AN INFORMATION THEORY APPROACH TO DECONVOLUTION

Bob Godfrey

Abstract

The Kullback information measure is a scale invariant measure of the amount of information available for discrimination between two distributions. Using Pearson's system of probability density functions to parameterize a time series, the Kullback measure is used to derive a norm measuring the information content of a time series relative to a Gaussian distribution. Maximizing the norm drives the time series away from Gaussian, the only assumption being that the current distribution of the time series can be represented by a particular system of Pearson densities. The maximization procedure used was steepest descent, and the above norm is used to derive a gradient.

The generalized Gaussian (see Gray, p. 123, this report) and Pearson system both can be used to parameterize the probability density function of a time series. A brief comparison between the two was made and the generalized Gaussian was found to have lower RMS error and lower maximum error than the Pearson.

A general method of parameterizing probability density functions using quantiles is presented in the appendix. Although results are almost identical to those obtained using a Cauchy parameterization, the method has the potential of being able to parameterize multi-modal distributions.

Introduction

In 1977, Wiggins introduced a novel form of deconvolution called MED, in which a norm measuring the spikiness of a seismic trace was used to iteratively deconvolve it. Two other norms, parsimony and variable norm, proposed by Claerbout and Gray respectively, have adjustable coefficients which must be empirically chosen. The present norm was developed in an attempt to replace the arbitrary selection of norm by a sound theoretical definition of information.

Shannon (1948) defined entropy, S, as the amount of information contained in a discrete channel consisting of an information source, message, transmitter, etc. Many authors equate entropy with misinformation or randomness to avoid confusion with the common meaning of the term, information. In this paper, maximizing information is equivalent to minimizing entropy. If a reflectivity sequence x_t is observed and the entropy is to be computed, Shannon's definition of entropy for continuous channels is

$$S = -\int f(x) \ln f(x)\, dx \tag{1}$$

where $f(x)$ is the probability density function (pdf) that the data is presumed to come from. S as defined in Equation (1) *is not* scale invariant since $f(x)$ scales inversely as x. To overcome this deficiency, Kullback (see Hobson and Cheng, 1972) defined information gain K as

$$K = \int f(x) \ln [f(x)/h(x)]\, dx \tag{2}$$

K measures the information in the pdf $f(x)$ relative to the pdf $h(x)$. Hobson and Cheng show that K is a generalization of S, and that K gives a consistent theory of information. Equation (2) will be used to derive a gradient that forces the current pdf $f(x)$ away from a prior pdf $h(x)$. K is always greater than or equal to zero (Jensen's Inequality).

Derivation of gradient

Assume a current estimate of reflectivity is available and denote it by x_t^k. A new estimate x_t^{k+1} is desired and obtained by

$$x_t^{k+1} = x_t^k + dx_t \tag{3}$$

$$dx_i = \alpha \frac{dK}{dx_i}$$

α = step-size parameter

Define term $dK/dx_i = g_i$ (gradient).

To derive an expression for g_i, rewrite Equation (2) as

$$K = \int f(x) \ln f(x)\, dx - \int f(x) \ln h(x)\, dx$$

$$K = E[\ln f(X)] - E[\ln h(X)] \tag{4}$$

Now, substitute for expectation E:

$$E(Q) = \frac{1}{N} \sum_j q_j \tag{5}$$

where Q is a random variable and q_j is a realization of Q. Substituting (5) into (4):

$$K = \frac{1}{N} \left[\sum_j \ln f(x_j) - \sum_j \ln h(x_j) \right] \tag{6}$$

The gradient g_i is found by differentiating Equation (6) with respect to x_i:

$$g_i = \frac{1}{N} \left[\frac{\frac{df}{dx_i}}{f(x_i)} - \frac{\frac{dh}{dx_i}}{h(x_i)} \right] \tag{7}$$

