the stimulus of the system. According to the crosscorrelation technique, the n-th order Wiener kernel can be determined directly from the n-th order crosscorrelation:

$$h_n(\tau_1, \tau_2, \ldots, \tau_n) = \frac{1}{n!P^n} \operatorname*{Lim}_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} y(t)x(t-\tau_1)x(t-\tau_2)\ldots x(t-\tau_n)dt$$

$$= E[y(t)x(t-\tau_1)\ldots x(t-\tau_n)] \qquad (2.3.18)$$

Of course, in practice, only a finite record length can be used to estimate these crosscorrelations; and consequently, we can only obtain an estimate of the kernel as:

$$\hat{h}_n(\tau_1, \ldots, \tau_n) = \frac{1}{n!P^n} \cdot \frac{1}{T-Tm} \int_{Tm}^{T} y(t)x(t-\tau_1)\ldots x(t-\tau_n)dt \qquad (2.3.19)$$

where, $T_m = \max [\tau_1, \ldots, \tau_n]$.

The expression above gives an estimate for the nondiagonal points, while to estimate the diagonal ones we must subtract from y(t) all the contributions of the G-functionals up to order (n-1) and crosscorrelate the resulting residue:

$$y^{(n)}(t) = y(t) - \sum_{k=0}^{n-1} G_k [h_k(\tau_1, \ldots, \tau_k); x(t'), t' \le t] \qquad (2.3.20)$$

$$\hat{h}_n(\tau_1, \ldots, \tau_n) = \frac{1}{n!P^n} \cdot \frac{1}{T-T_m} \int_{T_m}^{T} y^{(n)}(t)x(t-\tau_1)\ldots x(t-\tau_n)\,dt \qquad (2.3.21)$$

Therefore, the crosscorrelation technique provides a straightforward method to estimate the Wiener kernels of a system, which can be tested with GWN stimulus. However, there are some practical restrictions imposed in the application of the method. These restrictions result from the unrealistic nature of the ideal GWN, as well as from experimental and computational considerations. We discuss these restructions imposed by reality in the following section.

2.4 The Applicability of the Wiener Method

The Wiener formulation of the nonlinear system identification problem, in connection with the crosscorrelation technique, provides a general, straightforward and powerful approach to a subject of broad interest. However, the generality and the elegance of the method was bound to be moderated in the actual applications by limitations imposed by reality.

The very first reason for this, is that the GWN (for which the whole method has been designed) is not a physically realizable signal; since it has an infinite frequency bandwidth and, consequently, infinite power. Thus, in actual applications of the method, we have to use an approximate signal which exhibits the properties of the GWN within a certain range of interest and up to a determinable degree. This was initially done by using a band-limited GWN. The band-limited GWN has a flat power spectrum up to a characteristic cut-off frequency, which

is determined so to cover the bandwidth of the system under test. The use of band-limited GWN, instead of ideal GWN, introduces an estimation error, which varies according to the specific characteristics of every actual case, and which, in general, represents a loss of high frequencies in our kernel estimates.

A second practical restriction and source of estimation error in the application of the method is the finite record length. The experimentation time and the data recording time are naturally finite. Therefore, the stimulus and response records with which we compute the crosscorrelations are finite; and consequently, the averages are formed within some statistical error. This error, of course, is monotonically decreasing as the record length increases, and consequently, it can be supressed to a determinable degree.

A third practical restriction is the extent of the truncated Wiener series in connection with the computational capacity required to estimate higher order kernels. The computation of the crosscorrelation estimates involves a great number of multiplications and additions; and this computational burden is rapidly increasing as we move to higher order kernels. Therefore, we usually confine ourselves to the estimation of a few first terms of the Wiener series, which give a model of acceptable accuracy. This is the reason why the convergence pattern of the functional series is of great practical importance.

A fourth practical restriction is the truncation of the gaussian distribution, since the actual stimulus signal cannot attain infinite values. The error resulting from this restriction is usually insignificant.

Besides all these practical restrictions in the application of the method, there is a number of limitations and errors associated with the numerical operations involved and the capacity of the digital computer, that we are using to compute our estimates. We will discuss all these limitations and errors in chapter 3 and 5.

2.5 The Relation Between Wiener and Volterra Kernels

The Wiener series is equivalent to the Volterra series, in the sense that they both span the same function space. However, the Wiener kernels depend on the power level of the GWN with which they have been estimated.[6] The specific value of the power level determines the region of orthogonality of the Wiener G-functionals. Therefore, a system is completely described either by the set of Volterra kernels, or by the set of the Wiener kernels plus the corresponding power level. In this sense, the set of Volterra kernels is equivalent to the set of Wiener kernels plus the corresponding power level:

$$\{k_n\} \Leftrightarrow \{h_n, P\} \tag{2.5.1}$$

Clearly, the overall model given by the Wiener series is independent of P, however, both the individual Wiener kernels and the G-functionals depend on P. The Volterra kernels, on the other hand, must be thought of as a set of invariant characteristics of the system.

It must be emphasized that, in practice, we usually have to truncate the Wiener series and, consequently, the obtained model depends on P. This dependence on P is explicable in the sense that it

determines the range of the stimulus values within which the corresponding Wiener series is orthogonal.

The dependence of the Wiener kernels upon P can be nicely illustrated through the analytical relations between the Volterra and the Wiener kernels of a system. To this purpose, consider the Volterra series of a system:

$$y(t) = \sum_{n=0}^{\infty} \int_0^{\infty} \cdots \int k_n(\tau_1, \cdots \tau_n) \prod_{i=1}^{n} x(t-\tau_i) d\tau_i \qquad (2.5.2)$$

where, x(t) is GWN.

The Wiener kernels of the system can now be evaluated through the crosscorrelation technique:

$$h_{2n}(\sigma_1, \cdots \sigma_{2n}) = \sum_{m=n}^{\infty} \frac{2m!\, P^{m-n}}{2n!(m-n)!2^{m-n}} \int_0^{\infty} \cdots \int_0^{\infty} k_{2m}(\tau_1, \tau_1, \cdots, \tau_{m-n}, \tau_{m-n}, \sigma_1, \cdots, \sigma_{2n}) d\tau, \cdots d\tau_{m-n} \qquad (2.5.3)$$

$$h_{2n+1}(\sigma_1, \cdots, \sigma_{2n+1}) = \sum_{m=n}^{\infty} \frac{(2m+1)!P^{m-n}}{(2n+1)!(m-n)!2^{m-n}} \int_0^{\infty} \cdots \int_0^{\infty} k_{2m+1}(\tau_1, \tau_1, \cdots, \tau_{m-n}, \tau_{m-n}, \sigma_1, \cdots, \sigma_{2n+1}) d\tau_1 \cdots d\tau_{m-n} \qquad (2.5.4)$$

Clearly, the even order Wiener kernels are polynomials in P with coefficients depending on all the higher even order Volterra kernels. The analogous is true for the odd order Wiener kernels.

The derived expressions (2.5.3) and (2.5.4) clearly state the fact of the dependence of the Wiener kernels and the G-functionals on P. Similar expressions can be derived for the Volterra kernels of the system in terms of the Wiener kernels and the respective power level P by collecting the appropriate terms from the G-functionals (cf. eqns. 2.3.10 and 2.3.11).

One interesting implication of the derived relation between the Volterra and the Wiener kernels of the system is that the first order Wiener functional of a nonlinear system is, in general, different from the linear part of the system (first order Volterra functional) and it contains portions of the system nonlinearities. This demonstrates the efficiency of the Wiener representation, where even the first order functional term is probing into the nonlinear characteristics of the system.

We can also note that the impulse response of a nonlinear system:

$$p(t) = k_o + k_1(t) + k_2(t,t) + k_3(t,t,t) + \ldots \qquad (2.5.5)$$

is distinctly different from either the first order Volterra kernel or the first order Wiener kernel (unlike the case of a linear system).

2.6 The Case of Nonzero Stimulus Average Level

In a lot of practical applications, the necessity arises to use a stimulus signal, which cannot attain physically negative values. For example, if we want to study the behavior of part of the visual system of an insect under changes of light intensity, we have to use a stimulus signal (light intensity), which cannot attain physically negative values. In this case, the stimulus signal must have a nonzero average level, which we conventionally define to be our computational zero level. The question which arises in this case, is whether this is legitimate and what is the meaning of the obtained kernel estimates.

To answer this question, we recall the analogy between the

Volterra series and the Taylor series expansion, discussed in Sec. 2.3. As it is well known, the Taylor series can be written with respect to any reference point where the function is analytic. On the basis of the same argument the Volterra series expansion can be written with respect to any reference function, where the functional is analytic. One such function is a constant one (in our application the nonzero stimulus average level). Therefore, the computational reference to a conventional zero level is completely legitimate, as long as we stay within the region of analyticity of the functional at hand.

Of course, the obtained kernel estimates depend on the chosen reference level (which is here the average level of the stimulus), in the same way that the derivatives of a function depend on the specific value of the independent variable that is used as a reference point in the Taylor series expansion. Thus, the estimated Wiener kernels depend on the functional derivatives at the chosen reference level. It is interesting to see the analytical relations between the Volterra kernels of a system corresponding to a nonzero reference level c $\{k_n^c\}$ and the ones corresponding to the zero reference level $\{k_n\}$. Since:

$$y(t) = \sum_{n=0}^{\infty} \int_0^{\infty} \cdots \int_0^{\infty} k_n^c(\tau_1, \ldots, \tau_n) \prod_{i=1}^{n} x(t-\tau_i) d\tau_i$$

$$= \sum_{n=0}^{\infty} \int_0^{\infty} \cdots \int_0^{\infty} k_n(\tau_1, \ldots, \tau_n) \prod_{i=1}^{n} [x(t-\tau_i)+C] d\tau_i \qquad (2.6.1)$$

it follows that:

$$k_n^c(\tau_1, \ldots \tau_n) = \sum_{\ell=n}^{\infty} \binom{\ell}{n} C^{\ell-n} \int_0^{\infty} \cdots \int_0^{\infty} k_\ell(\tau_1, \ldots, \tau_n, \tau_{n+1}, \ldots \tau_\ell) d\tau_{n+1} \ldots d\tau_1 \qquad (2.6.2)$$

Notice that there is some morphological similarity between this expression and the one describing the relation between the Volterra and Wiener kernels; however, these expressions do not separate the odd from the even order kernels.

Conclusively, in those practical applications, where the stimulus signal is conventionally assigned a computational zero level, the obtained set of Wiener kernel estimates must be accompanied by the information of the value of this level (with respect to a global frame of reference); since the estimated model is good only for stimulus values referred to this reference level.

At this point, we would like to express the far reaching thought that, possibly, a proper choice of a reference function (not necessarily constant) may result in a strongly convergent Volterra series for a given system, which has a weak convergence for the null reference function.

2.7 The Multi-Input Case

In a lot of actual applications the system under study has more than one significant input. A subject of great interest is the study of the possible interactions between the several inputs; since, if there are no interactions each one of the inputs can be studied separately.

In order to study this multi-input case, the Wiener method was properly extended by P. Z. Marmarelis, G. D. McCann and K. I. Naka to include functional terms which depict the possible interactions among infinitesimal portions of the several inputs [4, 5]. According to this

method, an augmented functional series is written, which has functional terms of the multiple integral form that include all possible combinations of interactions among infinitesimal portions of the several inputs. These augmented functionals are constructed orthogonal to one another for the several inputs being independent GWN. They turn out to have the form: (for an m-input system)

$$Q_o[\, h_o; x_1(t'), \ldots, x_m(t'), t' \le t] = h_o \qquad (2.7.1)$$

$$Q_1[\, h_{x_1}, h_{x_2}, \ldots, h_{x_m}; x_1(t'), \ldots, x_m(t'), t' \le t] =$$
$$= \int_o^\infty h_{x_1}(\tau) x_1(t-\tau) d\tau + \int_o^\infty h_{x_2}(\tau) x_2(t-\tau) d\tau + \ldots +$$
$$\int_o^\infty h_{x_m}(\tau) x_m(t-\tau) d\tau \qquad (2.7.2)$$

$$Q_2[h_{x_1x_1}, h_{x_1x_2}, \ldots, h_{x_mx_m}; x_1(t'), x_2(t'), \ldots, x_m(t'), t' \le t] =$$
$$= \int_o^\infty\int h_{x_1x_1}(\tau_1, \tau_2) x_1(t-\tau_1) x_1(t-\tau_2) d\tau_1 d\tau_2 - P_{x_1} \int_o^\infty h_{x_1x_1}(\tau, \tau) d\tau$$

$$+ \int_o^\infty\int h_{x_2x_2}(\tau_1, \tau_2) x_2(t-\tau_1) x_2(t-\tau_2) d\tau_1 d\tau_2 - P_{x_2} \int_o^\infty h_{x_2x_2}(\tau, \tau) d\tau$$

$$+ \ldots$$

$$+ \int_o^\infty\int h_{x_mx_m}(\tau_1, \tau_2) x_m(t-\tau_1) x_m(t-\tau_2) d\tau_1 d\tau_2 - P_{x_m}$$
$$\int_o^\infty h_{x_mx_m}(\tau, \tau) d\tau$$

$$+ \int_o^\infty\int h_{x_1x_2}(\tau_1, \tau_2) x_1(t-\tau_1) x_2(t-\tau_2) d\tau_1 d\tau_2 + \ldots$$

$$+ \int_o^\infty\int h_{x_{m-1}x_m}(\tau_1, \tau_2) x_{m-1}(t-\tau_1) x_m(t-\tau_2) d\tau_1 d\tau_2 \qquad (2.7.3)$$

e.t.c.

where, $P_{x_1}, \ldots, P_{x_m}$ are the power levels of the m independent GWN inputs $x_1(t), \ldots, x_m(t)$ respectively. Since, the Q-functionals are orthogonal for independent GWN inputs, the estimation of the kernels can be done again through the crosscorrelation technique, in a way very similar to the one of the single input case. A certain kernel is now estimated by the crosscorrelation of the system output with the corresponding (to the kernel) inputs, after appropriate normalization with a factor involving the power levels of these inputs. For example, the second order cross-kernel $h_{x_k x_e}$ is obtained through the crosscorrelation technique as:

$$h_{x_k x_e}(\sigma_1, \sigma_2) = \frac{1}{P_{x_k} P_{x_e}} E[y(t) x_k(t-\sigma_1) x_e(t-\sigma_2)]$$

where, P_{x_k} and P_{x_e} are the power levels of the independent GWN input $x_k(t)$ and $x_e(t)$ respectively. Notice that, unlike the self-kernels, the cross-kernels are not symmetric functions of their arguments and they describe the pattern of interaction among the respective inputs.

It is apparent that the extension of the Wiener functional series to cover the multi-input case, is of great importance and usefulness in a variety of applications; since, it can be used to detect easily possible interactions among physical measures.

Besides the multi-input case, the case of more than one independent variable in the input-output measures appears as a challenging topic of study. Some notable work has been recently done towards that direction (spatio-temporal system kernels) by S. Yasui and D. H. Fender [6].

CHAPTER III

NONLINEAR SYSTEM IDENTIFICATION THROUGH BAND-LIMITED GAUSSIAN WHITE NOISE AND PSEUDORANDOM SIGNALS

3.1 Practical Restrictions of the Wiener Method and the Introduction of Quasi-White Signals

As it was discussed in sec. 2.4 the actual application of the Wiener method is confronted with a variety of problems. The main sources of problems are: the non-realizability of the white noise as a signal, the finite record lengths and the computational burden of the crosscorrelation technique. To overcome these problems, the scientists resorted to approximating signals which provided the biggest possible advantages with respect to these inherent practical restrictions of the method.

Naturally, the first approximating signal that comes to mind is the band-limited gaussian white noise, since it is physically realizable while still preserving the basic statistical properties of GWN. The band-limited gaussian white noise is a gaussian random process with a flat power spectrum over all frequencies lower than a particular one, which is chosen so as to cover the bandwidth of the system under test. The autocorrelation properties of this signal are fairly close to the ones of ideal gaussian white noise, and that allows its use as an approximating signal in the crosscorrelation technique.

Beyond the problem of physical realizability, an important remaining practical problem is the required long computations for the

statistical averages to form with tolerable deviations. Because of this problem, several scientists suggested the use of pseudorandom signals based on m-sequences. These signals being periodic (and deterministic in nature) were forming the desirable second order autocorrelation function and all the odd order ones in much shorter time, because they are constructed so that they eliminate the redundancy of a random signal; however, the higher even order autocorrelation functions exhibited inherent anomalies, which could cause serious estimation errors in identification of systems with higher order nonlinearities.

These two families of approximately white signals have been predominantly used so far in applications of the Wiener method in nonlinear system identification. When we say "Wiener method" we mean the general Wiener white-noise approach to the system identification problem in connection with the crosscorrelation technique. The choice between the band-limited gaussian white noise and the pseudorandom signals is made in each specific case on the basis of their relative advantages and disadvantages, as discussed in the following sections. Understandably, the principal concerns of the users of the Wiener method are: (a) the reduction of the inevitable estimation error to the minimum possible within the existing practical limitations, and (b) the efficiency of the estimation procedure.

In the following, we will call these approximating signals "quasi-white", since they are not really white but they are used as such in applications of the Wiener method. The use of the popular quasi-white families of band-limited gaussian white noise and pseudorandom signals

based on m-sequences will be discussed in the following sections 3.2 and 3.3.

Evidently, the construction of a new quasi-white signal is a matter of creative thinking and, therefore, their number is extensible. In fact, the major topic of the present dissertation is the introduction of a new family of quasi-white signals and the study of their properties in connection with nonlinear system identification. Undoubtedly, the search for quasi-white signals with more and more suitable properties for given applications remains a challenging subject of the scientific effort in the field of system identification.

3.2. The Band-Limited Gaussian White Noise (GWN)

The band-limited gaussian white noise (hereafter denoted simply by GWN) was naturally, the first quasi-white signal to be employed in applications of the Wiener-crosscorrelation method. The GWN is the first approximate signal that comes to mind in an attempt to physically realize the ideal gaussian white noise for which the whole theory had been developed through the remarkable work of a number of investigators [7, 8, 9, 10]. Successful applications of the GWN were delayed by several practical limitations and difficulties, and they finally took place mainly in the area of biological systems [11, 12, 13, 14, 15, 16, 17]. The successful application of GWN in biological systems identification predicated the potentiality of the Wiener-crosscorrelation method and it signaled its extension into other areas of scientific research [18, 19].

3.2.1 General Description and Generation of GWN

The band-limited gaussian white noise is a gaussian random process with a rectangular power spectrum (see Fig. 3.2.1), the bandwidth B of which ought to cover the bandwidth of the system under test, if we are to consider it an approximate of the ideal gaussian white noise. The mean of the GWN is zero and the variance is determined by the dynamic range within which we want to test the operation of the system at hand. Of course, in practice we have to truncate the tails of the gaussian distribution beyond a certain number of standard deviations (usually $\pm 3\sigma$). In Fig. 3.2.1, an actual band-limited distribution-truncated gaussian white noise signal is shown, along with the corresponding amplitude distribution.

As it is discussed in sec. 3.2.2, this quasi-white signal preserves the basic autocorrelation properties which allow it to be used in nonlinear system identification in connection with the crosscorrelation technique.

The generation of GWN in the laboratory is not a simple straightforward task. There is a variety of methods that have been used so far, but there is no universally accepted generation method of general use. These methods range from filtering of natural sources of approximately white noise (e.g. electron noise in vacuum tubes, solid state diodes noise, Zener diodes, radioactive materials etc.) to digital generation within the digital computer. The basic disadvantage of natural sources is that the spectral density usually drops significantly at low frequencies. Several techniques have been proposed to account for that

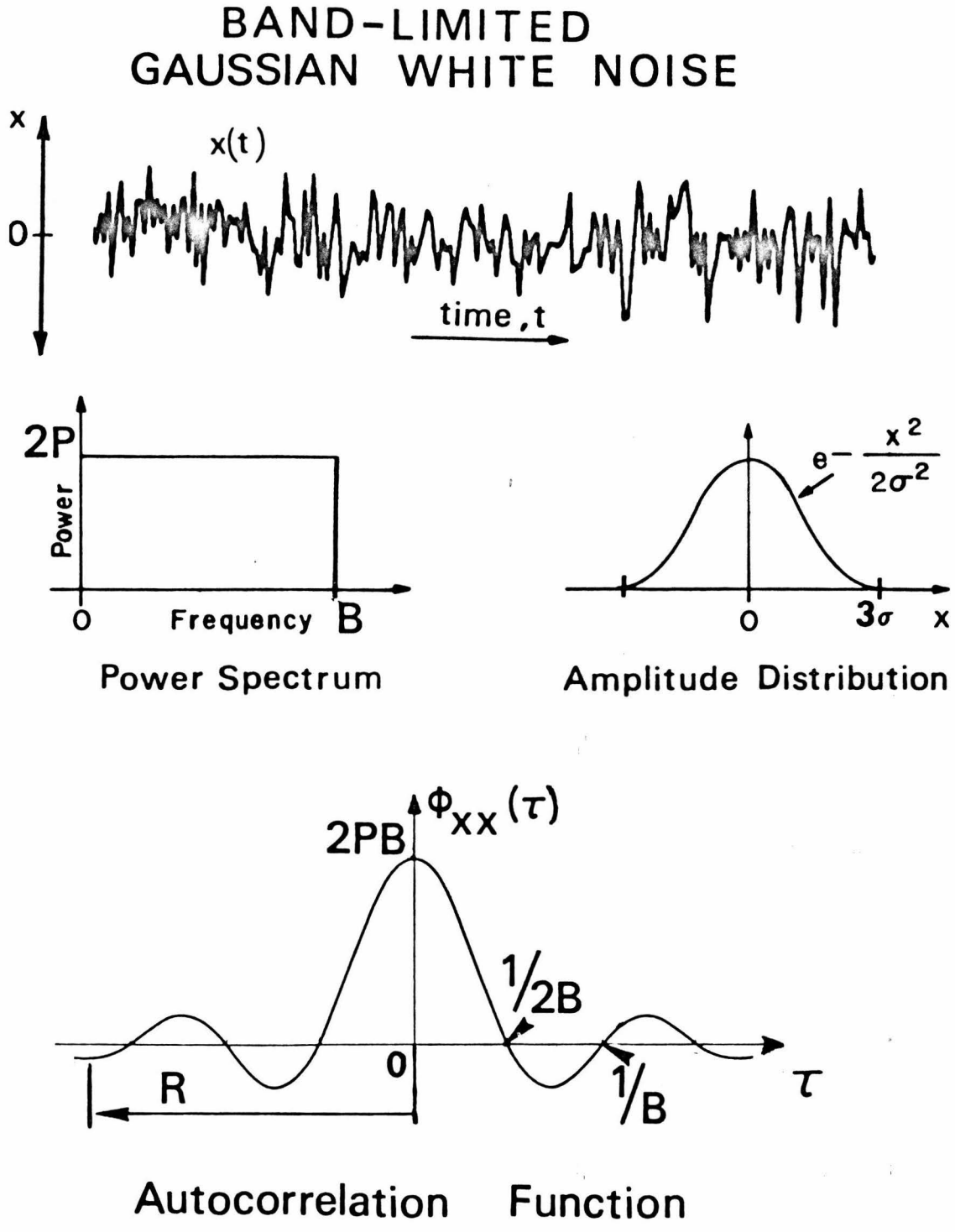


Fig. 3.2.1: Basic mathematical features of the band-limited distribution-truncated gaussian white noise

[20, 21].

A general digital generation method can be based upon a Fourier type of expansion:

$$x(t) = \sum_{n=1}^{N} a_n \sin n w_o t + b_n \cos n w_o t \; ; \; o \leq t \leq T \quad (3.2.1)$$

where, a_n and b_n are independent gaussian random variables with zero mean and variance σ^2; and w_o is a fundamental frequency determined by the specified record length T as:

$$w_o = 2\pi / T \quad (3.2.2)$$

Clearly, this representation is approximate and valid only within the interval [o, T], since it is periodic outside of it. This generation method has the advantage of being analytically well-posed and general in the sense that it enables us to generate GWN of any desirable bandwidth. Of course, it has also the disadvantage of possessing a discrete power spectrum instead of a continuous one, as well as employing the notion of the Fourier expansion which is inconsistent with the nature of a random process (aperiodic).

In any case, the generation of GWN in the laboratory is not a trivial task and it always constitutes a problem to reckon with, according to the practical requirements of the specific application at hand.

3.2.2 Autocorrelation Properties of GWN and Application in Nonlinear System Identification

Before we study the autocorrelation properties of GWN, we must state concisely what the autocorrelation properties of a signal ought to be, in order to be considered quasi-white and, consequently,

usable in connection with the crosscorrelation technique. Also, in order to facilitate the discussion around the autocorrelation functions, we have to make some basic definitions concerning the argument space of these functions.

Thus, consider the n-th order autocorrelation function of a stationary random signal x(t):

$$\phi_n(\tau_1, \tau_2, \ldots, \tau_n) = E[x(t-\tau_1)x(t-\tau_2)\ldots x(t-\tau_n)] \qquad (3.2.2)$$

Consider the point $(\tau_1, \tau_2 \ldots \tau_n)$ of the n-th dimensional space on which ϕ_n is defined. If at least two of the arguments are identical, the point is called "diagonal". If all the arguments τ_i form exhaustively pairs of identical values, the point is called "full-diagonal". If all the arguments are different, the point is called "nondiagonal". Clearly, the odd order autocorrelation functions do not have any full-diagonal points.

The white autocorrelation properties are those for which: (a) all the odd order autocorrelation functions are uniformly zero, and, (b) all the even order autocorrelation functions are zero everywhere except at the full-diagonal points.

A quasi-white signal must approximately possess these white autocorrelation properties. The approximation is due to the following moderation of requirement (b): the values of the even order autocorrelation functions of a quasi-white signal in the close neighborhood of the full-diagonal points ought to be much bigger than their values at the rest of the points of the space.

Evidently, the value of any even order autocorrelation function of ideal white noise at the full-diagonal points must be infinite, so that

its integral over the whole argument space is nonzero. This implies that the even order statistical moments of the ideal white noise are infinite, which is another manifestation of the infinite power and the physical unrealizability of this signal. Apparently, in the case of a quasi-white signal the values of the even order autocorrelation functions at the full-diagonal points are finite, and their integrals are also finite because of the moderation of requirement (b) that was described above.

Having defined the autocorrelation properties that a quasi-white signal must possess, the quasi-whiteness of GWN can be easily manifested. We only need to recall two basic properties of gaussian random variables: (1) the expected value of the product of an odd number of gaussian random variables with zero mean is zero; (2) the expected value of the product of an even number of gaussian random variables is equal to the sum of the products of the expected values of the products of all possible distinct pairs that can be formed. For example in the case of four gaussian random variables:

$$E[x_1x_2x_3x_4] = E[x_1x_2]\,E[x_3x_4] + E[x_1x_3]\,E[x_2x_4] + E[x_1x_4]\,E[x_2x_3] \qquad (3.2.3)$$

This is the well known decomposition property of the gaussian random variables, and it plays a major role in the construction of the Wiener series. Evidently, the number of terms of this sum in the case of 2m gaussian random variables is $\frac{2m!}{m!2^m}$.

The usefulness of this decomposition property in the case of the cross-

correlation technique is great. Clearly, the higher even order autocorrelation functions can be expressed in terms of only the second order autocorrelation function, as described by the decomposition property.

On the other hand, the second order autocorrelation function of GWN is: (Fig. 3.2.1)

$$\phi_2(\tau_1, \tau_2) = 2BP.\frac{\sin(2\pi B|\tau_1-\tau_2|)}{2\pi B|\tau_1-\tau_2|} \qquad (3.2.4)$$

where, B is the bandwidth of GWN and P its power level. Notice that the variance of the GWN is related to the bandwidth and the power level as:

$$\sigma^2 = 2PB \qquad (3.2.5)$$

Clearly, the values of the second order autocorrelation function are concentrated in a principal lobe around the origin. The width of the principal lobe is 1/B and it is double the width of any of the side lobes. The side lobes have clearly much smaller height than the principal lobe. In fact, the k-th side lobe extended from $|\tau_1-\tau_2| = k/2B$ to $|\tau_1-\tau_2| = (k+1)/2B$, attains its maximum for $|\tau_1-\tau_2| = (k+\frac{1}{2})/2B$ and this maximum value is $(-1)^k \cdot 2BP/(k+\frac{1}{2})\pi$. Therefore, the ratio of the heights of the principal and the k-th side lobe is (absolutely): $(k+\frac{1}{2})\pi$.

In conclusion, the form of the second order autocorrelation function of GWN in combination with the decomposition property of the gaussian random variables guarantees the fulfillment of the quasi-white autocorrelation requirements, that were previously stated, and manifests the quasi-whiteness of GWN.

The quasi-whiteness of GWN allows its use in nonlinear system identification in connection with the crosscorrelation technique. The obtained kernel estimates are approximately the Wiener kernels of the system. This approximation is usually satisfactory, if the bandwidth of the GWN covers the system bandwidth; however, several estimation errors exist, which make the kernel estimates deviate from the exact Wiener kernels of the system. This subject is discussed in the following section.

3.2.3 Estimation Errors Using GWN in Nonlinear System Identification

In order to understand the sources of the estimation errors, we must study the mechanism with which the system kernels are estimated through the crosscorrelation technique.

Consider, for example, the estimation of the first order kernel:

$$\hat{h}_1(\sigma) = \frac{1}{P} \cdot \frac{1}{T} \int_0^T y(t)x(t-\sigma)dt \qquad (3.2.5)$$

For x(t) being GWN, the Wiener functionals are only approximately orthogonal (because of the finite bandwidth), and also, the time integrations over the finite interval T does not assure the exact formation of the averages. Thus, we have two kinds of approximate orthogonality error, which, however, are usually negligible in practice for sufficiently large bandwith and record length.

Therefore, if we neglect the approximate orthogonality errors, we have:

$$\hat{h}_1(\sigma) = \frac{1}{P} \cdot \frac{1}{T} \int_0^\infty \int_0^T h_1(\tau)x(t-\tau)x(t-\sigma)dtd\tau$$

$$= \frac{1}{P} \cdot \int_0^\infty h_1(\tau) \hat{\phi}_2(\sigma-\tau)\, d\tau \qquad (3.2.6)$$

where,

$$\hat{\phi}_2(\sigma-\tau) = \frac{1}{T} \int_0^T x(t-\tau)x(t-\sigma)\, dt \qquad (3.2.7)$$

Clearly, $\hat{\phi}_2$ is an unbiased and consistent estimate of ϕ_2 (as long as the stimulus does not vanish for these values of τ and σ), and its statistical deviation from ϕ_2 is due to the finite record length. Thus we have an estimation error of statistical nature which is due to the finite record length. Evidently, this error decreases monotonically as the record length increases and it becomes negligible for sufficiently long record.

We can take an idea of the dependence of this error on the record length, by considering an infinite stimulus bandwidth and studying the variance of $\hat{h}_1$:

$$\text{Var}\,[\hat{h}_1(\sigma)] = E[\hat{h}_1^2(\sigma)] - h_1^2(\sigma)$$

$$= \frac{1}{P^2}\frac{1}{T^2} \int_0^\infty \int_0^\infty d\tau\, d\tau' h_1(\tau)h_1(\tau') \int_0^T \int_0^T E[x(t-\tau)x(t-\sigma)x(t'-\tau')x(t'-\sigma)]\, dt\, dt' - h_1^2(\sigma)$$

(for $g = t-t'$)

$$= \frac{1}{T} \cdot \int_0^\infty h_1^2(\tau)\, d\tau + \frac{2}{T} \int_0^\sigma (1-\frac{g}{T}) h_1(\sigma+g)h_1(\sigma-g)\, dg \qquad (3.2.8)$$

The first term is independent from σ, while the second term depends on σ. Clearly, the variance of the kernel estimate decreases monotonically as the record length increases, and it vanishes when T approaches infinity. This manifests the consistence of the kernel estimate in this case.

The previously derived expressions for the statistical estimation error due to the finite record length are approximately valid in the

case of GWN, if its bandwidth is broad enough with respect to the system bandwidth, so that its second order autocorrelation function resembles satisfactorily a delta function and it performs the deconvolution with acceptable accuracy for the system at hand. In any case, some loss of high frequencies in the kernel estimate will occur, which may or may not be of practical importance. It is important, however, to be aware of this kind of error and also have an idea of its extent. To this purpose, we will approximately evaluate this error in the following; assuming that the record length is infinite, in order to simplify the derivations by eliminating the statistical error. Thus, we have:

$$E[\hat{h}_1(\sigma)] = \frac{1}{P}\int_0^{\infty} h_1(\tau)\phi_2(\sigma-\tau)d\tau$$

$$= \frac{1}{P}\int_0^{\infty} h_1(\tau)2PB\,\frac{\sin(2\pi B(\sigma-\tau))}{2\pi B(\sigma-\tau)}\,d\tau \qquad (3.2.9)$$

We expand $h_1(\tau)$ in a Taylor series about σ:

$$h_1(\tau) = h_1(\sigma) + h_1^{(1)}(\sigma)(\tau-\sigma) + h_1^{(2)}(\sigma)\frac{(\tau-\sigma)^2}{2} + h_1^{(3)}(\sigma)\frac{(\tau-\sigma)^3}{6}$$

$$+ \ldots \qquad (3.2.10)$$

Since, ϕ_2 is an even function, the integration over a symmetric interval around σ will eliminate the odd order terms of the Taylor expansion. Thus, considering a symmetric interval around σ we have approximately that:

$$E[\hat{h}_1(\sigma)] \cong \frac{1}{\pi}\int_{\sigma-R}^{\sigma+R} h_1(\tau)\,\frac{\sin(2\pi B(\sigma-\tau))}{(\sigma-\tau)}\,d\tau \qquad (3.2.11)$$

where, R is an arbitrary constant considerably larger than 1/B, but still within the range of validity of the Taylor expansion (analyticity

region).

Finally,

$$E[\hat{h}_1(\sigma)] \cong \frac{1}{\pi}\sum_{n=o}^{\infty}\frac{h_1^{(2n)}(\sigma)}{2n!}\int_{-R}^{R}\xi^{2n-1}\cdot\sin(2\pi B\xi)d\xi$$

$$\cong h_1(\sigma) + \frac{2}{\pi}\sum_{n=1}^{\infty}C_n\cdot\frac{h_1^{(2n)}(\sigma)}{2n!} \qquad (3.2.12)$$

where,

$$C_n = \int_o^R \xi^{2n-1}\sin(2\pi B\xi)d\xi$$

$$= -\sum_{k=o}^{2n-1} k!\binom{2n-1}{k}\frac{R^{2n-k-1}}{(2\pi B)^{k+1}}\cos\left(2\pi BR + k\frac{\pi}{2}\right) \qquad (3.2.13)$$

Clearly, the deconvolution error that we commit because of the finite stimulus bandwidth is:

$$E[\hat{h}_1(\sigma)] - h_1(\sigma) \cong \frac{2}{\pi}\sum_{n=1}^{\infty}C_n\frac{h_1^{(2n)}(\sigma)}{2n!} \qquad (3.2.14)$$

where the coefficients C_n decrease rapidly as the bandwidth B of GWN increases. (eqn. 3.2.13) Consequently, the deconvolution error can become negligible for a sufficiently broad stimulus bandwidth.

For example, consider only the first term of the deconvolution error:

$$r_1(\sigma) = -\frac{1}{\pi}h_1^{(2)}(\sigma)\left[\frac{R}{2\pi B}\cos(2\pi BR) + \frac{1}{(2\pi B)^2}\cos\left(2\pi BR+\frac{\pi}{2}\right)\right]$$

$$\cong -\frac{1}{\pi}h_1^{(2)}(\sigma)\frac{R}{2\pi B}\cos(2\pi BR) \qquad (3.2.15)$$

it depends on the second derivative of the kernel and vanishes when B approaches infinity. Similar arguments hold in the higher order kernel cases but the analytical expressions that describe them are much more involved.

In conclusion, the deconvolution error, which is due to the finite stimulus bandwidth, can become practically negligible if a sufficiently broad stimulus bandwidth, with respect to the system bandwidth, is used. It must be emphasized, however, that we cannot extend in practice the stimulus bandwidth without cost; because, even though the increase of of the stimulus bandwidth reduces the deconvolution error, it also enhances significantly the statistical fluctuation error.

To illustrate the effect of the GWN bandwidth on the statistical fluctuation error, we will study the variance of a single sample $s_1(t, \sigma)$ of the crosscorrelation in the first order kernel case:

$$s_1(t, \sigma) = \int_0^\infty h_1(\tau)x(t-\tau)x(t-\sigma)d\tau \qquad (3.2.16)$$

Of course, the kernel estimate is obtained as a time average over these samples:

$$\hat{h}_1(\sigma) = \frac{1}{PT}\int_0^T s_1(t, \sigma)dt \qquad (3.2.17)$$

The derivations below are made for x(t) being band-limited gaussian white noise of bandwidth B:

$$\begin{aligned}
\text{Var}\,[s_1(t, \sigma)] &= \int_0^\infty\int_0^\infty h_1(\tau)h_1(\tau')\,E[x(t-\tau)x(t-\sigma)x(t-\tau')x(t-\sigma)]\,d\tau d\tau' \\
&\quad - \left\{\int_0^\infty h_1(\tau)\,E[x(t-\tau)x(t-\sigma)]\,d\tau\right\}^2 \\
&= \int_0^\infty\int_0^\infty h_1(\tau)h_1(\tau')\left\{\phi_2(\tau-\sigma)\phi_2(\tau'-\sigma)+\phi_2(o)\phi_2(\tau-\tau')\right. \\
&\qquad \left. + \phi_2(\tau-\sigma)\phi_2(\tau'-\sigma)\right\}\,d\tau d\tau' \\
&\quad - \int_0^\infty\int_0^\infty h_1(\tau)h_1(\tau')\phi_2(\tau-\sigma)\phi_2(\tau'-\sigma)d\tau d\tau' \qquad (3.2.18)
\end{aligned}$$

$$= \phi_2(o) \int_o^\infty \int_o^\infty h_1(\tau) h_1(\tau') \phi_2(\tau - \tau') d\tau d\tau' + \left[\int_o^\infty h_1(\tau) \phi_2(\sigma - \tau) d\tau \right]^2 \quad (3.2.18)$$

Consider now the Fourier transforms H_1 and Φ_2 of h_1 and ϕ_2 respectively. Then, the expression for the variance becomes:

$$\mathrm{Var}[s_1(t, \sigma)] = 2BP^2 \int_{-B}^{B} |H_1(f)|^2 \, df + P^2 \left[\int_{-B}^{B} H_1(f) e^{j2\pi f\sigma} df \right]^2 \quad (3.2.19)$$

given that:

$$\Phi_2(f) = \begin{cases} P & \text{for } -B \leq f \leq B \\ o & \text{for } |f| > B \end{cases} \quad (3.2.20)$$

Recalling that the estimation of $h_1(\sigma)$ requires the normalization of the first order crosscorrelation by $1/P$, we get approximately: (disregarding for the moment the dependence on T)

$$\mathrm{Var}[\hat{h}_1(\sigma)] \sim 2B \cdot \int_{-B}^{B} |H_1(f)|^2 df + \left[\int_{-B}^{B} H_1(f) \, e^{j2\pi f\sigma} df \right]^2 \quad (3.2.21)$$

The first term of this expression is clearly increasing with B and it is independent from σ; while, the second term depends on σ and no general statement can be made on the way it depends on B. If the GWN bandwidth B is broad enough with respect to the system bandwidth, it is reasonable to expect that the first term of the expression for the variance (eqn. 3.2.21) will be much more sensitive to changes of B than the second term. Consequently, the pattern of dependence of the kernel estimate variance upon the GWN bandwidth will be predominantly determined by the first term, which indicates a monotonically increasing pattern. Notice also that the first term is an upper bound of the second term for every value of σ, which means that the first term is also dominant in the formation of the variance of the kernel estimate.

In conclusion, considering eqns. 3.2.13 and 3.2.21, which describe the dependence of the deconvolution and the statistical fluctuation estimation errors on the GWN bandwidth, we come to the important realization that the overall effect of B upon the kernel estimation error is not monotonic. Unlike the effect of the record length, a change of the bandwidth triggers two antagonistic error producing mechanisms; the one relating to the statistical fluctuation error (cf. eqn. 3.2.21) and the other relating to the deconvolution error (cf. eqn. 3.2.13).

The existence of these two antagonistic mechanisms leads us to the expectation of an optimum value of B, for which the overall estimation error becomes minimum. Unfortunately, we cannot suggest, at present, a systematic optimization procedure with which the optimum values of the test parameters can be determined; because of the great complexity of the relevant expressions, especially in the higher order kernel cases. Notice, however, that an analysis of that sort is feasible in the case of the quasi-white signals that are introduced in chapter 4, and an optimization procedure for the test parameters can be practicably designed, as discussed in chapter 6.

Another cause of estimation error in the case of GWN is the fact that the tails of the gaussian distribution are inevitably truncated, and this results in somewhat distorting the statistical properties of the gaussian random signal. We will try here to give a quantitative idea of the error commited because of this truncation of the gaussian distribution.

Consider the truncated signal $x^*(t)$ as the superposition of a non-

truncated x(t) and an error signal $\epsilon(t)$:

$$x^*(t) = x(t) + \epsilon(t) \tag{3.2.23}$$

where,

$$\epsilon(t) = \begin{cases} 0 & \text{if} \quad |x(t)| \leq A \\ A-x & \text{if} \quad x(t) > A \\ -A-x & \text{if} \quad x(t) < -A \end{cases} \tag{3.2.24}$$

and, $\pm$ A are the truncation points of the distribution.

Clearly,

$$\text{Prob}\{x^*(t)\} = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^{*2}}{2}} & \text{if} \quad |x^*(t)| < A \\ 0 & \text{if} \quad |x^*(t)| > A \\ \text{erf}(-A)\cdot\delta(x^* \mp A) & \text{if} \quad x^*(t) = \pm A \end{cases} \tag{3.2.25}$$

for unit variance. In this formulation, the truncation can be handled as input noise $\epsilon(t)$.

Let us study the first order kernel case (assuming infinite frequency bandwidth to simplify the derivations):

$$\begin{aligned} \hat{h}_1(\sigma) &= \frac{1}{P}\int_0^\infty h_1(\tau)\, E[x^*(t-\tau)x^*(t-\sigma)]\, d\tau \\ &= \frac{1}{P}\int_0^\infty h_1(\tau)[\phi_{xx}(\sigma-\tau)+2\phi_{x\epsilon}(\sigma-\tau)+\phi_{\epsilon\epsilon}(\sigma-\tau)]\, d\tau \end{aligned} \tag{3.2.26}$$

but,

$$\phi_{xx}(\sigma-\tau) = P\cdot\delta(\sigma-\tau) \tag{3.2.27}$$

$$\begin{aligned} \phi_{x\epsilon}(\sigma-\tau) &= P\cdot\delta(\sigma-\tau)\cdot\int_{-\infty}^{\infty}\int x\epsilon P_{x\epsilon}(x,\epsilon)\,dx\,d\epsilon \\ &= P\cdot\frac{2\delta(\sigma-\tau)}{\sqrt{2\pi}}\int_A^\infty (Ax-x^2)e^{-x^2/2}dx \end{aligned} \tag{3.2.28}$$

$$\phi_{\epsilon\epsilon}(\sigma-\tau) = P\cdot\delta(\sigma-\tau)\int_{-\infty}^{\infty}\epsilon^2 p_\epsilon(\epsilon)\,d\epsilon = P\cdot\frac{2\,\delta(\sigma-\tau)}{\sqrt{2\pi}}\int_A^\infty (A-x)^2 e^{-x^2/2}dx \tag{3.2.29}$$

where, p_{ϵ} is the probability density of $\epsilon(t)$ and $P_{x\epsilon}$ the joint probability density of $\epsilon(t)$ and $x(t)$. Consequently,

$$\hat{h}_1(\sigma) = h_1(\sigma)[1+ E(A)] \tag{3.2.30}$$

where E(A) is the percentage error due to the truncation:

$$E(A) = \frac{2}{\sqrt{2\pi}} \int_A^{\infty} (A^2 - x^2)\, e^{-x^2/2} dx \tag{3.2.31}$$

Evidently, for A greater than three standard deviations, the percentage error E(A) is negligible. It must be emphasized, however, that it is crucial for the truncation to be symmetric, otherwise significant errors may be induced by the odd order autocorrelation functions.

The formulation of the truncation as an external noise signal in the input, allows the extension of the previous analysis to higher order kernel cases. Similar conclusions are derived there, but the analytical expressions become far more involved.

In conclusion, the distribution truncation error is, in most cases, practically negligible, if the truncation point is chosen to be at least three standard deviations away from the mean and the truncation is symmetric.

A last source of estimation error, that we will discuss here, is a disproportionality error of the kernel estimates which results from erroneous estimation of the power level of the stimulus used. Because of the erroneous estimation of the power level, there are deviations in the normalizing factors of the crosscorrelation estimates, which result in wrong scaling of the kernel estimates. This wrong scaling is increasingly worse in the higher order kernel estimates, because in their nor-

malizing factors the power level participates raised in some power (i. e. $Cn = 1/n!\, P^n$) and consequently the disproportionality error is magnified. To correct this error a final scaling adjustment procedure can be followed for the several estimated Wiener functionals, as discussed in sec. 6. 3.

3. 3 The Pseudorandom Signals Based on M-Sequences (PRS)

The computational burden accompanying the use of GWN in connection with the crosscorrelation technique motivated several scientists to search for specially structured signals that would reduce the natural redundancy of the random quasi-white processes, while still preserving the quasi-white autocorrelation properties.

This effort resulted in the introduction of the pseudorandom signals based on m-sequences, which are deterministic periodic signals with autocorrelation properties close to the quasi-white ones and of very high efficiency (very low redundancy). When we talk about redundancy here, we mean the repetition of identical waveform portions throughout the signal. In addition to the high efficiency, the pseudorandom signals have the advantage of easier and more reliable generation in the laboratory over the GWN. A further important advantage of the pseudorandom signals is the fact that their second order autocorrelation function is exactly zero in the region outside the origin neighborhood and within, of course, some limits determined by the period of the signal (since the autocorrelation functions are also periodic). This is an advantage over the random quasi-white signals, which exhibit small nonzero values in this region of their second order autocorrelation function that cause the

discussed statistical fluctuation error. Nevertheless, the pseudorandom signals exhibit significant imperfections in their higher even order autocorrelation functions, which spoil their superiority in the second order autocorrelation properties whenever the system possesses higher order nonlinearities.

Evidently, the pseudorandom signals are most advantageous in identification of linear systems, while the presence of nonlinearities in the system makes the choice between the random and the pseudorandom quasi-white test signals a complex one, depending upon the specific characteristics of the case at hand.

3.3.1 General Description and Generation of PRS

The pseudorandom signals based on m-sequences (hereafter denoted by PRS) have a special stair-like form. (Fig. 3.3.1) They remain constant within small finite time intervals and they switch abruptly at all the time instants which are integer multiples of a fundamental time interval Δt.

The values which they attain are determined by a linear recurrence formula of the form:

$$x_i = a_1 \otimes x_{i-1} \oplus a_2 \otimes x_{i-2} \oplus \ldots \oplus a_m \otimes x_{i-m} \qquad (3.3.1)$$

where, all values a_j and x_j correspond to the elements of a specified Galois field with finite population number, and the operations $\{\oplus, \otimes\}$ are defined in the proper way so as to be internal operations for this set of mathematical objects. For example, in the case of a binary pseudorandom signal (BPS) the Galois field has two elements, and consequently,

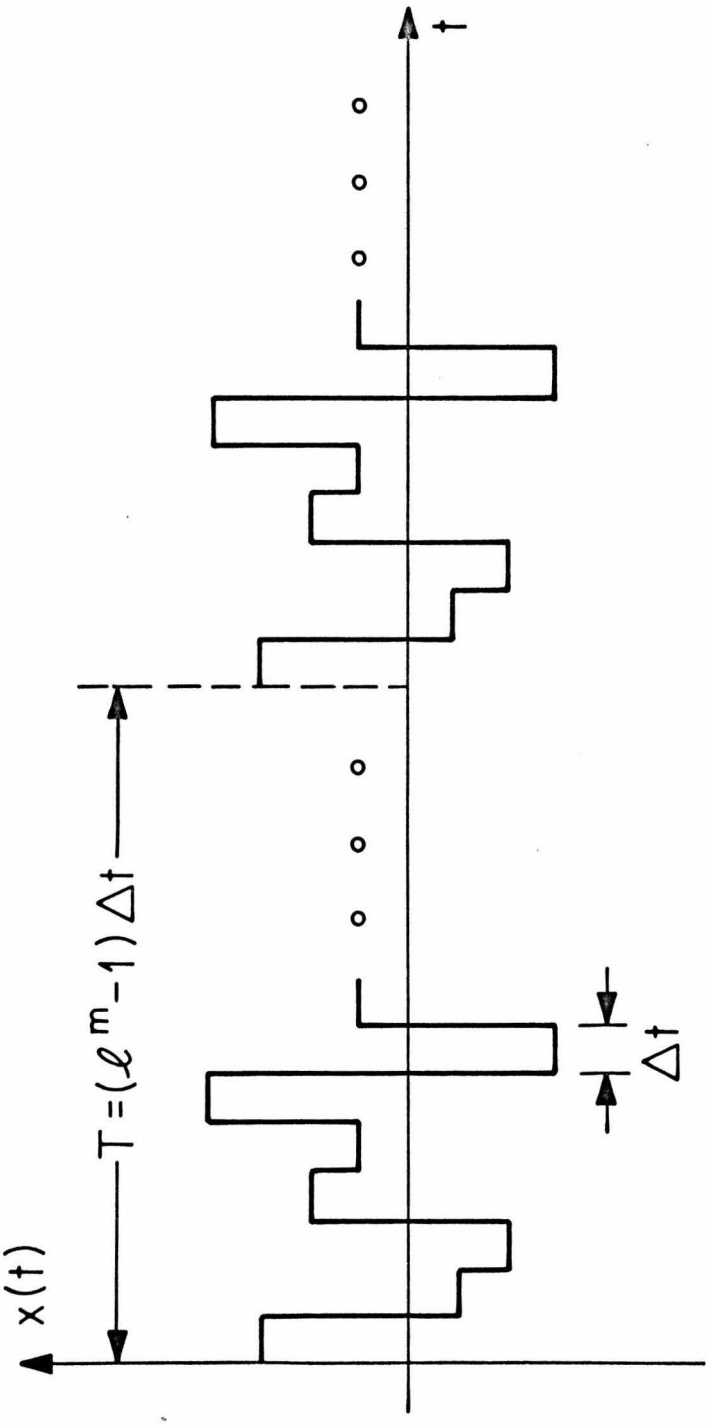


Fig. 3.3.1: Portion of pseudorandom signal

the operations $\oplus$ and $\otimes$ are defined modulo 2, so that the outcome of the recurrence formula (3.3.1) is also an element of the same Galois field.

It is evident that a sequence $\{x_i\}$, which is constructed on the basis of the linear recurrence formula (3.3.1), is periodic. The size of its period depends on the specific values of the coefficients a_j and the memory extent m of the recurrence formula (for a given Galois field). Among all the sequences $\{x_i\}$ which are constructed from the members of a certain Galois field and with linear recurrence formulas of memory extent (or order) m, there are some that have the maximum period. Evidently, this maximum period is $(\ell^m - 1)$, where ℓ is the number of elements of the Galois field; since ℓ^m is the number of all possible distinct arrangements with repetitions of ℓ elements in strings of m, and the null string is excluded.

These maximum period sequences are called m-sequences, and they correspond to a special choice of the coefficients $\{a_1, \ldots, a_m\}$. It was found that these special coefficients $\{a_1, \ldots, a_m\}$ coincide with the coefficients of a primitive (or irreducible) polynomial of degree (m-1) in the respective Galois field. [22] Thus, we can always determine the number of elements ℓ and the order of the recurrence formula m in such a way that we get an m-sequence with a desirable period (within the limitations posed by the integral nature of ℓ and m).

The initial string of m values of x_j with which the construction of the m-sequence originates is not of importance. Any initial string (except the null one) will give the same m-sequence (for a given set of coefficients a_j) merely shifted. In the following section, we will dis-

cuss why the m-sequences exhibit the desirable properties that allow us to use them in nonlinear system identification in connection with the crosscorrelation technique.

The generation of pseudorandom signals in the laboratory is a relatively simple task. Suppose we have decided upon the number of values that the signal will attain and the required maximum period (i. e. the order of the linear recurrence formula 3. 3. 1). Now, we simply need to know the coefficients of a primitive polynomial of the specified degree in the respective Galois field. Suppose that such a primitive polynomial is found (tables of such polynomials may be very helpful [23]). Then, we choose an initial string of values and we construct the corresponding m-sequence with a digital computer using the linear recurrence formula 3. 3. 1. The resulting sequence of numbers can subsequently be fed into a digital-to-analog transducer to generate the desired pseudorandom signal.

More specialized pieces of hardware can also be used for more efficient generation. For example, a binary m-sequence can be generated through a digital shift-register generator, which is briefly described below. (Fig. 3. 3. 2) The digital shift-register of n stages is a cascade of n flip-flops. Every flip-flop stores one binary "bit": 0 or 1. At every pulse of a clock the content of each flip-flop is forwarded to its neighbor to the right. Simultaneously, an "exclusive OR" gate combines the bits of two proper stages and feeds the outcome back to the first stage and the process continues. The output bits of the "exclusive OR" gate has been proven to constitute a binary pseudorandom m-

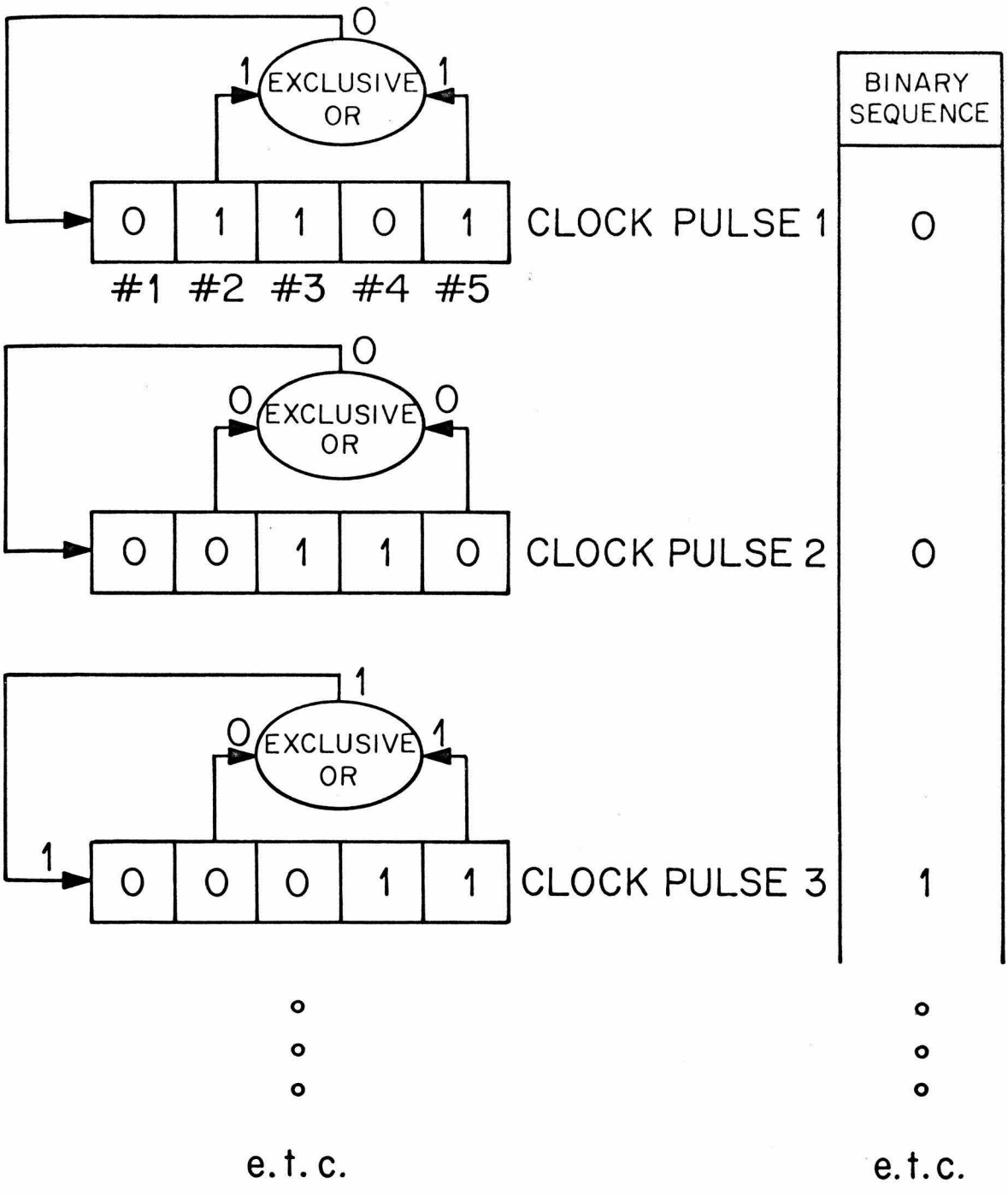


Fig. 3.3.2: Schematic representation of shift-register generator of a pseudorandom binary sequence

sequence. The period of this sequence is determined by the number of stages in the shift-register, as shown in Table 3.3.1, along with the order of the stage that provides (besides the right most stage) the other input to the "exclusive OR" gate.

3.3.2 Autocorrelation Properties of PRS and Application in Nonlinear System Identification

The quasi-whiteness of a signal (and consequently its use in connection with the crosscorrelation technique) is manifested through the proper autocorrelation properties, as they were stated in sec. 3.2.2. The pseudorandom signals based on m-sequences (PRS) are exhibiting these proper autocorrelation properties within some reasonable approximation.

This is due to the shift-and-add property of the m-sequences [24]. According to this property the product (of the proper modulo) of any number of sequence elements is another sequence element:

$$x_{k-j_1} \otimes x_{k-j_2} \otimes \cdots \otimes x_{k-j_m} = x_{k-\ell} \qquad (3.3.2)$$

where, ℓ depends on $j_1, j_2, \ldots, j_m$ but not on k. As a result of this shift-and-add property and the basic structural characteristics of the m-sequences (i.e. maximum period and antisymmetry), the odd order autocorrelation functions are uniformly zero everywhere and the even order ones approximate satisfactorily the quasi-whiteness requirements.

It must be noted that a slight modification must be made on the m-sequences with even number of levels in order to rectify the antisym-

TABLE 3.3.1

Number of stages in shift-register	Order of gate-feeding stage in addition to the rightmost one	Sequence length in bits
5	2	31
6	1	63
7	1 or 3	127
9	4	511
10	3	1,023
11	2	2,047
15	1, 4, or 7	32,767
18	7	262,143
20	3	1,048,575
21	2	2,097,151
22	1	4,194,303
23	5 or 9	8,388,607
25	3 or 7	33,554,431
28	3, 9, or 13	268,435,455
31	3, 6, 7, or 13	2,147,483,647

metric property. This slight modification has been illustrated, for example, by H. R. Simpson [25] in the case of a binary m-sequence, and simply comprises the inversion of every other bit of the m-sequence to obtain what he calls the corresponding n-sequence. This modification results in doubling the period of the sequence, thus alleviating the cause of the problem; namely, the fact that the number of elements in one period of a sequence with even number of levels is odd.

Because of the antisymmetric property, the odd order autocorrelation functions are perfect (i. e. uniformly zero). Nevertheless, the even order autocorrelation functions of order higher than the second exhibit some serious imperfections (anomalies), which constitute an important topic of study and source of controversy about the use of PRS in nonlinear system identification.

The second order autocorrelation function is zero everywhere (within a period) except in the neighborhood of the origin where it is triangularly shaped, as shown in Fig. 3.3.3. This establishes the PRS as a very effective tool in linear system identification. However, the higher even order autocorrelation functions exhibit some anomalies distributed all around their argument space and, therefore, the effectiveness of the PRS in nonlinear system identification is significantly reduced. E. P. Gyftopoulos and R. J. Hooper observed these anomalies [26, 27] and H. A. Barker and R. Pradisthayon studied their origin and properties [28]. They showed that these anomalies are due to existing linear relationships among the elements of the sequence, and their exact position and magnitude can be determined through a laborious algorithm

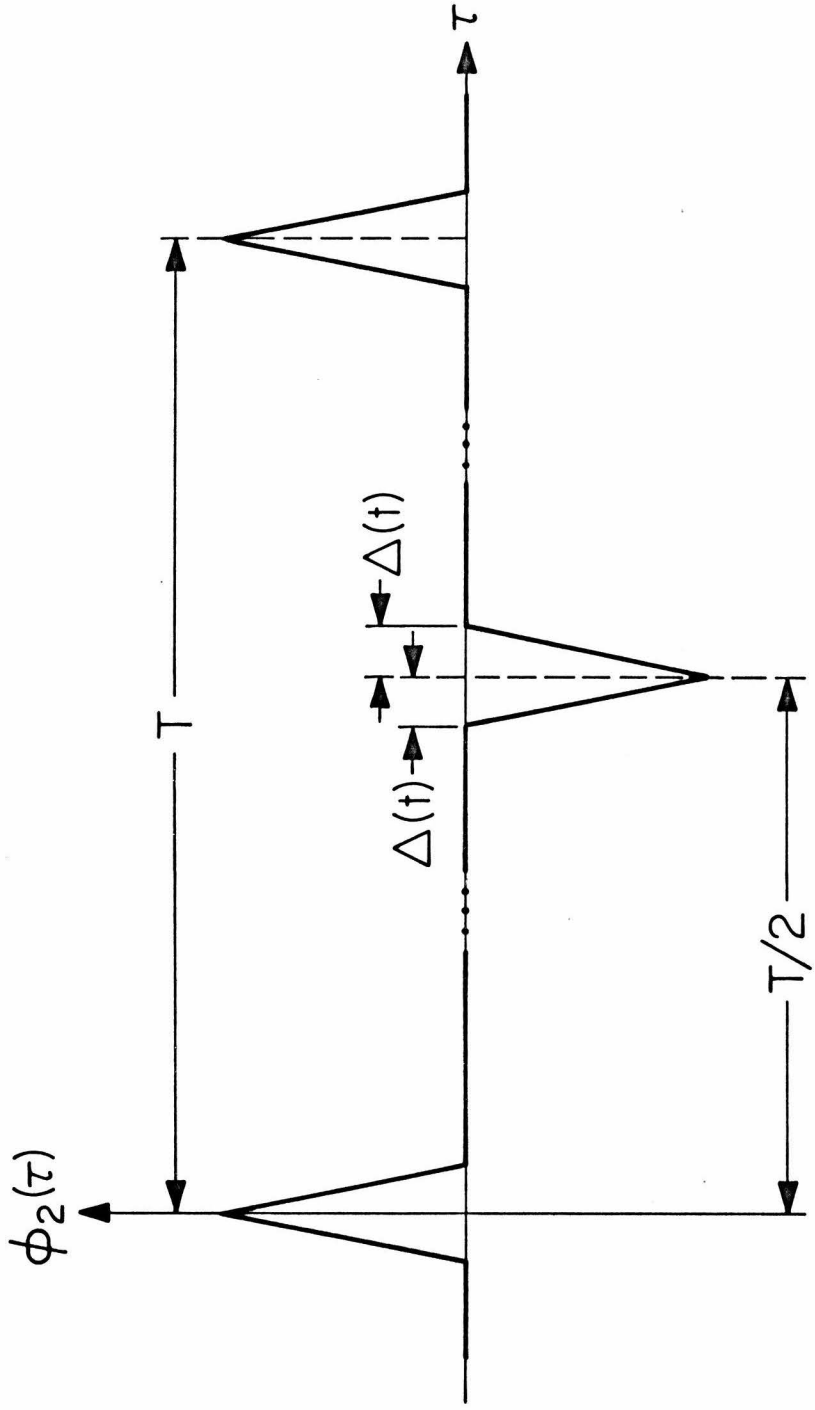


Fig. 3.3.3: Second order autocorrelation function of a pseudorandom signal

relating to polynomial division. In any case, these anomalies are proved to be inherent and inevitable characteristics of the m-sequences, directly and tightly related to their deterministic nature and their mathematical structure.

These anomalies cause estimation errors whenever nonlinearities are present. The magnitude of these estimation errors, as well as their relative severity (as compared to GWN) is still a subject under study. It is evident, however, that this depends on the specific system under study, as well as on the specific PRS or GWN which we happened to use. Barker et al. studied several PRS (binary, ternary and quinary) trying to compare their relative virtues, and they determined some optimum PRS, according to a criterion of performance weighting the amount of exhibited anomalies [29].

Despite the presence of anomalies, the PRS have been proven useful and efficient quasi-white signals, providing satisfactory results in some applications [30, 31, 32, 33]. Whether or not they should be preferable to other quasi-white signals depends on considerations of the specific case at hand.

A very important point, which ought to be emphasized, is that the functional series, which are orthogonal with respect to a specific PRS, are slightly modified, as compared to the original Wiener series, in order to accommodate the moments of the PRS (which are different, in general, from the moments of GWN). Consequently, the estimated kernels by the use of a PRS are different from the Wiener or Volterra kernels of the system and they are characteristic of the specific PRS and

the corresponding functional series. Of course, the overall model is equivalent to the Wiener or Volterra model; however, the individual kernels or functionals are different. As a result of this, if a truncated model is used to approximate the system (which is the usual case in practice) this model is different, in general, whenever the moments or the elementary step Δt of the PRS change.

This is easily seen from the relations of the estimated kernels (in each case) with the Volterra kernels of the system. To illustrate this, let us consider the estimation of the zero order kernel h_o of the zero-memory system:

$$y = e^{x} \tag{3.3.3}$$

using a PRS:

$$\begin{aligned} \hat{h}_o &= E\left[e^{x}\right] \\ &= \sum_{n=o}^{\infty} \frac{E[x^n]}{n!} \\ &= \sum_{n=o}^{\infty} \frac{m_{2n}}{2n!} \end{aligned} \tag{3.3.4}$$

where, m_{2n} is the 2n-th moment of the PRS.

Similar expressions hold for all orders of kernels and for the finite memory systems. They clearly demonstrate the dependence of the estimated kernels upon the moments of the PRS used. Further, if we attempt to orthogonalize the functional series with respect to a certain PRS, in the same way that Wiener orthogonalized the functional series with respect to GWN, we will realize that the resulting orthogonal

functionals depend upon the even order moments of the respective PRS. Thus, both the kernels and the functionals depend upon the moments of the respective PRS.

There are two basic ways in which the moments of a quasi-white signal can change. The one is by changing the probability distribution of the signal amplitude; the other is by simply scaling the signal amplitudes with the same factor without changing the probability distribution. These two cases are physically distinct, because the first one results in change of the information content of the signal (according to Shannon's definition), while the second one leaves the information content invariant.

We expect that these two distinct ways of changing the moments will have qualitatively different effects upon the corresponding functional series. This is a topic of great interest, since the structure and the convergence of the functional series is always a principal concern in the applications. We will study this subject more extensively in the following chapters in regard to the family of the random quasi-white signals that are introduced in chapter 4.

3.3.3 Estimation Errors Using PRS in Nonlinear System Identification

The PRS are periodic deterministic signals. Consequently, they do not exhibit any kind of statistical estimation error. However, they exhibit several other kinds of estimation error.

(1) Deconvolution error, i.e. the error due to the finite stimulus bandwidth, which results in some loss of high frequencies during the

deconvolution process. Because of the stair-like form of PRS, the principal lobes of the even order autocorrelation functions have a pyramidal form. Thus, if we expand the corresponding kernel into a multidimensional Taylor series we can analytically evaluate this deconvolution error $\Theta_n(\sigma_1, \ldots, \sigma_n)$ for the n-th order kernel estimate:

$$\Theta_n(\sigma_1, \ldots, \sigma_n) = \sum_{m=1}^{\infty} \frac{\Delta t^{2m}}{2m!(2m+1)(m+1)} \sum_{j_1, \ldots, j_{2m}=1}^{n} \frac{\partial^{2m} h_n(\sigma_1, \ldots, \sigma_n)}{\partial \sigma_{j_1} \ldots \partial \sigma_{j_{2m}}}$$

$$j_1 \leq j_2 \leq \ldots \leq j_{2m}$$

$$(\sigma_k \geq \Delta t \; ; \; k=1, 2, \ldots, n) \qquad (3.3.5)$$

where, Δt is the elementary step length of the PRS (Fig. 3.3.1) More details on the analytical evaluation of this error are given in sec. 5.1, in the analogous situation of a CSRS. This deconvolution error apparently becomes negligible for a sufficiently small Δt (as compared to the system bandwidth).

(2) Autocorrelation anomalies error, i.e. the error induced by the anomalies present in the higher even order autocorrelation functions, as discussed in sec. 3.3.2. It is presently very laborious to determine the actual magnitude of this error, and consequently it remains an individual concern in each specific case. It is definitely the most serious kind of error in the use of PRS and the most significant drawback in their use in nonlinear system identification. Several methods have been suggested [29] and others can be devised, to reduce the effect of these anomalies; nevertheless, the overall appreciation of the situation is still

a matter of case-to-case personal judgment.

(3) Finite transition time error, i. e. the error due to the finite response (rising) time of the input transducer. This subject has been studied by K. R. Godfrey et al. for the case of a binary pseudorandom signal [34, 35]. Their study suggests the negligibility of the resulting estimation error in the case of reversible transitions (i. e. when the patterns of upward and downward transitions are similar), and for reasonably small response time of the transducer (as compared to the system bandwidth). In conclusion, sufficient care must be taken of the transducer in order to make this kind of error negligible.

Finally, there are approximate orthogonality errors, as in the case of GWN, which are usually negligible for all practical purposes. The erroneous power level and computational errors are just as described later in the case of CSRS (cf. sec. 5.4 and 5.6). These errors become practically negligible under some easily implemented provisions described later in the case of CSRS.

CHAPTER IV

THE FAMILY OF CONSTANT-SWITCHING-PACE SYMMETRIC RANDOM SIGNALS (CSRS) AND THEIR USE IN NONLINEAR SYSTEM IDENTIFICATION

In the previous chapter we discussed the properties of two families of quasi-white signals that have been used so far in applications of the Wiener-crosscorrelation method.

In this chapter, we will introduce a new family of quasi-white random signals, that can be used in connection with the crosscorrelation technique for the identification of nonlinear systems. These new test signals combine the stair-like form of the PRS with the random character of GWN, and consequently, they exhibit a new blend of advantages and disadvantages.

We believe that the introduction of this new family of quasi-white signals will contribute largely in the better understanding of the underlying mechanisms of the crosscorrelation technique and will provide significant advantages in several applications of the Wiener method over the quasi-white signals that have been used so far.

4.1 General Description and Generation of CSRS

There are two basic defining characteristics of each member of the CSRS family:

(1) the value of the signal switches randomly and independently at all time instants which are integral multiples of an

elementary finite time interval Δt, attaining values according to a symmetric probability density function, and

(2) the value of the signal remains constant between two successive switching times. (Fig. 4.1.1)

The fundamental time interval Δt is called the "step" of the CSRS and it directly determines the bandwidth of the signal. Evidently, when the step Δt is decreasing, the bandwidth of the signal is increasing and it is asymptotically approximating the ideal white noise. The symmetric probability density function $p(x)$ (with which a CSRS $x(t)$ is generated) has zero mean, and consequently, all its odd order moments are zero. Additionally, the even order moments of $p(x)$ ought to exit. This is a condition that is always satisfied in practice, since we must have a finite domain for the probability density function when we generate the signal in the laboratory. Similarly, the probability density function becomes, in practice, always discrete, when a digital computer is used to generate the signal. The limits of this discretization are posed by the word length of the computer. In theory, however, the mathematical definition of a CSRS allows the probability density $p(x)$ to be continuous and of infinite domain, as long as the even order moments exist.

The stair-like form of a CSRS is a feature compatible with the use of digital computers in data processing. The handling of continuous signals with digital computers requires their discretization through a constant rate sampling procedure; while the complete recovery of the

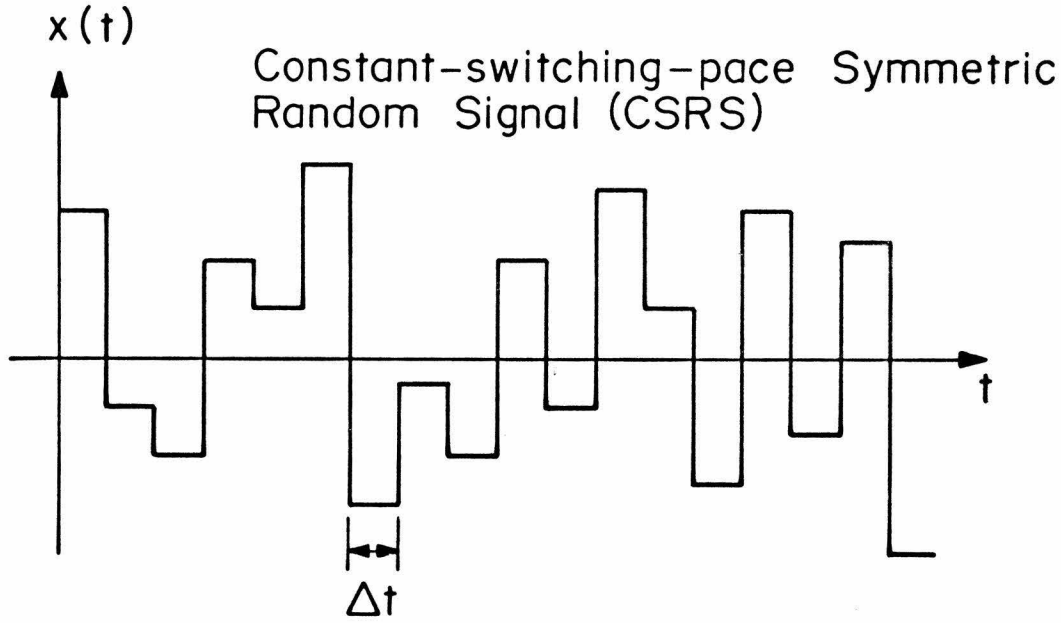


Fig. 4.1.1: Portion of a CSRS

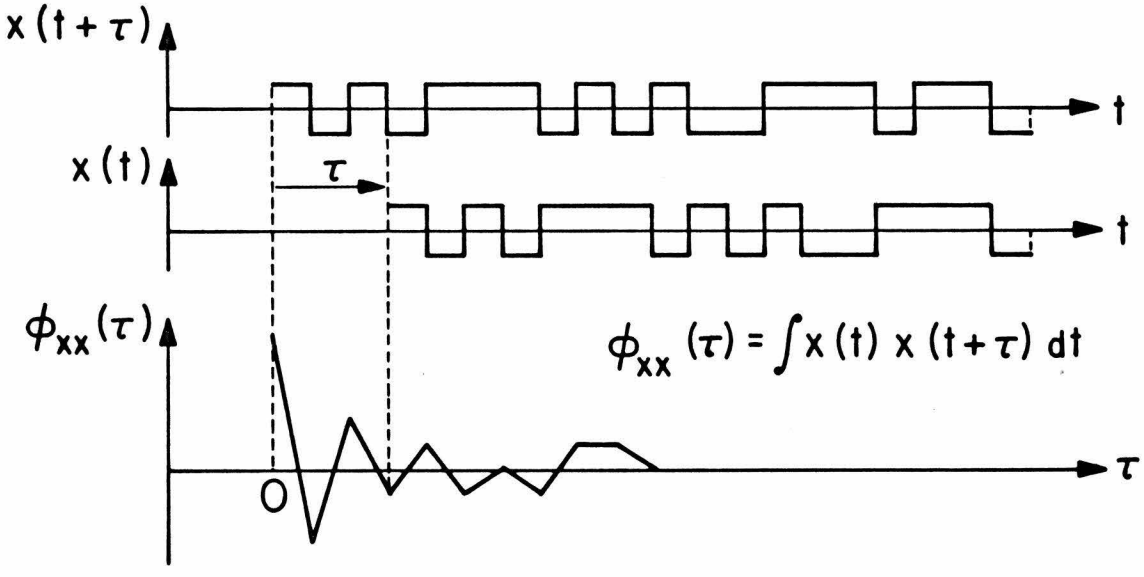


Fig. 4.1.2: Illustration of the linear form of the aggregated value of a product of stair-like signals

original signal is not (in general) possible, unless the signal has a special form (stair-like, piecewise linear etc.). Thus, the inevitable discretization of a continuous signal within the digital computer leads naturally to the conception of a stair-like constant-switching-pace signal; eliminating the errors resulting from the discretization and assuring the complete recovery of the original signal.

Clearly, the defining prescriptions of the CSRS family are fairly general and simple, providing the user with great flexibility in choosing the test signal that fits best the special considerations of a specific problem. In the following section, it will be shown that merely the statistical independence of any two steps of a CSRS, along with the fact that all the odd order moments of $p(x)$ are zero and the even order ones exist, are sufficient to guarantee the quasi-whiteness of $x(t)$ within the frequency range posed by the specified step length Δt.

