THE UNIVERSITY OF CALGARY

Probabilistic Error Regulating Functions, Their Properties and Design Within an Inversion Framework

by

Donald T. Easley

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF GEOLOGY AND GEOPHYSICS

CALGARY, ALBERTA

APRIL, 1987

ISBN 0-315-38005-5

THE UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled, "Probabilistic Error Regulating Functions, Their Properties and Design Within an Inversion Framework", submitted by Donald T. Easley in partial fulfillment of the requirements for the degree of Master of Science.

Justice

Dr. Jim Justice, Chairman
Dept. of Geology & Geophysics

R. J. Brown

Dr. R. J. Brown
Dept. of Geology & Geophysics

RBHicks

Dr. R. B. Hicks
Dept. of Physics

JAR. Blais

Dr. J. A. R. Blais
Dept. of Survey Engineering

May 29 1987

ABSTRACT

The use of ℓ^p norms in the regulation of errors in inversion is well known (least-squares corresponding to p=2, least-absolute value corresponding to p=1). In this thesis a new class of regulating function (PERFs) will be defined of which the ℓ^p norms are a subset. By consideration of inversion problems in a general framework, a set of defining properties of PERFs was arrived at. The form of PERFs thus defined allows a relationship to be drawn between a particular PERF and the underlying probability distribution of the gauged quantity. By variational techniques, plus a constraint, the extremum of a PERF is related to a specific probability distribution. Since extrema of PERFs are used in inversion to indicate an optimum, this provides a means to describe the statistical effect a particular PERF may have on the probabilistic part of an inversion procedure. The results applied to ℓ^p norms provide a testable family of probability distributions. Two properties of these derived distributions were tested using a couple of geophysical inversion problems. The first is constant value extraction, which is used to test the drift of the maximum of the distributions (the mode) from the origin as "p" is increased. The second is minimum entropy deconvolution, which tests the general shape of the distribution for different "p's". Both tests showed good qualitative agreement. A method to tailor PERFs based on <u>a priori</u> moment information is shown and used in conjunction

with deterministic properties in automated constant phase shift correction.

ACKNOWLEDGEMENTS

I wish to give the greatest thanks to "Mommy" who gave so much with no strings attached making this possible. I wish also to thank my wife and children for their forbearance during this period of time.

I wish to thank Dr. Jim Justice for the opportunity to get a master's degree and for the encouragement when I was about to "pack it in". I also wish to thank Jim for listening to my wild ideas and giving useful counsel. I am also grateful to the other members of my examining committee for their time and effort, specially Dr. Jim Brown for his careful proofreading and suggestions, Dr. Bart Hicks for taking time to discuss this thesis with me and Dr. Blaise for bringing to my attention many corrections that needed to be made.

Appreciation must also be expressed for Amoco Canada's support in all facets of this work. Special thanks go to Ken West and Amin Abdel-Kader of Amoco for providing the resources needed to accomplish this study during a period of much activity.

The many individuals who have contributed to this work are thanked here. There are two deserving of special mention; they are Mark Lane, for his patient reading of this thesis and his many suggestions; and Don Lacy, for reading a rather unpolished first draft.

Lastly, many thanks must be given to Cheryl Cross for her great effort in typing much of this thesis.

CONTENTS

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

(1.1) HISTORICAL BACKGROUND

The convolution model has proven to be very useful in geophysics. The forward process of sending a seismic pulse into the ground and recording its echoes from subsurface impedance boundaries results in a seismogram that can be modelled in this way. The seismogram can be viewed as the convolution of the pulse wavelet with a reflectivity sequence from the subsurface (Robinson and Treitel [23]). For an exploration seismologist the desired information is the reflectivity sequence. To extract this information, the effects of the forward convolution process must be undone. The undoing of the forward process to extract desired information is the inverse process. As a first step, the estimation of the wavelet and the removal of its effect is usually attempted. There are many ways to do this. Some examples are: Wiener-Levinson double inverse method (minimum-phase least-squares filtering), Wold-Kolmogorov factorization (log-Hilbert transform technique) and homomorphic deconvolution wavelet estimates (cepstral liftering) (Lines and Ulrych [18]). Of these the Wiener-Levinson method is most akin to inverses considered in this thesis. This method finds a filter which "best" shapes the wavelet to a spike (delta function) with a minimum phase assumption (Claerbout [4]). The term "best" means the differences between the filtered result and an actual spike have the least squared sum. At this point the inverse could be considered complete. But, even if the wavelet is known, information may still be lost because of the wavelet having zeros in its frequency spectrum. Techniques to attempt reconstruction of the reflectivity sequence within these zones of zero frequency spectrum also exist. One such technique reconstructs the reflectivity by making statistical assumptions about the additive noise and using the maximum likelihood principle (Ursin and Holberg [25]). Another technique, which is closer to the concerns of this thesis, finds the full-band (containing all frequencies) reflectivity which has the least absolute value sum, subject to linear constraints (Levy and Fullagar [15]). Techniques to simultaneously estimate both wavelet and reflectivity by minimizing the sum of squared differences between the modeled seismogram and the actual seismogram are available (Lines and Treitel [17]).

The solution of overdetermined systems of equations (OSE) is intimately related to inversion (Twomey [24]). The use of functions such as the sum of squares or the sum of absolute values, which are special cases of ℓ^p norms (Goffman and Pedrick [8]), as criteria of "best" solution has a long history in OSE problems. The use of least-squares techniques (minimization of the ℓ^2 norm) dates back to 1806, when Legendre suggested its use as a criterion in the solution of OSE; later Laplace in 1811 and Gauss during 1821-23 placed the least squares technique on firmer statistical grounds (Whittaker and Robinson [26]).

However, the least-squares criterion is not the only possible constraining criterion in the solution of OSE's. The least absolute value criterion (minimization of the ℓ^1 norm) can also be used for this problem (Barrodale and Roberts [1]). In fact there is no reason conceptually why any of the ℓ^p norms cannot be used. The use of ℓ^p norms is not restricted to geophysics and OSE problems. Another example, from many others, is in image restoration (Justusson and Tyan [13]). The popularity of least-squares algorithms is partly attributable to their ease of development. Least absolute value and other function based algorithms require more elaborate techniques, such as linear programming (Hadley [11]). Many functions can be used to regulate an inverse procedure, with different "regulating functions" giving differing solutions for the same input. A better understanding of regulating functions is useful in the design of inverse procedures.

(1.2) OVERVIEW

This thesis deals with the application of a particular class of regulating functions to inversion problems. It is apparent from observed results of inversion procedures and statements like, "minimization of this norm (ℓ^1 norm) favours solutions with as few non-zero values as possible" (Levy and Fullagar [15]), that the statistical properties of an inverse procedure can be affected by the choice of regulating functions.

A class of regulating functions will be defined, of which the previously described ℓ^p norms are a subset. The term "probabilistic error regulating functions" (PERFs) will be adopted for this class. An attempt will be made to quantify the aforementioned statistical effect of PERFs. A few geophysical inversion examples will be used to investigate the results obtained.

PERFs will be related to commonly used regulating functions, and their usefulness in solving practical geophysical problems will be shown. A technique to construct PERFs from a priori information will be given.

In particular, constant value extraction (as used in stacking) and minimum entropy deconvolution will be used to test the statistical properties of PERFs. Automated constant phase shift correction will be used to demonstrate how to construct a customized PERF from a priori information. The use of deterministic properties to influence the statistics will also be shown in this example.

CHAPTER 2
CONCEPTUAL DEVELOPMENT

(2.1) FORWARD/INVERSE PROCESSES AND REGULATING FUNCTIONS

To unify some of the many special situations that can arise, some of the terms used in this thesis will be given more general definitions.

Forward processes encrypt information in an output object. Inversion techniques attempt to deduce information from observed quantifiable characteristics of the object.

All forward processes considered in this thesis can be viewed as special cases of the following abstract experiment. The experiment consists of an input object "Y" which is corrupted by some process "S" resulting in an output object "X" (Figure 2.1.1).*

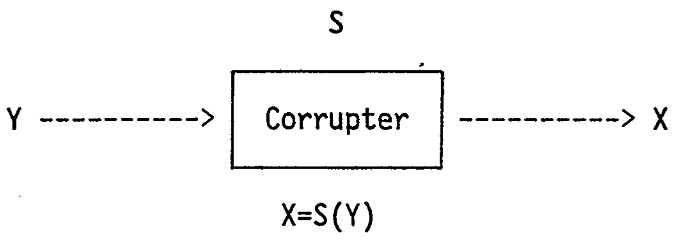


Figure 2.1.1
Forward Process as an Abstract Experiment

---

* At this point "Y" can be seen as an abstract embodiment of information, and "X" is a corrupted version of that information. "S" is just the recipe for how to corrupt "Y".

The inverse process attempts to deduce "Y" from some knowledge of "X".

Central to a broad group of inversion techniques is the use of regulating functions. Regulating functions are a set of rules by which characteristics of interest between pairs of objects can be mapped into the real numbers. A regulating function is then useful when the natural ordering of the real numbers relates meaningfully to our concept of similarity between the objects.

(2.2) THREE BROAD CLASSES OF PROCESSES

To see when regulating functions are useful, it is convenient to make some classifications based on the relationship between the forward process and its inverse. These classifications also provide a general framework from which desirable properties of PERFs can be drawn, so that special cases will not determine these properties. There are three broad classes. The first and last can be seen as end members, while the central classification can be seen as the spectrum bounded by these.

The first class will be called "wholly deterministic and invertible processes". In this class the object "Y" can be recovered from "X" without error. Equivalently, the inverse "S^{-1}" must be known exactly (Figure 2.2.1).

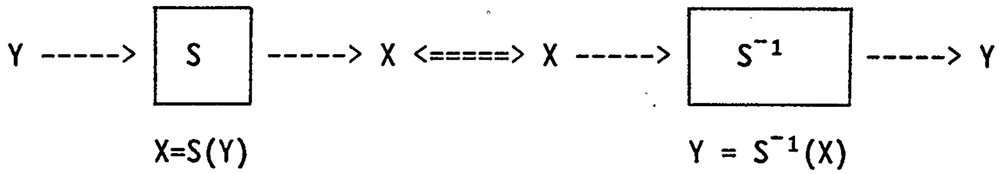


Figure 2.2.1
Wholly Deterministic and Invertible Processes

Examples:

(1) The forward process "S" consists of adding a known constant to the input "Y". The inverse "S^{-1}" is obviously the subtraction of this known constant from the output "X".

(2) "S" is discrete convolution, or Z-transform multiplication. "Y" is the input sequence to be convolved with known coefficients, resulting in the output sequence "X". "S^{-1}" is deconvolution or polynomial division. Here the coefficients are assumed to allow this. An obvious example where this cannot be done is when all the coefficients are zero (Claerbout [4]).

The second class will be termed "mixed processes". In this class, the exact inverse "S^{-1}" is unknown; only a close estimate to "Y" can be derived from "X" using an approximate inverse "$\tilde{S}^{-1}$" (Figure 2.2.2).

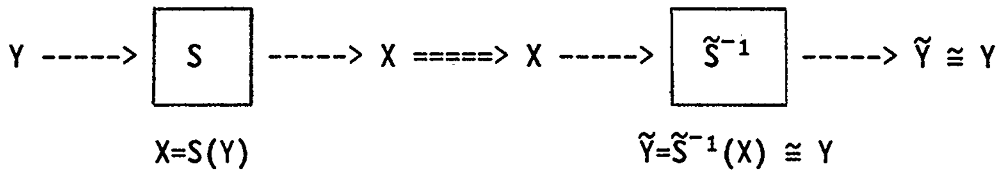


Figure 2.2.2
Mixed Processes

The inexact nature of "$\tilde{S}^{-1}$" provides the opportunity for the use of a regulating function. This regulating function provides some assurance that "$\tilde{Y}$" is in some manner close to "Y".

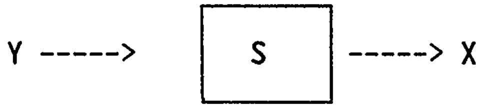
Examples:

(1) Extracting a linear trend in data that are unpredictably scattered (Neville and Kennedy [21]).

(2) Surface consistent static calculations in areas with near surface raypath anomalies (Musser, King and Wason [20]).

For some other special examples of this process refer to appendix E (E1-E5).

The third and last class will be called "totally uninvertible processes". This class contains two distinct subclasses. In each of these the output "X" is not related in any known manner to the input "Y". The first subprocess is "deterministic but uninvertible process". Here, the output "X" is not related to "Y", but "X" is always predictable (figure 2.2.3a).



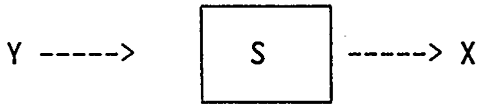
S(anything) = X such that X is predictable
or $S(.) = X \ni X$ is predictable.

Figure 2.2.3a
Deterministic but Uninvertible Processes

Example:

Regardless of the input "Y", the output is a known function.

The second, more interesting, subclass is "random processes". The output "X" for this process is again not related to the input "Y" in any <u>known</u> manner. In this case, "X" is totally unpredictable, although "Y" may in fact be hiding in the output "X" (Figure 2.2.3b).



$S(.) \Longrightarrow X \ni X$ is random

Figure 2.2.3b
Random Processes

Example:

The velocity of a particle exhibiting Brownian motion at constant temperature (Reif [22]).

Regulating functions, strangely enough, do enter this process in an indirect manner. The absolute sample moments of "X" can be determined. These may relate to the underlying probability density "P(x)". This is especially true if "P(x)" is symmetrical. As mentioned previously , the ℓ^p norms are a subset of regulating functions, and are defined for discrete "X" to be:

$$\|X\|_p \equiv \ell^p \text{ norm of } X = [\sum_{i=1}^{n} |x_i|^p]^{1/p}$$

The absolute sample moments for "X" are defined to be:

$$M_p \equiv p^{th} \text{ absolute sample moment } = \frac{1}{n} \sum_{i=1}^{n} |x_i|^p.$$

This gives the following relationship.

$$M_p(X) = \|X\|_p^p/n.$$

This ends the description of the three broad classes. As can be seen from the examples in appendix E (E1-E5), mixed processes in fact can be quite simple, almost falling into the first class considered (appendix E section E1) or they can be very complicated, and almost falling into the last class (appendix E section E2). Usually, it is possible to identify a mixed process as being comprised of deterministic (as in the first class, deterministic and invertible) and probabilistic (as in the third class, random) parts. The deterministic part is amenable to direct inversion as illustrated in the first case (figure 1.4.1). The probabilistic part can be controlled through regulating functions.

(2.3) DESIRABLE PROPERTIES OF PERFs

The usefulness of regulating functions in the inversion of mixed processes has been discussed. By consideration of mixed inversion schemes a few desirable qualities of regulating functions have emerged. These qualities will be couched in terms that may be a bit vague at present. This is because of the general background from which they are drawn. The terms will be made clear in chapter 3 (mathematical development).

As mentioned in section (2.2) regulating functions, in particular PERFs, are used to regulate the probabilistic part of an inversion procedure (refer to appendix E (E1-E4) for specific examples). This implies the object to be gauged by the regulating function is statistical in nature and cannot be known exactly. At best, an underlying probability distribution can be guessed at. This suggests the regulating function should respond to the magnitude of the elements comprising an object and not to the order which they occur.

From the hypothetical inversions considered in appendix E (E1-E5) the sets of objects compared by regulating functions are completely defined and manageable. In actual practice, the sets are usually infinitely large. Thus, objects are considered one at a time, with future objects derived from past objects in such a manner as to extremize the regulating function. This suggests that if the derived object is different from its predecessor only in global terms (this will be made clear in chapter 3) the regulating function should not attain an extremum, unless the compared objects are identical or some inbuilt deterministic constraint has been reached. This allows the search for an optimum to proceed smoothly away from choices that differ only in global terms. Also, the regulating function should not have properties which would make the search for an extremum ambiguous. This is also to ensure that the regulating function will be smoothly guidable towards an extremum.

Since the purpose of a regulating function is to compare two objects, when these two objects are identical the regulating function should attain an absolute extremum. The convention in this thesis is to let the regulating function become zero when this occurs and in all other cases let the regulating function be greater than zero.

The last condition is fundamentally different from the previous ones. The previous properties attempt to establish the form of regulating functions. The last condition relates extrema of regulating functions to their statistical effect. This is important since extrema of regulating functions are used to indicate optimum conditions in inversion.

To summarize, a regulating function should have the following characteristics.

(2.3.1) A regulating function should respond only to the magnitude of the elements comprising an object and not to the order in which they occur.

(2.3.2) Constant global changes to an object should not result in an extremum of a regulating function.

(2.3.3) A regulating function should not have properties which would make the search for an extremum ambiguous.

(2.3.4) A regulating function should attain an absolute extremum when the two objects compared are identical.

(2.3.5) An attempt should be made to relate the extremum of a regulating function to its statistical effects.

These conditions will be put in a mathematical framework in chapter 3.

CHAPTER 3

MATHEMATICAL DEVELOPMENT OF REGULATING FUNCTIONS

(3.1) FROM CONCEPTS TO MATHEMATICS

In the previous chapter, a conceptual framework was established in which regulating functions were of central importance. A set of desirable qualities that a regulating function should possess was outlined. In this section, the concepts developed will be put in a more usable mathematical form.

To begin, a few of the terms used in the last section will be paired with their mathematical analogs. The objects considered here will be vectors. A regulating function is then an indicator of similarity between vectors. Given a set "$\underline{V}$" of vectors over an appropriate field, the regulating function "L" maps pairs of elements of "$\underline{V}$" into the set of real numbers "R" or in symbols,

$$L: \underline{V} \times \underline{V} \dashrightarrow R \, .$$

The real number thus obtained will be called the distance between the two vectors. The distance between a vector and the origin will be called its length.

At this point, it is possible to incorporate condition (2.3.1) into a class of regulating functions. The condition states:

> A regulating function should respond only to the magnitude of the elements comprising an object [vector] and not to the order in which they occur.

This condition makes the regulating function a true statistical regulating function. The desired property is embodied in regulating functions of the following form. Let

$$X, Y \in \underline{V}$$

then

$$L(X,Y) = \Gamma[\sum_i \mu(y_i - x_i)] \quad : \underline{V} \equiv \text{discrete} \qquad (3.1.1)$$

$$= \Gamma[\int_a^b \mu(y(t)-x(t))dt] \quad : \underline{V} \equiv \text{continuous} \qquad (3.1.2)$$

where (a,b) represents the open interval on which "y" and "x" are defined.

"Γ" and "μ" are real-valued scalar functions whose actual forms need not yet be determined. It is clear from equations (3.1.1) and (3.1.2) that condition (2.3.1) is satisfied. A familiar regulating function of this form can be generated by allowing

$$\Gamma(.) = \mu^{-1}(.) \ ,$$[1]

then, for the discrete case equation (3.1.1) becomes

$$L(X,Y) = \mu^{-1}[\sum_i \mu(y_i - x_i)] \ .$$

Now let

$$\mu(.) = (.)^2 \implies \mu^{-1}(.) = \sqrt{\ .\ } \quad ,$$

---

[1] Recall, the superscript "$^{-1}$" indicates the inverse, not reciprocal of the function.

giving

$$L(X,Y) = \sqrt{\sum_i (y_i - x_i)^2} ,$$

which is just the common ℓ^2 norm.

To simplify further developments, the "Γ" function will be dropped by letting

$$\Gamma(x) = x .$$

This, as will be seen from the following two properties, is not a great loss, since for allowable functions, the extremum will not be altered (though the gradient of the function will). This is due to the monotonic increasing nature of allowed functions (Jeffreys and Jeffreys [12]).

The second property (2.3.2) states:

> Constant global changes to an object should not result in an extremum of a regulating function.

This can be accomplished by making $\mu(.)$ a monotonically increasing function away from a fixed point "x_0" (usually the origin or mean).[1] This means that if "x_0" is the fixed point, and "x_1"

---

[1] Throughout this thesis the fixed point will be the origin.

and "x_2" are any two points on one side of "x_0" having the property:

$$|x_1 - x_0| < |x_2 - x_0|,$$

then this will imply:

$$\mu(x_1) < \mu(x_2) .$$

To show that this in fact satisfies condition (2.3.2), the idea of constant global change must be made clear. In a vector, this means all elements are increased or decreased in absolute value. In other words, if "Y" is obtained from "X" by such a change, then

$$|y_i| < |x_i|, \quad \forall\ i$$

or

$$|y_i| > |x_i|, \quad \forall\ i.$$

Now, with no loss of generality, let "$x_0 = 0$". This means that if

$$|y_i| < |x_i|, \ \forall \ i$$

$$\Longrightarrow \mu(y_i) < \mu(x_i), \forall \ i$$

$$\Longrightarrow \sum_i \mu(y_i) < \sum_i \mu(x_i)$$

$$\Longrightarrow L(Y) < L(X) \ ;$$

or if

$$|y_i| > |x_i|, \ \forall \ i$$

$$\Longrightarrow L(Y) > L(X) \ .$$[1]

This means if the vector can be globally increased or decreased continuously, the regulating function will respond in kind without an extremum being reached.[2] Thus, condition (2.3.2) is satisfied. This can be done in another way. If a vector is globally increased or decreased and the measure is to respond in kind, the function "μ" must be monotonically increasing away from a fixed point. The proof is by contradiction.

It is given that if

$$|y_i| > |x_i|, \ \forall \ i \Longrightarrow L(Y) > L(X) \ .$$

---

[1] Note: L(Y) means L(Y,0) where "0" is the origin. This is not a special case since "Y" can be the difference of two vectors "$Y=X_1-X_2$".

[2] The limit point as "$y \to 0$" is not a problem since it represents certainty.