The Pearson system of densities will be used to obtain an analytic expression for f. Pearson (see Elderton and Johnson, 1969) defined the following family of densities:

$$\frac{df}{dx} = \frac{f(x - a)}{b_0 + b_1 x + b_2 x^2} \tag{8}$$

The family has the properties that

(i) $\frac{df}{dx} \to 0$ as $f \to 0$

(ii) $\frac{df}{dx} = 0$ when $x = a$

Property (ii) forces f to be bell-shaped near $x = a$. If the density is assumed to be symmetric with mode zero, then Equation (8) specializes to

$$\frac{df}{dx} = \frac{f \cdot x}{b_0 + b_2 x^2} \tag{9}$$

Substitute (9) into (7):

$$g_i = \frac{1}{N}\left[\frac{x_i}{b_0 + b_2 x_i^2} - \frac{\frac{dh}{dx_i}}{h(x_i)}\right] \tag{10}$$

To determine the pdf $h(x)$, make g_i orthogonal to x_i:

$$\sum_i g_i x_i = 0 \tag{11}$$

Substituting (10) into (11) results in a differential equation for h:

$$\frac{\frac{dh}{dx_i}}{h(x_i)} = \frac{-|x_i|^{\gamma-1}\ \mathrm{sgn}(x_i)}{\sigma^{\gamma}} \tag{12a}$$

$$\sigma^{\gamma} = \frac{-\sum_i |x_i|^{\gamma}}{\sum_i \left(\frac{x_i^2}{b_0 + b_2 x_i^2}\right)} \tag{12b}$$

Equation (12a) can be integrated for any γ. In particular, choosing $\gamma = 2$ results in $h(x)$ having the form,

$$h(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\ \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

$$\sigma^2 = \frac{-\sum_i x_i^2}{\sum\left(\frac{x_i^2}{b_0 + b_2 x_i^2}\right)} \tag{13}$$

The prior pdf h is Gaussian. The current pdf is the integral of Equation (9):

$$f(x) = f_0 \left(1 + \frac{x^2}{a^2}\right)^{-m} \tag{14}$$

$$f_0 = \frac{1}{a\sqrt{\pi}} \frac{\Gamma(m)}{\Gamma(m-1)}$$

$$m = -\frac{1}{2b_2}, \quad a^2 = \frac{b_0}{b_2}$$

Equation (14) defines the Cauchy family of densities (a subset of the Pearson system), where m = shape and a = scale parameter. Forming ratios of the moments of the pdf f enables "m" to be computed. The scheme used for the examples in this paper was

$$\frac{\mu_1^2}{\mu_2} = \frac{2}{\pi} \frac{\Gamma^2(m-1)}{\Gamma^2(m-\frac{1}{2})} (m - \tfrac{3}{2}) \tag{15a}$$

$$\mu_2 = \frac{a^2}{2(m - \frac{3}{2})} \tag{15b}$$

where $$\mu_i \simeq \left[\frac{1}{N} \sum_j |x_j|^i\right]$$

Equation (15a) is implicitly solved for "m," which enables "a" to be computed from Equation (15b).

A brief look at Equation (13) is in order. Assume the time series is Gaussian, then Equation (15a) becomes

$$\frac{2}{\pi} = \frac{2}{\pi} \frac{\Gamma^2(m-1)}{\Gamma^2(m-\frac{1}{2})} (m - \tfrac{3}{2}) \tag{16}$$

Using the asymptotic expansion for the ratio of two gamma functions, Equation (16) has the solution: $m \to \infty$ or $b_2 \to 0$. Using this limit in Equation (15b) gives $b_0 = -\mu_2$. Finally, substituting these values into

Equation (13) gives $\sigma^2 = \mu_2$, an intuitive result. Furthermore, Equation (10) gives $g_i = 0$, so if the data is Gaussian, it remains Gaussian.