4.2 The Autocorrelation Properties of CSRS

In sec. 3.2.2, it was stated what the autocorrelation properties of a signal ought to be in order to consider it a quasi-white signal. In this section, we will show that the CSRS family possesses these autocorrelation properties, and consequently, it can be used in nonlinear system identification through the crosscorrelation technique.

At first, we note that every member of the CSRS family is by construction a stationary and ergodic process. For a given step length Δt and a proper probability distribution $p(x)$, an ensemble of random processes $x(t)$ is defined within the CSRS family. The ergodicity of $x(t)$

allows us to define the autocorrelation functions in both the temporal and the probabilistic sense. Thus, the n-th order autocorrelation of x(t) is:

$$\phi_n(\tau_1, \tau_2, \ldots, \tau_n) = E[x(t-\tau_1)\, x(t-\tau_2) \ldots x(t-\tau_n)]$$
$$= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t-\tau_1)x(t-\tau_2)\ldots x(t-\tau_n)\,dt \qquad (4.2.1)$$

We will show that the autocorrelation functions, as defined by eqn. 4.2.1 , are the ones of a quasi-white signal. First, let us consider only the values of the arguments $\tau_1, \tau_2, \ldots, \tau_n$ which are integral multiples of Δt. In this case the several $x(t-\tau_1), x(t-\tau_2), \ldots, x(t-\tau_n)$ are random variables which are either statistically independent or completely identical (if their shifting times are equal). In order to facilitate the several derivations, let us make the following definitions:

The n-th order moment of the random variable x, which follows the probability distribution p(x):

$$m_n = \int_{-\infty}^{\infty} x^n\, p(x)\, dx \qquad (4.2.2)$$

The product of n time-shifted values of the random process x(t):

$$\rho_n = x(t-\tau_1)\ldots x(t-\tau_n) \qquad (4.2.3)$$

The groups $g_1, g_2, \ldots, g_k$ of the factors $x(t-\tau_i)$ of the product ρ_n, which are identical one to another, i.e., they have the same shift. The population numbers $n_1, n_2, \ldots, n_k$ of the groups $g_1, g_2, \ldots, g_k$. Clearly:

$$n = n_1 + n_2 + \ldots + n_k \qquad (4.2.4)$$

The subproducts of the members of each group g_j:

$$\pi_j = x(t-\tau_{j1})\, x(t-\tau_{j2})\ldots x(t-\tau_{jnj})$$

Clearly:
$$\rho_n = \pi_1 \cdot \pi_2 \cdot \ldots \cdot \pi_k \qquad (4.2.5)$$

Notice that the several π_i are statistically independent. A nodal point is a point $(\tau_1, \tau_2, \ldots, \tau_n)$ of the space, where the values of $\tau_1, \tau_2, \ldots, \tau_n$ are integral multiples of Δt. Order of a nodal point $(\tau_1, \tau_2, \ldots, \tau_n)$ (or of the corresponding product ρ_n) is the set $(n_1, n_2, \ldots, n_k)$ of the population numbers of the corresponding groups $g_1, g_2, \ldots, g_k$ in the product ρ_n. Diagonal point is a point $(\tau_1, \tau_2, \ldots, \tau_n)$ for which at least two of the arguments are equal. Full-diagonal point is a point $(\tau_1, \tau_2, \ldots, \tau_n)$ for which all the arguments form exhaustively pairs of equal values. Simple diagonal is a diagonal point which is not full-diagonal. The rest of the points in the space are called nondiagonal.

The following is the basic lemma of our derivations: Consider the population numbers $n_1, n_2, \ldots, n_k$ of the groups $g_1, g_2, \ldots, g_k$ of identical factors in the product: $\rho_n = x(t-\tau_1) \cdot x(t-\tau_2) \cdot \ldots \cdot x(t-\tau_n)$ where $x(t)$ is a CSRS and $(\tau_1, \tau_2, \ldots, \tau_n)$ is a nodal point of the n-dimensional space. The expected value of ρ_n is zero if and only if at least one of the n_i's is odd.

Proof: Consider the subproducts π_i as previously. Clearly:

$$\rho_n = \pi_1 \cdot \pi_2 \cdot \ldots \cdot \pi_k$$

Since the several π_i are statistically independent:

$$E[\rho_n] = E[\pi_1] \cdot E[\pi_2] \cdot \ldots \cdot E[\pi_k] \qquad (4.2.6)$$

But $$E[\pi_i] = m_{n_i} \qquad (4.2.7)$$

Therefore: $$E[\rho_n] = m_{n_1} \cdot m_{n2} \cdot \ldots \cdot m_{nk} \qquad (4.2.8)$$

Thus, if at least one n_i is odd, then the corresponding moment m_{n_i} will be zero (since all the odd moments are zero) and subsequently: $E[\rho_n] = 0$.

Note that all the even order moments are positive. Q. E. D.

This basic lemma manifests the quasi-whiteness of the members of the CSRS family -- at least at the nodal points -- because it guarantees that all the odd order autocorrelation functions are uniformly zero, while the even order ones are nonzero only at the full-diagonal points, i.e., where the arguments form exhaustively pairs of identical values.

In practice, however, we always have finite length sample signals and none of the expressions in eqn. 4.2.1 can be actually evaluated. Hence, we are practically restricted in obtaining only an estimate of the autocorrelation function, usually by time averaging, as:

$$\hat{\phi}_n(\tau_1, \tau_2, \ldots, \tau_n) = \frac{1}{T-Tm} \int_{Tm}^{T} x(t-\tau_1)x(t-\tau_2)\ldots x(t-\tau_n)\,dt \qquad (4.2.9)$$

where T is the temporal length of the sample signal x(t) and

$$T_m = \max [\tau_1, \tau_2, \ldots, \tau_n].$$

The estimate $\hat{\phi}_n(\tau_1, \ldots, \tau_n)$ is a random variable itself and its statistical properties must be studied in order to achieve an understanding of the kernel estimates obtained by the crosscorrelation technique.

Consider again only the nodal points of the argument space of $\hat{\phi}_n(\tau_1, \tau_2, \ldots, \tau_n)$. Clearly the expected value of $\hat{\phi}_n(\tau_1, \ldots, \tau_n)$ is $\phi_n(\tau_1, \ldots, \tau_n)$, which makes it an unbiased estimate. Also, the probability limit of $\hat{\phi}_n(\tau_1, \ldots, \tau_n)$ is $\phi_n(\tau_1, \ldots, \tau_n)$, which makes it a consistent estimate. Now, let's look at the second moment of $\hat{\phi}_n(\tau_1, \ldots, \tau_n)$:

$$E[\hat{\phi}_n^2(\tau_1, \ldots, \tau_n)] = \frac{1}{(T-Tm)^2} \int_{Tm}^{T}\int_{Tm}^{T} x(t-\tau_1) \cdot \ldots \cdot x(t-\tau_n) \cdot x(t'-\tau_1) \cdot \ldots \cdot x(t'-\tau_n)\,dt\,dt'$$

$$= \frac{1}{(T-Tm)^2} \int_{t=Tm}^{T} \int_{\lambda=-(T-t)}^{(T-\tau)} x(t-\tau_1) \cdot \ldots \cdot x(t-\tau_n) \cdot x(t-\tau_1+\lambda) \cdot \ldots \cdot x(t-\tau_n+\lambda)\,dt\,d\lambda \qquad (4.2.10)$$

where, $\lambda = t' - t$.

If the point $(\tau_1, \tau_2, \ldots, \tau_n)$ is not a full-diagonal point then the only full-diagonal point among the points $(\tau_1, \tau_2, \ldots, \tau_n, \tau_1 - \lambda, \ldots, \tau_n - \lambda)$, for all the values of λ which are integral multiples of Δt, is the one for $\lambda = 0$. In this case, if the point $(\tau_1, \tau_2, \ldots, \tau_n)$ is of order $(n_1, n_2, \ldots, n_k)$, the point $(\tau_1, \tau_2, \ldots, \tau_n, \tau_1, \tau_2, \ldots, \tau_n)$ will be of order $(2n_1, 2n_2, \ldots, 2n_k)$. The expected value of $\hat{\phi}_n(\tau_1, \tau_2, \ldots, \tau_n)$ for all the points which are not full-diagonal is zero. Therefore, the variance in this case becomes: (for $\Delta t << Tm << T$)

$$\text{Var}\left[\hat{\phi}_n(\tau_1, \tau_2, \ldots, \tau_n)\right] \cong \frac{\Delta t}{T} \cdot m_{2n_1} \cdot m_{2n_2} \cdot \ldots \cdot m_{2n_k} \qquad (4.2.11)$$

Clearly, the variance at these points depends on their order. If the point $(\tau_1, \tau_2, \ldots, \tau_n)$ is full-diagonal, then all the values of λ, which are integral multiples of Δt, give full-diagonal points $(\tau_1, \tau_2, \ldots, \tau_n, \tau_1 - \lambda, \tau_2 - \lambda, \ldots, \tau_n - \lambda)$ of several orders. An explicit and general expression for the variance of $\hat{\phi}_n(\tau_1, \ldots, \tau_n)$ cannot be derived because it depends on the relative position of the arguments.

In any case, the variance of $\hat{\phi}_n(\tau_1, \ldots, \tau_n)$ at all the nodal points tends to zero asymptotically with the record length, and the values of $\hat{\phi}_n(\tau_1, \ldots, \tau_n)$ tend to the ones determined by the basic lemma. The study of the first two moments of $\hat{\phi}_n(\tau_1, \tau_2, \ldots, \tau_n)$ suffices for the description of its whole statistical behavior, because the probability distribution of the values of $\hat{\phi}_n(\tau_1, \tau_2, \ldots, \tau_n)$ at any nodal point $(\tau_1, \tau_2, \ldots, \tau_n)$ is approximately gaussian.

This can be easily seen by considering the way of actual evaluation of $\hat{\phi}_n(\tau_1, \tau_2, \ldots, \tau_n)$:

$$\hat{\phi}_n(\tau_1=k_1\cdot\Delta t, \tau_2=k_2\cdot\Delta t, \ldots, \tau_n=k_n\cdot\Delta t) = \frac{1}{N}\sum_{k=1}^{N} x(k\,\Delta t-k_1\Delta t) \cdot x(k\Delta t-k_2\Delta t)\cdot\ldots\cdot x(k\Delta t-k_n\Delta t) \quad (4.2.12)$$

where, $T = N \cdot \Delta t$.

Consider one of the terms of the sum:

$$v(k) = x(k\Delta t-k_1\Delta t)\cdot x(k\Delta t-k_2\Delta t)\cdot\ldots\cdot x(k\,\Delta t-k_n\Delta t) \quad (4.2.13)$$

The samples $x(k\Delta t-k_j\Delta t)$ are following the probability distribution $p(x)$, and they are either statistically independent or completely identical to one another (depending on whether their arguments $(k\Delta t-k_j\Delta t)$ are respectively different or identical). For given $(k_1, k_2, \ldots, k_n)$, the several $v(k)$ are uncorrelated but not necessarily statistically independent. Actually, $v(k)$ and $v(k^*)$ can be statistically dependent if $|k^*-k| \le \max[k_1, k_2, \ldots, k_n] = L$.

Suppose now we form the partial sums $U(\ell)$:

$$U(\ell) = \sum_{i=0}^{M-1} v(\ell + i\cdot L) \quad (4.2.14)$$

where, $\ell = 1, 2, 3, \ldots, L$ and M = integer part of $[N/L]$.

The several terms of these partial sums are clearly statistically independent and consequently the Central Limit Theorem is applicable, dictating that, if M is sufficiently large, then each $U(\ell)$ is approximately a gaussian random variable. It is also evident that all the $U(\ell)$ are uncorrelated. Therefore, since:

$$\hat{\phi}_n(\tau_1, \tau_2, \ldots, \tau_n) = \frac{1}{N}\sum_{\ell=1}^{L} U(\ell) \quad (4.2.15)$$

and $U(\ell)$ are uncorrelated gaussian random variables, $\hat{\phi}_n(\tau_1, \tau_2, \ldots, \tau_n)$ is also a gaussian random variable.

So far, our discussion has been limited to the nodal points of the space. To complete our study we have to prove that the other points

of the space of the autocorrelation function have also the desirable values. To this purpose, we only need to employ a basic property of the aggregated values of products of shifted stair-like signals. A stair-like signal is one which remains constant during finite time intervals and jumps abruptly from one value level to another. It is a piecewise continuous time function and its first derivative is a series of impulses.

Consider now this signal being shifted in time by τ and form the product of the original and the shifted signal. The integral of this product over time (it is shown in Fig. 4.2.1 for a binary stair-like signal) is a piecewise linear function of the shifting time with discontinuous first derivatives at all the shifting times for which at least two of the jumping points coincide. The derivative remains constant between two successive points of discontinuity. Consider now n shifted stair-like signals $x(t-\tau_1)$, $x(t-\tau_2), \ldots, x(t-\tau_n)$. Form the product of them and integrate this product over time. Similar arguments hold in this generalized case. The aggregate value is of first degree (sectionally linear) with respect to the shifting times $\tau_1, \tau_2, \ldots, \tau_n$, and subsequently it has piecewise continuous first order partial derivatives with constant values between two successive points of discontinuity (i.e., the partial derivatives are stair-like functions).

Employing this basic property of the stair-like functions, we conclude that the (n + 1)-dimensional surface $\hat{\phi}_n(\tau_1, \ldots, \tau_n)$ is of first degree with respect to the arguments $\tau_1, \ldots, \tau_n$ and its first order partial derivatives are stair-like functions. Employing also the fact that a point of derivative discontinuity appears at those shifting times

at which at least two jumping points of the signals coincide, we conclude that the locus of the points of derivative discontinuity are hyperplanes of the (n+1)-dimensional space determined by analytical relations of the form:

$$\tau_i - \tau_j = k \cdot \Delta t \qquad (4.2.16)$$

where $i, j = 1, 2, \ldots, n$, and k is an integer (positive, negative or zero). Clearly, each of these hyperplanes is parallel to (n-1) of the (n+1) axes of the space.

Notice that the nodal points of the space are in the cross sections of these hyperplanes, and consequently they are the apexes of the surface $\hat{\phi}_n(\tau_1, \ldots, \tau_n)$. To illustrate this, consider the aggregate value:

$$\hat{\phi}_3(\tau_1, \tau_2, \tau_3) = \frac{1}{T - \max[\tau_1, \tau_2, \tau_3]} \int_{\max[\tau_1, \tau_2, \tau_3]}^{T} x(t-\tau_1)\, x(t-\tau_2)\, x(t-\tau_3)\, dt \qquad (4.2.17)$$

which is a four-dimensional surface.

In Fig. 4.2.2., the locus of the points of derivative discontinuity for a cross section of the space at $\tau_3 = 0$ is shown with dotted line.

Therefore, the general morphology of a surface $\hat{\phi}_n(\tau_1, \ldots, \tau_n)$ appears to have apexes corresponding to the nodal points of the n-dimensional space, interconnected with (n+1)-dimensional hyperplane segments. The direct implication of this morphology is that the "extrema" of the surface $\hat{\phi}_n(\tau_1, \ldots, \tau_n)$ must be sought among its apexes (i.e., among the nodal points). Thus, the stair-like form of the CSRS assures that the n-th order autocorrelation estimate is an (n+1)-dimensional surface of the described morphology.

As a result of this morphology, the statistical behavior of the

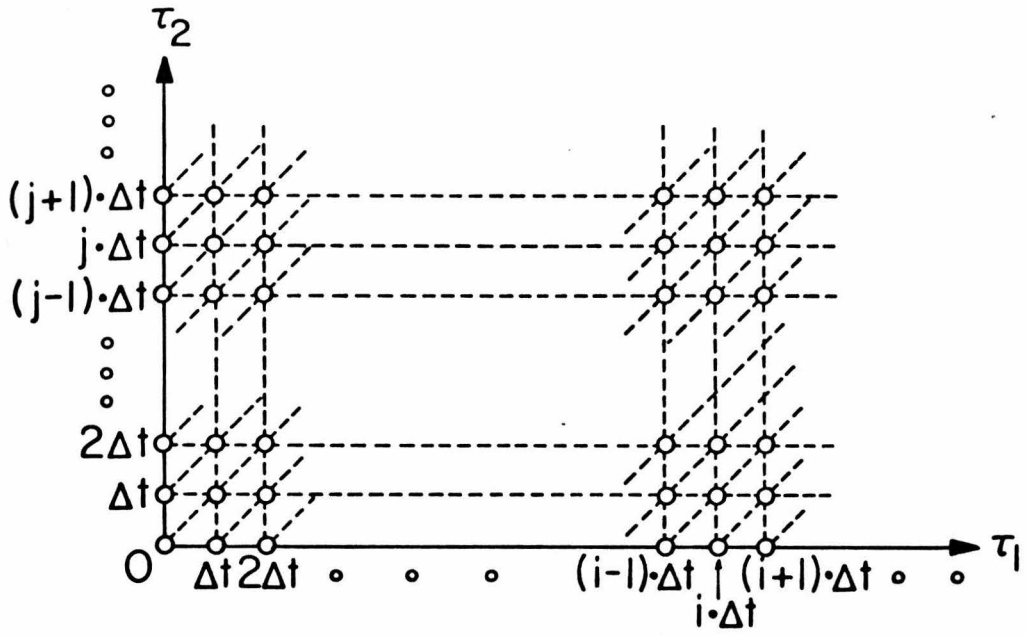


Fig. 4.2.2: Locus of derivative discontinuity points of $\hat{\phi}_3(\tau_1, \tau_2, 0)$.

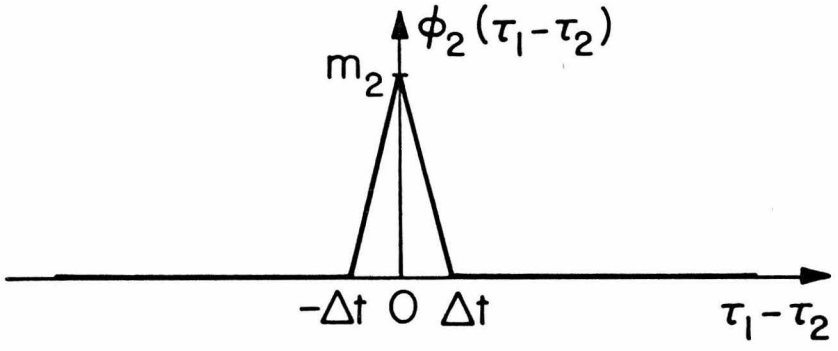


Fig. 4.2.3: Second order autocorrelation function of a CSRS.

intermediate points is determined by the statistical behavior of the neighboring nodal points through a multi-dimensional linear relationship. Consequently, the odd order autocorrelation functions of a CSRS are uniformly zero and the even order ones are zero everywhere except around the full-diagonal points, where they exhibit pyramoidal nonzero values.

To illustrate this, we consider the second order autocorrelation function of a CSRS: (Fig. 4.2.3)

$$\phi_2(\tau_1, \tau_2) = m_2\left(1 - \frac{|\tau_1 - \tau_2|}{\Delta t}\right) \qquad \text{for } |\tau_1 - \tau_2| \le \Delta t$$
$$= 0 \qquad \text{for } |\tau_1 - \tau_2| > \Delta t \quad (4.2.18)$$

We also show in Fig. 4.2.4 the form of the fourth order autocorrelation function $\phi_4(\tau_1, \tau_2, \sigma_1, \sigma_2)$ of a CSRS, for given values σ_1 and σ_2. Notice the anticipated pyramoidal shape of the autocorrelation function surface. With this brief analysis, we have completed the study of the autocorrelation properties of the CSRS family and its quasi-whiteness has been manifested.

4.3 The Use of CSRS in Nonlinear System Identification

In the previous section, we established the quasi-whiteness of the CSRS family, which justifies their use of nonlinear system identification through the crosscorrelation technique. The kernel estimates, that are obtained through the use of CSRS, correspond to a functional series which is different in structure from the Wiener series. This structural difference is due to the statistical properties of the CSRS, as expressed by the moments of the amplitude probability distribution,

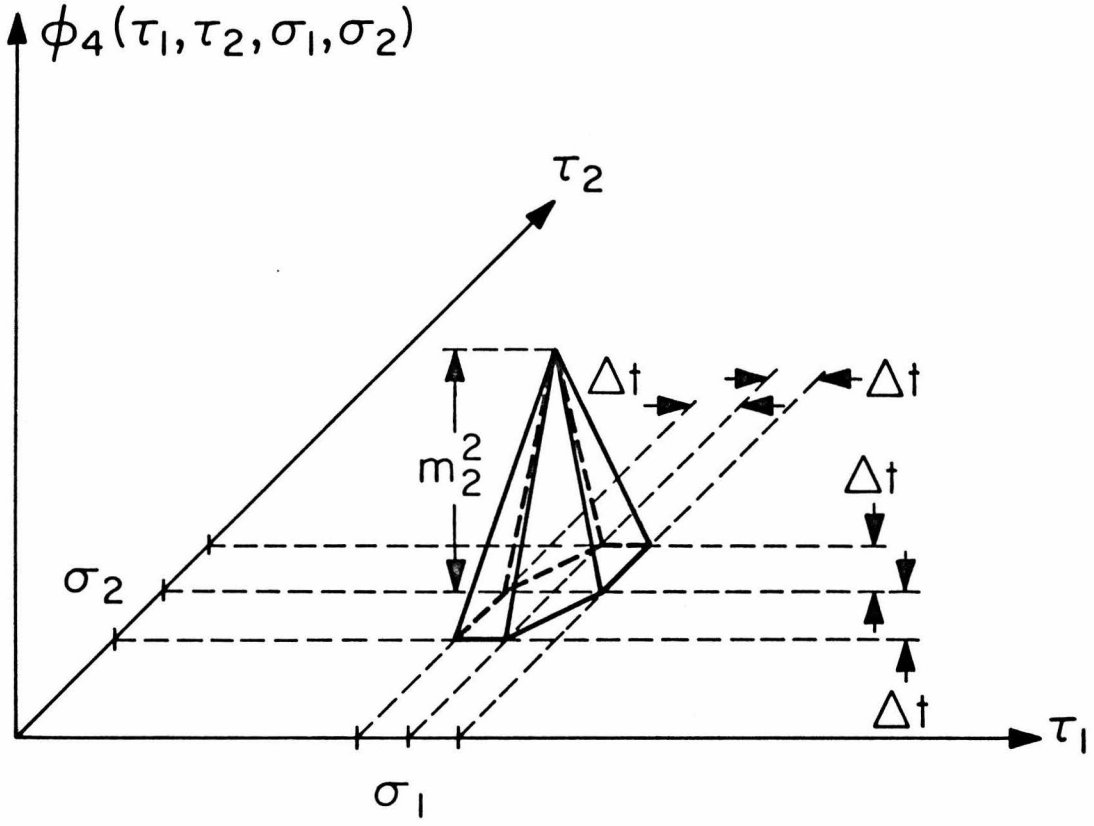


Fig. 4.2.4: Pyramidal form of the principal lobe of an autocorrelation function

which are different from the ones of GWN.

More specifically, the gaussian amplitude distribution possesses the important property of being describable by only its first two moments. This, along with the decomposition property (cf. sec. 3.2.2), results in great simplification of the expressions describing the orthogonal Wiener functionals. In the general case of a CSRS, however, the complete description of the amplitude probability distribution requires all of its moments. This results in a certain complexity of the form of the CSRS orthogonal functionals. Nevertheless, the construction of the CSRS functionals is made routinely on the basis of an orthogonalization procedure similar to the one that was used in the construction of the Wiener series. It must be also noted that the orthogonality of the CSRS functionals is only approximate and based on the assumption that the bandwidth of the respective CSRS is broad enough with respect to the system bandwidth so that the deconvolution takes place with acceptable accuracy.

Evidently, the structural form of the CSRS functional series depends on the even order moments and the step length of the associated CSRS. In the special case where a gaussian amplitude distribution is chosen for the CSRS, the CSRS functional series takes exactly the form of the Wiener series, where the power level of the GWN is equal to the product of the second moment and the step length of the CSRS. Thus, it becomes clear that the CSRS functional series is a more general orthogonal functional expansion than the Wiener series, extending the basic idea of orthogonal expansion using Volterra-type functionals throughout the space of symmetric probability distributions, in the possible

expense of more complexity in the functional expressions. The advantages that such a generalization of the orthogonal functional series provides are the same as the ones of any optimization problem where the parameter space is augmented. The augmentation of the parameter space provides greater flexibility and allows the search (and usually the achievement) of new global maxima of the "gain function". This is analytically demonstrated through an example later in this section.

It must be emphasized that, even though each one of these functional series represents a formally different description of the system, all of them are equivalent in the sense that they are functional expansions of the same response signal. What is generally different in all these functional series is the convergence pattern. As a result of this, the estimated truncated models, that we usually obtain in practice, give different accuracy in the model predicted response.

The subject of the functional series convergence is apparently of great theoretical and practical interest, although it is also a subject of great analytical complexity. Later in this section, we study analytically the convergence of a first order model for a third order nonlinear system. This example will demonstrate the basic theoretical and practical aspects of interest, even though it is limited by the specific choice of system.

In chapter 9, we will also give an illustration of the series convergence, using several CSRS in computer simulated applications.

The orthogonal functions that correspond to a CSRS have the form:

$$G_o^*[\, g_o;\, x(t'), t' \le t \,] = g_o \qquad (4.3.1)$$

$$G_1^*[g_1(\tau_1);\ x(t'),\ t' \le t] \int_0^\infty g_1(\tau_1)\, x(t-\tau_1)\, d\tau_1 \qquad (4.3.2)$$

$$G_2^*[g_2(\tau_1,\tau_2);\ x(t'), t' \le t] \int_0^\infty \int_0^\infty g_2(\tau_1,\tau_2)\, x(t-\tau_1)\, x(t-\tau_2)\, d\tau_1 d\tau_2 - (m_2 \Delta t) \int_0^\infty g_2(\tau_1,\tau_1)\, d\tau_1 \qquad (4.3.3)$$

$$G_3^*[g_3(\tau_1,\tau_2,\tau_3);\ x(t'),\ t' \le t] \int_0^\infty \int_0^\infty \int_0^\infty g_3(\tau_1,\tau_2,\tau_3)\, x(t-\tau_1)\, x(t-\tau_2) x(t-\tau_3)\, d\tau_1 d\tau_2 d\tau_3 - 3(m_2 \Delta t) \int_0^\infty \int_0^\infty g_3(\tau_1,\tau_2,\tau_2)\, x(t-\tau_1)\, d\tau_1 d\tau_2 - [(m_4/m_2 - 3m_2)\Delta t] \int_0^\infty g_3(\tau_1,\tau_1,\tau_1)\, x(t-\tau_1)\, d\tau_1 \qquad (4.3.4)$$

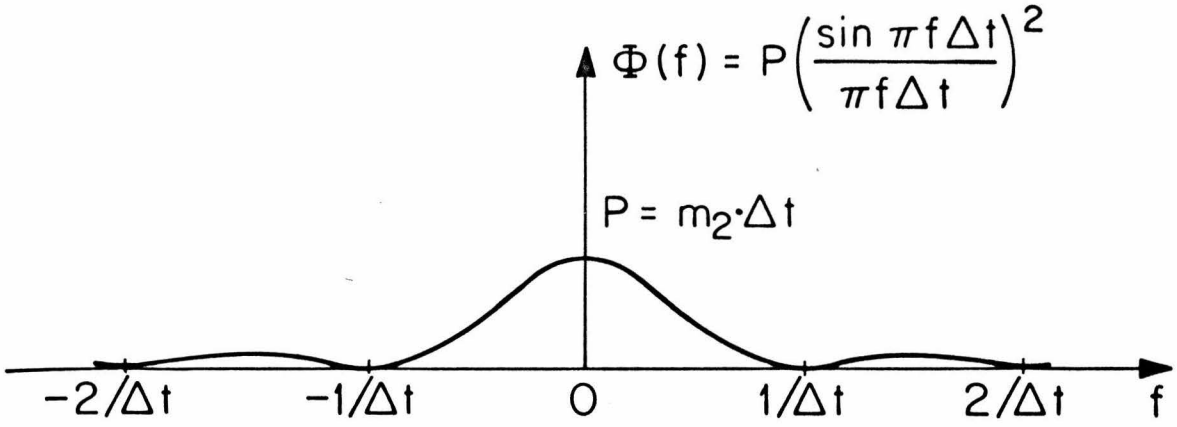
etc.

where, $x(t)$ is a CSRS, Δt is its step length and m_2, m_4, etc. are the second, the fourth, etc. moments of its amplitude probability density function $p(x)$.

The expressions become quite involved as we go to higher order functionals, but their derivation can be done routinely on the basis of a Gram-Schmidt type of orthogonalization procedure.

Of course, the orthogonality of the G*-functionals is only approximate, within the range of frequencies posed by the specified step length Δt.

The power spectrum of a CSRS with step length Δt and second moment m_2 is shown in Fig. 4.3.1. Clearly, the bandwidth of the signal is inversely proportional to Δt, and it approaches the ideal white noise as Δt approaches zero (provided that the power level $P = m_2 \cdot \Delta t$ remains finite). Apparently, the degree of orthogonality of the G*-functionals monotonically increases as Δt is decreased. However, for all practical purposes, it usually suffices that the bandwidth of the CSRS covers the bandwidth of the system under study.



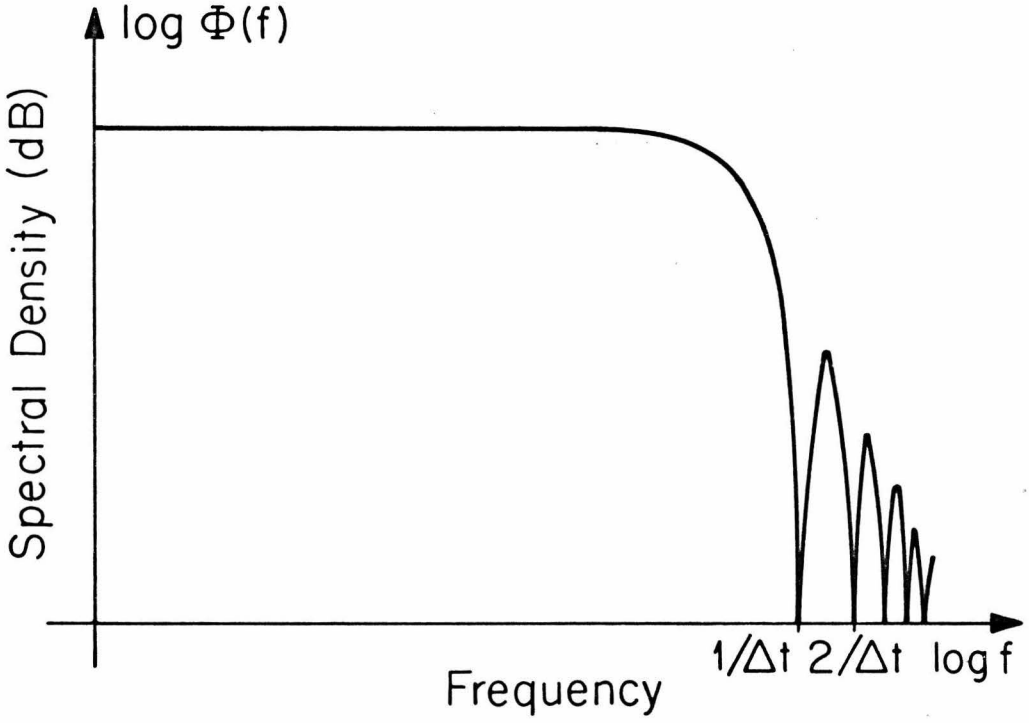


Fig. 4.3.1: The power spectrum of a CSRS

Notice that the basic structural form of the G*-functionals is the same as in the Wiener G-functionals, i.e. the odd order functionals have all and only odd order integral terms of equal and lower order; and analogously the even order functionals have all and only even order integral terms of equal and lower order. Notice also that if a gaussian probability density function is chosen for a CSRS, then $m_4 = 3m_2^2$ and, consequently, G_3^* becomes exactly like the G_3 Wiener functional, with power level $P = m_2 \cdot \Delta t$.

The actual procedure of the crosscorrelation technique is basically the same as in the GWN case. That is, in order to estimate the n-th order kernel, we crosscorrelate the n-th order response residual with n time-shifted versions of the stimulus and we properly normalize the outcome: (cf. sec. 2.3)

$$\hat{h}_n(\sigma_1, \ldots, \sigma_n) = C_n \cdot E[\, y^{(n)}(t)\, x(t-\sigma_1) \ldots x(t-\sigma_n)] \qquad (4.3.5)$$

where,

$$y^{(n)}(t) = y(t) - \sum_{i=0}^{n-1} G_i^* [\, g_i(\tau_1, \ldots, \tau_i);\, x(t'), t' \leq t] \qquad (4.3.6)$$

and C_n is the proper normalizing factor.

In the case of GWN, the normalizing factor C_n is $1/n!P^n$, where P is the power level of the GWN. In the case of CSRS, the normalizing factor differ for the diagonal and nondiagonal points. For the nondiagonal points, the normalizing factor is just like in the GWN case:

$$C_n = \frac{1}{n!P^n} \qquad (4.3.7)$$

where, $P = m_2 \cdot \Delta t$.

However, for the diagonal points the normalizing factor depends on the order of the specific point. For this reason, it is impossible to give a

concise and general expression for the normalizing factors, which could cover all the diagonal points of all kernels.

This is not a problem in practice, because we always confine ourselves in estimating the first few kernels (usually up to the second order) and thus we only need the normalizing factors for these limited cases. For example, in the second order case the normalizing factor can be easily evaluated on the basis of our knowledge about the form and structure of the autocorrelation functions of the CSRS. (cf. sec. 4.2) Thus, the second order crosscorrelation gives:

$$\begin{aligned}\psi_2(\sigma_1,\sigma_2) &= E[\,y(t)\,x(t-\sigma_1)\,x(t-\sigma_2)] \qquad (4.3.8)\\ &= E[\,G_2^{*}(t)\,x(t-\sigma_1)\,x(t-\sigma_2)]\\ &= \int_0^{\infty}\int g_2(\tau_1,\tau_2)\,E[\,x(t-\tau_1)\,x(t-\tau_2)\,x(t-\sigma_1)\,x(t-\sigma_2)]\,d\tau_1 d\tau_2\\ &\qquad -(m_2\Delta t)\int_0^{\infty} g_2(\tau,\tau)\,d\tau\cdot E[\,x(t-\sigma_1)x(t-\sigma_2)]\end{aligned}$$

$$= \begin{cases} 2(m_2\Delta t)^2\cdot g_2(\sigma_1,\sigma_2) & \text{for the nondiagonal points}\\ (m_4-m_2^{\,2})\,\Delta t^2\cdot g_2(\sigma,\sigma) & \text{for the diagonal points}\end{cases}$$

Therefore the normalizing factors for the second order kernel are:

$$C_2 = \begin{cases} 1/2(m_2\Delta t)^2 & \text{for the nondiagonal points} \qquad (4.3.9)\\ 1/(m_4-m_2^{\,2})\cdot\Delta t^2 & \text{for the diagonal points} \qquad (4.3.10)\end{cases}$$

The determination of the appropriate normalizing factors for the diagonal points of other kernels will require basically the evaluation of the volume under the corresponding autocorrelation function surface.

We will study now the relation between the Volterra and the CSRS kernels of a system. We will attempt an analytical approach in such a way that the role of the basic parameters affecting the estimated model will be depicted in the derived expressions. Our objective is to

demonstrate the dependence of the CSRS kernels upon the moments and the step length of the specific CSRS that is used to estimate the kernels.

The expressions that describe analytically the relation between the CSRS and the Volterra kernels of the system are very complicated in their generality. Since, our objective is to give simply a qualitative idea of the existing relation, we will confine ourselves to the nondiagonal points of the CSRS kernels and to general nonexplicit analytical expressions.

Thus, consider the Volterra expansion of the system response:

$$y(t)=\sum_{n=0}^{\infty}\int_{0}^{\infty}\cdots\int_{0}^{\infty}k_n(\tau_1,\ldots,\tau_n)x(t-\tau_1)\ldots x(t-\tau_n)d\tau_1\ldots d\tau_n \qquad (4.3.11)$$

Let x(t) be a CSRS with a probability density p(x) and step length Δt. Let us evaluate now the CSRS kernels through the crosscorrelation technique. The resulting expressions for the CSRS kernels in terms of the Volterra kernels are:

$$g_{2n}(\sigma_1,\ldots,\sigma_{2n})=\sum_{\ell=n}^{\infty}\sum_{i}\Delta t^{(\ell-n)}C_i(m_2,\ldots,m_{2\ell-2n})\int_{0}^{\infty}\cdots\int_{0}^{\infty}k_{2\ell}\left(\underline{\sigma_i}^{(2n)},\underline{\tau_i}^{(2\ell-2n)}\right)d\underline{\tau_i}^{(2\ell-2n)} \qquad (4.3.12)$$

$$g_{2n+1}(\sigma_1,\ldots,\sigma_{2n+1})=\sum_{\ell=n}^{\infty}\sum_{i}\Delta t^{(\ell-n)}C_i(m_2,\ldots,m_{2\ell-2n})\int_{0}^{\infty}\cdots\int_{0}^{\infty}k_{2\ell+1}\left(\underline{\sigma_i}^{(2n+1)},\underline{\tau_i}^{(2\ell-2n)}\right)d\underline{\tau_i}^{(2\ell-2n)} \qquad (4.3.13)$$

where, the index i is summing over all possible combinations of even population number groupings of $(2\ell-2n)$ arguments. To each one of these combinations corresponds a proper coefficient C_i, which depends on the even order moments from the second up to the $(2\ell-2n)$-th one. The arguments of the Volterra kernels in the expressions also attain proper values for each one of these combinations. The vectors $\sigma_i^{(2n)}$ and

$\underline{\tau_i}^{(2\ell-2n)}$ symbolize these argument arrangements, which are similar to the ones appearing in the relations between the Wiener and Volterra kernels (cf. sec. 2.5).

Clearly, the even order CSRS kernels depend on all the even order Volterra kernels of equal or higher order (and analogously the odd order CSRS kernels), demonstrating a notable similarity to the structure of the Wiener kernels.

Another interesting remark is that the expressions (4.3.12) and (4.3.13) can be written as power series in Δt:

$$g_{2n}(\sigma_1,\ldots,\sigma_{2n}) = \sum_{\ell=n}^{\infty} \Delta t^{\ell-n} \cdot f_\ell^{(2n)}(\sigma_1,\ldots,\sigma_{2n}) \qquad (4.3.14)$$

$$g_{2n+1}(\sigma_1,\ldots,\sigma_{2n+1}) = \sum_{\ell=n}^{\infty} \Delta t^{\ell-n} \cdot f_\ell^{(2n+1)}(\sigma_1,\ldots,\sigma_{2n+1}) \qquad (4.3.15)$$

where,

$$f_\ell^{(2n)}(\sigma_1,\ldots,\sigma_{2n}) = \sum_i C_i(m_2,\ldots,m_{2\ell-2n}) \int_0^\infty \cdots \int_0^\infty k_{2\ell}(\underline{\sigma_i}^{(2n)}, \underline{\tau_i}^{(2\ell-2n)}) \, d\underline{\tau_i}^{(2\ell-2n)} \qquad (4.3.16)$$

$$f_\ell^{(2n+1)}(\sigma_1,\ldots,\sigma_{2n+1}) = \sum_i C_i(m_2,\ldots,m_{2\ell-2n}) \int_0^\infty \cdots \int_0^\infty k_{2\ell+1}(\underline{\sigma_i}^{(2n+1)}, \underline{\tau_i}^{(2\ell-2n)}) \, d\underline{\tau_i}^{(2\ell-2n)} \qquad (4.3.17)$$

These expressions of the CSRS kernels as power series in Δt can be useful in studying the effect of changing Δt upon the estimated CSRS kernels and the corresponding functional series. For example, we can observe that for very small values of Δt the CSRS kernel of some order is mainly affected by the Volterra kernel of the same order, while the contribution of the higher order Volterra kernels is negligible. This may be utilized in order to estimate approximately Volterra kernels of the system. Also, information about the order of nonlinearity of the system can be derived by proper utilization of eqns. 4.3.14 and 4.3.15 .

As an illustrative example on the parametric dependences of the CSRS functional series, consider a third order Volterra system (i.e. a system for which $k_n(\tau_1, \ldots, \tau_n) \equiv 0$ for any $n > 3$). The system response to a CSRS x(t) is: (also consider $k_o = 0$).

$$y(t) = \int_0^\infty k_1(\tau_1)x(t-\tau_1)d\tau_1 + \iint_0^\infty k_2(\tau_1, \tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2 + \iiint_0^\infty k_3(\tau_1, \tau_2, \tau_3)x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)d\tau_1 d\tau_2 d\tau_3 \quad (4.3.18)$$

and the CSRS kernels of the system are:

$$g_o = (m_2 \Delta t)\int_0^\infty k_2(\tau_1, \tau_1)d\tau_1 \quad (4.3.19)$$

$$g_1(\sigma_1) = k_1(\sigma_1) + 3(m_2 \Delta t)\int_0^\infty k_3(\sigma_1, \tau_1, \tau_1)d\tau_1 + \left[\left(\frac{m_4}{m_2} - 3m_2\right)\Delta t\right] \cdot k_3(\sigma_1, \sigma_1, \sigma_1) \quad (4.3.20)$$

$$g_2(\sigma_1, \sigma_2) = k_2(\sigma_1, \sigma_2) \quad (4.3.21)$$

$$g_3(\sigma_1, \sigma_2, \sigma_3) = k_3(\sigma_1, \sigma_2, \sigma_3) \quad (4.3.22)$$

Evidently, the second and third order CSRS kernels are identical to the Volterra ones because of the absence of nonlinearities of order higher than third. The zero order CSRS kernel depends upon the second order Volterra kernel, and the first order CSRS kernel depends upon the first and third order Volterra kernels in accordance with eqns. 4.3.14 and 4.3.15 .

The difference between the Volterra and the CSRS kernels of zero and first order indicates the fact that a CSRS functional of some order probes into the higher order Volterra functionals, thus providing a different span of the function space of the system response. This different span is expectedly more efficient for expanding the system response to a CSRS stimulus, because of the orthogonality of the CSRS functionals.

Recalling that the Volterra kernels constitute a nonparametric description of the system (i.e. they depend only upon the functional derivatives of the system operator and not upon any stimulus characteristics, like the power level etc.) the expressions above state very eloquently the dependence of the CSRS kernels upon the even moments and the step length of the associated CSRS.

If we generalize the concept of the power level so that we call the expression:

$$P_n = m_{2n} \cdot \Delta t^n \tag{4.3.23}$$

the n-th order power level of a CSRS; then we can generally state that the CSRS kernels depend upon the power levels of all orders up to the order of nonlinearity of the system minus one.

What is of importance in the system modeling business, is the accuracy (in the mean square sense) of the model predicted response to a given stimulus. We will study now the improvement of this accuracy when we use a CSRS model, as opposed to a Volterra model of the same order. To study this, we will simply consider the zero order model of the system above. The m.s. error of the zero order Volterra model response is:

$$e_o^2 = E\{y^2(t)\} \tag{4.3.24}$$

The m.s. error of the zero order CSRS model response is:

$$\begin{aligned} \epsilon_o^2 &= E\left\{\left[y(t)-g_o\right]^2\right\} \\ &= E\left\{y^2(t)\right\}+g_o^2-2g_o \cdot E\left\{y(t)\right\} \end{aligned} \tag{4.3.25}$$

Therefore, the improvement in accuracy of the zero order model predicted response, using a CSRS model instead of a Volterra model, is:

$$i_o = e_o^2 - \epsilon_o^2$$

$$= 2 g_o \cdot E\{y(t)\} - g_o^2 \qquad (4.3.26)$$

If x(t) is the CSRS stimulus with which the model was estimated then:

$$i_o = g_o^2$$

$$= \{P_1 \int_o^\infty k_2(\tau_1, \tau_1) d\tau_1\}^2 \geq 0 \qquad (4.3.27)$$

Therefore, we always have a positive (or zero) improvement in accuracy with the zero order CSRS model in predicting the system response to the CSRS stimulus. This improvement clearly depends on the first order power level of the CSRS stimulus.

Let us study now this measure of improvement (or deterioration) in the accuracy of the model predicted response to several other stimuli, when we are using the previous CSRS zero order model instead of the Volterra model of the same order.

If x(t) is another CSRS with first order power level P_1^*, then:

$$i_o = P_1 [2P_1^* - P_1] \cdot \{\int_o^\infty k_2(\tau_1, \tau_1) d\tau_1\}^2 \qquad (4.3.28)$$

Thus, we can have improvement or deterioration in the accuracy of the model response depending on the relative size of the power levels. A similar expression holds in the case of GWN, PRS or any other quasi-white signal.

If x(t) is an arbitrary signal, then:

$$i_o = P_1 \int_o^\infty k_2(\tau_1 \tau_1) d\tau_1 \cdot \left\{ 2 \int_o^\infty k_1(\tau_1) \phi_1(\tau_1) d\tau_1 + 2 \int\int\int_o^\infty k_3(\tau_1, \tau_2, \tau_3) \phi_3(\tau_1, \tau_2, \tau_3) \right.$$

$$\left. d\tau_1 d\tau_2 d\tau_3 + \iint_o^\infty k_2(\tau_1, \tau_2) \left[2\phi_2(\tau_1, \tau_2) - P_1 \cdot \delta(\tau_1 - \tau_2)\right] d\tau_1 d\tau_2 \right\} \qquad (4.3.29)$$

where,

$$\phi_i(\tau_1, \ldots, \tau_i) = \frac{1}{T}\int_0^T x(t-\tau_1)\ldots x(t-\tau_i)dt \quad ; i=1, 2, 3 \qquad (4.3.30)$$

Eqn. 4.3.29 clearly demonstrates the fact that the improvement (or deterioration) in this case depends upon the autocorrelation functions ϕ_i of the stimulus. This fact leads us to a basic conclusion with respect to the system models that are obtained by using quasi-white signals: if we estimate the kernels of all existing orders in the system, then the obtained model is complete and capable to predict exactly (with some small numerical inaccuracies) the system response to any stimulus; however, if the obtained model is of lower order than the system (which is often the case in practice), then the performance of the model (in the m. s. error sense) depends crucially on the relation of the autocorrelation functions of the stimulus with the ones of the associated quasi-white signal.

This basic conclusion is illustrated by computer simulated examples in sec. 8.3.

Another important remark is that the improvement in accuracy of the zero order model in the considered example becomes maximum if and only if:

$$E\left[y(t)\right] = g_o \qquad (4.3.31)$$

as it is evident from the derivatives:

$$\frac{\partial i_o}{\partial P_1} = 2\int_0^\infty k_2(\tau_1, \tau_1)d\tau_1 \cdot \left\{E[y(t)] - g_o\right\} \qquad (4.3.32)$$

$$\frac{\partial^2 i_o}{\partial P_1^2} = -2\left\{\int_0^\infty k_2(\tau_1, \tau_1)d\tau_1\right\}^2 \qquad (4.3.33)$$

Thus, the improvement is maximized in the case of the CSRS stimulus

with which the model was estimated (as it was expected), but this may also happen with other stimuli for which eqn. 4.3.31 holds.

Evidently, for a given arbitrary stimulus there is always, in this case, a CSRS which gives the optimum zero order model. This CSRS is specified by its first order power level:

$$P_1 = E\left[y(t)\right] \Big/ \int_0^\infty k_2(\tau_1, \tau_1)d\tau_1 \qquad (4.3.34)$$

Of course, as the order of the system nonlinearity and of the model increases, the analytical expressions describing the dependence of the model accuracy upon the power levels of the CSRS become rapidly very complicated. However, the basic remark can be generalized to state that: for a given system and a given stimulus there are CSRS, specified by some values of generalized power levels, which give the model of a certain order with the highest accuracy in predicting the system response to the given stimulus.

It must be emphasized that these optima CSRS are specified strictly and solely by their generalized power levels; as long as the stimulus frequency bandwidth remains broad enough so that the CSRS functional series can be considered approximately orthogonal.

In order to illustrate a case where more power levels than the first order one participate in determining the accuracy of the CSRS model, consider the first order CSRS model of the previous system, where for simplification of the derivations we take:

$$k_1(\vartheta_1) = h(\vartheta_1) \qquad (4.3.35)$$

$$k_2(\vartheta_1, \vartheta_2) = h(\vartheta_1)h(\vartheta_2) \qquad (4.3.36)$$

$$k_3(\vartheta_1, \vartheta_2, \vartheta_3) = h(\vartheta_1)h(\vartheta_2)h(\vartheta_3) \qquad (4.3.37)$$

The m. s. error of the first order Volterra model is:

$$e_1^2 = E\left\{\left[y(t) - \int_o^\infty h(\partial_1)x(t-\partial_1)d\partial_1\right]^2\right\} \tag{4.3.38}$$

The m. s. error of the first order CSRS model is:

$$\epsilon_1^2 = E\left\{\left[y(t) - g_o - \int_o^\infty g_1(\partial_1)x(t-\partial_1)d\partial_1\right]^2\right\} \tag{4.3.39}$$

Therefore, the improvement in accuracy of the first order CSRS model is: (for the CSRS stimulus)

$$\begin{aligned} i_1 &= e_1^2 - \epsilon_1^2 \\ &= -\{F_1(P_1) + F_2(P_1) \cdot F(P_1, P_2) + F_3(P_1) \cdot F^2(P_1, P_2)\} \end{aligned} \tag{4.3.40}$$

where,

$$F_1(P_1) = 9a^3P_1^3 + (5a^2 - 18a^3)P_1^2 - 6a^2P_1 \tag{4.3.41}$$

$$F_2(P_1) = [6aP_1^2 + 2(1-6a)P_1 - 2]b \tag{4.3.42}$$

$$F_3(P_1) = (P_1 - 2)c \tag{4.3.43}$$

$$F(P_1, P_2) = \frac{P_2}{P_1} - 3P_1 \tag{4.3.44}$$

$$a = \int_o^\infty h^2(\tau)d\tau \tag{4.3.45}$$

$$b = \int_o^\infty h^4(\tau)d\tau \tag{4.3.46}$$

$$c = \int_o^\infty h^6(\tau)d\tau \tag{4.3.47}$$

Clearly, the accuracy of the first order CSRS model depends upon the first and second order power levels in this case; and the dependence is not monotonic like the accuracy of the zero order model (cf. eqn. 4.3.27).

Trying to determine the extrema of the surface $i_1(P_1, P_2)$, we can take (for derivation convenience) without loss of generality a = 1. Notice also that the positions of the extrema are invariant to scaling changes of the signal amplitude, which allows us to make the substitution:

$$P_2 = (\lambda+3)P_1^2 \tag{4.3.48}$$

where changing λ corresponds to changing probability distribution. With these simplifications, we get:

$$i_1 = -(c\lambda^2 + 6b\lambda + 9)P_1^3 + (2c\lambda^2 + 10b\lambda + 13)P_1^2 + (2b\lambda + 6)P_1 \qquad (4.3.49)$$

where, $o < c < b < 1; \lambda > -2; b^2 < c$. All these conditions are derived by using the Schwartz inequality. Notice that:

$$(c\lambda^2 + 6b\lambda + 9) > o \qquad (4.3.50)$$

$$(2c\lambda^2 + 10b\lambda + 13) > o \qquad (4.3.51)$$

$$(2b\lambda + 6) > o \qquad (4.3.52)$$

for all values of b, c and λ.

Consequently, for a given system (b, c) and a given probability distribution (λ), the improvement in accuracy of the first order CSRS model is maximized for a single value of P_1, which is easily determined as the single positive root of the binomial:

$$3(c\lambda^2 + 6b\lambda + 9)P_1^2 - 2(2c\lambda^2 + 10b\lambda + 13)P_1 - (2b\lambda + 6) = o \qquad (4.3.53)$$

The function $i_1(P_1)$ is zero at $P_1 = o$ and has positive derivative (since $2b\lambda + 6 > o$). As we increase P_1 the function $i_1(P_1)$ rises monotonically to its maximum and subsequently it steadily declines to minus infinity. The fact that i_1 becomes increasingly negative for very large values of P_1 must not disturb us, since the second order functional term is probably compensating for that. Therefore, the efficacy of the first order (or any other order) model must be viewed within a certain limited range of values of the associated power levels, and not throughout the whole range from zero to infinity.

Clearly, the position and the magnitude of the maximum

improvement in accuracy of the first order CSRS model depend on the first and second order power levels independently. The actual determination of this maximum requires sizable calculations comprising the solution of the two variable algebraic system:

$$3(c\lambda^2+6b\lambda+9)P_1^2-2(2c\lambda^2+10b\lambda+13)P_1-(2b\lambda+6) = o$$

$$\lambda = -\frac{b(3P_1^2-5P_1-1)}{cP_1(P_1-2)} \qquad (4.3.54)$$

Evidently, the function $i_1(\lambda, P_1)$ exhibits several extrema; but we seek its global maximum. The exact position and magnitude of the global maximum in this case is not of interest in our study and, therefore, it is omitted. What is of great interest is the realization of the continuous way in which the variable parameters λ and P_1 determine this maximum. This implies that the acquisition of the optimum model in each case, strictly determines the power levels of the CSRS that ought to be used for the estimation of this model. This is our fundamental conclusion from all this discussion. Let us generalize and restate this fundamental conclusion: if we are given a system and the order of the CSRS model that is to be used in order to predict approximately the system response, then there is a finite number of determinable sets of generalized power levels, which correspond to the CSRS stimuli that ought to be used to estimate the model with the highest possible accuracy. The space of signals within which these specific CSRS possess this optimal property is the space of all the quasi-white signals.

The band-limited gaussian white noise (GWN) and the pseudo-random signals (PRS) are included in this space because their corresponding models are found as special cases within the CSRS models.

For example, in the case of GWN the corresponding model is similar to the CSRS model with $\lambda = 0$, which simply manifests the statistical property of the gaussian distribution that all of its even moments are expressible in terms of the second order one. Thus, in the case of GWN the parameter space for the search of the maximum accuracy model reduces to one-dimensional. Consequently, the use of CSRS provides greater flexibility and the definite potentiality (if it is to be used) of achieving a model with higher accuracy in predicting the system response. This is certainly one of the principal virtues of the CSRS family in connection with their use in nonlinear system identification.

4.4 Discussion on Relative Advantages and Disadvantages of GWN, PRS and CSRS

The quasi-white signals that have been used so far in nonlinear identification through the crosscorrelation technique are: the GWN, the PRS and the newly introduced CSRS. In the process of deciding which one of these signals is preferable in a specific application, one is faced with a complicated problem, because each one of these signals exhibits its own characteristic set of advantages and disadvantages.

The appreciation of the relative advantages and disadvantages is a complex task and often highly subjective. Of course, there are some objective criteria that can be used in this relative appreciation; nevertheless, the special characteristics of a specific application can introduce a variety of influential factors in the decision making.

For example, if a system is predominantly linear and the com-

putational burden is a principal concern, then the PRS are apparently most advantageous. However, if there are significant higher order nonlinearities then the PRS may do a fairly poor job, giving significantly erroneous kernel estimates (because of the anomalies in their higher even order autocorrelation functions). To mention a more singular but still possible situation, suppose that in order to generate the PRS we need a linear recurrence formula of such a high order that it is not given by the available tables, and we happen not to have the capacity to determine it by ourselves, then the generation of the PRS test signal becomes problematic. As another example, suppose the specific system under test has significant higher order nonlinearities and the computational burden is not a principal concern, then the CSRS are apparently most advantageous.

Of course, the determinants of the decision making are usually not so simply and clearly outlined; and the whole decision making process is a more complicated function, in which a variety of factors participate, being appreciated by the human subject in a way that is not always free of errors or even psychological bias.

To mention one of the most important factors, which can be possibly ignored (because of accidental ignorance or because of the present difficulty in determining the effectual pattern of this factor), we can refer to the different convergence patterns of the several functional series corresponding to each one of these signals. Recalling also the considerable number of test parameters (stimulus bandwidth, record length etc.) that affect the actual result of the identification procedure,

we get a good idea of the complexity of the problem.

In any case, since we are bound to make a choice, we might as well try to do the best out of what the practical conditions permit. To this end, we will discuss in the following the most basic considerations that the investigator ought to have in mind in attempting to choose the test signal, that fits best (according to a limited number of basic criteria) his specific application.

In review, the principal advantages and disadvantages of each one of these families of quasi-white signals are:

(I) For the GWN

The main advantages of the GWN derive from its gaussian nature and the fact that it has traditionally established a solid reputation among the users of the Wiener approach and the crosscorrelation technique, so that more relevant literature can be found if wanted. Its gaussian nature secures the simplest and most elegant expressions for the orthogonal functional series (in this case the original Wiener series) and related matters (like relations between estimated kernels and Volterra kernels, the normalizing factors of the crosscorrelation estimates, etc.), simply because of the decomposition property of the gaussian random variables (which allows all the higher even moments to be expressed in terms of the second one). Additionally, the GWN is a signal with rich information content and it expectedly provides a good model for a great variety of inputs.

The main disadvantages of the GWN are: First, the actual generation of GWN in the laboratory is a relatively complex task with

respect to the straightforward way of generating PRS or CSRS. Second, there are some imperfections in the autocorrelation functions due to the side lobes (cf. sec. 3.2.2) and the truncation of the gaussian distribution. Third, the computational burden can become heavy in order to reduce the statistical fluctuation error down to an acceptable degree. Fourth, the error analysis is complicated by the inconvenient analytical form that the autocorrelation functions attain and the involved actual method of GWN generation.

(II) For the PRS:

The main advantages of the PRS are: First, they are generated in a relatively simple way in the laboratory. Second, they require relatively short records in order to form the desirable autocorrelation functions and consequently, they reduce significantly the computational burden. This reduction becomes even more dramatic when a binary or a ternary PRS is used,and we are exploiting the special form of these signals using proper pieces of software and/or hardware.

The main disadvantages of the PRS are: First, they exhibit anomalies in the higher (>2) even order autocorrelation functions, which may induce considerable estimation errors if the system contains significant nonlinearities. Second, the analytical expressions concerning the corresponding functional series and related matters (i.e. relation of PRS kernels with Volterra kernels, normalizing factors of the crosscorrelation estimates etc.) are fairly complicated. Third, the error analysis is quite difficult, because of the involved and laborious method with which the anomalies can be determined in position and

magnitude.

(III) For the CSRS:

The main advantages of the CSRS are: First, they are generated in a simple way, easily implemented in the laboratory. Second, their autocorrelation functions do not exhibit any deterministic kind of imperfections, within the frequency limits posed by the signal bandwidth (which must cover the system bandwidth). Third, the error analysis is greatly facilitated by the convenient form that the autocorrelation functions attain and the simple way of CSRS definition and generation. Fourth, they provide the user with a wide variety and great flexibility in choosing the signal with the number of levels and probability distribution that fits best the specific case at hand and provides the model with the highest accuracy (if pursued).

The main disadvantages of the CSRS are: First, they require fairly long records in order to reduce the statistical fluctuation error down to acceptable limits, which result in heavy computational burden (as in the case of GWN). However, the computational burden can be considerably reduced by using binary or ternary CSRS and exploiting the special form of these signals employing proper pieces of hardware and/or software. Second, the analytical expressions concerning the corresponding functional series and related matters (i.e. relation of CSRS kernels with Volterra kernels, normalizing factors of the cross-correlation estimates etc.) are fairly complicated. Nevertheless, they are a little simpler than in the case of PRS, since it is easier to evaluate the even order moments of the CSRS rather than the ones of the

PRS.

Besides the mentioned advantages and disadvantages of GWN, PRS and CSRS, there may be other factors which become important in a specific situation because of special experimental or computational (or other) considerations. Therefore, the choice of the proper test quasi-white signal in a specific case is still considerably a case-to-case matter relying on subjective appreciation and relative weighing of a variety of effectual factors.

In any case, we consider it essential for the prospective user of the crosscorrelation method to know and consider seriously the mentioned principal advantages and disadvantages of the presently available quasi-white signals, as a basic necessary scientific approach in a task which retains still some artistic flavor.

CHAPTER V

ESTIMATION ERRORS USING CSRS IN NONLINEAR SYSTEM IDENTIFICATION

As it was discussed in sec. 2.4, the actual application of the Wiener method encounters a variety of problems. The method, being theoretically general, elegant and powerful, is designed for a stimulus signal which is not physically realizable: the gaussian white noise. Therefore, we are promptly forced in practice to make use of quasi-white signals, which approximate the ideal white noise within a range of tolerable error.

The use of quasi-white signals, as well as the limited experimental and computational capacity of the "finite time - finite machine" research, induces a variety of inevitable estimation errors in the nonlinear system identification through the Wiener method. It is a principal task in the study of the Wiener identification method to look into the error producing mechanisms and try to determine both the optimal test conditions of a specific application and the magnitude of the commited errors.

In this section, we will study the several error producing mechanisms in the case of the CSRS. The main kinds of estimation errors, in the case of the CSRS, are: (1) The "deconvolution" error: this is due to the finite stimulus bandwidth and amounts to a loss of high frequencies in the kernel estimates. (2) The "statistical fluctuation" error: this is due to the finite record length and the random

nature of the CSRS. It amounts to some random deviations in the kernel estimates. (3) The "approximate orthogonality" error: this refers to the approximate orthogonality of the CSRS functional series, because of the finite stimulus bandwidth and the finite record length. The finite stimulus bandwidth makes the CSRS functional series only approximately orthogonal, while the finite record length makes the estimated functionals only approximately orthogonal to one another. (4) The "erroneous power level" error: this is due to erroneous estimation of the power level (and consequently of the normalizing factors) and amounts to scalar distortion (disproportionality) of the kernel estimates. (5) The "finite transition time" error: this is due to the finite response time of the stimulus transducer and it usually amounts to a small change of the power level. (6) The "computational" errors: these are due to the finite capacity of a digital computer and the digital processing of continuous data. They usually amount to negligible deviations of the kernel estimates, if proper operational procedures are followed.

In the following, the mechanisms that are producing each of these kinds of estimation errors will be studied, so as to suggest the optimum procedure which ought to be followed in every case in order to maximize the accuracy of the obtained system model.

5.1 The Deconvolution Error: (θ - error)

In order to study the mechanism that is producing the deconvolution estimation error, we first have to review the general deconvolution mechanism with which the kernels are estimated from the cross-correlation functions.

As an example, consider the estimation of the first order kernel from the first order crosscorrelation: (to simplify derivations, assume here an infinite record length).

$$\psi_1(\sigma) = E[y(t)x(t-\sigma)]$$

$$= \int_0^\infty h_1(\tau)\ E[x(t-\tau)x(t-\sigma)]\,d\tau$$

$$= P\int_0^\infty h_1(\tau)\delta(\sigma-\tau)d\tau = P \cdot h_1(\sigma) \qquad (5.1.1)$$

where $h_1(\tau)$ is the first order Wiener kernel and x(t) is ideal white noise. Now, if x(t) is a CSRS, the second order autocorrelation function is not a δ-function but an approximating triangle (Fig. 5.2.1a):

$$\psi_1(\sigma) = \int_0^\infty h_1(\tau)\phi_2(\sigma-\tau)d\tau$$

$$= m_2 \int_{\sigma+\Delta t} h_1(\tau)\left[1-\frac{|\sigma-\tau|}{\Delta t}\right]d\tau \qquad (5.1.2)$$

Thus, the crosscorrelation expression takes the form of a convolution integral, where one of the convolving functions has its significant values highly concentrated in a narrow area; therefore, approaching the form of an impulse function and subsequently possessing the ability to deconvolve approximately this convolution integral:

$$\psi_1(\sigma) \cong (m_2 \cdot \Delta t) \cdot h_1(\sigma) \qquad (5.1.3)$$

The deconvolution error that we intend to study here is exactly the deviation of this approximation. To evaluate this error, we evaluate the exact convolution integral from eqn. 5.1.2. Assuming $h_1(\tau)$ is analytic in the neighborhood of σ, we expand it in a Taylor series:

$$\psi_1(\sigma) = m_2 \int_{\sigma-\Delta t}^{\sigma+\Delta t} \left[h_1(\sigma) + h_1^{(1)}(\sigma)(\tau-\sigma) + \frac{h^{(2)}(\sigma)}{2}(\tau-\sigma)^2 + \dots \right] \cdot \left[1 - \frac{|\tau-\sigma|}{\Delta t} \right] d\tau$$

$$= (m_2 \cdot \Delta t) \sum_{n=o}^{\infty} \frac{\Delta t^{2n}}{2n!\,(2n+1)(n+1)} h_1^{(2n)}(\sigma) \quad (5.1.4)$$

Therefore,

$$\hat{h}_1(\sigma) = \frac{1}{(m_2 \cdot \Delta t)} \psi_1(\sigma) = \sum_{n=o}^{\infty} \frac{\Delta t^{2n}}{2n!(2n+1)(n+1)} h_1^{(2n)}(\sigma) \; ;$$

$$\sigma \geq \Delta t \quad (5.1.5)$$

Notice that since ϕ_2 is an even function, all the odd order terms vanish, so long as the kernel h_1 is analytic within the interval $[\sigma-\Delta t, \sigma+\Delta t]$.

In a lot of cases the kernel $h_1(\tau)$ is not analytic at $\tau=o$. In these cases, ϕ_2 is not symmetric in the area of analyticity of $h_1(\tau)$, and consequently, the odd order terms of the Taylor series expansion do not vanish with the integration. More specifically, this is the case where $o \leq \sigma \leq \Delta t$, and the convolution integral gives then the estimate:

$$\hat{h}_1(\sigma) = \sum_{n=o}^{\infty} \alpha_n(\lambda) \frac{\Delta t^n}{n!} h_1^{(n)}(\sigma) \quad ; \quad o \leq \sigma \leq \Delta t \quad (5.1.6)$$

where,

$$\alpha_n(\lambda) = \frac{1}{(n+1)(n+2)} + (-1)^n \frac{\lambda^{n+1}}{(n+1)} + (-1)^{n+1} \frac{\lambda^{n+2}}{(n+2)} \; ; \; o \leq \lambda = \frac{\sigma}{\Delta t} < 1$$

(5.1.7)

Therefore, special correctional procedure must be employed for the points of the initial region ($o \leq \sigma_i < \Delta t$) of the kernel estimates.

The expressions become more involved in higher order kernel cases; nevertheless, some basic remarks can be easily extended,

like: (1) because of the symmetry of the autocorrelation functions, only the even order partial derivatives are included in the expressions (for the points outside the initial region), (2) for sufficiently small Δt, the higher order terms decline very rapidly, and the second order terms become the principal error part, as:

$$\hat{h}_n(\sigma_1, \ldots, \sigma_n) = h_n(\sigma_1, \ldots, \sigma_n) + \frac{\Delta t^2}{12} \cdot \sum_{\substack{i,j,=1 \\ i \leq j}}^{n} \frac{\partial^2 h_n(\sigma_1, \ldots, \sigma_n)}{\partial \sigma_i \partial \sigma_j} + o(\Delta t^4)$$

$$\text{for } \sigma_k \geq \Delta t \quad ; k=1, 2, \ldots, n. \tag{5.1.8}$$

The complete expression for the n-th order kernel estimate is:

$$\hat{h}_n(\sigma_1, \ldots, \sigma_n) = \sum_{m=0}^{\infty} \frac{\Delta t^{2m}}{2m!\,(2m+1)(m+1)} \sum_{\substack{j_1, \cdots j_{2m}=1 \\ j_1 \leq j_2 \cdots \leq j_{2m}}}^{n} \frac{\partial^{2m} h_n(\sigma_1, \ldots, \sigma_n)}{\partial \sigma_{j1} \partial \sigma_{j2} \cdots \partial \sigma_{j2m}}$$

$$(\sigma_k \geq \Delta t \quad ; \quad k=1, 2, \ldots, n) \tag{5.1.9}$$

Therefore the deconvolution error for the n-th order kernel estimate is:

$$\theta_n(\sigma_1, \ldots, \sigma_n) = \sum_{m=1}^{\infty} \frac{\Delta t^{2m}}{2m!(2m+1)(m+1)} \sum_{\substack{j_1, \cdots, j_{2m}=1 \\ j_1 \leq j_2 \leq \cdots \leq j_{2m}}}^{n} \frac{\partial^{2m} h_n(\sigma_1, \ldots, \sigma_n)}{\partial \sigma_{j_1} \cdots \partial \sigma_{j2m}}$$

$$(\sigma_k \geq \Delta t \quad ; \quad k=1, 2, \ldots, n) \tag{5.1.10}$$

and, in first approximation, for Δt fairly small:

$$\theta_n(\sigma_1, \ldots, \sigma_n) \cong \frac{\Delta t^2}{12} \sum_{\substack{i,j=1 \\ i \leq j}}^{n} \frac{\partial^2 h_n(\sigma_1, \ldots, \sigma_n)}{\partial \sigma_i \, \partial \sigma_j}$$

$$(\sigma_k \geq \Delta t \; ; \quad k=1, 2, \ldots, n) \tag{5.1.11}$$

For later use, let us establish the notation:

$$\theta_n(\sigma_1, \ldots, \sigma_n) = \Delta t^2 \cdot D_{h_n}(\Delta t, \sigma_1, \ldots, \sigma_n) \tag{5.1.12}$$

where,

$$D_{h_n}(\Delta t, \sigma_1, \ldots, \sigma_n) = \sum_{m=1}^{\infty} \frac{\Delta t^{2m-2}}{2m!(m+1)(2m+1)} \sum_{\substack{j_1, \ldots j_{2m}=1 \\ j_1 \leq \ldots \leq j_{2m}}}^{n} \frac{\partial^{2m} h_n(\sigma_1, \ldots, \sigma_n)}{\partial \sigma_{j_1} \ldots \partial \sigma_{j_{2m}}} \tag{5.1.13}$$

Clearly, D_{h_n} depends very little on Δt, when Δt is much smaller than one. So, whenever Δt is very small, we accept approximately that:

$$D_{h_n}(\sigma_1, \sigma_2, \ldots, \sigma_n) = \frac{1}{12} \sum_{\substack{j_1, j_2=1 \\ j_1 \leq j_2}}^{n} \frac{\partial^2 h_n(\sigma_1, \ldots, \sigma_n)}{\partial \sigma_{j_1} \, \partial \sigma_{j_2}} \tag{5.1.14}$$

i.e. D_{h_n} is independent of Δt, for Δt in the range of values of practical interest.

Notice, however, that the initial region may disturb the square relation between Δt and θ_n. This is because in the initial region θ_n depends on Δt according to polynomial relation including all powers of Δt (cf. eqn. 5.1.6). For example, in the first order kernel case the

deconvolution error for the initial region ($o \leq \sigma < \Delta t$) is:

$$\theta_1(\sigma) = [\alpha_o(\lambda)-1]h_1(\sigma)+\alpha_1(\lambda)h^{(1)}(\sigma)\cdot\Delta t + \frac{\alpha_2(\lambda)}{2} h^{(2)}(\sigma)\Delta t^2 \ldots \tag{5.1.15}$$

where $\lambda = \sigma/\Delta t$.

Thus,

$$D_{h_1}(\sigma) = \beta_o(\sigma)+\beta_1(\sigma)\cdot\Delta t+\beta_2(\sigma)\cdot\Delta t^2+\beta_3(\sigma)\Delta t^3+\beta_4(\sigma)\Delta t^4+\ldots \tag{5.1.16}$$

where, $\beta_o(\sigma), \beta_1(\sigma), \beta_3(\sigma), \beta_5(\sigma), \ldots$ (and the rest of the odd order terms) are nonzero only when $o \leq \sigma < \Delta t$.

Therefore, the dependence of $\theta_1(\sigma)$ on Δt becomes analytically complicated if the initial region is to be counted in. For this reason, it is suggested that, in practice, we compute the deconvolution error over the space of the kernel outside the initial region (i. e. $o \leq \sigma_i \leq \Delta t$), in order to keep the analytical relation between θ_n and Δt in the simple form described by eqn. 5.1.12. This point is essential in the design of the optimum test as discussed in sec. 6.5.

5.2 The Statistical Fluctuation Error: (ϵ-error)

In the previous section, the autocorrelation functions were computed as statistical averages. However, in practice, we have to compute the averages by integrating in time over finite time intervals. Thus, in practice, we can only attain an estimate of the autocorrelation functions, as:

$$\hat{\phi}_n(\tau_1, \ldots, \tau_n) = \frac{1}{T-Tm} \int_{Tm}^{T} x(t-\tau_1)\ldots x(t-\tau_n)dt$$

$$Tm = \max [\tau_1, \ldots, \tau_n] \tag{5.2.1}$$

Usually, the record length T is much longer than the memory of the system (and, subsequently, much bigger than Tm) and therefore we can write:

$$\hat{\phi}_n(\tau_1, \ldots, \tau_n) \simeq \frac{1}{T} \int_0^T x(t-\tau_1) \ldots x(t-\tau_n)\, dt \qquad (5.2.2)$$

This autocorrelation estimate $\hat{\phi}_n$ is a random variable itself (cf. sec. 4.2) and, consequently, its convolution with a kernel will give a kernel estimate with random characteristics. Therefore, the kernel estimates, that we obtain in practice by averaging over finite records, are deviating randomly from the exact kernels. This random deviation of the kernel estimates constitutes the statistical fluctuation error.

5.2.1 The Dependence of the Statistical Fluctuation Error on T and Δt.

In the following, we will try to evaluate this error in the case of a CSRS stimulus. At first, let us study the first order kernel case:

$$\hat{h}_1(\sigma) = \frac{1}{(m_2 \cdot \Delta t)} \int_0^\infty h_1(\tau) \hat{\phi}_2(\sigma-\tau)\, d\tau \qquad (5.2.3)$$

where,

$$\hat{\phi}_2(\sigma-\tau) = \frac{1}{T - \max[\tau, \sigma]} \int_{\max[\sigma,\tau]}^T x(t-\tau)x(t-\sigma)\, dt$$

$$\cong \frac{1}{T} \int_0^T x(t-\tau)x(t-\sigma)\, dt \quad ; \quad \text{since } T >> \sigma, \tau. \qquad (5.2.4)$$

Every value of $\hat{\phi}_2(\sigma-\tau)$ is a random variable with:

$$E[\hat{\phi}_2(\sigma-\tau)] = \phi_2(\sigma-\tau) \qquad (5.2.5)$$

$$\text{Var}[\hat{\phi}_2(\sigma-\tau)] = \frac{1}{T^2} \int_0^T \int_0^T E[x(t-\tau)x(t-\sigma)x(t'-\tau)x(t'-\sigma)]\, dt\, dt'$$

$$- \phi_2{}^2(\tau-\sigma)$$

$$= \begin{cases} \frac{\Delta t}{T} m_2{}^2 & (\text{for } \tau \neq \sigma) \\ \frac{\Delta t}{T} (m_4 - m_2{}^2) & (\text{for } \tau = \sigma) \end{cases} \qquad (5.2.6)$$

These expressions for the variance refer only to the nodal points $(\sigma-\tau)$, and they are accurate under the condition that the memory of the studied system is much smaller than the record length T, and much bigger than the step length Δt. $(\Delta t << \mu << T)$ From the expression (5.2.6), we clearly see that the variance of the second order autocorrelation estimate is inversely proportional to the number of steps of the stimulus signal. The same holds true for the higher order autocorrelation estimates.