Let "μ" fail to be monotonically increasing away from the origin. Then on one side of the origin

$$\exists \text{ an "a" and "b" with } |a| < |b| \ni \mu(a) \geq \mu(b) .$$

Now let

$$x_i = a, \forall i \; \& \; y_i = b, \forall i$$

$$\Longrightarrow |y_i| > |x_i|, \quad \forall i ;$$

but,

$$\mu(y_i) \leq \mu(x_i), \forall i$$

$$\Longrightarrow \sum_i \mu(y_i) \leq \sum_i \mu(x_i)$$

$$\Longrightarrow L(Y) \leq L(X)$$

====> <==== (a contradiction).

Therefore, "$\mu(.)$" must be monotonically increasing away from the origin. The function "μ", for want of a better name, will be called the ruler function.

By making the ruler function monotonically increasing away from the origin, another desirable property (2.3.3) is realized. (2.3.3) states:

> A regulating function should not have properties which would make the search for an extremum ambiguous.

Even though it is obvious that this condition is satisfied if "μ" is monotonically increasing away from the origin, I feel there is a need to firm up the rather loosely stated ideas that gave rise to property

(2.3.3) in section (2.3). The reader is directed to the next page starting from the statement "Condition (2.3.4) states:" if further clarification is not necessary. Ambiguity occurs when the ruler function has local extrema or flat spots. These spots would cause ambiguous results to the search for an extremum. This is because a move in either direction can produce the same result in the regulating function. To see this mathematically, let "μ" have an extremum at "x_0". Then, there exist points to either side of "x_0" that will produce the same value in the ruler function (refer to Figure 3.1.1). Let these points be "$x_0+\delta a$" and "$x_0-\delta b$", then

$$\mu(x_0+\delta a) = \mu(x_0-\delta b) .$$

This would mean if a vector Y was such that

$$y_i = x_0, \forall i ,$$

then

$$\mu(y_i+\delta a) = \mu(y_i-\delta b)$$

$$\Longrightarrow \sum_i \mu(y_i+\delta a) = \sum_i \mu(y_i-\delta b)$$

$$\Longrightarrow L(Y+A) = L(Y-B)$$

where "A" and "B" are vectors whose elements are all δa and δb, respectively.

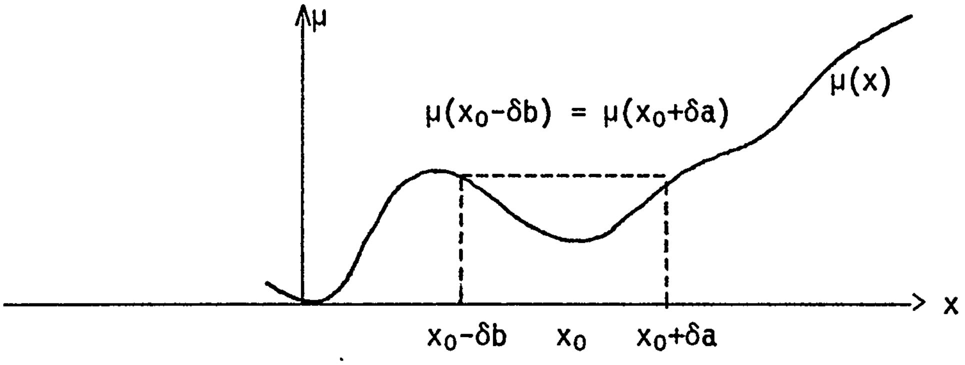


Figure 3.1.1
Local Extremum of Ruler Function

This is an undesirable state. Since allowed ruler functions are free of these extrema except at the origin, these ambiguities will not exist. The origin is not a problem as will be made clear in the following paragraph.

Condition (2.3.4) states:

> A regulating function should attain an absolute extremum when the two objects compared are identical.

This condition is satisfied by making the ruler function a minimum at the origin. To see this, consider the case where

$$X = Y ,$$

and

$$\mu(0) = 0 .$$

Then

$$\mu(x_i - y_i) = 0, \forall\ i,$$

$$\Longrightarrow L(X,Y) = 0 .$$

If "X" or "Y" were perturbed so that "$X \neq Y$", then knowing "μ" to be monotonically increasing away from the origin, it can be said

$$\mu(x_i - y_i) > 0 \text{ for some } i \text{ ,}$$

$$\text{===>} \; L(X,Y) > 0 \text{ .}$$

This means "$L(X,Y)$" is an absolute minimum when "$X=Y$". This condition arises in an inversion procedure when there is no probabilistic part.

When all the conditions developed so far have been satisfied, then the regulating function is a probabilistic error regulating function (PERF). In other words, a PERF is of the form given by equations (3.1.1) and (3.1.2), in which "Γ" is a monotonically increasing function and "μ" is a monotonically increasing function away from the origin.

(3.2) STATISTICAL EFFECTS OF PERFs

This section differs from the previous one in a fundamental way. The previous section established the form of PERFs; this section attempts to relate the extrema of PERFs to their statistical effects. This is basically condition (2.3.5), which states:

> An attempt should be made to relate the extremum of a regulating function to its statistical effects.

This is important since extrema of PERFs are used to indicate an optimum has been reached in an inversion procedure. In other words, the goal is to ascertain the statistics of the elements of a vector "X" which is gauged by a PERF "L" when "L(X)" is an extremum within an inverse procedure. In this situation, let the elements of "X" be a realization of a probability distribution "P(x)". If the dimension of "X" is large and given the form of PERFs with "$\Gamma(x) = x$" then we will have an approximate proportionality between the value of "L" and the statistical expectation of "$\mu(x)$" taken over the distribution "P(x)" (See appendix A):

$$L \mathrel{\tilde{\propto}} \langle \mu \rangle$$

(3.2.1)

$\langle . \rangle \equiv$ expected value of the argument.

The search for an extremum of "L" is now shifted to the expected value of the ruler function "μ".

The expected value of the ruler function is given by

$$\langle \mu \rangle = \sum_i P_i \, \mu(x_i) \quad : \text{discrete} \tag{3.2.1a}$$

$$= \int_{\infty}^{\infty} P(x) \, \mu(x) dx \; : \text{continuous .} \tag{3.2.1b}$$

From this point on, the continuous case will be considered exclusively. This will not be overly restrictive since, if we let

$$P(x) = \sum_i P_i \, \delta(x-x_i) \; ,$$

where

$\delta(.)$ is the Dirac delta function,

then

$$\langle\mu\rangle = \int_{\infty}^{\infty} P(x) \; \mu(x) \; dx$$

$$= \sum_i P_i \int_{\infty}^{\infty} \delta(x-x_i) \; \mu(x) \; dx$$

$$= \sum_i P_i \; \mu(x_i) \; .$$

Since it is desirable for "$\langle\mu\rangle$" to be an extremum, it would be tempting to apply variational techniques directly to equation (3.2.1b). The problem with such a direct attack is that "μ" has too much freedom, and the solution would be

$$\mu(x) = 0 \; .$$

This ensures "$\langle\mu\rangle$" would be a stable minimum for any P(x). While this result is correct, it is far from useful. The variational technique essentially allows "$\mu(x)$" to take on many shapes and from these, finds the one which makes (3.2.1b) an extremum. To exclude useless trivial cases as mentioned above, it is necessary to introduce a function. Its general form will be a function of the ruler, its derivatives and the variable "x", that is,

$$G = G(x,\mu,\mu',\mu'',\ldots) \; .$$

There are two ways to see how "G" can be used. The first is to use "G" in a constraint, Where

$$\int_{\infty}^{\infty} G\ dx = M, \tag{3.2.2}$$

the constant "M" is chosen to ensure that

$$\int_{\infty}^{\infty} P(x)\ dx = 1.$$

Care must be taken with the end points of integration and will be discussed shortly. This last condition is necessary since the variational technique will pair a "μ" with a "P" and not all choices of "M" will ensure the unit area property of "P(x)". The constraint (3.2.2) can be introduced into (3.2.1b) in the context of a variational integral "I", as given by

$$I = \int_{\infty}^{\infty}[P(x)\ \mu(x) + \alpha\ G(x,\mu(x),\mu'(x),\ldots)]\ dx. \tag{3.2.3}$$

The Lagrange multiplier "α" is chosen to satisfy (3.2.2) (Morse and Feshbach [19]). The second way to view "G" is as a point-wise description of some quality of the ruler function "μ". This description can be incorporated into equation (3.2.1b) as done in (3.2.3), with "α" being a parameter which can be varied from zero to infinity. Obviously when "$\alpha \dashrightarrow \infty$" extremizing (3.2.3) leads to an extremum of the quality embodied in "G" in a global sense; α=0, on the other hand leads back to an extremum of (3.2.1b). Thus we can control how much of each is honored by the choice of α. In this thesis "α" is chosen to honor the unit area property of "P(x)". This is very similar to the technique used in constrained linear inversion, where "G" would play the role of a measure of smoothness (Twomey [24]).

The specification of "G" will determine a pairing of "P" and "μ". The exercise is then to find a "G" which makes the pairing useful in describing the statistical effects of extrema of "L".

Integrals in the form of equation (3.2.3) are quite general, and variational techniques provide solutions to the extrema problem of quite general cases. The case considered in this thesis will be relatively simple. The cost function "G" will be a function of "x", "μ" and "μ'" only. This, as will be shown, is sufficient to incorporate properties (2.3.1) to (2.3.4). This also produces a useful pairing of "P(x)" and "$\mu(x)$" that attempts to satisfy property (2.3.5). With this in mind, equation (3.2.3) can be rewritten as

$$I = \int_{\infty}^{\infty}[P(x)\ \mu(x) + \alpha\ G(x,\mu(x),\mu'(x))]\ dx. \qquad (3.2.4)$$

Let the integrand be represented by

$$J(x,\mu,\mu') = P(x)\ \mu(x) + \alpha\ G(x,\mu(x),\mu'(x))\ . \qquad (3.2.5)$$

This allows equation (3.2.4) to be rewritten as

$$I = \int_{\infty}^{\infty} J(x,\mu,\mu')\ dx\ . \qquad (3.2.6)$$

With due consideration for the end points of integration, as discussed in appendix B, equation (3.2.6) is in a form that is well known in variational calculus. A good reference is Oskar Bolza's [2] work. In classical mechanics, "J" is known as the Lagrange density function.[1] Equation (3.2.6) must obey the Euler equation if "I" is to be an extremum. A brief derivation of this can be found in appendix C. The Euler equation is:

$$\frac{d}{dx}\left[\frac{\partial J}{\partial \mu'}\right] - \frac{\partial J}{\partial \mu} = 0 \ . \qquad (3.2.7)$$

Direct substitution of (3.2.5) into (3.2.7) results in

$$\frac{d}{dx}\left[\frac{\partial G}{\partial \mu'}\right] - \frac{\partial G}{\partial \mu} = \frac{1}{\alpha}\, P(x) \ .^{2} \qquad (3.2.8)$$

The exact form for "G" must now be specified. The specification of the exact form of "G" is approached heuristically in this thesis. Though heuristic in nature, the process was guided by a few points. The first is that PERFs are to be monotonically increasing away from a point so "G" should be chosen to predispose "μ" towards monotonic solutions. The second comes from the statement "the ℓ^1 norm will tend to maximize the number of measured zeros" which paraphrases a quote from Levy

---

[1] Good references for this are: Goldstein [9] and Morse & Feshbach [19].

[2] Note that to this point "P(x)" could be replaced by "β P(βx)", with "β" being a constant, and the development would be unaltered. For this thesis only the shape of "P(x)" is explored so "β" will be dropped. But "β" is important in describing a particular gauged quantity in an inverse procedure when "L" is an extremum.

and Fullagar [15]. This suggests the pairing of the ℓ^1 norm to "P(x) = δ(x)". This, though not entirely true, is a good approximation and should provide useful results. An example is the estimation of the expected value of a probability distribution. The ℓ^1 norm extracts the median while the mode actually maximizes the number of measured zeros. The mode and median need not be the same (Neville and Kennedy [21]). The final point is as follows: since "μ" is to be monotonically increasing away from the origin and the observation is that changes in "μ" result in changes in "P", then if "P(x) = δ(x)", one would suspect, "μ" would increase away from the origin with no change. Since we are only concerned with extrema of the PERF and that is not altered by changes in the slope of "μ" then this is just a restatement of the second point. From this backdrop and from trial and error the form of "G" is chosen to be:

$$G(x,\mu,\mu') = \sqrt{1 + \mu'^2}\,, \tag{3.2.9}$$

and this choice can be shown to be appropriate for other choices of "P(x)", in other words it is not restricted to the ℓ^1 norm. Analogous to the previous more general discussion of "G" the special form (3.2.9) can also be viewed in two ways. In the first way, since the integral of "G" can be seen as the length of "μ", we can use "G" to fix the length of "μ" then find a shape of "μ" which will extremize (3.2.1b). In the second way, the incorporation of "G" as in integral (3.2.4) can be seen as forcing "μ's" length to be shortened; this would tend to stop "μ" from "kinking up" and tend to predispose "μ" towards monotonic solutions which will extremize (3.2.1b). This choice of "G" has paired each "μ" with a "P", in particular it pairs the ℓ^1 norm with "P(x) = δ(x)", as will be shown explicitly.

Substitution of (3.2.9) into (3.2.8) gives

$$P(x) = \frac{\alpha\, \mu''\,(x)}{[\sqrt{1 + \mu'^2\,(x)}\,]^3}\ . \tag{3.2.10}$$

This equation allows easy calculation of a probability density given a ruler function. It would be nice to be able to write "μ" as a function of "P". This is done in appendix D, and the result is

$$\mu(x) = \int_{xo}^{x} \frac{F(y)}{\sqrt{1 - F^2\,(y)}}\, dy\ , \tag{3.2.11a}$$

where

$$F(x) = \frac{1}{\alpha} \int_{xo}^{x} P(y) dy\ . \tag{3.2.11b}$$

Here, x_o plays the role of the fixed zero point (which is usually the origin and sometimes the mean).

To see that equation (3.2.10) has the desired properties, it is only necessary to notice that

$$P(x) \geq 0\ ,$$

and the denominator is always positive. This assures

$$\mu''(x) \geq 0 \ . \qquad (3.2.12a)$$

The fixed point "x_0" as defined in equations (3.2.11a) and (3.2.11b) makes

$$\mu(x_0) = 0 \ . \qquad (3.2.12b)$$

Equations (3.2.11a) and (3.2.11b) make "μ" positive away from "x_0". Thus, combined with conditions (3.2.12a) and (3.2.12b), "μ" must be a positive monotonically increasing function away from its zero point "x_0". There is one last matter which should be shown explicitly. When "$P(x) = \delta(x)$" and the fixed point is just the origin, then the associated ruler function is just the ℓ^1 norm. Since "$P(x) = \delta(x)$" is zero between the origin and the positive and negative infinities, then in this region equation, (3.2.10) becomes

$$\mu''(x) = 0 \ .$$

This implies

$$\mu(x) = a \ x + b$$

where "a" and "b" are constants. If

$$\mu(0) = 0 \text{ then } b = 0.$$

Since the extrema will not be changed by the choice of slope "a" there is no problem in letting "$\mu(x_1) = x_1$" for some large "x_1",which means the following must be true:

$$\mu(x) = x : 0 < x < \infty .$$

If "$\delta(x)$" can be considered the limit of symmetric functions about "0", then by the symmetry of the variational integral in appendix B about "x_0", one can write

$$\mu(x) = |x| \text{ and } L(x) = \|x\|_1 .$$

This is just the ℓ^1 norm desired.

(3.3) ℓ^p NORMS AS PERFs

Now that a relationship has been established between PERFs and probability densities, it would be prudent to see how these manifest themselves for some well known PERFs. A group of well studied norms are the ℓ^p norms, which are defined as

$$\|x\|_p = \sum_i [|x_i| ^p]^{1/p} . \tag{3.3.1}$$

Since $[.]^{1/p}$ is an increasing function of the positive argument, the extrema should not change if this part is dropped, giving:

$$L = \sum_i |x_i|^p . \tag{3.3.2}$$

The proof of (3.3.1) and (3.3.2) having the same extrema can be found in appendix F. Since the concern of this thesis is with the extreme points, there should be no confusion if equation (3.3.2) is called the ℓ^p norm. Equation (3.3.2) is in the form considered in this work. The ruler function for a particular ℓ^p norm is

$$\mu(x) = |x|^p \; . \tag{3.3.3a}$$

Substitution of (3.3.3a) into (3.2.10) and setting "$\alpha = 1/2$" to ensure the unit area property, the following probability distributions result:

$$P(x,p) = \frac{1/2\ p(p-1)\ |x|^{p-2}}{[\sqrt{1 + p^2|x|^{2p-2}}\,]^3} \; . \tag{3.3.3b}$$

Note that

$$\lim_{p \to 1} P(x,p) = 0 \;\forall\; x \neq 0$$

and

$$\int_{\infty}^{\infty} P(x,p)dx = 1 \;\forall\; p \geq 1 \; .$$

This implies

$$\lim_{p \to 1} P(x,p) = \delta(x)$$

(Reif [22]) which is consistent with the choice of (3.2.9) as the cost function.

It would be interesting to see how (3.3.3a) and (3.3.3b) appear graphically. Figure 3.3.1 shows that when $p = 2$, which leads to the PERF used in least-squares techniques, "P(x)" takes on a Gaussian-like shape. This may explain partially why least-squares techniques are used so widely. Another reason is the ease with which least-squares algorithms can be developed. A very good discussion of this can be found in Twomey's book [24]. The corresponding curves for the other ℓ^p norms are shown in figure 3.3.2 (because of symmetry only one side is plotted). Note that for "$p < 1$", the probability distributions are negative. This is due to the ruler function breaking the concave upwards property demanded by the constraint.

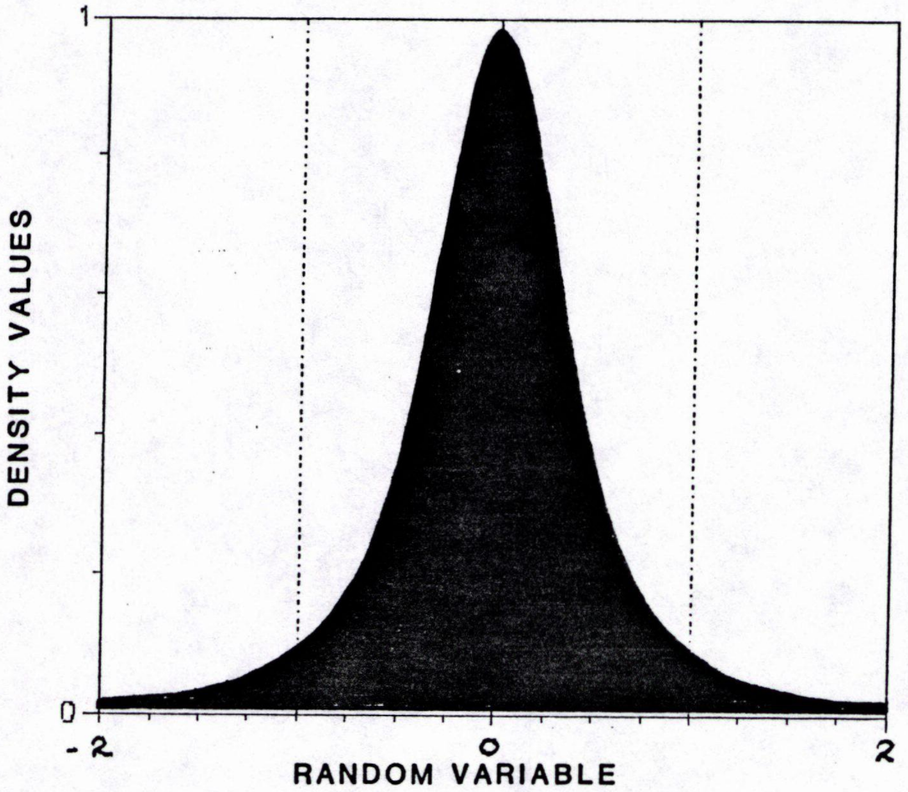


Figure 3.3.1

ℓ^2 Prob. Densities

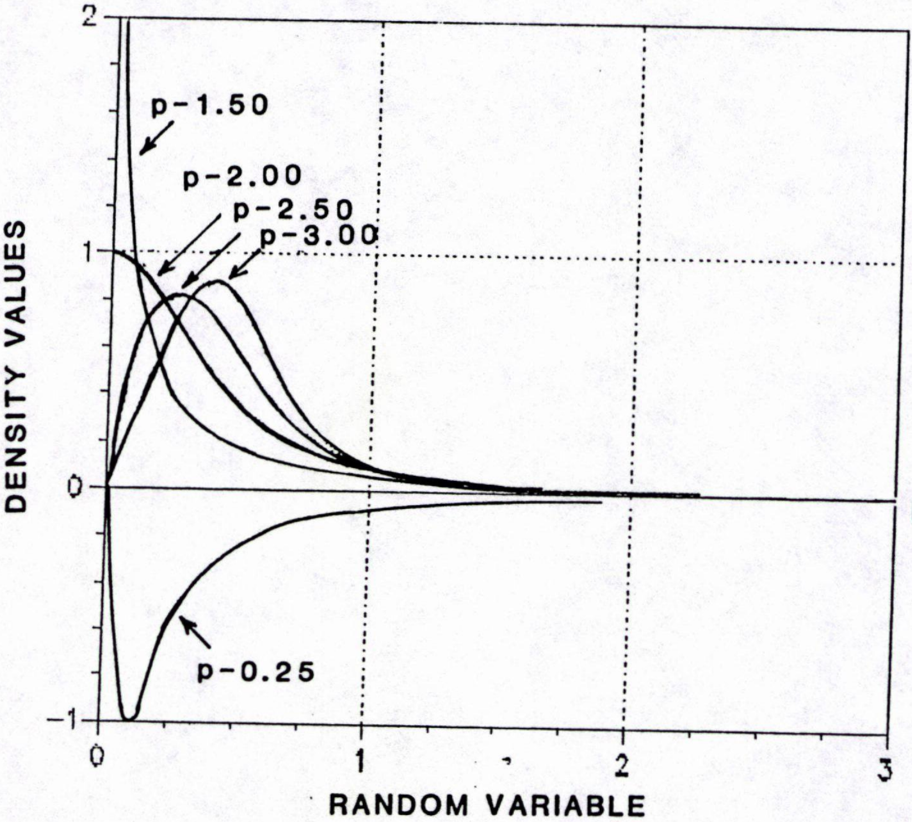


Figure 3.3.2

ℓ^p Prob. Densities

Thus far, the following has been accomplished:

- Desirable properties of regulating functions have been outlined.
- Mathematical manifestation of the properties has been established.
- PERFs have been defined.
- A relationship between PERFs and probability distributions that extremize the PERFs has been found (by the imposition of the least-length constraint).
- The relationship has been applied to the well known ℓ^p norms.