The expression for computing dx_i is

$$dx_i = \alpha g_i \tag{18a}$$

$$g_i = \frac{1}{N}\left[\frac{x_i}{b_0 + b_2 x_i^2} + \frac{x_i}{\sigma^2}\right] \tag{18b}$$

The denominator of the first term in the RHS of Equation (18b) contains a zero when $b_0/b_2 < 0$. From Equation (15b) this means $m < 3/2$. This forms a natural limit for the Cauchy density since $m = 3/2$ gives a scale equal to zero.

Multichannel algorithm

In the following discussion, capitals refer to frequency domain and lower case to time domain. Subscripts refer to channel and superscripts to iteration.

The convolutional model of a seismic trace is represented as

$$y_c = x_c * b \tag{19}$$

y_c = c-th channel of raw data

x_c = c-th channel of reflectivity

b = stationary wavelet

$*$ = convolution in time

In the frequency domain, (19) is

$$Y_c = X_c B \tag{20}$$

An equation for db, a change in wavelet, as a function of dx will now be derived. The derivation is similar tc that given in SEP-14 (Godfrey and Claerbout). Denote by X_c^K and B^K, the K-th estimate of reflectivity and wavelet respectively. Rewrite Equation (20) as

$$(Y_c + dY_c) = (X_c^K + dX_c^K)\ (B^K + dB^K) \tag{21}$$

Setting $dY_c = 0$ and dropping second order terms in (21) gives

$$X_c^K\ dB = -B^K\ dX_c^K \tag{22}$$

An expression for dX_c^K is found by transforming Equation (18a) into frequency domain:

$$dX_c^K = \alpha_c\ G_c^K \tag{23}$$

Using (23) in (22) gives

$$X_c^K \left(\frac{dB}{B}\right)^K = -\alpha_c\ G_c^K \tag{24}$$

There are several ways to proceed. The technique that gave good results using synthetics was to multiply (24) by $\overline{X}_c^K$ and sum over channel:

$$\left(\frac{dB}{B}\right)^K \sum_c (X_c\ \overline{X}_c)^K = -\sum_c \alpha_c (G_c\ \overline{X}_c)^K \tag{25}$$

In the time domain, the RHS is the crosscorrelation of g and x, whereas the LHS is the autocorrelation of x. The solution for $(dB/B)^K$ in the frequency domain is

$$\left(\frac{dB}{B}\right)^K = \frac{-\sum_c \alpha_c (G_c\ \overline{X}_c)^K}{\sum_c (X_c\ \overline{X}_c)^K + \varepsilon_1} \tag{26}$$

ε_1 = "prewhitening coefficient"

Using $(dB/B)^K$ as computed from (26), the new estimate of wavelet b is

$$B^{K+1} = B^K \left[1 + \left(\frac{dB}{B}\right)^K\right] \tag{27}$$

Finally, the new estimate of reflectivity is computed via

$$X_c^{K+1} = \frac{Y_c \, \overline{B}^{K+1}}{(B \, \overline{B})^{K+1} + \varepsilon_2} \qquad (28)$$

ε_2 = "prewhitening coefficient"

The sequence (26), (27), (28) is repeated until the shape parameter m has converged for each channel.

Notes

(1) The standard method of calculating $\varepsilon_1, \varepsilon_2$ was used, e.g.

$$\varepsilon_1 = 0.001 \text{ (zero-th lag of autocorrelation of } \sum_c X_c \, \overline{X}_c)$$

$$\varepsilon_2 = 0.001 \text{ (zero-th lag of autocorrelation of } B \, \overline{B})$$

(2) Equation (28) can be viewed as a three-stage process in the time domain (see Wood, et al, 1978). Rewrite (28) as

$$X_c = [(Y_c \, \overline{B})F \,]\overline{F} \qquad \text{where} \qquad F \, \overline{F} = \frac{1}{B \, \overline{B} + \varepsilon}$$

(i) Matched filtering of data by crosscorrelation with wavelet.

(ii) Filtering output of (i) with zero-delay Weiner inverse filter of wavelet.

(iii) Matched filtering of output of (ii) by crosscorrelation with the inverse wavelet of (ii).