Consider now the "deviation function":

$$r_2(\sigma-\tau) = \hat{\phi}_2(\sigma-\tau) - \phi_2(\sigma-\tau) \qquad (5.2.7)$$

which is the source of the statistical fluctuation error. Clearly,

$$\hat{h}_1(\sigma) = \frac{1}{m_2 \cdot \Delta t} \int_o^\infty h_1(\tau)\phi_2(\sigma-\tau)d\tau + \frac{1}{m_2 \cdot \Delta t} \int_o^\infty h_1(\tau) r_2(\sigma-\tau)d\tau$$

$$= [\, Ip_1(\sigma) + Is_1(\sigma) \,] / (m_2 \cdot \Delta t) \qquad (5.2.8)$$

We call these two integrals, the principal and the side part integral respectively. The principal part integral is deterministic, while the side part integral is random in nature. In order to evaluate the statistical fluctuation error ϵ_1, we simply have to evaluate the side part integral Is_1, which is the convolution of the deviation function r_2 with the kernel h_1: (see Fig. 5.2.1)

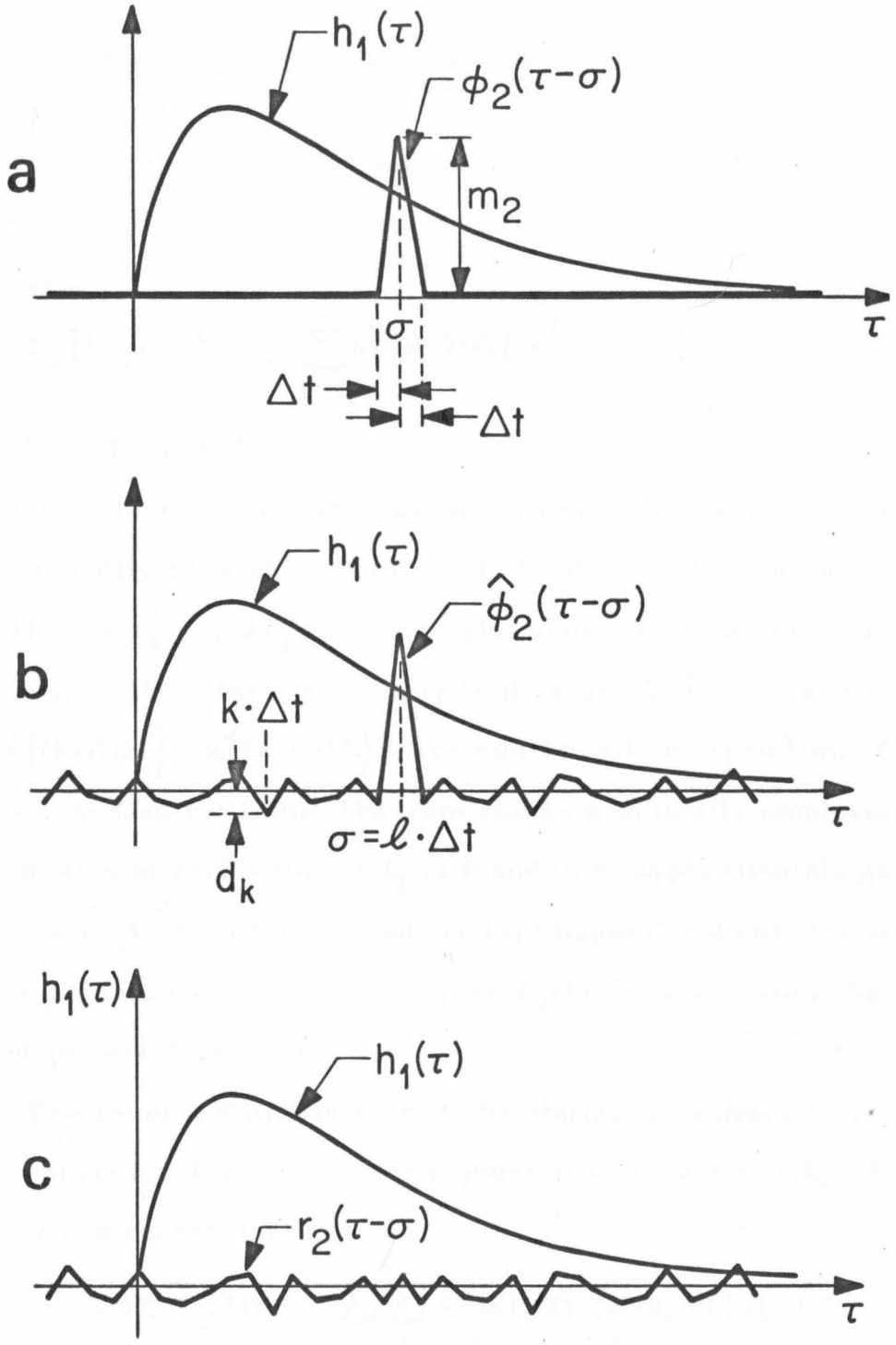


Fig. 5.2.1: Convolution of 1st order kernel with:
(a) 2nd order CSRS autocorrelation function,
(b) 2nd order CSRS autocorrelation estimate,
(c) 2nd order CSRS deviation function.

$$\epsilon_1(\sigma) = Is_1(\sigma)/(m_2 \cdot \Delta t) = \frac{1}{(m_2 \cdot \Delta t)} \int_o^\infty h_1(\tau) r_2(\sigma - \tau) d\tau \quad (5.2.9)$$

But first, we have to study the statistical structure of the deviation function r_2. To this purpose, notice that $\phi_2(\sigma - \tau)$ is computed in practice as:

$$\hat{\phi}_2\Big((i-j)DT\Big) = \frac{DT}{T} \sum_{k=1}^{N} x\Big((k-i)DT\Big)\, x\Big((k-j)DT\Big)$$

$$(t = kDT,\ T = N \cdot DT) \quad (5.2.10)$$

where, DT is the sampling interval, not necessarily equal to the step length Δt (usually Δt is an integral multiple of DT). Of course, $x\Big((k-i)DT\Big)$ and $x\Big((k-j)DT\Big)$ are uncorrelated for any k, as long as $|(i-j)DT| \geq \Delta t$. Also, for a given pair (i, j), a great number of the products $[\, x\Big((k-i)DT\Big) \cdot x\Big((k-j)DT\Big)\,]$ are statistically independent. Consequently, the Central Limit Theorem can be practically employed and the distribution of each value of $\hat{\phi}_2$ is found to be approximately gaussian. (cf. sec. 4.2) In the case where i=j (diagonal points), the same reasoning holds; and thus, all the values $\hat{\phi}_2(kDT)$, are found to be gaussian random variables.

The several $\hat{\phi}_2(kDT)$ seem to be statistically dependent, since they are constructed with the same population of values x(iDT). However, they are uncorrelated, since:

$$E[\,\hat{\phi}_2(kDT)\hat{\phi}_2(\ell DT)\,] = \sum_i \sum_j E[\, x(iDT)\, x\Big((i+k)DT\Big)\, x(jDT)\, x\Big((j+\ell)DT\Big)\,]$$

$$= o \quad \text{for } k \neq \pm \ell \quad (5.2.11)$$

In conclusion, if we call d_k the values of the deviation function r_2 at the nodal points, then the several d_k (of one side of the symmetric deviation function) are found to be uncorrelated gaussian random variables with zero mean and variance m_2^2/N for $k \neq o$ and $(m_4 - m_2^2)/N$ for $k=o$. (for $N=T/\Delta t$ and $T \gg \mu$)

In the following analysis, we will consider that the variance of d_o is the same with the variance of all other d_k. This approximation is not expected to cause significant errors while it facilitates greatly the derivations.

Let us consider again the principal and the side part integrals (Fig. 5.2.1) The principal part integral was evaluated in the previous section 5.1. We will now evaluate the side part integral, which is, as exposed above (eqn. 5.2.9), the source of the statistical fluctuation error.

The deviation function $r_2(\sigma - \tau)$ is a piecewise linear function, therefore, we first evaluate a portion δ_k of $Is_1(\sigma)$ integrating between two successive nodal points:

$$\delta_k = \int_{k\Delta t}^{(k+1)\Delta t} h_1(\xi) \left[d_k\left(k+1-\frac{\xi}{\Delta t}\right) - d_{k+1}\left(k-\frac{\xi}{\Delta t}\right) \right] d\xi \qquad (5.2.12)$$

Let,

$$H_1(k) = \int_{-\infty}^{k\Delta t} h_1(\sigma)\, d\sigma \qquad (5.2.13)$$

$$G_1(k) = \int_{-\infty}^{k\Delta t} g(\sigma) \int_{-\infty}^{\sigma} h_1(\lambda)\, d\lambda \qquad (5.2.14)$$

Then,

$$\delta_k = \left[(k+1)\alpha_k - \beta_k \right] d_k + \left[\beta_k - k\alpha_k \right] d_{k+1} \qquad (5.2.15)$$

where,

$$\alpha_k = H_1(k+1) - H_1(k) \tag{5.2.16}$$

$$\beta_k = (k+1) \cdot H_1(k+1) - k \cdot H_1(k) - \frac{1}{\Delta t}[G_1(k+1) - G_1(k)] \tag{5.2.17}$$

and since $r_2(k\Delta t)$ is symmetric about $\sigma = \ell \cdot \Delta t$:

$$I_{s1}(\sigma) = \sum_{k=o}^{M} \delta_k = \sum_{k=\ell-M}^{\ell-1} (\delta_k + \delta_{2\ell-k-1}) \tag{5.2.18}$$

where M is the number of steps covering the memory of the kernel, i.e. $M = \mu/\Delta t$, and $\delta_k = o$ if $k < o$ or $k > M$. Substituting the values of δ_k from eqn. 5.2.15 into eqn. 5.2.18, we have:

$$I_{s1}(\sigma) = \sum_{k=o}^{M} \gamma_k \cdot d_{\ell-k} \tag{5.2.19}$$

where,

$$\gamma_k = \begin{cases} C_{\ell+k} + C_{\ell-k} & \text{for } k = 1, \ldots, M \\ C_\ell & \text{for } k = o \end{cases} \tag{5.2.20}$$

and,

$$C_k = \frac{1}{\Delta t}[G_1(k+1) - 2G_1(k) + G_1(k-1)]$$
$$= \sum_{m=1}^{\infty} \frac{2}{2m!} \Delta t^{2m-1} h_1^{(2m)}(k \cdot \Delta t) \tag{5.2.21}$$

Clearly, for sufficiently small Δt, $\frac{C_k}{\Delta t}$ is an estimate of the second derivative of $G_1(k)$, which is the kernel $h_1(k \cdot \Delta t)$. Therefore, the dependence of C_k on Δt is primarily dominated by a linear relation. The direct implication of this fact is that, if we define:

$$S_k \triangleq \frac{1}{\Delta t}[C_{\ell+k} + C_{\ell-k}] \tag{5.2.22}$$

then,

$$I_{s_1}(\sigma) = \Delta t \cdot \sum_{k=o}^{M} S_k \cdot d_{\ell-k} \tag{5.2.23}$$

where, S_k depends mainly on h_1 and very slightly on Δt (for values of Δt reasonably small).

Notice also that $I_{s_1}(\sigma)$ is a gaussian random variable (as weighted sum of uncorrelated gaussian random variables) with:

$$E[I_{s_1}(\sigma)] = o \tag{5.2.24}$$

$$\text{Var}[I_{s_1}(\sigma)] \triangleq q_1^2(\ell, \Delta t) = \Delta t^2 \cdot \sigma^2_{d_1} \cdot \sum_{k=o}^{M} S_k^2 \tag{5.2.25}$$

where, $\sigma^2_{d_1}$ is the variance of each nodal point of the deviation function r_2:

$$\sigma^2_{d_1} = \frac{m_2^2 \cdot \Delta t}{T} \tag{5.2.26}$$

Thus, we can write:

$$q_1^2(\ell, \Delta t) = \frac{(m_2 \cdot \Delta t)^2}{T} C_{h_1}^2(\ell, \Delta t) \tag{5.2.27}$$

where,

$$C_{h_1}^2(\ell, \Delta t) = \Delta t \cdot \sum_{k=o}^{M} S_k^2 \tag{5.2.28}$$

Notice that $C_{h_1}^2$ depends very slightly on Δt, for Δt fairly small, which is usually the case in practice.

Going back to eqn. 5.2.9, we see that the statistical fluctuation error in the first order kernel case is a gaussian random variable with:

$$E[\epsilon_1(\sigma)] = o \tag{5.2.29}$$

$$\text{Var}\left[\epsilon_1(\sigma)\right] = \frac{1}{T} C_{h_1}^2(\ell, \Delta t) \quad ; \quad \sigma = \ell \cdot \Delta t \tag{5.2.30}$$

where $C_{h_1}^2(\ell, \Delta t)$ depends very slightly on Δt, for $\Delta t << \mu$, which is the case of practical interest.

A much simpler but less rigorous way to arrive at the same result is by studying the statistical behavior of the deviation function. The side part integral is the convolution of the kernel (of some order) with the corresponding deviation function. The statistical character of the side part integral is attributed to the statistical character of the deviation function, since the kernel is a deterministic function. In that sense, the side part integral is a gaussian random variable produced by a weighted summation of uncorrelated gaussian random variables (namely, the values of the deviation function at the nodal points). The weights in this summation depend on the kernel and on Δt. We are trying to determine the pattern of dependence of the side part integral on Δt. To this purpose, it is necessary to study the pattern of dependence of the deviation function on Δt. After this is done, it is reasonable to assert that the pattern of dependence of the side part integral on Δt is similar to the one of the deviation function on Δt.

The side part integral, in the first order case, is a gaussian random variable with:

$$E\left[I_{s_1}(\sigma)\right] = o \tag{5.2.31}$$

and,

$$\text{Var}\left[I_{s_1}(\sigma)\right] = E\left\{\int_o^\infty\int h_1(\tau)h_1(\tau')r_2(\sigma-\tau)r_2(\sigma-\tau')d\tau d\tau'\right\}$$

$$\cong \alpha_1 \Delta t \sigma_{d1}^2 \int_o^\infty h_1^2(\tau)d\tau \tag{5.2.32}$$

since,

$$E[\, r_2(\sigma-\tau) r_2(\sigma-\tau') \,] \cong \alpha_1 \cdot \Delta t\, \sigma_{d_1}^{\,2} \cdot \delta(\tau-\tau')$$

$$(\text{for } \Delta t << \mu << T) \qquad (5.2.33)$$

where, α_1 is a constant of proportionality, which depends on the area under the parabolic principal lobe of $E[\, r_2(\sigma-\tau) r_2(\sigma-\tau') \,]$.

Therefore, the variance of the statistical fluctuation error is found again to be approximately independent from Δt.

$$q_1^{\,2}(\sigma, \Delta t) \simeq \frac{\alpha_1}{T} \int_0^{\infty} h_1^{\,2}(\tau)\, d\tau \qquad (5.2.34)$$

comparing expressions 5.2.30 and 5.2.34 , we see that:

$$C_{h_1}^{\,2}(\sigma, \Delta t) \equiv \alpha_1 \int h_1^{\,2}(\tau) d\tau \qquad (5.2.35)$$

which implies that the dependence of $C_{h_1}^{\,2}$ on both σ and Δt is very weak, for $\Delta t << \mu << T$.

In these two ways, we have illustrated the dependence of the statistical fluctuation error on the important test parameters (record length and step length), for the first order kernel case. Unfortunately, these two ways of reasoning are not easily extendible to the higher order kernel cases. For this reason, a new approach has to be followed, which is easily extendible to the higher order kernel cases.

This new approach is based upon the following basic remark: The side part integral is the convolution of the kernel with the corresponding deviation function. The convolution integral is a linear operation. Consequently, the gaussian random characteristics of the deviation function are going to be preserved, by this operation. Thus, if

characteristics depend in some way on a parameter, then the random characteristics of the convolution integral are going to depend on that parameter in the same way. This is more evident in the case where the random function does not have any directionality, i.e. when the statistical properties of the convolution of this random function with any unitary (i.e. the integral of the square of the function is one) deterministic function are the same.

According to eqns. 5.2.31 and 5.2.34 , the side part integral I_{s1} clearly belongs to this category, since its gaussian statistical properties are completely described by the first two moments, and these moments are shown independent from Δt.

Notice that this convolution integral is analogous to the inner product of a random and a deterministic unitary vector, in a linear vector space of infinite dimensionality. In this sense, the "random vector" r_2 has spherical directionality, i.e. it does not exhibit any preferable directionality but it attains all directions with the same likelihood. The direct implication of this, is that the inner product of the "random vector" r_2 with the diagonal vector (step function) can be used as a reliable indicator of the statistical behavior of the inner product of r_2 with any other deterministic unitary vector of the space.

In conclusion, the parametric study of the statistical behavior of the side part integral can be interchanged, for the purposes of the present analysis, with the study of the statistical behavior of the integral of the deviation function (inner product with the step function).

Thus, for the first order case this quantity (hereafter called

the "integrated deviation") is:

$$R_2 = \int_o^{\mu_1} r_2(\lambda)d\lambda \qquad (5.2.36)$$

where μ_1 is the first order memory of the system and $r_2(\lambda)$ is taken to be symmetric about zero. Since r_2 is piecewise linear, the evaluation of the integral R_2 requires the evaluation of each linear portion:

$$U_k = \int_{k\Delta t}^{(k+1)\Delta t} r_2(\lambda)d\lambda \quad ; \quad k=o, 1, 2, \ldots, M_1-1 \; ; \; M_1 = \mu_1/\Delta t. \qquad (5.2.37)$$

where, $r_2(\lambda) = d_k\left[(k+1) - \frac{\lambda}{\Delta t}\right] - d_{k+1}\left(k - \frac{\lambda}{\Delta t}\right)$, for $k\Delta t \leq \lambda \leq (k+1)\Delta t$ (5.2.38)

Evaluating the integral in eqn. 5.2.37:

$$U_k = \frac{\Delta t}{2}(d_k + d_{k+1}) \qquad (5.2.39)$$

and the integrated deviation becomes:

$$R_2 = \sum_{k=o}^{M_1-1} U_k \quad \cong \quad \sum_{k=1}^{M_1} \Delta t \cdot d_k \qquad (5.2.40)$$

Therefore, R_2 is a gaussian random variable with:

$$E[R_2] = o \qquad (5.2.41)$$

$$Var[R_2] \cong \Delta t^2 \cdot \sigma_{d_1}^{\;2} \cdot M_1$$

$$\cong (m_2 \cdot \Delta t)^2 \cdot \frac{\mu_1}{T} \qquad (5.2.42)$$

Note that the integrated deviation must be normalized by the normalizing factor of the respective order of kernel. Thus, finally:

$$q_1^{\;2}(\sigma, \Delta t, T) \sim \frac{1}{T} \qquad (5.2.43)$$

So, we conclude again that the variance of the statistical fluctuation

error of first order is independent from Δt.

As we said before, this approach is easily extendible to the higher order kernel cases. For example, consider the second order case. The second order crosscorrelation estimate at a nondiagonal point (σ_1, σ_2) is:

$$\hat{\psi}_2(\sigma_1, \sigma_2) = \frac{1}{T}\int_o^T y(t)x(t-\sigma_1)x(t-\sigma_2)dt$$

$$= \int_o^\infty\int h_2(\tau_1, \tau_2)\hat{\phi}_4(\sigma_1-\tau_1, \sigma_2-\tau_2)d\tau_1 d\tau_2 \qquad (5.2.44)$$

where,

$$\hat{\phi}_4(\sigma_1-\tau_1, \sigma_2-\tau_2) = \frac{1}{T}\int_o^T x(t-\tau_1)x(t-\tau_2)x(t-\sigma_1)x(t-\sigma_2)dt \qquad (5.2.45)$$

The function $\hat{\phi}_4$ was discussed in sec. 4.3, and it is a sectionally linear (with respect to τ_1 and τ_2) three dimensional surface, for a given point (σ_1, σ_2). Introducing the deviation function r_4:

$$r_4(\sigma_1-\tau_1, \sigma_2-\tau_2) = \hat{\phi}_4(\sigma_1-\tau_1, \sigma_2-\tau_2) - \phi_4(\sigma_1-\tau_1, \sigma_2-\tau_2) \qquad (5.2.45)$$

we have:

$$\hat{\psi}_2(\sigma_1, \sigma_2) = \int_o^\infty\int h_2(\tau_1, \tau_2)\phi_4(\sigma_1-\tau_1, \sigma_2-\tau_2)d\tau_1 d\tau_2$$

$$+ \int_o^\infty\int h_2(\tau_1, \tau_2)r_4(\sigma_1-\tau_1, \sigma_2-\tau_2)d\tau_1 d\tau_2$$

$$= Ip_2(\sigma_1, \sigma_2) + I_{s_2}(\sigma_1, \sigma_2) \qquad (5.2.47)$$

Again, the side part integral $I_{s2}(\sigma_1, \sigma_2)$ is the random part of the crosscorrelation estimate $\hat{\psi}_2(\sigma_1, \sigma_2)$. In order to study the statistical behavior of I_{s_2}, we will simply study the statistical behavior of the integrated deviation:

$$R_4 = \int\int^{\mu_2} r_4(\lambda_1, \lambda_2)d\lambda_1 d\lambda_2 \qquad (5.2.48)$$

where μ_2 is the second order memory of the system and r_4 is symmetric about the diagonal.

The deviation function r_4 consists of triangular plane segments, as indicated in Fig. 5.2.2. Let us denote the deviation values at the nodal points as:

$$d_{i,j} = r_4(i\Delta t, j\Delta t) \tag{5.2.49}$$

To evaluate the integral (5.2.48), we have to evaluate the integral over each triangular plane segment. Let us denote by $D^R_{i,j}$ the triangular plane segment with its right angle at ($\lambda_1 = i\Delta t$, $\lambda_2 = j\Delta t$) and on the left of its hypotenuse (analogously we define $D^L_{i,j}$).

In this way, the deviation function for $D^R_{i,j}$ is: (Fig. 5.2.2)

$$r_4(\lambda_1, \lambda_2) = \left[i-j+1 + \frac{\lambda_2 - \lambda_1}{\Delta t}\right] d_{ij} + \left[j - \frac{\lambda_2}{\Delta t}\right] d_{i,j-1} + \left[\frac{\lambda_1}{\Delta t} - i\right] d_{i+1,j} \tag{5.2.50}$$

and consequently:

$$U^R_{i,j} = \int_{\lambda_1 = i\Delta t}^{(i+1)\Delta t} d\lambda_1 \int_{\lambda_2 = \lambda_1 - (i-j+1)\Delta t}^{j\Delta t} r_4(\lambda_1, \lambda_2)\, d\lambda_2$$

$$= \frac{\Delta t^2}{6} \left(d_{i,j} + d_{i,j-1} + d_{i+1,j}\right) \tag{5.2.51}$$

Therefore,

$$R_4 = \sum_{i=o}^{M_2-1} \sum_{j=o}^{M_2-1} \left(U^R_{i,j} + U^L_{i,j}\right)$$

$$\cong \sum_{i=1}^{M_2} \sum_{j=1}^{M_2} \Delta t^2 \cdot d_{i,j} \quad ; \quad M_2 = \mu_2/\Delta t \tag{5.2.52}$$

In conclusion, R_4 is a gaussian random variable with:

$$E[R_4] = o \tag{5.2.53}$$

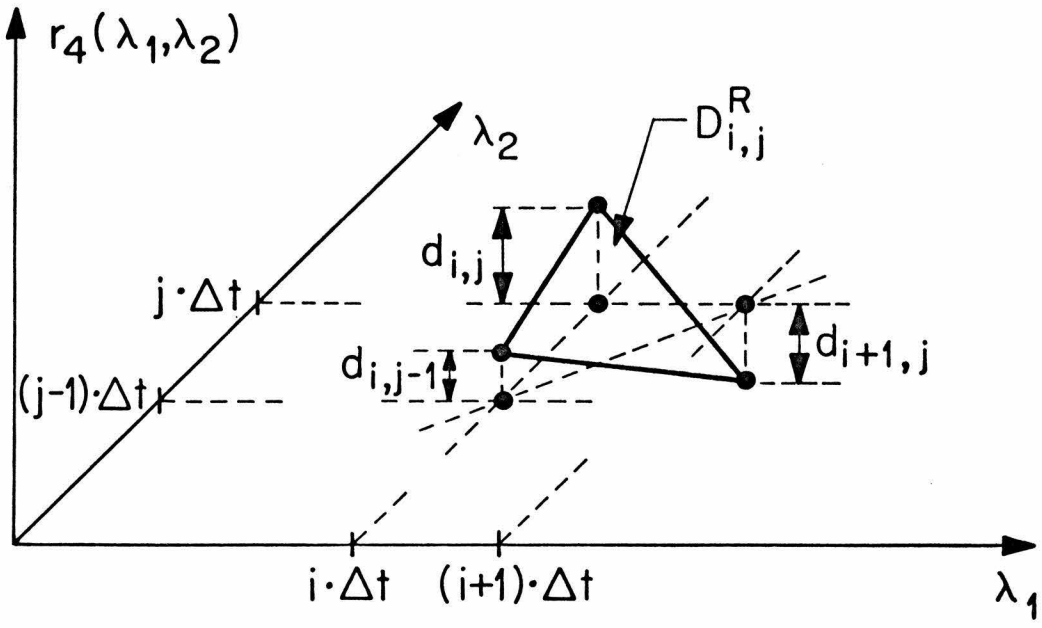


Fig. 5.2.2: Triangular plane segment of fourth order deviation function

$$\text{Var}\,[R_4] = \Delta t^4 \cdot \sigma_{d_2}^2 \cdot M_2^2$$

$$\equiv (m_2 \cdot \Delta t)^4 \cdot \frac{\mu_2^2}{T \cdot \Delta t} \qquad (5.2.54)$$

$$q_2^2(\sigma_1, \sigma_2, \Delta t, T) \sim \frac{1}{T \cdot \Delta t} \qquad (5.2.55)$$

Therefore, in the second order case, the variance of the statistical fluctuation error is inversely proportional to Δt.

It must be noted that for these derivations, a basic simplifying assumption is always made: that the variance of the deviation function at all the nodal points is the same; while, we saw in sec. 4.3 that it depends on the order of the nodal point. This assumption is not expected to cause any significant distortion to the derived expressions and conclusions.

Following the same basic steps, this approach can be extended to the n-th order case. There, the integrated deviation is:

$$R_{2n} = \int_0^{\mu_n} \cdots \int_0^{\mu_n} r_{2n}(\lambda_1, \ldots, \lambda_n) d\lambda_1 \cdots d\lambda_n$$

$$\equiv \sum_{k_1=1}^{Mn} \cdots \sum_{k_n=1}^{Mn} \Delta t^n \cdot d_{k_1 k_2 \cdots k_n} \quad ; \quad M_n = \mu_n / \Delta t \qquad (5.2.56)$$

where μ_n is the n-th order memory of the system and $r_{2n}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ a symmetric function with respect to the arguments $(\lambda_1, \lambda_2, \ldots \lambda_n)$. From eqn. 5.2.56 follows that R_{2n} is a gaussian random variable with:

$$E[R_{2n}] = o \qquad (5.2.57)$$

$$\text{Var}\,[R_{2n}] = \Delta t^{2n} \cdot \sigma_d^2 \cdot M_n^n$$

$$\simeq (m_2 \cdot \Delta t)^{2n} \cdot \frac{\mu_n^{\,n}}{T \cdot \Delta t^{n-1}} \qquad (5.2.58)$$

since, $\sigma_{d_n}^2 = m_2^{2n} \cdot \frac{\Delta t}{T}$ (5.2.59)

Consequently,

$$q_n^2(\sigma_1, \ldots, \sigma_n, \Delta t, T) \sim \frac{1}{T \cdot \Delta t^{n-1}} \qquad (5.2.60)$$

that is, the variance of the statistical fluctuation error, in the n-th order case, is inversely proportional to the record length T and to the (n-1) -th power of the step length Δt.

Of course, this is an approximate result, which illustrates the principal pattern of dependence of the statistical fluctuation error on T and Δt, in the cases where $\Delta t << \mu << T$ (which are the cases of interest in practice). Evidently, the exact pattern of dependence of q_n^2 on Δt is related to the specific kernel (system) involved; however, this is of secondary importance for all practical purposes, as it was previously illustrated in the first order kernel case.

In general, it can be stated that the variance of the statistical fluctuation error in the n-th order kernel case and in all situations of practical interest ($\Delta t << \mu << T$) is:

$$q_n^2(\sigma_1, \ldots, \sigma_n, \Delta t, T) = \frac{C_{h_n}^2(\sigma_1, \ldots, \sigma_n, \Delta t)}{T \cdot \Delta t^{n-1}} \qquad (5.2.61)$$

where, the function $C_{h_n}^2(\sigma_1, \ldots, \sigma_n, \Delta t)$ depends very slightly on Δt.

A special note must be made with respect to the zero order kernel estimate $\hat{g}_o$; since the eqn. 5.2.61 does not hold in that case. The estimation error of $\hat{g}_o$ includes both deconvolution and statistical fluctuation errors. The dependence of these errors upon the test

parameters Δt and T can be studied with the help of the original Volterra series.

Recalling that the estimation of the CSRS zero order kernel comprises simply the time averaging of the system response to the CSRS stimulus, we reason that the mean square error of $\hat{g}_o$ is given by an expression of the form:

$$q_o^2 = \sum_{n=2}^{\infty} \alpha_{o,n} \cdot \Delta t^{2n} + \frac{1}{T} \sum_{m=1}^{\infty} \beta_{o,m} \cdot \Delta t^m \qquad (5.2.62)$$

where, the coefficients $\alpha_{o,n}$ depend on the even order partial derivatives of the even order Volterra kernels at the origin of the respective argument space, and the coefficients $\beta_{o,m}$ depend on the Volterra kernels of all orders higher than zero.

The pattern of dependence of q_o^2 on Δt, as described by eqn. 5.2.62, is very complicated to allow any general conclusion which would be of practical usefulness. However, it is evident that, if Δt is very small, then the primary dependence of q_o^2 on Δt is linear. It must be also noted here, that the accuracy of the zero order kernel estimate is usually of secondary importance in the applications.

5.2.2 Conclusions and Discussion on the Statistical Fluctuation Error (ϵ-error)

The ϵ-error is due to the random fluctuations of the estimates of the autocorrelation functions, and consequently it is of statistical nature. The practical limitation that causes this error is the finite record length. The ϵ-error is monotonically decreasing as the record length increases. However, the record length is not the only param-

eter affecting the ϵ-error. It was shown in the previous subsection, that the stimulus bandwidth is an even more crucial factor affecting this error in the higher order kernel estimates. The degree of this dependence increases rapidly with the order of the kernel estimate under study.

The ϵ-error at each point of the n-th order kernel estimate has been found to be a gaussian random variable with zero mean and variance given by eqn. 5.2.61. The function $C^2_{h_n}(\sigma_1, \ldots, \sigma_n, \Delta t)$, that appears in eqn. 5.2.61, depends heavily on h_n and very slightly on Δt (for all cases of practical interest, where $\Delta t << \mu << T$).

We can obtain an expression of the mean square ϵ-error of the n-th order kernel estimate by integrating $q_n^2(\sigma_1, \ldots, \sigma_n)$ over the whole memory space of the kernel:

$$\overline{\epsilon_n^2} = \int_0^{\mu_n} \cdots \int_0^{\mu_n} q_n^2(\sigma_1, \ldots, \sigma_n, \Delta t) d\sigma_1 \cdots d\sigma_n$$

$$= \frac{B_n(\Delta t)}{T \cdot \Delta t^{n-1}} \qquad (5.2.63)$$

where,

$$B_n(\Delta t) = \int_0^{\mu_n} \cdots \int_0^{\mu_n} C_{h_n}^2(\sigma_1, \ldots, \sigma_n, \Delta t) d\sigma_1 \cdots d\sigma_n \qquad (5.2.64)$$

Clearly, $B_n(\Delta t)$ depends very slightly on Δt, in a way to be considered constant for all practical purposes, as long as Δt remains in the range of values of interest. ($\Delta t << \mu_n$) Of course, B_n is characteristic of the kernel h_n, and it also depends on the specified memory extent μ_n.

5.3 The Approximate Orthogonality Errors

There are two kinds of approximate orthogonality errors. The

one refers to the approximate orthogonality of the G*-functionals as they are constructed (sec. 4.3) and it is due to the finite stimulus bandwidth. The other refers to the approximate orthogonality of the estimated G*-functionals, because of the finite records that are used in the estimation procedure (crosscorrelation technique).

5.3.1 Approximate Orthogonality Error Due to the Finite Stimulus Bandwidth: (ξ-error)

From the way that the G*-functionals are constructed, it is evident that their orthogonality is complete only when the stimulus bandwidth is infinite (ideal white noise). In practice, of course, the bandwidth of a quasi-white signal is finite and the orthogonality of the G*-functionals (of order higher than one) holds only approximately. The degree of this approximation becomes higher as the stimulus bandwidth becomes broader with respect to the sytem bandwidth.

To illustrate the error of this sort, let us consider the case of the second order G*-functional:

$$G_2^*(t) = \int_0^\infty\int g_2(\tau_1, \tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2 - (m_2\Delta t)\int_0^\infty g_2(\tau, \tau)d\tau \qquad (5.3.1)$$

However, the exact orthogonal form for the CSRS is:

$$\overline{G_2^*}(t) = \int_0^\infty\int g_2(\tau_1, \tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2 - \int_0^\infty\int g_2(\tau_1, \tau_2)\phi_2(\tau_2-\tau_1)d\tau_1 d\tau_2 \qquad (5.3.2)$$

where, ϕ_2 is the second order autocorrelation function of the CSRS. Thus, G_2^* and $\overline{G_2^*}$ differ at the second integral term.

The second integral term of G_2^* can be written as:

$$\int_0^\infty\int g_2(\tau_1,\tau_2)\phi_2(\tau_2-\tau_1)d\tau_1 d\tau_2 = \int_0^\infty d\tau \int_{-\Delta t}^{\Delta t} g_2(\tau,\tau+\lambda)m_2\left(1-\frac{|\lambda|}{\Delta t}\right)d\lambda$$

$$= \int_0^\infty d\tau \int_{-\Delta t}^{\Delta t}\left[g_2(\tau,\tau) + \left[\frac{\partial g_2(\tau,\sigma)}{\partial\sigma}\right]_{\sigma=\tau}\cdot\lambda + \left[\frac{\partial^2 g_2(\tau,\sigma)}{\partial\sigma^2}\right]_{\sigma=\tau}\cdot\frac{\lambda^2}{2} + \ldots\right] m_2\left(1-\frac{|\lambda|}{\Delta t}\right)d\lambda \qquad (5.3.3)$$

if we expand $g_2(\tau,\tau+\lambda)$ in a Taylor series in the region of its analyticity. Since the interval of integration of λ is symmetric, the odd power terms of λ will vanish with the integration, and therefore:

$$\int_0^\infty\int g_2(\tau_1,\tau_2)\phi_2(\tau_2-\tau_1)d\tau_1 d\tau_2$$

$$= (m_2\Delta t)\int_0^\infty d\tau \cdot \sum_{n=0}^{\infty}\frac{\Delta t^{2n}}{2n!\,(2n+1)(n+1)}\left[\frac{\partial^{2n} g_2(\tau,\sigma)}{\partial\sigma^{2n}}\right]_{\sigma=\tau} \qquad (5.3.4)$$

Clearly, the approximate orthogonality error of first kind in the second order case is:

$$\xi_2 = G_2^* - \overline{G_2^*} = (m_2\cdot\Delta t)\sum_{n=1}^{\infty}\frac{\Delta t^{2n}}{2n!(2n+1)(n+1)}\int_0^\infty\left[\frac{\partial^{2n} g_2(\tau,\sigma)}{\partial\sigma^{2n}}\right]_{\sigma=\tau} d\tau \qquad (5.3.5)$$

which obviously becomes negligible when Δt becomes very small (very broad stimulus bandwidth). Similar expressions can be derived for the higher order G^*-functionals.

5.3.2 Approximate Orthogonality Error Due to the Finite Record Length: (ζ-error)

The estimation of the G^*-functionals through the crosscorrelation technique is based upon the orthogonality of these functionals, expressed as:

$$E[G_m^* G_n^*] = 0 \qquad (m \neq n) \qquad (5.3.6)$$

However, the statistical average is substituted, in practice, by a time average over a finite time period. This has as a result the incomplete formation of this average, and subsequently some estimation error occurs in terms of remnants from the incomplete implementation of the orthogonality.

For example, the estimation of the n-th order kernel involves the crosscorrelation:

$$\hat{\psi}_n(\sigma_1, \ldots, \sigma_n) = \frac{1}{T}\int_0^T y(t)x(t-\sigma_1)\ldots x(t-\sigma_n)dt$$

$$= \sum_{m=0}^{\infty} \frac{1}{T}\int_0^T G_m^*[g_m; x(t'), t' \leq t]\, x(t-\sigma_1)\ldots x(t-\sigma_n)dt \qquad (5.3.7)$$

And the remnants of the incomplete orthogonality are of the form:

$$r_{n,m}(\sigma_1 \ldots, \sigma_n) = \frac{1}{T}\int_0^T G_m^*[g_m; x(t'), t' \leq t]\cdot x(t-\sigma_1)\ldots$$

$$x(t-\sigma_n)dt \quad (m \neq n) \qquad (5.3.8)$$

These remnants attain small nonzero values, which appear as estimation errors in our n-th order kernel estimate. Notice that, because of the use of the residual response in the crosscorrelation technique (cf. eqn. 4.3.5), the remnants $r_{n,m}$ with $m < n$ are very much smaller than the ones with $m > n$.

For illustrative purposes, let us study the ζ-error contribution of the leading integral term of G_m^*:

$$r'_{n,m}(\sigma_1, \ldots, \sigma_n) = \frac{1}{T}\int_0^T dt \int_0^\infty \ldots \int_0^\infty g_m(\tau_1, \ldots, \tau_m)x(t-\tau_1)\ldots x(t-\tau_m)$$

$$x(t-\sigma_1)\ldots x(t-\sigma_n)\, d\tau_1 \ldots d\tau_m$$

$$= \int_0^\infty \ldots \int_0^\infty g_m(\tau_1, \ldots, \tau_m)q_{n,m}(\tau_1, \ldots, \tau_m, \sigma_1, \ldots \sigma_n)\, d\tau_1 \ldots d\tau_m \qquad (5.3.9)$$

where,

$$q_{n,m}(\tau_1,\ldots,\tau_m,\sigma_1,\ldots,\sigma_n) = \frac{1}{T}\int x(t-\tau_1)\ldots x(t-\tau_m)x(t-\sigma_1)\ldots x(t-\sigma_n)dt \qquad (5.3.10)$$

Clearly, $q_{n,m}$ is an (m+n)-th order autocorrelation estimate, and as such, it has zero expected value and variance which depends on the order of the point $(\tau_1,\ldots,\tau_m,\sigma_1,\ldots,\sigma_n)$. In any case, this variance is inversely proportional to the record length and it becomes practically negligible for sufficiently long records.

It is very difficult to derive general and explicit expressions for the ζ-error commited in the estimation of the n-th order kernel, which would hold for any CSRS; because of the complexity of description of the higher order autocorrelation functions in the general case of an arbitrary CSRS. However, a basic qualitative remark can be made: the mean square ζ-error depends upon the even order moments of the CSRS test signal and the integrals of the squares of the system kernels.

5.4 The Erroneous Power Levels Error

As it was discussed in sec. 4.3, the normalizing factors of the crosscorrelation estimates depend upon the even moments and the step length of the respective CSRS or, in other words, upon the generalized power levels of the CSRS (cf. eqn. 4.3.23). Consequently, the evaluation of these normalizing factors requires the knowledge of the appropriate even moments and of the step length.

This knowledge is either *a priori* possessed or it must be

obtained through estimation. In the case where it is "a priori" and accurately possessed (for example, in computer simulations where all the parameters are accurately known), there is no additional estimation error caused by the normalization procedure. However, in a lot of actual applications in the laboratory, some test parameters are not accurately known and they must be measured or estimated.

In that case, measurement or estimation errors in these parameters may induce a disproportionality estimation error in the kernel estimates due to the erroneous normalizing factors. An additional error may result from the fact that the G^*-functionals of order higher than the first depend explicity upon the generalized power levels, which are possibly erroneously estimated.

To illustrate these two kinds of error, consider the case where we are trying to estimate the first order power level:

$$P_1 = m_2 \cdot \Delta t \qquad (5.4.1)$$

in the laboratory. The step length Δt can usually be measured with high accuracy, because it is equal to the distance between two successive points of derivative discontinuity in the second order autocorrelation function estimate. The measurement of the second moment m_2 usually involves more inaccuracies, since an estimate of it, obtained as the mean square of the signal samples, exhibits a variance of significant magnitude. More specifically:

$$\hat{m}_2 = \frac{1}{N} \sum_{i=1}^{N} x_i^2 \quad ; \quad N = T/\Delta t \qquad (5.4.2)$$

$$E[\hat{m}_2] = m_2 \qquad (5.4.3)$$

$$\text{Var}\left[\hat{m}_2\right] = \frac{\Delta t}{T}(m_4 - m_2^2) \tag{5.4.4}$$

and $\hat{m}_2$ follows approximately a gaussian distribution. Clearly, the accuracy of this estimate increases as the number of steps in the signal increases.

Another popular way of estimation of the first order power level is by computing the integral of the second order autocorrelation function estimate from $\tau=-\mu_1$ to $\tau=+\mu_1$, where μ_1 is the first order memory extent of the system. In this case the power level estimate is approximately a gaussian random variable with: (cf. eqns. 5.2.36, 5.2.41, 5.2.42)

$$E\left[\hat{P}_1\right] = m_2 \cdot \Delta t \tag{5.4.5}$$

$$\text{Var}\left[\hat{P}_1\right] \cong (m_2 \cdot \Delta t)^2 \frac{4\mu_1}{T} \tag{5.4.6}$$

Notice that the accuracy of the estimate improves as the record length increases.

In the first method of estimation the expected percentage error is:

$$\epsilon_1 = \sqrt{\frac{\Delta t}{T}\left(\frac{m_4}{m_2^2} - 1\right)} \tag{5.4.7}$$

where, we assume that Δt is measured with complete accuracy.

In the second method of estimation the expected percentage error is:

$$\epsilon_2 = 2\sqrt{\frac{\mu_1}{T}} \tag{5.4.8}$$

The disproportionality percentage error π_n in the n-th order kernel estimate (for the nondiagonal points), caused by a percentage

error ϵ in the estimation of the first order power level is:

$$\pi_n = \frac{(1+\epsilon)^n - 1}{(1+\epsilon)^n} \tag{5.4.9}$$

and, if $n\epsilon << 1$ (which is usually the case in practice), then:

$$\pi_n \cong n\epsilon \tag{5.4.10}$$

The effect of the erroneous estimation of the power levels upon the leading integral terms of the G^*-functionals is apparently (percentage-wise) exactly the same with the disproportionality error. However, the effect upon the following integral terms is smaller, because of the presence of the power levels in the coefficients of these terms.

5.5 The Finite Transition Time Error

The generation of the stimulus signal in the laboratory comprises two basic steps. In the first step, a string of proper numbers corresponding to the values of the CSRS at each step is generated by the digital computer. In the second step, this sequence of numbers is fed into a digital-to-analog transducer, which, in turn, generates an analog continuous signal in the physical dimensions required for the actual test of the system under study.

Expectedly, the response (or rising) time of the transducer will be finite, because of some inevitable inertia of the driving system. Of course, this finite response time can be reduced down to a desirable degree by using the proper piece of hardware. The practical requirements concerning the response time of the transducer are set by the frequency bandwidth of the system under test.

In any case, the finite response time results in finite transition time at the switching points of the signal. It can be stated right from the beginning that the practical effect of this finite transition time is negligible if the inverse of the transition time is greater than the frequency bandwidth of the system under test. This is a practical requirement, which (in most of the cases) can be easily met, if we design and use the proper piece of hardware.

Nevertheless, we will study a little more this effect for two general cases of practical interest:

(1) the reversible transitions

(2) the irreversible transitions

The reversibility of a transition is defined according to the patterns followed in an upwards or a downwards transition. If those patterns are the same the transition is called reversible, otherwise it is called irreversible.

Consider now the ideal (i.e. zero rising time) CSRS stimulus $x(t)$ and the actual output of a transducer $x^*(t)$, which has finite reversible transition times. Consider their difference as being an error signal: (Fig. 5.5.1)

$$\epsilon(t) = x(t) - x^*(t) \qquad (5.5.1)$$

Since the transition is assumed reversible, the operation of the transducer is described by the same differential equation both in upwards and downwards movements. This differential equation is usually linear with constant coefficients, and with all its roots on the negative

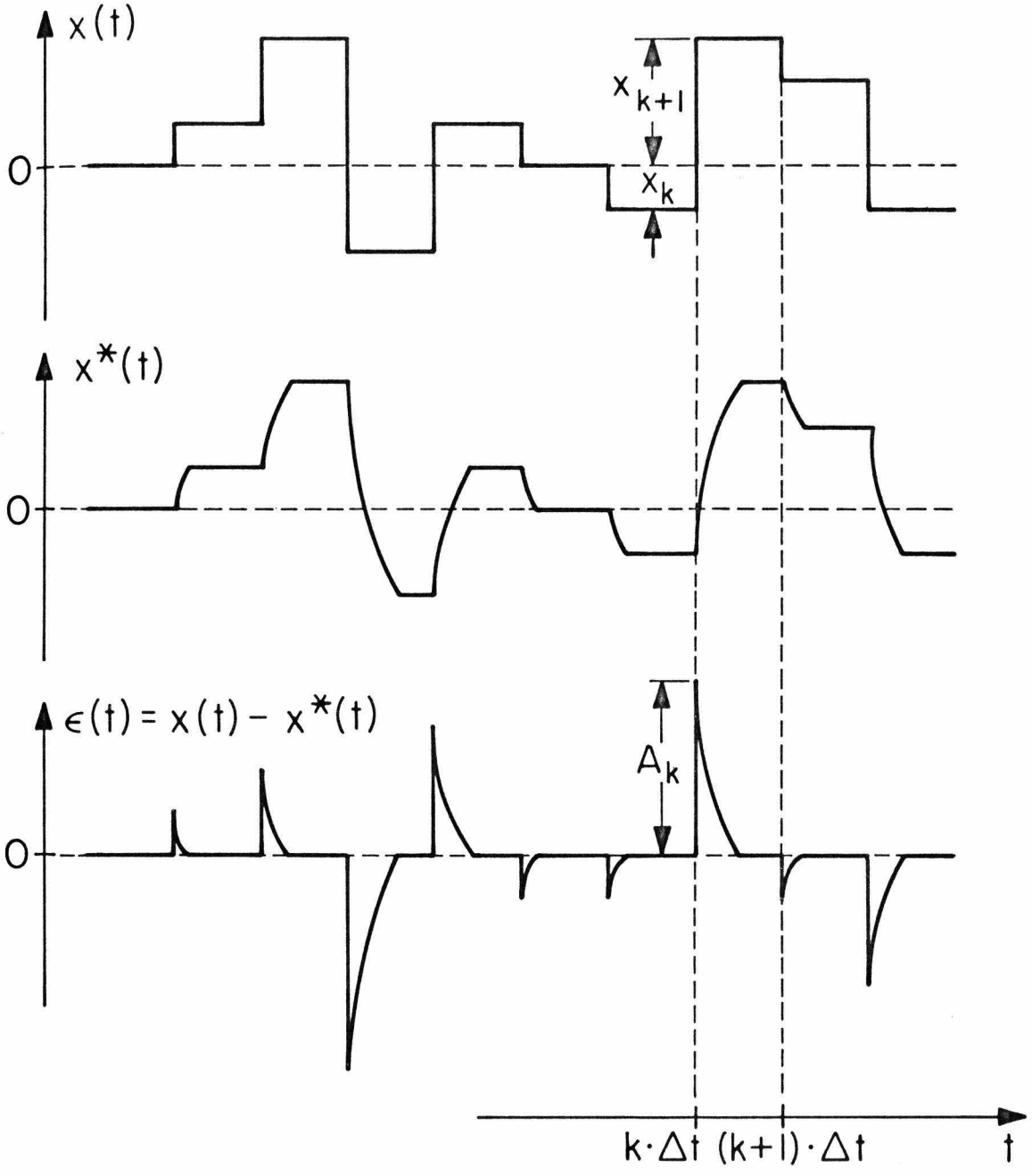


Fig. 5.5.1: Illustration of the error signal resulting from finite reversible transition time

real axis.

For illustrative purposes, let us consider the very simple case where the operation of the transducer is described by the first order linear differential equation:

$$T_r \frac{dx^*(t)}{dt} + x^*(t) = x(t) \qquad \text{for } k\Delta t \leq t < (k+1)\Delta t \qquad (5.5.2)$$

and T_r is the characteristic response time of the transducer, chosen to be much smaller than the step length Δt. The solution of the differential equation (5.5.2) is, for every step of the CSRS signal x(t):

$$x^*(t) = [x_k - x_{k+1}]e^{-(t-k\Delta t)/T_r} + x_{k+1} \text{ ; for } k\Delta t \leq t < (k+1)\Delta t \qquad (5.5.3)$$

where, x_k is the constant value of x(t) for $(k-1)\Delta t < t \leq k\Delta t$. Notice that, since $T_r << \Delta t$, the error signal $\epsilon(t)$ is practically zero for most of the time.

Let us study the error signal:

$$\epsilon(t) = [x_{k+1} - x_k]e^{-(t-k\Delta t)/T_r}$$

$$= A_k \cdot e^{-(t-k\Delta t)/T_r} \quad ; \text{ for } k\Delta t \leq t < (k+1)\Delta t \qquad (5.5.4)$$

Clearly, $\epsilon(t)$ is a sequence of random pulse-like exponentials (see Fig. 5.5.1). Their random character is attributed to their random height A_k. From the generation mechanism of the CSRS family (namely, the statistical independence of each step and the symmetric amplitude probability density function) it follows directly that the random amplitude A_k is distributed according to a symmetric probability density function q(A), which is the convolution integral of the amplitude probability density function p(x) of the CSRS with itself. The mean of q(A) is zero and the variance

is double the variance of p(x).

This brief analysis of the error signal $\epsilon(t)$ will enable us to study its autocorrelation properties, as well as its crosscorrelation with the ideal CSRS x(t). The study of these properties is necessary for the appreciation of the effect of the finite transition time upon our kernel estimates, since we have to show that the quasi-whiteness is preserved in $x^*(t)$ to an acceptable degree. We have:

$$\phi_n^*(\tau_1, \ldots, \tau_n) = E[x^*(t-\tau_1) \cdot \ldots \cdot x^*(t-\tau_n)]$$
$$= \sum_{k=o}^{n} \binom{n}{k} \xi_{k,n}(\tau_1, \ldots, \tau_n) \qquad (5.5.5)$$

where,

$$\xi_{k,n}(\tau_1, \ldots, \tau_n) = E[x(t-\tau_1) \cdot \ldots \cdot x(t-\tau_k) \cdot \epsilon(t-\tau_{k+1}) \cdot \ldots \cdot \epsilon(t-\tau_n)] \qquad (5.5.6)$$

At first, let us study the second order autocorrelation function of the error signal:

$$\phi_{\epsilon\epsilon}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \epsilon(t)\epsilon(t-\tau)dt \qquad (5.5.7)$$

With the help of eqn. 5.5.4 we derive the values of $\phi_{\epsilon\epsilon}(\tau)$: (see Fig. 5.5.2)

$$\epsilon\epsilon(\tau) = \begin{cases} \frac{m_2 \cdot T_r}{\Delta t} \cdot \left\{ e^{-\tau/T_r} - e^{-\Delta t/T_r} \cdot \left[e^{-(\Delta t-\tau)/T_r} + \sinh\left(\frac{\tau}{T_r}\right) \right] \right\} & \text{for } o \leq |\tau| \leq \Delta t \\ \frac{m_2 \cdot T_r}{2\Delta t} \cdot e^{-(\tau-\Delta t)/T_r} \left[1 - e^{-2(2\Delta t-\tau)/T_r} \right] & \text{for } \Delta t \leq |\tau| \leq 2\Delta t \\ \simeq o & \text{for } |\tau| > 2\Delta t \end{cases} \qquad (5.5.8)$$

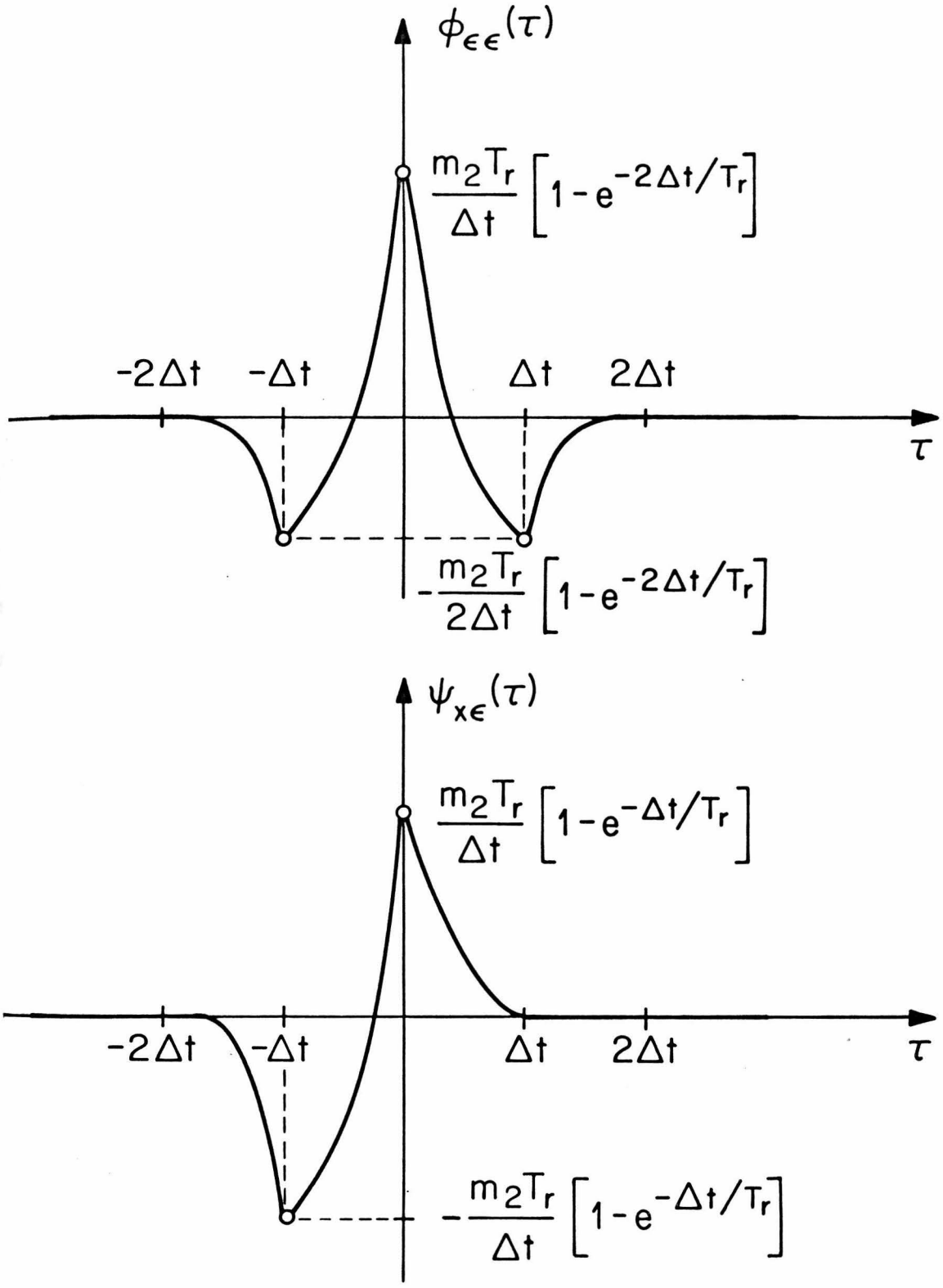


Fig. 5.5.2: Second order autocorrelation function of the error signal

Fig. 5.5.3: Crosscorrelation function of the ideal and the error signal

Notice that:

$$-\frac{m_2 T_r}{2\Delta t} < \phi_{\epsilon\epsilon}(\tau) < \frac{m_2 T_r}{\Delta t} \quad ; \text{ for any } \tau \tag{5.5.9}$$

Consequently, the values of this autocorrelation function become negligible if the response time T_r of the transducer is chosen much smaller than the step length Δt.

Similarly, we derive the expressions for the crosscorrelation function:

$$\psi_{x\epsilon}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)\epsilon(t-\tau)dt \tag{5.5.10}$$

using eqn. 5.5.4 (see Fig. 5.5.3), in terms of the second moment of the CSRS m_2, the step length Δt and transducer response time T_r:

$$\psi_{x\epsilon}(\tau) = \begin{cases} \frac{m_2 T_r}{\Delta t}\left[1-e^{-(\Delta t-\tau)/T_r}\right] & \text{for } o \leq \tau \leq \Delta t \\ \frac{m_2 \cdot T_r}{\Delta t}\left[2e^{\tau/T_r} - 1 - e^{-\Delta t/T_r}\right] & \text{for } -\Delta t \leq \tau \leq o \\ \frac{m_2 \cdot T_r}{\Delta t}\left[e^{-\Delta t/T_r} - e^{(\tau+\Delta t)/T_r}\right] & \text{for } -2\Delta t \leq \tau \leq \Delta t \\ \simeq o & \text{for } \tau < -2\Delta t \text{ and } \tau > \Delta t. \end{cases} \tag{5.5.11}$$

We note again that:

$$-\frac{m_2 \cdot T_r}{\Delta t} < \psi_{x\epsilon}(\tau) < \frac{m_2 \cdot T_r}{\Delta t} \quad ; \quad \text{for any } \tau \tag{5.5.12}$$

which implies that the values of this crosscorrelation become negligible, if T_r is chosen to be much smaller than Δt.

From the study of the autocorrelation and crosscorrelation properties of $\epsilon(t)$ and $x(t)$, we derive that the functions $\xi_{k,n}(\tau_1, \ldots \tau_n)$

of eqn. 5.5.5 do not deviate significantly from the quasi-whiteness, since their nonzero values remain concentrated within diagonal strips with width of the order of magnitude of Δt (see Figs. 5.2.2 and 5.2.3). In addition to that, it became evident from expressions (5.5.9) and (5.5.12) that the values of the functions $\xi_{k,n}(\tau_1, \ldots, \tau_n)$, for $k < n$, are much smaller than the values of the function $\xi_{n,n}(\tau_1, \ldots, \tau_n)$, if T_r is chosen to be much smaller than Δt.

Thus, if we choose T_r to be much smaller than Δt, then the functions $\xi_{k,n}(\tau_1, \ldots, \tau_n)$, for $k < n$, become practically negligible as compared to $\xi_{n,n}(\tau_1, \ldots, \tau_n)$. Consequently, eqn. 5.5.5 reduces approximately to:

$$\phi_n^*(\tau_1, \ldots, \tau_n) \cong \xi_{n,n}(\tau_1, \ldots, \tau_n) = \phi_n(\tau_1, \ldots, \tau_n)$$

Notice that even if the functions $\xi_{k,n}$ (for $k < n$) are not negligible the quasi-whiteness of $x^*(t)$ is not seriously affected. The only thing that is probably affected in this case is the value of the autocorrelation function at the full-diagonal points. For this reason, special care must be taken in order to use the correct normalizing factors; since they are crucially depending on the values of the autocorrelation functions at the full-diagonal points.

In the case of an irreversible transition the differential equation which describes the operation of the transducer is not the same for upwards and downwards transitions. Apparently, this complicates the analytical derivations to a considerable degree. For this reason, we will not attempt to derive analytical expressions, but we will confine

ourselves to some qualitative remarks: (1) The significant (note that we do not use in this case the word "nonzero", that we used for the reversible transitions) values of the autocorrelation functions of the error signal will be concentrated within diagonal strips with width of the order of magnitude of Δt. The same is true for the crosscorrelation functions of the CSRS and the error signal. (2) The values of these autocorrelation and crosscorrelation functions can become practically negligible, if both the response times (of upwards and downwards transitions) are chosen to be much smaller than Δt.

The validity of these two remarks assures the preservation of the quasi-whiteness of $x^*(t)$ to a certain degree, which is definitely lower than in the reversible transitions case.

Notice that in the case of irreversible transitions, the expected value of the error signal (as well as all its odd order moments) is not zero; while in the case of reversible transitions, it is zero (along with all its odd order moments). This is a clear indication of the fact that the case of irreversible transitions is disadvantageous, because it causes more significant deviations from the quasi-whiteness than the case of reversible transitions.

Therefore, it is highly advisable that we pursue in our applications to have a transducer with reversible transitions.

5.6 Computational Errors

There are three main sources of computational errors in connection with the crosscorrelation technique. These sources are:

(1) Discrete representation of continuous data

(2) Discrete integration in computing the crosscorrelations

(3) Numerical round-off errors

From these categories of computational errors, we will further discuss the first two as having some effect of practical interest; while the effect of the last category is, for all practical purposes, negligible, for the kind of digital computer that has been used in the present applications (PDP 11/45).

5.6.1 Discrete Representation of Continuous Data

The input and output signals, that we deal with in the present study, are continuous in time. However, the handling of any signal, as a dataset within a digital computer, requires the "digitalization" (or "discretization") of the signal, with which the continuous signal becomes a string of numbers. Thus, a continuous signal has to be sampled, at some proper sampling rate, and be represented thereafter within the digital computer as a discrete time series. This operation of "discretization" involves several pitfalls, which have received considerable attention in the literature.

The most important of these pitfalls is the "aliasing problem". This problem is the natural consequence of the simple mathematical fact, that there is an infinite number of sinusoidal curves (of different frequencies) which pass through a set of equidistant points of a line, when the fundamental distance between any two points is finite. (Fig. 5.6.1)

It is evident, that the several frequencies of these sinusoidal curves are integral multiples of a fundamental frequency:

$$f_o = {}^{1}/_{2d} \qquad (5.6.1)$$

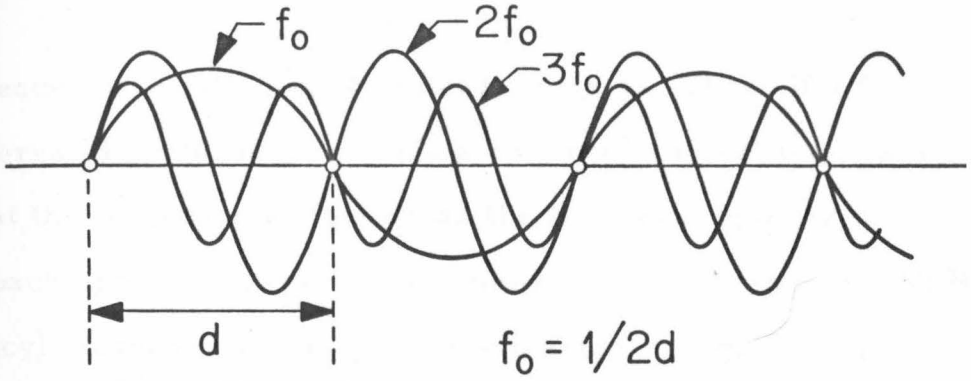


Fig. 5.6.1: Illustration of the aliasing problem

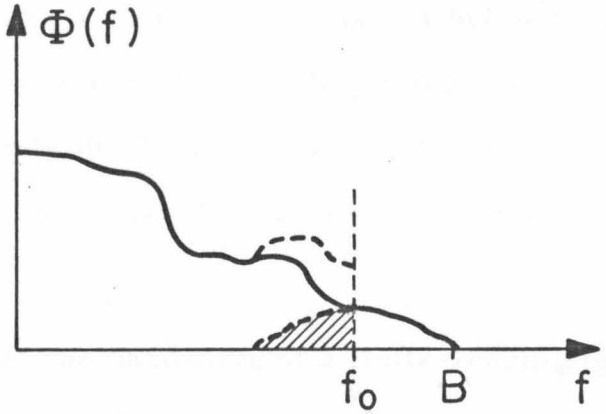


Fig. 5.6.2: Power spectrum folding due to aliasing

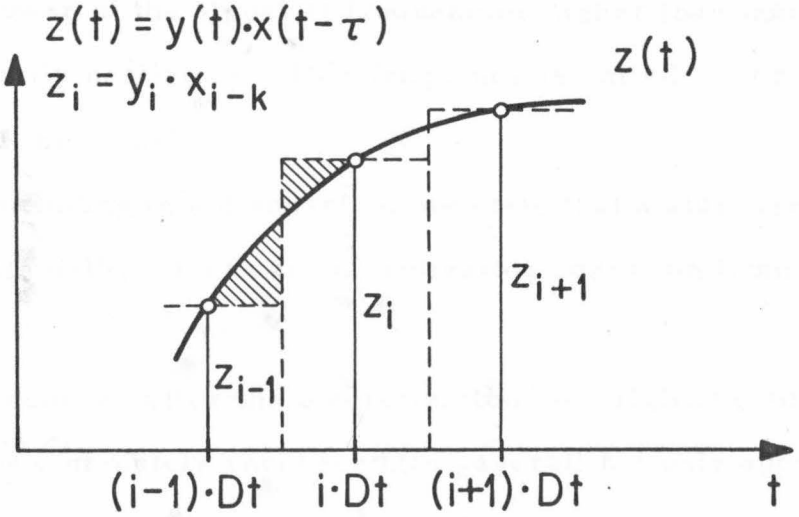


Fig. 5.6.3: Illustration of the deviation between continuous integration and discrete summation

This frequency is called the Nyquist frequency. Thus, if a finite sampling interval d is chosen to sample a continuous signal the power of the signal at the frequencies higher than the Nyquist frequency will be folded back upon the power of the symmetric (with respect to the Nyquist frequency) frequencies of the power spectrum. (Fig. 5.6.2)

This very well known and studied fact constitutes the main pitfall of discretizing a continuous signal. Obviously, if the sampling rate f_s (i.e. the inverse of the sampling interval) is chosen to be greater than double the bandwidth B of the signal, then the aliasing effect is eliminated. Thus, we must always use in practice:

$$f_s \geq 2B \qquad (5.6.2)$$

Of course, the bandwidth of a finite-length signal is not strictly finite, but there can be always determined a frequency for which the aggregate power of the signal at frequencies higher than that is considered practically negligible. This frequency is called, in practice, the bandwidth of the signal.

Concluding this discussion, we state that a sufficiently high sampling rate will protect our discretization operation from the aliasing effect.

Of course, after the discretization the original continuous signal cannot be completely recovered (in general) but only approximated through some interpolation scheme. It is evident that in the case of a CSRS the full and complete recovery of the original continuous signal from the discrete dataset is possible. This is due to the special stair-like form of the CSRS. Nevertheless, the aliasing problem can be

severe, even in the case of a CSRS, if the sampling rate is not sufficiently high.

5.6.2 Discrete Integration Error

The computation of a crosscorrelation for the estimation of a kernel involves time integration of the product of the output signal with the proper number of time-shifted versions of the input signal:

$$\hat{\psi}_n(\sigma_1, \ldots, \sigma_n) = \frac{1}{T} \int_0^T y(t)x(t-\sigma_1) \cdot \ldots \cdot x(t-\sigma_n)dt \qquad (5.6.3)$$

where T is the record length (taken constant in this case).

For the reasons that we explained previously, the input and output signals are available within the computer only as discrete datasets. Therefore, the integration can never be completely accurate. However, we can conceivably achieve almost any specified degree of accuracy by using the proper interpolation-integration numerical procedure. A prerequisite, of course, for this is that the discrete representation of the continuous signals is practically free of the aliasing effect, and the accuracy capacity (word length, floating-point processor etc.) of the digital computer that is used conforms with the pursued degree of overall accuracy.

In practical applications of the crosscorrelation technique the computational burden is, by the method itself, very heavy, and consequently we try to keep the numerical procedures that are used as simple and efficient as possible. Thus, in the trade off between computational accuracy and overall efficiency of the method, the balancing point is usually found on the side of simple numerical procedures with reason-

able (but not the best possible) computational accuracy.

The simplest integration procedure for the computation of the crosscorrelation with the discrete input and output datasets $\{y_i\}$ and $\{x_i\}$ is to compute the summation: (consider the first order case as an example)

$$\hat{\psi}_1(k \cdot Dt) = \frac{1}{N} \sum_{i=1}^{N} y_i \cdot x_{i-k} \qquad (5.6.4)$$

where, Dt is the sampling interval.

We will now evaluate the difference of the integral portions between $(i-1) \cdot Dt$ and $i \cdot Dt$, obtained by the analytical and the numerical methods. (Fig. 5.6.3) The integral portion obtained by the simple numerical summation is:

$$I_i = (z_{i-1} + z_i) \frac{Dt}{2} \qquad (5.6.5)$$

The integral portion obtained analytically is:

$$J_i = \int_{(i-1)Dt}^{iDt} z(t)dt$$

$$= \int_{(i-1)Dt}^{(i-\frac{1}{2})Dt} z_I(t)dt + \int_{(i-\frac{1}{2})Dt}^{iDt} Z_{II}(t)dt \quad = L_1 + L_2 \qquad (5.6.6)$$

where $z_I(t)$ and $z_{II}(t)$ are two Taylor expansions of $z(t)$ about z_{i-1} and z_i respectively:

$$z_I(t) = z_{i-1} + z_{i-1}^{(1)}(z-z_{i-1}) + \frac{z_{i-1}^{(2)}}{2}(z-z_{i-1})^2 + \frac{z_{i-1}^{(3)}}{3!}(z-z_{i-1})^3 + \cdots \qquad (5.6.7)$$

$$z_{II}(t) = z_i + z_i^{(1)}(z-z_i) + \frac{z_i^{(2)}}{2}(z-z_i)^2 + \frac{z_i^{(3)}}{3!}(z-z_i)^3 + \ldots \qquad (5.6.8)$$

thus,

$$L_1 = \sum_{n=o}^{\infty} \frac{z_{i-1}^{(n)}}{(n+1)!} (Dt/2)^{n+1} \qquad (5.6.9)$$

$$L_2 = \sum_{n=o}^{\infty} (-1)^n \frac{z_i^{(n)}}{(n+1)!} (Dt/2)^{n+1} \qquad (5.6.10)$$

Consequently, the difference between the numerical and the analytical outcome of the i-th integral portion is:

$$D_i = J_i - I_i$$

$$= \sum_{n=1}^{\infty} \frac{\left[z_{i-1}^{(n)} + (-1)^n z_i^{(n)}\right]}{(n+1)!} \left(\frac{Dt}{2}\right)^{n+1} \qquad (5.6.11)$$

Now, consider the total difference between the two integrals: (N is the number of samples in the dataset)

$$D = \sum_{i=1}^{N} D_i \qquad ; \quad N = T/Dt$$

$$\cong \sum_{m=1}^{\infty} \sum_{i=1}^{N} \frac{2z_i^{(2m)}}{(2m+1)} \left(\frac{Dt}{2}\right)^{2m+1} \qquad (5.6.12)$$

and,

$$z_i^{(2m)} = \sum_{j=o}^{2m} \binom{2m}{j} y_i^{(j)} x_{i-k}^{(2m-j)} \qquad (5.6.13)$$

Clearly, the exact value of $z_i^{(2m)}$ (and consequently the value of D) depends on the relative values of the derivatives of $y(t)$ and $x(t-\tau)$ at the sampling points, and the only general statement that can be made is that the difference D decreases very fast as Dt decreases.

If $x(t)$ is band-limited gaussian white noise then the values of

its derivatives at the sampling points are usually significant and, of course, they depend mainly on the bandwidth of the signal. On the other hand, the values of the derivatives of y(t) depend on the bandwidths of both the input x(t) and the system under test.

Notice that if x(t) is a CSRS all its derivatives at the sampling points are zero! Consequently, the expression (5.6.13) for $z_i^{(2m)}$ is dramatically simplified:

$$z_i^{(2m)} = y_i^{(2m)} \cdot x_{i-k} \qquad (5.6.14)$$

and, in most of the cases, it is expected to be much smaller than in the case of gaussian white noise. One more reason for this, is that the values of the derivatives of y(t) are generally much smaller than the ones of x(t), because the bandwidth of x(t) is usually much bigger than the one of y(t).

For all these reasons, the computational error resulting from the numerical integration is expected to be smaller in the case of x(t) being a CSRS rather than GWN.

In most of the cases, the difference D (and consequently the computational error) is approximately proportional to the square of the sampling interval; since, the sampling interval Dt is always fairly small (for aliasing reasons) and therefore:

$$D \cong \sum_{i=1}^{N} \frac{z_i^{(2)}}{12} \cdot Dt^3 \cong Dt^2 \cdot \left(Dt \sum_{i=1}^{N} \frac{z_i^{(2)}}{12} \right) \cong Dt^2 \cdot \text{const} \qquad (5.6.15)$$

because, the number of samples N is inversely proportional to Dt (for a constant record length) and the values of $z_i^{(2)}$ at closely neighboring

points are only slightly different.

In conclusion, the basic factor affecting the numerical integration error is the sampling interval. For sufficiently small sampling interval, this error becomes usually negligible for all practical purposes.

CHAPTER VI

ERROR MANAGEMENT AND DESIGN OF THE OPTIMUM TEST

6.1 General Error Management

In chapter 5, we studied the several kinds of estimation error in the use of CSRS for nonlinear system identification in connection with the crosscorrelation technique. In this chapter, we will discuss the optimum strategy which ought to be followed in practical applications, in order to achieve the highest possible accuracy in our kernel estimates.

In the previous chapter, we diagnosed eight main sources of estimation error. These are:

(1) The deconvolution error (θ-error)

(2) The statistical fluctuation error (ϵ-error)

(3) The approximate orthogonality error due to the finite stimulus bandwidth (ξ-error)

(4) The approximate orthogonality error due to the finite record length (ζ-error)

(5) The erroneous power levels error (π-error)

(6) The finite transition time error

(7) The continuous signals discretization error

(8) The discrete integration numerical error

From all these categories of error, some can be neglected as insignificant in practical applications, some can be corrected, and the rest, which cannot be either corrected or neglected, can be optimized with respect to the controllable test parameters.

Evidently, some errors can be neglected only if the controll-

able test parameters attain values within a proper range. More specifically:

(a) The discrete integration error can be neglected, if the specified sampling interval Dt is sufficiently small. (cf. sec. 5.6.2)

(b) The discretization error can be neglected, if Dt is smaller than 1/2B, where B is the bandwidth of the system under study. (cf. sec. 5.6.1)

(c) The finite transition time error is similar to the erroneous power levels error (cf. sec. 5.5). This error can be corrected to a great extent, provided that the transitions are reversible, by computing the actual power levels corresponding to the imperfect experimental stimulus signal.

(d) The erroneous power levels error, in the cases that it occurs (cf. sec. 5.4), can be largely corrected by employing a final scaling procedure, which is presented in sec. 6.3.

(e) The approximate orthogonality error due to the finite record length (ζ-error) can be usually neglected if the record length is reasonably long. However, if the record length happens to be fairly short (a case which is unlikely in practice), a final correctional procedure can be followed on the basis of the analysis made in sec. 5.3.2 to compensate for part (at least) of the error commited. This procedure is expected to be quite cumbersome and inefficient with respect to the positive return that it brings, therefore, it is hardly suggested in practice.

(f) The approximate orthogonality error due to the finite stimulus bandwidth (ξ-error) can be either corrected, on the basis of the analysis made in sec. 5.3.1, or neglected, if the step length Δt is sufficiently small. It must be noted that, like in the previous case of the ζ-error, this correctional procedure is expected to be cumbersome and inefficient with respect to the achieved positive return.

(g) Finally, the statistical fluctuation error (ϵ-error) as well as the deconvolution error (θ-error) can neither be neglected (in most cases) nor satisfactorily corrected. Of course, the θ-error can be somewhat corrected, on the basis of the analysis made in sec. 5.1, by estimating the even order derivatives of the kernels from the original

kernel estimates, however, a short preliminary study of the numerical problems involved pointed to the limited efficiency of this correctional procedure. Also, the ϵ-error can be somewhat corrected by averaging several sample kernel estimates, however the efficiency of this procedure is again low.

Because of these reasons, the ϵ and θ errors mount to the principal types of estimation error, which have to be optimized with respect to the controllable test parameters (namely, the record length and the step length of the CSRS stimulus). This optimization procedure is illustrated in the following section, and it is based upon the analysis made in secs. 5.1 and 5.2.

In conclusion, the basic test parameters (i.e. the record length T, the step length Δt and the sampling interval DT) are usually such that the most of the estimation errors diagnosed in chapter 5 can be neglected, for all practical purposes, in actual applications. The only types of estimation error that usually retain significant size are the ϵ and θ errors. Therefore, the optimization of the controllable test parameters Δt and T, which influence the ϵ and θ errors (cf. secs. 5.1 and 5.2), in order to minimize the combined effect of these two estimation errors becomes the principal target of the error management in the case of the CSRS.

6.2 Optimization of ϵ and θ Errors - The Fundamental Error Equation

As it was discussed in sec. 5.1, the θ-error is deterministic in nature and it is due to the finite stimulus bandwidth. It amounts to some loss of high frequencies in our kernel estimates.

The explicit expression of this error for the n-th order kernel estimate is: (the initial region is omitted)

$$\theta_n(\sigma_1, \ldots, \sigma_n) = \Delta t^2 \cdot D_{h_n}(\Delta t, \sigma_1, \ldots \sigma_n) \tag{6.2.1}$$

where,

$$D_{h_n}(\Delta t, \sigma_1, \ldots \sigma_n) = \sum_{m=1}^{\infty} \frac{\Delta t^{2m-2}}{2m!(2m+1)(m+1)} \sum_{\substack{j_1, \ldots, j_{2m}=1 \\ j_1 \leq \ldots \leq j_{2m}}}^{n} \frac{\partial^{2m} h_n(\sigma_1, \ldots, \sigma_n)}{\partial \sigma_{j1} \ldots \partial \sigma_{j_{2m}}} \tag{6.2.2}$$

Clearly, D_{h_n} depends very little on Δt, since Δt attains in practice fairly small values, and consequently:

$$D_{h_n}(\sigma_1, \ldots, \sigma_n) \cong \frac{1}{12} \sum_{\substack{j_1, j_2=1 \\ j_1 \leq j_2}}^{n} \frac{\partial^2 h_n(\sigma_1, \ldots, \sigma_n)}{\partial \sigma_{j_1} \partial \sigma_{j_2}} \tag{6.2.3}$$

Thus, the θ_n error depends, in first approximation, on Δt^2 and the second partial derivatives of the kernel h_n.

On the other hand, as it was discussed in sec. 5.2, the ϵ-error is random in nature and it is due to the finite record length. It amounts to random deviations of the kernel estimates, according to a gaussian probability law. The first two moments of the gaussian distribution of these deviations at the nodal points of the n-th order kernel are:

$$E[\epsilon_n(\sigma_1, \ldots, \sigma_n)] = o \tag{6.2.4}$$

$$\mathrm{Var}[\epsilon_n(\sigma_1, \ldots, \sigma_n)] \triangleq q_n^2(\sigma_1, \ldots, \sigma_n) = \frac{1}{T \cdot \Delta t^{n-1}} C_{h_n}^2(\Delta t, \sigma_1, \ldots, \sigma_n) \tag{6.2.5}$$

Again, it must be emphasized that $C_{h_n}^2$ depends very slightly on Δt, for Δt being in the range of values of practical interest.