This application to the ℓ^p norms has given some testable relationships. The testing of these relationships is the core of material in the next chapters.

CHAPTER 4
(TRIAL I): CONSTANT VALUE EXTRACTION

The first numerical trial of the previously developed methods is also the simplest. It is also the simplest form of the mixed process in appendix E section E1. The process "S" consists of a constant "c" imbedded in additive noise "n_i",

$$S(c) = c + n_i \ .^*$$

An optimal estimate of the constant "c" is desired. Many physical situations can be described by this formulation. One geophysical example is the stacking process. Two common stacking techniques are the mean and median stacks. As is well known, these correspond to using the ℓ^2 and ℓ^1 criteria, respectively (Whittaker and Robinson [26]).

To see how to proceed toward an estimate of "c", consider "N" realizations of the process "S". This gives the set of "N" outcomes below,

$$\underline{X} \equiv \{x_i : x_i = c + n_i; \ i=1,2,\ldots,N\}$$

The representation of this set with a single value "d" is the goal. In accordance with the discussion in appendix E section E1, the constant

---

* The subscript "i" is used to index a specific outcome of "S". It is necessary to uniquely identify the outcomes since the noise is different for repeats of "S".

chosen as the deterministic part of the process and the additive noise the probabilistic part. This choice is the natural one, and is obvious, but the choice may not be obvious in other situations. Conceptually, the noise can be viewed as a realization of a probability distribution "P(x)". When a constant "c" is added, this becomes a realization of a new distribution

$$\tilde{P}(x) = P(x-c) \; .$$

The idea is then to shift "$\tilde{P}(x)$" by constant amounts "d" until a match is achieved, as shown below:

$$\tilde{P}(x+d) = P(x) \; .$$

The difficulty with this is usually two-fold. First, P(x) is not always known. Secondly, since "N" is finite, $\tilde{P}(x)$ cannot be completely defined.

At this point, a regulating function "L" is invoked. A value "d" which minimizes the regulating function is called an optimal guess of "c". But, as previously mentioned, there are an infinite number of regulating functions which can be employed. To aid in the choice, the relationships (3.2.10) and (3.2.11a) may be employed. To see how these relations can help, the ℓ^p norms will be used as a test case. The relationships developed for the ℓ^p norms were (3.3.2) and (3.3.3b) and are reproduced below as:

$$L = \|X\|_p^p = \sum_i |x_i|^p \qquad (4.1a)$$

and

$$P(x,p) = \frac{1/2\ p(p-1)\ |x|^{p-2}}{[\sqrt{1 + p^2\ |x|^{2p-2}}\ ]^3} \ . \qquad (4.1b)$$

One obvious characteristic of equation (4.1b), as seen in figure 3.3.2, is that the maximum of the probability distribution drifts away from the origin as "p" is increased. This would suggest, if an "ℓ^p" norm were to be used in selecting an optimum representation of the set "$\underline{X}$", this value "d" should have the property of making the point which has the highest density of occurrences of elements of "$\underline{X}$" to drift away from the origin, unless the set was totally symmetrical in which case all the "ℓ^p" norms would pick the same value. In other words, if "x" was such a point where elements of "$\underline{X}$" are clustered, then the point of maximum density resulting from the subtraction of "d" from the elements of "$\underline{X}$", namely:

$$\text{resulting point of max. density} = x - d = \delta,$$

would also drift from the origin with increasing values of "p". To show this is in fact the case, the following data set was contrived:

$$\underline{X} = \{x_i\} = \{5,2,5,4,5\} \ .$$

In this set

$$x = 5$$

and

$$\delta = 5 - d .$$

The value "d" is said to be optimal in the "ℓ^p" sense if it minimizes the following expression:

$$\left\|\underline{X}-d\right\|_p^p = \sum_i \left|x_i-d\right|^p .$$

The "d" values which satisfy this condition were found by a simple bisection method. The "d" values and the resulting maximum density points are tabulated below in Table 4.1:

ℓ^p Norm Power (p)	Extracted Optimum (d_p)	Resulting Point of Max. Density (δ)	ℓ^p Norm Power (p)	Extracted Optimum (d_p)	Resulting Point of Max. Density (δ)
1.0	5.000	0.000	5.5	3.682	1.318
1.5	4.446	0.554	6.0	3.664	1.336
2.0	4.200	0.800	6.5	3.649	1.351
2.5	4.025	0.975	7.0	3.636	1.364
3.0	3.903	1.097	7.5	3.626	1.374
3.5	3.825	1.175	8.0	3.617	1.383
4.0	3.772	1.228	8.5	3.610	1.390
4.5	3.734	1.266	9.0	3.603	1.397
5.0	3.704	1.296	9.5	3.597	1.403

Table 4.1
ℓ^p Constant Value Extraction Results

The results are also represented graphically in figure 4.1. As a comparison, the derivative of equation (4.1b) was taken and set to zero; solving for "x" gives the maxima of "P(x,p)" (appendix H), which is

$$x = [(2-p)/(p^2[4-3p])]^{1/(2p-2)} .$$

This result is represented in figure 4.2. As mentioned before the distributions are defined only to a scaling factor "β". This scaling factor is necessary to describe a particular outcome, but since it is only the general shape of the curve that is being compared "β"is not incorporated. There is another point to consider before a comparison is made between figures 4.1 and 4.2; it should be noted that the set "$\underline{X}$" sets physical limits on the outcome. One such limit is the value "d" is constrained mathematically to lie between 5 and 2. A proof of this can be found in appendix G. Given these physical constraints, figures 4.1 and 4.2 cannot be expected to be identical, but their form should be similar. Comparison of the figures shows good agreement indicating the usefulness in pairing equations 4.1a and 4.1b. This was tried on several other sets of numbers giving similar results.

Resulting Point of Max. Density (δ)
as a Function of Norm Power (p)

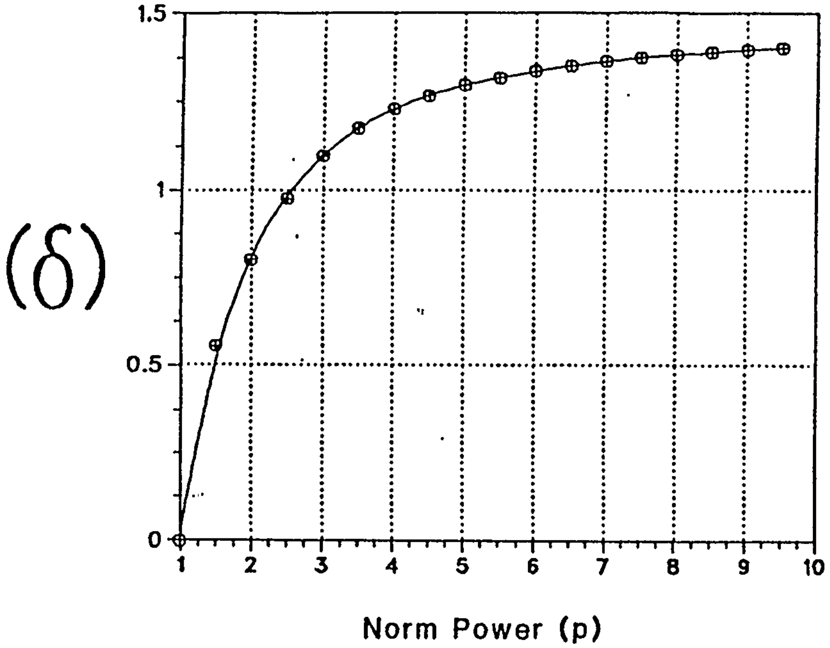


Figure 4.1

Maximum Point of ℓ^p Probabilities
as a Function of Norm Power (p)

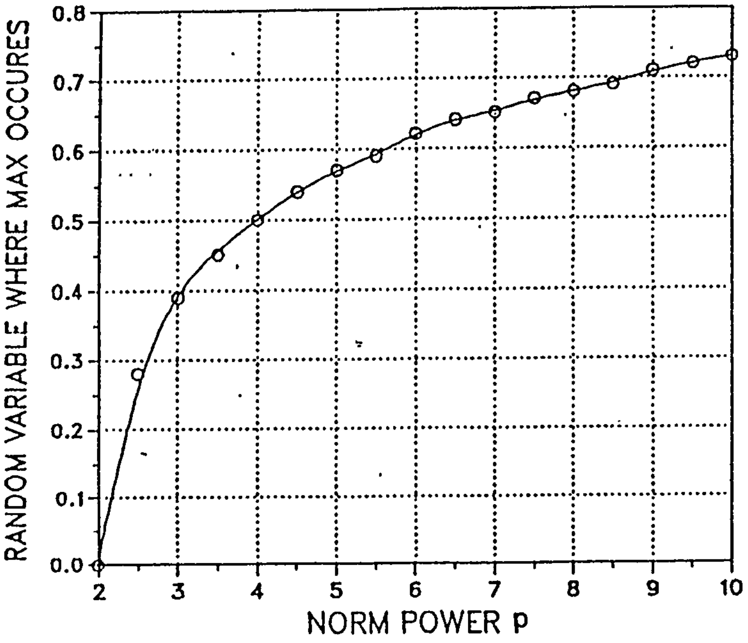


Figure 4.2

This trial tests a specific aspect of the probability distributions derived (drift of δ). The next trial will be designed to test the overall shape of the predicted distributions.

CHAPTER 5
(TRIAL II): MINIMUM ENTROPY DECONVOLUTION

This numerical trial deals with a well known, though not overworked, deconvolution procedure. The procedure is known as "minimum entropy deconvolution" or simply M.E.D.[1] This technique was established by Ralph Wiggins [27]. He proposed the use of the varimax norm as a indicator of spikiness in his M.E.D. algorithm. Later, T. J. Deeming [5] generalized the technique to other appropriate indicators of spikiness. A good general discussion of some deconvolution techniques used in geophysics and M.E.D. in particular can be found in William Gray's thesis [10]. He introduced a class of regulating functions (variable norms) which allows one to distinguish if a sample is closer to one member of a family of distributions (generalized Gaussians) than to another member of the same family. These regulating functions were used in his M.E.D. implementation to provide some control. Another good discussion of M.E.D., especially its statistical properties, can be found in Donoho's work [6]. For a discussion of geometrical properties of M.E.D., Cabrelli's paper [3] is a good source of information.

I will give a brief discussion of M.E.D., but for greater detail, the reader is urged to look at the above references. This technique presupposes that the data consists of a wavelet "W" convolved with a

---

[1] David Donoho gave partial justification for this terminology [6].

sparse spike sequence "R" (the reflectivity hopefully). This allows the forward process "S" to be represented by figure 5.A1.

$$R \text{ -----> } \boxed{S(R) = W * R} \text{ -----> } X$$

$$x_j = S(R) = \sum_i w_i r_{j-i} \ .$$

Figure 5.A1
Forward Convolution Model

The approximate inverse "$\tilde{S}^{-1}$" consists of finding a filter "F" which best removes the effects of the wavelet "W", thus recovering the sparse spike sequence "R". Since it is assumed that "R" is spiky, an indicator of spikiness can be used to gauge how well the choice of "F" is doing towards giving such an output. In any case, the approximate inverse "$\tilde{S}^{-1}$" is dependent on unknown parameters, the filter coefficients "$\{f_i\}$". It can be represented as follows in figure 5.A2.

$$X \text{ -----> } \boxed{\tilde{S}^{-1}(f_i;X)} \text{ -----> } Y \cong R$$

$$y_i \doteq \tilde{S}^{-1}(f_i;X) = \sum_i f_i x_{j-i} \cong r_j .$$

Figure 5.A2
Inverse Convolution Model

This is very reminiscent of mixed processes in appendices E2 and E3. As mentioned, a regulating function is needed to indicate whether the output "Y" is indeed spiky in the right way. Wiggins [27], as mentioned in the first paragraph, chose the varimax norm below:

$$V(y) = \frac{\sum_i y_i^4}{[\sum_j y_j^2]^2} = \sum_i \left[\frac{y_i^2}{\sum_j y_j^2}\right]^2 ,$$

where "i" and "j" both are indexing sample number in a time series. But, as shown by Deeming [5], any positive increasing function of the scalar and sign-invariant mapping of the data,

$$z_i = \frac{y_i^2}{\sum_j y_j^2} , \tag{5.1}$$

will tend to minimize the number of spikes present on the M.E.D. output.[1] With this in mind, the PERF can be written as:

$$L(Z) = \sum_i \mu(z_i) . \tag{5.2a}$$

This is of the form discussed in this paper. If "μ" takes on the form

$$\mu(x) = |x|^p , \tag{5.2b}$$

then equation (3.3.3b) can be used again to describe the distribution of the final "z_i's". The equation is:

$$P(x,p) = \frac{1/2\ p(p-1)\ |x|^{p-2}}{[\sqrt{1 + p^2\ |x|^{2p-2}}\ ]^3} . \tag{5.3}$$

---

[1] Other scalar and sign invariant mappings exist but will not be considered here.

As shown by Cabrelli [3], the scalar and sign-invariant mapping (5.1) constrains "Z" to be a vector of the set from the origin to the intersection of the hyperplane perpendicular to the barycenter "$B=(m^{-1},...m^{-1})$", where "m" is the dimension of "Z", and the region defined by the the positive axis along the natural basis for R^m (m-dimensional vector space). This set of vectors must also contain the natural basis. He also showed that the varimax norm ($\mu(x) = |x|^2$) is a minimum at the barycenter and a maximum at the natural basis for "R^m". Thus, maximizing the varimax gauges the withdrawal from the barycenter or the proximity to the natural basis. This produces spiky results. But, Deeming [5] showed any monotonically increasing "$\mu(x)$" will also gauge this withdrawal form the barycenter or the proximity to the natural basis. So all "μ's" will produce spiky "Z's" but they need not be equal. This implies the spikes in the "Z's" should take on realizations of the probability distributions (5.3) if ruler functions of the form (5.2b) are used and the PERFs given by (5.2a) are extremized. In other words, all rulers will give a spiky output, but the distribution of the spikes will differ from ruler to ruler. If the spike sequence is Gaussian, undifferentiability problems arise from filtering, since linear combinations of Gaussian variables are again Gaussian. A good discussion of this can be found in Donoho's paper [6].

An overview of how the filter coefficients are found is now in order. The details can be found in appendix I, where the derivation is for the multichannel case. Here, only the highlights of the single channel case will be developed. To recap, a filter "F" is to be found which, when convolved with the data "X", is to produce an output "Y" that is spiky like the input "R". A spiky output should maximize the PERF (5.2a). In order for (5.2a) to be such an extremum, the following condition must be satisfied:

$$\frac{\partial L}{\partial f_i} = 0, \forall\ i \ . \tag{5.4}$$

Solving equation (5.4) results in the following:

$$\frac{\partial L}{\partial f_j} = \sum_i u_i x_{j-i} - \sum_i f_i \phi_{j-i} = 0 \ , \tag{5.5}$$

where

$$u_i = \frac{\mu_i' y_i}{\langle \mu' z \rangle} \text{ and } \phi_{j-i} = \sum_k x_{k-i}\ x_{k-j} \ .$$

Equation (5.5) gives rise to what has been termed the iteration equation in appendix I and listed below as:

$$\sum_i f_i \phi_{j-i} = \sum_i u_i x_{j-i} \ . \tag{5.6}$$

Since "u_i" is dependent on the filter "f_i", equation (5.6) is highly non-linear. It cannot be solved directly for "F"; rather, an iterative procedure is employed to converge on a solution. To begin, cast (5.6) in a more familiar matrix form,

$$F \cdot \Phi = V \tag{5.7}$$

where

$$f_i = f_i,\ \Phi_{i,j} = \phi_{j-i} \text{ and } v_i = \sum_k u_k\ x_{i-k} \ .$$

The rows and columns of these matrices are defined in the obvious way. Equation (5.7) can be solved formally for "F" as:

$$F = \Phi^{-1} V \; . \tag{5.8}$$

Since "Φ" is an autocorrelation matrix, (5.8) is similar to the familiar normal equations; therefore is amenable to solution by Levinson recursion. The total method consists of making an initial guess of the filter, calculating "u_i" from equation (5.5), getting an update filter from equation (5.8) and iterating this until convergence occurs. Convergence can be determined either by monitoring the regulating function "L" using the Cauchy criterion, or by using the method as suggested in Deeming's paper [5] and shown in appendix I. The convergence characteristics of this procedure are discussed in Deeming's paper [5].

Now, the test can begin. The first trial data set consists of a wavelet (figure 5.1) convolved with a spike sequence (figure 5.2) resulting in an output (figure 5.3). M.E.D. was applied to this trace with differing ruler functions, where

$$\mu(x) = |x|^p \; .$$

The results are shown on figures 5.4-5.8. All deconvolved traces are normalized so that the largest spike has a value of ten. As can be seen, all measures (differing in values for "p") extracted the spike sequence rather well. It is also obvious that as "p" was increased, the solution became more and more dominated by the largest spike. It is also interesting to speculate what happens to the spike estimates prior to the actual convergence of the technique, but that is not the thrust of this thesis.

To see how the distribution of the measured quantity "Z" is actually altered, a second spike sequence was generated, which after being mapped by equation (5.1), has a flat distribution. This data set was then input into the M.E.D. algorithm. The output measured quantity was then plotted as a function of the input measured quantity.

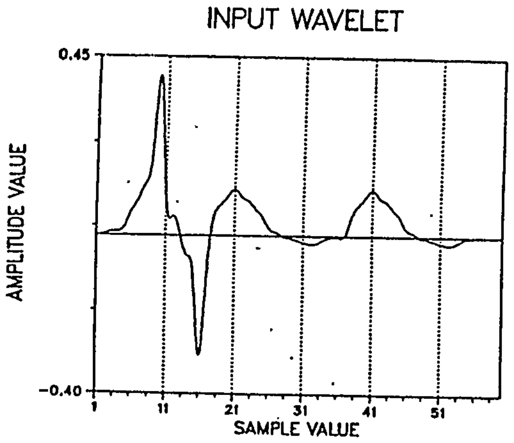


Figure 5.1

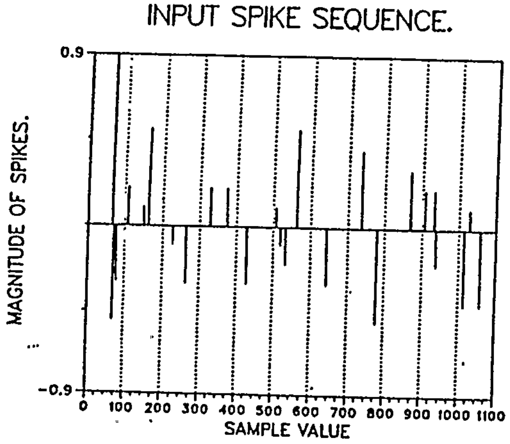


Figure 5.2

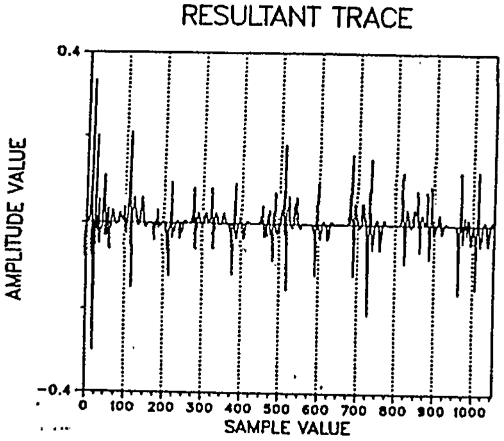
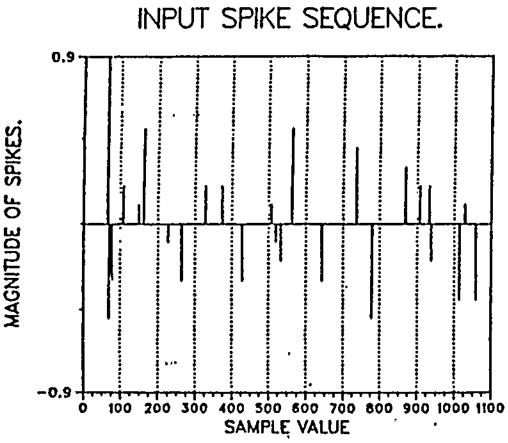
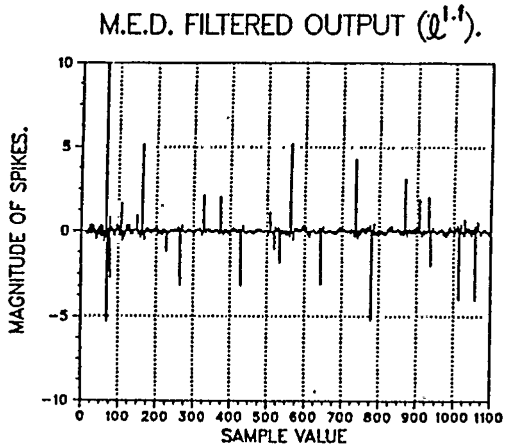


Figure 5.3



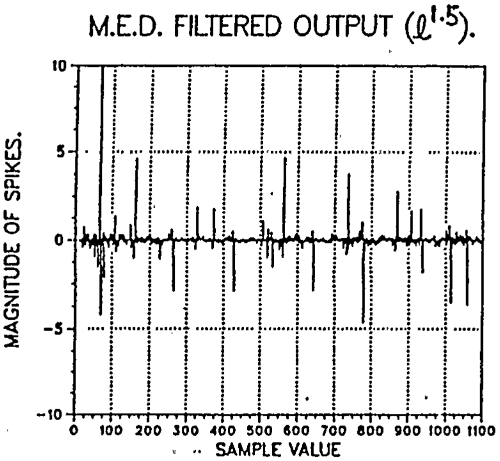
INPUT SPIKE SEQUENCE.

Figure 5.2
Repeated



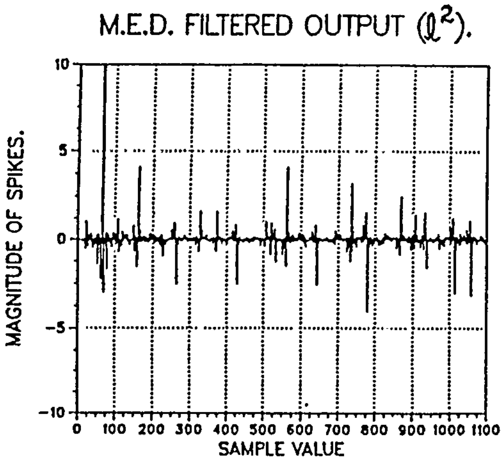
M.E.D. FILTERED OUTPUT ($\ell^{1.1}$).

Figure 5.4



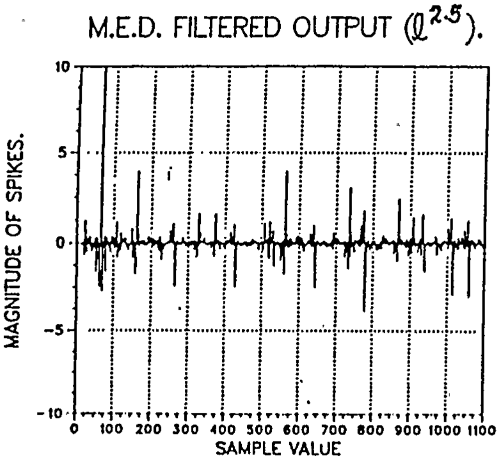
M.E.D. FILTERED OUTPUT ($\ell^{1.5}$).

Figure 5.5



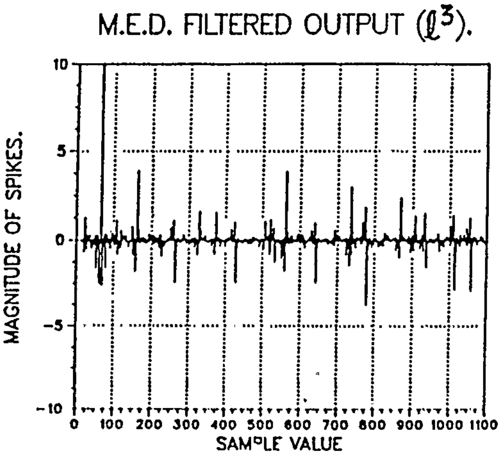
M.E.D. FILTERED OUTPUT (ℓ^{2}).

Figure 5.6



M.E.D. FILTERED OUTPUT ($\ell^{2.5}$).

Figure 5.7



M.E.D. FILTERED OUTPUT (ℓ^{3}).

Figure 5.8

The results are shown in figures 5.9-5.14. For powers greater than two, the scatter is great. This results from the dominance of the largest estimated spike as can be seen circled on figures 5.11 and 5.13. The spike is so large that small fluctuations of the others do not play a very important role. This large spike was excluded in figures 5.12 and 5.14 , and a general one-one mapping was inferred through the scatter, as was done with the other plots.

To see the significance of the plots mentioned above, the following development is necessary. If a random variable "x" is monotonically mapped into "y", and the probability distribution of "x" is "f(x)", then from statistical theory (Freund [7]), the distribution of "y" will be

$$g(y) = \left| \frac{dx}{dy} \right| f(x) \ . \tag{5.9}$$

"g(y)" is the distribution of the random variable "y". If "x" has a flat distribution, then "g(y)" will be proportional to the inverse of the mapping slope,

$$g(y) \propto \left| \frac{dy}{dx} \right|^{-1} . \tag{5.10}$$

As stated, the mapping curve was interpreted on figures 5.9-5.14. A pictorial representation of "g(z)", the final distribution of mapped quantity, based on a few hand calculated slopes obtained from the mapping curve is directly below each of these curves. This, in turn, can be compared to the probability distributions on figure 3.3.2. Again, there appears to be good qualitative agreement.

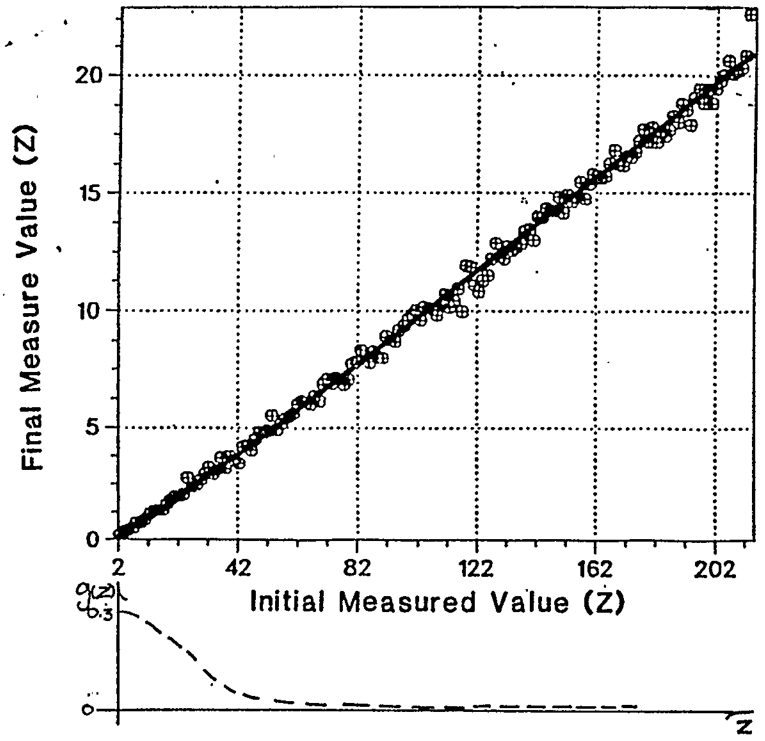


Figure 5.9

MAPPING WHEN USING $\ell^{1.5}$

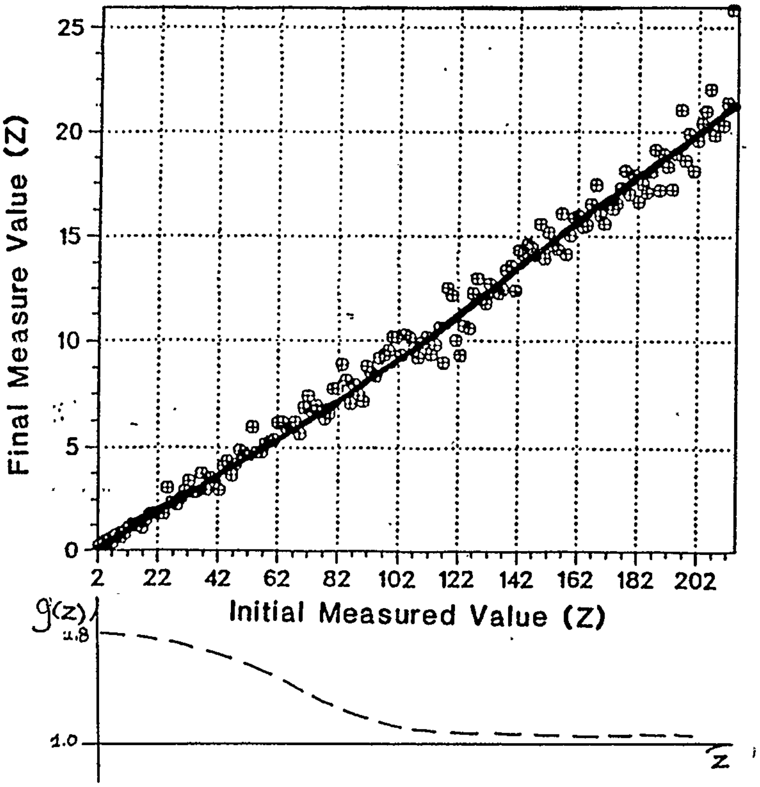


Figure 5.10

MAPPING WHEN USING $\ell^{2.0}$

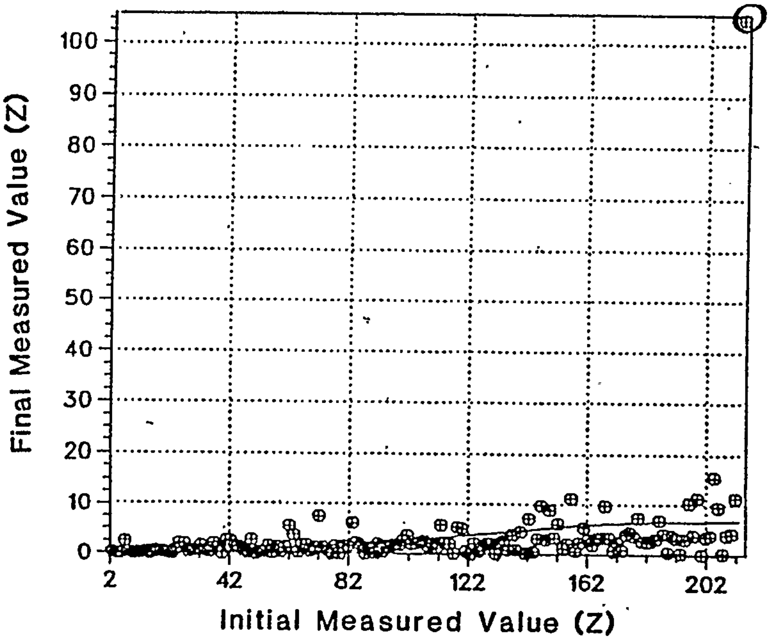


Figure 5.11
MAPPING WHEN
USING $\ell^{2.5}$

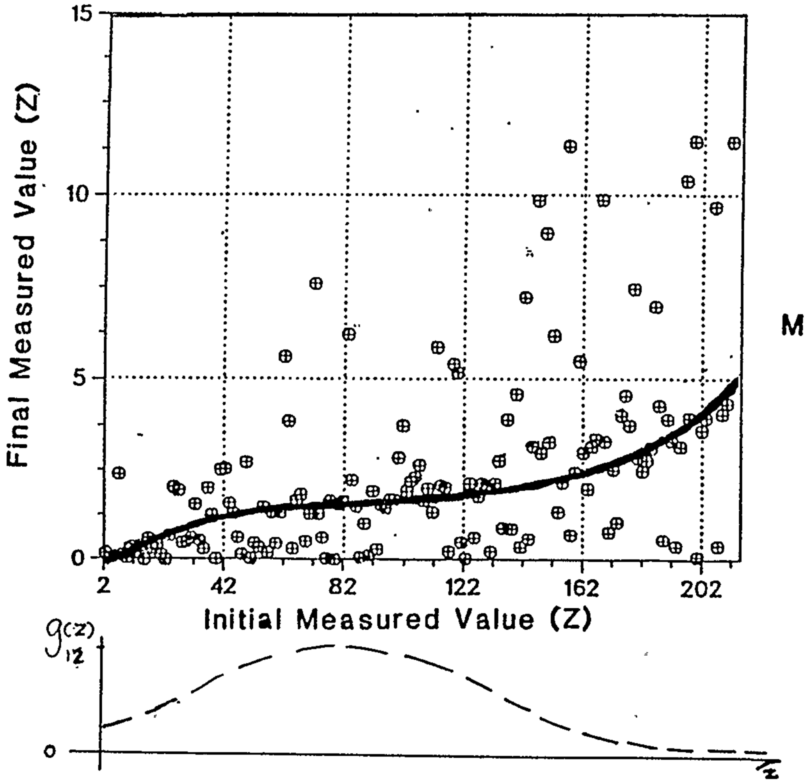


Figure 5.12
MAPPING WHEN
USING $\ell^{2.5}$
(Expanded)

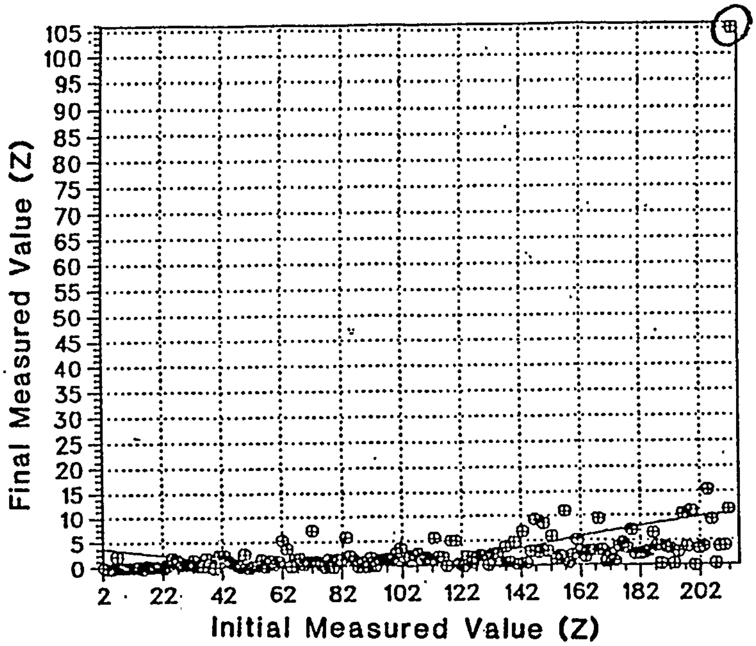


Figure 5.13
MAPPING WHEN
USING $\ell^{3.0}$

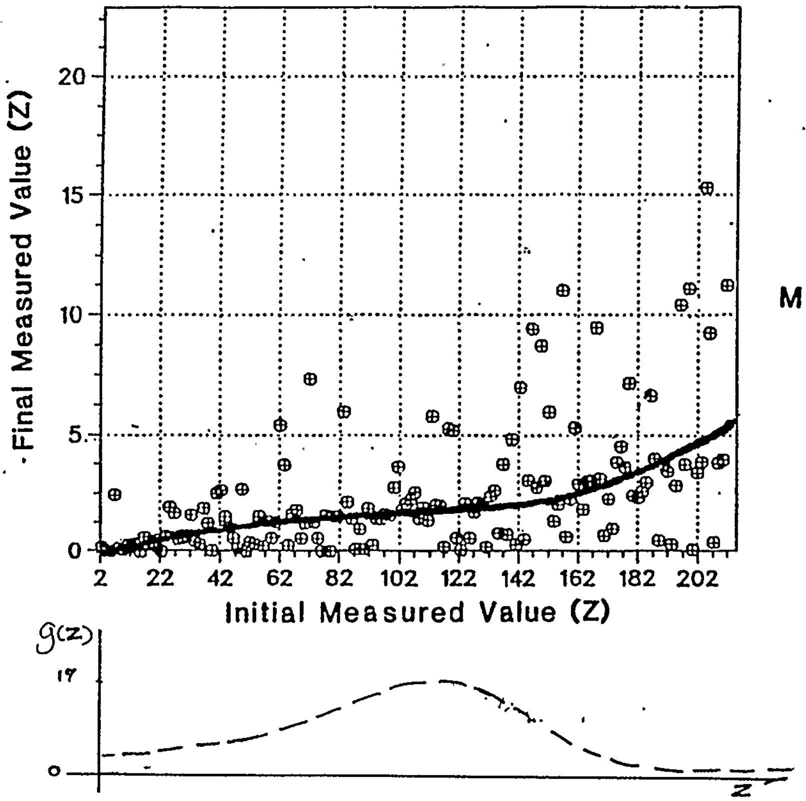


Figure 5.14
MAPPING WHEN
USING $\ell^{3.0}$
(Expanded)

This test has shown qualitative agreement to results expected from equation 3.2.9. Thus, if one desired results to be emphasized in a particular fashion in this inversion procedure, equation 3.2.9 can be used to help him choose a particular PERF. This shows how the choice of a PERF might influence a particular inversion procedure. The next trial will be used to show how a PERF can be custom designed for a particular inversion scheme. This could have been applied in M.E.D. as well.

CHAPTER 6

(TRIAL III): AUTOMATED CONSTANT PHASE SHIFT CORRECTION

In this section, two objectives will be accomplished: first, a new process will be introduced to test our ideas; second, a technique will be proposed which allows one to custom-design a PERF given a few estimates of moments for the probabilistic part of our process.

The forward process "S" will play a very small role in this trial. It can be considered a black box which produces an output "X" which is not what we want. The undesirable aspect can be adequately modelled by a constant phase shift (Levy and Oldenberg [16]). The approximate inverse "$\tilde{S}^{-1}$" is, of course, to apply the negative phase shift, bringing "X" to more desirable form "$\tilde{Y}$". Pictorially, this can be represented by figure 6.1.

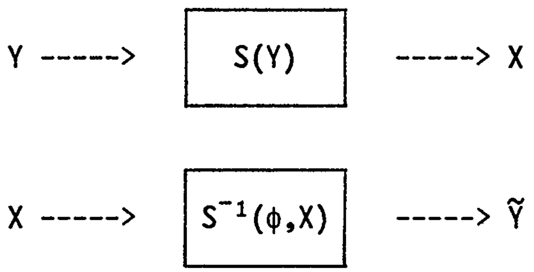


Figure 6.1
Constant Phase Shift Correction

The symbol "ϕ" represents the unknown phase shift that must be applied to "X" to yield an optimum "$\tilde{Y}$". This process is, of course, a special example of the mixed process in appendix E section E3. To make this discussion more concrete, an actual geophysical application will be considered.

Constant phase shift as a model for discrepancy between seismic traces and synthetic seismograms from well logs has long been used in exploration seismology. The use of constant phase shift, in a laterally varying sense, as a final adjustment towards a zero-phase section is, strangely enough, relatively new. The use of a regulating function to automate the process was brought to my attention by Doug Oldenberg in a paper he presented at the 1985 C.S.E.G. convention and recently published [16]. The technique was used, in the absence of well information, as a final dephasing attempt before full-band spike-sequence extraction. The regulating function used in his discussion was the varimax norm. The normalized varimax norm is defined to be:

$$V(N) = \frac{\frac{1}{m}\sum_{i=1}^{m} n_i^4}{\left[\frac{1}{m}\sum_{i=1}^{m} n_i^2\right]^2} \tag{6.1}$$

A good discussion of the properties of this measure can be found in Wiggins' [27] and Deeming's [5] papers as mentioned in chapter 5. The property of concern is that this norm attains a maximum when "N" consists of only one value or spike. Since a zero-phase wavelet is very peaked in comparison to other phase shifts, one would expect "V" to be a high number for a zero-phase wavelet. Figures 6.2 and 6.3 try to show this schematically. Figure 6.2 represents a zero-phase wavelet and has one dominant peak; whereas, figure 6.3 represents a ninety-degree phase-shifted wavelet with two large peaks.