Summing up these two errors, we end up with the following total estimation error of appreciable size in practical applications of the CSRS:

$$\delta_n(\sigma_1, \ldots, \sigma_n) = \theta_n(\sigma_1, \ldots, \sigma_n) + \epsilon_n(\sigma_1, \ldots, \sigma_n) \qquad (6.2.6)$$

Clearly, the total error δ_n is a gaussian random variable with:

$$E[\delta_n(\sigma_1, \ldots, \sigma_n)] = \theta_n(\sigma_1, \ldots, \sigma_n) \qquad (6.2.7)$$

$$\mathrm{Var}[\delta_n(\sigma_1, \ldots, \sigma_n)] = q_n^2(\sigma_1, \ldots, \sigma_n) \qquad (6.2.8)$$

Consequently, the mean square error of the n-th order kernel estimate is:

$$\begin{aligned} Q_n &= \int_0^{\mu_n} \cdots \int_0^{\mu_n} E[\delta_n^2(\sigma_1, \ldots, \sigma_n)]\, d\sigma_1 \cdots d\sigma_n \\ &= \int_0^{\mu_n} \cdots \int_0^{\mu_n} \theta_n^2(\sigma_1, \ldots, \sigma_n) d\sigma_1 \cdots d\sigma_n \\ &\qquad + \int_0^{\mu_n} \cdots \int_0^{\mu_n} q_n^2(\sigma_1, \ldots, \sigma_n) d\sigma_1 \cdots d\sigma_n \\ &= \Delta t^4 \cdot \int_0^{\mu_n} \cdots \int_0^{\mu_n} D_{h_n}^2(\Delta t, \sigma_1, \ldots, \sigma_n) d\sigma_1 \cdots d\sigma_n \\ &\qquad + \frac{1}{T \cdot \Delta t^{n-1}} \int_0^{\mu_n} \cdots \int_0^{\mu_n} C_{h_n}^2(\Delta t, \sigma_1, \ldots, \sigma_n) d\sigma_1 \cdots d\sigma_n \\ &= A_n(\Delta t) \cdot \Delta t^4 + \frac{B_n(\Delta t)}{T \cdot \Delta t^{n-1}} \end{aligned} \qquad (6.2.9)$$

In the applications, Δt is always much smaller than the kernel memory, and for this range of values A_n and B_n become approximately

independent from Δt, thus:

$$Q_n \cong A_n \cdot \Delta t^4 + \frac{B_n}{T \cdot \Delta t^{n-1}} \qquad (6.2.10)$$

Notice, however, that A_n and B_n depend on the specified kernel memory μ_n. The equation above will be called "fundamental error equation" (FEE) and it will constitute the principal instrument of the optimization procedure for Δt and T.

The function Q_n described by the FEE has always a single minimum for A_n and B_n positive (which is actually the case) and for a given T. (Fig. 6.2.1) This single minimum is the optimum test position, corresponding to the optimum Δt (for a given T).

The value of the optimum Δt can be easily determined analytically from FEE:

$$(\Delta t_{opt})_n = \left[\frac{(n-1)\, B_n}{4T\, A_n}\right]^{1/(n+3)} \qquad (6.2.11)$$

and the resulting optimum m. s. error is:

$$(Q_{opt})_n = A_n^{(n+2)/(n+3)} \cdot \left[\frac{(n-1)B_n}{4T}\right]^{1/(n+3)} + \left[\frac{4A_n}{n-1}\right]^{(n-1)/(n+3)} \cdot \left[\frac{B_n}{T}\right]^{4/(n+3)} \qquad (6.2.12)$$

Of course, the expressions above are accurate only if the optimum Δt is in the neighborhood of very small Δt where A_n and B_n can be considered approximately independent of Δt. Otherwise, the optimum Δt results as the solution in Δt of the equation:

$$\Delta t^{n+4} \cdot \frac{dA_n}{d\Delta t} + \frac{\Delta t}{T} \cdot \frac{dB_n}{d\Delta t} + 4\Delta t^{n+3} A_n - \frac{(n-1)}{T} B_n = o \quad (\Delta t \neq o) \qquad (6.2.13)$$

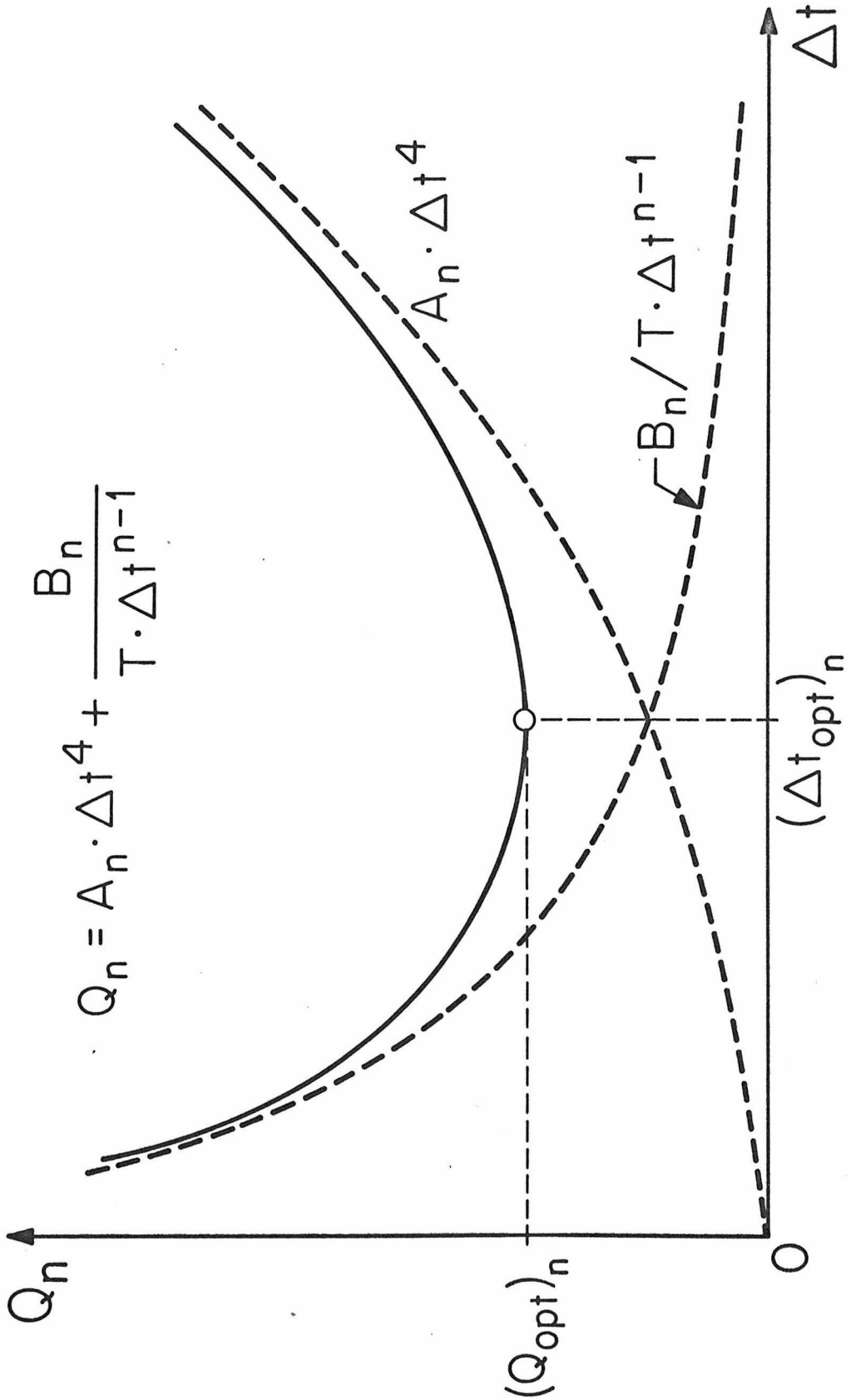


Fig. 6.2.1: Graphic representation of the fundamental error equation (FEE)
(Drawing for n=5)

However, we always anticipate, in practice, having such a range of Δt values that A_n and B_n are approximately constant with respect to Δt. In this case, it is evident from eqns. 6.2.11 and 6.2.12 that the optimum Δt and the corresponding Q_n are monotonically decreasing as T is increasing (Fig. 6.2.2). The continuous curve in Fig. 6.2.2 is the locus of the points $[(\Delta t_{opt})_n, (Q_{opt})_n]$ of the plane for changing T.

In the first order kernel case, the optimum Δt, as determined by eqn. 6.2.11, is zero. Of course, zero is not a realistic value, therefore Δt is determined in the first order case either by eqn. 6.2.13 or by signal-to-noise ratio considerations.

In the former case, eqn. 6.2.13 reduces for n=1 to:

$$\Delta t^4 \frac{dA_1}{d\Delta t} + 4\Delta t^3 \cdot A_1 + \frac{1}{T} \frac{dB_1}{d\Delta t} = o \qquad (6.2.14)$$

and since the optimum Δt will be found in the range of very small values, the solution of eqn. 6.2.14 in Δt will be very close to the one of the equation:

$$4\Delta t^3 A_1 + \frac{1}{T} \frac{dB_1}{d\Delta t} = o \qquad (6.2.15)$$

In the latter case, we note that the power level of any CSRS diminishes as Δt goes to zero, therefore in cases where external seriously contaminating noise is present, the optimum Δt is determined as the minimum Δt for which the stimulus power level assures an acceptable signal-to-noise ratio in the response of the system.

Thus, we see that in some cases factors relating to characteristics of the experiment also participate in the determination of the

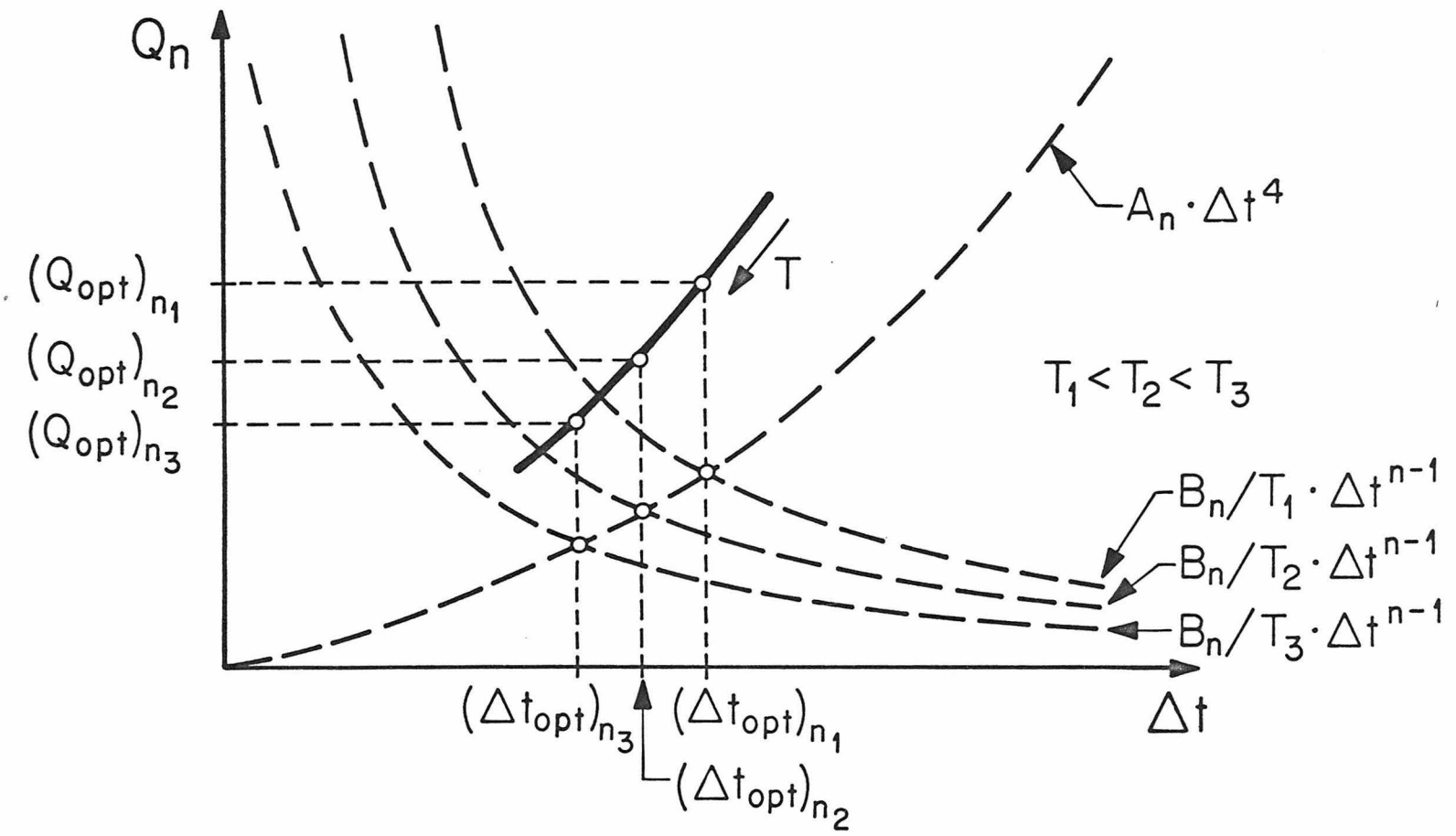


Fig. 6.2.2: Dependence of the optimum test position on the record length T (Drawing for n=5)

optimum test parameters. Apparently, one of the most important of these factors is the external noise. In most of the applications, the external noise is a serious source of error in our kernel estimates. The effect of the external noise will be discussed in sec. 6.4. In that respect also, the crosscorrelation identification method appears to be most advantageous.

Closing the discussion on the FEE, we want to point out that the optimum Δt is probably increasing as we move towards higher order kernel estimates.

6.3 Scaling Adjustment of Estimated Functional Terms

It is expected that the final estimated model of the system will predict the system response within a certain deviation. This deviation is due to all kinds of estimation errors that we discussed in the previous sections.

In this section, we will describe a simple mathematical procedure with which we can correct any substantial disproportionalities of the kernel estimates resulting mainly from erroneous power level and finite transition time errors. This mathematical procedure is simply an appropriate scalar weighting of the estimated G^*-functionals, so that the mean square error of the model predicted response is minimized. This is basically a least square fit procedure in the space of the G^*-functionals.

For example, let us consider the case of a second order CSRS model. From the crosscorrelation technique we have estimated the

CSRS kernels g_o, $g_1(\tau)$, $g_2(\tau_1, \tau_2)$, and now we want to minimize through a simple rescaling procedure the mean square error of the model predicted response $\hat{y}(t)$ with respect to the actual system response $y(t)$. Clearly:

$$\hat{y}(t) = \hat{G}_o^* + \hat{G}_1^*(t) + \hat{G}_2^*(t)$$

$$= \hat{g}_o + \int_o^\infty \hat{g}_1(\tau)x(t-\tau)d\tau + \int_o^\infty\int_o^\infty \hat{g}_2(\tau_1, \tau_2)x(t-\tau_1)x(t-\tau_2)$$

$$d\tau_1 d\tau_2 - P\int_o^\infty \hat{g}_2(\tau, \tau)d\tau \qquad (6.3.1)$$

We are looking for the scalars $\lambda_o, \lambda_1, \lambda_2$ for which:

$$q^2 = \int_o^T [y(t) - \lambda_o \hat{G}_o^* - \lambda_1 \hat{G}_1^*(t) - \lambda_2 \hat{G}_2^*(t)]^2 dt \qquad (6.3.2)$$

becomes minimum.

From the linear spaces theory we know that these values of $\lambda_o, \lambda_1, \lambda_2$ are determined by solving the matrix equation:

$$\begin{bmatrix} (\hat{G}_o^*, \hat{G}_o^*) & (\hat{G}_o^*, \hat{G}_1^*) & (\hat{G}_o^*, \hat{G}_2^*) \\ (\hat{G}_1^*, \hat{G}_o^*) & (\hat{G}_1^*, \hat{G}_1^*) & (\hat{G}_1^*, \hat{G}_2^*) \\ (\hat{G}_2^*, \hat{G}_o^*) & (\hat{G}_2^*, \hat{G}_1^*) & (\hat{G}_2^*, \hat{G}_2^*) \end{bmatrix} \begin{bmatrix} \lambda_o \\ \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} (\hat{G}_o^*, y) \\ (\hat{G}_1^*, y) \\ (\hat{G}_2^*, y) \end{bmatrix} \qquad (6.3.3)$$

where the symbol ( ) denotes inner product:

$$(f, g) = \int_o^T f(t)g(t)dt \qquad (6.3.4)$$

Evidently, if the several $\hat{G}_n^*$ are orthogonal, then the solution of eqn. 6.3.3 gives $\lambda_o = \lambda_1 = \lambda_2 = 1$. However, the several $\hat{G}_n^*$ are not exactly orthogonal because of the several estimation errors. Following this correctional procedure, we compensate for part of these errors and

primarily for the disproportionality errors (cf. chapter 5). This procedure aims at making the response residual orthogonal to the model response. According to the linear spaces theory, this is the best linear fit (in the mean square error sense) of the model response that we can get.

The matrix of eqn. 6.3.3 can be extended to the n-th order CSRS model case:

$$\begin{bmatrix} (\hat{G}_o^*, \hat{G}_o^*) & (\hat{G}_o^*, \hat{G}_1^*) & \dots & (\hat{G}_o^*, \hat{G}_n^*) \\ (\hat{G}_1^*, \hat{G}_o^*) & (\hat{G}_1^*, \hat{G}_1^*) & \dots & (\hat{G}_1^*, \hat{G}_n^*) \\ \vdots & & & \\ (\hat{G}_n^*, \hat{G}_o^*) & (\hat{G}_n^*, \hat{G}_1^*) & \dots & (\hat{G}_n^*, \hat{G}_n^*) \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix} = \begin{bmatrix} (\hat{G}_o^*, y) \\ (\hat{G}_1^*, y) \\ \vdots \\ (\hat{G}_n^*, y) \end{bmatrix} \tag{6.3.5}$$

and the final kernel estimates are:

$$\begin{cases} \hat{\hat{g}}_o = \lambda_o \hat{g}_o \\ \hat{\hat{g}}_1(\tau) = \lambda_1 \cdot \hat{g}_1(\tau) \\ \quad \vdots \\ \hat{\hat{g}}_n(\tau_1, \dots, \tau_n) = \lambda_n \cdot \hat{g}_n(\tau_1, \dots, \tau_n) \end{cases} \tag{6.3.6}$$

6.4 External Noise Effect

The presence of external noise in the performance of a real system is usually a serious concern, because of the errors that it may induce in the identification process. In some systems, e.g. the biological systems, the contaminating effect of the external noise is very sig-

nificant and special care is required in the identification of the system.

The Wiener formulation of the identification problem and the crosscorrelation technique provide a significant advantage in that respect, if the external noise has zero expected value.

The external noise is a label under which we cover the contribution of a great variety of external factors to the performance of our system. These factors are usually uncontrollable in a specific experiment, and their total effect on the system has naturally a random character. There may be, however, external factors in some cases which have a dominant, distinguishable effect that attains a systematic character. In these cases, it is strongly suggested that those factors are identified and included in the system analysis as distinct inputs. Of course, this cannot always be done, because sometimes we are not able to identify or to control those factors. In any case, as long as this external noise is independent from the stimulus, the crosscorrelation technique provides a great comfort from its hazardous effect.

We can have noise present in the input or in the output of the system. It is shown here that the case of the input contaminating noise is much worse than the case of the output contaminating noise from the identification point of view.

Consider the system S and the case of output contaminating noise (Fig. 6.4.1). Then, if the external output noise n(t) is independent from x(t) and has average zero, the kernel estimates are not affected at all since:

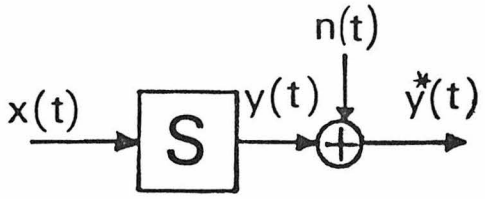


Fig. 6.4.1: Output contaminating noise

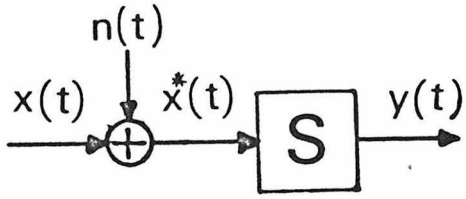


Fig. 6.4.2: Input contaminating noise

$$\hat{g}_n^*(\sigma_1, \ldots, \sigma_n) = C_n \cdot E\left[[y(t)+n(t)]\, x(t-\sigma_1) \ldots x(t-\sigma_n)\right]$$

$$= C_n \cdot E[y(t)x(t-\sigma_1) \ldots x(t-\sigma_n)]$$

$$= \hat{g}_n(\sigma_1, \ldots, \sigma_n) \qquad (6.4.1)$$

If n(t) has a nonzero average, then only the full-diagonal points of the even order kernels are affected to a degree described by eqn. 6.4.1. The independence of n(t) from x(t) is usually true in practice.

Consider now the case of input contaminating noise (Fig. 6.4.2). Here things are far more complicated, because the orthogonality of the G^*-functionals is obviously distorted, even if n(t) is independent from x(t) and has average zero.

To illustrate that, consider the zero and the second order functionals:

$$E[G_o^* \cdot G_2^*] = g_o \int_0^\infty \int_\infty g_2(\tau_1, \tau_2) E\left[[x(t-\tau_1)+n(t-\tau_1)]\,[x(t-\tau_2)+n(t-\tau_2)]\right]$$

$$d\tau_1 d\tau_2 - g_o \cdot P \int_0^\infty g_2(\tau, \tau) d\tau$$

$$= g_o \cdot \int_0^\infty \int_0^\infty g_2(\tau_1, \tau_2) \cdot E[n(t-\tau_1)n(t-\tau_2)]\, d\tau_1 d\tau_2 \qquad (6.4.2)$$

which is, in general, nonzero.

In conclusion, we state that the case of input contaminating external noise is far more hazardous than the case of output contaminating external noise; and on the other hand, the input noise is usually more controllable. Therefore, good care must be taken in order to be avoided up to the highest possible degree in the applications.

6.5 Design of the Optimum Test

In this section, we will outline the basic steps of the procedure which ought to be followed in practical applications of the CSRS in non-linear system identification.

The very first task is the system formulation of the problem. We define the input and output measures so that the basic assumptions of the Volterra series expansion (cf. sec. 2.3) are satisfied. A discussion on how to practically check the fulfillment of these assumptions is given in sec. 6.5.1.

After the problem has been formulated as an one input - one output Volterra system, our effort is directed towards determining the basic characteristics of the CSRS model to be estimated. These basic characteristics are: (a) the number of the CSRS functionals that are required in order to obtain an acceptable model for the system, (b) the extent of the memory of each one of the required kernels.

The determination of these characteristics can be done by successive trials, but a more systematic and rigorous method is discussed in chapter 7. The discussion of chapter 7 refers to the Volterra series of the system, but the conclusions can be extended to the CSRS series of the system on the basis of their relations as discussed in chapter 4.

After the extent of the functional series and the memory of the individual kernels have been determined (within the scope of practical interest), our effort is directed towards determining the basic test parameters: the record length T and the step length Δt; in such way that

the resulting m. s. error in our estimates is minimum. This procedure is discussed in sec. 6. 5. 2.

Finally, an overall account of the performance of the optimum test is given in sec. 6. 5. 3 and the design of the optimum test is concluded.

6. 5. 1 Checking the Defining Characteristics of a Volterra System

As we discussed in sec. 2. 3, the defining characteristics of a Volterra system are: (a) Stationarity (b) Finite-memory (c) Analyticity.

The stationarity of a system can be checked through the statistics of the system response to a stationary random stimulus.

One practical way to do that is by using a stationary gaussian random process as an input and then test the stationarity of the first two moments (i. e. the mean and the variance) of the output signal with a "run test".

Of course, if the system is nonlinear the output signal is not, in general, a gaussian process, which means that we do not cover the whole statistical behavior of the signal by testing only the first two moments. However, the test of these two moments is the minimum requirement to establish the weak stationarity of the signal and, consequently, of the system.

The "run test" is a special purpose statistical test, which applies in cases of stationarity testing. According to the "run test", given a random time series $\{x_1, x_2 \ldots x_n\}$, we form k groups of M successive values each, so that $N = k \cdot M$. Then we compute a statistic

d_i (like the sample mean or the sample variance etc.) from each one of the k groups using its M samples.

In this way, we have the sequence of the statistics $\{d_1, d_2 \ldots d_k\}$. We compute the median d_m of the population $\{d_1 \ldots d_k\}$, and the differences:

$$f_i = d_i - d_m \quad ; \quad i = 1, 2, \ldots, k \tag{6.5.1}$$

Finally, we form the sequence of the signs (positive or negative) of these differences and we count the number of times the sign is changing in the sequence. This number plus one is the number of runs in the sequence. Then, we compare this number to the bounds given in the table 6.5.1, which are the bounds of a confidence interval that the hypothesis of stationarity of the time series, with respect to the statistic d_i, is not rejected.

The finite-memory of a system is mainly a matter of definition of the input and output measures; inasmuch as differentiation can reduce the degree of a polynomial function. However, in some cases (e.g. undamped oscillatory memory), there is no proper definition of the input and output measures that gives a finite-memory system. We consider these cases very unusual in the physical world.

In any case, the final manifestation of the finite-memory of the system is done through actual estimation and inspection of the system kernels.

The analyticity of a system cannot be, in general, easily checked. The failure of the Volterra, Wiener or CSRS functional series to give satisfactory, sensible or consistent results is always a

TABLE 6.5.1

Confidence Bounds in "Run Test"

Total number of statistics d_i	Lower confidence bound in number of "runs"	Upper confidence bound in number of "runs"
10	2	10
12	3	11
14	3	13
16	4	14
18	5	15
20	6	16
22	7	17
24	7	19
26	8	20
28	9	21
30	10	22
32	11	23
34	11	25
36	12	26
38	13	27
40	14	28

supporting indication of the nonanalyticity of the system. Therefore, the analyticity of a system is usually judged on the basis of the final performance of the estimated system model.

6.5.2 The Optimal Determination of T and Δt

The optimal determination of T and Δt is based mainly upon the fundamental error equation of sec. 6.2. Notice that:

$$T = \alpha \cdot N \cdot \Delta t \qquad (6.5.2)$$

where, N is the number of samples in the record and α the inverse of the number of samples per stimulus step.

First, we determine the record length T. In this process, we distinguish two main cases:

Case I: The experimentation time is limited by the nature of the specific system under study or by the specific experimentation procedure; while our computational capacity is practically unlimited. For example, the limited physical endurance of a nerve cell being probed by a measurement electrode puts a strict limitation on the experimentation time and, consequently, on the input-output data records.

In this case, the record length is determined at the maximum experimentation time, which gives reliable data. The determination of Δt then follows, as the optimum value estimated with the help of the fundamental error equation and the relevant discussion of sec. 6.2. This optimization procedure aims at the value of Δt which gives the least mean square for the combined ϵ and θ errors: (cf. sec. 6.2)

$$Q_n = A_n \cdot \Delta t^4 + \frac{B_n}{T \cdot \Delta t^{n-1}} \qquad (6.5.3)$$

We assume that for all practical purposes, A_n and B_n are independent from Δt, in the range of Δt values of practical interest.

Of course, the determination of the optimum Δt in the n-th order kernel case requires the knowledge of the values of A_n and B_n. These two constants depend on the n-th order kernel. Apparently, there are a lot of ways to estimate these constants using the basic facts discussed in sec. 6.2. In any case, we will have to utilize the different nature of the θ and $\in$ errors (deterministic and random) in order to distinctly estimate A_n and B_n. We outline below a procedure of that sort.

We perform k preliminary tests with independent CSRS stimuli of the same T_1 and Δt_1. Of course, Δt_1 attains a value from the range of practical interest, as determined by the system bandwidth. The record length T_1 of these preliminary tests can be short, so that the experimental and computational burden is light. Actually, it is usually most practicable to obtain these k independent short records as segments of a single longer record.

In any case, from each one of these k preliminary tests we estimate the n-th order kernel by crosscorrelation. According to sec. 6.2 we have:

$$\hat{h}_n(\sigma_1, \ldots, \sigma_n) = h_n(\sigma_1, \ldots, \sigma_n) + \theta_n(\sigma_1, \ldots, \sigma_n) + \in_n(\sigma_1, \ldots, \sigma_n) \tag{6.5.4}$$

where,

$$\theta_n(\sigma_1, \ldots, \sigma_n) = \Delta t^2 \cdot D_{h_n}(\sigma_1, \ldots, \sigma_n) \tag{6.5.5}$$

and, $\epsilon_n(\sigma_1, \ldots, \sigma_n)$ is a gaussian random variable with:

$$E[\epsilon_n(\sigma_1, \ldots, \sigma_n)] = o \tag{6.5.6}$$

$$\mathrm{Var}[\epsilon_n(\sigma_1, \ldots, \sigma_n)] \triangleq q_n^2(\sigma_1, \ldots, \sigma_n) = \frac{1}{T \cdot \Delta t^{n-1}} \cdot C_{h_n}^2(\sigma_1, \ldots, \sigma_n) \tag{6.5.7}$$

From the k obtained kernel estimates $\{\hat{h}_{n_i}\}$, we can estimate, for each point $(\sigma_1, \ldots, \sigma_n)$, the mean and the variance of the random variable $\hat{h}_n(\sigma_1, \ldots, \sigma_n)$:

$$\hat{m}_{h_n}(\sigma_1, \ldots, \sigma_n) = \frac{1}{k} \sum_{i=1}^{k} \hat{h}_{n_i}(\sigma_1, \ldots, \sigma_n) \tag{6.5.8}$$

$$\hat{v}_{h_n}(\sigma_1, \ldots, \sigma_n) = \frac{1}{(k-1)} \sum_{i=1}^{k} [\hat{h}_{n_i}(\sigma_1, \ldots, \sigma_n) - \hat{m}_{h_n}(\sigma_1, \ldots, \sigma_n)]^2 \tag{6.5.9}$$

Clearly, (cf. eqn. 6.5.4)

$$m_{hn}(\sigma_1, \ldots, \sigma_n) = h_n(\sigma_1, \ldots, \sigma_n) + \theta_n(\sigma_1, \ldots, \sigma_n) \tag{6.5.10}$$

$$v_{h_n}(\sigma_1, \ldots, \sigma_n) = q_n^2(\sigma_1, \ldots, \sigma_n) \tag{6.5.11}$$

The estimation of B_n is now straightforward:

$$\hat{C}_{h_n}^2(\sigma_1, \ldots, \sigma_n) = T_1 \cdot \Delta t_1^{n-1} \cdot \hat{v}_{h_n}(\sigma_1, \ldots, \sigma_n) \tag{6.5.12}$$

and finally:

$$\hat{B}_n = \int_o^{\mu_n} \cdots \int_o^{\mu_n} \hat{C}_{h_n}^2(\sigma_1, \ldots, \sigma_n) d\sigma_1 \ldots d\sigma_n \tag{6.5.13}$$

For the estimation of A_n, we need another set of k preliminary tests with different step length Δt_2. From this second set, we obtain another estimate $\hat{m}_{h_n}^*$ of the mean of the n-th order kernel estimates.

Subtracting these two estimates of the mean, we obtain an estimate of A_n as:

$$\hat{D}_{h_n}(\sigma_1, \ldots, \sigma_n) = \frac{\hat{m}_{h_n}(\sigma_1, \ldots, \sigma_n) - \hat{m}^*_{h_n}(\sigma_1, \ldots, \sigma_n)}{\Delta t_1^2 - \Delta t_2^2} \qquad (6.5.14)$$

and finally:

$$\hat{A}_n = \int_0^{\mu_n} \ldots \int_0^{\mu_n} \hat{D}^2_{h_n}(\sigma_1, \ldots, \sigma_n) d\sigma_1 \ldots d\sigma_n \qquad (6.5.15)$$

Now that the basic parameters A_n and B_n have been estimated, the fundamental error equation can be used to determine the optimum step length for the estimation of the n-th order kernel of the system at hand. This actual determination can be done either graphically (see Fig. 6.5.1) or analytically (see eqn. 6.2.11). We discuss the graphical determination of $(\Delta t_{opt})_n$ later in this section.

Case II: The experimentation time is not very limited and our principal practical restriction now becomes the computational burden. In this case, our principal concern is the number N of data points in the record (see eqn. 6.5.2), and it is chosen to conform with our maximum computational capacity.

The optimization procedure for Δt is now performed on the basis of the modified FEE:

$$Q_n = A_n \cdot \Delta t^4 + \frac{B_n}{\alpha \cdot N \cdot \Delta t^n} \qquad (6.5.16)$$

The steps for the estimation of A_n and B_n are similar to the ones in case I, with some small modifications in the formulae resulting from the substitution of T with $(\alpha \cdot N \cdot \Delta t)$.

After A_n and B_n have been estimated, the optimum value of Δt is determined again either graphically (see Fig. 6.5.1) or analytically as:

$$(\Delta t_{opt})_n = \left[\frac{n}{4\alpha N} \cdot \frac{B_n}{A_n}\right]^{1/(n+4)} \qquad (6.5.17)$$

Notice that in the expression above, α is another determinable parameter. For a given N, α determines the number of steps in the stimulus signal. Clearly, α can only attain values in the range between zero and 1. In most of the cases α is chosen to be 1/2 i.e. two sampling points per stimulus step, which is the case where the sampling rate is equal to the Nyquist frequency of the stimulus.

However, α can be possibly used to achieve the optimum solution which exhausts the remaining margins of computational capacity or experimentation time of the previous cases I and II; since, it is a regulating parameter between the number of samples and the number of stimulus steps (and consequently the record length). In other words, the optimum choice of α, which exhausts the possible margins of both the "quantities in scarcity" (namely, computational capacity and experimentation time), is:

$$(\alpha_{opt})_n = {}^{T_{max}}\!/_{N_{max}} \cdot (\Delta t_{opt})_n \qquad (6.5.18)$$

where, T_{max} and N_{max} are determined by the practical limitations as discussed in cases I and II. It is also provided that:

$$o < (\alpha_{opt})_n \leq 1 \qquad (6.5.19)$$

where only the upper bound is of practical importance.

If $(\alpha_{opt})_n$, as computed from eqn. 6.5.18, is bigger than 1,

then the optimum feasible choice is 1. Of course in this case we encounter significant aliasing problems.

It must be noted that in most applications we have an understanding of which one is the principal practical limitation (the experimentation time or the computational burden) without being able to determine both of them quantitatively in a way that would allow the determination of $(\alpha_{opt})_n$ through eqn. 6.5.18. Thus, we usually choose the value of α on a more or less arbitrary fashion from a seemingly reasonable range of values (usually from 1/4 to 1/2).

The graphical determination of $(\Delta t_{opt})_n$ is much simplified if logarithmic paper is used (Fig. 6.5.1). The optimum position is determined from the Δt value corresponding to the cross section of the ϵ-error and θ-error lines by properly scaling it with $K_n=[4/(n-1)]^{1/(n+3)}$.

The ϵ-error line is described in the previously discussed two cases by the equations:

$$\text{Case I:} \qquad \log Q_n=\log B_n-\log T-(n-1)\cdot\log\Delta t \tag{6.5.20}$$

$$\text{Case II:} \qquad \log Q_n=\log B_n-\log(\alpha N)-n\cdot\log\Delta t \tag{6.5.21}$$

The θ-error line is described by the same equation in both cases I and II:

$$\log Q_n=\log A_n+4\cdot\log\Delta t \tag{6.5.22}$$

This graphical representation is very illustrative. Firstly, it illustrates clearly how $(\Delta t_{opt})_n$ is decreasing when T is increasing. Secondly, it shows nicely the dependence of $(\Delta t_{opt})_n$ on the order n. Thirdly, it illustrates the role of A_n and B_n, and consequently, the

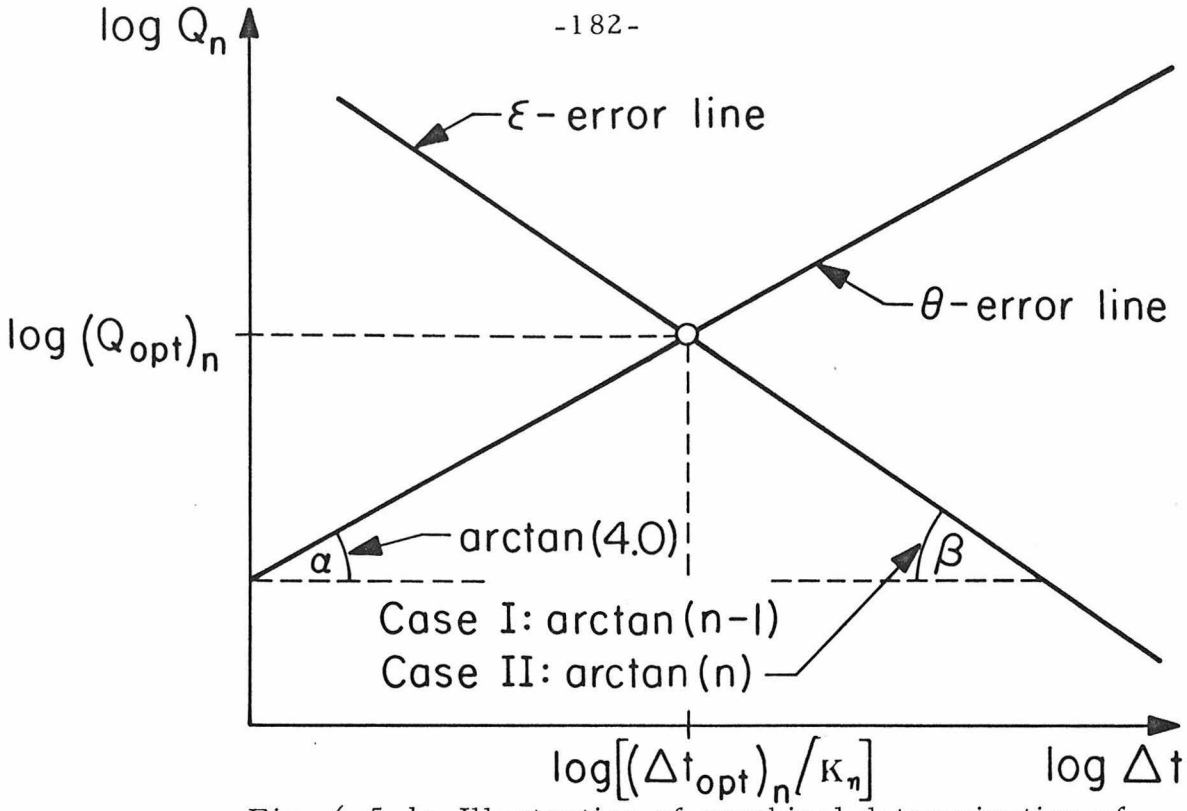


Fig. 6.5.1: Illustration of graphical determination of $(\Delta t_{opt})_n$

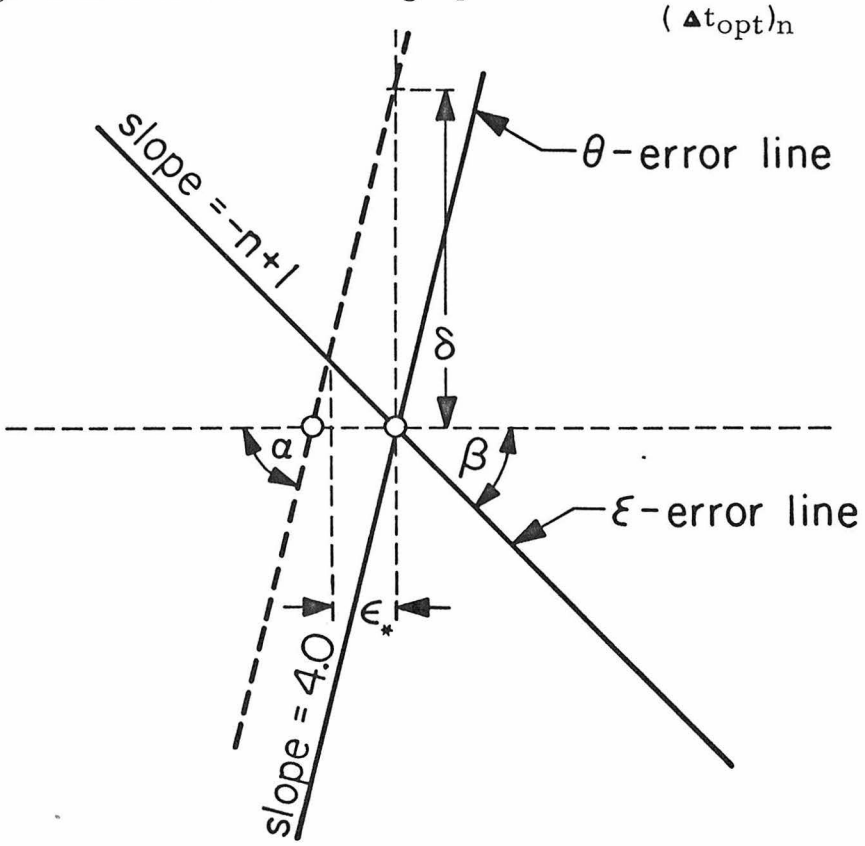


Fig. 6.5.2: Illustration of the sensitivity of the determined $(\Delta t_{opt})_n$.

effect of small estimation errors of $\hat{A}_n$ and $\hat{B}_n$ upon the determination of $(\Delta t_{opt})_n$.

Suppose, for example, that we have an estimation error δ of $\log A_n$ in the case I; then, as it is shown geometrically in Fig. 6.5.2, the induced error in the determination of $\log(\Delta t_{opt})_n$ is $\epsilon_* = \delta/(n+3)$ and towards the opposite direction (increase-decrease).

Similarly, if an estimation error δ occurs to $\log B_n$, then it can be shown easily that, the induced error in the determination of $\log (\Delta t_{opt})_n$ is again $\epsilon_* = \delta/(n+3)$ but towards the same direction with δ.

Therefore, the sensitivity of the determined $(\Delta t_{opt})_n$ seems lower in the higher order kernel cases, and its deviation is, in any case, smaller than the estimation error of $\hat{A}_n$ or $\hat{B}_n$ which causes it.

6.5.3 An Overall Account of the Performance of the Optimum Test

For a sufficiently long record length and a reasonable value of α ($\leq 1/2$), we eliminate, for all practical purposes, the effect of the external noise and the computational errors due to aliasing and discrete integration.

With the described procedure of the scaling adjustment of the estimated functional terms, we can practically eliminate any significant error due to finite transition time (reversible transitions) or erroneous power levels.

The approximate orthogonality errors are usually negligible, for all practical purposes, and if greater accuracy is desired, the suggestions given in sec. 5.3 can be followed.

Finally, the statistical fluctuation error and the deconvolution error are minimized through the optimal determination of T and Δt, as discussed in the previous section. In this way, the optimum test (in the mean square error sense) is achieved, using CSRS for nonlinear system identification.

CHAPTER VII

SPECIAL TECHNIQUES FOR THE STUDY OF THE VOLTERRA SERIES CONVERGENCE AND THE KERNELS EFFECTIVE MEMORY EXTENT

In this chapter, we present two special techniques which can be employed in the study of two aspects of particular practical interest in the crosscorrelation approach of nonlinear system indentification: the convergence of the system Volterra series and the effective memory extent of the system kernels.

These two aspects are associated with two basic preliminary decisions that the user of the crosscorrelation method has to make. The one is the decision concerning the number of functional terms that are required to be estimated in order to achieve a system model of reasonable accuracy (in accordance, of course, with the possessed experimental and computational capacity which may even make the actual identification of a system infeasible). The other decision concerns the extent to which the kernels have actually to be computed in order to modelize their significant effect (memory length).

The computational time depends very drastically on this extent of the computed kernel memory. In fact, if a constant sampling interval is used, the computational time is approximately proportional to the number of data points in one dimension raised to a power equal to the

order of the kernel. For example, if the memory length of the second order kernel is doubled, then the computational time is quadrupled.

We will refer to the first technique as the "general nonlinearity test," because it is intended to allow the practical determination of the degree of system nonlinearity, in terms of the number of estimated functionals that are required to give a model of reasonable accuracy. We will refer to the second technique as the "memory extent test," for obvious reasons.

It must be noted that the first of these tests deals with the Volterra series, and the obtainable results concern the Volterra kernels and functionals of the system. It is understood that we can extend the conclusions derived for the system Volterra series to any other of the equivalent orthogonal series, on the basis of the discussed relations between the several functional series (cf. sec. 2.5 and 4.3).

7.1 The General Nonlinearity Test

The general nonlinearity test (GNT) is a special purpose test aimed at examining the pattern of convergence of a system Volterra series, in order to allow the practical determination of a reasonable truncation point of the series.

It is well understood that a chosen truncation point (i.e., the order of the model) of the Volterra series will cover any other orthogonal functional series; in the sense that any truncated orthogonal functional series of the same order will give, in general, better accuracy within the specified operational and frequency range.

The GNT employs a special test signal (Fig. 7.1.1):

$$x(t) = A \cdot t \cdot e^{-\alpha t} u(t) \qquad (7.1.1)$$

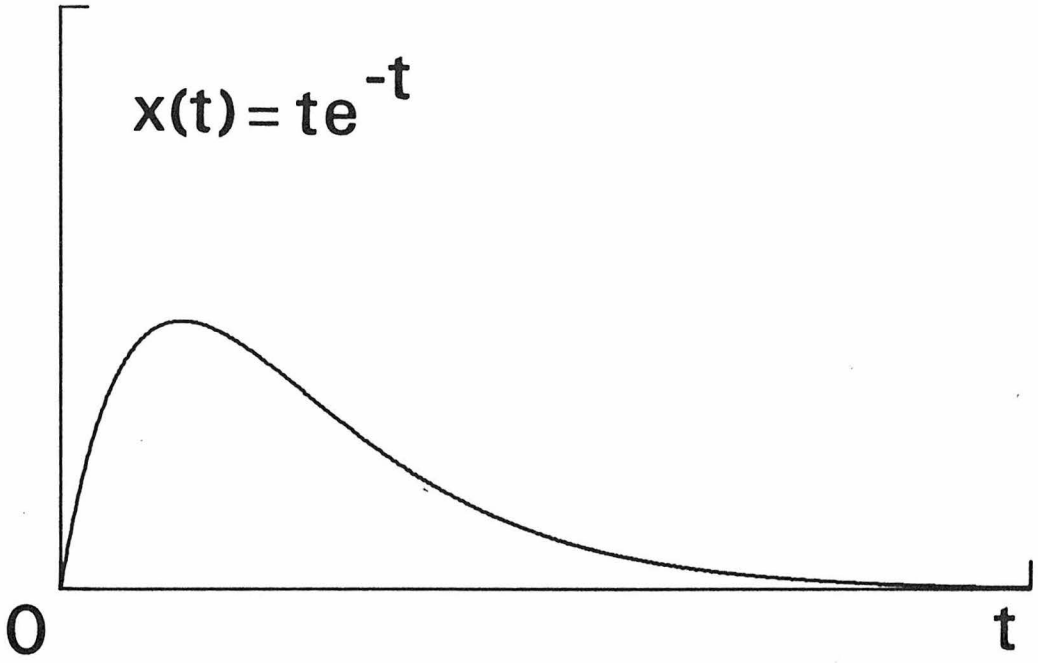


Fig. 7.1.1: Form of stimulus used in the general nonlinearity test (GNT).

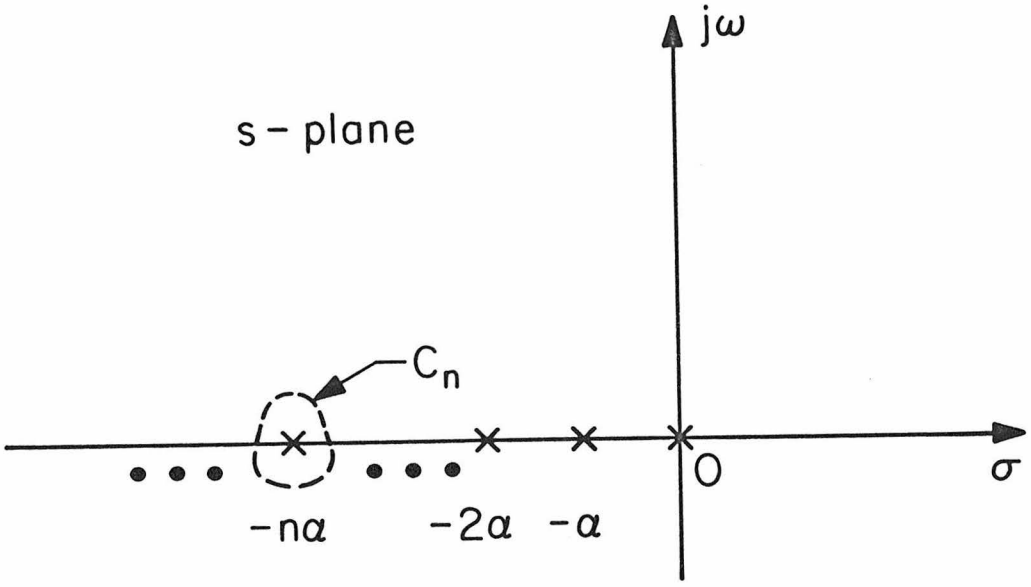


Fig. 7.1.2: Pole diagram of the Laplace transform of the system response in the GNT.

where A and α are positive constants and u(t) is the step function:

$$u(t) = \begin{cases} 1 & t \geq 0 \\ 0 & t < 0 \end{cases} \tag{7.1.2}$$

Clearly, x(t) is a deterministic signal of finite energy:

$$E = \int_0^\infty x^2(t)dt = A^2 \int_0^\infty t^2 e^{-2\alpha t} dt = \frac{A^2}{4\alpha^3} \tag{7.1.3}$$

The system response to this test signal is:

$$y(t) = \sum_{n=0}^{\infty} I_n(t) \tag{7.1.4}$$

where $I_n(t)$ is the n-th order Volterra functional of the system:

$$I_n(t) = \int_0^\infty \cdots \int_0^\infty k_n(\tau_1, \ldots, \tau_n) x(t-\tau_1)\, x(t-\tau_2) \cdots x(t-\tau_n) d\tau_1 \cdots d\tau_n \tag{7.1.5}$$

If we call:

$$\begin{aligned} X_n(t, \tau_1, \ldots, \tau_n) &= x(t-\tau_1) \cdots x(t-\tau_n) \\ &= A^n (t-\tau_1) \cdots (t-\tau_n) e^{-\alpha n t + \sum_{i=1}^{n} \alpha \tau_i} \cdot u(t-\tau_1) \cdots u(t-\tau_n) \end{aligned} \tag{7.1.6}$$

then, X_n is the only function that contains t within the functional $I_n(t)$.

Now, let us apply the one-sided Laplace transform:

$$Y(s) = S\{y(t)\} = \int_0^\infty y(t) e^{-st} dt = \sum_{n=0}^{\infty} \int_0^\infty I_n(t) e^{-st} dt \tag{7.1.7}$$

But,

$$L_n(s) \triangleq \int_0^\infty I_n(t) e^{-st} dt = \int_0^\infty \cdots \int_0^\infty k_n(\tau_1, \ldots, \tau_n) d\tau_1 \cdots d\tau_n \int_0^\infty X_n(t, \tau_1, \ldots, \tau_n) e^{-st} dt \tag{7.1.8}$$

and,

$$Q_n(s,\tau_1,\ldots,\tau_n) \triangleq \int_0^\infty X_n(t,\tau_1,\ldots,\tau_n)e^{-st}dt$$

$$= A^n e^{\alpha\sum_{i=1}^{n}\tau_i} \int_0^\infty (t^n + b_{n-1}t^{n-1} + \ldots + b_0)e^{-(s+\alpha n)t}dt \tag{7.1.9}$$

where $T_n = \max[\tau_1, \tau_2, \ldots, \tau_n]$ and the coefficients b_0, b_1, ..., b_n depend on τ_1, τ_2, ..., τ_n in the well-known fashion (relations between roots and coefficients of rational polynomials):

$$b_k = (-1)^{n-k} \sum_{\substack{\text{all possible}\\ \text{combinations}\\ \text{of } (j_1,j_2,\ldots,j_{n-k})}} \tau_{j_1}\tau_{j_2}\cdots\tau_{j_{n-k}} \tag{7.1.10}$$

where $j_1, j_2, \ldots, j_{n-k}$ attain the values 1, 2, ..., n. Consequently,

$$Y(s) = \sum_{n=0}^{\infty} \int_0^\infty \cdots \int_0^\infty k_n(\tau_1,\ldots,\tau_n)\, Q_n(s,\tau_1,\ldots,\tau_n)d\tau_1\cdots d\tau_n \tag{7.1.11}$$

Let us evaluate now Q_n:

$$Q_n(s,\tau_1,\ldots,\tau_n) = A^n \cdot e^{\alpha\sum_{i=1}^{n}\tau_i} \cdot \sum_{k=0}^{n} b_k \int_{T_n}^\infty t^k e^{-(s+\alpha n)t}dt \tag{7.1.12}$$

but

$$\int_0^\infty t^k e^{-(s+\alpha n)t}dt = \left[-e^{-(s+\alpha n)t} \sum_{j=0}^{k} \frac{k!}{j!} \frac{t^j}{(s+\alpha n)^{k+1-j}} \right]_{t=T_n}^{t=\infty}$$

$$= e^{-(s+\alpha n)T_n} \sum_{j=0}^{k} \frac{k!}{j!} \frac{T_n}{(s+\alpha n)^{k+1-j}} \tag{7.1.13}$$

$$(\text{for } \mathrm{Re}\{s+\alpha n\} > 0)$$

thus,

$$Q_n(s, \tau_1, \ldots \tau_n) = A^n \cdot e^{\alpha(\tau_1 + \ldots + \tau_n) - (s+\alpha n)T_n} \sum_{k=0}^{n} b_k \sum_{j=0}^{k} \frac{k!}{j!} \frac{T_n}{(s+\alpha n)^{k+1-j}}$$

(7.1.14)

Finally the expression for $Y(s)$ becomes:

$$Y(s) = \sum_{n=0}^{\infty} A^n \int_0^{\infty} \ldots \int_0^{\infty} k_n(\tau_1, \ldots \tau_n) T_n e^{-\sum_{i=1}^{n} \alpha(T_n - \tau_i) - sT_n} \sum_{k=0}^{n} \sum_{j=0}^{k} \frac{k!}{j!} \frac{b_k}{(s+\alpha n)^{k+1-j}}$$

(7.1.15)

Clearly, the complex function $Y(s)$ is analytic over the entire s-plane except at the points of the real axis:

$$s = -n\alpha \quad ; \quad n = 0, 1, 2, \ldots \tag{7.1.16}$$

where it exhibits regular singularities (poles). This is the special feature that we are going to utilize in the GNT.

Notice that the transformation $L_n(s)$ of each one of the Volterra functionals $I_n(t)$, is an analytic function over the entire s-plane except at the point $s = -n\alpha$, where it exhibits a regular singularity (pole) (Fig. 7.1.2). If we take the Cauchy integral on a contour C_n which surrounds the pole at $s = -n\alpha$ (and no other pole of $Y(s)$), then the resulting residue will correspond only to $L_n(s)$, while the contribution of the other terms in the series will vanish.

In this way, we isolate one term of the series at a time by changing the contour of integration. The resulting residues can be used as indicators of the relative contribution of each one of the terms in the series, and, consequently, they provide a picture of the convergence of the system Volterra series.

Let us evaluate these residues. Consider the Cauchy integral around the contour C_n (Fig. 7.1.2):

$$R_n = \frac{1}{2\pi i} \oint_{Cn} Y(s)ds = \frac{1}{2\pi i} \oint_{Cn} L_n(s)ds \tag{7.1.17}$$

and

$$R_n = A^n \int_0^\infty \cdots \int_0^\infty k_n(\tau_1, \ldots, \tau_n) T_n \cdot e^{\alpha(\tau_1 + \cdots \tau_n)} \cdot$$

$$\cdot \sum_{k=0}^{n} \sum_{j=0}^{k} \frac{k! b_k}{j!} \frac{1}{2\pi i} \oint_{Cn} \frac{e^{-(s+\alpha n)T_n}}{(s+\alpha n)^{k+1-j}} ds \cdot d\tau_1 \cdots d\tau_n$$

$$= A^n \int_0^\infty \cdots \int_0^\infty k_n(\tau_1, \ldots, \tau_n e^{\alpha(\tau_1 + \cdots \tau_n)} \cdot$$

$$\cdot \sum_{k=0}^{n} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} b_k T_n^{k-j} d\tau_1 \cdots d\tau_n \tag{7.1.18}$$

and, because of the symmetry of $k_n(\tau_1, \ldots \tau_n)$ we can write (accepting conventionally that $\tau_1 \geq \tau_2 \geq \ldots \geq \tau_n$):

$$R_n = A^n \cdot n! \cdot \int_0^\infty d\tau_1 \int_0^{\tau_1} d\tau_2 \cdots \int_0^{\tau_{n-1}} d\tau_n \cdot k_n(\tau_1, \ldots, \tau_n) \tau_1 e^{\alpha(\tau_1 + \ldots + \tau_n)} \cdot$$

$$\cdot \sum_{k=0}^{n} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} b_k \tau_1^{k-j} \tag{7.1.19}$$

For convenience of notation let us call:

$$\rho_n(\tau_1, \ldots, \tau_n) \triangleq \sum_{k=0}^{n} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} b_k \tau_1^{k-j+1} \tag{7.1.20}$$

Thus,

$$R_n = A^n n! \int_0^\infty d\tau_1 \int_0^{\tau_1} d\tau_2 \cdots \int_0^{\tau_{n-1}} d\tau_n \cdot k_n(\tau_1, \ldots, \tau_n) \rho_n(\tau_1, \ldots, \tau_n) e^{\alpha(\tau_1 + \ldots + \tau_n)} \tag{7.1.21}$$

The quantity R_n is characteristic of the n-th order Volterra kernel of the system. It also depends on the function $\rho_n(\tau_1, \ldots, \tau_n)$, which, however, is independent of the system. The function $\rho_n(\tau_1, \ldots, \tau_n)$ depends only on the arguments $\tau_1, \ldots, \tau_n$ and consequently it is a weighting function invariant through different system tests.

It must be noted that, in practice, the several arguments $\tau_1, \ldots, \tau_n$ attain values from a finite domain $[0, \mu_n]$, where μ_n is the effective memory of the n-th order kernel (which is a finite number for a finite-memory system). Consequently, the function ρ_n is bounded by the quantity:

$$\rho_n \leq \sum_{k=0}^{n} \sum_{j=0}^{k} (-1)^{n-j} \binom{k}{j} \binom{n}{k} \mu_n^{n-j+1} \tag{7.1.22}$$

Since,

$$|b_k| \leq \binom{n}{k} \mu_n^{n-k} \tag{7.1.23}$$

and

$$b_k = (-1)^{n-k} \cdot |b_k| \tag{7.1.24}$$

However, even if we do not allow ourselves to assume the finiteness (for all practical purposes) of μ_n, the function ρ_n is still bounded by the quantity $(\beta \cdot \tau_1^{n+1})$, where β is a finite positive constant appropriately chosen so that:

$$\sum_{k=0}^{n} \sum_{j=0}^{k} (-1)^{n-j} \binom{k}{j} \binom{n}{k} \tau_1^{n-j+1} \leq \beta \cdot \tau_1^{n+1} \tag{7.1.25}$$

(for $0 \leq \tau_1 \leq \infty$)

Therefore, the residue R_n exists if the kernel k_n is of exponential order:

$$k_n(\tau_1, \ldots, \tau_n) \leq M e^{-\theta_1\tau_1 - \theta_2\tau_2 - \cdots - \theta_n\tau_n} \qquad (7.1.26)$$

where $M, \theta_1, \theta_2, \ldots, \theta_n$ are the positive constants, and

$$\alpha < \max[\theta_1, \theta_2, \ldots, \theta_n] \qquad (7.1.27)$$

We see then that the absolute integrability of k_n, as required for the convergence of the series (cf. sec. 2.3), is not sufficient to assure the existence of R_n; but a stricter condition is required, as described by eqns. 7.1.26 and 7.1.27.

We studied the conditions under which R_n exists. Now, we will study one way in which the knowledge of R_n can be used to assist the choice of the truncation point of the system functional series.

To this end, we write eqn. 7.1.18 as:

$$R_n = \int_0^\infty d\tau_1 \int_0^\infty d\tau_2 \cdots \int_0^\infty d\tau_n \cdot k_n(\tau_1, \ldots, \tau_n) \cdot \beta_n(\tau_1, \ldots, \tau_n) \qquad (7.1.28)$$

where

$$\beta_n(\tau_1, \ldots, \tau_n) = A^n e^{\alpha(\tau_1 + \ldots + \tau_n)} \sum_{k=0}^{n} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} b_k \, T_n^{k+1-j} \qquad (7.1.29)$$

and

$$T_n = \max[\tau_1, \tau_2, \ldots, \tau_n] \quad .$$

Notice that the function β_n does not depend on the specific system under test, but it does depend on the parameters A and α of the specific stimulus used.

As we will see later, the dependence of β_n on α attains major importance in connection with a proposed identification

procedure. In general, the dependence of β_n on α is more important than the dependence on A, because the former affects the directionality of β_n (if β_n is considered a vector of infinite components), while the latter has a simple scaling effect without affecting the directionality of β_n.

Let us write:

$$\beta_n(\tau_1, \ldots, \tau_n) = A^n e^{\alpha(\tau_1 + \ldots + \tau_n)} \gamma_n(\tau_1, \ldots, \tau_n) \qquad (7.1.30)$$

where

$$\gamma_n(\tau_1, \ldots, \tau_n) = \sum_{k=0}^{n} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} b_k T_n^{k-j+1} \qquad (7.1.31)$$

Clearly the function γ_n does not depend either on the system or on the stimulus used, but only on the order n. Thus, it is an invariant function, which can be evaluated once and forever and be used in all different applications. This facilitates the study of the dependence of β_n on α and A.

Viewing the functions k_n and β_n as generalized vectors of infinite components, we can say that the residue R_n is the inner product of two generalized vectors; the one of which depends on the system under study and the other on the stimulus used.

Employing the Schwartz inequality we get:

$$|R_n| \leq \left[\int_0^\infty \ldots \int_0^\infty k_n^2(\tau_1, \ldots, \tau_n) \cdot d\tau_1 \ldots d\tau_n \cdot \int_0^\infty \ldots \int_0^\infty \beta_n^2(\tau_1, \ldots, \tau_n) d\tau_1 \ldots d\tau_n \right]^{\frac{1}{2}} \qquad (7.1.32)$$

This inequality involves the euclidean norms of two generalized vectors (k_n and β_n) and a scalar (R_n):

$$\|R_n\| \leq \|k_n\| \cdot \|\beta_n\| \qquad (7.1.33)$$

where,

$$\| R_n \| = | R_n |^2 \tag{7.1.34}$$

$$\| k_n \| = \int_0^\infty \cdots \int_0^\infty \cdot k_n^2(\tau_1, \ldots, \tau_n) d\tau_1 \cdots d\tau_n \tag{7.1.35}$$

$$\| \beta_n \| = \int_0^\infty \cdots \int_0^\infty \cdot \beta_n^2(\tau_1, \ldots, \tau_n) d\tau_1 \cdots d\tau_n \tag{7.1.36}$$

Finally, since the quantities $\| R_n \|$ and $\| \beta_n \|$ are known (the first is computed from the contour integration and the other is given by eqn. 7.1.29), a lower bound for $\| k_n \|$ can be obtained as:

$$\| k_n \| \geq \frac{\| R_n \|}{\| \beta_n \|} \tag{7.1.37}$$

This lower bound can be used as a reliable indicator in choosing the truncation point of the system functional series.

Clearly, this indicator (being a lower bound) does not provide information about where we can truncate the series, but about where we cannot do so. For example, if the quantity $\frac{\| R_n \|}{\| \beta_n \|}$ is of considerable size, then we know that we have to include the n-th order functional in our truncated model, since the n-th order kernel will be also sizeable. On the other hand, if the quantity $\frac{\| R_n \|}{\| \beta_n \|}$ is negligible, then this does not necessarily imply that the n-th order kernel is negligible. It may or may not be, since the inner product attains small values either because the vectors are small or because they are almost orthogonal.

This can be checked in some cases by changing the directionality of β_n (changing the parameter α) and re-evaluating $\frac{\| R_n \|}{\| \beta_n \|}$. If the new lower bound has changed significantly, then k_n is probably not

negligible, but β_n and k_n happened to be fairly "orthogonal." Thus, in some cases, we can detect an accidental orthogonal arrangement between k_n and β_n .

There is a more sophisticated and conclusive way to utilize the information obtained through the evaluated residues. Nonetheless, this new way is more involved and laborious. It aims at the actual evaluation of the system Volterra kernels through the information contained in the residues. For this reason, its actual implementation could have far reaching consequences in the general approach of the identification problem.

The method is based on the expansion of the kernel k_n in an appropriate basis of functions:

$$k_n(\tau_1, \ldots, \tau_n) = \sum_{i=1}^{\infty} a_i \psi_{i,n}(\tau_1, \ldots, \tau_n) \cong \sum_{i=1}^{M} a_i \psi_{i,n}(\tau_1, \ldots, \tau_n) \qquad (7.1.38)$$

Then by evaluating the integrals:

$$p_{i,n} = \int_0^{\infty} \cdots \int_0^{\infty} \psi_{i,n}(\tau_1, \ldots, \tau_n) \beta_n(\tau_1, \ldots, \tau_n) d\tau_1 \cdots d\tau_n \qquad (7.1.39)$$

We obtain a linear expression of the residue R_n in terms of the unknown expansion coefficients a_i (assuming that M terms of the expansion suffice for a satisfactory accuracy):

$$R_n = \sum_{i=1}^{M} a_i \cdot p_{i,n} \qquad (7.1.40)$$

Notice that the quantities R_n and $\{p_{i,n}\}$ depend on the parameter α of the stimulus signal. Thus, by changing α , we obtain M independent linear equations:

$$\left\{ R_n(\alpha_j) = \sum_{i=1}^{M} a_i \cdot p_{i,n}(\alpha_j) \right\} \qquad j = 1, 2, \ldots, M \qquad (7.1.41)$$

which can be solved to obtain the values of the unknown coefficients $\{a_i\}$.

In this way, we obtain an approximate of the kernel k_n in terms of the expansion of eqn. 7.1.38, and up to a desirable degree of accuracy. Apparently, there are some computational problems associated with the proposed method; however, the achievable results are of such a great importance for the nonlinear system identification problem as to justify a sizable investment in the study and solution of the computational aspects of the method.

7.2 The Memory Extent Test

This test is designed so as to allow the determination of the effective memory extent of the kernels of a system, before these kernels are actually estimated. These kernels may correspond to any of the quasi-white signals that we discussed in the previous chapters.

This is of major importance in practical applications, because the memory length of an estimated kernel is a crucial factor affecting the computational burden. This becomes increasingly dramatic as we go to higher order kernels.

In fact, for a certain sampling interval, the computational burden for the estimation of the n-th order kernel is approximately proportional to the n-th power of the computed kernel memory length. Therefore, it is evident that some a priori knowledge of the extent of

the kernels effective memory can be of great practical usefulness, since it will optimize the use of our computational capacity.

The test is based upon the following analysis which involves the second order autocorrelation function of the system response to a quasi-white signal. We know that:

$$y(t) = \sum_{n=0}^{\infty} G_n^* [h_n; x(t'),\ t' \leq t] \tag{7.2.1}$$

where, $x(t)$ is a quasi-white stimulus signal and G_n^* are the corresponding orthogonal functionals of the system under study. The second order autocorrelation function of the system response is:

$$\begin{aligned} \phi_{yy}(\sigma) &= E[y(t)\, y(t-\sigma)] \\ &= E\left[\sum_{n=0}^{\infty} G_n^*(t) \cdot \sum_{m=0}^{\infty} G_m^*(t-\sigma)\right] \\ &= \sum_{n=0}^{\infty} E[G_n^*(t)\, G_n^*(t-\sigma)] \end{aligned} \tag{7.2.2}$$

since,

$$E[G_n^*(t)\, G_m^*(t)] = 0 \quad \text{if} \quad n \neq m \quad . \tag{7.2.3}$$

Let us now evaluate some of the first terms of the series on the right hand side of eqn. 7.2.2 in the special case where the amplitude probability distribution of the quasi-white stimulus signal is gaussian. We consider this special case to simplify the derived analytical expressions; however, our basic conclusions hold qualitative for any amplitude probability distribution. We can easily derive that:

$$E[G_0^{*2}] = h_0^2 \tag{7.2.4}$$

$$E[G_1^*(t)\, G_1^*(t-\sigma)] = P \cdot \int_0^\infty h_1(\tau) h_1(\tau-\sigma) d\tau \tag{7.2.5}$$

$$E[G_2^*(t)\, G_2^*(t-\sigma)] = 2P^2 \int_0^\infty \int_0^\infty h_2(\tau_1, \tau_2)\, h_2(\tau_1-\sigma, \tau_2-\sigma) d\tau_1 d\tau_2 \tag{7.2.6}$$

$$E[G_3^*(t)\, G_3^*(t-\sigma)] = 6P^3 \int_0^\infty \int_0^\infty \int_0^\infty h_3(\tau_1,\tau_2,\tau_3)\, h_3(\tau_1-\sigma,\tau_2-\sigma,\tau_3-\sigma) d\tau_1 d\tau_2 d\tau_3 \tag{7.2.7}$$

etc.

where, P is the power level of the quasi-white signal.

Extending our derivations to the n-th order functional , we get:

$$E[G_n^*(t) G_n^*(t-\sigma)] = n! P^n \int_0^\infty \cdots \int_0^\infty h_n(\tau_1,\tau_2,\ldots,\tau_n) h_n(\tau_1-\sigma,\tau_2-\sigma,\ldots,\tau_n-\sigma) d\tau_1 d\tau_2 \cdots d\tau_n \tag{7.2.8}$$

This expression 7.2.8 along with eqn. 7.2.2 are the key formulas for the memory extent test.

We have found that the second order autocorrelation function of the system response to a quasi-white stimulus with gaussian amplitude distribution is:

$$\phi_{yy}(\sigma) = \sum_{n=0}^{\infty} A_n(\sigma) \cdot P^n \cong \sum_{n=0}^{N} A_n(\sigma) \cdot P^n \tag{7.2.9}$$

where,

$$A_n(\sigma) = n! \int_0^\infty \cdots \int_0^\infty h_n(\tau_1, \tau_2, \ldots \tau_n)\, h_n(\tau_1-\sigma, \tau_2-\sigma, \ldots, \tau_n-\sigma) d\tau_1 d\tau_2 \cdots d\tau_n \tag{7.2.10}$$

Notice that the memory extent of the h_n is the same in all of its

dimensions because of its symmetry. Consequently, if there is some finite effective memory μ_n for the kernel h_n, then the quantity $A_n(\sigma)$ for $\sigma > \mu_n$ has to negligible, in exactly the same way that $h_n(\tau_1, \ldots, \tau_n)$ is negligible for $\tau_i > \mu_n$.

Therefore, for successive values of σ we check which ones of the coefficients $A_n(\sigma)$ of the polynomial in P of eqn. 7.2.9 become negligible. The least value of σ for which $A_n(\sigma)$ becomes and remains negligible (while σ is increasing) determines the effective memory extent μ_n of the kernel h_n.

Of course, a practical problem of this method is the actual evaluation of the coefficients $A_n(\sigma)$ of the polynomial in P that $\phi_{yy}(\sigma)$ is for a given value of σ. Two ways to approach this problem are by using the various methods of polynomial interpolation or by straightforward linear expansions on a basis of powers of P.

Evidently, a reliable implementation of this method could also be used in connection with the problem of the functional series truncation by detecting no-memory system kernels.

CHAPTER VIII

COMPUTER SIMULATED APPLICATIONS OF THE CSRS IN NONLINEAR SYSTEM IDENTIFICATION

In chapter 4, we discussed the autocorrelation properties and the use of the CSRS in nonlinear system identification in connection with the crosscorrelation technique. In that chapter, we presented the theoretical arguments that manifest the quasi-whiteness of the CSRS.

In this chapter, we will illustrate the quasi-whiteness of the CSRS through the actual identification of computer simulated systems. In addition to that, we will illustrate several basic features of the CSRS test signals and several important aspects of their use in nonlinear system identification (i. e. estimation errors, convergence of the CSRS functional series etc.) as they have been theoretically discussed throughout the chapters 4, 5 and 6.

8. 1 Illustration of Basic Features of CSRS Test Signals

In this section, we will illustrate with the help of the computer (PDP 11/45) some basic features of the CSRS test signals.

In this illustration, we will use members of the equirandom CSRS group. A multi-level CSRS is called equirandom, when it attains a finite number of discrete equidistant amplitude levels with the same probability.

The basic features that we will illustrate in this section are: the form of the test signal; its amplitude histogram; its second order autocorrelation function estimate; and the histogram of the side values

($\tau > |\Delta t|$) of the autocorrelation estimate. The autocorrelation estimate has been obtained from a stimulus record of 6600 independent steps.

These basic features of equirandom CSRS with number of levels: 2, 3, 4, 8, 16 and 64 are shown in Fig. 8.1.1, 8.1.2, 8.1.3, 8.1.4, 8.1.5 and 8.1.6. respectively.

Notice that the operational range is the same for all these signals: $[-A, +A]$. Therefore, the second moment m_2 (and consequently the power level $P=m_2 \cdot \Delta t$) of them decreases as the number of levels increases. The explicit relation between the second moment m_2 and the number of levels M of an equirandom CSRS is:

$$m_2 = \frac{A^2}{3} \cdot \frac{M+1}{M-1} \qquad (8.1.1)$$

Clearly, this is a monotonically decreasing function of M, which asymptotically approaches the value $m_2 = \frac{A^2}{3}$ (Fig. 8.1.7). The gradual decrease of the second moment (and the power level) is demonstrated in Figs. 8.1.1 through 8.1.6 by the gradual decrease of the peak of the corresponding autocorrelation estimates.

Another interesting remark is that the histogram of the side values of the autocorrelation estimates (i.e. the values of the deviation function r_2 introduced in sec. 5.2 for $\tau \neq o$) is close to a gaussian distribution (as it was theoretically determined in sec. 4.2). Naturally, the variance of these histograms is decreasing as the number of levels increases, since the second moment of the respective signals is also decreasing (cf. eqn. 5.2.6). According to eqn. 5.2.6 the variance is proportional to the square of the second moment. This is manifested by the findings of our computational analysis (Fig. 8.1.8).

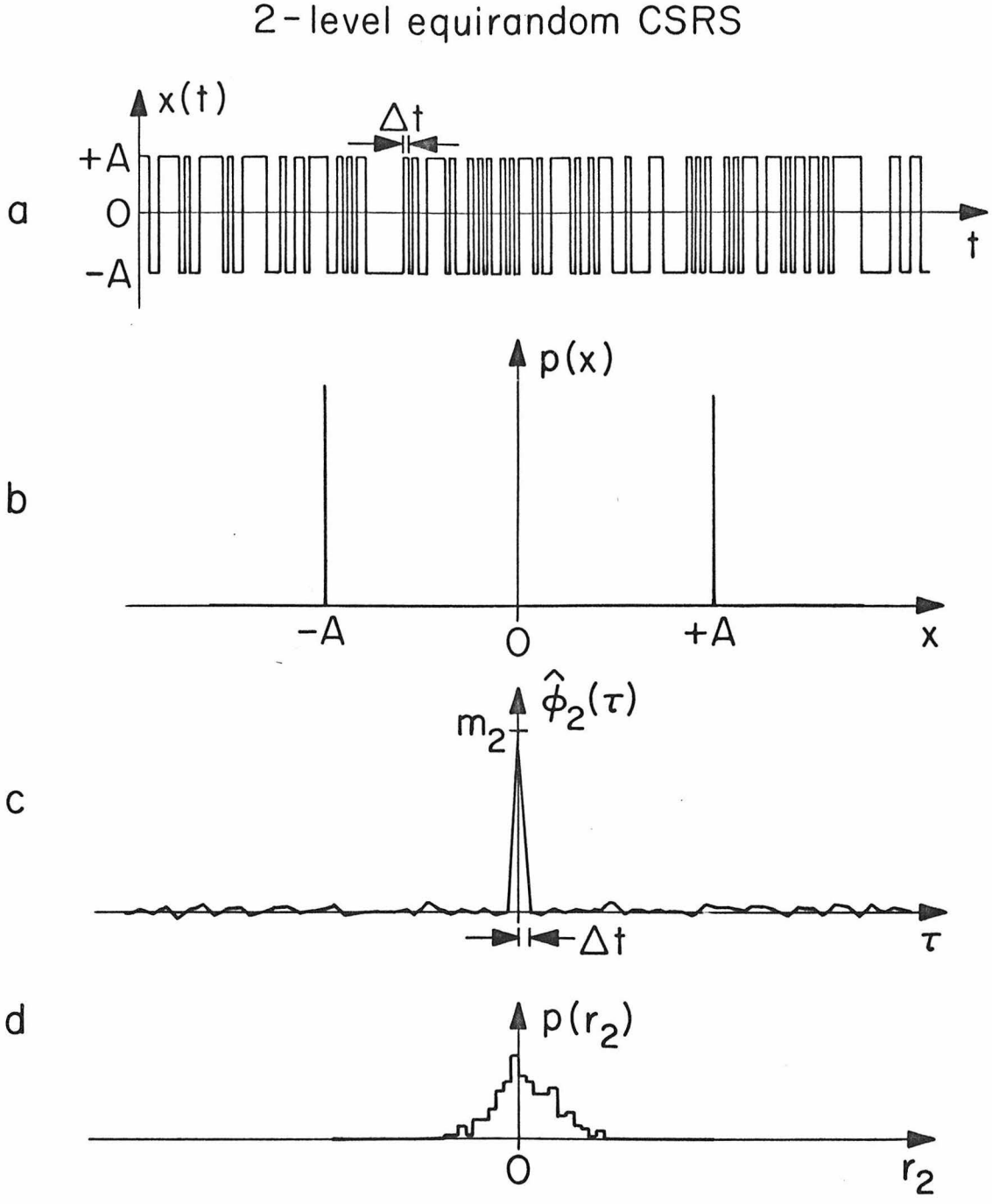


Fig. 8.1.1: (a) Form of a 2-level equirandom CSRS.
(b) Its amplitude histogram.
(c) Its second order autocorrelation function estimate.
(d) Histogram of the side values ($\tau > \Delta t$) of $\hat{\phi}_2(\tau)$.