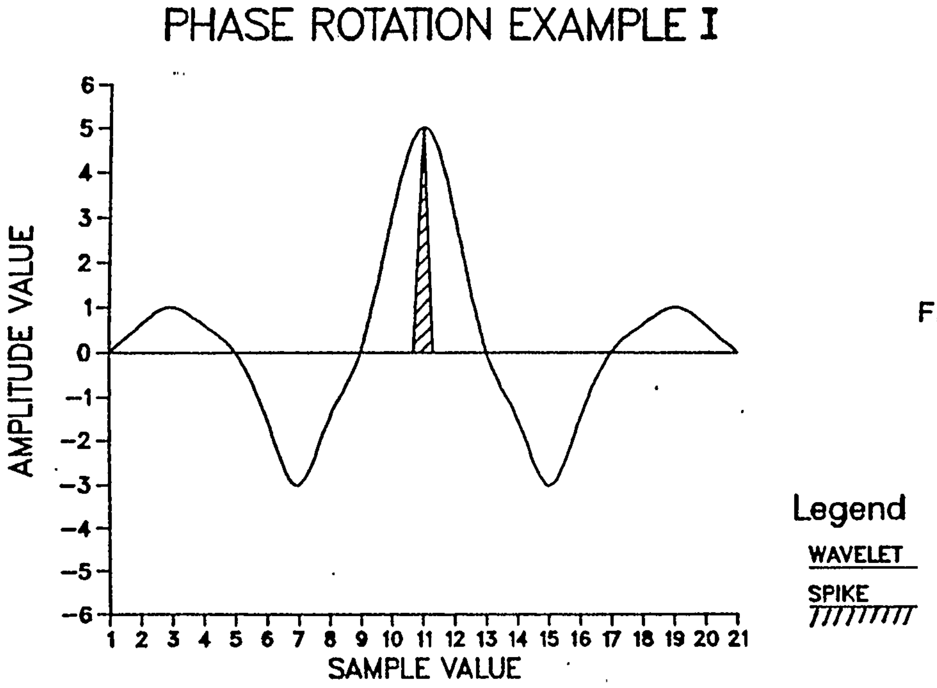


Figure 6.2

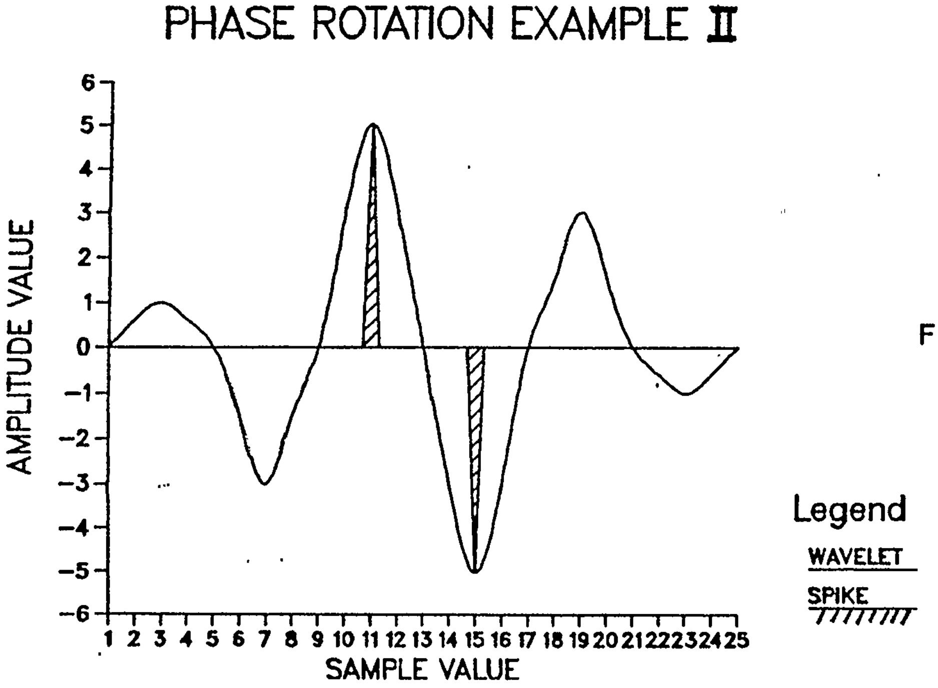


Figure 6.3

The varimax of the second wavelet would then tend towards a lower value. This technique would then be plausible for isolated wavelets, and from discussions with one of the seismic processing contractors, appears to do well in that case. The difficulty arises from non-isolated wavelets, as happens when a wavelet is convolved with a spike sequence. Figures 6.4 and 6.5 try to show this pictorially. Here, we have a zero-phase and ninety-degree phase-shifted wavelets convolved with two spikes of equal magnitude. It is no longer evident what the varimax will do here, since it is always possible to construct cases where the varimax will be maximum at phase shifts other than zero. To alleviate this problem, I propose a two-fold attack: first, a constant phase shift invariant representation will be introduced to stabilize the procedure; secondly, a measure will be introduced that automates the procedure based on prior knowledge. These concepts will be clarified in the following development. Two obvious constant phase invariant representations are the energy envelope and instantaneous frequency. These are calculated from the analytic trace. The analytic trace "$\beta(t)$" is related to the real trace "$\alpha(t)$" in the following manner:

$$\beta(t) = \alpha(t) + i\,H[\alpha(t)] \tag{6.2}$$

$i = \sqrt{-1}$

$H \equiv$ Hilbert transform
(Morse and Feshbach [19]).

"$\beta(t)$" is called the analytic extension of "$\alpha(t)$". It is shown in appendix J that if a constant phase shift "σ" is applied to the trace "$\alpha(t)$" giving "$\tilde{\alpha}(t)$",. the analytic extension of "$\tilde{\alpha}(t)$" will be related to "$\beta(t)$" in the following manner:

$$\tilde{\beta}(t) = \tilde{\alpha}(t) + i\ H[\tilde{\alpha}(t)] = e^{-i\sigma}\ \beta(t)\ . \tag{6.3}$$

Therefore, the energy envelope as defined below,

$$\tilde{A}^2(t) = \tilde{\beta}^2(t) = [e^{-i\sigma}\ \beta(t)]^2 = \beta^2(t) = A^2(t), \tag{6.4}$$

is obviously constant phase shift invariant. The instantaneous phase of "$\alpha(t)$" is defined to be

$$\phi(t) = \arctan[\mathrm{Im}(\beta)/\mathrm{Re}(\beta)] = \arctan[H(\alpha)/\alpha] \tag{6.5}$$

and is related to the instantaneous phase of "$\tilde{\alpha}(t)$" in the following way:

$$\tilde{\phi}(t) = \phi(t) - \sigma\ . \tag{6.6}$$

The instantaneous frequency is defined to be the time derivative of the instantaneous phase. Using equation (6.6) to calculate instantaneous frequencies we get

$$\tilde{w}(t) = \frac{d\tilde{\phi}}{dt} = \frac{d\phi}{dt} = w(t)\ . \tag{6.7}$$

This shows instantaneous frequency to be constant phase shift invariant as well.

Now that two constant phase shift invariant representations have been established, they can be used to help select the proper phase shift in correcting the data. For an isolated wavelet, the process consists of finding a phase shift which aligns the central peak of the wavelet with the maximum of the envelope. Note that by comparison of envelope to wavelet it is now possible to distinguish the zero-phase and inverted wavelets whereas strict use of the varimax method would not allow this. A schematic representation of this can be seen on figures 6.6 thru 6.8. A PERF can, of course, be used to automate this. If we do not have an isolated wavelet, the procedure will not be as simple, as can be seen on figures 6.4 and 6.5. Here, it would be nice to incorporate more information if available. If well information is present in an area, its information should be used. We do not want to overconstrain the problem, recognizing the fact that geology changes. One way to do this is to band-pass the spike sequence from the well to be as close as possible to the spectrum of the data. Then, take the difference of the resultant synthetic and the envelope of the synthetic. The resultant values are normalized so that the maximum is less than or equal to one. The sample moments are calculated for these normalized values and incorporated into a PERF as shown in appendix L. This PERF should tend towards a maximum if the data is indeed close in some sense to the well data. The procedure is then to rotate the data, calculate the difference between the value of the rotated trace to the energy envelope, normalize the values, enter these values into the PERF, and do this until a maximum is found, at which point the result will be termed optimum.

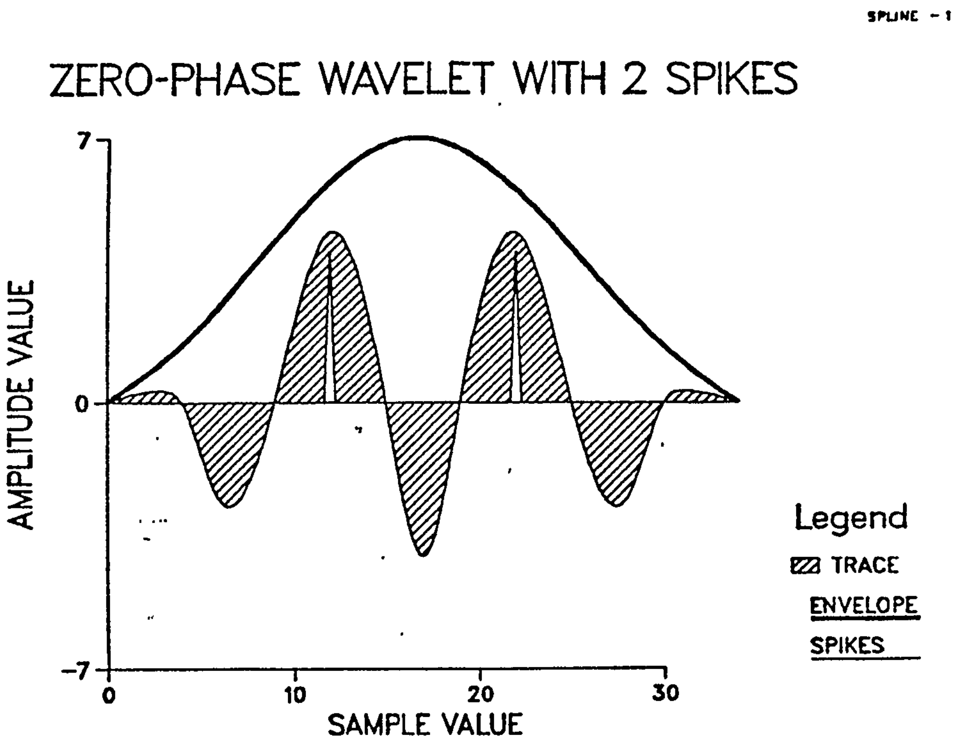


Figure 6.4

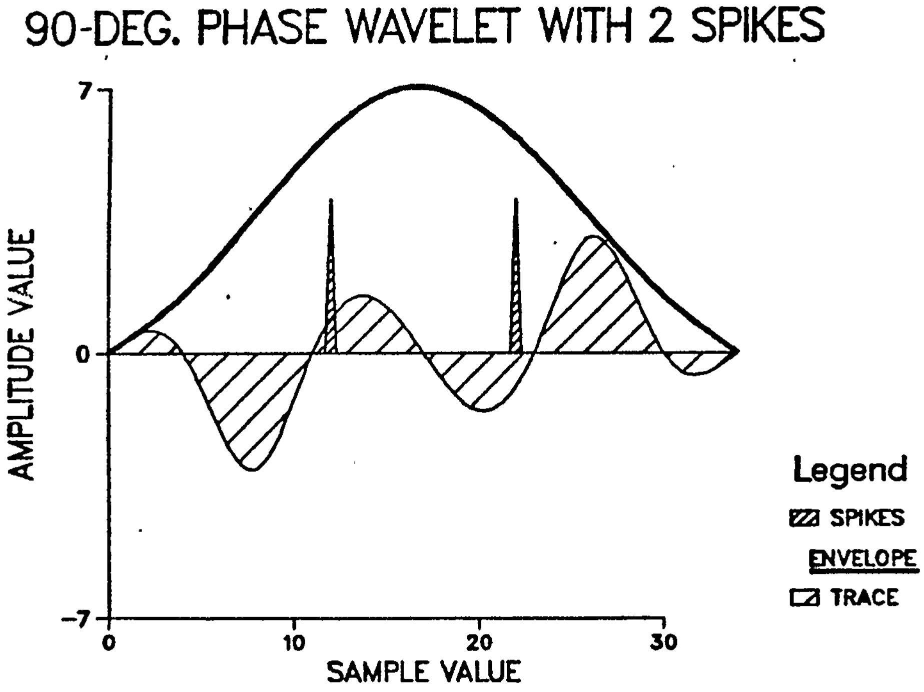


Figure 6.5

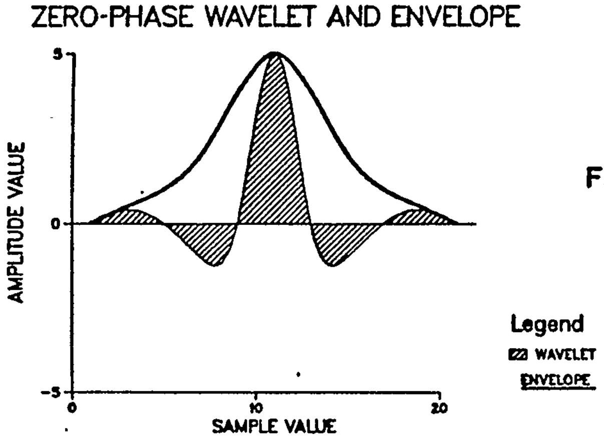


Figure 6.6

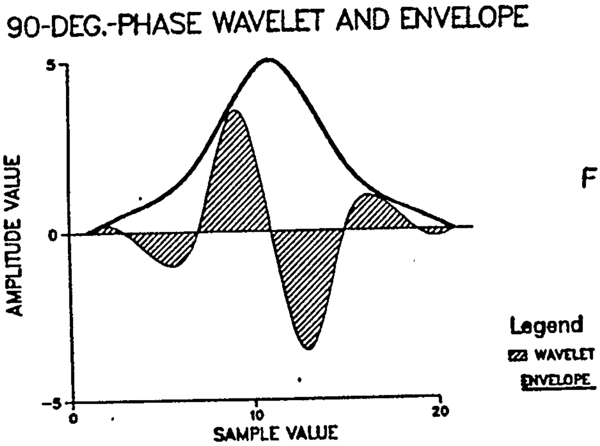


Figure 6.7

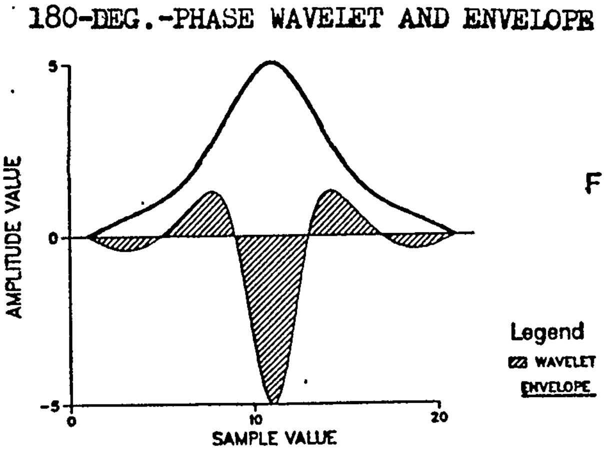


Figure 6.8

SUMMARY AND CONCLUSIONS

In this thesis a class of regulating functions (PERFs) is defined, of which the ℓ^p norms are a subset. The defining properties of PERFs have been drawn from a general inversion framework; the use of PERFs has also been shown within that framework. The form of PERFs allows them to be paired to probability distributions. This attempts to describe the statistical effect that the extrema of PERFs impose on a gauged quantity in inversion. The pairing was achieved by variational techniques and the imposition of a length constraint on the ruler function. The choice of the constraint was arrived at heuristically. More work needs to be done to put the choice on firmer mathematical grounds. The pairing was then conducted for the ℓ^p norms resulting in a family of probability distributions. Two properties of these distributions were tested on a couple of geophysical inversion problems. The first inversion problem was constant value extraction. This was used to test the drift of the maximum (the mode) of the distributions away from the origin as "p" is increased. The second inverse problem was minimum entropy deconvolution. This was used to test the general shape of the distributions for different values of "p". Both test showed good qualitative agreement. These tests were qualitative in nature, more tests are needed to quantify the results and also to indicate the bounds of applicability for these relationships. A method to tailor PERFs based on a priori moment information was shown and used in conjunction with deterministic properties (constant phase shift invariance) in automated constant phase shift correction.

BIBLIOGRAPHY

[1] Barrodale, and Roberts, F.D.K., (1974), Solution of an Overdetermined System of Equations in the ℓ^1 Norm, Communications of the Association of Computing Machines, V. 19, No. 6, p 319-320.

[2] Bolza, O., (1904), Lectures on the Calculus of Variations, Dover Publications, Inc., New York, N.Y.

[3] Cabrelli, C. A.,(1984), Minimum Entropy Deconvolution and Simplicity: A Noniterative Algorithm, Geophysics, V. 50, No. 3, p 394-413.

[4] Claerbout, J. R., (1976), Fundamentals of Geophysical Data Processing: With Application to Petroleum Prospecting, International Series in the Earth and Planetary Sciences, McGraw-Hill Book Co., Inc., N.Y.

[5] Deeming, T. J., (1981), Deconvolution and Reflection Coefficient Estimation using the Minimum Entropy Principle, Presented at the 51st Ann. Mtg., and Expos., Soc. Explor. Geophys., Los Angeles.

[6] Donoho, D., (1981), On Minimum Entropy Deconvolution, Applied Time Series Analysis II, Academic Press, Inc., New York.

[7] Freund, J. E., (1971), Mathematical Statistics, 2nd Ed., Prentice-Hall, Inc., Englewood Cliffs, N.J.

[8] Goffman C., and Pedrick G., (1965), First Course in Functional Analysis, Prentice-Hall, Inc., Englewood Cliffs, N.J.

[9] Goldstein, H., (1950), Classical Mechanics, Addison-Wesley Pub. Co., Inc., Reading, Mass.

[10] Gray, W. C., (1979), Variable Norm Deconvolution, Ph.D. Thesis, Stanford University, S.E.P., Report 19.

[11] Hadley, G., (1962), Linear Programming, Addison-Wesley Pub. Co., Inc., U.S.A.

[12] Jeffreys, H., and Jeffreys, B. S., (1980), Methods of Mathematical Physics, 3rd Ed., Cambridge University Press, Cambridge.

[13] Justusson, B. I., and Tyan, S. G., (1981), in: 2-D Digital Signal Processing II, Transforms and Median Filters, T.S. Huang Ed., Springer Verlag, Berlin, Heidelberg, N.Y.

[14] Lee, Y. W., (1967), Statistical Theory of Communication, John Wiley and Sons, Inc., N.Y.

[15] Levy, S., and Fullagar, P. K., (1981), Reconstruction of a Sparse Spike Train from a Portion of its Spectrum and Application to High-Resolution Deconvolution, Geophysics, V. 46, No. 9, p 1235-1243.

[16] Levy, S., and Oldenburg, D. W., (1987), Automatic Phase Correction of Common-Midpoint Stacked Data, Geophysics, V. 52, No. 1, p 51-59.

[17] Lines, L. R., and Treitel, S., (1984), Tutorial, A Review of Least-Squares Inversion and its Application to Geophysical Problems, Geophysical Prospecting, V. 32, p 159-186.

[18] Lines, L. R., and Ulrych, T.J., (1977), The Old and New in Seismic Deconvolution and Wavelet Estimation, Geophysical Prospecting, V. 25, No.3, p 512-540.

[19] Morse, P. M., and Feshbach, H., (1953), Methods of Theoretical Physics, McGraw-Hill Book Co., Inc., U.S.A.

[20] Musser, T. J., King, D., and Wason, C. B., (1986), Total Differential Statics: a Comparative Evaluation with Other Statics Procedures, Presented at the 56th Ann. Mtg., Soc. Explor. Geophys., Houston, Texas.

[21] Neville, A. M., and Kennedy, J. B., (1968), Basic Statistical Methods for Engineers and Scientists, International Textbook Co., Scranton, Penn.

[22] Reif, F., (1965), Fundamentals of Statistical and Thermal Physics, McGraw-Hill, Inc., N.Y.

[23] Robinson, E. A., and Treitel, S., (1980), Geophysical Signal Analysis, Prentice-Hall, Inc., Englewood Cliffs, N.J.

[24] Twomey, S., (1977), Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements, Elsevier Scientific Pub. Co., N.Y.

[25] Ursin, B., and Holberg, O.,(1985), Maximum-Likelihood Estimation of Seismic Impulse Response, Geophysical Prospecting, V. 33, p 233-251.

[26] Whittaker, E., and Robinson, E., (1944), The Calculus of Observations, 4th Ed., Blackie and Son Ltd., London and Glasgow.

[27] Wiggins, R. A., (1978), Minimum Entropy Deconvolution, Geoexploration 16, p 21-35.

APPENDIX A

PROOF OF $L \tilde{\propto} \langle\mu\rangle$

Proof of the relation "$L \tilde{\propto} \langle\mu\rangle$" will follow the development in Y. W. Lee's book [14]. Let "n" be the random variable we are to measure, and let the set

$$\{\tilde{n}_i : i=1, \ldots, N\}$$

be the exhaustive mutually exclusive values it can assume. Performing the experiment "M" times furnishes the resulting set

$$\{n_i : i=1, \ldots, M\} ,$$

with the following proportions:

$\tilde{N}_1$ occurrences of $n=\tilde{n}_1$
$\tilde{N}_2$ occurrences of $n=\tilde{n}_2$
$\cdot$
$\cdot$
$\cdot$
$\tilde{N}_N$ occurrences of $n=\tilde{n}_N$

$$\sum_{i=1}^{N} \tilde{N}_i = M .$$

The resulting value of the regulating function will be:

$$L = \sum_{i=1}^{M} \mu(n_i) = \sum_{j=1}^{N} \tilde{N}_j \, \mu(\tilde{n}_j) .$$

Division by "M" gives

$$L/M = \sum_{j=1}^{N} (\tilde{N}_j/M)\ \mu(\tilde{n}_j)\ . \tag{A.1}$$

"Bernoulli's theorem states that, if in a series of M independent trials of a conceptual random experiment the number of successes of an event is $\tilde{N}_j$ (subscript "j" uniquely determines one such event) and the probability of the event is P_j, then the probability that the frequency ratio $\tilde{N}_j/M$ differs from P_j by less than a preassigned quantity ϵ, however small, tends to unity as M tends to infinity. In symbolic form,

$$P(|\tilde{N}_j/M - P_j| < \epsilon) \dashrightarrow 1\ , \tag{A.2}$$

as M -----> ∞" (Lee [14]).

Using result (A.2), (A.1) can be written as:

$$\frac{L}{M} \dashrightarrow \sum_{j=1}^{N} P_j \mu(\tilde{n}_j) = \langle\mu\rangle\ ,$$

as M -----> ∞. Therefore, for M sufficiently large, we can write:

$$\frac{L}{M} \cong \langle\mu\rangle$$

or

$$L \mathrel{\overset{\sim}{\propto}} \langle\mu\rangle$$

If "M ---> ∞"; then the proportionality is strict, "$L \propto \langle\mu\rangle$".

To see how to go from the discrete to the continuous form, refer to Y. W. Lee's book [14].

APPENDIX B

LIMITS OF INTEGRATION w.r.t. EULER'S EQUATION

The integral equation

$$I = \int_{-\infty}^{\infty} L(x,\mu,\mu') \, dx \tag{B.1}$$

is not cast in the exact form required by the classical derivation of Euler's equation, which requires the end points to be finite and fixed.

A simple device can be used to remedy this problem. Since we want "μ" to be zero at some point "n_0" (usually the mean), then the integral (B.1) can be broken into two integrals in the following way:

$$I = I_1+I_2 = \int_{-\infty}^{n_0} L \, dx + \int_{n_0}^{\infty} L \, dx \, . \tag{B.2}$$

Let the probability distribution "P" become negligible beyond a certain value " n_1 ". In other words, "P(x)" can be assumed to be zero beyond "$\pm n_1$ "(which implies the same for "L"). It would not be too limiting if we set "μ" to a constant value "M" at "$\pm n_1$ ". The reason for this is that "μ" is still allowed to vary with relative impunity where "P" has greatest effect. With this in mind, (B.2) can be rewritten as:

$$I \cong I_1+I_2 = \int_{-n_1}^{n_0} L \, dx + \int_{n_0}^{n_1} L \, dx \, . \tag{B.3}$$

Integrals I_1 and I_2 are now in the exact form required by the classical derivation of Euler's equations. Of course, a more elaborate technique may be applied by limiting procedures in the classical derivation which will allow for infinite end points as well, but this would add little to the development.

APPENDIX C

DERIVATION OF EULER'S EQUATION

Here, the calculus of variations will be introduced and Euler's equation will be derived. An excellent account of this can be found in Bolza's book [2]. Consider the following set of curves

$$\underline{M} = \{C_i : C_i \text{ representable as } y=f_i(x) \text{ for } x_0 \leq x \leq x_1\},^*$$

and a function of three independent variables

$$F(x,y,y'),$$

such that the integral

$$J_i = \int_{x_0}^{x_1} F(x,y,y')\, dx \tag{C.1}$$

taken along curve "C_i" has determinate finite values. This gives the natural pairing of a curve with a specific integral value

$$(J_i, C_i) .$$

---

* This can be relaxed by parametric representation of curves, see Bolza [2].

Let "J_0" be an extremum of J_i; this means "$C_0 \in \underline{M}$" extremizes (C.1). Create a neighborhood around "C_0" of all points $\leq \rho$ from "C_0", as indicated by figure C.1.

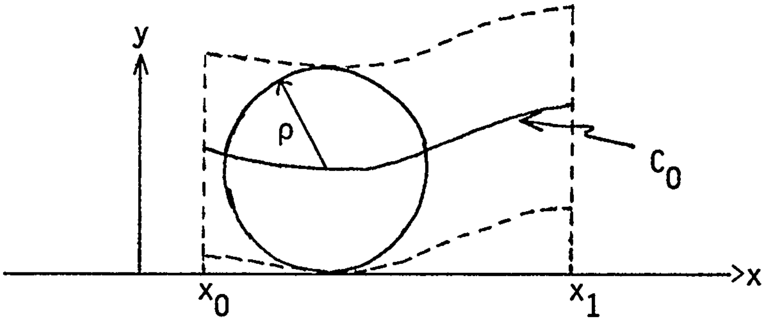


Figure C.1
Neighborhood Around C_0

Replace curve "C_0" with curve "C_m" where

$$C_m \in \eta \ni \eta \subset \underline{M}$$

and

$$\eta \equiv \{C_i : \text{which lie everywhere within the neighborhood established by "}C_0\text{" and "}\rho\text{"}\}$$

To simplify matters, let the extremum be a minimum. (Note: ρ is chosen to ensure that all other $C_m \in \eta$ will not be extremal.) The increment in "y" will be defined as:

$$\Delta y = y_m - y_0 = f_m(x) - f(x) = w(x)$$
$$\Longrightarrow |w(x)| < \rho .$$

The total variation of the integral, which must be positive, will be:

$$\Delta J = J_m - J_0 = \int_{x_0}^{x_1} [F(x,y+w,\ y'+w') - F(x,y,y')]\ dx. \qquad (C.2)$$

Expanding $F(x,y+w, y'+w')$ as a two-variable Taylor series around the point (y,y') gives:

$$F(x,y+w, y'+w') = F(x,y,y') + [F_y w + F_{y'} w'] + 1/2! \, [F_{yy} w^2 + 2F_{yy'} ww' + F_{y'y'} w'^2] + \theta(w^3) \; . \qquad (C.3)$$

Substitution of (C.3) into (C.2) gives

$$\Delta J = \int_{x_0}^{x_1} [F_y w + F_{y'} w'] dx + 1/2! \int_{x_0}^{x_1} [F_{yy} w^2 + 2F_{yy'} ww' + F_{y'y'} w'^2] dx + \ldots \qquad (C.4)$$

Now we adopt a device introduced by Lagrange. Let the variation be represented in the following form:

$$w(x) = \delta\psi(x) \; , \qquad (C.5)$$

where "ψ" has similar properties to "w", and "δ" is a constant that is small enough to ensure

$$\delta\psi < \rho.$$

Using (C.5), we can write (C.4) as:

$$\Delta J = \delta[\int_{x_0}^{x_1} F_y \psi + F_{y'} \psi' dx + \theta(\delta)] \; . \qquad (C.6)$$

Since "J" is a minimum for "δ=0", then for "δ" sufficiently small $\theta(\delta)$ will become insignificant and (C.4) becomes:

$$\Delta J \cong \delta \int_{x_0}^{x_1} [F_y \psi + F_{y'} c'] \; dx. \qquad (C.7)$$

Since ΔJ must be positive, independent of our choice of sign for "δ", then the following must be true:

$$\int_{x_0}^{x_1} [F_y \psi + F_{y'} \psi'] \; dx = 0. \qquad (C.8)$$

In other words, condition (C.8) must be satisfied if (C.1) is to be an extremum. To cast (C.8) in a more useful form, we integrate by parts resulting in:

$$[\psi F_{y'} \Big|_{x_0}^{x_1} + \int_{x_0}^{x_1} [\psi(F_y - \frac{d}{dx} F_{y'})] \; dx = 0. \qquad (C.9)$$

Now, we use the condition that the end points are fixed and not allowed to vary. This means that $\psi = 0$ for $x = x_0$ and x_1. Then the first term of (C.9) is zero. Since ψ is arbitrary in the second term, the only way this term can vanish is if

$$F_y - \frac{d}{dx} F_{y'} = 0 .^* \qquad (C.10)$$

This is known as Euler's equation.

---

* To see that this assertion is true, let

$$M(x) = F_y - \frac{d}{dx} F_{y'} .$$

Now if $M(x) \neq 0$, set $\psi(x)$ to be a smooth function $\neq 0 \ni$ the sign of $\psi(x)$ = sign of $M(x)$, $\forall x \quad (x_0, x_1)$ ==>

$$\int_{x_0}^{x_1} \psi M dx > 0 \qquad ==><==$$

$$\therefore M(x) = 0, \forall x. \square$$

APPENDIX D

SOLVING FOR $\mu(x)$ AS A FUNCTION OF $P(x)$

Given:

$$P(x) = \frac{\alpha\, \mu''(x)}{[\sqrt{1 + \mu'^{2}(x)}\,]^{3}} \;\&\; \mu(x_0) = \mu'(x_0)=0.$$

Prove:

$$\mu(x) = \int_{x_0}^{x} \frac{F(y)}{\sqrt{1 - F^{2}(y)}}\, dy,$$

where

$$F(x) = \frac{1}{\alpha}\int_{x_0}^{x} P(y)\, dy.$$

Proof:

$$\int_{x_0}^{x} \frac{\alpha\mu''}{[\sqrt{1 + u'^{2}}\,]^{3}}\, dy = \int_{x_0}^{x} P(y)\, dy.$$

Let $\beta=\mu'$ ==> $d\beta=\mu'' dy$, assuming μ'' exists, then

$$\int_{y=x_0}^{x} \frac{d\beta}{(1+\beta^{2})^{3/2}} = \frac{1}{\alpha}\int_{x_0}^{x} P(y)dy.$$

Let $F(x) = \frac{1}{\alpha}\int_{x_0}^{x} P(y)dy,$

and knowing

$$\int (x^2 + a^2)^{-3/2} dx = x/[a^2 \ (x^2+a^2)^{1/2}]$$

we arrive at

$$\left[\frac{\mu'(y)}{\sqrt{1 + \mu'^2(y)}}\right|_{y=x_0}^{x} = F(x) \ ,$$

$$\Longrightarrow \quad \left[\frac{\mu'(x)}{\sqrt{1 + \mu'^2(x)}}\right] = F(x)$$

$$\Longrightarrow \quad \mu'(x) = \frac{F(x)}{\sqrt{1 - F^2(x)}}$$

$$\Longrightarrow \quad \left[\mu(y)\right|_{y=x_0}^{x} = \int_{y=x_0}^{x} \left[\frac{F(y)}{\sqrt{1 - F^2(y)}}\right] dy$$

$$\Longrightarrow \quad \mu(x) = \int_{x_0}^{x} \left[\frac{F(y)}{\sqrt{1 - F^2(y)}}\right] dy \ .$$

□

APPENDIX E

MIXED PROCESSES

This appendix is broken into five sections. Each section is meant to show a particular mixed process. The complexity of the process increases generally with section number.

SECTION E1

NOISE CONTAMINATION

The first process considered is also the simplest. Here, we assume the process consists of two parts: a wholly deterministic and invertible part (figure 2.2.1) "F" and a random part (figure 2.2.3b) "N_i", which will be called noise.

These parts are related by addition. Thus, the whole process "S" can then be written as:

$$S(.) = F(.) + N_i \ .$$

The additive noise N_i, being random and unpredictable in nature, is different for repeats of the experiment with the same input "Y". This repetition gives rise to a set of "m" observations, or outputs:

$$\underline{X} \equiv \{X_i : X_i = S(y) = F(y) + N_i; \ i=1,2,...m\} \ .$$

The goal is to find an inverse procedure that will allow the estimation of "Y", the input. There are two simple ways to accomplish this. These correspond to attacking the additive noise in either the output or input domains.

The first inversion procedure starts with trying to estimate "X", where:

$$X = F(Y) .$$

Since the set "$\underline{X}$" consists of elements of the form

$$X + N_i ,$$

an estimate of "X" can be arrived at by finding an object "$\tilde{X}$" which is "most similar" to all the elements of "$\underline{X}$".

As mentioned in (2.1), regulating functions provide a quantitative indicator of similarities between two objects. Let "L" be such a regulating function. Given a regulating function "L" and a domain of search "$\underline{D}$", the set of real numbers,

$$\underline{A} \equiv \{a_{ij} : a_{ij} = L(\hat{X}_i - X_j) \ni \hat{X}_i \in \underline{D} \; \& \; x_j \in \underline{X}\} ,^1$$

---

[1] Note the use of the notation $L(\hat{X}_i - X_j)$ instead of $L(\hat{X}_i, X_j)$. This was done to emphasize the inverse aspect of this case. This is also consistent with the mathematical definition in the next section.

can be generated. For a successful regulating function, these numbers should reflect the similarity between what we think "X" might be (an element of $\underline{D}$), to what we know is "X" distorted by additive noise (an element of $\underline{X}$). But, "$\underline{A}$" does not uniquely determine a "$\tilde{X} \in \underline{D}$" which is an optimum representation of "X". As an intermediate step towards finding an optimum, sort the set "$\underline{A}$" into a set of sets "$\bar{A}$" in the following manner:

$$\bar{A} \equiv \{\bar{a}_i : \bar{a}_i \equiv \{a_{i1},a_{i2},\dots,a_{iM}\} \ni a_{ij} \in \underline{A}\} .$$

The elements of set "$\bar{A}$" are uniquely paired to elements in set "$\underline{D}$" as indicated below:

$$(\bar{a}_i, \hat{X}_i) .$$

A special situation arises when an element of set "$\bar{A}$", say "$\bar{a}_R$", has all zero elements. This can only occur when the following condition is satisfied:

$$\hat{X}_R = \tilde{X} = X .$$

This corresponds to a noise-free case; though of little interest in itself, it does provide an intuitively reasonable ideal which the desired optimum to be found should approach.[1] To meet this end, the element of "$\bar{A}$" should be ascertained which is most similar to such a set. Let "$\bar{0}$" be such a set,[2] and let the regulating function "L'" indicate the similarity between "$\bar{0}$" and elements of "$\bar{A}$". Generate a second set of real numbers "$\underline{B}$", where:

$$\underline{B} \equiv \{b_i : b_i = L'(\bar{a}_i - \bar{0}) = L'(\bar{a}_i) \ni a_i \in \bar{A}\} .$$

Elements of sets "$\underline{B}$", "$\bar{A}$" and "$\underline{D}$" can now be grouped as follows:

$$(b_i, \bar{a}_i, \hat{x}_i) \ni b_i \in \underline{B}, \bar{a}_i \in \bar{A} \ \& \ \hat{x}_i \in \underline{D}.$$

This grouping allows one to go from choosing the smallest element in "$\underline{B}$" to an element in "$\underline{D}$", which will be termed optimum with respect to regulating functions "L" and "L'". Again, let "$\tilde{X}$" be this optimum estimate of "x". Now, if the inverse "F^{-1}" has the very necessary property of mapping objects close to "X" into objects close to "Y", the approximate inverse can be completed by writing:

$$F^{-1}(\tilde{X}) = \tilde{Y} \cong Y .$$

---

[1] This noise-free case is no longer a mixed process. It is wholly deterministic and invertible.

[2] $\bar{0} \equiv$ set of "m" elements all of which are zero.

This case corresponds to attacking the noise in the output domain, but one could also attack the noise in the input domain. To do this, we first map all elements of the output set "$\underline{X}$" using "F^{-1}" into a set of potential inputs "$\underline{Y}$". "$\underline{Y}$" is defined as:

$$\underline{Y} \equiv \{\hat{Y}_i : \hat{Y}_i = F^{-1}(X_i) \ni X_i \in \underline{X}\} .$$

To simplify this problem, let "F^{-1}" be linear. This would make the elements of "Y" have the following form:

$$\begin{aligned} Y_i &= F^{-1}(X_i) \\ &= F^{-1}(S(Y)) = F^{-1}(F(Y) + N_i) \\ &= Y + F^{-1}(N_i) \end{aligned}$$

let

$$F^{-1}(N_i) = \hat{N}_i ,$$

then

$$Y_i = Y + \hat{N}_i .$$

Here we have the same situation as in the first case, where a set of results is available that is contaminated by additive noise. The same technique can, therefore, be applied. Choose regulating functions "L" and "L´" to find an optimum estimate of "Y", given the set "$\underline{Y}$" and a domain of search "$\bar{Y}$". The technique perfectly mirrors the procedure utilized to find "$\tilde{X}$" and will not be elaborated upon.

Let "$\tilde{Y}$" be the optimum found, which is assumed to be close to "Y", or:

$$\tilde{Y} \cong Y .$$

This ends the second approximate inverse procedure.

Note that in both cases the regulating function "L" basically acted upon the noise and is, therefore, by our convention, the probabilistic part, and "F" with its inverse, the deterministic part. The dominance of the deterministic part was assured by our choice of the smallest element of set "$\underline{B}$". Note how important it is for a regulating function to be able to determine when what it sees is actually a manifestation of the noise.

The process just discussed is summarized in figure E1.1.

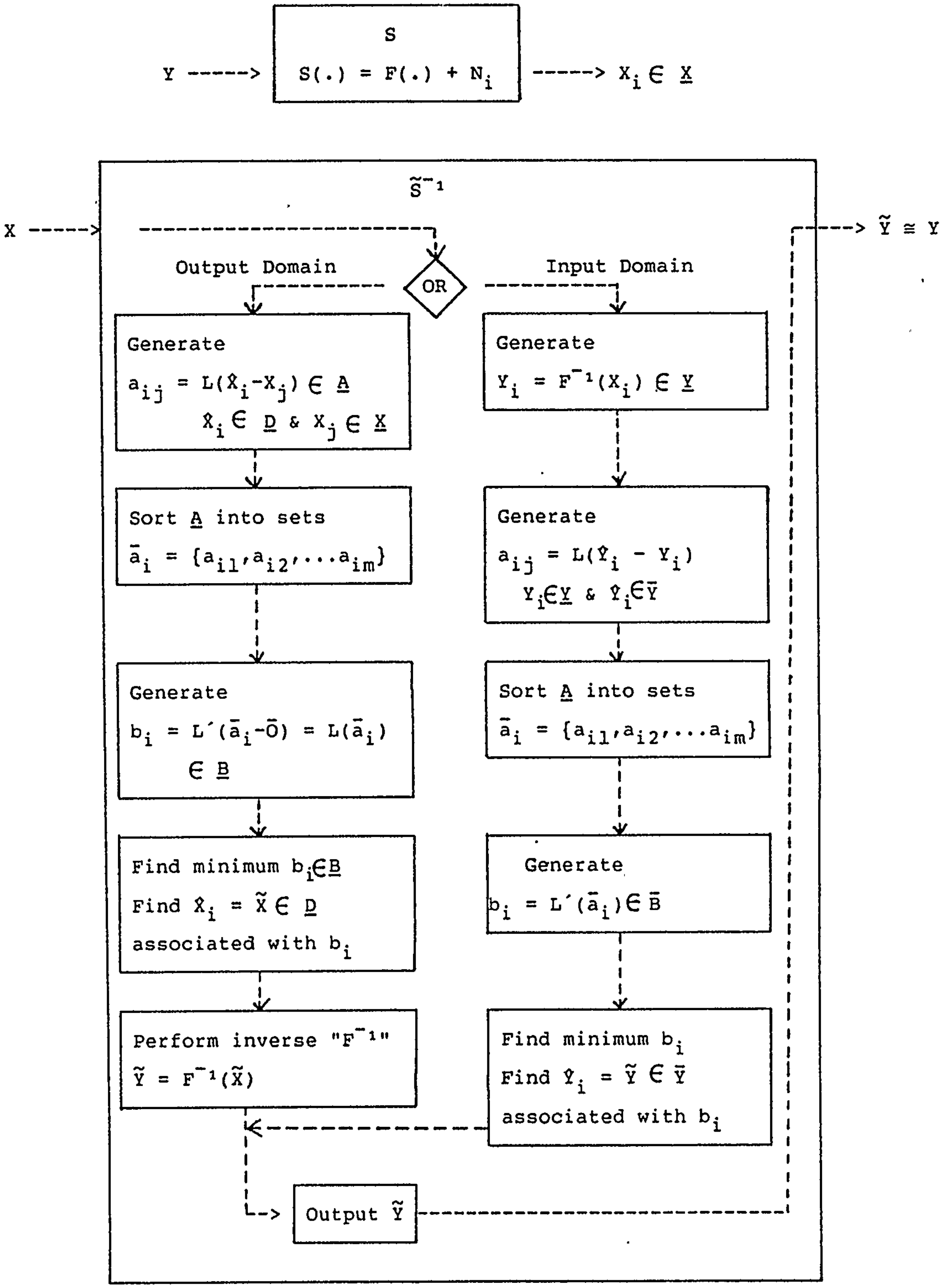


Figure E1.1
Noise Contamination

SECTION E2

NON-UNIQUENESS

The second process to be examined consists of a forward process "S" which is non-unique. Non-uniqueness implies many inputs to "S" can have the same output. The non-uniqueness will be attributed to a set of unknown parameters "p_i". In other words, if the "p_i's" are known, the output can then be uniquely paired to an input. The exact inverse, given a set of parameters "p_i'", must be known. Let "$S^{-1}(p_i';.)$" be this inverse.

As in case I, there are two obvious ways to frame an approximate inverse. These correspond to finding an optimum in the input or output domains. In both domains a set of allowable parameters "$\underline{P}$" must be defined. In the "Output Domain" procedure, which will be considered first, an additional set must be defined. This set "$\bar{Y}$" consists of possible inputs. With this and the knowledge of the forward process "S", a set of trial outputs "$\underline{X}$" can be generated, where:

$$\underline{X} \equiv \{X' : X' = S(p_i';Y') \ni Y' \in \bar{Y} \; \& \; p_i \in \underline{P}\} .$$

The elements of sets $\underline{X}$, $\bar{Y}$ and $\underline{P}$ are naturally grouped as:

$$(X', Y', p_i') .$$

Thus, if an element of "$\underline{X}$" can be found to be optimum, a corresponding element in "$\bar{Y}$" can then be termed an optimum representation of "Y". To this end, a regulating function "L" is introduced to indicate closeness of elements of "$\underline{X}$" to the actual output "X".

Let "$\tilde{X} \in \underline{X}$" be this element; in other words:

$$L(\tilde{X},X) \leq L(X',X), \forall\, X' \in \underline{X} \,.$$

By the natural grouping mentioned, we arrive at an optimum representation of "Y", namely, "$\tilde{Y}$". This concludes the first method.

The second approximate inverse starts by defining the set of potential inputs "$\underline{Y}$" as:

$$\underline{Y} \equiv \{Y' : Y' = S^{-1}(p_i';X) \ni p_i' \in \underline{P}\} \,.$$

At this point, there must be some concept of desirability the input should have.[1] This desirability is then embodied in a regulating function, which attains an extremum as the gauged object approaches the desired state. Since the extremum can be either a minimum or maximum,

---

[1] Desirability at this point is necessarily a nebulous concept, since it encompasses many manifestations. The importance is the ability to incorporate it in a regulating function. Desirability can be a probability distribution to which the input should be close, a trait like spikiness, or it could be closeness to some reference object.

depending on its design, the choice of maximum will be made with no loss in generality. Given such a regulating function "L", the optimum "$\tilde{Y}$" will be signified by the following condition:

$$L(\tilde{Y}) \geq L(Y'), \ \forall \ Y' \in \underline{Y} \ .$$

This concludes the second "Input Domain" procedure. The entire process can be summarized pictorially in figure E2.1.