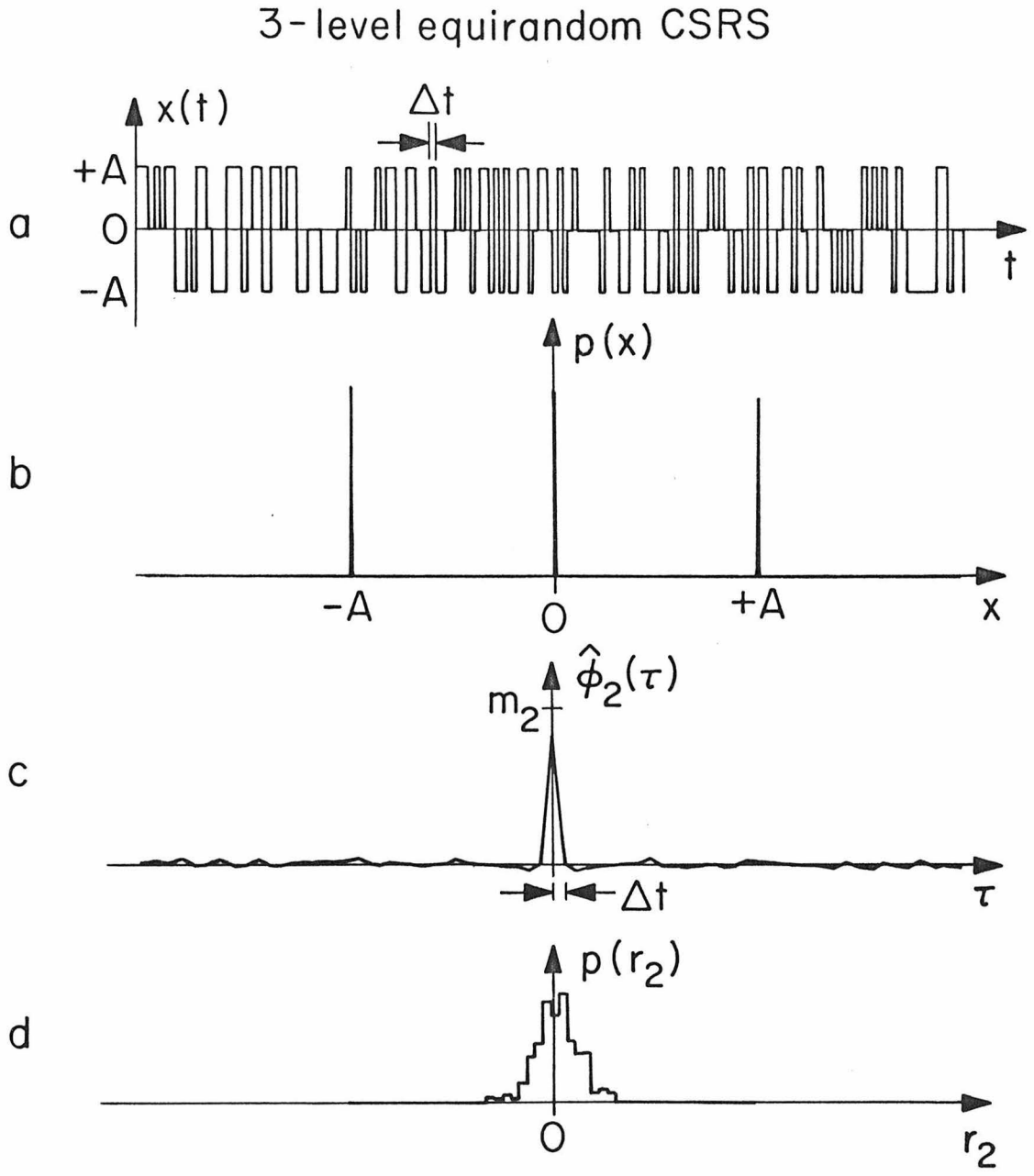


Fig. 8.1.2: (a) Form of a 3-level equirandom CSRS.
(b) Its amplitude histogram.
(c) Its second order autocorrelation function estimate.
(d) Histogram of the side values ($\tau > \Delta t$) of $\hat{\phi}_2(\tau)$.

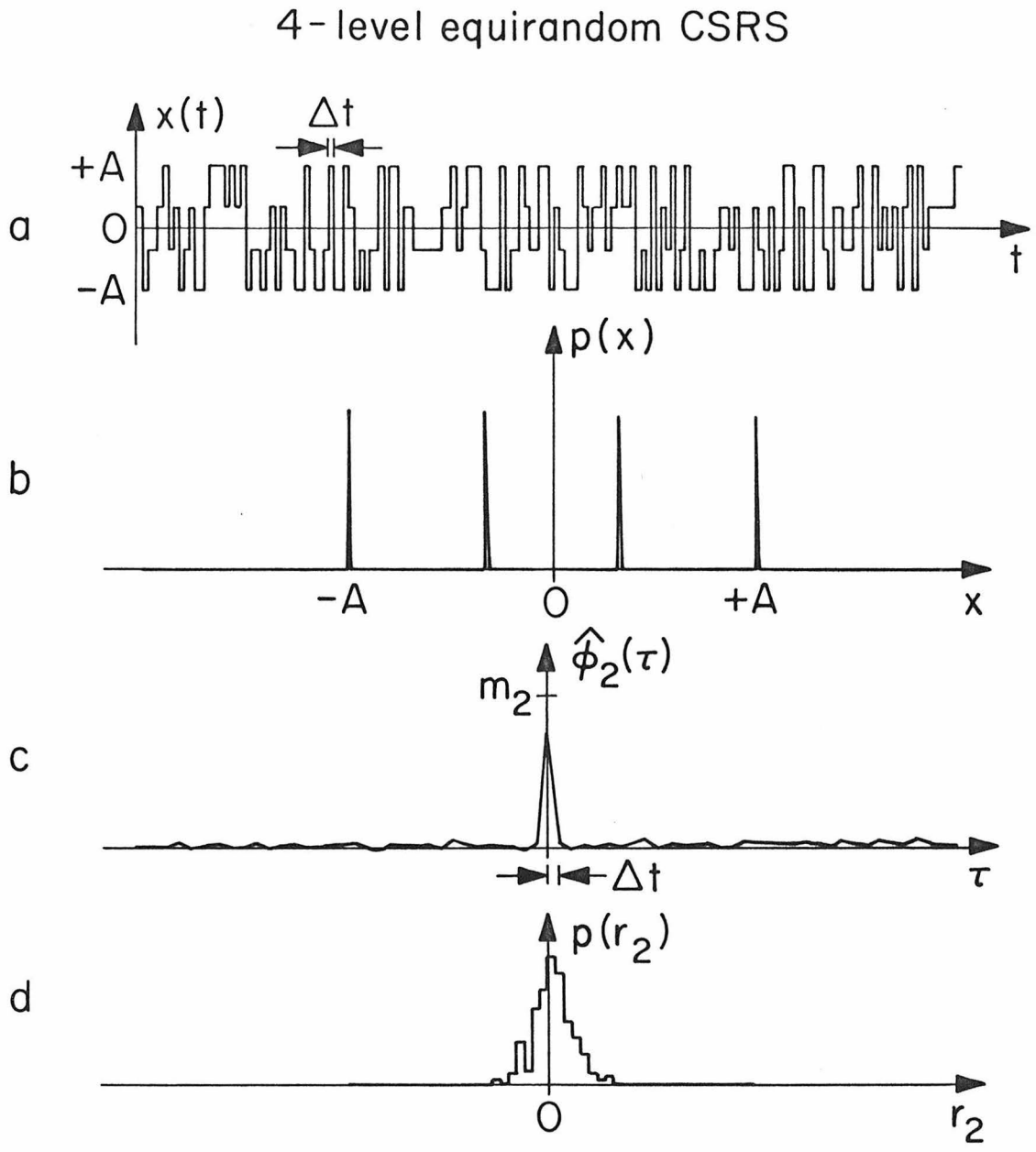


Fig. 8.1.3: (a) Form of a 4-level equirandom CSRS.
(b) Its amplitude histogram.
(c) Its second order autocorrelation function estimate.
(d) Histogram of the side values ($\tau > \Delta t$) of $\hat{\phi}_2(\tau)$.

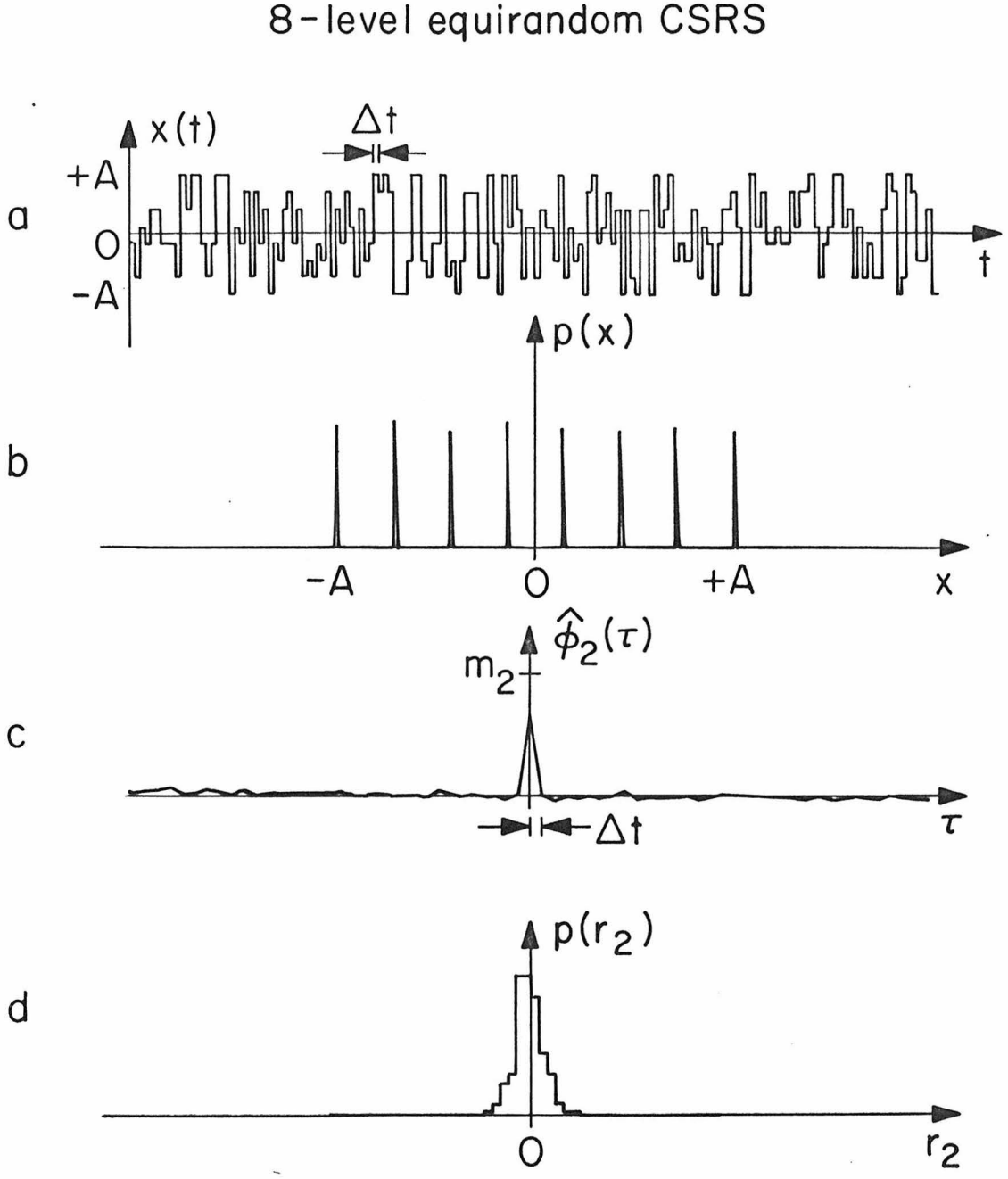


Fig. 8.1.4: (a) Form of a 8-level equirandom CSRS.
(b) Its amplitude histogram.
(c) Its second order autocorrelation function estimate.
(d) Histogram of the side values ($\tau > \Delta t$) of $\hat{\phi}_2(\tau)$.

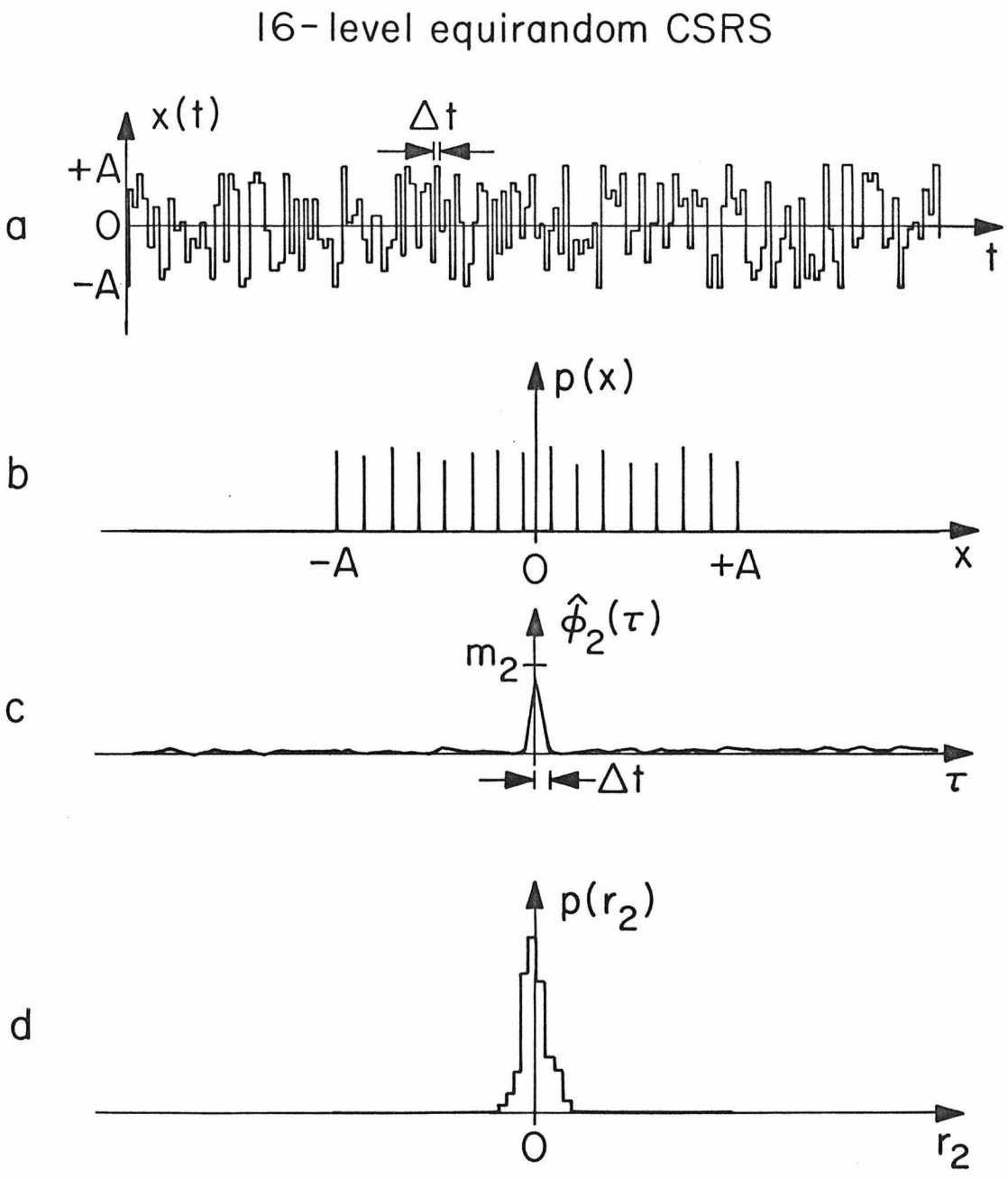


Fig. 8.1.5: (a) Form of a 16-level equirandom CSRS.
(b) Its amplitude histogram.
(c) Its second order autocorrelation function estimate.
(d) Histogram of the side values ($\tau > \Delta t$) of $\hat{\phi}_2(\tau)$.

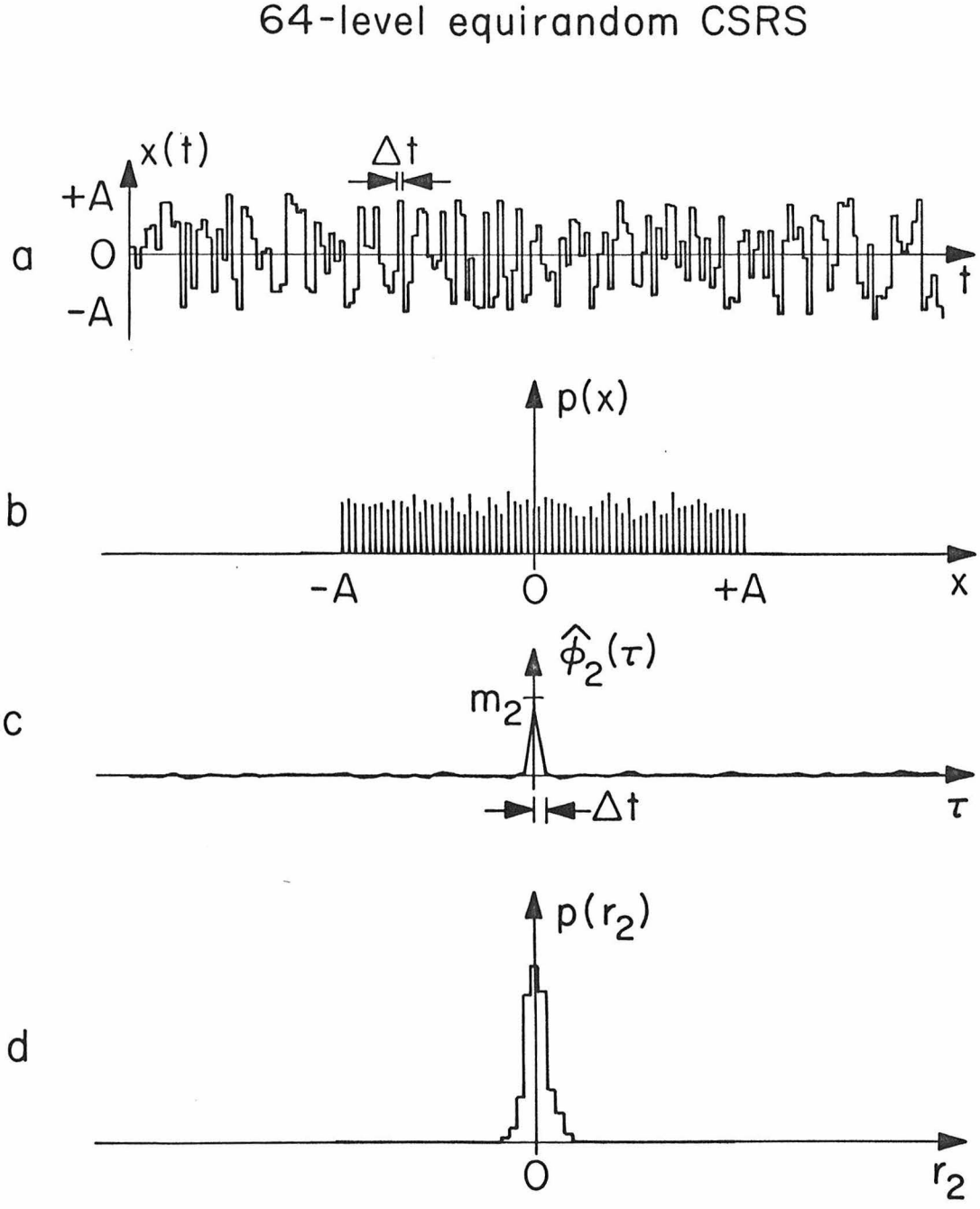


Fig. 8.1.6: (a) Form of a 64-level equirandom CSRS.
(b) Its amplitude histogram.
(c) Its second order autocorrelation function estimate.
(d) Histogram of the side values ($\tau > \Delta t$) of $\hat{\phi}_2(\tau)$.

M	2	4	8	16	32	64	128	256
m_2/A^2	1	.555	.428	.378	.355	.344	.338	.336

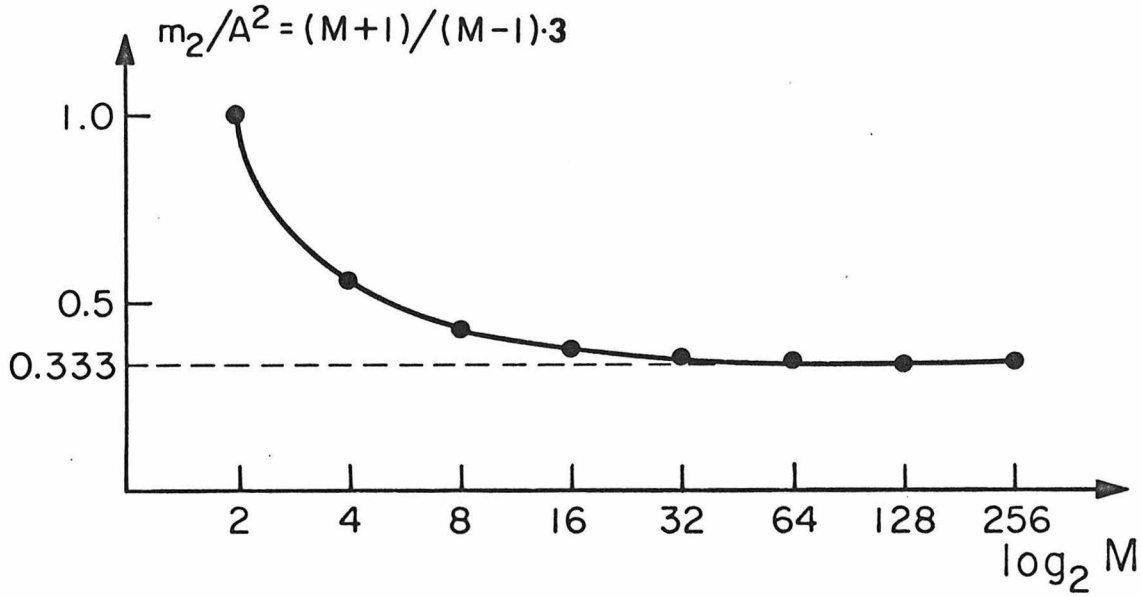


Fig. 8.1.7: Dependence of second moment m_2 of an equirandom CSRS with half operational range A upon the number of levels M.

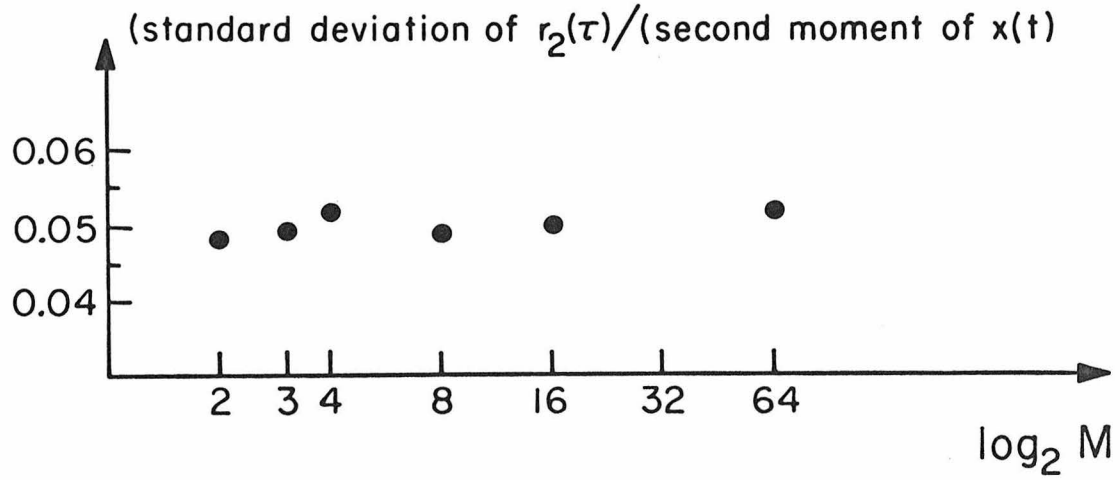


Fig. 8.1.8: Illustration of the dependence of the standard deviation of the autocorrelation estimate side values on the second moment of the CSRS.

It must be emphasized that the smaller variance of the side autocorrelation values in the case of signals with a great number of levels, must not be misread as an advantage of these signals over the ones with fewer levels (from the statistical fluctuation point of view). What is of importance with respect to the statistical fluctuation error is the ratio of this variance to the expected peak value of the respective autocorrelation estimate. And this ratio is independent from the second moment (cf. eqns. 5.2.58 through 5.2.60).

Another important remark concerning the side values of the autocorrelation estimates is the fact that those values are uncorrelated. This is illustrated in Fig. 8.1.9 by the autocorrelation of these side values for a ternary equirandom CSRS. The data set used had only 200 sample points, thus the obtained graph demonstrates strongly the lack of correlation between the nodal side values. Notice the parabolic shape of the principal and the side lobes, which was theoretically anticipated, as being the product of the linear segments of the autocorrelation estimate.

8.2 Illustration of CSRS Kernel Estimates

In this section we will actually compute the first and second order kernel estimates of a computer simulated nonlinear system, using several CSRS. This computer simulated nonlinear system has nonlinearities of all orders, so that possible deviations of the autocorrelation functions of the several CSRS test signals from quasi-whiteness would induce serious estimation errors. This is an indirect but

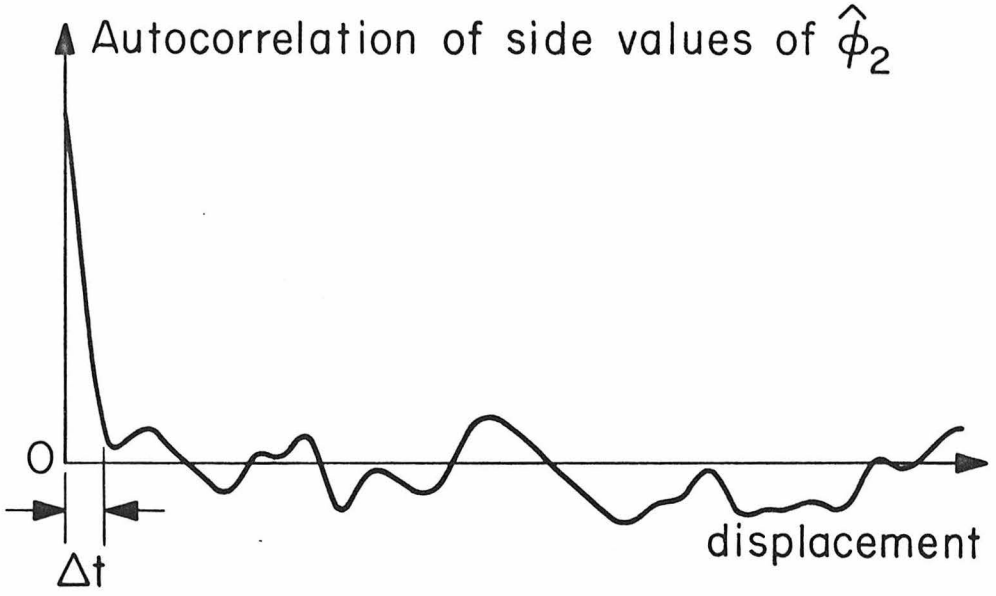


Fig. 8.1.9: Illustration of the lack of correlation between the nodal side values of the autocorrelation estimates.

reliable way to check the quasi-whiteness of the autocorrelation functions of the CSRS test signals and the validity of their use in nonlinear system identification.

This actual computation of the several CSRS kernel estimates will also demonstrate the dependence of these kernels upon the moments and the step length of the corresponding CSRS; and it will also illustrate the form of the CSRS kernels as compared to the Wiener and Volterra kernels of the system.

The system that we will use in this illustration is the cascade of a linear and a zero-memory nonlinear system, as shown in Fig. 8.2.1. This system has nonlinearities of all orders. This becomes evident when we expand the function $y(v)=e^{v}$ in a Taylor series about $v=o$.

The n-th order Volterra kernel of this system is:

$$k_n(\tau_1, \ldots, \tau_n) = \frac{1}{n!} \cdot h(\tau_1) \cdot h(\tau_2) \ldots \cdot h(\tau_n) \qquad (8.2.1)$$

Consequently, its first and second order Wiener kernels are:

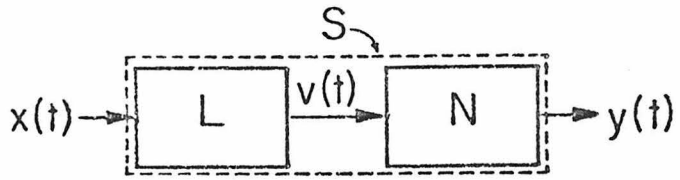
$$h_1(\tau_1) = \alpha_1 \cdot h(\tau_1) \qquad (8.2.2)$$

$$h_2(\tau_1, \tau_2) = \alpha_2 \cdot h(\tau_1) \cdot h(\tau_2) \qquad (8.2.3)$$

where α_1, α_2 are scalars depending upon the power level P.

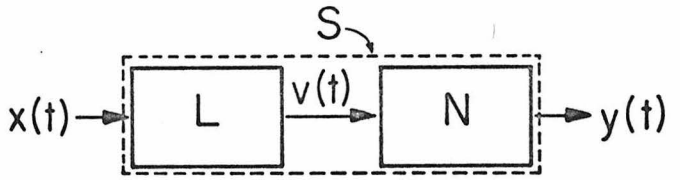
The first and second order kernels corresponding to a certain CSRS have the form:

$$g_1(\tau_1) = \sum_{k=o}^{\infty} \beta_{1,k} h^{2k+1}(\tau_1) \qquad (8.2.4)$$



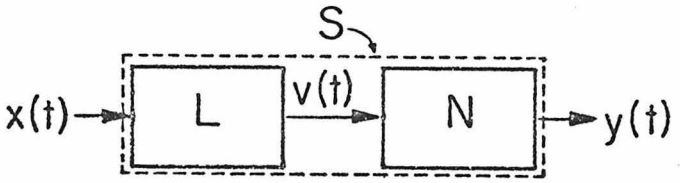
$$v(t) = \int g(\tau)x(t-\tau)\,d\tau \qquad y = e^{v}$$

Fig. 8.2.1: Cascade nonlinear system of infinite order.



$$v(t) = \int g(\tau)x(t-\tau)\,d\tau \qquad y = v+v^2+v^3$$

Fig. 8.3.1: Cascade nonlinear system of third order.



$$v(t) = \int g(\tau)x(t-\tau)\,d\tau \qquad y = v+v^2$$

Fig. 8.5.1: Cascade nonlinear system of second order.

$$g_2(\tau_1, \tau_2) = \begin{cases} \sum_{k,\ell=1}^{\infty} \beta_{2,k,\ell}\, h^k(\tau_1) h^\ell(\tau_2) & \text{for } |\tau_1 - \tau_2| \geq \Delta t \\ \sum_{k,\ell=1}^{\infty} \beta^*_{2,k,\ell}\, h^k(\tau_1) h^\ell(\tau_2) & \text{for } |\tau_1 - \tau_2| < \Delta t \end{cases} \qquad (8.2.5)$$

and k, ℓ attain those values for which $(k+\ell)$ is even. The coefficients $\beta_{1,k}$, $\beta_{2,k,\ell}$ and $\beta^*_{2,k,\ell}$ depend on the moments and the step length of the respective CSRS.

Apparently, the expressions for the CSRS kernels are less elegant than the ones for the Wiener kernels. The simplicity and the elegance of the expressions of the Wiener kernels are due to the decomposition property of the gaussian random variables (as it was discussed in sec. 3.2), which results in relatively simple analytical relations among the moments of a gaussian random variable.

In the case of the CSRS, however, the several moments of the amplitude probability density function p(x) may have any kind of relation, and consequently, the resulting analytical expressions are more complicated - in their generality.

We can also notice that the form of the CSRS kernels is generally different than the form of the Wiener or the Volterra kernels (which in this case, differ only by a scalar factor).

Another thing to be noticed is the idiomorphic form of the CSRS kernels at the diagonal points (cf. eqn. 8.2.5).

In the present illustration, we will use six equirandom CSRS (introduced in the previous section) of different number of levels,

namely 2, 3, 4, 8, 16, 64. All of this stimuli have half operational range A=1 stimulus unit, step length $\Delta t = .3$ sec. and record length T=4000 sec.

The first and second order kernel estimates obtained by each one of these CSRS stimuli are shown in Figs. 8.2.2 through 8.2.7. For comparison purposes the first and second order Volterra kernels of the system are shown in Fig. 8.2.8. Notice the idiomorphic form of the CSRS second order kernels at the diagonal points and the distinction of the 2-levels case.

From the Figs. 8.2.2 through 8.2.7, it is evident that the used CSRS stimuli possess the anticipated quasi-white autocorrelation properties, and that the magnitude of the obtained kernel estimates depends on the moments of the respective test stimulus signal. Noting that a moment of any order of the used CSRS stimuli is monotonically decreasing with the number of levels of the signal, we verify the decreasing magnitude of the corresponding kernel estimates.

As it was discussed in sec. 4.3, different kernel estimates (in magnitude) would be obtained, if a stimulus with a different operational range or step length was used. To illustrate this point, the first and second order kernel estimates obtained through the use of a ternary equirandom signal with A=.5, $\Delta t = .3$ sec. and T=4000 sec. are shown in Fig. 8.2.9. Also, the first and second order kernel estimates obtained through the use of a ternary equirandom signal with A=1, $\Delta t = .15$ sec. and T=4000 sec. are shown in Fig. 8.2.10. Both of these figures verify the anticipated results; i.e., in both cases the magnitude of the respective kernels is changed according to eqns. 4.3.12 and 4.3.13. Notice

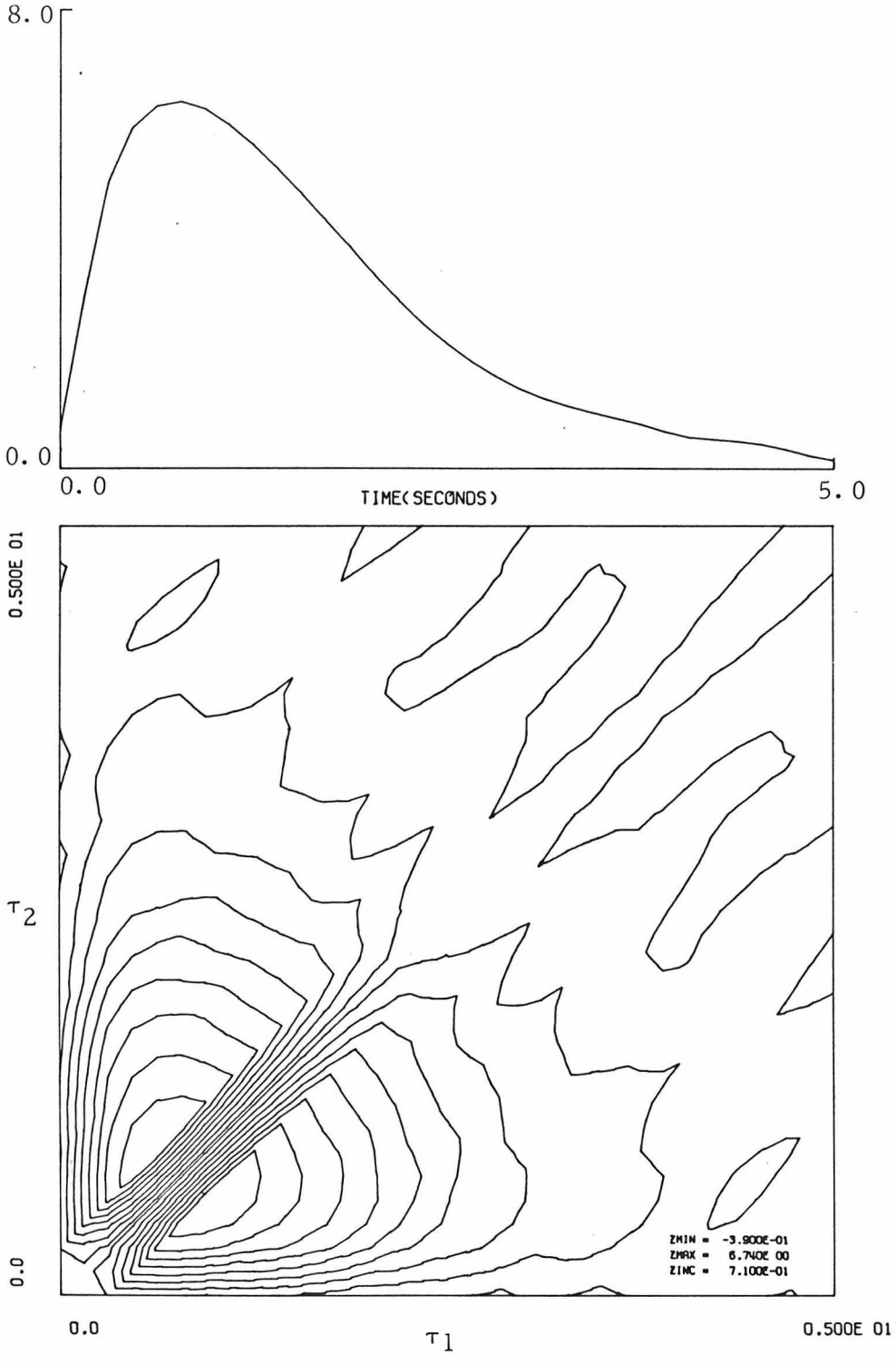


Fig. 8.2.2: First and second order kernel estimates obtained through the use of a 2-level equirandom CSRS.

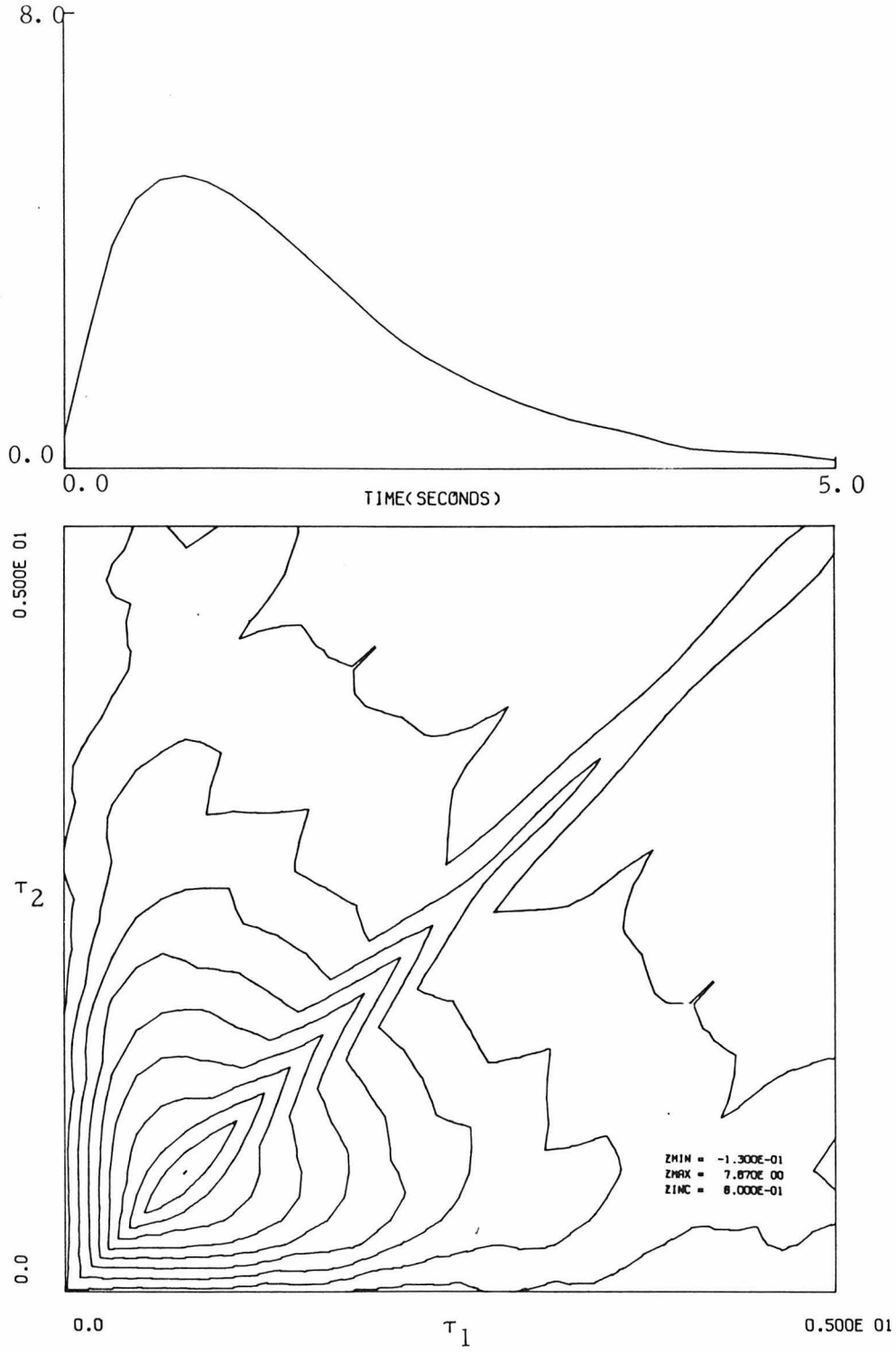


Fig. 8.2.3: First and second order kernel estimates obtained through the use of a 3-level equirandom CSRS.

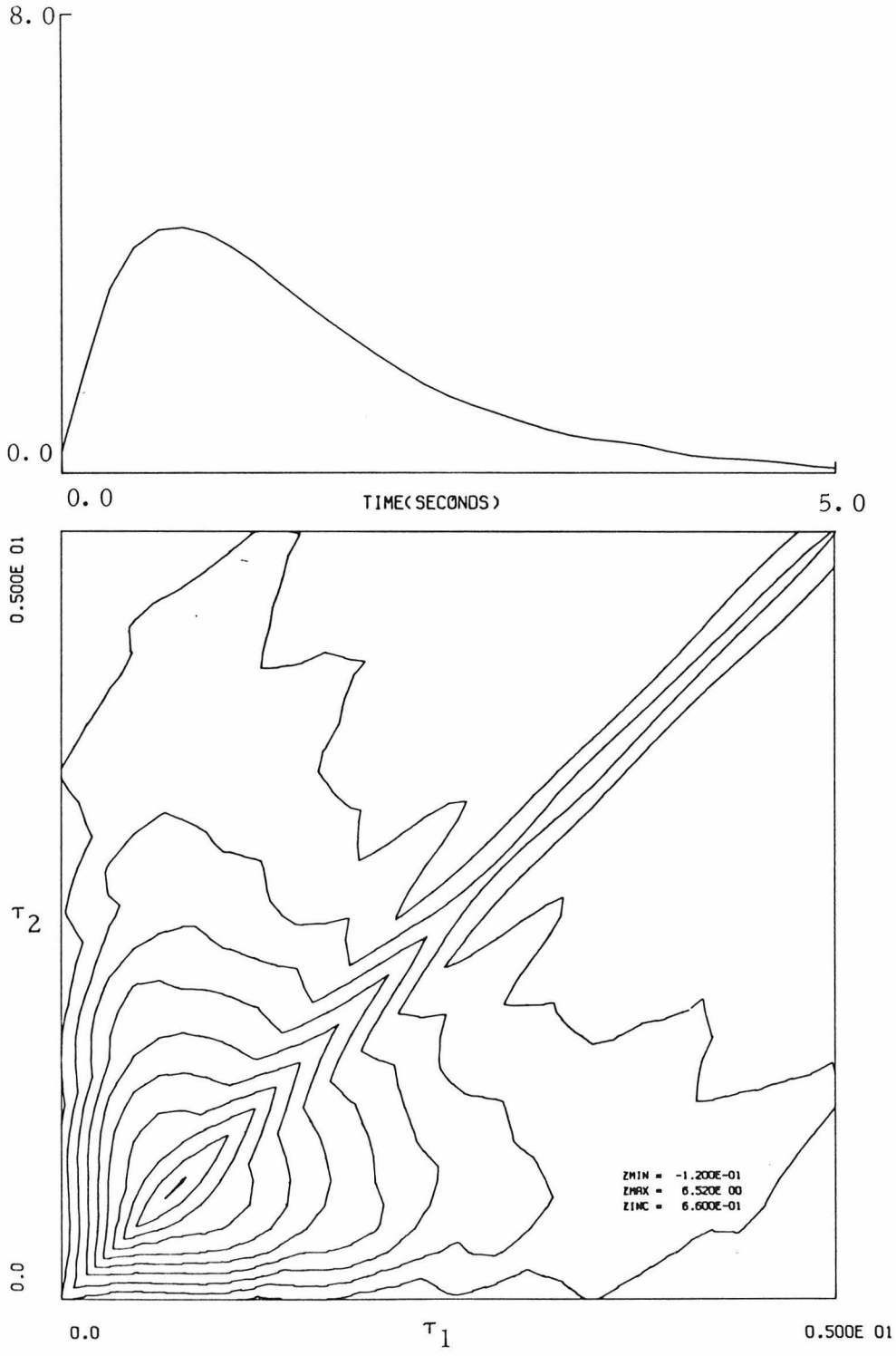


Fig. 8.2.4: First and second order kernel estimates obtained through the use of a 4-level equirandom CSRS.

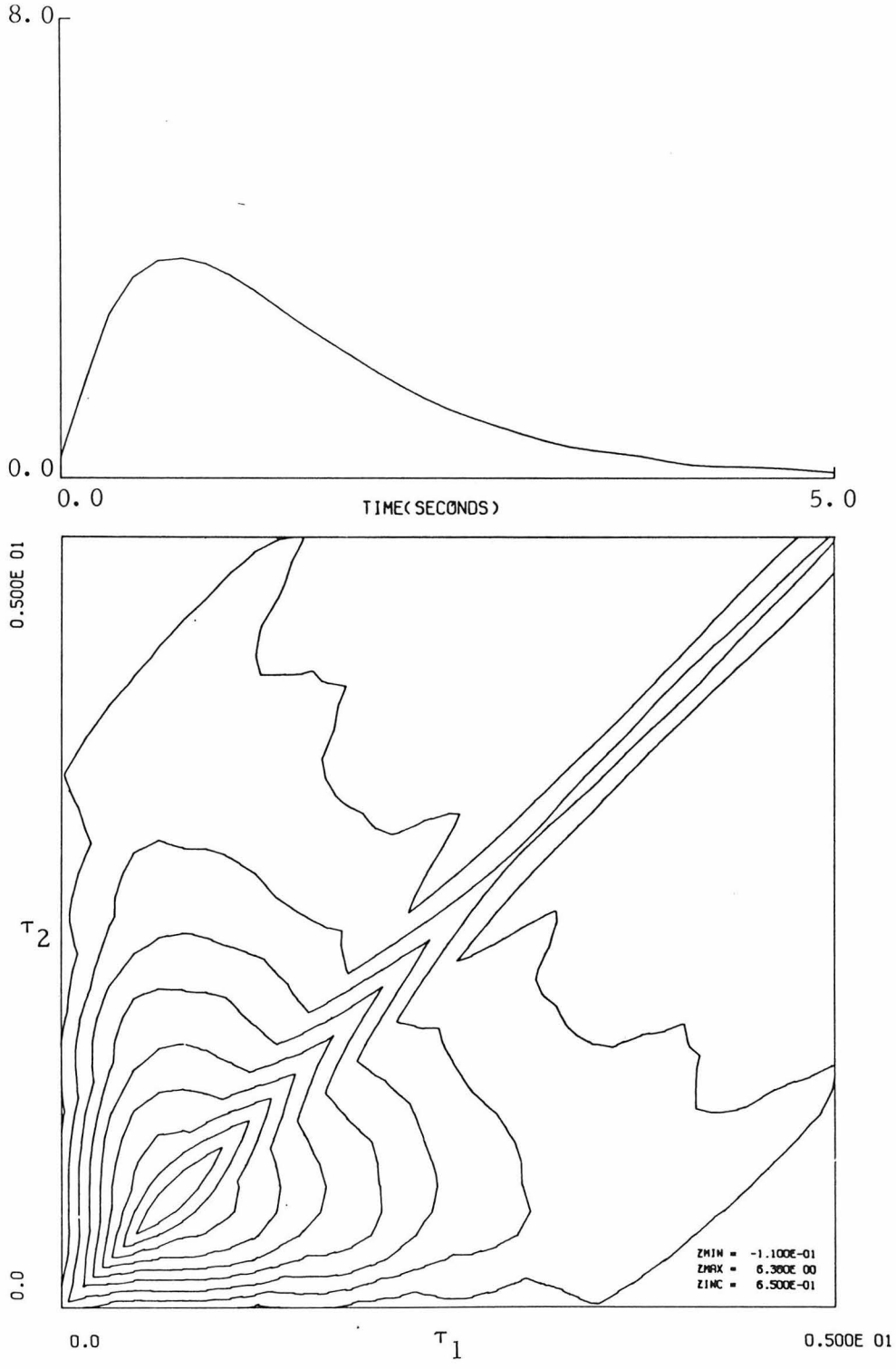


Fig. 8.2.5: First and second order kernel estimates obtained through the use of a 8-level equirandom CSRS.

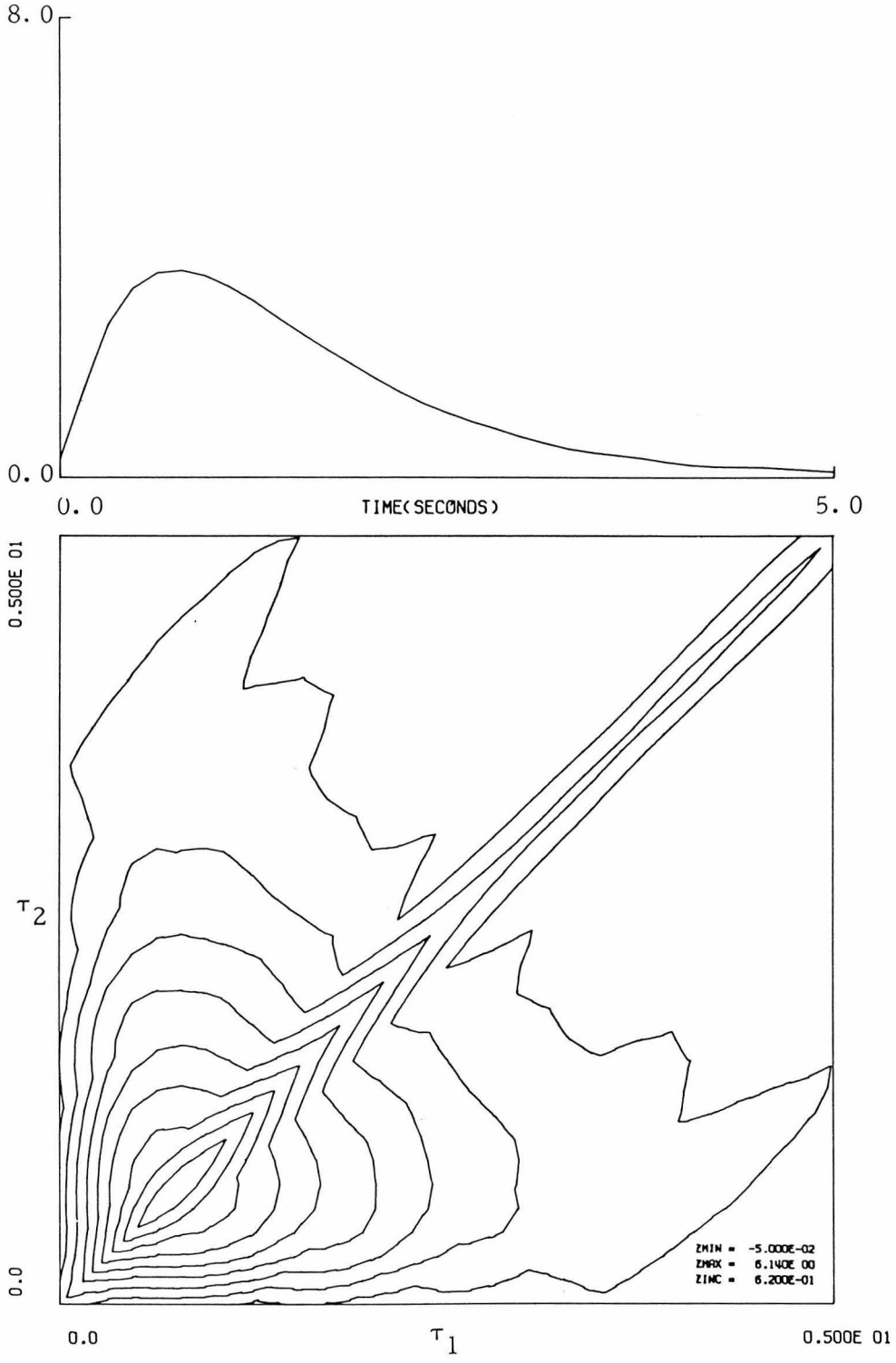


Fig. 8.2.6: First and second order kernel estimates obtained through the use of a 16-level equirandom CSRS.

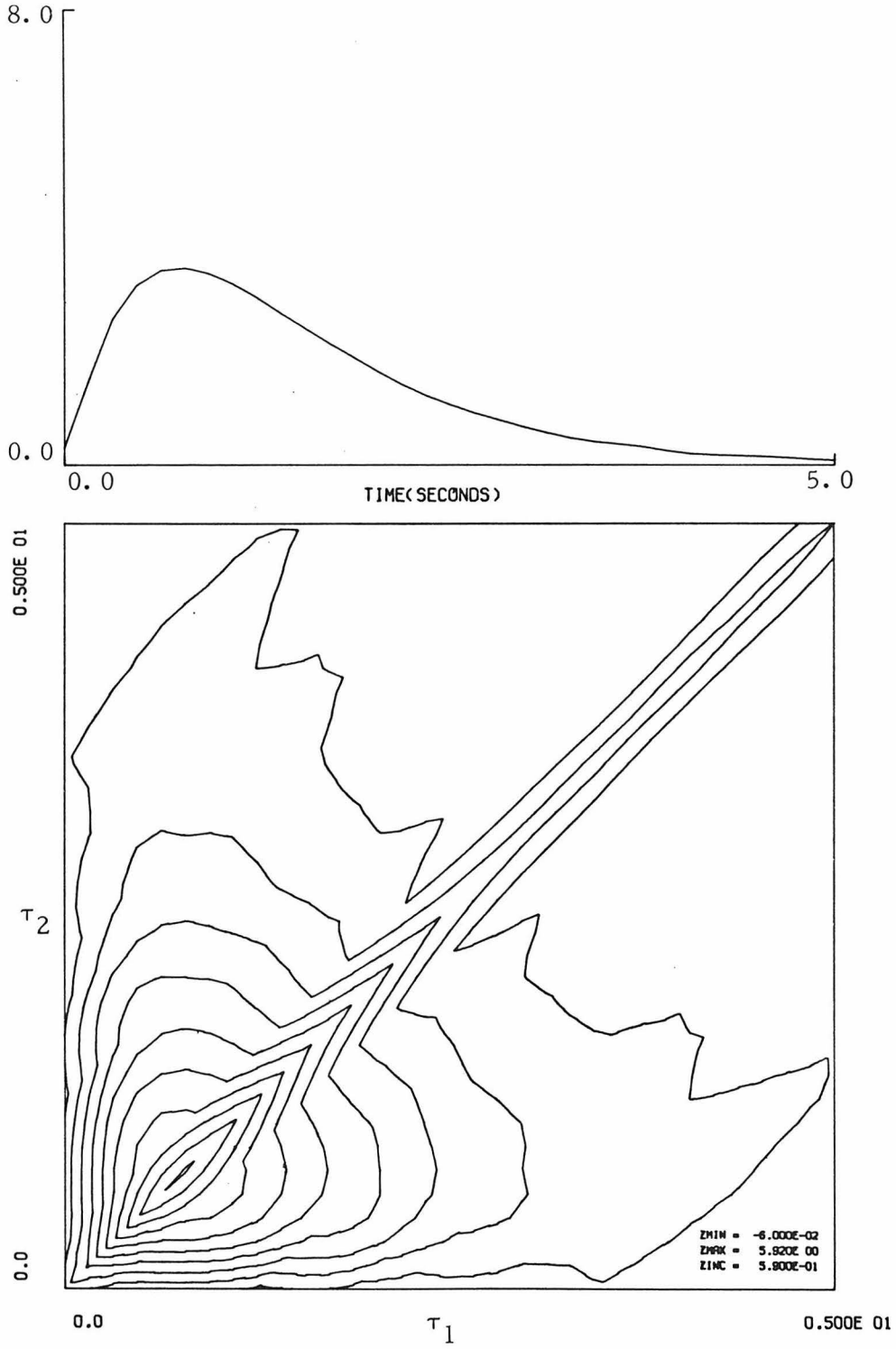


Fig. 8.2.7: First and second order kernel estimates obtained through the use of a 64-level equirandom CSRS.

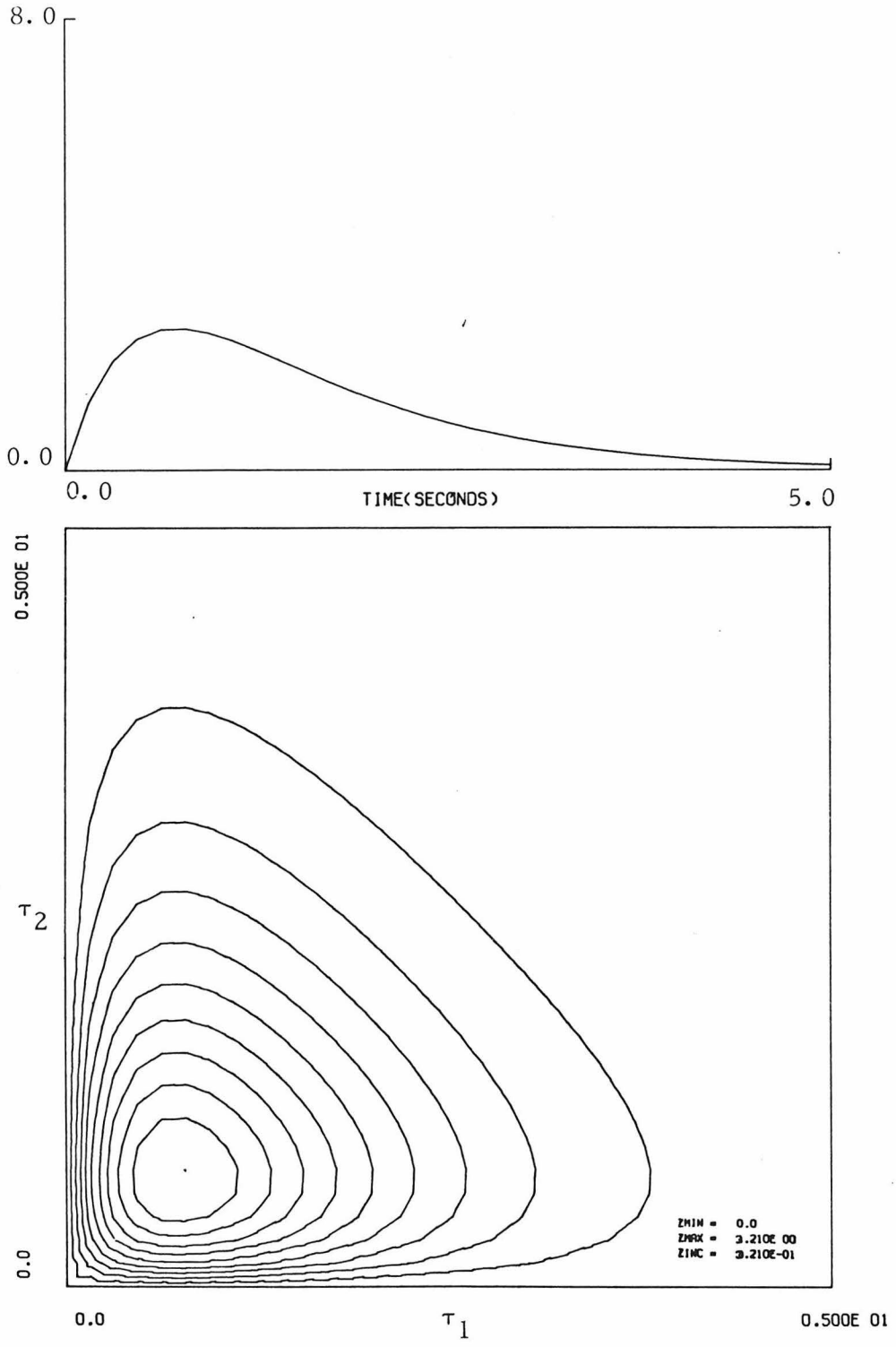


Fig. 8.2.8: First and second order Volterra kernels of the system in Fig. 8.2.1.

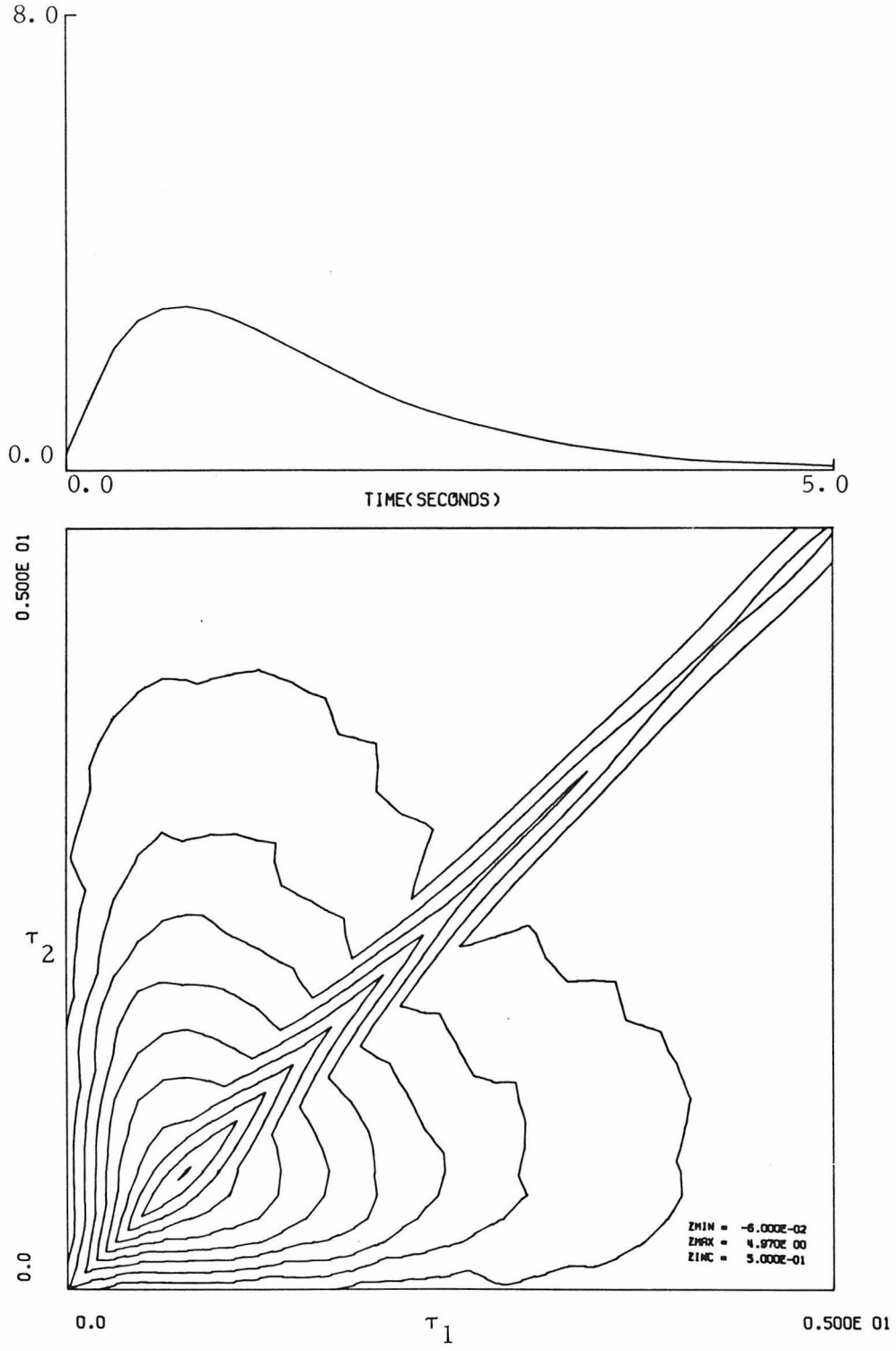


Fig. 8.2.9: First and second order kernel estimates obtained through the use of a 3-level equirandom CSRS with reduced operational range.

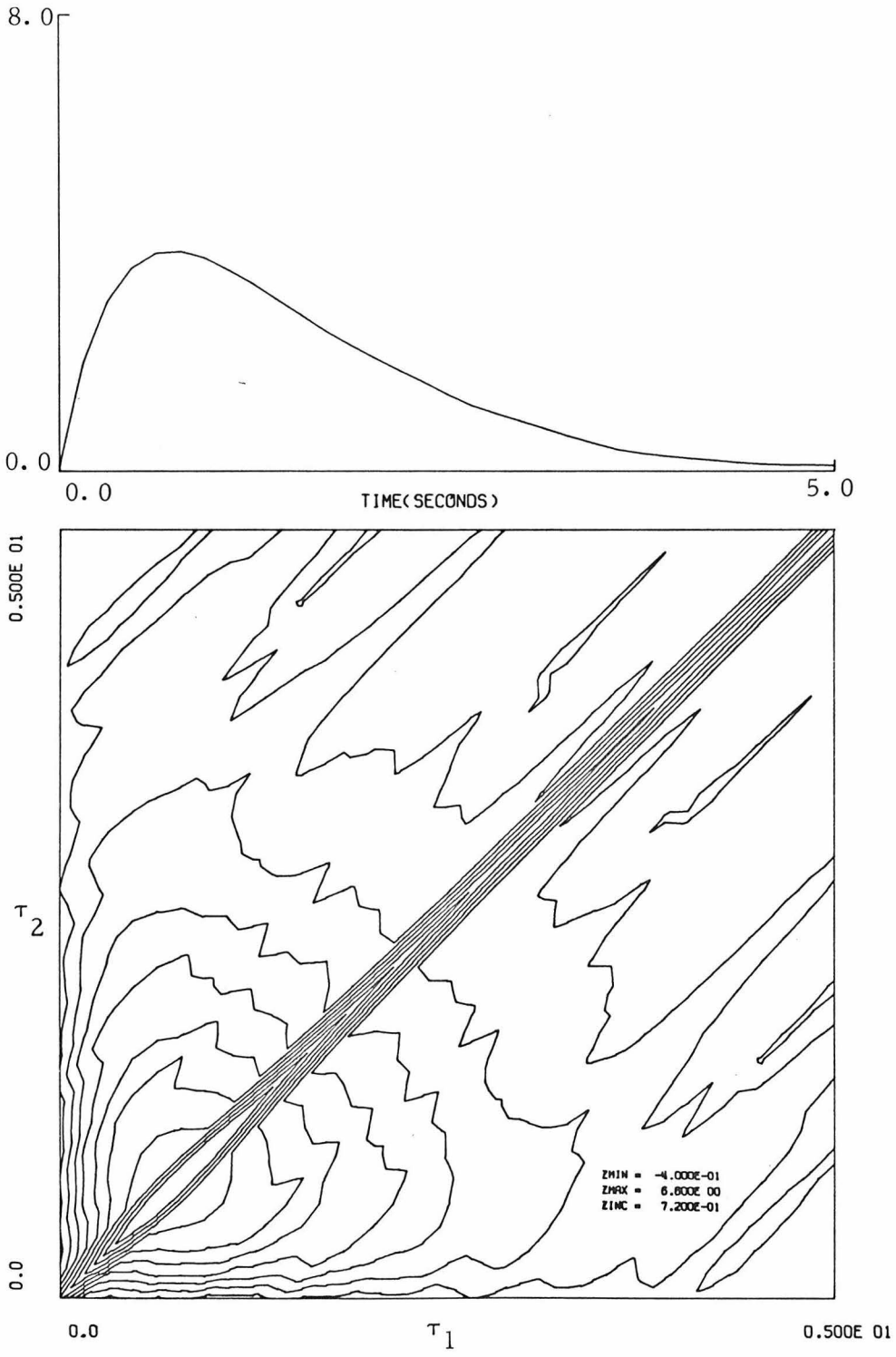


Fig. 8.2.10: First and second order kernel estimates obtained through the use of a 3-level equirandom CSRS with reduced step length.

that the power level in the first case is P=.05, while in the second case it is P=.10.

8.3 Study of the Accuracy of Truncated CSRS Estimated Models

As we discussed in the previous section the several CSRS kernels depend on the moments and the step length of the respective CSRS stimulus with which they are estimated.

As we also discussed in sec. 4.3 the modified functional series depend themselves on the moments and the step length of the associated CSRS.

Evidently, the convergence pattern of the modified functional series (and consequently the accuracy of the estimated truncated models) depends on the moments and the step length of the associated CSRS. It is a question of great practical interest in what way and to what extent the accuracy (always in the mean square error sense) of the model predicted responses is affected by the moments and the step length of the associated CSRS. We attempt to investigate this important point in this section by actually computing the accuracy of several model predicted responses and comparing the results. In the present study a third order nonlinear system has been considered, consisting of the cascade of a linear and a zero-memory nonlinear subsystem, as shown in Fig. 8.3.1.

The Volterra kernels of this system are:

$$k_0 = 0 \tag{8.3.1}$$

$$k_1(\sigma_1) = h(\sigma_1) \tag{8.3.2}$$

$$k_2(\sigma_1, \sigma_2) = h(\sigma_1)\, h(\sigma_2) \tag{8.3.3}$$

$$k_3(\sigma_1, \sigma_2, \sigma_3) = h(\sigma_1)\, h(\sigma_2)\, h(\sigma_3) \tag{8.3.4}$$

$$k_n(\sigma_1, \ldots, \sigma_n) = 0 \tag{8.3.5}$$

The zero, first and second order CSRS kernels of this system are (cf. sec. 4.3):

$$g_0 = (m_2 \Delta t) \int_0^\infty h^2(\tau) d\tau = \frac{16}{3} (m_2 \Delta t) \tag{8.3.6}$$

$$\begin{aligned} g_1(\sigma_1) &= h(\sigma_1) + 3(m_2 \Delta t) \int_0^\infty h^2(\tau) d\tau \cdot h(\sigma_1) + \left[\left(\frac{m_4}{m_2} - 3m_2 \right) \Delta t \right] h^3(\sigma_1) \\ &= [1 + 16(m_2 \Delta t)]\, h(\sigma_1) + \left[\left(\frac{m_4}{m_2} - 3m_2 \right) \Delta t \right] h^3(\sigma_1) \end{aligned} \tag{8.3.7}$$

$$g_2(\sigma_1, \sigma_2) = h(\sigma_1)\, h(\sigma_2) \tag{8.3.8}$$

$$g_3(\sigma_1, \sigma_2, \sigma_3) = h(\sigma_1)\, h(\sigma_2)\, h(\sigma_3) \tag{8.3.9}$$

$$g_n(\sigma_1, \ldots, \sigma_n) = 0 \qquad n = 4, 5, \ldots \tag{8.3.10}$$

where Δt is the step length and m_2, m_4 are the second and fourth moments of the associated CSRS.

In our present study, we will use several equirandom CSRS (introduced in the previous sections), for which:

$$m_2 = \frac{A^2}{3} \cdot \frac{M+1}{M-1} \tag{8.3.11}$$

$$m_4 = \frac{A^4}{15} \cdot \frac{(M+1)(3M^2-7)}{(M-1)^3} \tag{8.3.12}$$

where, A is the half operational range and M is the number of the signal levels.

We notice that the CSRS kernels of order higher than one are identical to the Volterra kernels of the system. This is due to the absence of nonlinearities higher than third order in the system.

It must be remembered that the diagonal values of the CSRS kernels have been estimated from the respective crosscorrelations by using different normalizing factors than in the case of nondiagonal values (cf. sec. 4.3). More specifically the diagonal values of the second order CSRS kernel have been normalized by the factor:

$$C_2^* = \frac{1}{(m_4 - m_2^2)\Delta t^2} \tag{8.3.13}$$

while the nondiagonal values of the same kernel have been normalized by the factor:

$$C_2 = \frac{1}{2(m_2 \Delta t)^2} \tag{8.3.14}$$

These factors are derived by evaluating the crosscorrelation with the respective modified CSRS functional for diagonal and for nondiagonal points. Notice that in the special case of a binary signal ($m_4 = m_2^2$) the diagonal values are directly evaluated as: $\hat{g}_2(\sigma,\sigma) = 0$ (cf. sec. 4.3).

In general, the normalizing factors of the diagonal values of the n-th order kernel may involve up to the 2n-th moment, while the normalizing factor for the nondiagonal values only involves the second moment.

The difference between the Volterra and CSRS kernels of zero and first order, manifests the different way in which the two functional

bases span the function space of the system responses. The Volterra functional series is an expansion based upon the functional derivatives of the system operator at the null function, while the CSRS functional series is a specially constructed expansion so to be orthogonal with respect to the associated CSRS.

Evidently, the CSRS expansion is expected to have stronger convergence for stimuli that assimilate the associated CSRS. This is illustrated in Table 8.3.1 where the m.s.e. of first and second order Volterra and CSRS model responses to the corresponding CSRS stimuli are compared.

Apparently, the CSRS models are spanning this function space more efficiently than the Volterra models; since (as indicated by eqns. 8.3.6 and 8.3.7) the zero and first-order CSRS functionals are probing into the space of the second and third order Volterra functionals as well.

The results shown in Table 8.3.1 were obtained by using equirandom CSRS with number of levels $M = 2, 3, 4, 8, 16, 64$ and 256. For all of these stimuli the step length was $\Delta t = .3$ sec, the sampling interval was $DT = .15$ sec, the half operational range was $A = 1$ stimulus unit and the record length was $T = 4000$ sec.

Notice that the decrease of the m.s.e. with the number of levels is not indicative of the "goodness" of the respective model but rather a logistic result due to the polynomial type of the zero-memory nonlinearity; because the polynomial operation tends to exaggerate the higher values, and the average squared magnitude of the signal that enters the zero-memory nonlinearity relates to the signal power level, which, in turn, relates to the number of levels (for constant operational range and step length).

TABLE 8.3.1

Percentage Mean Square Error of Model Predicted Responses

Number of levels of CSRS Stimulus	1st order CSRS model	1st order Volterra model	2nd order CSRS model	2nd order Volterra model
2	31.62	74.72	19.99	55.12
3	29.93	69.57	17.44	43.49
4	27.14	64.18	15.02	40.32
8	26.38	59.61	12.89	33.94
16	27.41	58.52	12.33	30.99
64	26.17	55.62	11.60	29.42
256	26.11	55.31	11.39	28.85

For this reason, we observe a decrease of the m. s. e. with increasing number of levels in both the CSRS and the Volterra models, even though it is milder in the case of the CSRS models. This happens because the CSRS models contain information from the third order Volterra functional as well, thus moderating the emphasis given to the higher signal values by the third degree term of the polynomial non-linearity.