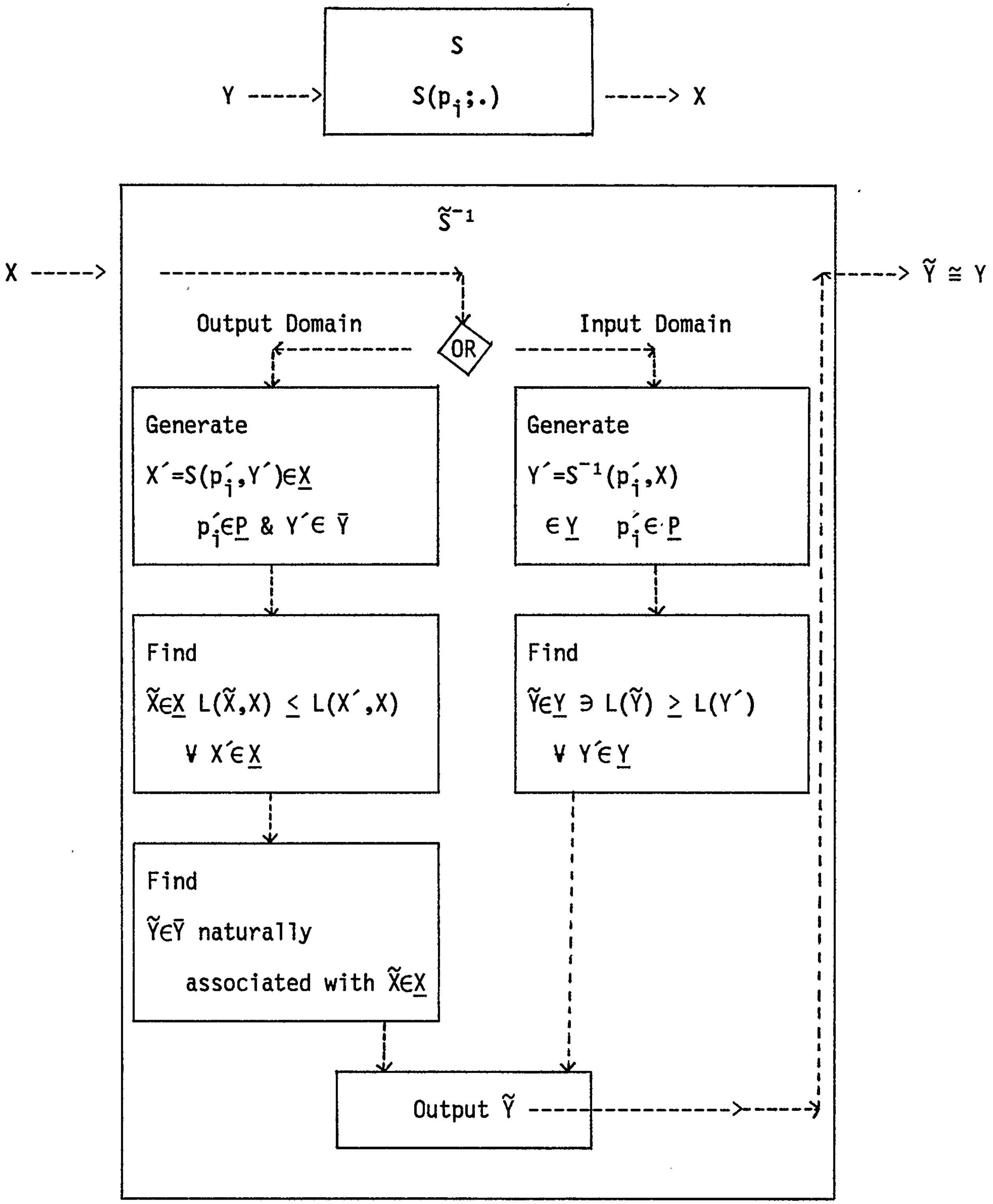


Figure E2.1
Non-Uniqueness

The benefit of the "Output Domain" procedure comes from not needing to know "S^{-1}". The disadvantage arises from the chance of increased non-uniqueness due to the additional parameters introduced by the set $\bar{Y}$.

A good example of this case is band-passed data. Within the pass-band the original data are preserved. Outside of the pass-band the data are altered. There are infinitely many data sets which are identical within the pass-band and differing outside. To choose the most desirable, a regulating function can be introduced to pick one out of the many. This is basically what is done by Levy and Fullager [15]. They assume zero-phase, band-passed data and choose the full band representation which minimizes the ℓ^1 norm.

SECTION E3

SHAPING PROCESSES

The third process has much in common with the process considered in appendix E2. The difference between the two is basically conceptual since, in actual implementation, they are identical.

To begin, consider a process "S" where the exact inverse "S^{-1}" is not available. An approximate inverse "$\tilde{S}^{-1}(p_i;.)$" is introduced. The parameters "p_i" allow more flexibility in the approximation. "$\tilde{S}^{-1}$" can be seen as a shaper, and what it does is shape output "X" to a desirable form "Y".

The actual procedure, as can be anticipated, is identical to the previous case. The only change to case two is to substitute "$\tilde{S}^{-1}$" for all occurrences of "S^{-1}". To complete the analogy in the "Output Domain", consider the situation where only an approximate forward process "$\tilde{S}(p_i;.)$" is on hand. By the substitution of "$\tilde{S}$" in lieu of "S", the analogy between this process and the one considered in appendix E2 is complete.

A good example of this case, which will be examined in greater detail later in this paper, is the use of constant phase shifts as a final effort in dephasing a time series. One automated technique for doing this can be found in Levy and Oldenburg's [16] paper. Here, it is assumed the residual phase can be approximated by a frequency independent shift. The approximate inverse "$\tilde{S}^{-1}$" is, therefore, constant phase-shifting. The variable parameter is the phase shift.

SECTION E4

CASCADED PROCESSES

The fourth process is an example of how the processes considered in appendix E1 and E2 can be used together. The forward process "S" in this case consists of a non-unique forward process "$F(p_j;.)$" and additive noise "N_i". Symbolically, this can be written as:

$$S(.) = F(p_j;.) + N_i \ .$$

As in case II, the inverse of "F" for any set of parameters "p'_j" is known. Let "$F^{-1}(p'_j;.)$" be this inverse.

This case can be seen as two processes cascaded together. Figure E4.1 attempts to show this.

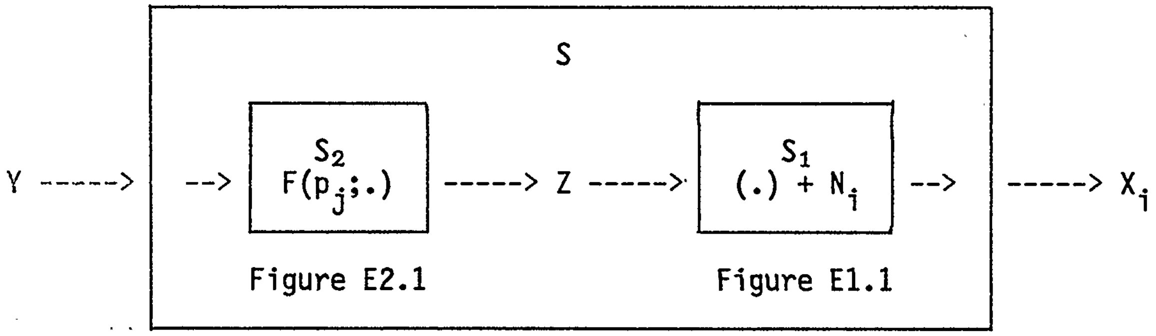


Figure E4.1
Cascaded Process

Figure E4.1 demonstrates how this process can be seen as the combination of processes represented in figures E2.1 and E1.1. The input "Y" passes through the forward process in appendix E2 "S_2", giving the intermediate result "Z". The resultant "Z" is then subjected to the forward process in appendix E1 "S_1", producing the output "X_i". Symbolically, this can be represented by:

$$Z = S_2(Y) = F(p_j;Y)$$
$$X_i = S_1(Z) = Z + N_i = F(p_j;Y) + N_i \ ,$$

or simply:

$$X_i = S_1[S_2(Y)] \ .$$

This immediately suggests the cascading of inverses from appendices E1 and E2. Let "$\tilde{S}_1^{-1}$" be the approximate inverse from appendix E1, as shown in figure E1.1. Let "$\tilde{S}_2^{-1}$" be the approximate inverse from appendix E2, as shown in figure E2.1. Then "$\tilde{S}_1^{-1}$" can be used to deduce an optimum representation of "Z", namely, "$\tilde{Z}$", from the outputs "X_i". "$\tilde{Z}$" can then be input to "$\tilde{S}_2^{-1}$" to get an optimum estimate of "Y", namely, "$\tilde{Y}$. Figure E4.2 attempts to show this pictorially.

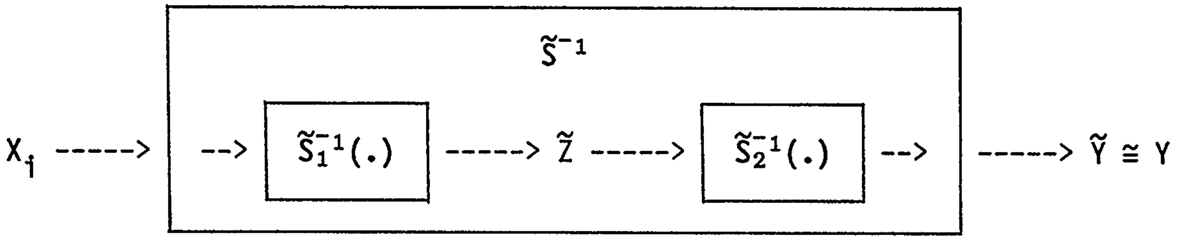


Figure E4.2
Cascaded Inverse

Symbolically, this can be represented by:

$$\tilde{Z} = \tilde{S}_1^{-1}(X_i) = \tilde{S}_1^{-1}(S_1[S_2(Y)]) = \tilde{S}_2(Y)$$

$$\tilde{Y} = \tilde{S}_2^{-1}(\tilde{Z}) = \tilde{S}_2^{-1}(\tilde{S}_2[Y]) \cong Y \ .$$

An example of this process is full-band inversion of band-limited data in the presence of noise.[1]

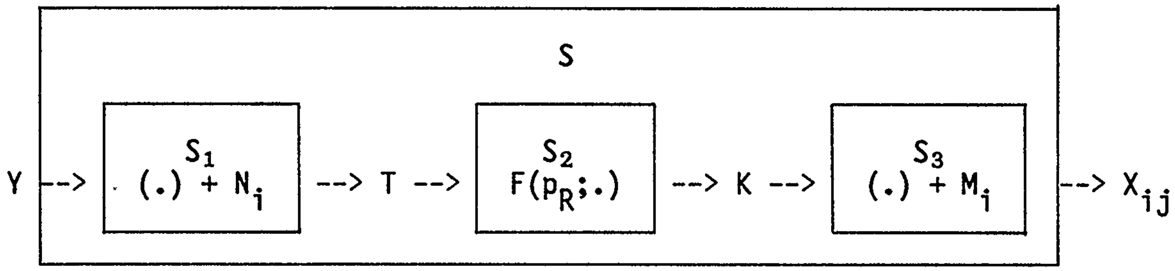
---

[1] Redundancy is assumed for the noise cancellation process to work.

SECTION E5

SUPER CORRUPTER

Lastly, to show how complex a process can be, consider the following case represented in figure E5.1.



$$X_{ij} = S(Y) = F(p_R;Y+N_i) + M_j$$

Figure E5.1
Super Corrupter

This case contains additive noise "N_i" and "M_j" before and after the non-unique forward process "F". An approximate inverse procedure can be designed based on appendix E4, but due to the complexity in this case, the result may not be very desirable. A method here to obtain good results is anybody's guess. It would depend on how "F" treats what is called noise and on our knowledge of what "$\tilde{Y}$" should look like.

The complexity for these cases can be increased <u>ad infinitum</u>, but if the errors are bounded, an approximation can be used. Otherwise, "S" will tend towards a totally uninvertible process (figure 2.2.3b).

APPENDIX F

PROOF OF COMMON EXTREMUM

Here, we will prove that the regulating functions

$$\|x\|_p = [\sum_i |x_i|^p]^{1/p} \tag{F.1}$$

and

$$L = \sum_i |x_i|^p \tag{F.2}$$

indeed have the same extrema. Since, in this paper regulating functions are used to indicate extremum conditions, this will allow us the flexibility of calling (F.2) the ℓ^p norm in place of the classical form given by (F.1). The difference will be the gradient away from the extremum.

The proof begins by setting up the extremal condition for both equations then showing that they are indeed identical. For equation (F.1),

$$\frac{\partial}{\partial x_i}\|x\|_p = \frac{\partial}{\partial x_i}\left[\sum_i |x_i|^p\right]^{1/p} = \frac{1}{p}\left[\sum_i x_i^{\ p}\right]^{(1/p)-1} \frac{\partial}{\partial x_i} \sum_i |x_i|^p = 0$$

$$\Longrightarrow \frac{\partial}{\partial x_i}\left[\sum_i |x_i|^p\right] = 0 \ , \tag{F.3}$$

is the condition required for an extremum. But, that is exactly the same condition for equation (F.2) since

$$\frac{\partial L}{\partial x_i} = \frac{\partial}{\partial x_i}\left[\sum_i |x_i|^p\right] = 0 \ . \tag{F.4}$$

Therefore, combining (F.3) and (F.4), we get:

$$\frac{\partial}{\partial x_i}\|x\|_p = \frac{\partial}{\partial x_i}\left[\sum_i x_i\right]^p = \frac{\partial L}{\partial x_i} = 0 \ . \tag{F.5}$$

This equivalence of extremal conditions proves our contention. □

It should be stated, though, that even with equal extrema, the two regulating functions will have different gradients away from the extrema, and techniques which use the gradient could have different convergence characteristics.

APPENDIX G

PROOF OF $x_{max} \geq d \geq x_{min}$

Given:

the set

$$\underline{X} \equiv \{x_i : x_{max} \geq x_i \geq x_{min} \; \forall \; i\} \; ,$$

and "d" is the value which minimizes the expression

$$L = \sum_i \left| d-x_i \right|^p \quad p \geq 1 \; .$$

Prove:

$$x_{max} \geq d \geq x_{min} \; .$$

Proof:

Assume the contrary, let "L" be minimized by $d \ni d \geq x_{max}$.

$\Longrightarrow \exists \; \delta \in \{\text{positive reals}\} \ni d \geq d-\delta = x_{max}$,

$\Longrightarrow \left| d-\delta-x_i \right| < \left| d-x_i \right|$,

and since $|y|^p$ is an increasing function $\forall \; p \geq 1$,

$$\Longrightarrow |y_1| > |y_2| \Longleftrightarrow |y_1|^p > |y_2|^p \; .$$

$$\left| d-\delta-x_i \right|^p < \left| d-x_i \right|^p \; .$$

$$\Longrightarrow \sum_i \left|d-\delta-x_i\right|^p < \sum_i \left|d-x_i\right|^p = L .$$

===> d does not minimize "L" ===><===.

our assumption is wrong and we must have

$$d \leq x_{max} .$$

□

The proof for $d \geq x_{min}$ is exactly the same as above.

APPENDIX H

MAXIMUM POINT OF ℓ^p PROBABILITY DENSITIES

Given:

Ruler function

$$\mu(x) = |x|^p ,$$

and its associated probability distribution

$$P(x,p) = \frac{1/2 \; p(p-1) \; |x|^{p-2}}{[\sqrt{1 + p^2 \; |x|^{2p-2}} \;]^3} .$$

Prove:

The point of maximum probability density is given by

$$x_{max} = \left[\frac{2-p}{p^2(4-3p)}\right]^{1/2p-2}$$

$$\forall \; x \geq 0 \text{ and } p > 1 .$$

Proof:

For $x \geq 0$ the absolute value sign can be dropped in $P(x,p)$. The maximum density point will have zero first derivative w.r.t. x.

$$\frac{\partial P(x,p)}{\partial x} = \frac{1}{2} \, p(p-1) \left[\frac{p-2x^{p-3}}{[1+p^2x^{2p-2}]^{3/2}} - \frac{3/2 \; p^2(2p-2)x^{3p-5}}{[1+p^2x^{2p-2}]^{5/2}}\right]$$

$$= \frac{1}{2}\, p(p-1) \left[\frac{(p-2)x^{p-3}+p^2x^{3p-5}-3/2\ p^2(2p-2)x^{3p-5}}{[1+p^2x^{2p-2}]^{5/2}} \right] .$$

$$= \frac{1}{2}\, p(p-1) \left[\frac{(p-2)x^{p-3} + p^2(4-3p)x^{3p-5}}{\left[\sqrt{1 + (px^{p-2})^2}\right]^5} \right]$$

Setting this last relation to zero gives the maximum density point.

$$(p-2)x^{p-3} + p^2(4-3p)\ x^{3p-5} = 0$$

$$\Longrightarrow (p-2) + p^2(4-3p)\ x^{2p-2} = 0$$

$$\Longrightarrow \qquad x^{2p-2} = \frac{2-p}{p^2(4-3p)}$$

$$\Longrightarrow \qquad x = \left[\frac{2-p}{p^2(4-3p)}\right]^{1/(2p-2)} .$$

□

APPENDIX I

MINIMUM ENTROPY DECONVOLUTION

What follows is a deconvolution technique consolidated from T. J. Deeming's paper presented at the 1981 S.E.G. Convention [5]. The procedure consists of finding a filter "f_i" which, when convolved with trace "x_i", produces an output "y_i" that maximizes the regulating function "L".

This procedure is motivated by the forward model which considers the output "x_i" as the convolution of a wavelet "w_i" with a sparse spike sequence "r_i", or

$$x_i = \sum_j w_j r_{i-j} \ . \tag{I.1}$$

The inverse is then to find a filter "f_i" which gives a result "y_i" that is spiky and hopefully a good representation of "r_i". The regulating function "L" is chosen to indicate the spiky nature of the result "y_i", where

$$y_i = \sum_{j=0}^{\ell-1} f_j x_{i-j} \ . \tag{I.2}$$

Deeming showed that the sum of any positive increasing function of the scaler and sign invariant mapping of the trace,

$$z_i = \frac{y_i^2}{\sum_j y_j^2} , \tag{E.3}$$

will tend to minimize the number of measured spikes present in the M.E.D. output. Thus, the regulating function can be written as

$$L(y) = \sum_i \mu(z_i) . \tag{E.4}$$

Here, "μ" is the ruler function as considered in this thesis.

For any given result "y_i", (E.4) will be the resulting value of the regulating function. To generalize to a multi-channel case, (E.4) can be rewritten as:

$$V = \sum_t \alpha_t L_t = \sum_t \alpha_t \left[\sum_i \mu(z_i)\right]_t . \tag{E.5}$$

Where the subscript "t" indicates different traces, and the "α's" are weighting factors for each trace. The idea is to find a maximum value for (E.5). For (E.5) to be a maximum, it must obey the following condition:

$$\frac{\partial V}{\partial f_j} = \sum_t \alpha_t \left[\left\{\sum_i \frac{\partial \mu}{\partial z_i} \frac{\partial z_i}{\partial f_j}\right\}_t\right] = 0 . \tag{E.6}$$

Aside:

$$\frac{\partial z_i}{\partial f_j} = \frac{\partial}{\partial f_j}\left[\frac{y_i^2}{\langle y^2\rangle}\right] = \frac{1}{\langle y^2\rangle}\, 2y_i \frac{\partial y_i}{\partial f_j} + y_i^2 \frac{\partial}{\partial f_j}\left[\frac{1}{\langle y^2\rangle}\right]$$

since

$$y_i = \sum_j f_j x_{i-j} \Longrightarrow \frac{\partial y_i}{\partial f_j} = x_{i-j} ,$$

and

$$\frac{\partial}{\partial f_j}\left[\frac{1}{\langle y^2\rangle}\right] = -\langle y^2\rangle^{-2}\left[\frac{\partial\langle y^2\rangle}{\partial f_j}\right] = -\frac{1}{\langle y^2\rangle^2}\frac{\partial}{\partial f_j}\left[\frac{1}{n}\sum_k y_k^2\right]$$

$$= -\frac{1}{n\langle y^2\rangle^2}\sum_k 2y_k \frac{\partial y_k}{\partial f_j} = \frac{-1}{n\langle y^2\rangle^2}\sum_k 2y_k x_{k-j}$$

$$\Longrightarrow \frac{\partial z_i}{\partial f_j} = \frac{2}{\langle y^2\rangle}\left\{y_i x_{i-j} - \frac{z_i}{n}\left[\sum_k y_k x_{k-j}\right]\right\}.$$

Let $\beta_i = \dfrac{\partial\mu}{\partial z_i} = \mu_i'$.

End Aside

Thus

$$\frac{\partial V}{\partial f_j} = \sum_t \alpha_t\left\{\sum_i \beta_i\left[\frac{2}{\langle y^2\rangle}\quad y_i x_{i-j} - \frac{z_i}{n}\sum_k y_k x_{k-j}\right]\right\}$$

$$= \sum_t \alpha_t\left\{\frac{2}{\langle y^2\rangle}\left[\sum_i \beta_i y_i x_{i-j} - \frac{1}{n}\sum_i \beta_i z_i \sum_k y_k x_{k-j}\right]\right\} .$$

But

$$\langle \beta z \rangle = \frac{1}{n} \sum_i \beta_i z_i \ , \tag{E.7}$$

and letting

$$\mu_i = \frac{\beta_i y_i}{\langle \beta z \rangle} \tag{E.8}$$

one can write

$$\frac{1}{2} \frac{\partial V}{\partial f_j} = \sum_t \alpha_t \frac{\langle \beta z \rangle}{\langle y^2 \rangle} \left[\sum_i \mu_i x_{i-j} - \sum_k y_k x_{k-j} \right] . \tag{E.9}$$

Aside:

Note the bracketed portion of (E.9) must equal zero when "$\partial a/\partial f_j = 0$", or

$$\sum_i (\mu_i - y_i) \ x_{i-j} = 0 \ ,$$

up to the lag of the filter; this can be used as a stopping condition.