Now that the difference in convergence between the CSRS and the Volterra functional series has been illustrated, the remaining important question is the relative accuracy of the several CSRS model responses to a stimulus other than the associated CSRS. In other words, given that our objective is to predict the system response to an arbitrary stimulus, what is the CSRS model that will give us the best accuracy.

Obviously, a CSRS model depends on the moments and the step length of the associated CSRS, since the CSRS kernel estimates, which make up the model, depend themselves on these moments and this step length.

The question thus becomes: what is the determining factor of the accuracy of the predicted response by a certain model to a given stimulus?

It is evident that the less the introduction of the given stimulus distorts the orthogonal arrangement among the estimated CSRS functionals, the better the accuracy of the model predicted response will be. Furthermore, the form of the CSRS functionals indicates that the "angles" (as defined in the theory of linear spaces from the inner product) between them depend on the autocorrelation functions of the stimulus involved.

Therefore, it is reasonable to assert that the accuracy of the predicted response by a certain CSRS model to a given stimulus depends upon the "proximity" of the autocorrelation functions of the specific stimulus with the autocorrelation functions of the CSRS associated with the model.

Apparently, if the system has up to n-th order nonlinearities, then only the autocorrelation functions of order up to the 2n-th affect the accuracy of the model predicted response.

The term "proximity" in our stated conclusion is not rigorously defined here for any arbitrary stimulus. However, it is understood that the criterion of proximity should be based on a metric norm of integrated difference (for example, the integral of the absolute value of the difference of the respective autocorrelation functions).

In any case, the reach of the present study is limited to a qualitative conclusion, which is valuable as an orientation instrument towards more advanced and detailed research on the subject that may follow.

To illustrate our stated conclusion,we consider a band-limited gaussian white noise stimulus (GWN) of band width 3.33 Hz and we compute the m. s. e. of the second order CSRS model response from the previously estimated models. The standard deviation of the GWN is $\sigma = 1$; therefore, its power level is $P = .15$ (cf. eqn. 3.2.22).

Considering the fact that the system under test (Fig. 8.3.1) is of third order, we conclude that the principal determining factor of the "proximity" of the autocorrelation functions of the GWN with respect to the ones of the CSRS is the power level. Consequently, we expect the

least m. s. e. of the model predicted response in the case of a CSRS with power level close to .15. In fact, the obtained results verify our expectation, as shown in Table 8. 3. 2.

To check this experimental result, the same series of experiments is repeated after changing only the standard deviation of the GWN to $\sigma = .75$. The power level of the GWN is now $P = .084$ and the least m. s. e. occurs again in the case which is anticipated according to our basic conclusion (see Table 8. 3. 2).

To reconfirm the same conclusion, we perform a series of experiments with several GWN stimuli of different power levels for the previously obtained model by using the binary equirandom signal. The obtained results are shown in Table 8. 3. 3 and they are in accordance with our basic conclusion.

To examine a case where the CSRS is not equirandom, we consider a CSRS with 255 levels and a triangular amplitude distribution profile. The model corresponding to this CSRS is tested against the model corresponding to an equirandom CSRS with 256 levels for two different cases. In the first case, the two CSRS have the same operational range $A = 1$; while in the second case, they have the same power level $P = .1$ (i. e., the same second moment and step length). In both cases, they have the same step length $\Delta t = .3$ sec.

The two models are compared for a GWN stimulus of bandwidth 3. 33 Hz and of several power levels. The results obtained in both cases are shown in Table 8. 3. 4 and they illustrate the basic assertion that the determining factor of the model accuracy with respect to a certain stimulus is the proximity of the autocorrelation functions (in this case the power levels) of the given stimulus and the

TABLE 8.3.2

Percentage Mean Square Error of 2nd Order CSRS Model
Predicted Response to GWN Stimulus

Parameters of CSRS model		Parameters of GWN stimulus*	
Number of levels of associated CSRS	Corresponding power level	P = .15 σ = 1	P = .084 σ = .75
2	.300	38.61	114.60
3	.200	21.90	58.90
4	.166	16.21	27.91
8	.128	16.17	16.91
16	.113	17.18	13.67
64	.103	18.27	12.05
256	.100	18.52	11.75

*P: power level of GWN.

σ: standard deviation of GWN.

TABLE 8.3.3

Percentage Mean Square Error of Binary Equirandom Model Response to Several GWN Stimuli

Parameters of GWN stimulus		M. s. e. of predicted response by binary model with corresponding power level .30
Standard deviation	Power level	
.75	.084	114.60
1.00	.150	38.61
1.25	.234	21.82
1.50	.337	24.19
1.75	.459	31.95
2.00	.600	40.45

TABLE 8.3.4

Percentage Mean Square Error of 2nd Order CSRS Model
Predicted Response to GWN Stimulus
For a Third Order Nonlinear System

GWN stimulus testing the model		Models estimated by CSRS with the same operational range (A = 1)		Models estimated by CSRS with the same power level (P = .1)	
Standard deviation	Power level	Uniform A. D. P. * CSRS model (P = .100)	Triangular A. D. P. * CSRS model (P = .050)	Uniform A. D. P. * CSRS model (P = .100)	Triangular A. D. P. * CSRS model (P = .100)
.25	.001	113.70	24.59	113.70	165.50
.50	.037	30.33	6.19	30.33	43.12
.75	.084	11.75	13.17	11.75	14.15
1.00	.150	18.52	27.47	18.52	17.30

*A. D. P. = Amplitude Distribution Profile.

CSRS associated with the model.

In conclusion, a crucial factor in choosing optimally the CSRS test signal with which to identify a nonlinear system is the environment of stimuli to which the system is exposed during its lifetime. The optimal choice is a CSRS that has autocorrelation functions "proximal" to the ones of the stimuli with respect to which we seek an optimum (in the m. s. e. sense) model.

This point has been repeatedly illustrated in the case of stochastic stimuli (see Tables 8. 3. 2, 8. 3. 3 and 8. 3. 4). Now, we will try to illustrate the same point in the case of deterministic stimuli.

For example, consider a pulse stimulus and compute the m. s. e. of the model response for the models that were previously estimated (i. e. the second order CSRS model of the third order system of Fig. 8. 3. 1). Changes of the pulse height and duration are also considered and their effect on the m. s. e. of the model predicted response is studied. The obtained results are shown in Table 8. 3. 5 and they confirm our basic conclusion. A decrease of the height or the duration of the pulse results in the move of the least m. s. e. position at a model corresponding to a CSRS with larger number of levels (and, consequently, with smaller power level).

Similar experiments are performed with sinusoidal stimuli of different amplitude and frequency. The obtained results are shown in Table 8. 3. 6, and they suggest once more our basic conclusion. A decrease of the amplitude or the period of the sinusoidal wave results in the move of the least m. s. e. position at a model corresponding to a CSRS with larger number of levels (and, consequently, with smaller power level).

TABLE 8.3.5

Percentage Mean Square Error of 2nd Order CSRS Model
Predicted Response to Pulse Stimuli
for a Third Order Nonlinear System

M	P	H = 1 D = 5	H = .5 D = 5	H = .5 D = 10
2	.300	26.50	5.77	2.23
3	.200	26.71	5.31	1.94
4	.166	37.46	4.23	3.69
8	.128	38.59	5.67	5.53
16	.113	39.69	6.90	6.70
64	.103	41.30	8.33	8.27

M: number of levels of CSRS associated with the model.

P: power level of CSRS associated with the model.

H: pulse stimulus height.

D: pulse stimulus duration.

TABLE 8.3.6

Percentage Mean Square Error of 2nd Order CSRS Model Predicted Response to Sinusoidal Stimuli for a Third Order Nonlinear System

M	P	F = .2 A = 1	F = .5 A = 1	F = .2 A = .75
2	.300	8.13	493.8	43.94
3	.200	7.21	249.9	11.93
4	.166	11.85	162.6	6.04
8	.128	16.15	94.63	4.79
16	.113	19.40	70.12	5.86
64	.103	21.94	53.71	7.06

M: number of levels of CSRS associated with the model.

P: power level of CSRS associated with the model.

F: sinusoidal stimulus frequency (in Hz).

A: sinusoidal stimulus amplitude.

In both cases of the deterministic stimuli a decrease in the energy of the signals results in better accuracy for the models associated with smaller power levels. This is in accordance with our basic conclusion on the accuracy of the model response to an arbitrary stimulus.

8.4 Dependence of the Deconvolution and Statistical Fluctuation Errors upon the Record and the Step Length of the CSRS Test Signal

In this section we will illustrate through computer simulated examples the dependence of the deconvolution (θ-error) and statistical fluctuation (ε-error) estimation errors upon the record length T and the step length Δt of a CSRS test signal. These errors have been analytically evaluated in secs. 5.1 and 5.2, and the present illustration aims at the justification of the assumptions made throughout those derivations.

Our special concern with these types of estimation errors springs from the fact that they determine in an actual case the optimum values of the two basic test parameters: the record length T and the step length Δt of the CSRS test signal.

It is of great practical importance that the validity (within reasonably small deviations) of the fundamental error equation 6.2.10 is manifested, since the design of the optimum test heavily relies on that formula.

At first, we illustrate the dependence of the θ-error upon Δt. Recalling eqn. 6.2.1 we have approximately:

$$\theta_n(\sigma_1, \ldots, \sigma_n) \cong \Delta t^2 \cdot D_{hn}(\sigma_1, \ldots, \sigma_n) \qquad (8.4.1)$$

where the approximation results from the assumption that D_{hn} does

not depend considerably upon Δt , for values of Δt of practical interest (i.e., Δt small). Notice that this relation holds strictly for points outside the initial region ($\sigma_i \geq \Delta t$), while for the points within the initial region the relation involves also terms of zero and first degree with respect to Δt (cf. sec. 5.1).

It has been also shown in sec. 5.1, that for values of Δt of practical interest, D_{h_n} depends only on the second partial derivatives of the respective kernel (cf. eqn. 5.1.14).

We illustrate this in Fig. 8.4.1 by plotting the deconvolution error of the first order kernel estimates of the system of Fig. 8.3.1 versus the second derivative of the kernel. The three plots a, b and c correspond to kernel estimates obtained by using ternary equirandom test signals of half operational range $A = 1$, record length $T = 8000$ sec and step lengths $\Delta t = .30$ sec, $.45$ sec and $.60$ sec respectively.

The linear form of the plots validates the assumption of independence of D_{h_1} from Δt ; and the slope of these lines demonstrates the square relation between θ_1 and Δt . Notice that the initial regions of the kernels have been excluded from these plots.

The dependence of the deconvolution error upon the second derivative and the step length can also be seen at the actual kernel estimates. In Fig. 8.4.2, the previously discussed first order kernel estimates are shown, along with the exact first order kernel of the system. It can be seen that the deviation of the estimates from the exact kernel is positive or negative according to whether the second derivative is positive or negative. However, this is not true in the initial region ($\sigma < \Delta t$) because of the derivative discontinuity of the kernel at the origin. Thus, the deviation of the estimates within the initial region is positive even though the second derivative is negative.

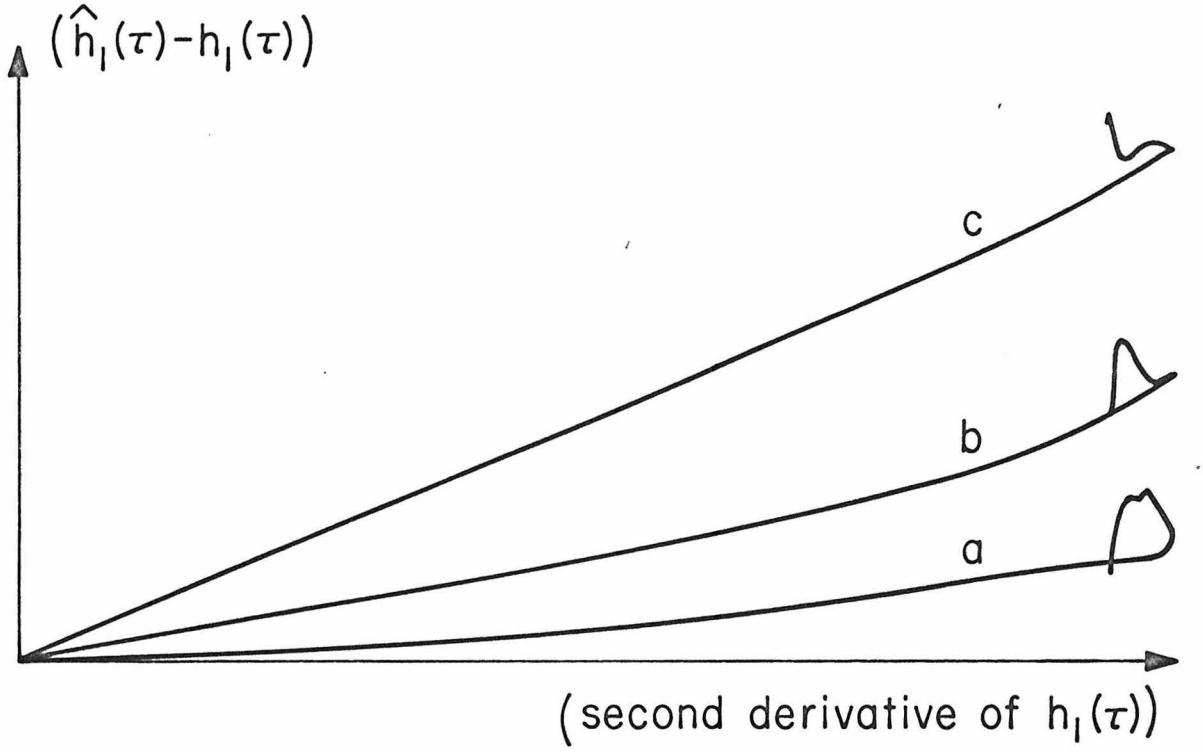


Fig. 8.4.1: Illustration of the dependence of the deconvolution error on the kernel second derivative.

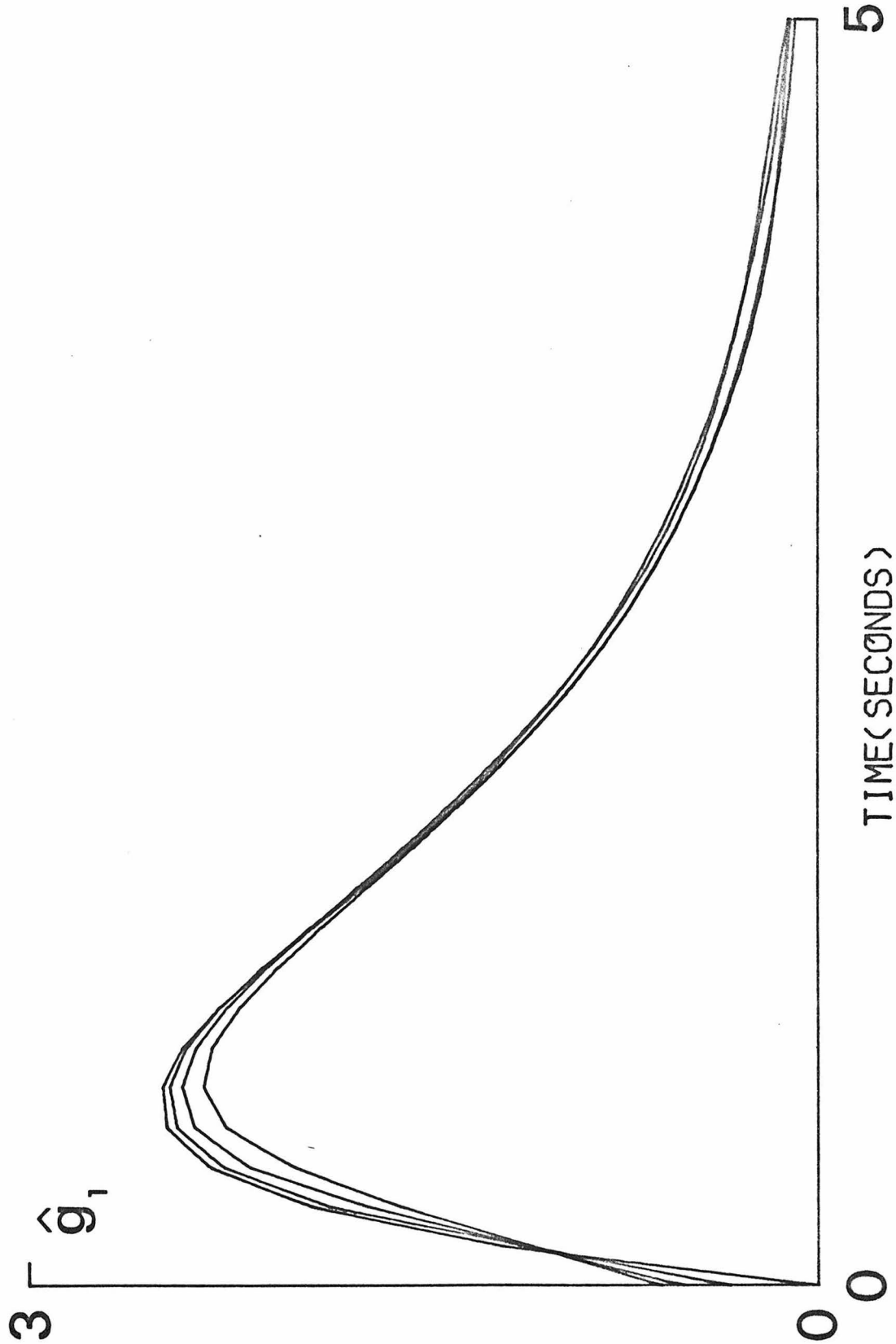


Fig. 8.4.2: Illustration of the deconvolution error in first order kernel estimates.

These simple illustrations demonstrate the validity of the assumptions made in connection with the deconvolution error in the design of the optimum test (sec. 6.5).

In order to illustrate the dependence of the statistical fluctuation error upon the record length T, we compute the percentage m. s. e. of the first and second order kernel estimates for several ternary equirandom stimuli with various record lengths.

More specifically, all the ternary equirandom stimuli used have half operational range $A = 1$ and step length $\Delta t = .15$ sec. We compute the average percentage m. s. e. using six estimates for each one of three different record lengths: 500 sec, 1000 sec and 2000 sec.

The obtained results are shown in Fig. 8.4.3 along with some representative kernel estimates, and they verify our expectations. Notice that the total estimation error can be approximately identified with the statistical fluctuation error because of the considerably small step length $\Delta t = .15$ sec.

In order to illustrate the dependence of the statistical fluctuation error upon the step length, we follow the following procedure:

(a) We obtain four independent kernel estimates of first and second order for each one of four different step lengths: .15 sec, .30 sec, .45 sec and .60 sec; while keeping the length of each one of these records constant at 2000 sec.

(b) We compute the integrated squared differences between any two of the four kernel estimates of the same order and step length. Notice that these integrated squared differences are quantities independent from the deconvolution error and they only depend on the statistical fluctuation error.

(c) We compute the averages of these quantities for each group cor-

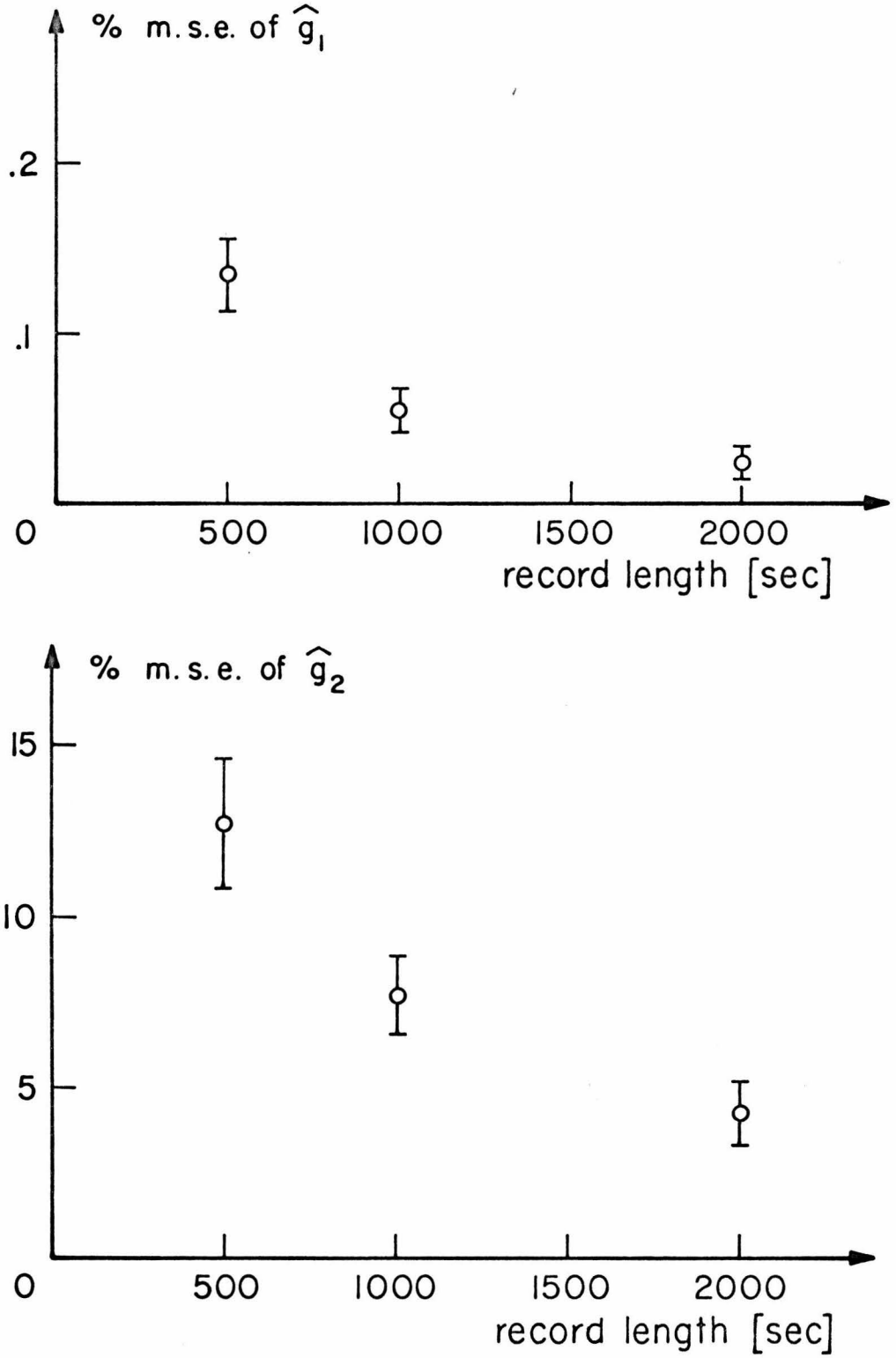


Fig. 8.4.3a: Illustration of the dependence of the CSRS kernel estimation error on the record length.

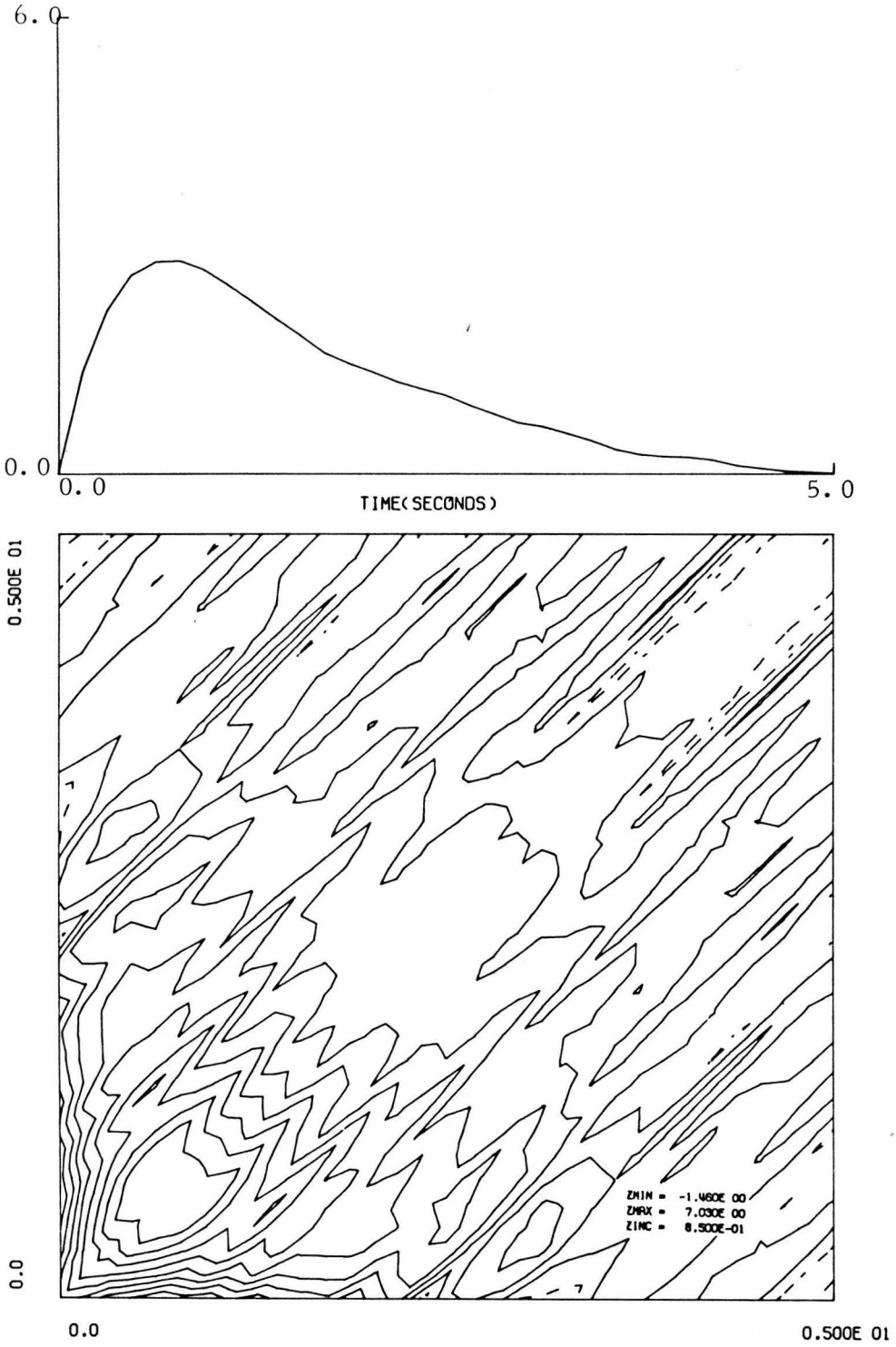


Fig. 8.4.3b: First and second order kernel estimates obtained from a 500 sec long record.

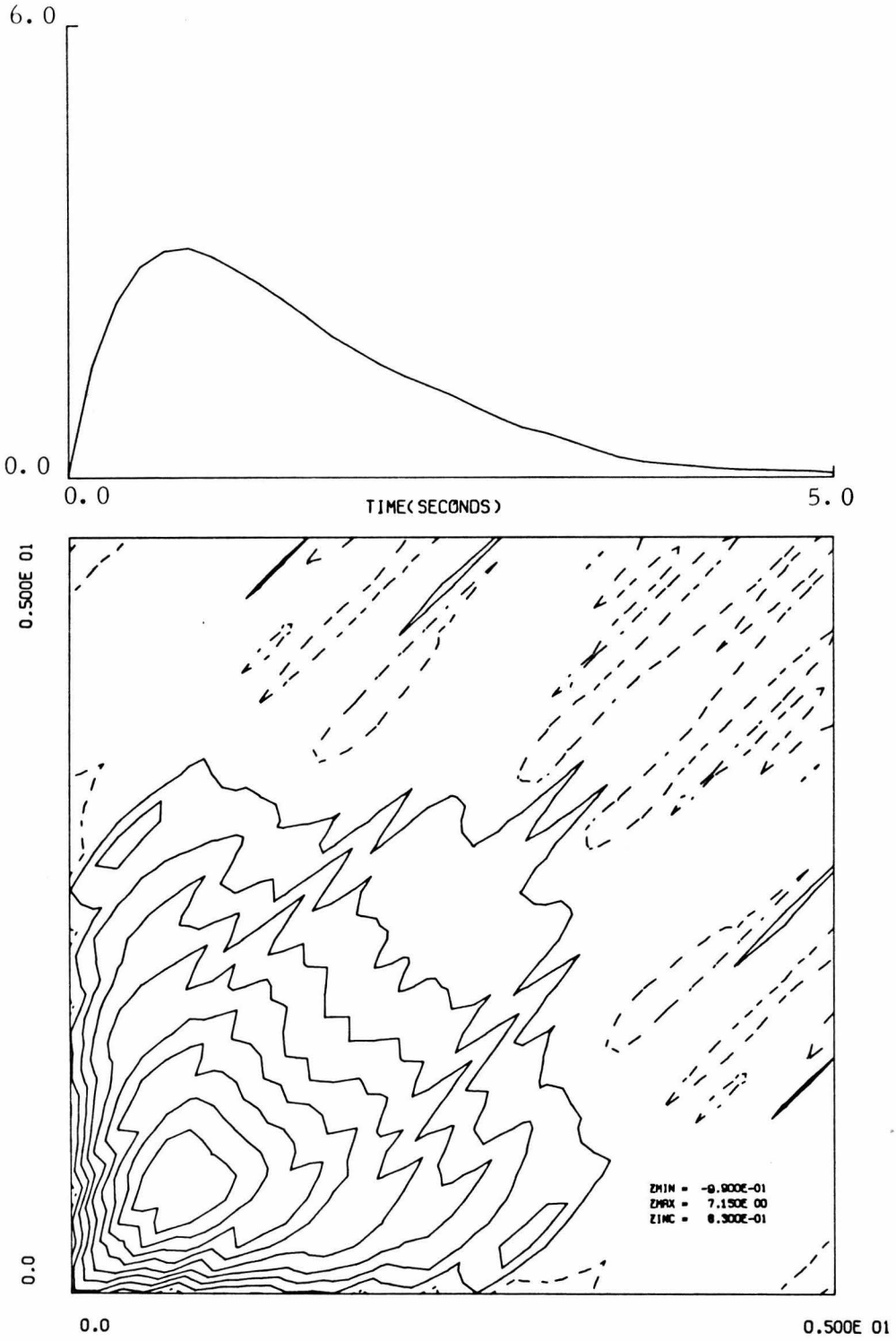


Fig. 8.4.3c: First and second order kernel estimates obtained from a 1000 sec long record.

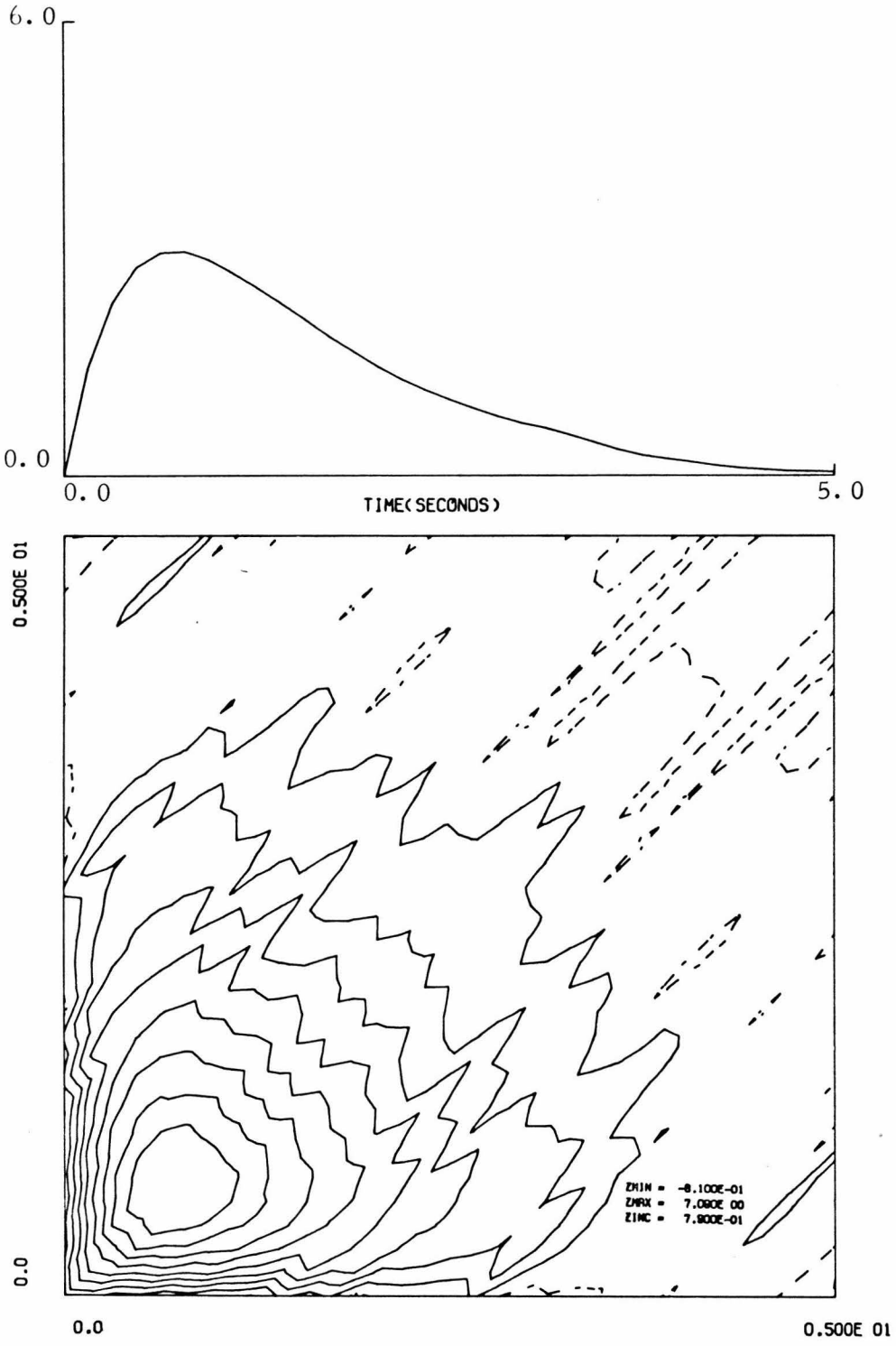


Fig. 8.4.3d: First and second order kernel estimates obtained from a 2000 sec long record.

responding to a specific step length.

The obtained results are shown in Fig. 8.4.4, along with representative second order kernel estimates, and they demonstrate the anticipated dependence of the statistical fluctuation error on the stimulus step length (cf. eqn. 6.2.10).

With these simple illustrations, we demonstrated the validity of the approximating assumptions made in connection with the statistical fluctuation error in the design of the optimum test (sec. 6.5).

8.5 Illustration of the Fundamental Error Equation Curves

In this section, we will illustrate the form of the fundamental error equation (FEE) curves, which were discussed in sec. 6.2, in an actual case of identification of a computer simulated nonlinear system.

To this purpose, we consider a cascade system consisting of a linear subsystem followed by a second degree polynomial zero-memory nonlinearity, as shown in Fig. 8.5.1 Our intention is to compute the mean square error of the CSRS kernel estimates as a function of the test parameters Δt and T, and verify the analytical relation described by the FEE (cf. eqn. 6.2.10).

We recall that the total estimation error for the n-th order kernel consists of two significant parts. The one is due to the deconvolution error and depends on Δt, and the other is due to the statistical fluctuation error and depends on Δt and T. The quantitative relation of all these is described by the FEE.

In the present example, the system has up to second order nonlinearities. Therefore, the m. s. error of the zero, first and second order kernel estimates must be used to determine the optimum test parameters. In fact, the optimization procedure takes, in practice,

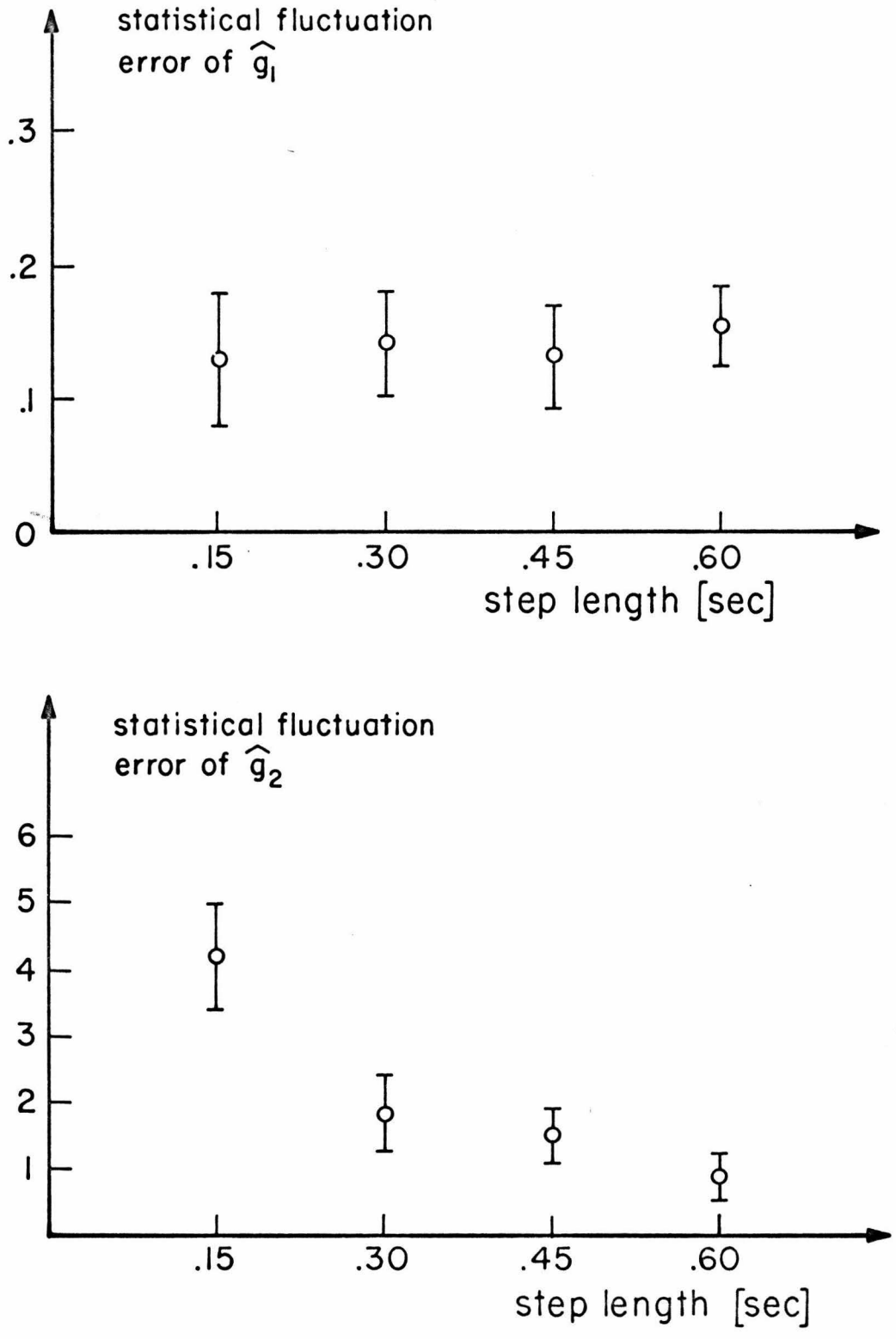


Fig. 8.4.4a: Illustration of the dependence of the statistical fluctuation estimation error on the step length.

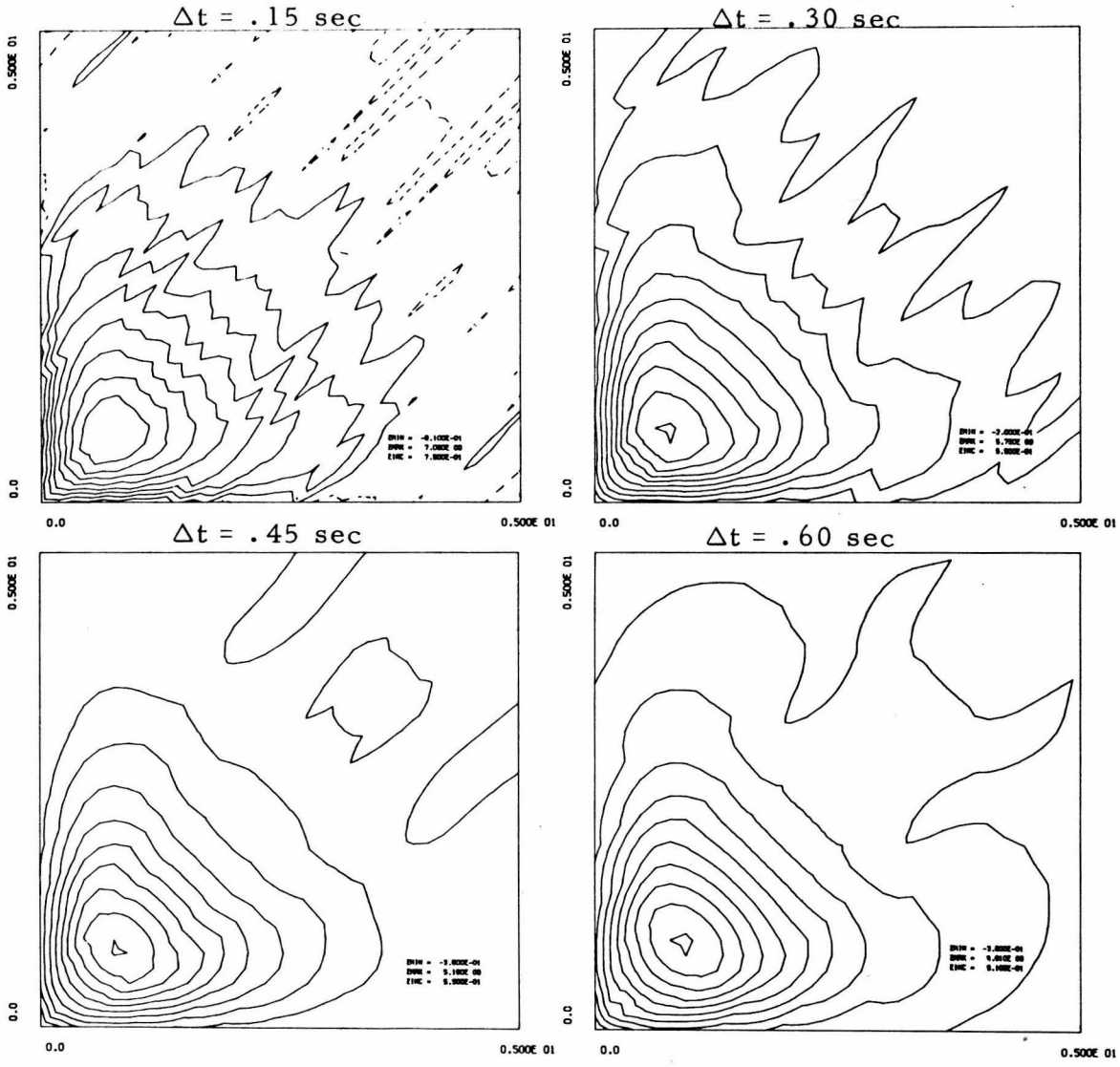


Fig. 8.4.4b: Second order CSRS kernel estimates obtained through stimuli of various step lengths.

the form of determining the optimum Δt for a given T (cf. sec. 6.2).

There are two distinct ways to do that. The one is to determine the optimum Δt by using the total estimation error for all the kernels. The other is to determine a different optimum Δt for each individual kernel estimate by using only the corresponding estimation error. It must be emphasized that in the second case (where we use stimuli with different Δt to estimate each one of the kernels) the operational ranges of the several stimuli must be adjusted in such a way that the power levels of all these stimuli are the same and, consequently, they can be used compatibly in the same model.

In Fig. 8.5.2, we show the results obtained by computer simulations of the system of Fig. 8.5.1, using ternary equirandom CSRS stimuli of several record lengths (T = 500, 1000 and 2000 sec) and step lengths (Δt = .15, .30, .45 and .60 sec). The FEE curves for the zero, first and second order kernel estimates are shown along with the curve corresponding to the total estimation error. The total estimation error is computed as the weighted sum of the errors corresponding to the individual kernels:

$$Q_{tot} = Q_0 + m_2 Q_1 + m_2^2 Q_2 \qquad (8.5.1)$$

where m_2 is the second moment of the CSRS stimulus (cf. eqn. 8.6.5).

The data points corresponding to record lengths of 500 sec are computed as averages from eight independent kernel estimates. The data points corresponding to record lengths of 1000 sec and 2000 sec are computed as averages from six and four independent kernel estimates respectively. This is justified on the basis of the fact that the statistical variation of the kernel estimates decreases as the corresponding stimulus record length increases.

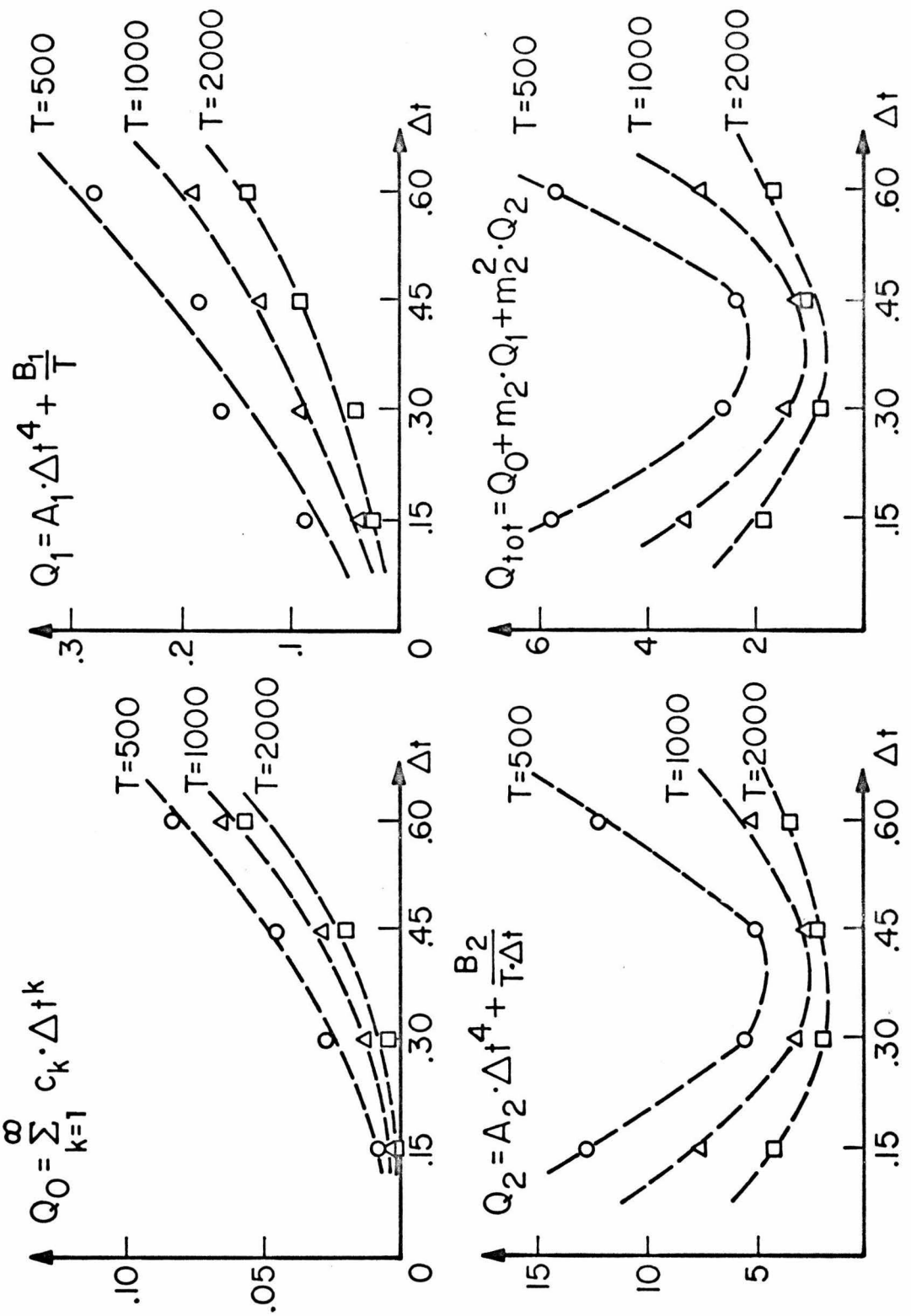


Fig. 8.5.2: Illustration of the FEE curves of the system in Fig. 8.5.1.

The FEE curves shown in Fig. 8.5.2 illustrate the validity of the analytical relations derived in sec. 6.2, within a degree of tolerable deviation due to the statistical fluctuation of the data points.

8.6 Determination of the Optimum Step and Record Length

The determination of the optimum step and record length is done on the basis of the fundamental error equation (FEE), as it was discussed in secs. 6.2 and 6.5.

According to the FEE, the knowledge of the constants A_n and B_n of a given system suffices for the determination of the optimum pair $[\Delta t, T]$, which provides the best (in the m.s.e. sense) estimate of the n-th order kernel of the system. Therefore, the determination of the optimum pair $[\Delta t, T]$ for the n-th order kernel estimate simply requires the determination (estimation) of the constants A_n and B_n.

This is strictly true for kernel estimates of order higher than the first. In the case of the first order kernel, the use of the FEE in the simplified form of eqn. 6.2.10 would result in determining a zero optimum Δt. This is, of course, unrealistic because zero Δt implies zero power level for the signal. Consequently, we have to have, in practice, a finite nonzero Δt, which cannot be smaller than a lower bound determined by signal-to-noise ratio considerations of the specific system at hand.

On the other hand, the "constants" A_1 and B_1 are not strictly constants. In sec. 6.2, they were considered approximately constant for practical convenience, and this approximation is a valid one (for all practical purposes) as long as we deal with kernels of order higher than one. However, in the case of the first order kernel, and due to the absence of Δt from the denominator of the statistical fluctua-

tion error term, the slight dependence of B_1 on Δt assumes a considerable importance in determining the optimum Δt for a given T.

Thus, we can seek to determine in this case the optimum pair $[\Delta t, T]$ on the basis of the relations:

$$4A_1\Delta t^3 + \frac{1}{T}\frac{\partial B_1}{\partial \Delta t} = 0 \qquad (8.6.1)$$

$$12A_1\Delta t^2 + \frac{1}{T}\frac{\partial^2 B_1}{\partial \Delta t^2} > 0 \qquad (8.6.2)$$

Notice that the terms involving derivatives of A_1 have been deleted as being insignificant compared to the other two terms (cf. sec. 6.2). The actual evaluation of the first and second partial derivatives of B_1 with respect to Δt is not a very complicated task, as it is outlined in sec. 5.2. Of course, their evaluation is based on a kernel estimate instead of the exact kernel, and it is done numerically from finite differences, therefore, it involves several inaccuracies. However, these inaccuracies can be significantly reduced if proper care is taken in the evaluation procedure; in terms of a fairly accurate kernel estimate, a sufficiently small sampling interval and a good numerical differentiation procedure.

Apparently, eqns. 8.6.1 and 8.6.2 may not yield a positive solution for Δt and then the optimum Δt is determined on the basis of signal-to-noise ratio considerations.

In the case of the zero order kernel estimate, the FEE does not hold in the form of eqn. 6.2.10. As it was discussed in sec. 5.2, the statistical fluctuation error for $\hat{g}_0$ depends upon Δt in a polynomial fashion involving all powers of Δt of degree higher than zero. The deconvolution error has still the same form as in the FEE.

The study of the dependence of the $\hat{g}_0$ estimation error upon Δt becomes in practice quite involved, while its impact on the total error of the model is of secondary importance. For this reason, the determination of the optimum pair $[\Delta t, T]$ is done, in practice, without taking into consideration the estimation error of the zero order kernel. In any case, the analytical relations for that type of error are given in sec. 5.2, and they can be utilized if a certain application requires so.

A very important remark must be made in connection with the several optima Δt which correspond to the several individual kernels. As it was discussed in sec. 4.3, the CSRS kernels depend on the moments and the step length of the respective CSRS stimulus. More specifically, if we call the expression:

$$P_n = m_{2n} \cdot \Delta t^n \qquad (8.6.3)$$

the "generalized n-th order power level" of a CSRS, then we can state that: the CSRS kernels depend upon the generalized power levels of the respective CSRS of all orders up to the order of nonlinearity that the system possesses.

The implication of this fact, in connection with the different optima Δt that we can determine for each individual kernel of the system, is that we have to adjust the moments of our CSRS in such a way that the generalized power levels remain the same despite the changes of the step length.

In other words, if different Δt are to be used to estimate the several kernels (which means that different stimuli and experiments must be performed in each case), then the probability density function

or the operational range of the CSRS must be modified in such a way that the required generalized power levels remain the same in all cases. This is absolutely necessary for the several kernel estimates to be compatible within the same model.

Clearly, the simplest way to readjust the moments of the CSRS according to the changing step length is by simply multiplying the signal with a scalar:

$$\beta = \frac{x_i(t)}{x_j(t)} = \sqrt{\frac{\Delta t_j}{\Delta t_i}} \qquad (8.6.4)$$

In that case, the operational ranges are different. If we want the operational ranges to be the same in all cases, then a more sophisticated modification procedure is required regarding the probability distributions of the signals. Evidently, such a procedure would also depend on the order of nonlinearity of the specific system at hand. Because of the apparent complications of such a procedure, we usually use the same step length for all kernel estimates, if the operational range is needed to remain constant.

Naturally, in that case, we determine the optimum pair $[\Delta t, T]$ from the total estimation error of all kernels in the model and not from the estimation error of individual kernels.

It must be noted that the total estimation error must be computed as a weighted sum of the estimation errors of the individual kernels. The weights of this sum reflect the actual contribution of the respective functional terms into the total model response.

Thus, in case we want to optimize the model response for a quasi-white stimulus, these weights become some metric of the stimulus dependent part of each functional term. If a Euclidean metric is adopted, then these weights become:

$$W_n = \int_0^\infty \cdots \int_0^\infty x^2(t-\tau_1) \cdots x^2(t-\tau_n)\, d\tau_1 \cdots d\tau_n \qquad (8.6.5)$$

Evidently, the values of these weights are expressible in terms of the second moment of the CSRS stimulus $x(t)$:

$$W_n \cong m_2^n \qquad (8.6.6)$$

To illustrate the determination of the optimum pair $[\Delta t, T]$, we will follow the optimum test procedure for the system of Fig. 8.5.1 under the requirement of having the same operational range (and consequently the same step length) for all orders of kernels. Thus, the determination of the optimum pair $[\Delta t, T]$ will be done on the basis of the total estimation error. In fact, we will only consider the estimation errors of the first and second order functional terms; thus deleting the zero order term, which is of secondary importance and may complicate the optimization procedure because of its distinct pattern of dependence upon Δt (as previously discussed).

In the evaluation of the total estimation error, the second order term will be weighted by m_2 according to eqn. 8.6.6. Therefore, the total weighted estimation error in this case will be:

$$Q_{tot} = (A_1 + m_2 \cdot A_2)\Delta t^4 + \frac{1}{T}\left[B_1 + \frac{m_2 B_2}{\Delta t}\right] \qquad (8.6.7)$$

And the optimum step length for a given record length T is:

$$(\Delta t)_{opt} = \left[\frac{m_2 B_2}{4(A_1 + m_2 \cdot A_2) \cdot T} \right]^{\frac{1}{5}} \qquad (8.6.8)$$

The optimization procedure aims at estimating the constants A_1, A_2, and B_2 in this case. This can be done in a variety of ways, always based upon the form of the FEE and the nature of the deconvolution and statistical fluctuation error (cf. sec. 6.2).

In this illustration, we will consider the record length T as the quantity "in scarcity" (cf. sec. 6.5) and, consequently, the optimum Δt will be determined from eqn. 8.6.8.

The methods that we will use to estimate the constants A_1, A_2 and B_2 in this example are more practicable than the general methods presented in sec. 6.5. The estimation of A_1 and A_2 requires the previous estimation of B_1 and B_2 respectively.

The estimation of B_1:

We compute the first order kernel estimates from several independent records of three different lengths. In this case, we used eight independent records of 500 sec long; six independent records of 1000 sec long; and four independent records of 2000 sec long; with step length .3 sec and half operational range 1 stimulus unit.

For each group of kernel estimates of the same record length, we compute the integral of the squared difference of any two of them.

We find the mean of these values and we consider it an estimate $\hat{\epsilon}_1$ of the quantity $2B_1/T$.

We map these estimates on a $\log \epsilon_1$-$\log T$ chart, and we draw a least squares fitted line with 45 degrees negative slope. This line is

a graphic approximation of the analytical relation:

$$\log \epsilon_1 = \log (2B_1) - \log T \qquad (8.6.9)$$

which results from the FEE.

Finally, we estimate B_1 from any convenient point of this line (see Fig. 8.6.1). Considering the point where $\epsilon_1 = .01$, we have $T \cong 10400$ and consequently: $\hat{B}_1 \cong 52$.

The estimation of B_2:

We compute the second order kernel estimates from the same records as previously, and we compute similarly the integrals of the squared differences of any two kernel estimates within each group. The mean of these values for each group is an estimate $\hat{\epsilon}_2$ of the quantity $2B_2/T\Delta t$.

We map these values on a $\log \epsilon_2$-$\log T$ chart, and we draw a least squares fitted line with 45 degrees negative slope. This line is a graphic approximation of the analytical relation:

$$\log \epsilon_2 = \log \left(\frac{2B_2}{\Delta t} \right) - \log T \qquad (8.6.10)$$

which results from the FEE.

Finally, we estimate B_2 from any convenient point of this line (see Fig. 8.6.2). Considering the point where $\epsilon_2 = 1$, we have $T \cong 4800$ and consequently: $\hat{B}_2 \cong 720$.

The estimation of A_1:

We compute four groups of first order kernel estimates corresponding to four different step lengths $\Delta t = .15$ sec, .3 sec, .45 sec and .6 sec. Each group comprises four independent kernel estimates and the length of each independent record is $T = 2000$ sec. The half operational range is again $A = 1$ stimulus unit.

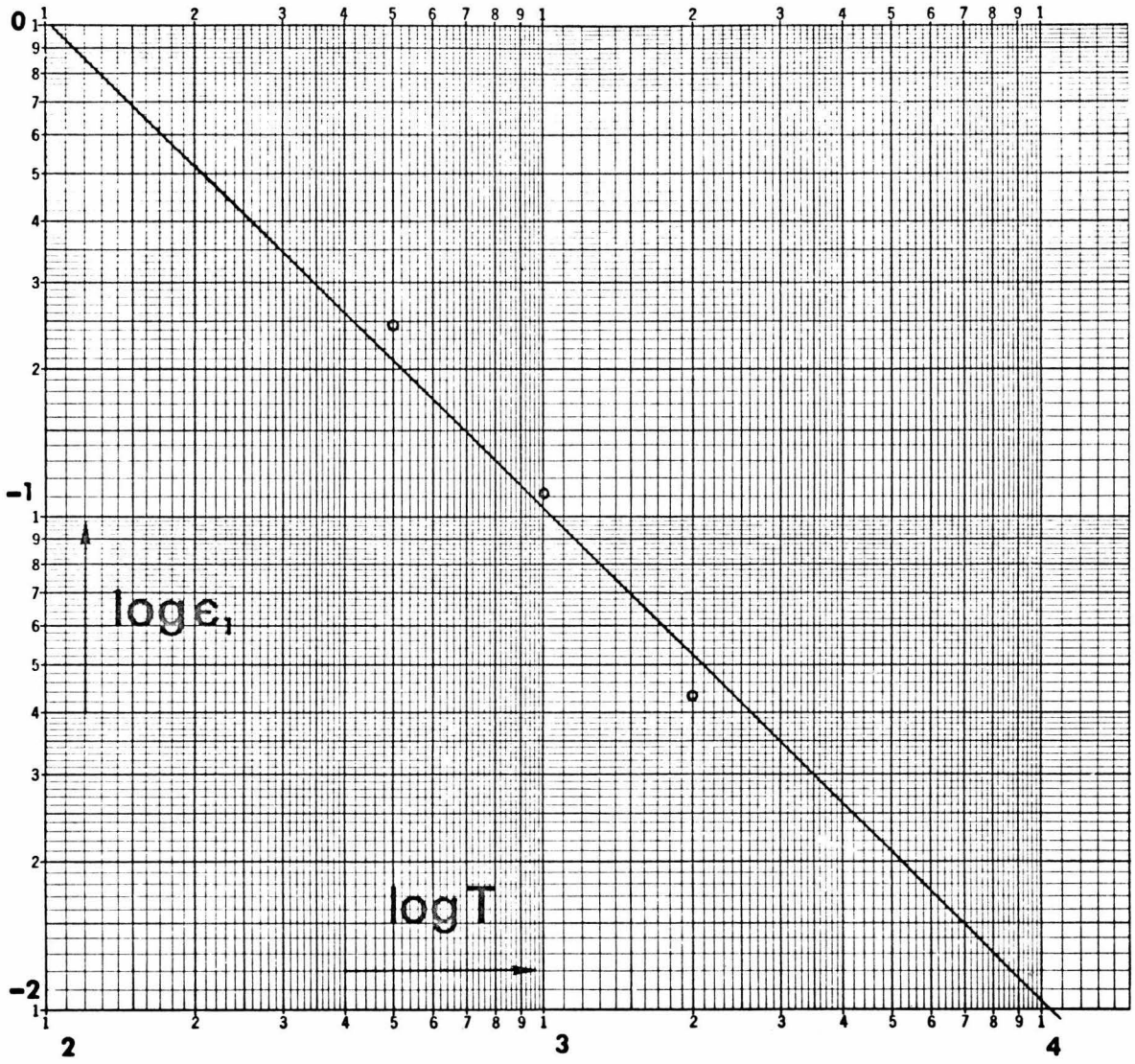


Fig. 8.6.1: Graphic estimation of B_1 .

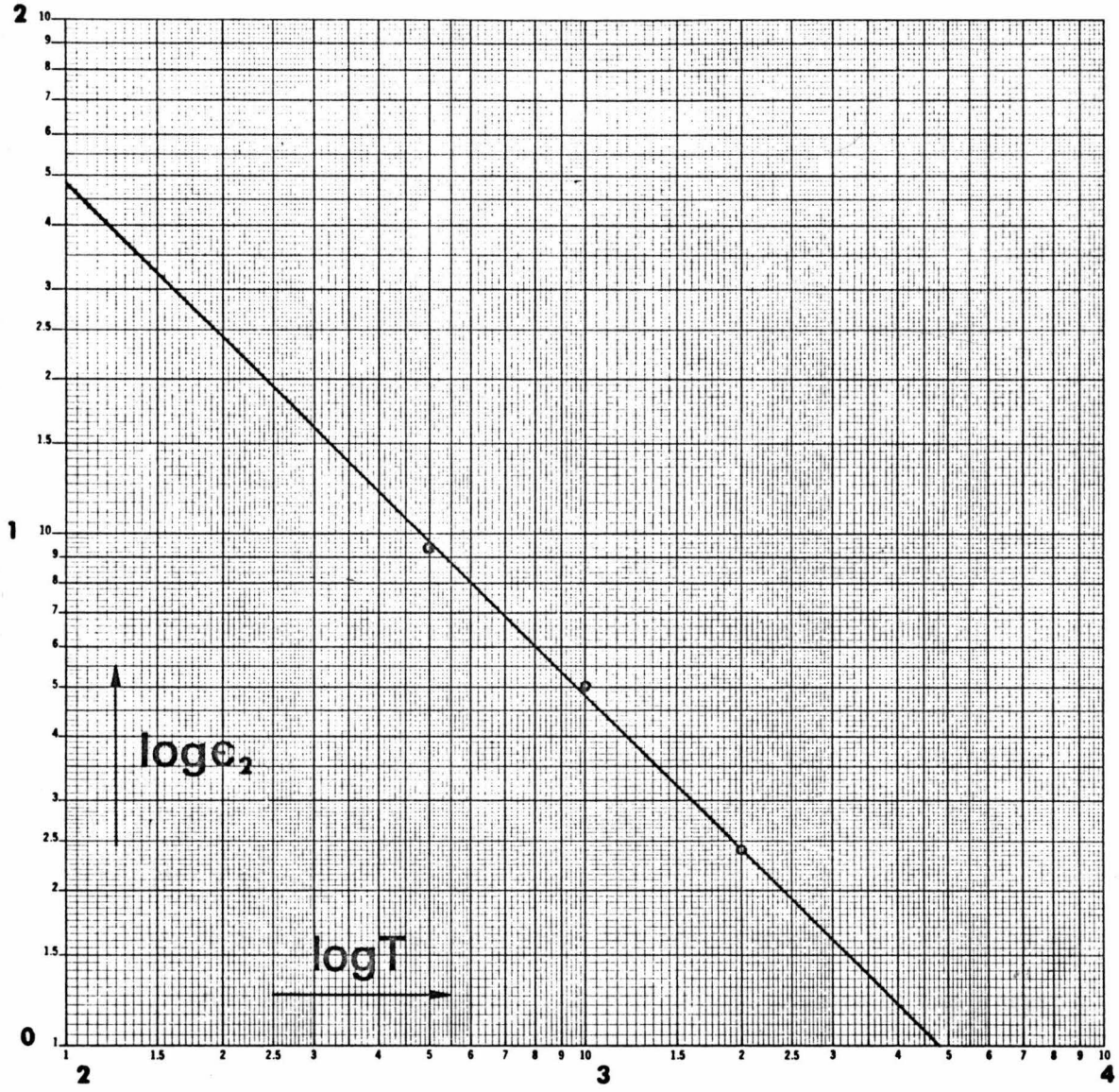


Fig. 8.6.2: Graphic estimation of B_2 .

We compute the integrals of the squared differences between any two members of the first and each one of the rest three groups successively; thus forming three (4X4) matrices of values.

We compute the mean of the entries of each one of these three matrices. The computed means are:

$$\mu_1 = .058 \quad ; \quad \mu_2 = .108 \quad ; \quad \mu_3 = .167$$

They correspond to estimates of the quantity:

$$\mu = A_1 \Delta t_1^4 (1-\beta^2)^2 + \frac{2B_1}{T} \qquad (8.6.11)$$

in three different cases where $\beta = 2$, 3 and 4, respectively; while $\Delta t_1 = .15$ sec and $T = 2000$ sec. This is following directly from the FEE.

Now, using the value of B_1 that we previously estimated, we obtain three estimates of A_1 from eqn. 8.6.11. Finally, we compute the average of these three estimates and take that average as an estimate of A_1: $\hat{A}_1 \cong 1.34$.

Apparently, a graphical solution could be also employed by drawing the line of the quantity μ versus $(1-\beta^2)^2$ (cf. eqn. 8.6.11).

The estimation of A_2:

We repeat the same sequence of computations as before, using the second order kernel estimates. The obtained means of the entries of the three matrices are:

$$\mu_1 = 4.54 \quad ; \quad \mu_2 = 6.28 \quad ; \quad \mu_3 = 8.16$$

They correspond to estimates of the quantity:

$$\mu = A_2 \Delta t_1^4 (1-\beta^2)^2 + \frac{B_2}{T \cdot \Delta t_1} \left(1 + \frac{1}{\beta}\right) \qquad (8.6.12)$$

for three different cases where $\beta = 2$, 3 and 4 respectively; and $\Delta t_1 = .15$ sec, $T = 2000$ sec. This is again a direct derivation from the FEE.

Using the value of B_2 that we previously estimated, we obtain three estimates of A_2 from eqn. 8.6.12. The average of these three estimates is taken to be an estimate of A_2: $\hat{A}_2 \cong 115$.

A graphical solution could be also employed (like in the case of $\hat{A}_1$) provided that the record length is also changed along with the step length, in such a way that the product $(T \cdot \Delta t)$ remains constant.

Having estimated the constants A_1 , A_2 and B_2 , the determination of the optimum step length is a simple task using eqn. 8.6.8. In the present illustration $m_2 = \frac{2}{3}$ and consequently:

$$(\Delta t)_{opt} \cong \frac{1.09}{T^{1/5}} \qquad (8.6.13)$$

Thus, the optima Δt for several record lengths are:

T (sec)	$(\Delta t)_{opt}$ (sec)
500	.314
1000	.275
2000	.238
4000	.209
8000	.181

It must be noted that the effect of the first order estimation error upon the determination of the optimum Δt is negligible, since

A_1 is two orders of magnitude smaller than A_2. In fact:

$$\frac{m_2 A_2}{A_1} \cong 58 \quad .$$

The outlined methods in this section for the determination of the optimum Δt are not in any sense unique or most efficient. The statistical analysis of data is a wide discipline comprising a variety of advanced and sophisticated techniques, which can be possibly employed in this case to provide more efficient or more accurate procedures for the determination of the optimum Δt.

The basic idea presented in this section is the use of the FEE for the determination of the optimum Δt.

In fact, the term "determination" is somewhat an exaggeration since the evaluated optimum Δt is only an approximation based upon the obtained estimates of the constants A_1, A_2 and B_2. Therefore, the material presented in this section must be regarded with a certain degree of statistical tolerance and granted the potentiality of procedural improvements.

CHAPTER IX

APPLICATION OF CSRS IN PHYSIOLOGICAL SYSTEM IDENTIFICATION

In this chapter, we will illustrate the application of CSRS in identification of a real system. The system that we have chosen is a physiological one. More specifically, it is part of the visual neuron chain of the fly Calliphora Erythrocephala with white eyes. This system exhibits the basic characteristics, which made the crosscorrelation technique suitable for its identification; as it is discussed in the following section.

9.1 The Physiological System: Light-Potential Transducer in the Photoreceptor Cell of Calliphora Erythrocephala

The physiological system under study is the very initial portion of a visual neuron chain, which operates as light-potential transducer in the photoreceptor cell 1-6 of the fly Calliphora Erythrocephala with white eyes [36].

The input of the system is temporally varying light intensity applied upon the eye of the fly and the output is a slow (continuous) potential recorded in the photoreceptor cell body.

The function of this system comprises the transformation of the photon energy to a slow potential through the isomerization of rhodopsin. Our effort aims at describing the functional characteristics of this physico-chemical process through mathematical models of the Volterra-Wiener type estimated by the use of CSRS.

This physico-chemical process takes place within the membranes of microvilli, which constitute the rhabdomere of the photoreceptor. The rhabdomere constitutes the site of the physiological processes that occur within the system.

In this study, we do not utilize any information about the physiological or anatomical properties of the rhabdomere, but we attempt to describe its function by regarding the rhabdomere as a black-box, and formulating the problem as a system identification one.

In order to be able to employ the crosscorrelation technique in the identification of this system, we must show that the system satisfies the three basic requirements of the Volterra series expansion (cf. sec. 2.3); namely, the stationarity, finite-memory and analyticity requirements.

As far as the stationarity of the system is concerned, we test the stationarity hypothesis using the "run test" for the system response to a stationary gaussian process, as described in sec. 6.5.1. The outcome of the test is that we do not reject the stationarity hypothesis; however, this does not mean that we confirm it either. In fact, we chose the Calliphora Erythrocephala with white eyes instead of the one with red eyes, in order to avoid possible significant nonstationarities in the system behavior resulting from the presence of shielding pigment (light intensity adaptation).

The finite-memory of the system is checked through the computation of the autocorrelation of the system response to a quasi-white stimulus (cf. sec. 7.2); and in a final check through the inspection of the system kernels themselves.

The analyticity of the system is assumed on the basis of the

expectable differential smoothness of a physical system, and it can only be checked indirectly through the obtained identification and synthesis results.

Finally, it is found that the chosen physiological system can be satisfactorily described (within the testing stimulus range) by only a second order CSRS model, and consequently the required computational burden for our study is of reasonable size.

9.2 Estimated CSRS Models of the Physiological System

In this section, we will illustrate the use of CSRS in the identification of nonlinear systems. The obtainment of "legitimate" system models through the use of CSRS (and in connection with the crosscorrelation technique) demonstrates the quasi-whiteness of the CSRS. The "legitimacy" of a system model is judged upon its ability to predict with satisfactory accuracy the response of the system to a given stimulus.

The CSRS, that we are going to use in this illustration, are six members of the equirandom CSRS group that we presented in chapter 8. As a reminder: the members of this group are multi-level CSRS with a uniform probability distribution profile. These stimuli have various numbers of levels (namely, 2, 3, 4, 8, 16, 64 levels), the same operational range, the same step length and the same record length.

We stimulate the physiological system described in the previous section with temporally varying light intensity stimuli modulated in the CSRS fashion. We record the response of the system and we compute the zero, first and second order kernels of the system under test.

The test parameters that we have used for all the stimuli are:

Step length	$\Delta t = .005$ sec.
Record length	$T = 100$ sec.
Sampling interval	$Dt = .001$ sec.
Half operational range	$A = 1.85$ Volts (measured through a photo-cell)
First order kernel memory	$\mu_1 = .05$ sec.
Second order kernel memory	$\mu_2 = .05$ sec.

The obtained CSRS kernel estimates for the CSRS stimuli with 2, 3, 4, 8, 16, 64 levels are shown in Fig. 9.2.1 through 9.2.6 respectively, along with portions of the stimulus, the system response and the model predicted response.

The corresponding m. s. errors of the model predicted responses are given in table 9.2.1. Clearly, the accuracy of all these estimated models is comparable, and the small differences in the corresponding m. s. errors of the model predicted responses are probably of statistical nature.

An interesting remark is that the size of the CSRS kernel estimates does not change significantly as the number of levels (and consequently the generalized power levels) change. This indicates the minor role of higher than the second order nonlinearities in the system.