End Aside

To simplify equation (E.9), the second term in the brackets can be written as

$$\begin{aligned} \sum_k y_k x_{k-j} &= \left[\sum_k \sum_{i=0}^{\ell-1} f_i x_{k-i} \quad x_{k-j} \right] \\ &= \sum_{i=0}^{\ell-1} f_i \sum_k x_{k-i} x_{k-j} \ . \end{aligned}$$

Now define the following terms:

$$g_j^t = \left[\sum_i \mu_i x_{i-j\ t}\right] \equiv \text{crosscorrelation of n \& x for trace "t"} .$$

$$G_j = \sum_t \alpha_t \left[\frac{\langle\beta z\rangle}{\langle y^2\rangle}\right]_t g_j^t$$

$$c_{ij}^t = \sum_k [x_{k-i}\ x_{k-j}]_t \equiv \text{autocorrelation of "x" for trace "t".}$$

$$C_{ij} = \sum_t \alpha_t \left[\frac{\langle\beta z\rangle}{\langle y^2\rangle}\right]_t c_{ij}^t .$$

Given these definitions, equation (E.9) can be rewritten as

$$\frac{1}{2}\frac{\partial V}{\partial f_j} = G_j - \sum_{i=0}^{\ell-1} f_i C_{ij} . \tag{E.10}$$

As stated previously, a maximum is reached when $\partial a/\partial f_j$ is identically zero, which gives the equation

$$\sum_{i=0}^{\ell-1} C_{ij} f_i = G_j . \tag{E.11}$$

As mentioned in an aside, the stopping condition is

$$H_j = 0 \tag{E.12}$$

where

$$H_j = \sum_t \alpha_t \left[\frac{\langle\beta z\rangle}{\langle y^2\rangle}\right]_t h_j^t$$

and

$$h_j^t = \left[\sum_i (\mu_i - y_i) x_{i-j}\right]_t .$$

The total iterative procedure is to first get an initial estimate of the filter "f_i^o", which can be used to get an initial result "y_i^o", giving "z_i^o", "$<y_i>^o$" and "$<\beta z>^o$", which in turn allows the calculation of "C_{ij}^o", and "G_j^o". At this point, a new filter can be solved for through equation (E.11) giving "$f_i^{(1)}$", the updated filter. This procedure is continued until "$H_j = 0$", or convergence by the Cauchy criterion.

Summary of symbols and equations:

x_i ≡ Input trace.

y_i ≡ Output filtered trace.

f_i ≡ filter $(i=0,1,\ldots,\ell-1)$.

$z_i = y_i^2/<y^2>$ ≡ Sign and scaler invariant mapping.

$L_t = \left[\frac{1}{n} \sum_i \mu(z_i)\right]_t$ ≡ regulating function for trace "t".

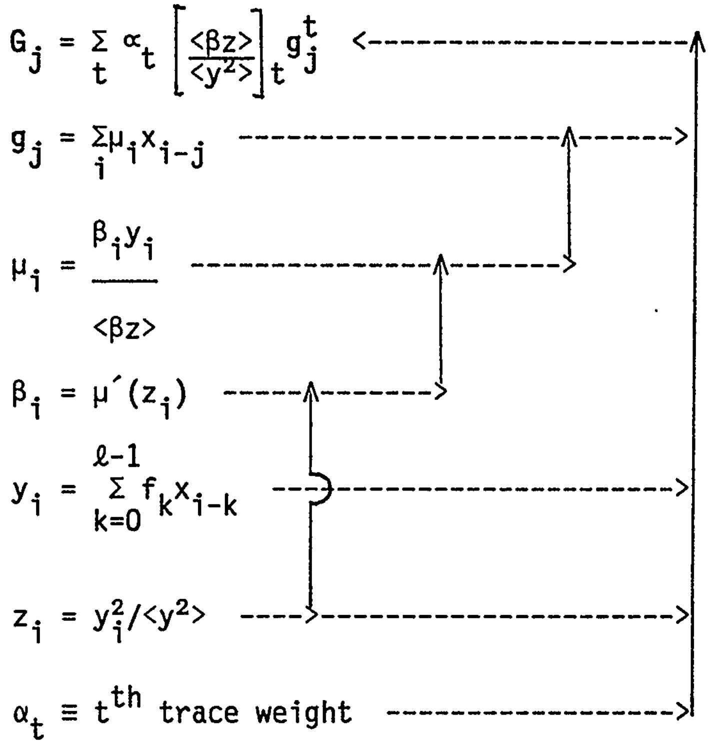
$V = \sum_t L_t$ ≡ multi-channel M.E.D. regulating function.

---

f_i^o ≡ Initial filter estimate.

$\sum_{i=0}^{\ell-1} f_i^{(k)} C_{ij} = G_j$ ≡ Iteration equation (k^th Iteration).

$H_j = 0$: Stopping condition $j=0,1,\ldots\ell-1$.



$$G_j = \sum_t \alpha_t \left[\frac{\langle\beta z\rangle}{\langle y^2\rangle}\right]_t g_j^t$$

$$g_j = \sum_i \mu_i x_{i-j}$$

$$\mu_i = \frac{\beta_i y_i}{\langle\beta z\rangle}$$

$$\beta_i = \mu'(z_i)$$

$$y_i = \sum_{k=0}^{\ell-1} f_k x_{i-k}$$

$$z_i = y_i^2/\langle y^2\rangle$$

$\alpha_t \equiv t^{th}$ trace weight

$$C_{ij} = \sum_t \alpha_t \left[\frac{\langle\beta z\rangle}{\langle y^2\rangle}\right]_t c_{ij}^t$$

$$c_{ij} = \sum_k x_{k-i}\, x_{k-j}$$

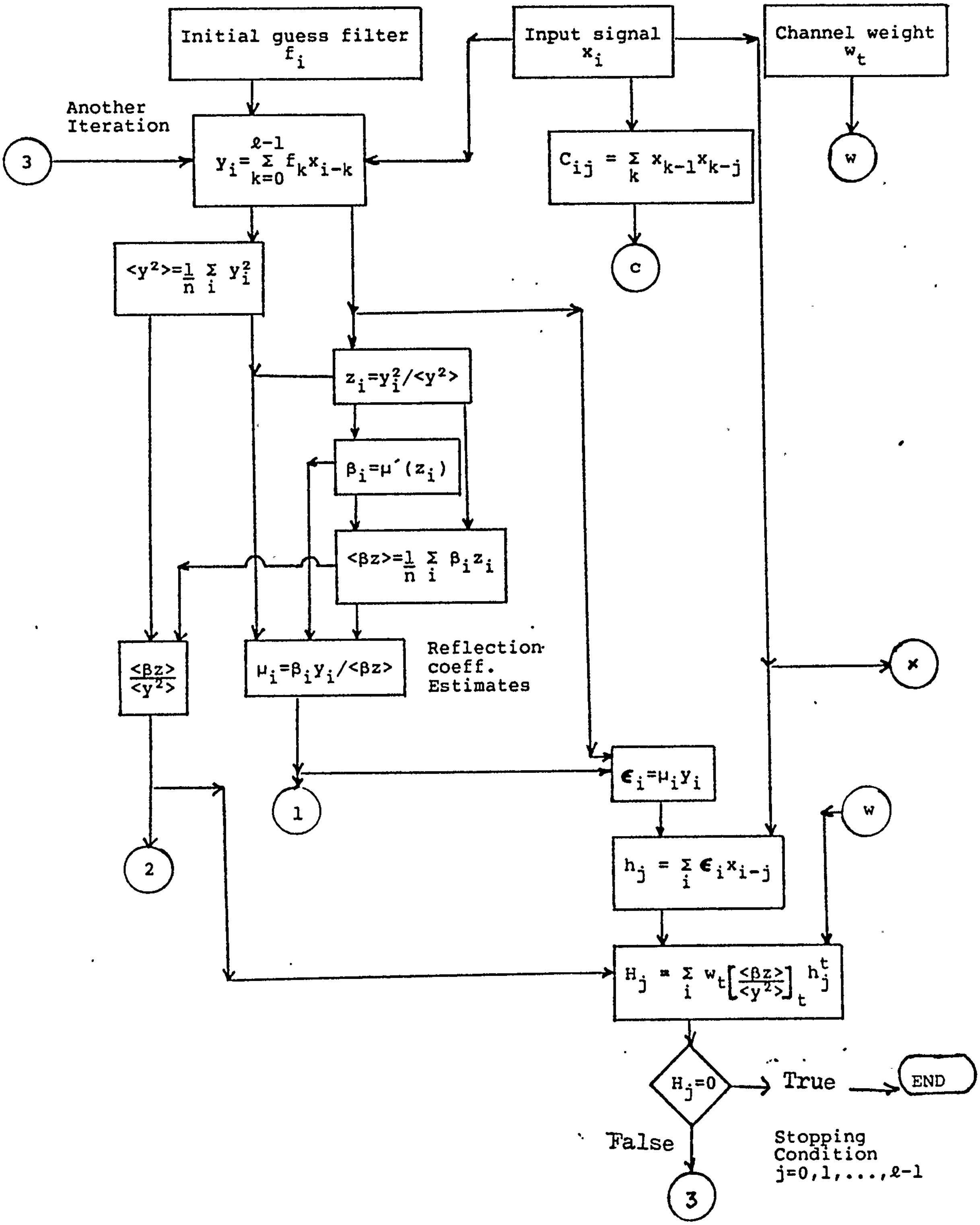
$$H_j = \sum_t w_t \left[\frac{\langle\beta z\rangle}{\langle y^2\rangle}\right]_t h_j^t$$

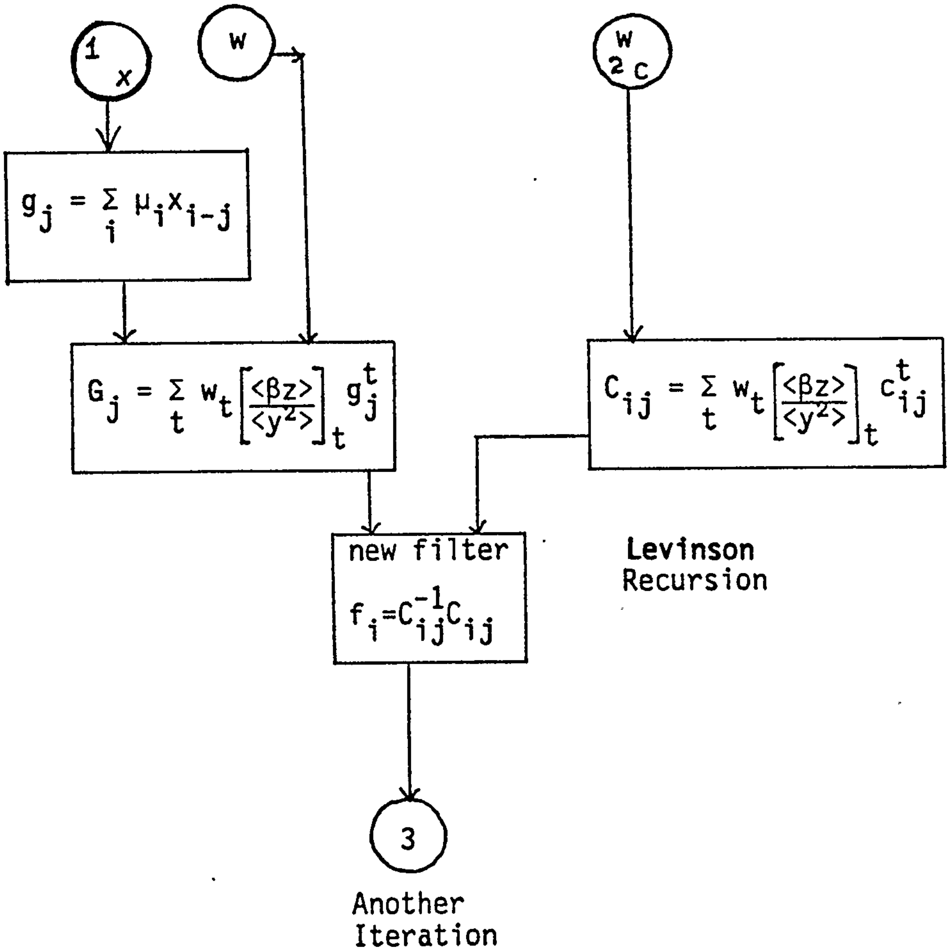
$$h_j = \sum_i \varepsilon_i x_{i-j}$$

$$\varepsilon_i = \mu_i - y_i \ .$$

Flow Chart for Decon Algorithm

The horizontal axis represents order to perform procedures. When two procedures appear at the same level, it means there is no apparent reason to perform one or the other first. Since the algorithm is iterative, it obviously wraps around.




1
x
w
w
2
c
$g_j = \sum_i \mu_i x_{i-j}$
$G_j = \sum_t w_t \left[\frac{\langle\beta z\rangle}{\langle y^2\rangle}\right]_t g_j^t$
$C_{ij} = \sum_t w_t \left[\frac{\langle\beta z\rangle}{\langle y^2\rangle}\right]_t c_{ij}^t$
new filter
$f_i = C_{ij}^{-1} C_{ij}$
Levinson
Recursion
3
Another
Iteration

APPENDIX J

PROOF OF $\underline{\beta}(t) = e^{-i\sigma}\,\beta(t)$

Given:

Real trace "$\alpha(t)$" and its analytic extension

$$\beta(t) = \alpha(t) + i\,H[\alpha(t)]$$

$$i = \sqrt{-1}$$

$H \equiv$ Hilbert transform .

Prove:

The constant phase-shifted version of "$\alpha(t)$", namely, "$\underline{\alpha(t)}$", is related by the following expression:

$$\underline{\beta}(t) = \underline{\alpha}(t) + i\,H[\underline{\alpha}(t)] = e^{-i\sigma}\,\beta(t)$$

$\sigma \equiv$ constant phase shift .

Proof:

Let "$A(w)$" be the Fourier transform of "$\alpha(t)$". Then by definition of Fourier transforms and Hilbert transforms, we have

$$\alpha(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} A(w)e^{-iwt}\, dw \tag{J.1}$$

and

$$H[\alpha(t)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{Sgn}(w)\ A(w)e^{-i[wt + \pi/2]}\, dw\ . \tag{J.2}$$

$$\mathrm{Sgn}(w) = \begin{cases} 1 : w > 0 \\ -1 : w < 0 \end{cases}$$

If a constant phase shift "σ" is added to "$\alpha(t)$", expressions (J.1) and (J.2) become:

$$\underline{\alpha(t)} = \frac{1}{2\pi} \int_{-\infty}^{\infty} A(w)e^{-i[wt + \sigma]}\, dw \tag{J.3}$$

and

$$H[\underline{\alpha}(t)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{Sgn}(w)\ A(w)e^{-i[wt + \sigma + \pi/2]}\, dw\ . \tag{J.4}$$

By direct comparison of (J.1) to (J.3) and (J.2) to (J.4), we have:

$$\underline{\alpha(t)} = e^{-i\sigma}\, \alpha(t) \text{ and } H[\underline{\alpha}(t)] = e^{-i\sigma} H[\alpha(t)]\ .$$

This, in turn, allows one to write:

$$\underline{\beta(t)} = \underline{\alpha(t)} + i\ H[\underline{\alpha}(t)] = e^{-i\sigma}\beta(t) \quad \square$$

APPENDIX K

PROOF OF $\underline{\phi}(t) = \phi(t) - \sigma$

Given:

$$\beta(t) = \alpha(t) + i\, H[\alpha(t)]$$

and

$$\underline{\beta(t)} = \underline{\alpha(t)} + i\, H[\underline{\alpha(t)}] \; ,$$

$$H \equiv \text{Hilbert transform}$$

$$i = \sqrt{-1} \; ,$$

where "$\underline{\beta(t)}$" is a constant phase-shifted version of "$\beta(t)$".

Prove:

$$\phi(t) = \underline{\phi(t)} - \sigma$$

where

$$\tan\phi(t) = \frac{\text{Im}[\beta(t)]}{\text{Re}[\beta(t)]} \tag{K.1}$$

and

$$\tan\underline{\phi(t)} = \frac{\text{Im}[\underline{\beta(t)}]}{\text{Re}[\underline{\beta(t)}]} \; . \tag{K.2}$$

Proof:

$$\tan\underline{\phi(t)} = \frac{\text{Im}[\underline{\beta(t)}]}{\text{Re}[\underline{\beta(t)}]} = \frac{H[\underline{\alpha(t)}]}{\underline{\alpha(t)}} \; .$$

Using the result from Appendix J, we can write:

$$\tan\underline{\phi(t)} = \frac{H[\alpha(t)]\cos\sigma - \alpha(t)\ \sin\sigma}{\alpha(t)\ \cos\sigma + H[\alpha(t)]\sin\sigma}$$

$$= \frac{\dfrac{H[\alpha(t)]}{\alpha(t)}\cos\sigma - \sin\sigma}{\cos\sigma + \dfrac{H[\alpha(t)]}{\sigma(t)}\sin\sigma}.$$

Substituting in relation (K.1), we get:

$$\tan\underline{\phi(t)} = \frac{\tan\phi(t)\ \cos\sigma - \sin\sigma}{\cos\sigma + \tan\phi(t)\ \sin\sigma}$$

$$= \frac{\sin\phi(t)\ \cos\sigma - \sin\sigma}{\cos\phi(t)\cos\sigma + \sin\phi(t)\sin\sigma}$$

$$= \frac{\sin(\phi-\sigma)}{\cos(\phi-\sigma)} = \tan(\phi-\sigma)$$

$$\Longrightarrow \qquad \underline{\phi(t)} = \phi(t) - \sigma\ . \quad \square$$

APPENDIX L
CONSTRUCTION OF PERFs BASED ON A PRIORI MOMENT INFORMATION

A technique will be introduced here which allows one to incorporate statistical information into a PERF. This information will be in the form of sample moments

$$M_R = \frac{1}{n} \sum_{i=1}^{n} x_i^k \,, \qquad \text{(L.1)}$$

where the set

$$\underline{X} = \{x_i : i=1,2,...n\}$$

is a given realization of the probabilistic part of the process. To simplify the discussion, we will require the elements of $\underline{X}$ to have the following property:

$$0 \leq x_i \leq 1 \,. \qquad \text{(L.2a)}$$

This ensures that

$$M_{R+\ell} \leq M_R \quad \forall\ \ell \geq 0 \,. \qquad \text{(L.2b)}$$

Property (L.2a) is not overly restrictive, since by simple normalization, we can ensure all elements to be less than unity and scaler independent. The forthcoming arguments which require the elements to be positive can be made for negative values independently with no ambiguity.

To begin, expand an arbitrary ruler function, ∀y>0, in terms of a series

$$\mu(y) = \frac{1}{n} \sum_{j=1}^{M} a_j \, y^j \, . \tag{L.3}$$

Incorporating (L.3) into a PERF "L" and using it to gauge a set

$$\underline{Y} = \{y_i \; : \; i=1,2,\ldots n\}$$

we get

$$L = \sum_{i=1}^{n} \mu(y_i) = \sum_{i=1}^{n} \frac{1}{n} \sum_{j=1}^{M} a_j \, y_i^j$$

$$= \sum_{j=1}^{M} a_j \quad \frac{1}{n} \sum_{i=1}^{n} y_i^j \quad . \tag{L.4}$$

But, the sample moments for the set $\underline{Y}$ are identically

$$\tilde{M}_j = \frac{1}{n} \sum_{i=1}^{n} y_i^j \, . \tag{L.5}$$

Substitute (L.5) into (L.4) to get

$$L = \sum_{j=1}^{M} a_j \, \tilde{M}_j \, . \tag{L.6}$$

If we let $a_j = M_j$, as defined in equation (L.1), and define the following vectors:

$$\tilde{M} = (\tilde{M}_1, \tilde{M}_2, \ldots, \tilde{M}_M) \qquad \text{(L.7a)}$$

and

$$M = (M_1, M_2, \ldots, M_M) \ , \qquad \text{(L.7b)}$$

then

$$L = M \quad \tilde{M} \ . \qquad \text{(L.8)}$$

This means that as Y approaches X, statistically, we would expect (L.8) to become a maximum. This allows us to use "L" as an indicator of this distribution in an inversion procedure.

LIST OF SYMBOLS

Symbol	Definition	Page of first occurrence
S	A process	5
S^{-1}	An inverse process	6
~	Over a symbol indicates approximation	7
<====>	if and only if	7
=====>	(which) implies	8
==><==	a contradiction	19
$\ni$	such that	9
$\in$	(is) an element of	15
$\exists$	there exists	18
$\propto$	is proportional to	22
$\tilde{\propto}$	is approximately proportional to	71
{ }	A set	36
{t:P}	Set of elements "t" with properties "P"	34
X,Y	Uppercase letters are vectors	15
$\underline{X},\underline{Y}$	Underscored or overscored uppercase letters are sets	14
x,y	Lowercase letters are real numbers	16
x_i,y_i	Indexed real numbers, as in sets or vectors	15
x	Cartesian product	14
L:A-->B	A mapping of "A" into "B"	14
$\|X\|_p$	ℓ^p norm	10

Symbol	Definition	Page of first occurrence
$L(X,Y)$	Regulating functions or PERFs	15
$\subset$	a subset of	79
μ	Ruler function	19
R^m	Real m-dimensional vector space	46
$\square$	Completion of proof	82
$\therefore$	therefore	82
$\forall$	for all	17
$<$	less than	17
$\leq$	less than or equal to	106
$>$	greater than	17
$\geq$	greater than or equal to	105
$<x>$	expected value of x	23