Another important remark concerning the physiological system under study is that any cross section of a second order kernel estimate obtained through a quasi-white test signal with gaussian amplitude distribution along the τ_1 or the τ_2 axis has a form similar to the first order kernel estimate, being simply multiplied by a negative scalar

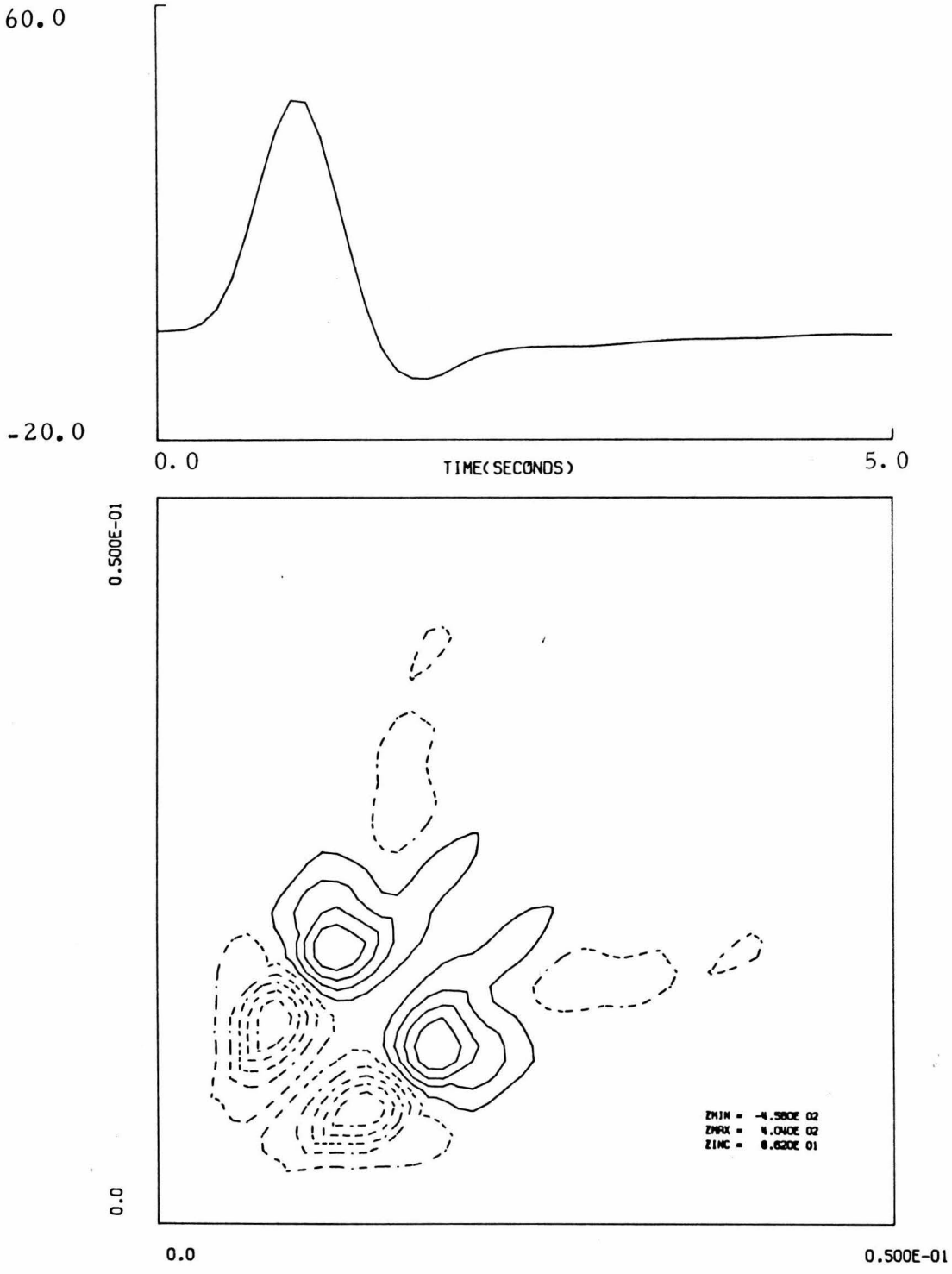


Fig.9.2.1a: First and second order kernel estimates obtained through a 2-level equirandom CSRS (continued)

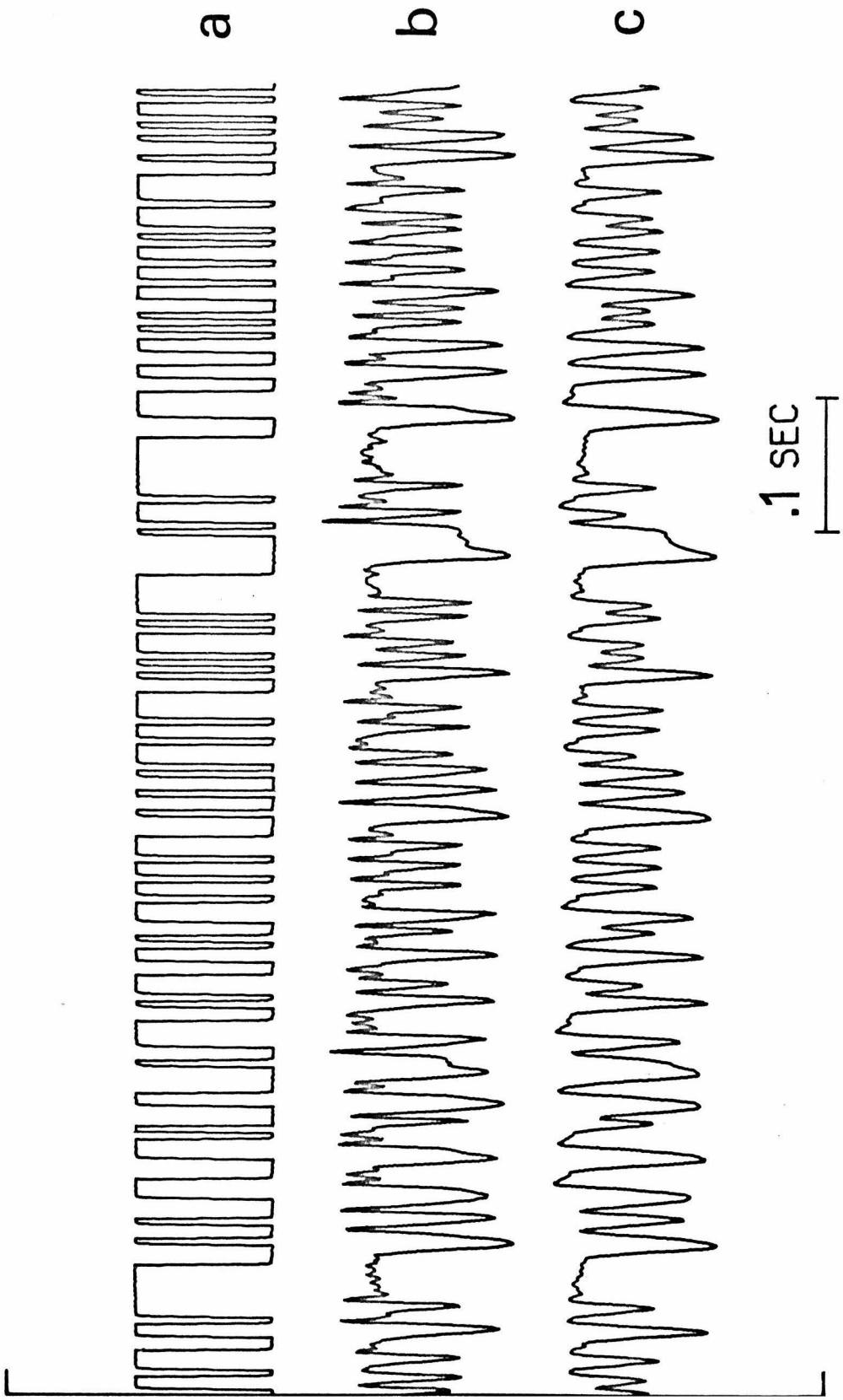


Fig. 9.2.1b: Stimulus (a), response (b) and model predicted response (c) for the 2-level equirandom CSRS case

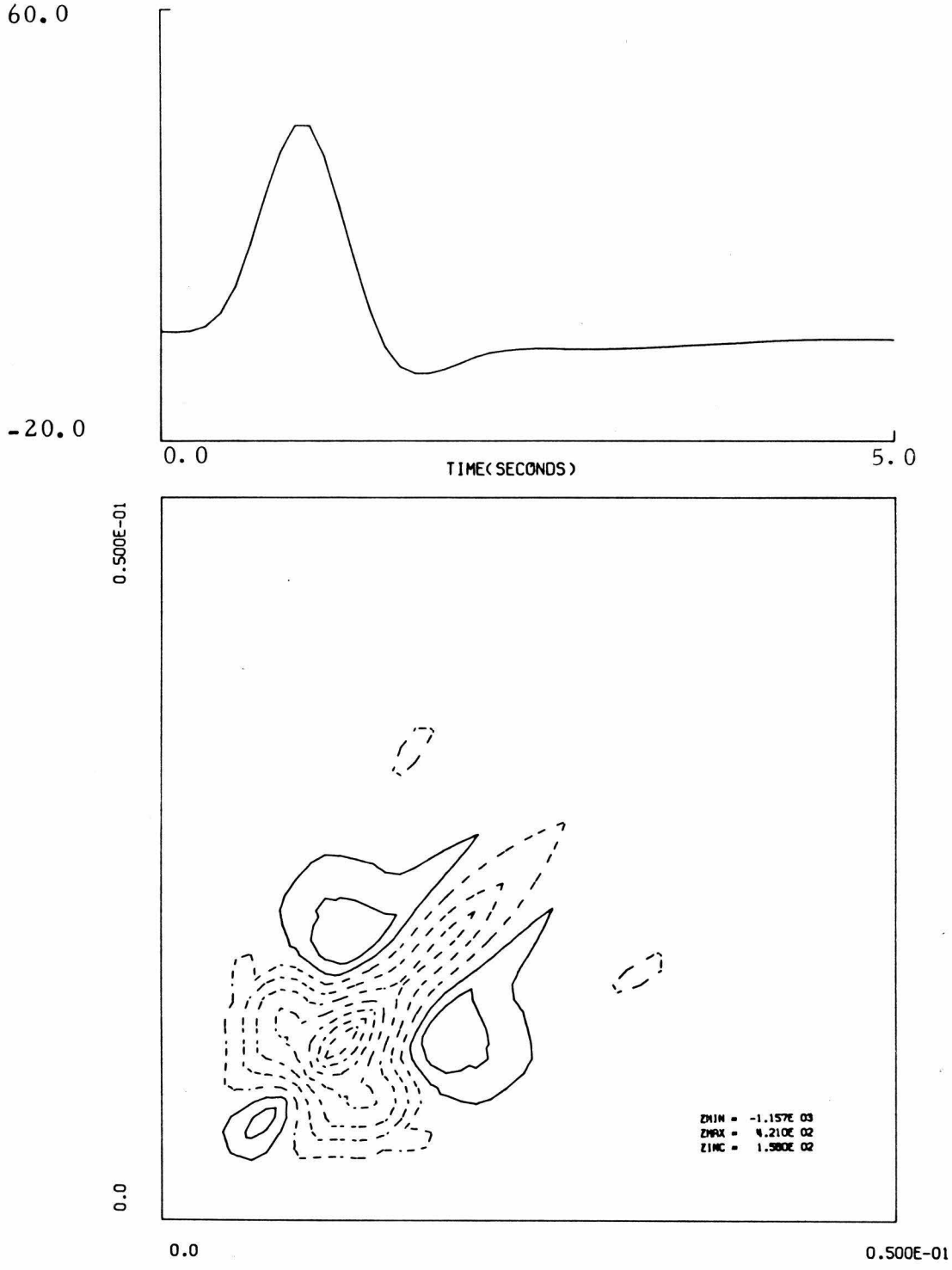


Fig. 9.2.2a: First and second order kernel estimates obtained through a 3-level equirandom CSRS (continued)

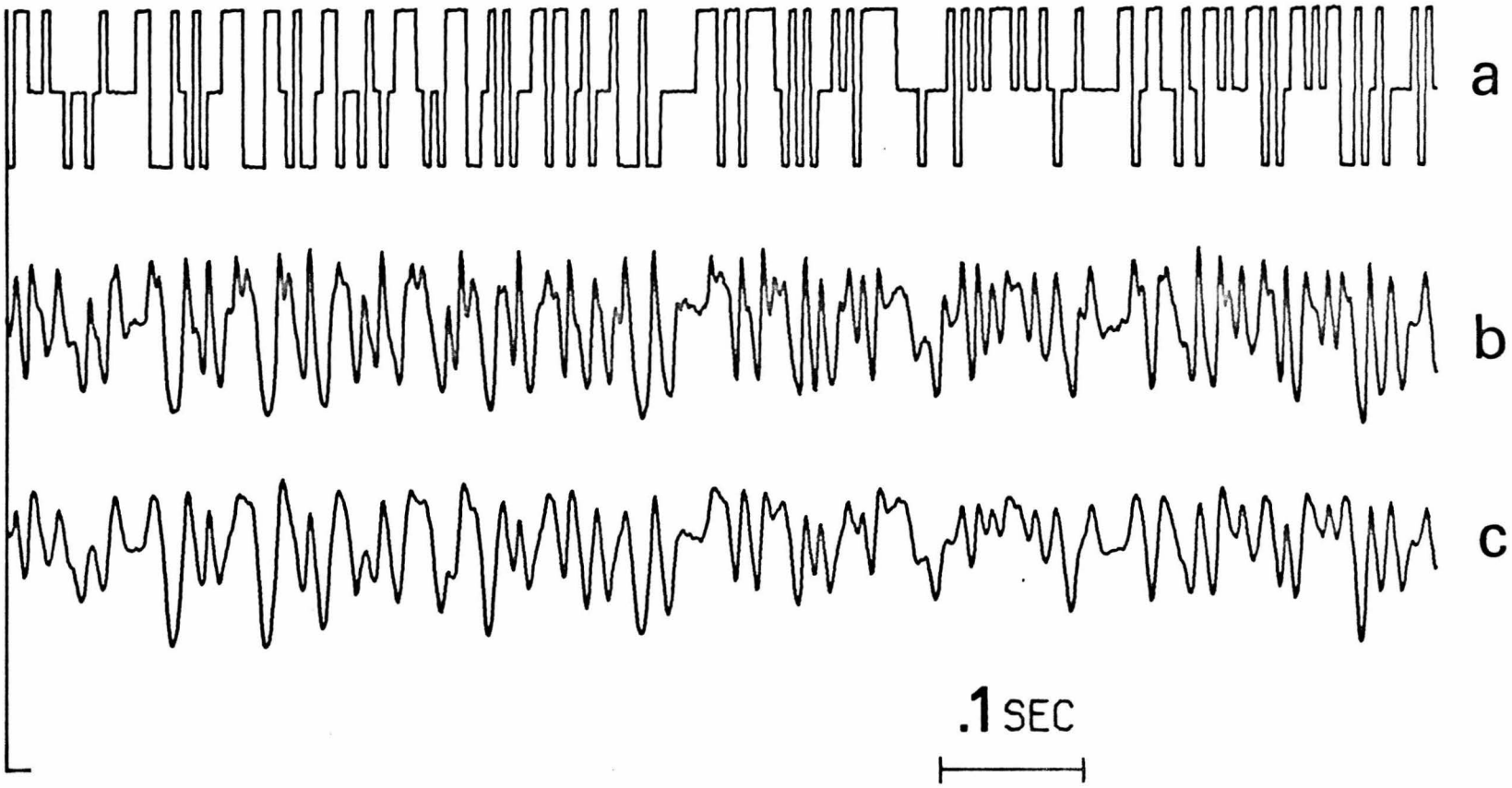


Fig. 9.2.2b: Stimulus (a), response (b) and model predicted response (c) for the 3-level equirandom CSRS case

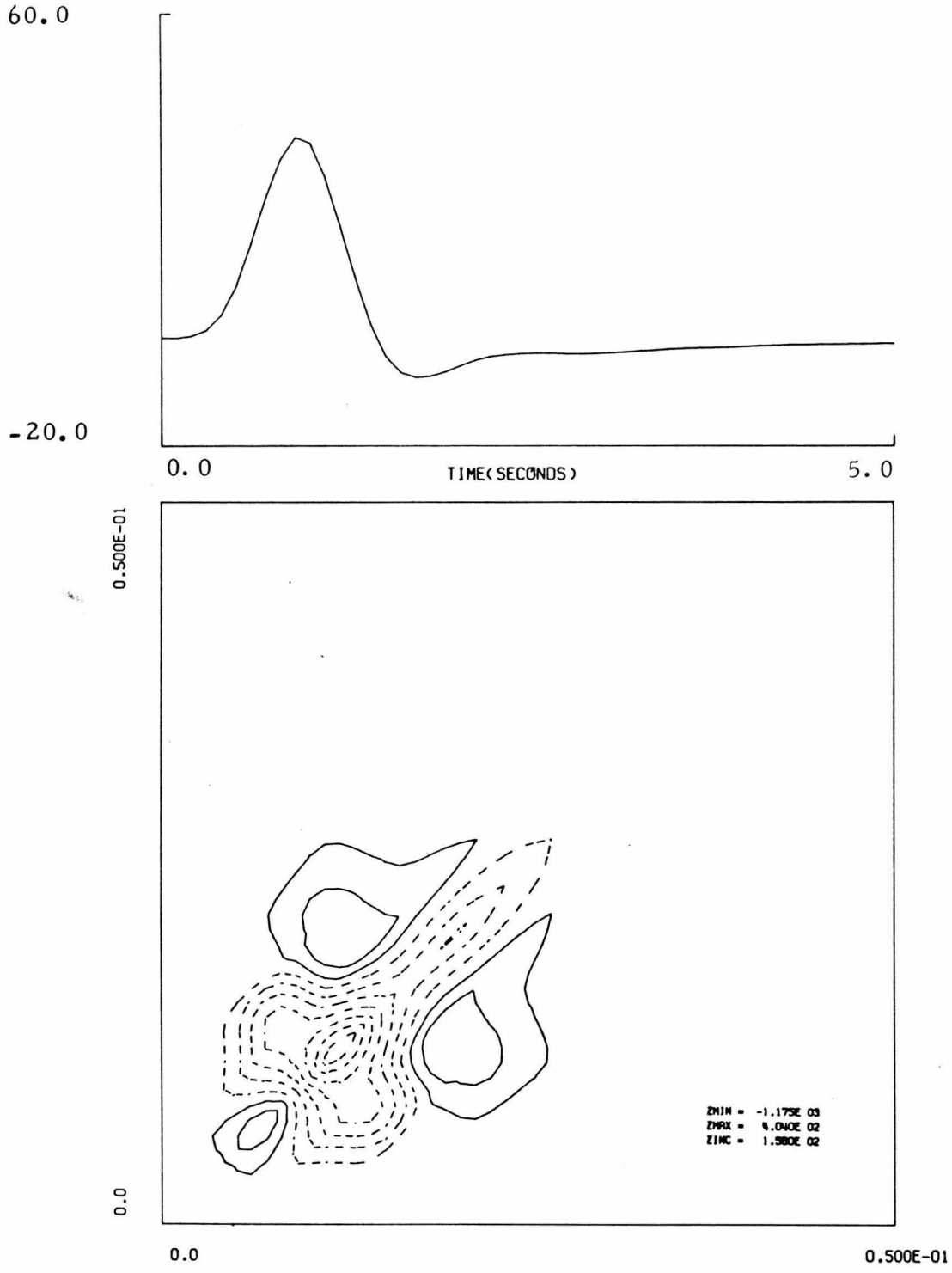


Fig. 9.2.3a: First and second order kernel estimates obtained through a 4-level equirandom CSRS (continued)

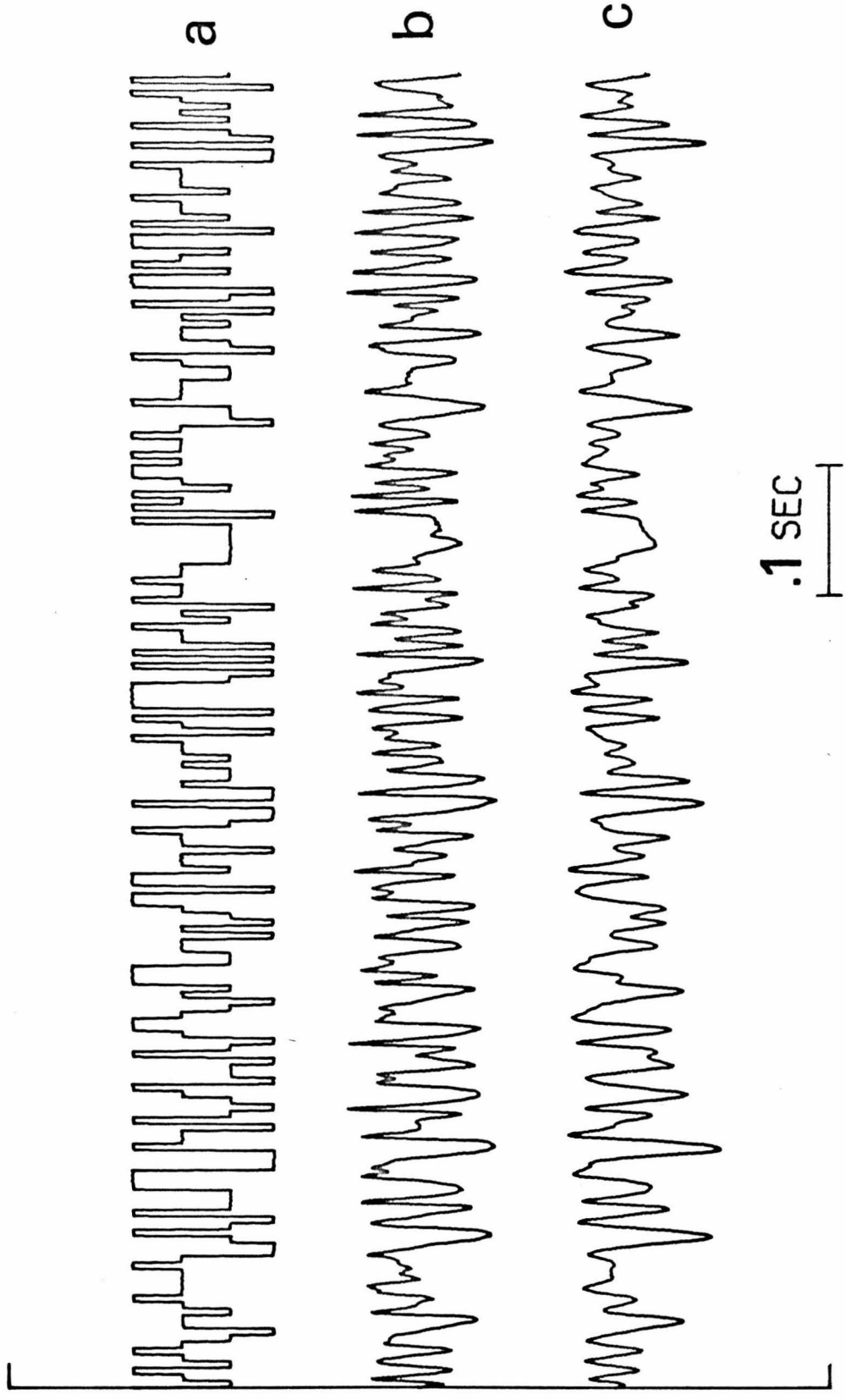


Fig. 9.2.3b: Stimulus (a), response (b) and model predicted response (c) for the 4-level equirandom CSRS case

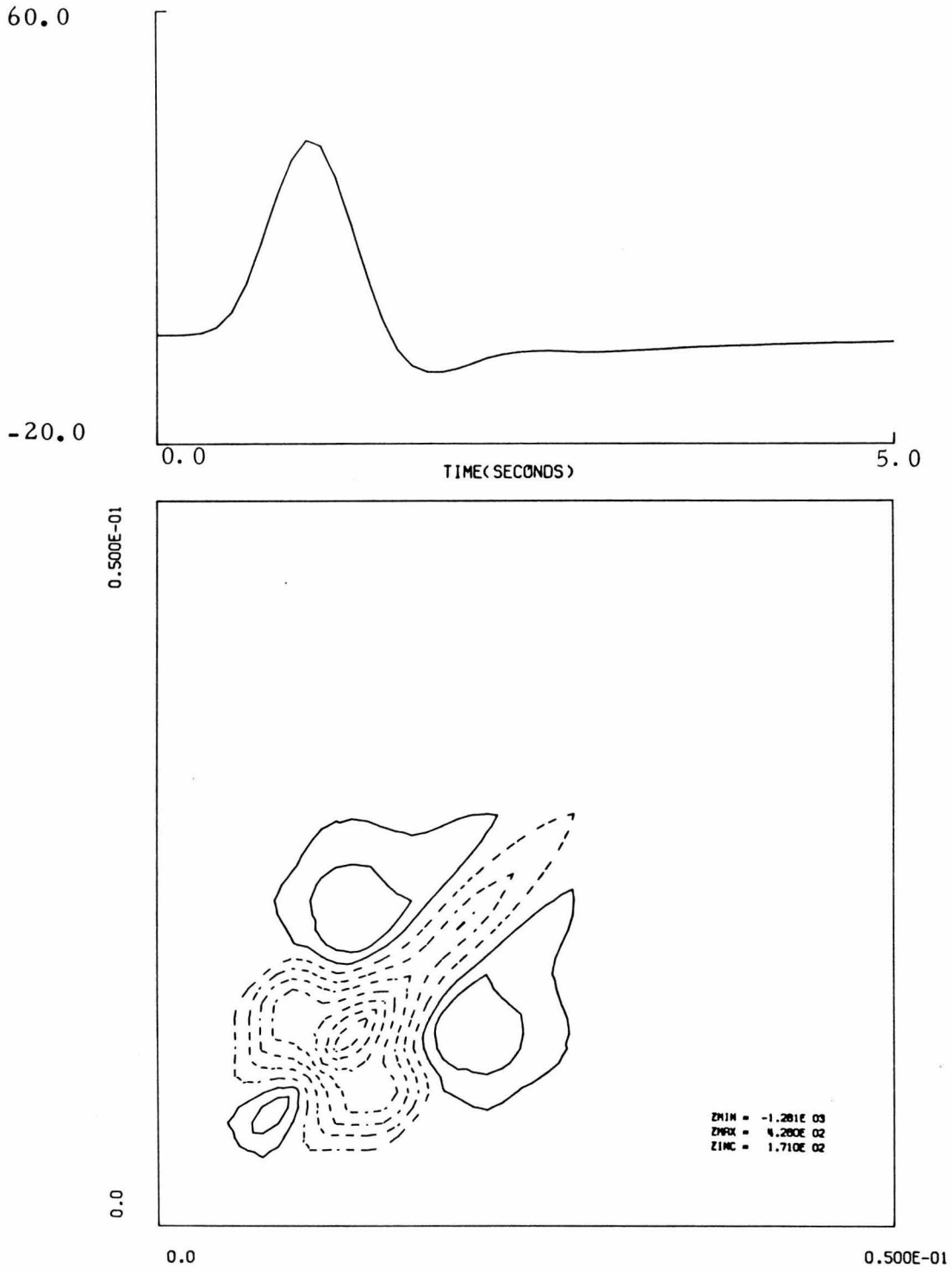


Fig. 9.2.4a: First and second order kernel estimates obtained through a 8-level equirandom CSRS (continued)

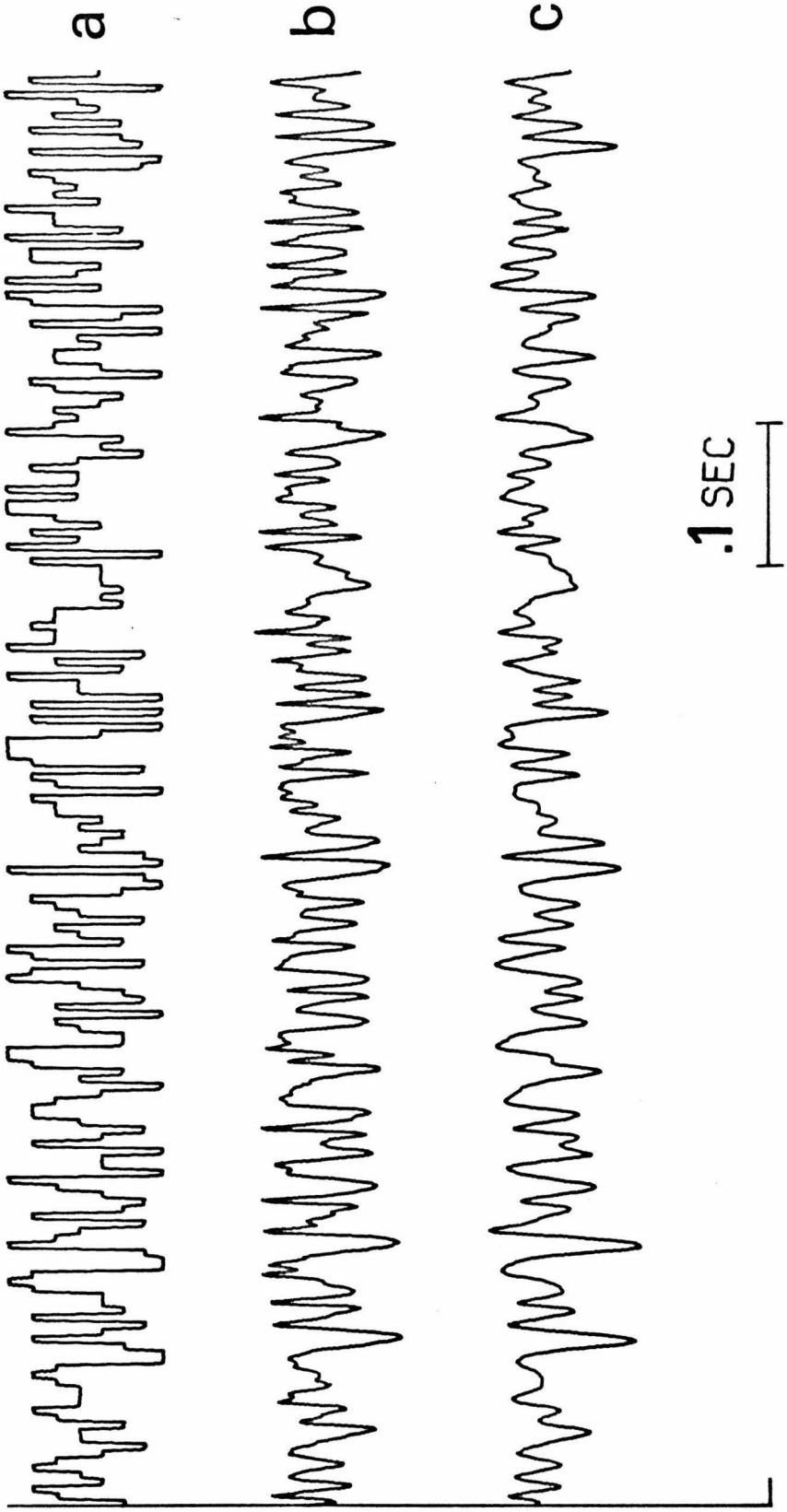


Fig. 9.2.4b: Stimulus (a), response (b) and model predicted response (c) for the 8-level equirandom CSRS case

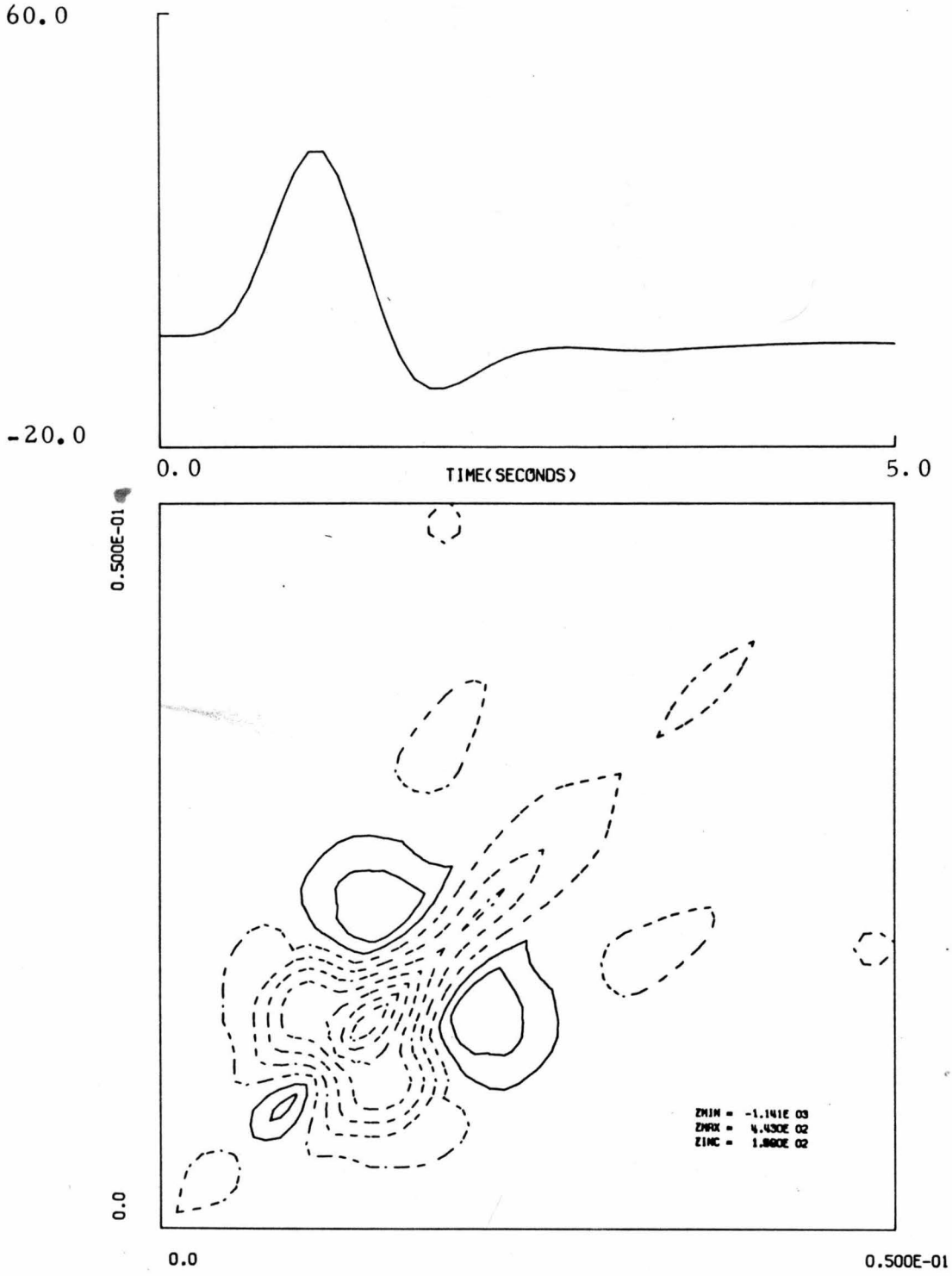


Fig. 9.2.5a: First and second order kernel estimates obtained through a 16-level equirandom CSRS (continued)

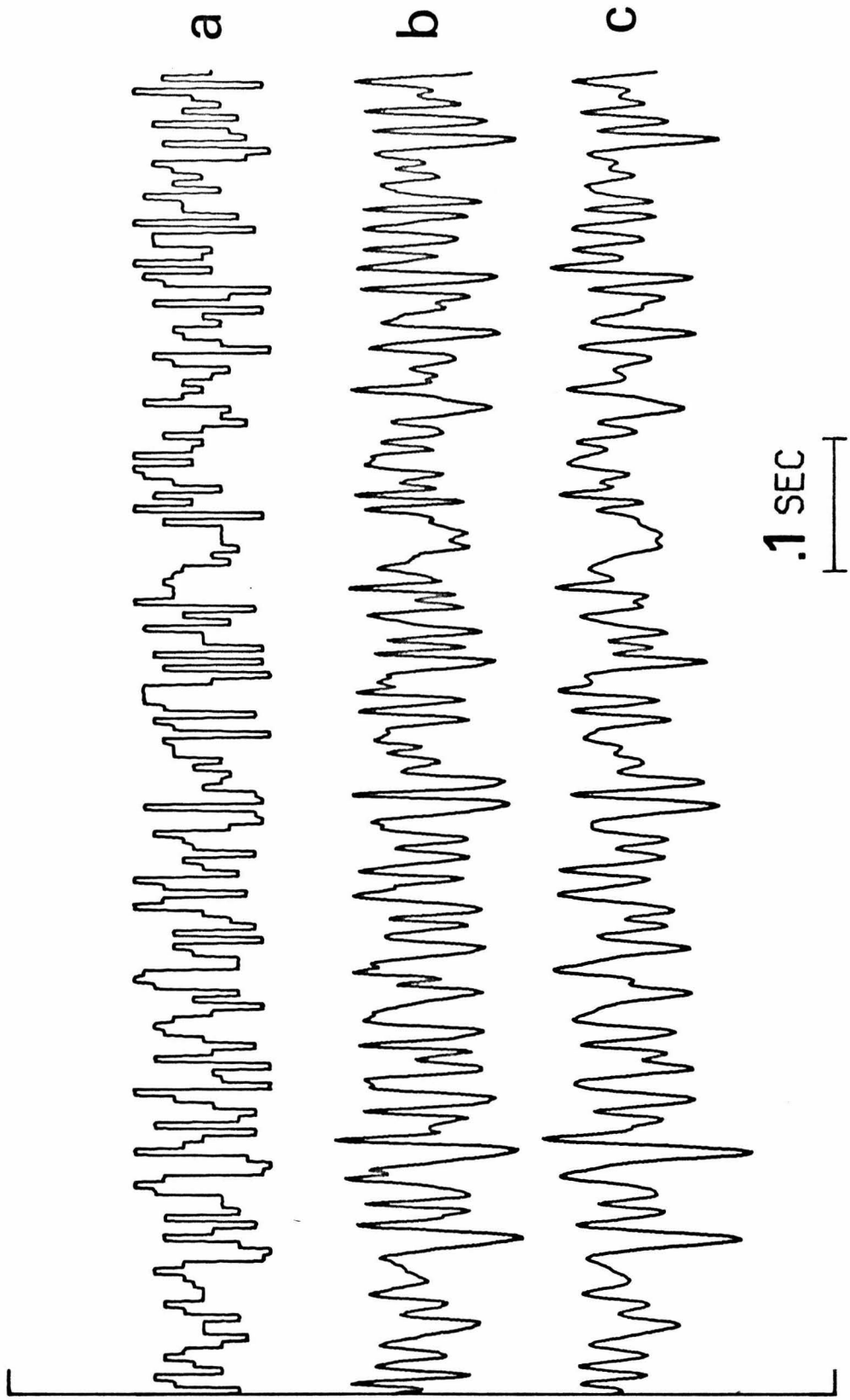


Fig. 9.2.5b: Stimulus (a), response (b) and model predicted response (c) for the 16-level equirandom CSRS case

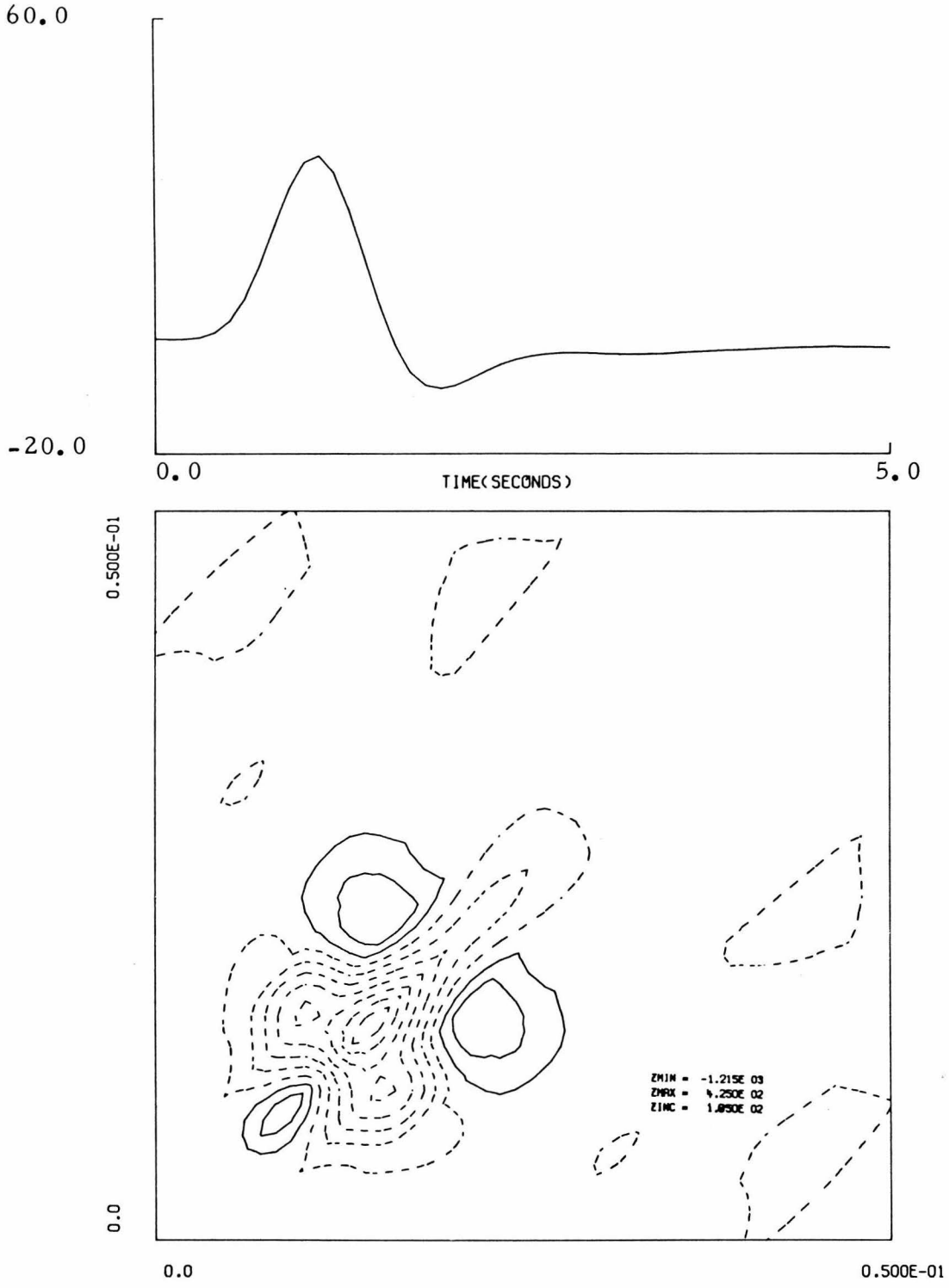


Fig. 9.2.6a: First and second order kernel estimates obtained through a 64-level equirandom CSRS (continued)

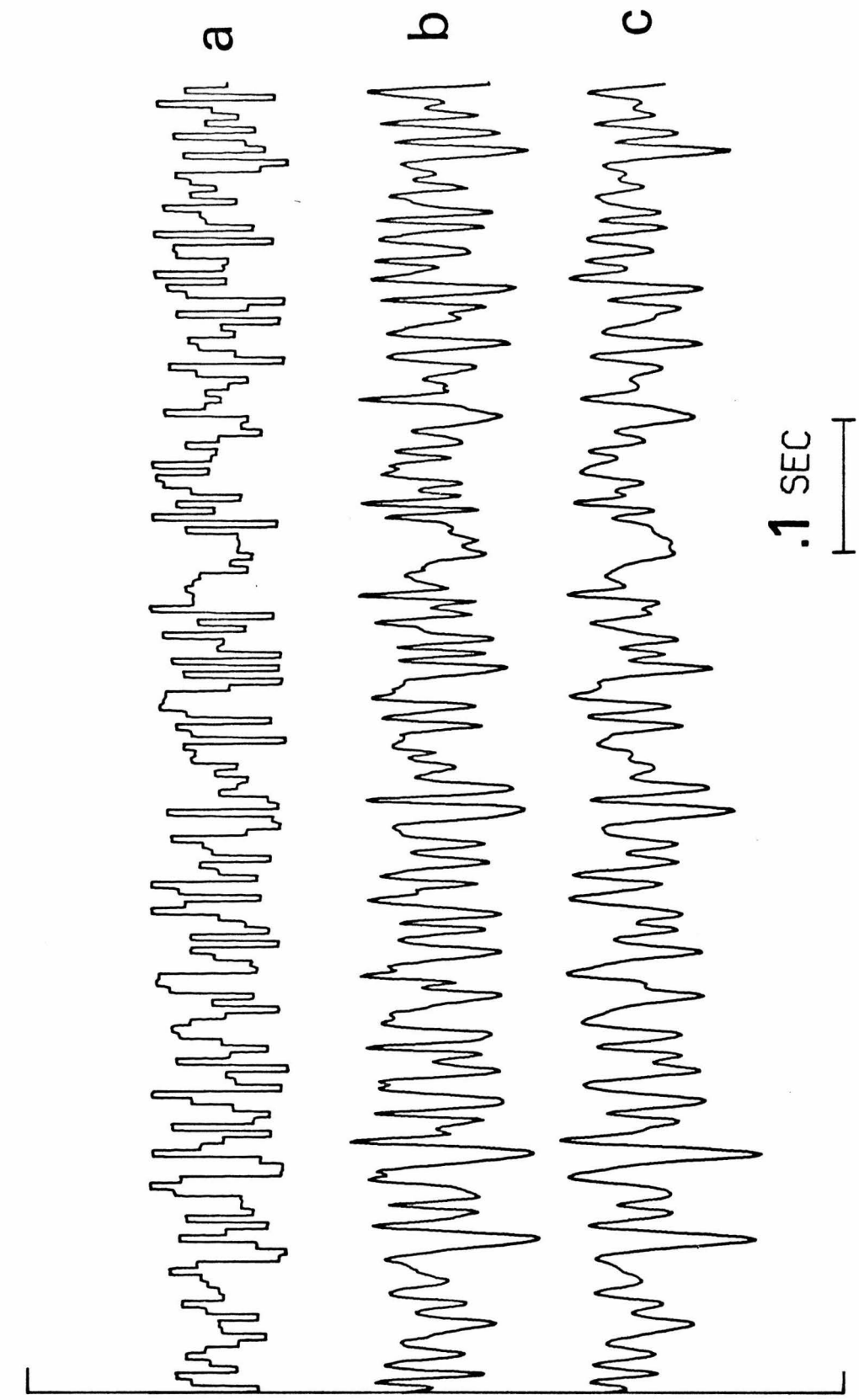


Fig. 9.2.6b: Stimulus (a), response (b) and model predicted response (c) for the 64-level equirandom CSRS case

TABLE 9.2.1

Number of levels of equirandom CSRS test signal	Percentage m.s.e. of model predicted response to the respective stimulus
2	11.90%
3	15.67%
4	16.33%
8	16.12%
16	11.71%
64	10.95%

(see Fig. 9.2.7).

This is strong evidence that the physiological system under study can be formalized as a cascade of a linear subsystem followed by a zero-memory nonlinearity. In addition, we can state that the second derivative of the zero-memory nonlinearity at the origin is negative.

If our assertion about the cascade arrangement of the system at hand is true, then we should be able to identify completely the zero-memory nonlinearity by relating a properly phase-shifted sinusoidal stimulus with the system response to that sinusoidal stimulus.

This is actually done in this case with a sinusoidal stimulus of frequency 2 Hz and amplitude 1.80 Volts (voltage corresponding to light intensity measured through a photo-cell). The obtained results is shown in Fig. 9.2.8, and it reveals a convex zero-memory nonlinearity with negative second derivative (as it was predicted by the cross section of the estimated second order kernel).

It must be emphasized that under the cascade arrangement the physiological system under study has been completely identified; since the linear subsystem is completely describable by its impulse response, which in turn is identifiable through the first order kernel estimate, and the zero-memory nonlinear subsystem is identified as demonstrated above (Fig. 9.2.9). Thus, this special cascade formalization of the system makes its complete identification a feasible and relatively simple task.

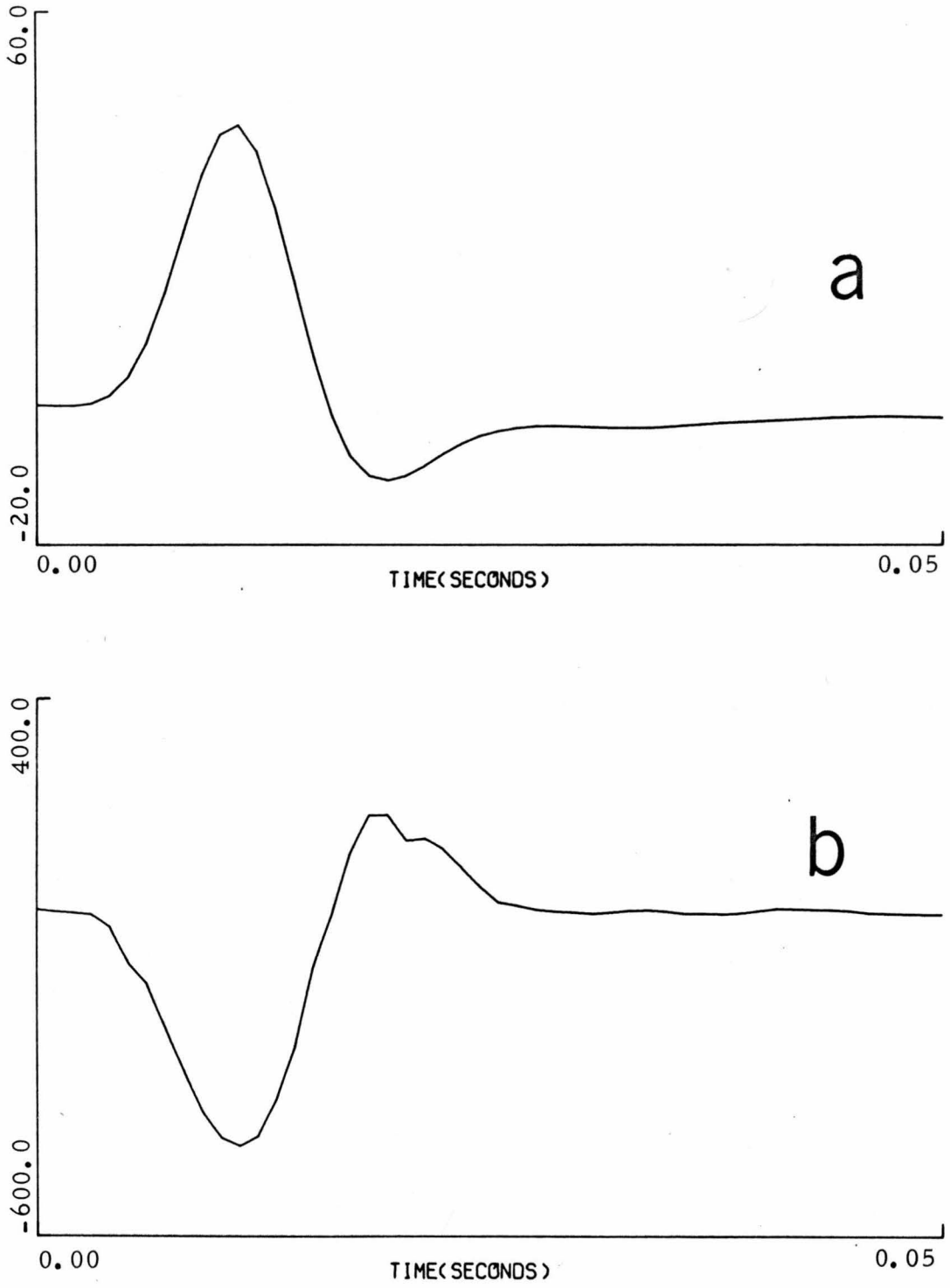


Fig. 9.2.7: (a) First order kernel estimate
(b) Cross section of second order kernel estimate

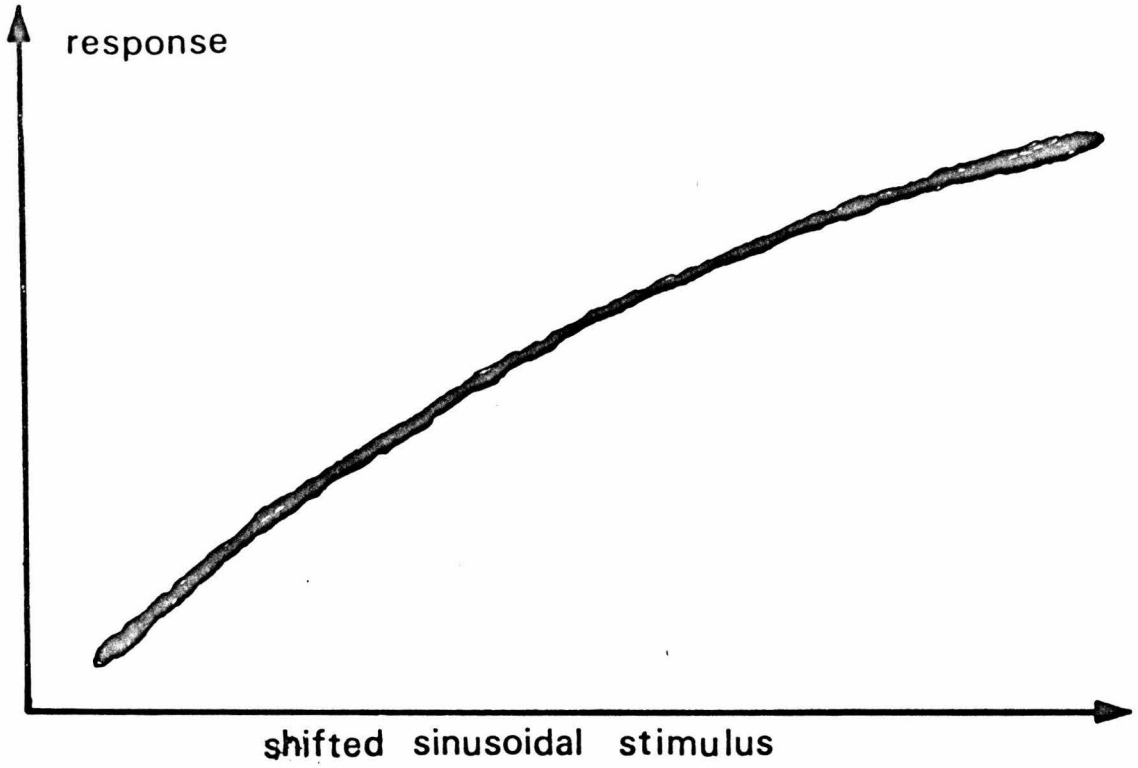


Fig. 9.2.8: Determination of the form of the zero-memory nonlinearity

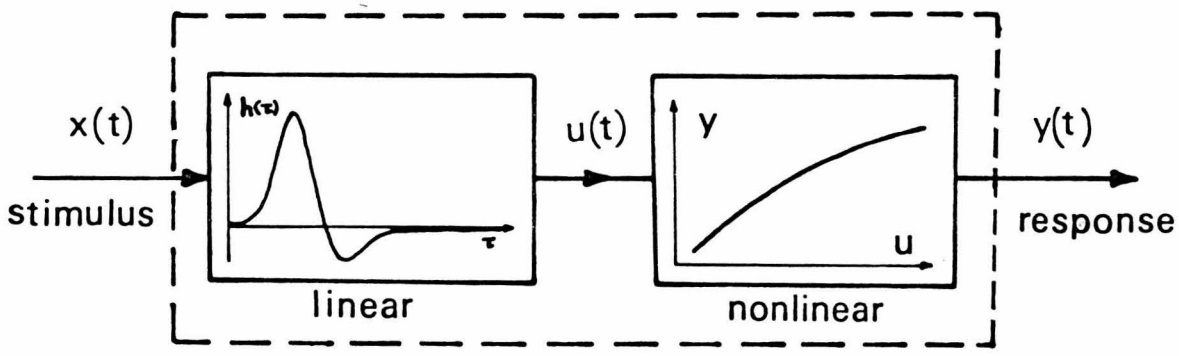


Fig. 9.2.9: Cascade system formalization of the photoreceptor light-potential transduction

9.3 Design of the Optimum Test

In this section, we will design the optimum test for the physiological system at hand by determining the optimum pair of test parameters $[T, \Delta t]$.

To this purpose, we follow a procedure similar to the one outlined in sec. 8.6 for the computer simulated application. Our immediate aim is to estimate the constants A_1, A_2, B_1, and B_2.

In order to estimate the constants B_1 and B_2, we consider a pair of experimental ternary equirandom CSRS stimulus-response records of 100 sec. long. The step length of the stimulus is .005 sec., the half operational range 1.59 Volts and the sampling interval .001 sec.

By segmenting these records, we create three groups of independent records of several lengths. More specifically:

Group 1: 2 records of 50 sec. long

Group 2: 4 records of 25 sec. long

Group 3: 8 records of 12.5 sec. long

From each one of these records, we estimate the first and second order kernel of the system. Then, we compute the integrated squared difference between any two kernels of the same group and order. Finally, we compute the averages of these integrated squared differences for each one of the groups and orders of kernel.

Using a graphical method, as demonstrated in sec. 8.6, we estimate the constants B_1 and B_2, as it is shown in Figs. 9.3.1 and 9.3.2 respectively. The obtained estimates are:

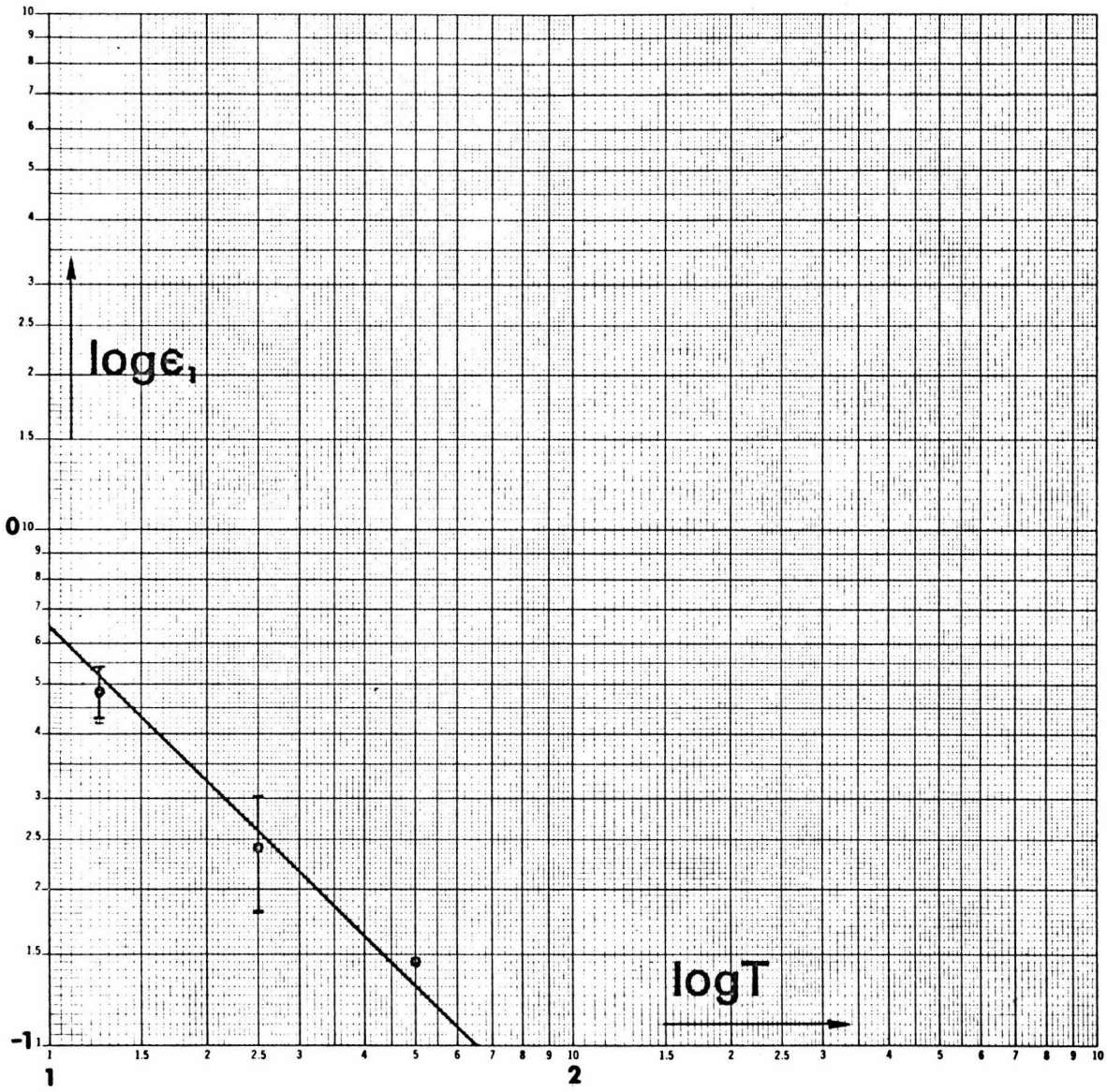


Fig. 9.3.1: Graphic estimation of B_1 in the photoreceptor

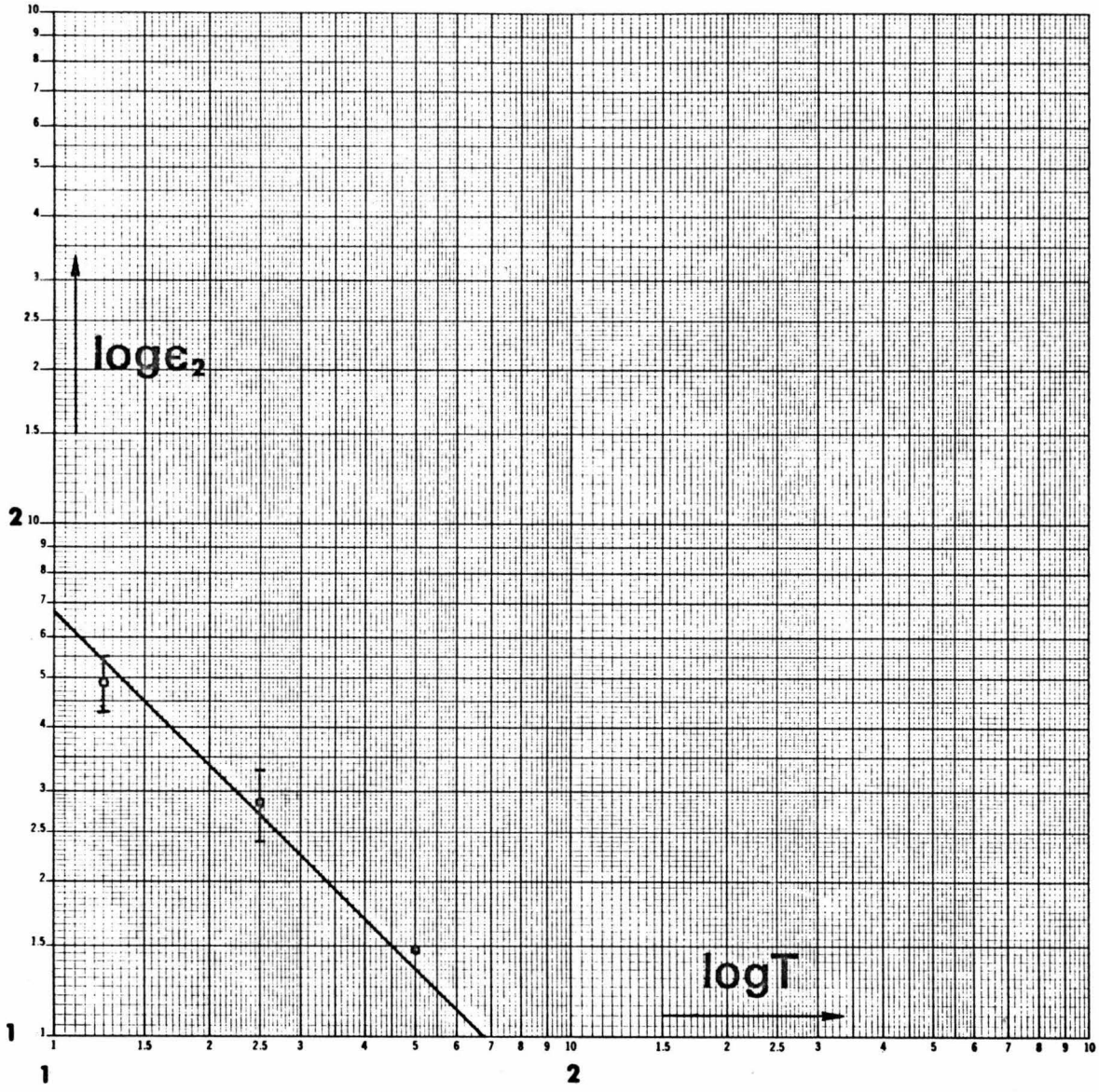


Fig. 9.3.2: Graphic estimation of B_2 in the photoreceptor

$$\hat{B}_1 \cong 3.24$$

$$\hat{B}_2 \cong 1.68$$

Using the method illustrated in sec. 8.6, for step lengths Δt = .002, .004, .006 and .008 sec. and record length T = 100 sec., we also estimate the constants A_1 and A_2. The obtained estimates are:

$$\hat{A}_1 \cong 2.12 \cdot 10^{10}$$

$$\hat{A}_2 \cong 8.24 \cdot 10^{11}$$

Now that the constants A_1, A_2 and B_2 have been estimated, the optimum Δt, for the record length T=100 sec., can be estimated with the help of eqn. 8.6.8: (note that m_2=1.04)

$$(\Delta t)_{opt} = \left[\frac{(1.04)(1.68)}{400(2.12 \cdot 10^{10} + 8.24 \cdot 10^{11})}\right]^{1/5}$$

$$\simeq 1.38 \cdot 10^{-3} \text{ sec.}$$

The optimum step length as a function of the record length is:

$$(\Delta t)_{opt} \simeq .00347 \Big/ T^{1/5} \qquad \text{(T and } \Delta t \text{ in sec.)} \qquad (9.3.1)$$

With the derivation of eqn. 9.3.1, the estimation of the optimum test parameters T and Δt has been concluded for the physiological system under study. It must be reminded, however, that all these results must be viewed with a statistical perspective and the becoming tolerance.

CHAPTER X

COMPARATIVE STUDY OF THE USE OF GWN, PRS AND CSRS IN NONLINEAR SYSTEM IDENTIFICATION

In this chapter, we will illustrate through actual applications the relative efficiency of GWN, PRS and CSRS in nonlinear system identification. These actual applications include several computer simulated systems (sec. 10.1) and a real physiological system (sec. 10.2).

In sec. 4.4, we discussed the relative advantages and disadvantages of GWN, PRS and CSRS on a theoretical basis. This chapter attempts to provide some examples illustrating the basic assertions made in that section.

10.1 Computer Simulated Applications of GWN, PRS and CSRS

The systems that we consider in these computer simulated applications are the ones shown in Figs. 8.5.1, 8.3.1 and 8.2.1. We consider those systems, since they cover the main cases of system nonlinearities that are of interest in practice, given that we usually confine ourselves in estimating upto the second order kernel.

More specifically, the system of Fig. 8.5.1 has up to second order nonlinearities; therefore, a complete model of it can be estimated.

The system of Fig. 8.3.1 has up to third order nonlinearities with the third order term being of comparable size with the first and second order terms; thus, the incomplete second order model that is

estimated will contain only part of the quantity corresponding to the third degree polynomial term (cf. sec. 8.3), and the effectiveness of this probing into the third degree term depends upon the fourth order autocorrelation function (recall that the fourth order autocorrelation function is the first one where the pseudorandom signals exhibit anomalies).

Finally, the system of Fig. 8.2.1 has nonlinearities of all orders, while the size of the higher order terms is rapidly decreasing; thus, the incomplete second order model that is estimated will contain parts of the quantities corresponding to terms of all orders with the relative contribution decreasing as the order of the term is increasing.

The test signals, that are used in these applications, have the appropriate characteristics that make them completely comparable from the system identification point of view.

More specifically, they all have the same record length T= 1968 sec. and sampling interval DT=.15 sec. The record length corresponds to one period of a ternary pseudorandom signal being generated by an eighth order linear recurrence formula ($3^8-1=6560$ steps) and having a step length $\Delta t=.3$ sec. ($6560 \times .3 = 1968$ sec.).

The CSRS family is represented in these applications by a ternary equirandom signal of the same record and step length with the pseudorandom ternary signal. The operational range for both the PRS and the CSRS is A=1 stimulus unit.

Due to the fact that the amplitude histogram of one period of the pseudorandom ternary signal has approximately uniform profile

(i. e. 2187 step values are +1's, 2187 step values are -1's and 2186 step values are 0's), the pseudorandom and the equirandom ternary test signals have approximately the same power level P=. 2. That makes them completely comparable from the system identification point of view (cf. sec. 4. 3).

Things are a little more complicated trying to make the GWN comparable to the PRS and the CSRS. For one thing, there is no exact correspondence between the frequency bandwidths in the two cases. The power spectrum of the PRS and the CSRS is shown in Fig. 4. 3. 1, and the frequency where it first vanishes in this application is 3. 33 Hz. Thus, it is reasonable to take a bandwidth of approximately 2. 5 Hz for the GWN test signal. For another thing, due to the different statistical moments of a gaussian and an equirandom ternary amplitude probability distribution, the requirement of having the same power level in all cases will force us to have an operational range for the GWN which is different than the one of PRS and CSRS. Thus, the appropriate operational range for the GWN is determined from the specified GWN bandwidth B=2. 5 Hz and power level P=. 2 (cf. eqn. 3. 2. 5).

Summarizing the criteria that we used to determine the GWN, PRS and CSRS test signals, so that they are comparable from the system identification point of view: (1) the same temporal record length; (2) the same number of sample points; (3) the same power level (of first order); and, (4) approximately the same frequency bandwidth.

Apparently, the comparability criteria (3) and (4) allow a certain degree of arbitrariness, which in turn points to the approximate

nature of this study. The suggested approximate prospect of this illustration is further enhanced by the fact that the GWN and CSRS test signals are random in structure, and consequently they insert an additional degree of randomness in the obtained results.

Notice that we selected the specific GWN and CSRS test signals, that have been used in this illustration, as the best performing ones from an ensemble of four randomly chosen sample signals in each case.

Having described the test signals and the systems that are going to be used in the present illustration we proceed with presenting the obtained results.

For the system of Fig. 8.5.1 (second order nonlinearity):

The percentage m.s. errors of the first order kernel estimates are:

GWN: .73%

PRS: .37%

CSRS: .67%

The percentage m.s. errors of the second order kernel estimates are:

GWN: 3.93%

PRS: 5.49%

CSRS: 3.36%

These first and second order kernel estimates are shown in Figs. 10.1.1 through 10.1.3.

The percentage m.s. errors of the second order model predicted responses are:

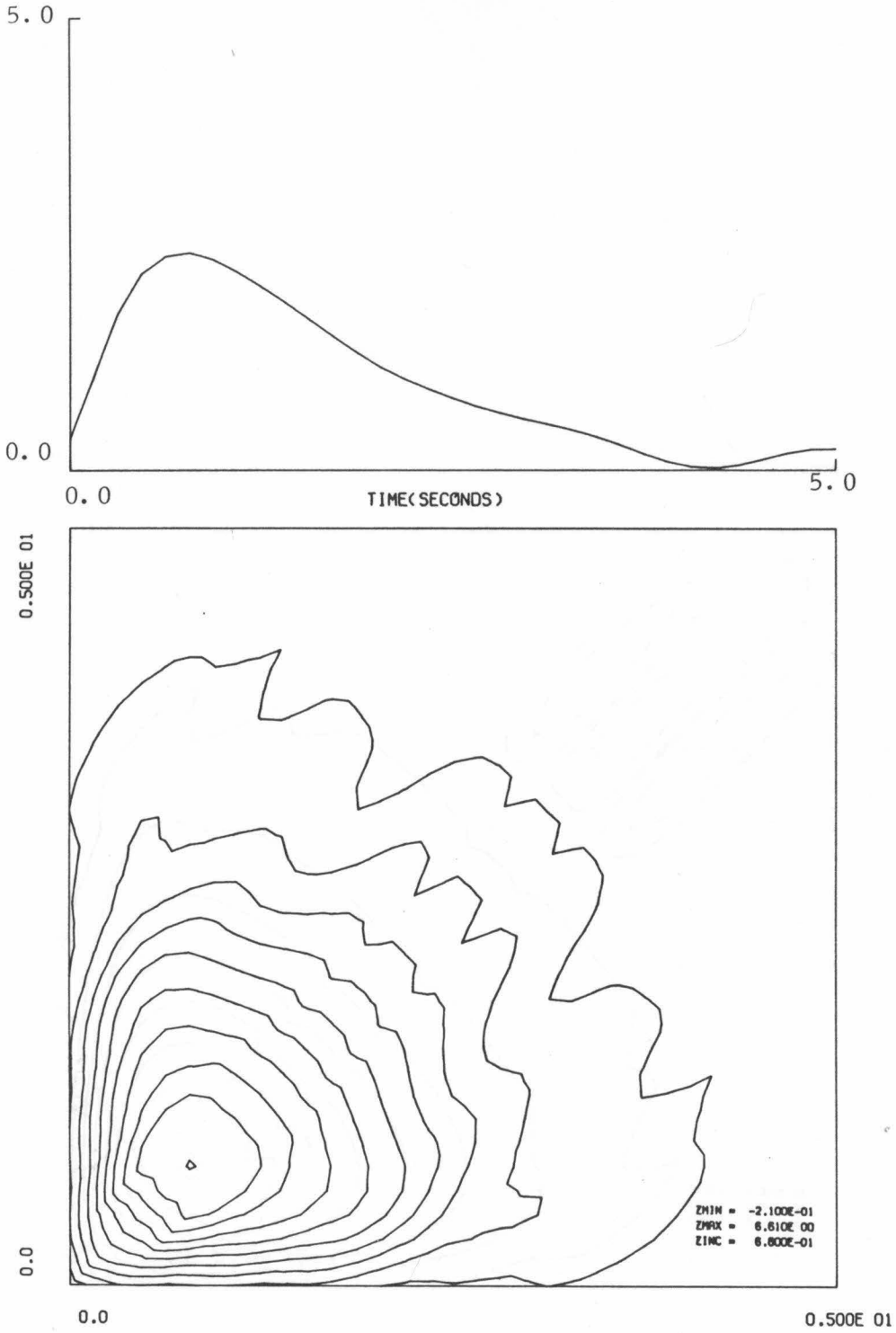


Fig. 10.1.1: First and second order GWN kernel estimates for system of Fig. 8.5.1.

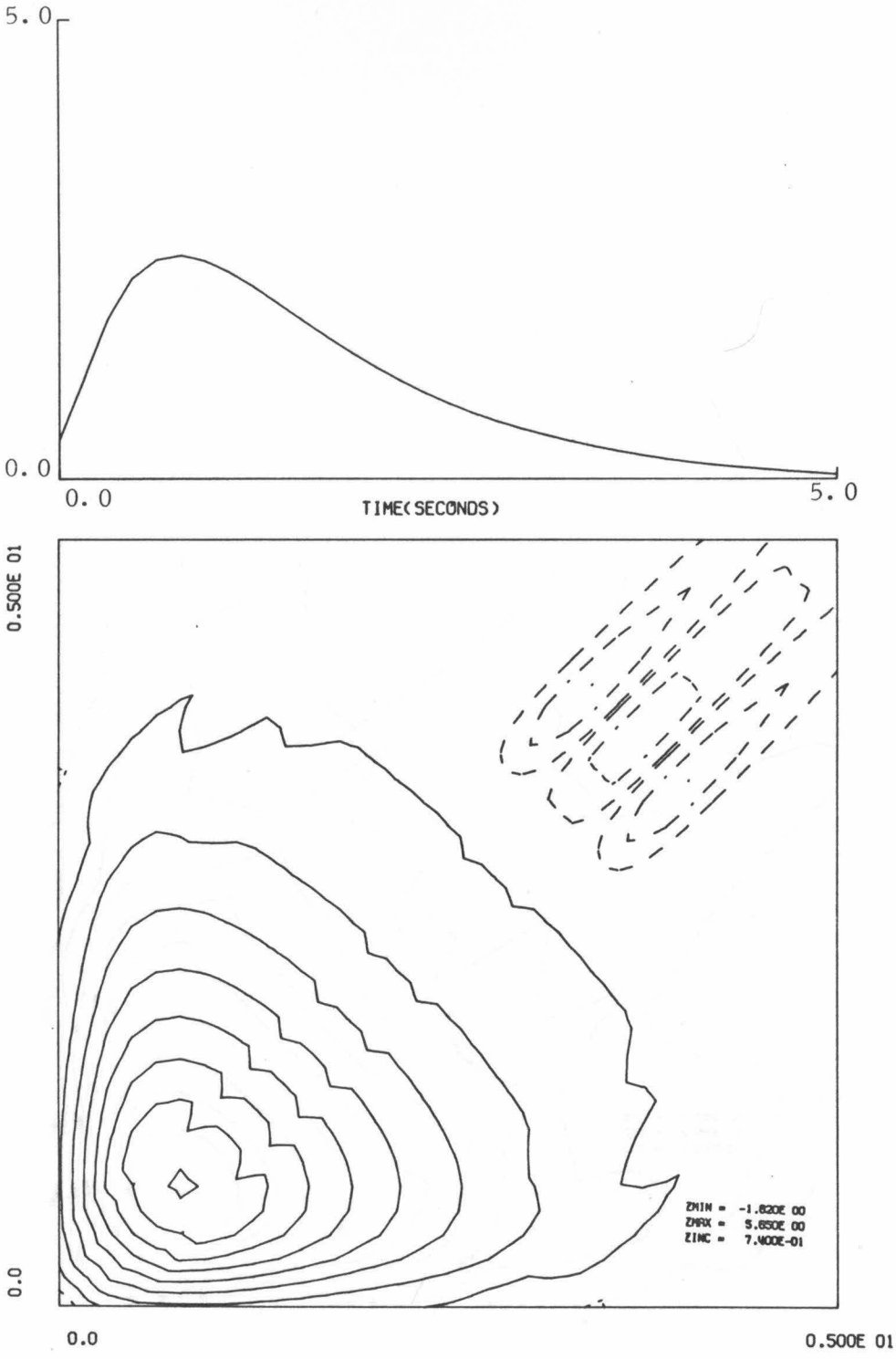


Fig. 10.1.2: First and second order PRS kernel estimates for system of Fig. 8.5.1.

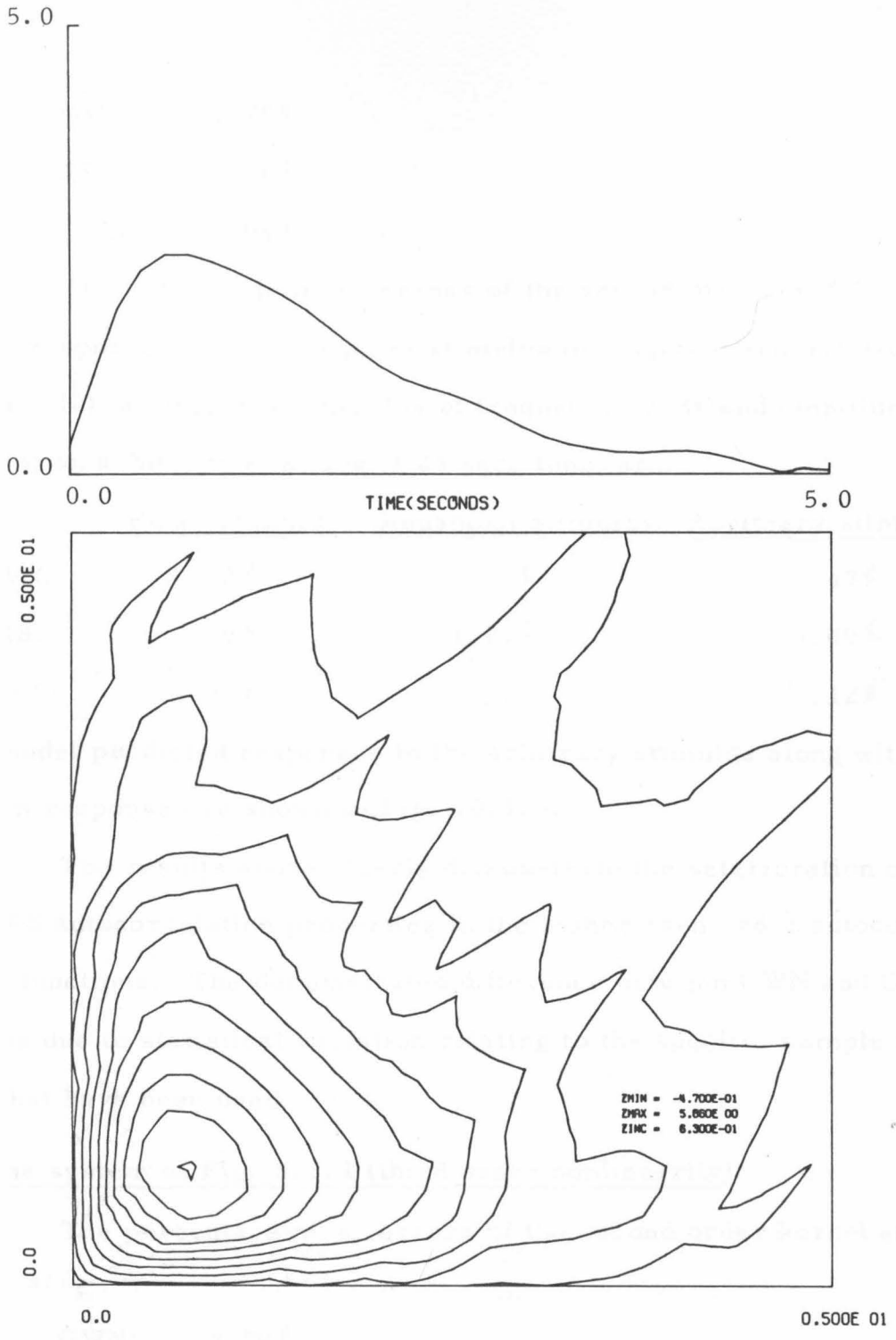


Fig. 10.1.3: First and second order CSRS kernel estimates for system of Fig. 8.5.1.

GWN: 1.95%

PRS: 1.88%

CSRS: 1.04%

The percentage m. s. errors of the second order model predicted responses to: (1) a pulse stimulus of height .5 and duration 10 sec.; (2) a sinusoidal stimulus of frequency .2 Hz and amplitude 1; and, (3) an arbitrary stimulus of 25 sec. long; are:

	Pulse stimulus	Sinusoidal stimulus	Arbitrary stimulus
GWN:	.62%	1.51%	.47%
PRS:	.69%	1.13%	1.49%
CSRS:	.35%	.79%	.12%

The model predicted responses to the arbitrary stimulus along with the system response are shown in Fig. 10.1.4.

The results above clearly demonstrate the deterioration of the PRS autocorrelation properties in the higher even order autocorrelation functions. The demonstrated difference between GWN and CSRS may be due to statistical variation relating to the specific sample signals that have been used.

For the system of Fig. 8.3.1 (third order nonlinearity):

The percentage m. s. errors of the second order kernel estimates are:

GWN: 8.09%

PRS: 5.48%

CSRS: 5.93%

The first and second order kernel estimates are shown in Fig. 10.1.5

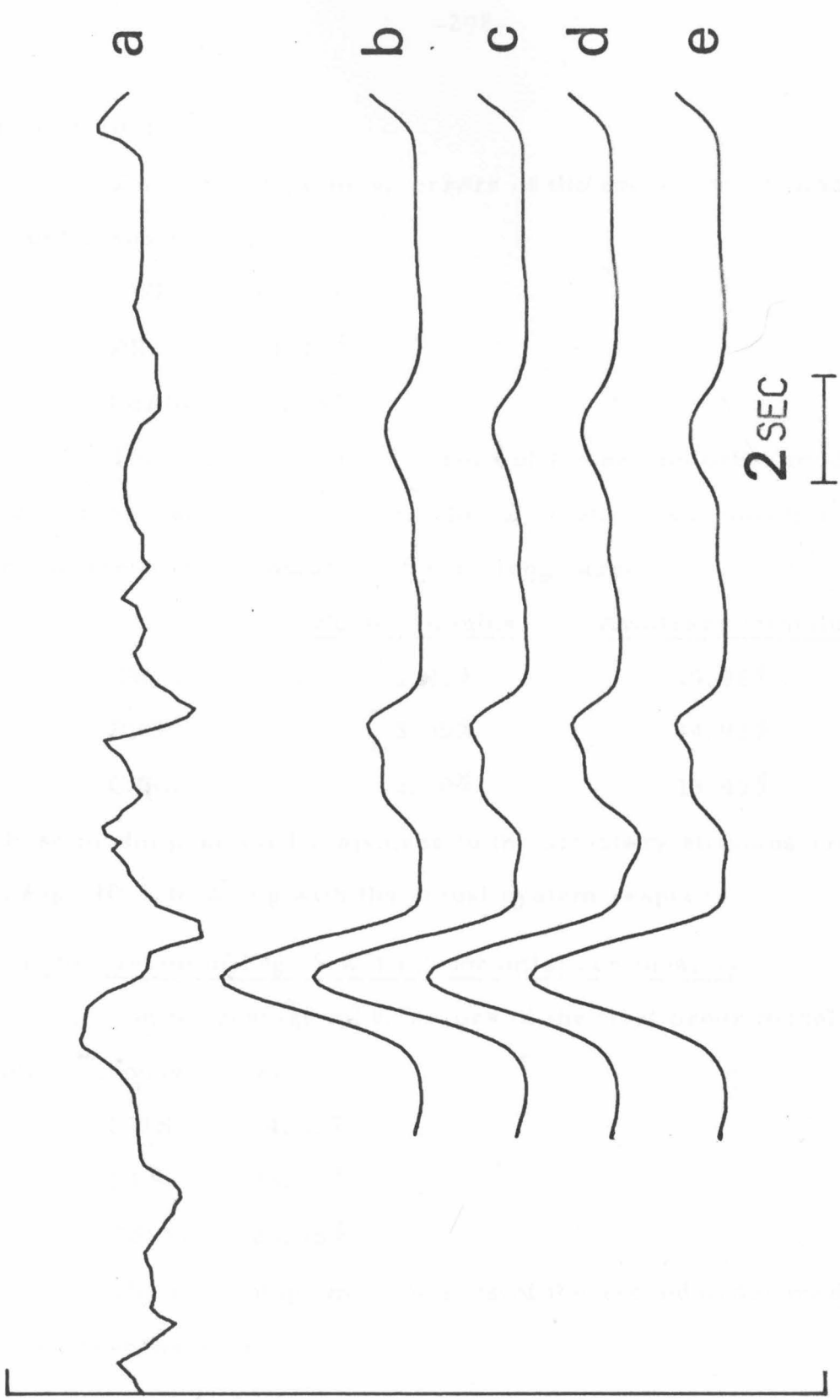


Fig. 10.1.4: (a) arbitrary stimulus, (b) system (of Fig. 8.5.1) response to the arbitrary stimulus, (c) GWN model predicted response, (d) PRS model predicted response, (e) CSRS model predicted response.

through 10. 1. 7.

The percentage m. s. errors of the second order model predicted responses are:

GWN: 35.02%

PRS: 30.08%

CSRS: 24.88%

The percentage m. s. errors of the second order model predicted responses to a pulse stimulus of height . 5 and duration 10 sec., and an arbitrary stimulus of 25 sec. long, are:

	Pulse stimulus	Arbitrary stimulus
GWN:	1.99%	19.98%
PRS:	3.80%	14.92%
CSRS:	2.10%	13.45%

These model predicted responses to the arbitrary stimulus are shown in Fig. 10. 1. 8, along with the actual system response.

For the system of Fig. 8. 2. 1 (Exponential nonlinearity):

The percentage m. s. errors of the first order model predicted responses are:

GWN: 54.33%

PRS: 33.31%

CSRS: 28.48%

The percentage m. s. errors of the second order model predicted responses are:

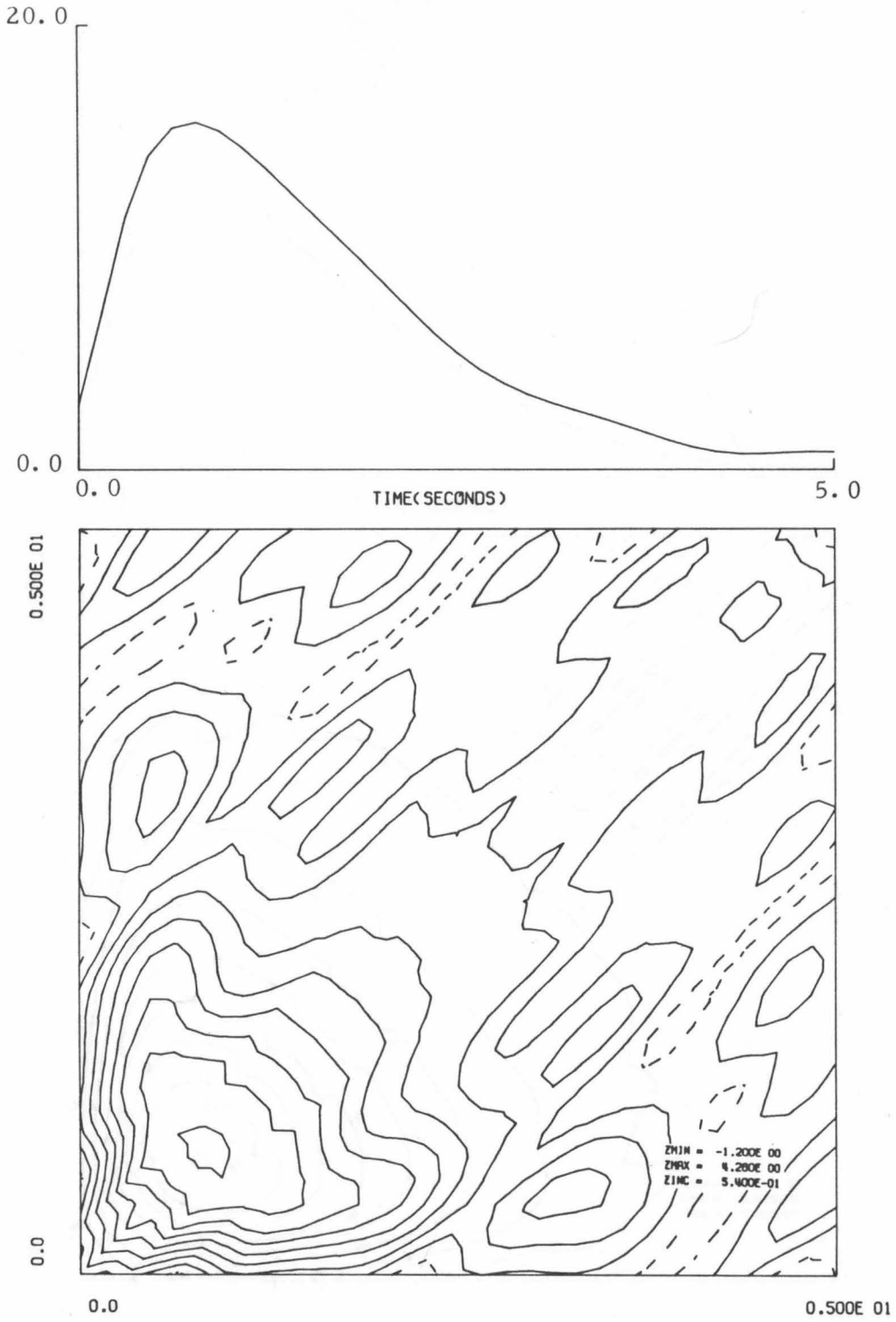


Fig. 10.1.5: First and second order GWN kernel estimates for system of Fig. 8.3.1.

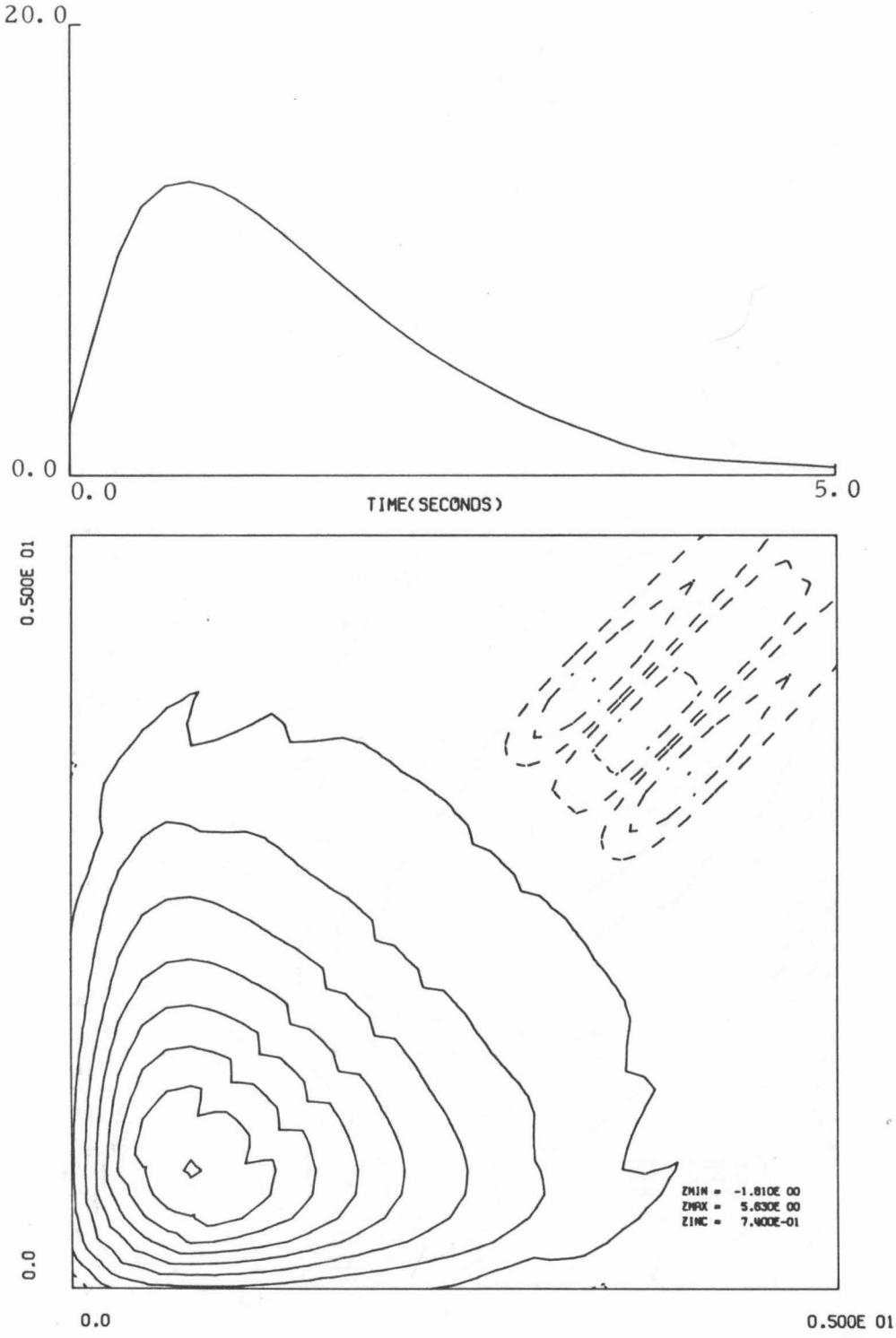


Fig. 10.1.6: First and second order PRS kernel estimates for system of Fig. 8.3.1.

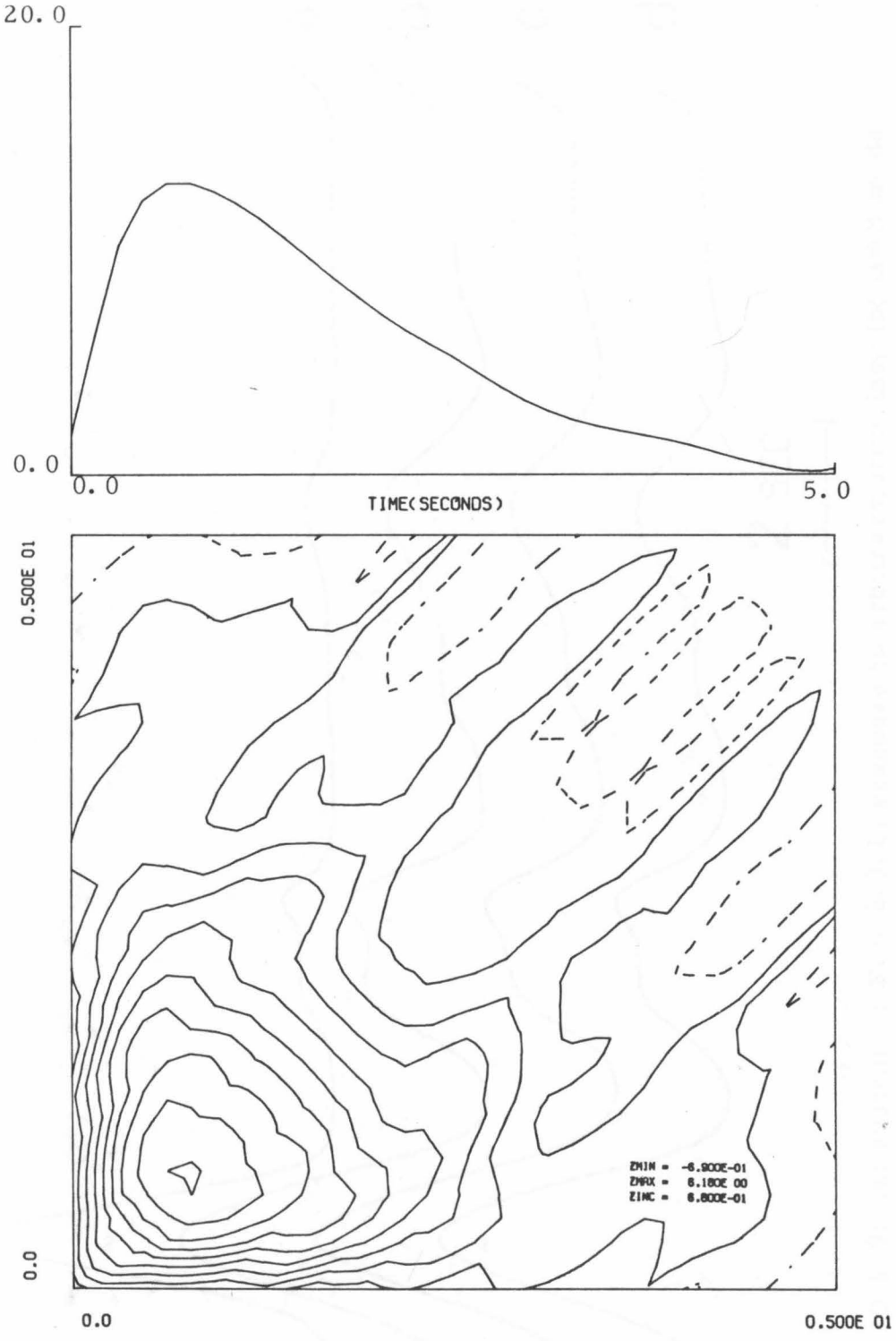


Fig. 10.1.7: First and second order CSRS kernel estimates for system of Fig. 8.3.1.

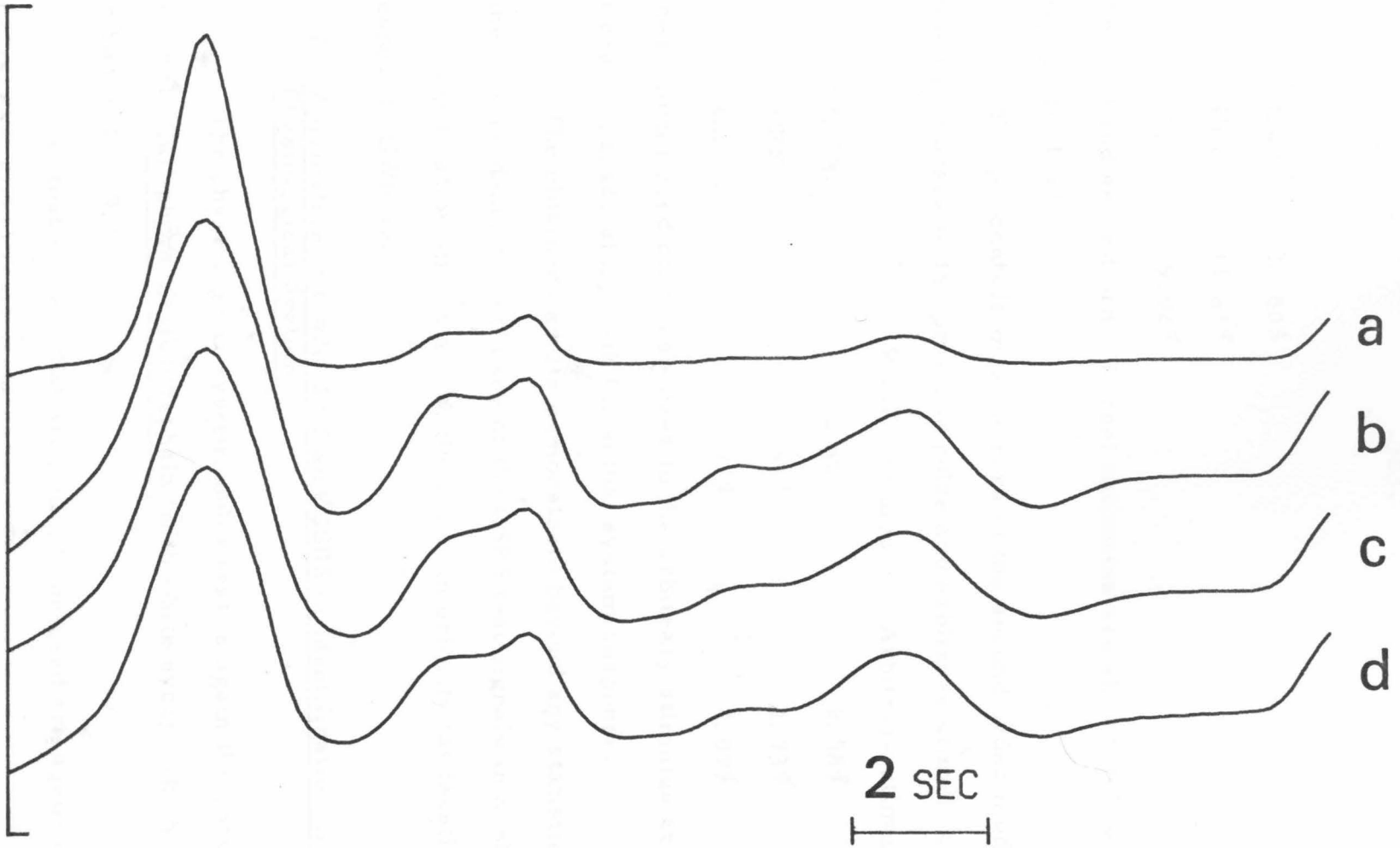


Fig. 10.1.8: (a) system (of Fig. 8.3.1) response to arbitrary stimulus, (b) GWN model predicted response, (c) PRS model predicted response, (d) CSRS model predicted response.

GWN: 26.80%

PRS: 11.83%

CSRS: 9.02%

The first and second order kernel estimates are shown in Fig. 10.1.9 through 10.1.11.

The percentage m.s. errors of the second order model predicted responses to the previous pulse and arbitrary stimuli are:

	Pulse stimulus	Arbitrary stimulus
GWN:	1.90%	8.58%
PRS:	.83%	4.73%
CSRS:	.77%	4.07%

These model predicted responses to the arbitrary stimulus are shown in Fig. 10.1.12, along with the actual system response.

The obtained results demonstrate beyond any statistical (or other) coincidence the efficacy of the CSRS test signals in nonlinear system identification and establish their non-inferiority (at least) with respect to GWN and PRS.

10.2 Application of GWN, PRS and CSRS in Identification of a Physiological System

The physiological system under test is again the photo-receptor of the fly Calliphora Erythrocephala (with white eyes), which was described in sec. 9.1.

The test signals that are going to be used are again a ternary equirandom signal (representing the CSRS family), a ternary pseudo-random signal (representing the PRS family) and a GWN signal.

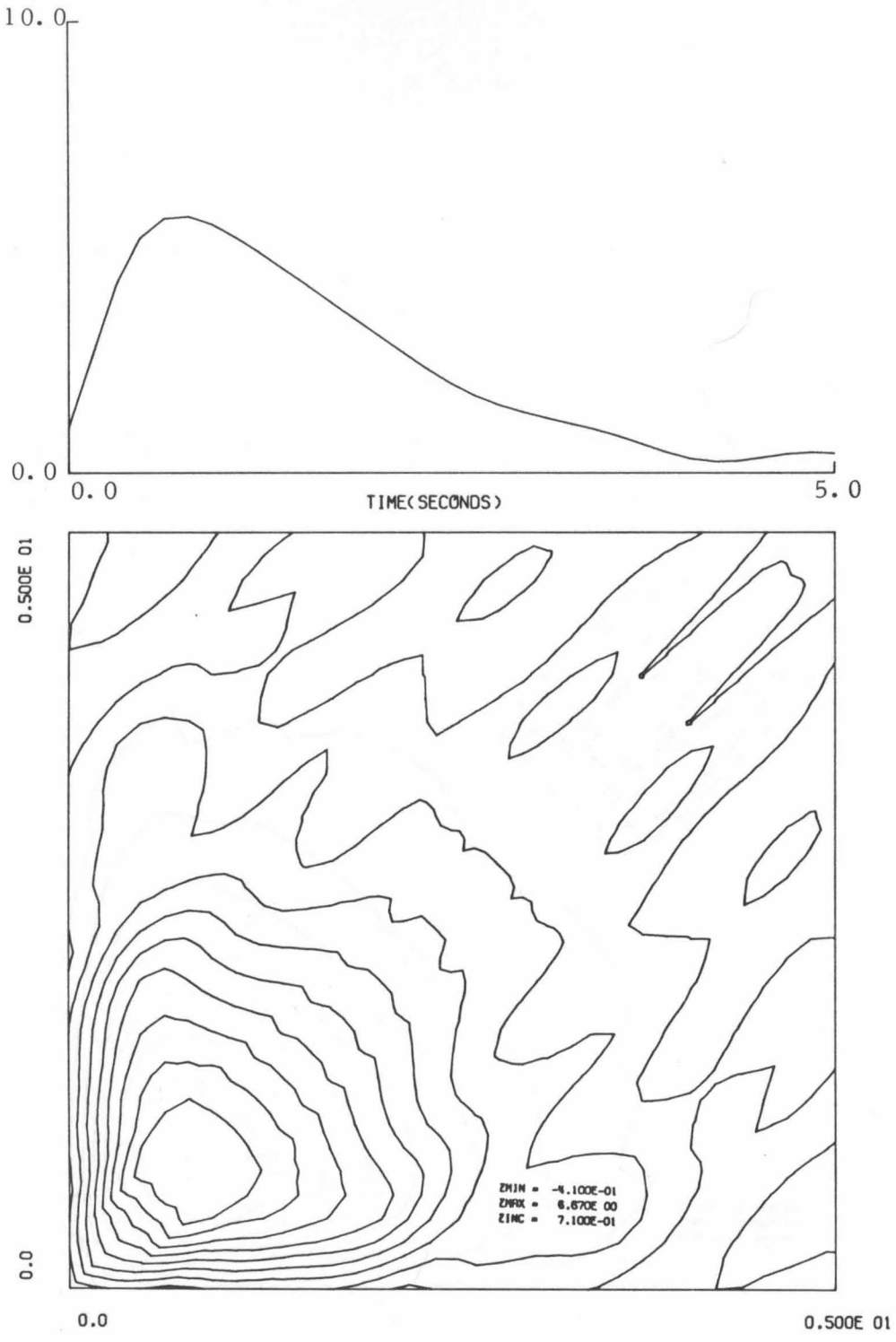


Fig. 10.1.9: First and second order GWN kernel estimates for system of Fig. 8.2.1.

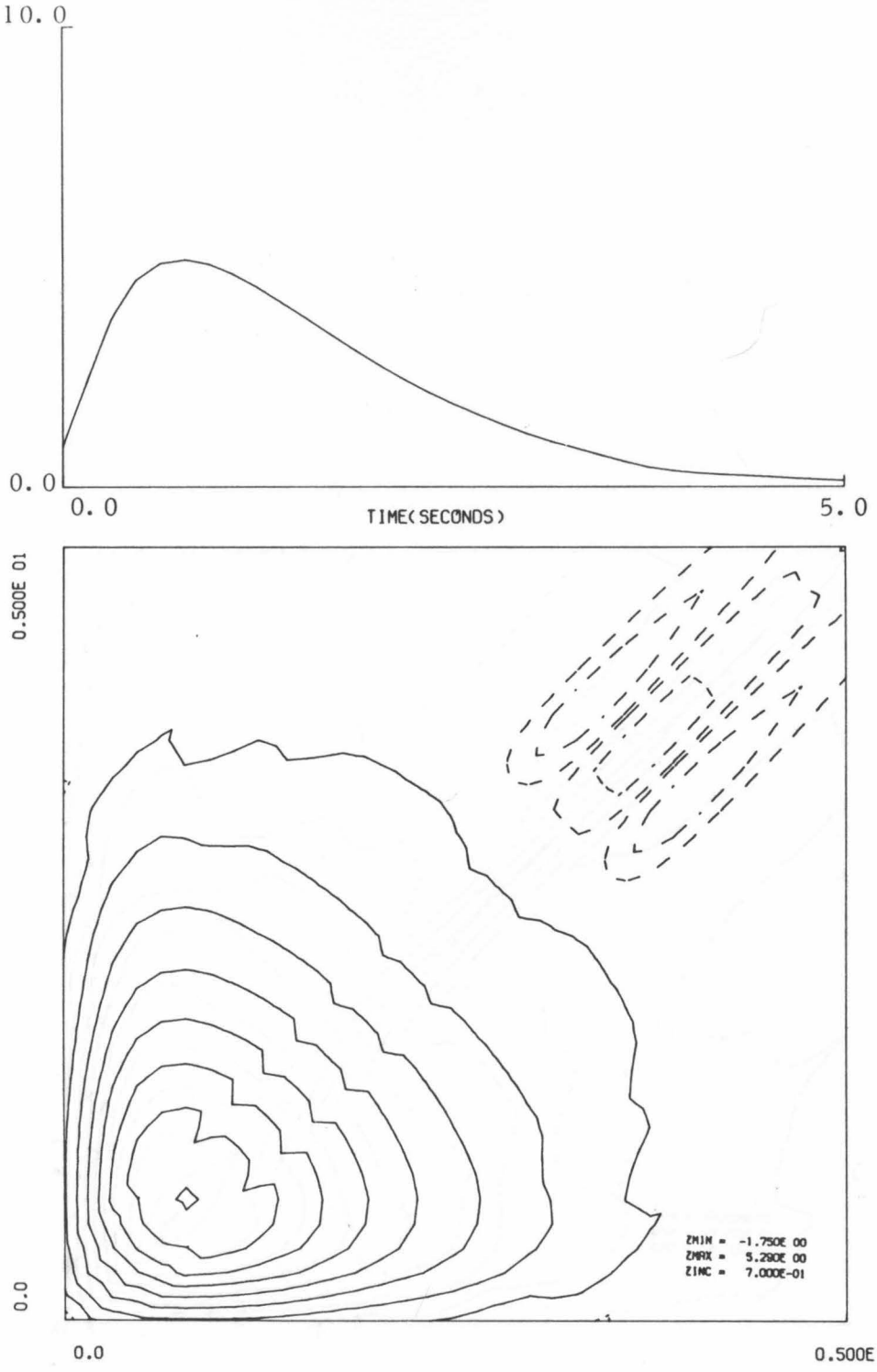


Fig. 10.1.10: First and second order PRS kernel estimates for system of Fig. 8.2.1.

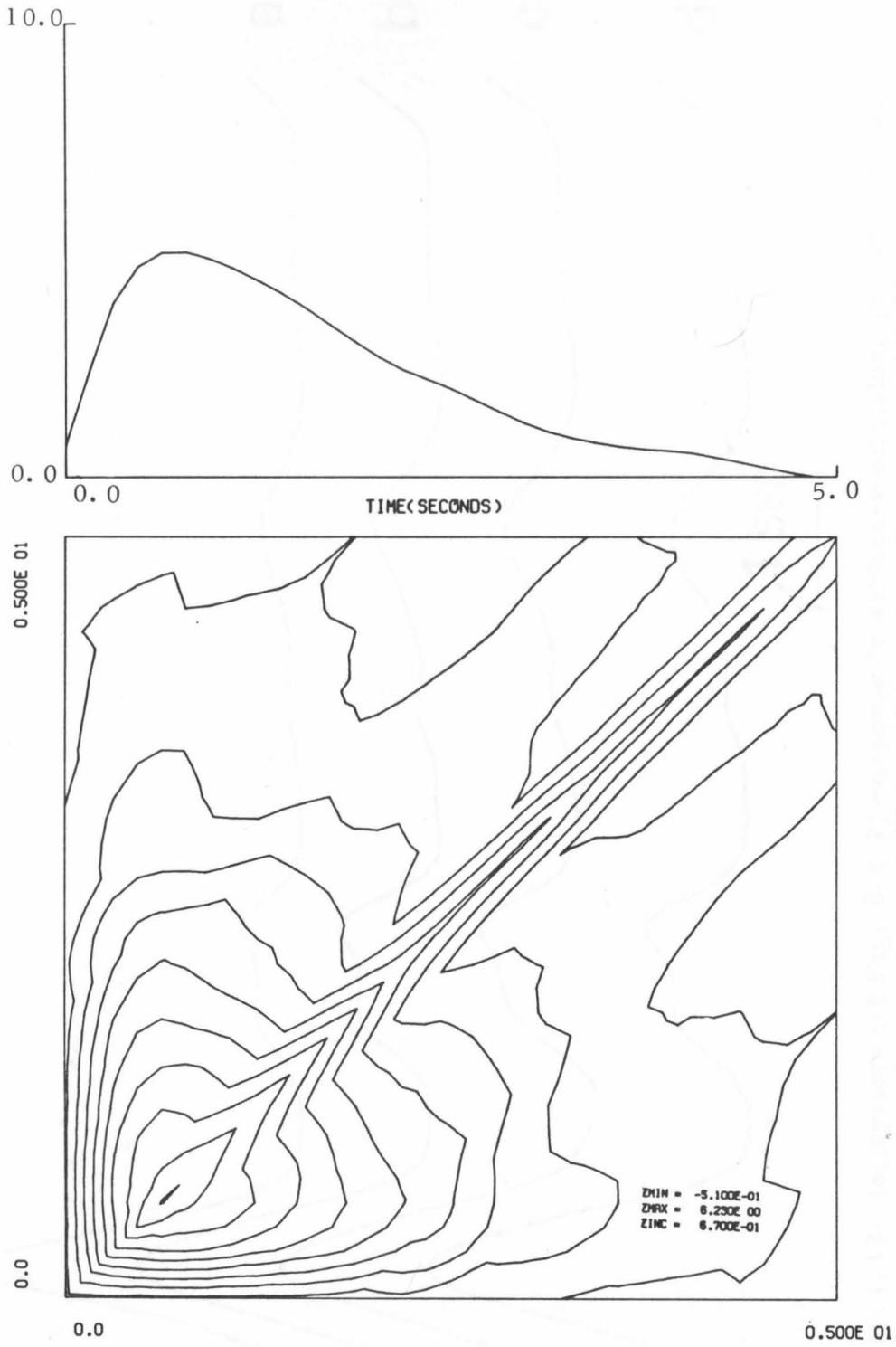


Fig. 10.1.11: First and second order CSRS kernel estimates for system of Fig. 8.2.1.

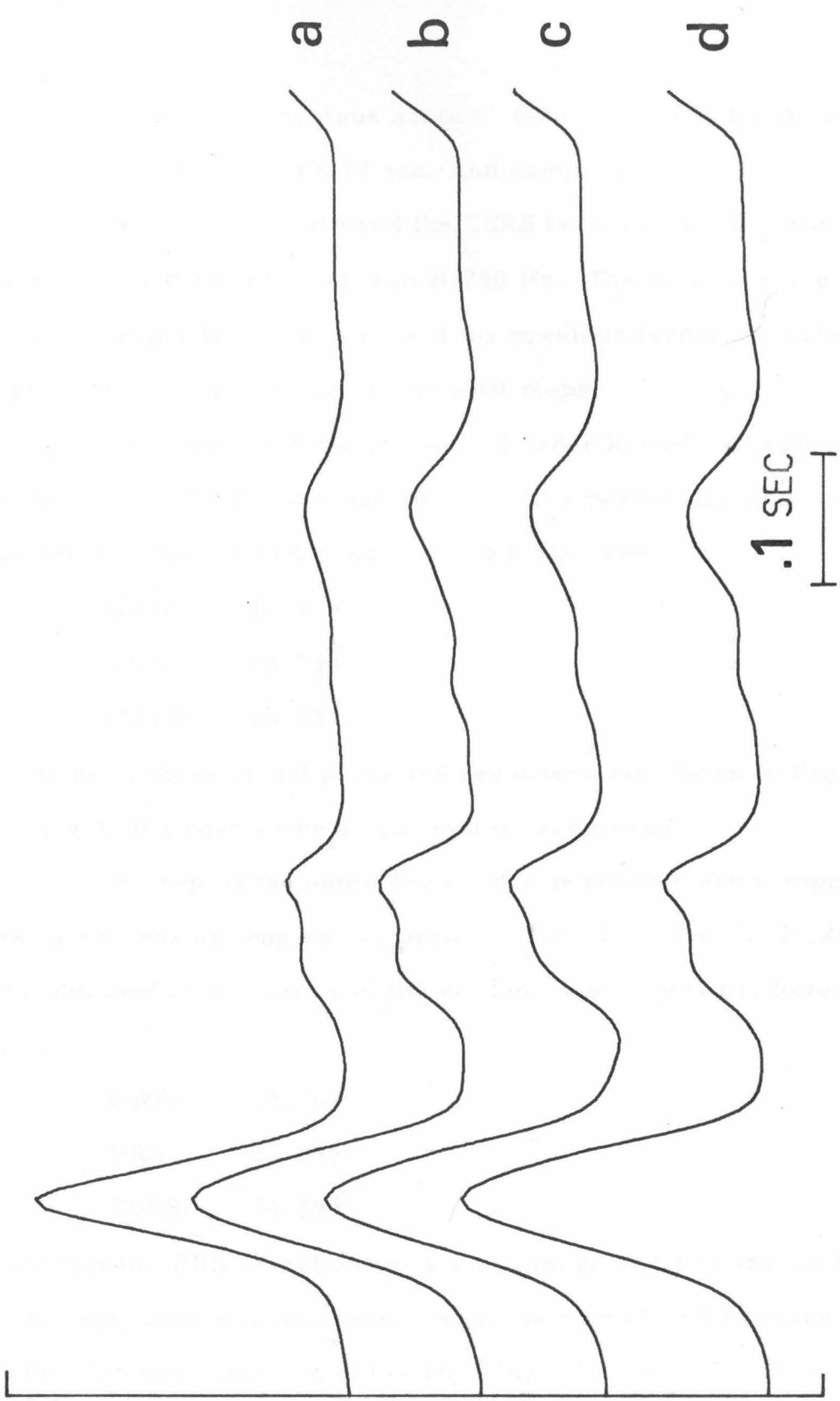


Fig. 10.1.12: (a) system (of Fig. 8.2.1) response to arbitrary stimulus, (b) GWN model predicted response, (c) PRS model predicted response, (d) CSRS model predicted response.

As in the previous section, all these test signals have the same record length T=13.12 sec. and sampling interval DT=.001 sec. The step length of the PRS and the CSRS is taken Δt=.002 sec. The bandwidth of the GWN is chosen B=250 Hz. The record length T=13.12 sec. corresponds to one period of the previous ternary pseudorandom signal (cf. sec. 10.1), which has 6560 steps.

The obtained first and second order kernel estimates are shown in Fig. 10.2.1 through 10.2.3. The percentage m.s. errors of the second order model predicted responses are:

GWN: 16.41%

PRS: 23.03%

CSRS: 22.35%

Portions of these model predicted responses are shown in Fig. 10.2.4, along with the respective actual system responses.

We repeat the same series of experiments and computations, taking records as long as two periods of the PRS (i.e. T=26.24 sec.). The obtained m.s. errors of the second order model predicted responses are:

GWN: 12.76%

PRS: 23.03%

CSRS: 14.39%

Note that the PRS kernel estimates and m.s. error of the model predicted response remain exactly the same no matter how many periods of the PRS are contained in the stimulus. This is, of course, theoretically anticipated; however, in a lot of actual applications, the accuracy

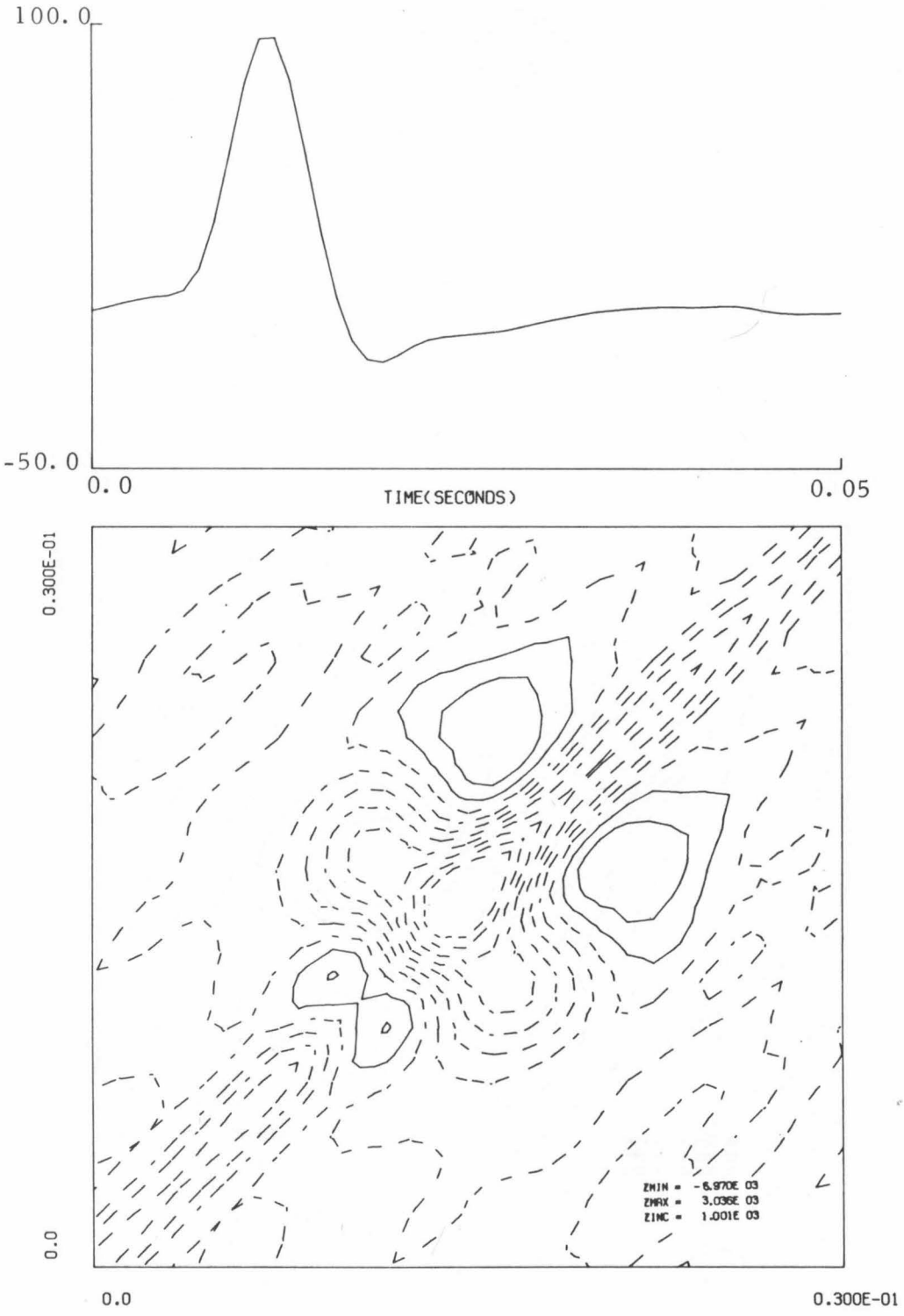


Fig. 10.2.1: First and second order GWN kernel estimates of photoreceptor.

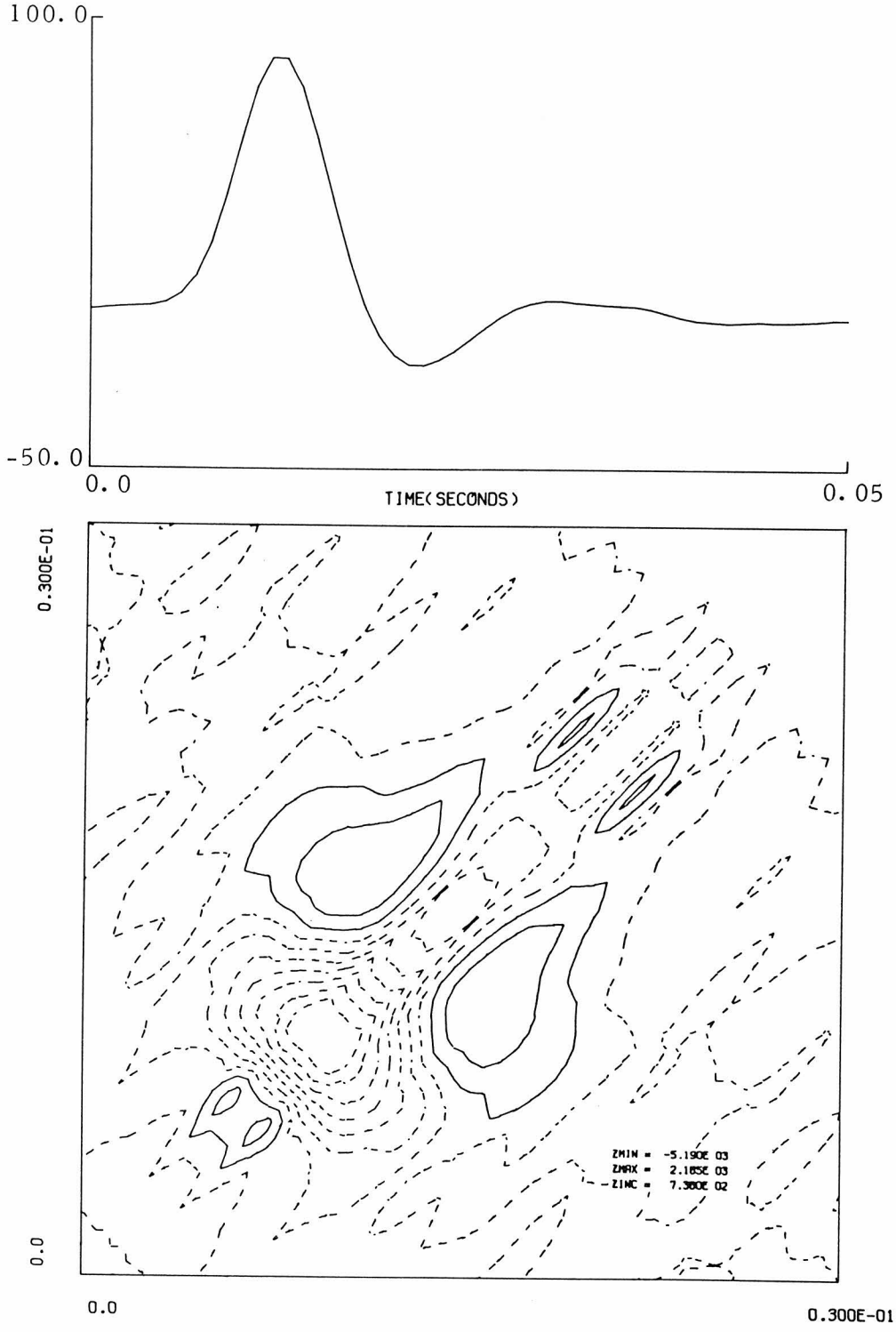


Fig. 10.2.2: First and second order PRS kernel estimates of photoreceptor.

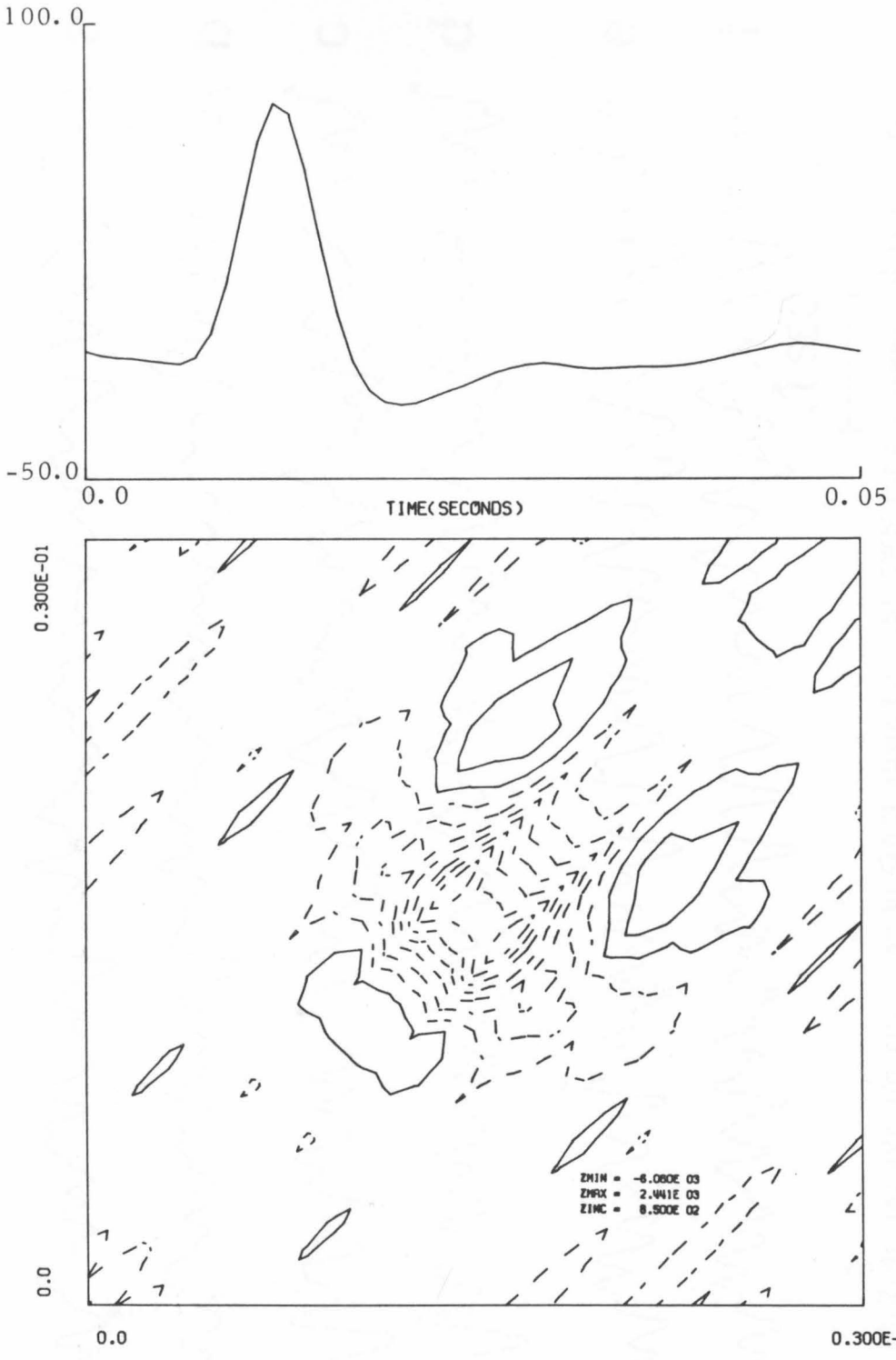


Fig. 10.2.3: First and second order CSRS kernel estimates of photoreceptor.

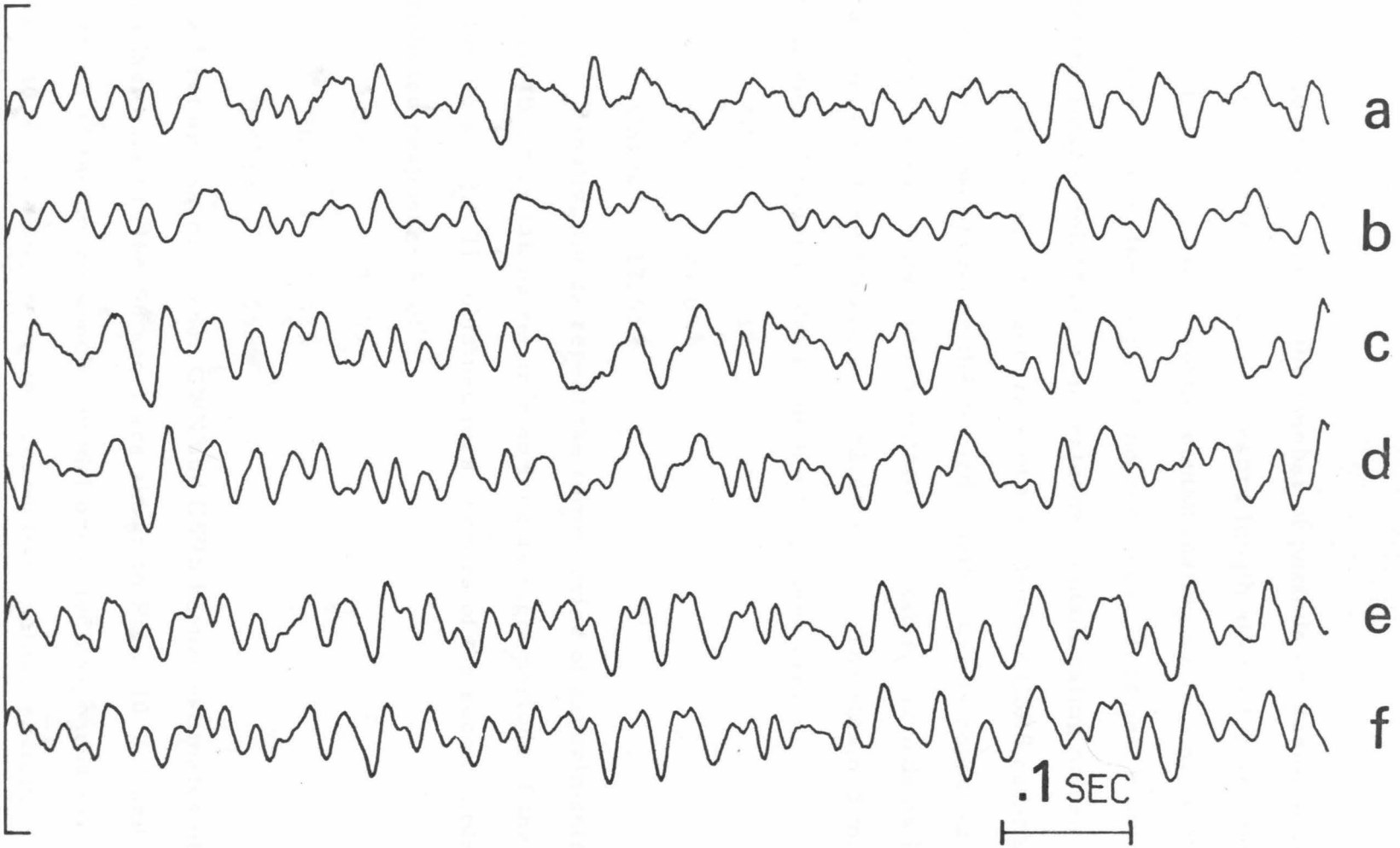


Fig. 10.2.4: (a) system response to GWN stimulus, (b) GWN model predicted response, (c) system response to PRS stimulus, (d) PRS model predicted response, (e) system response to CSRS stimulus, (f) CSRS model predicted response.

improves as we increase the number of periods contained in the stimulus signal, because the longer record length allows for greater reduction of the effect of the external contaminating noise. Apparently, in the present case, the length of one period (T=13.12 sec.) suffices for the practical elimination of the external contaminating noise.

Of course, the accuracy of the GWN and CSRS estimated model increases as we increase the record length. Let us repeat the same series of experiments and computations by taking records as long as four periods of the PRS (i.e. T=52.48 sec.). The obtained m.s. errors of the second order model predicted responses are:

GWN: 10.93%

PRS: 23.03%

CSRS: 12.55%

Finally, let us repeat the same series of experiments and computations by taking records as long as eight periods of the PRS (i.e. T=104.96 sec.). The obtained m.s. errors of the second order model predicted responses are:

GWN: 10.85%

PRS: 23.03%

CSRS: 5.55%

The first and second order GWN and CSRS kernel estimates obtained in this last case (T=104.96 sec.) are shown in Figs. 10.2.5 and 10.2.6. Portions of the corresponding model predicted responses are shown in Fig. 10.2.7, along with the respective actual system responses.

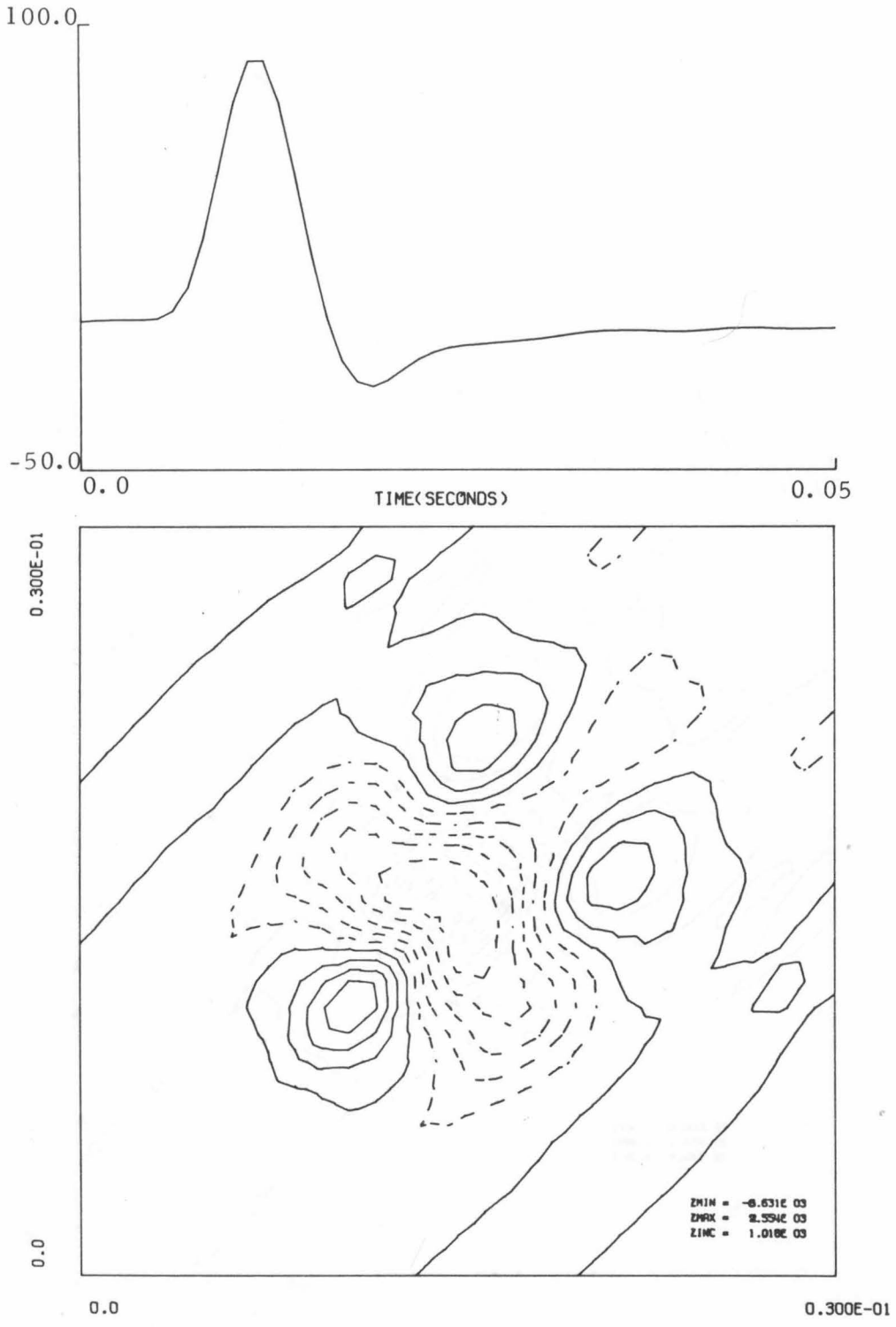


Fig. 10.2.5: First and second order GWN kernel estimates of the photoreceptor.

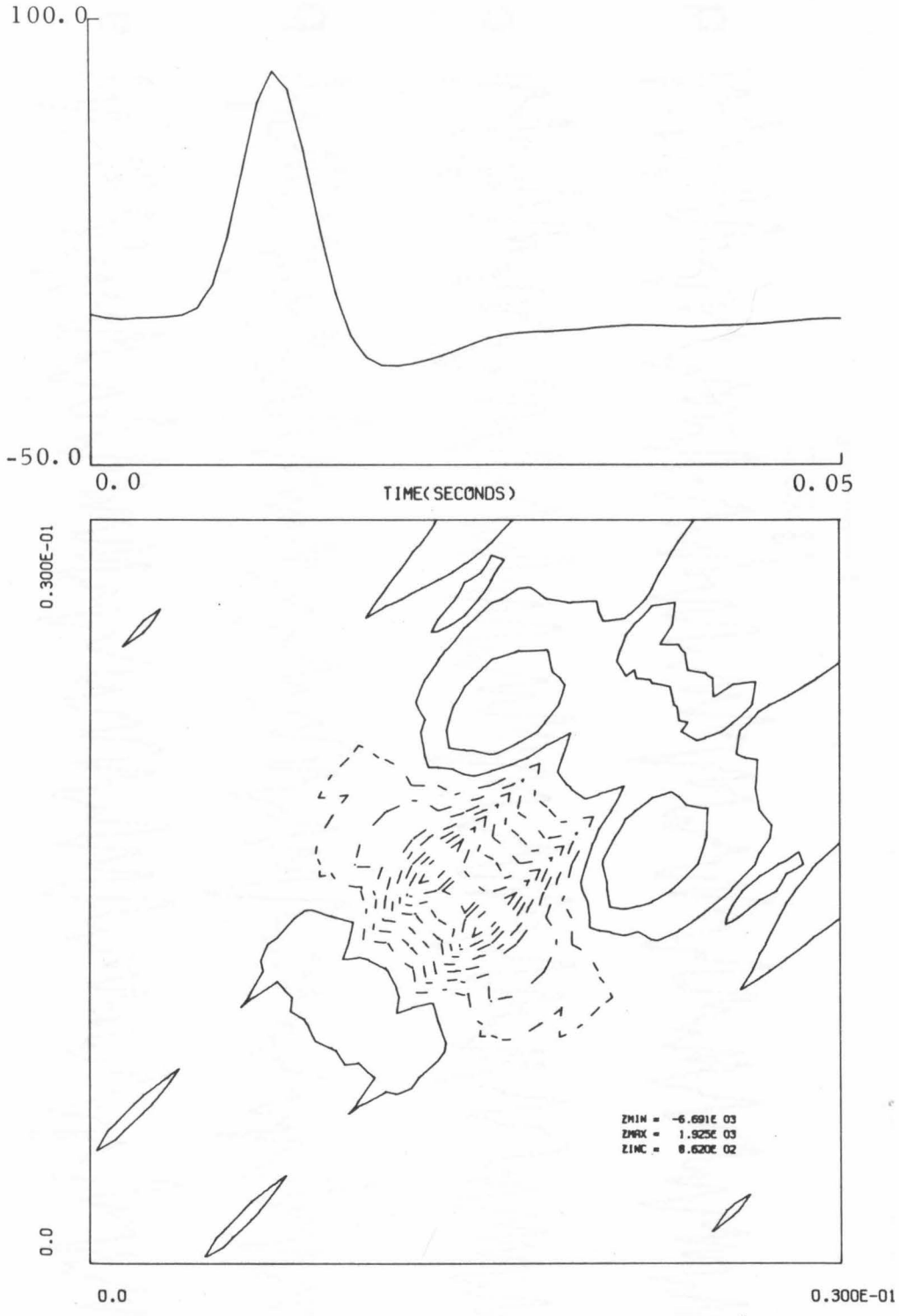


Fig. 10.2.6: First and second order CSRS kernel estimates of the photoreceptor.

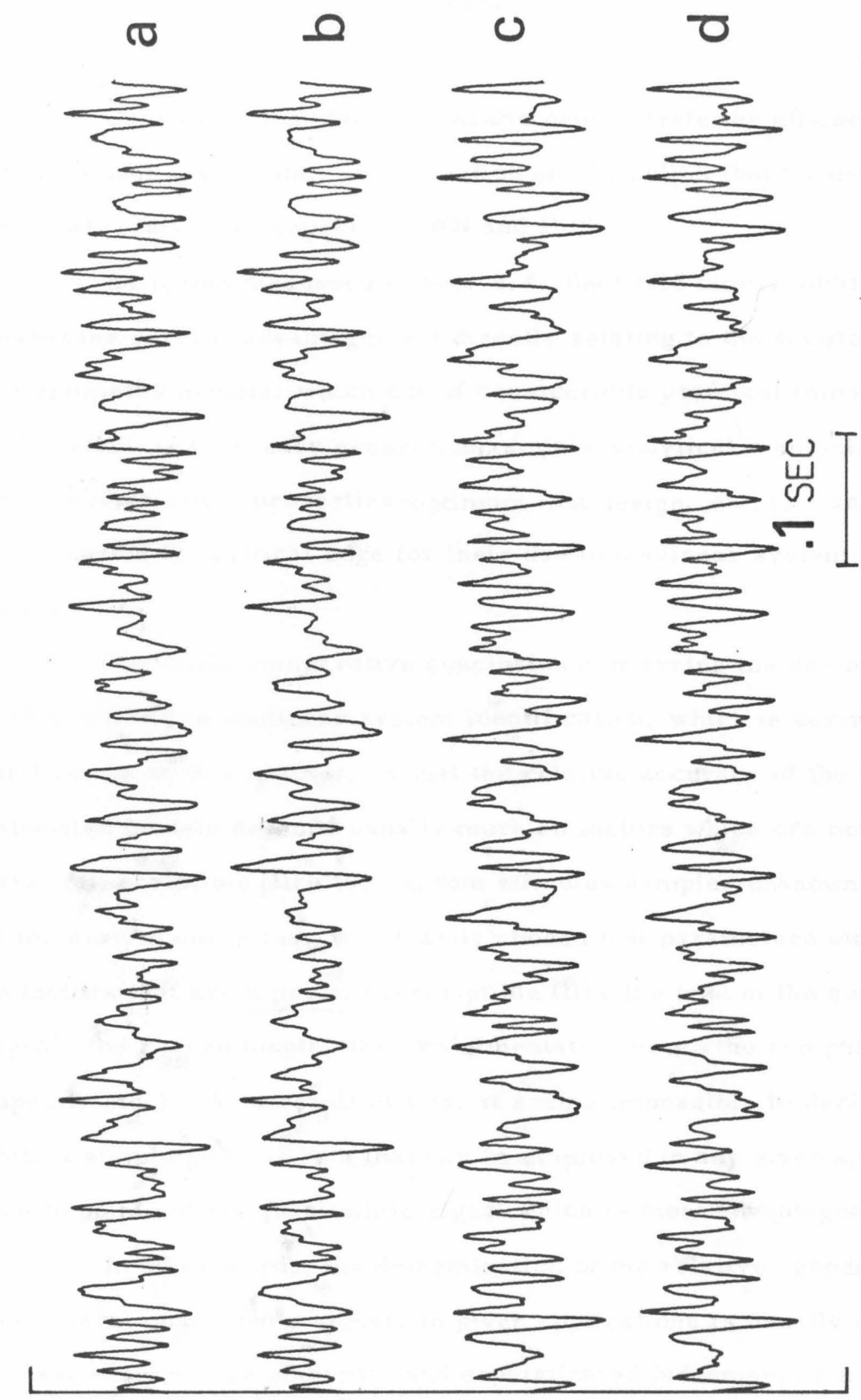


Fig. 10.2.7: (a) System response (photoreceptor) to GWN stimulus, (b) GWN model predicted response, (c) system response (photoreceptor) to CSRS stimulus, (d) CSRS model predicted response.

Once again the obtained results demonstrate the efficacy of the CSRS in nonlinear system identification and establish their non-inferiority (at least) with respect to GWN and PRS.

As it was discussed in sec. 4.4, the CSRS family additionally possesses several advantages not directly relating to the accuracy of the estimated models, which are of considerable practical importance and usefulness (i.e. easy generation, simple analytical manipulation of their mathematical properties-optimum test design, etc.), and which may provide the critical edge for their use in nonlinear system identification.

The basic comparative conclusion concerning the use of GWN, PRS and CSRS in nonlinear system identification, which is derived from the findings of this chapter, is that the relative accuracy of the several estimated models depends usually more on factors which are not a priori prescriptible (like the random stimulus sample, unknown features of the system under test, arbitrarily chosen test parameters etc.) than on factors that are a priori prescriptible (like the type of the quasi-white signal, the record length, the experimentation time, the computational capacity etc.). As a result of this, it seems impossible to derive and state a simple general rule that can be employed in any given application to point out the quasi-white signal which is most advantageous.

In other words, the determination of the relative "goodness" of the several quasi-white signals in given applications is usually a close contest requiring an elaborate and sophisticated judgement on a case-to-case basis.

It must be noted, however, that the CSRS provide a unifying general and systematic approach to the quasi-white test signal design problem; which, beyond the advantages in easy generation, good auto correlation properties and optimum determination of test parameters, also allows the optimization of the estimated system model with respect to a given environment of stimuli by an appropriate choice of the testing CSRS amplitude probability distribution and step length.

Important note:

The accuracy of the kernel estimates that are given as illustrations throughout chapters 8, 9, and 10, can be appreciably improved if the length of the record is increased accordingly. This is of course, desired when practical applications are being made of the method.

Since the intent of the present work is to study the several subjects pertinent to the "mathematics" of this specific identification method and not to provide accurate kernel estimates, we have used records of medium length, which give kernel estimates of reasonable accuracy without creating a heavy computational burden hardly justifiable for a study of that sort. To partly compensate for this intentional and justifiable deficiency of the given illustrations, we must note that in the case of the physiological system, that was used in chapters 9 and 10, the m. s. error can be reduced down to the neighborhood of 5% if records of several hundreds seconds long are used.

REFERENCES

1. Volterra, V.: "Theory of functional and of integral and integro-differential equations". Blackie, 1930.
2. Wiener, N.: "Nonlinear problems in random theory". New York: Wiley, 1958.
3. Lee, Y. W. and M. Schetzen: "Measurement of the kernels of a nonlinear system by crosscorrelation". Int. J. Contr., Vol. 2, pp. 237-254, 1965.
4. Marmarelis, P. Z. and G. D. McCann: "Development and application of white noise modeling techniques for studies of insect visual nervous systems". Kybernetik, Vol. 12, pp. 74-89, 1973.
5. Marmarelis, P. Z. and K. I. Naka: "Identification of multi-input biological systems". IEEE, Transactions on Biomedical Eng., Vol. 21, No. 2, 1974.
6. Yasui, S., Fender, D. H.: "Methodology for measurement of spatio-temporal Volterra and Wiener kernels for visual systems". Proc. of 1st Symp. on Testing and Identification of Nonlinear Systems, Calif. Inst. of Technology, March 1975, pp. 366-383.
7. Bose, A. G.: "A theory of nonlinear systems". Res. Lab. Electron. MIT, Cambridge, Mass., Tech. Rep. 345, 1958.
8. Brilliant, M. B.: "Theory of the analysis of nonlinear systems". Res. Lab. Electron., MIT, Cambridge, Mass., Tech. Rep. 345, 1958.
9. George, D. A.: "Continuous nonlinear systems". Res. Lab. Electron. MIT, Cambridge, Mass., Tech. Rep. 355, 1958.
10. Barrett, J. F.: "The use of functionals in the analysis of nonlinear physical systems". J. Electron. Contr., Vol. 15, pp. 567-615, 1963.

11. Stark, L.: "Neurological control systems, Studies in bioengineering"., Sec. II, Ch. 4, Plenum Press, N.Y., 1968.

12. Marmarelis, P.Z., Naka, K.-I.: "White-noise analysis of a neuron chain: An application of the Wiener theory". Science 175, 1276-1278, 1972.

13. Marmarelis, P.Z.: "Nonlinear identification of bioneuronal systems through white-noise stimulation". Proc. 13th Joint Automatic Control Cont., Stanford, Calif., pp. 117-126, 1972.

14. Marmarelis, P.Z., Naka, K.-I.: "Nonlinear analysis and synthesis of receptive-field responses in the catfish retina. I: Horizontal cell → ganglion cell chain". J. Neurophysiol. Vol. 36, pp. 605-618, 1973.

15. Marmarelis, P.Z., Naka, K.-I.: "Nonlinear analysis and synthesis of receptive-field responses in the catfish retina. II: One-input white-noise analysis". J. Neurophysiol. Vol. 36, pp. 619-633, 1973.

16. Marmarelis, P.Z., Naka, K.-I.: "Nonlinear analysis and synthesis of receptive-field responses in the catfish retina. III: Two-input white-noise analysis". J. Neurophysiol. Vol. 36, pp. 634-648, 1973.

17. McCann, G.D.: "Nonlinear identification theory model for successive stages of visual nervous systems of flies". J. of Neurophysiol. Vol. 37(5), pp. 869-895, 1974.

18. Udwadia, F., Marmarelis, P.Z.: "Identification of building structural systems; Part I: the linear case". Bulletin of Seismological Society of America. Feb., 1976 (to be published).

19. Udwadia, F., Marmarelis, P.Z.: "Identification of building structural systems; Part II: the nonlinear case". Bulletin of Seismological Society of America. Feb., 1976 (to be published).

20. Korn, G. A.: "Random process simulation and measurements". McGraw-Hill, 1966.

21. French, A. S.: "Synthesis of low-frequency noise for use on biological experiments". Trans. IEEE Biom. Eng., Vol. 21, pp. 251-252, 1974.

22. Zierler, N.: "Linear recurring sequences", J. Soc. Industr. Appl. Math. Vol. 7, No. 1, pp. 31-48, 1959.

23. Church, R.: "Tables of irreducible polynomials for the first four prime moduli", Ann. Math., Vol. 36, p. 198, 1935.

24. Ream, N.: "Nonlinear identification using inverse-repeat m-sequences". Proc. Inst. Elec. Eng., Vol. 117, pp. 213-218, 1970.

25. Simpson, H. R.: "Statistical properties of a class of pseudorandom sequences". Proc. Inst. Elec. Eng., Vol. 113, pp. 2075-2080, 1966.

26. Gyftopoulos, E. P. and R. J. Hooper: "Signals for transfer function measurement in nonlinear systems". Noise analysis in nuclear systems, USAEC symposium, series 4, TID-7679, 1964.

27. Gyftopoulos, E. P. and R. J. Hooper: "On the measurement of characteristic kernels of a class of nonlinear systems", Neutron noise, waves and pulse propagation: USAEC Conference Report 660206, 1967.

28. Barker, H. A. and T. Pradisthayon: "High-order autocorrelation functions of pseudorandom signals based on m-sequences", Proc. Inst. Elec. Eng., Vol. 117, pp. 1857-1863, 1970.

29. Barker, H. A., Obidegwn, S. N., Pradisthayon, T.: "Performance of antisymmetric pseudorandom signals in the measurement of 2nd-order Volterra kernels by Crosscorrelation", Proc. IEE, Vol. 119, No. 3, pp. 353-362, 1972.

30. Balcomb, J. E., H. B. Demuth, and E. B. Gyftopoulos: "A cross-correlation method for measuring the impulse response of reactor systems". Nuclear Sci. Eng., Vol. 11, pp. 159-166, 1961.

31. Briggs, P. A. N., Hammond, P. H., Hughes, M. T. G., and Plumb, G. O.: "Correlation analysis of process dynamics using pseudorandom binary test perturbations", Proc. Instn. Mech. Eng., 1964-1965, Vol. 179, p. 37.

32. Ream, N.: "Testing a 2-input linear system with periodic binary sequences", Proc. IEE, Vol. 114, No. 2, pp. 305-307, 1967.

33. O'Leary, D. P., R. F. Dunn, and V. Honrubia: "Functional and anatomical correlation of afferent responses from the isolated semicircular canal", Nature, Vol. 251, Sept. 20, 1974, pp. 225-227.

34. Godfrey, K. R. and N. Murgatroyd: "Input transducer errors in binary crosscorrelation experiments", Proc. IEE, Vol. 112 (3), p. 565, 1965.

35. Godfrey, K. R., Everett, D., and P. R. Bryant: "Input transducer errors in binary crosscorrelation experiments - 2", Proc. IEE, Vol. 113, No. 1, p. 185, 1966.

36. Aidley, D. J.: "The physiology of excitable cells", Cambridge Univ. Press, N. Y., 1971.