It is necessary to pad zeroes onto y_c to ensure that step (i) is done correctly in the frequency domain.

(3) The algorithm has one arbitrary parameter, the step-size scalar α_c. This was computed by fixing the length of dx_c to be one-fifth the length of x_c:

$$\sqrt{|| \, dx_c \, ||_2} = \frac{1}{5}\sqrt{|| \, x_c \, ||_2}$$

or

$$\alpha_c = \frac{1}{5}\sqrt{\frac{|| \, x_c \, ||_2}{|| \, g_c \, ||_2}}$$

Examples

Figure 1 shows a group of five traces, before and after deconvolution. Twenty-one iterations were used, and Table 1 shows the values of "m" for each channel as iteration proceeds. Along with the trace (plotted as a function of time) is the gradient, plotted as a function of amplitude (x) and as a function of time. The g-x plots (middle column) illustrate that as iteration proceeds, the force with which small values are driven towards zero increases. This behavior was noted when using the parsimony norm, except there the norm coefficient had to be arbitrarily adjusted. A change in statistics is reflected in a change of "m" and "a" and hence, the present norm automatically adjusts. Even though small values are driven toward zero, a precursor persists on all the deconvolved traces. It can be eliminated by constraining the wavelet to be one-sided (no anticipation component), but the deconvolved traces are less spiky. In this example, each trace was 512 points long, and was padded with 512 zeroes to yield a 1024-point trace. The initial wavelet was assumed to be a spike at $t = 0$, and subsequent wavelets were free to contain 1024 points. No constraints were imposed on the wavelet other than to zero the dc and Nyquist frequency.

An assumption made during the derivation of the algorithm was that the Cauchy family adequately describes the pdf of the data. Figure 2 contains two plots: the top figure shows the actual cumulative distribution function cdf (continuous curve that changes in jumps) of a trace from Figure 1 vs. the cdf of a Cauchy (dotted curve); the bottom plot shows the actual cdf vs. the cdf for a generalized Gaussian (see Gray, this report). On both plots, the middle curve is the error between actual and theoretical curve. Although both Cauchy and generalized Gaussian fit well, the latter was found to have lower RMS error and lower maximum error.

The final example, Figure 3, is a test prepared by Gray. Figure 3a contains eight groups of six channels, where each group has undergone a different exponential gain. Figure 3b shows a good deconvolution of the first four groups, a single spike deconvolution of the fifth group, and a noisy deconvolution of the remaining three groups. The average amplitude spectrum of groups 4 and 5 before and after deconvolution is plotted in Figures 3d and 3c respectively.

Before exponential gain was applied to groups 5 through 8, the S/N ratio was probably small for times greater than one second. Hence, exponential gain would have blown up the noise in this time range. The bottom, right-hand corner of Figure 3a appears very noisy with respect to the rest of the plot. This observation is also reflected in that the "m" values for groups 1 through 4 ranged from 3.3 to 5.5, whereas the range for groups 5 through 8 was 3.3 to >10. So the deconvolution of groups 5 through 8 may be influenced more by noise than by signal.

A subroutine to compute g_t as given by Equation (18b) follows Figure 3.

Conclusion

The Kullback information measure, measures the information contained in a time series relative to a prior distribution. Using the Pearson system of densities to parameterize the current distribution of a time series and selecting a Gaussian prior distribution, an algorithm that forced the distribution of the time series to be maximally far from Gaussian was developed. The particular family of Pearson's system used was Cauchy, and a comparison was made between Cauchy and Gray's generalized Gaussian in order to determine which gave the best fit for seismic data. A few tests indicate the generalized Gaussian has lower RMS and lower maximum error than Cauchy.

The Kullback measure, however, can be used for any parameterization of data. A general algorithm that models any data distribution is presented in the appendix. This algorithm is more expensive than the previous one; however, it has the potential of being able to parameterize multi-modal distributions, whereas two-parameter families are intrinsically uni-modal. Currently, some impedance logs are being investigated to test for multi-modality.

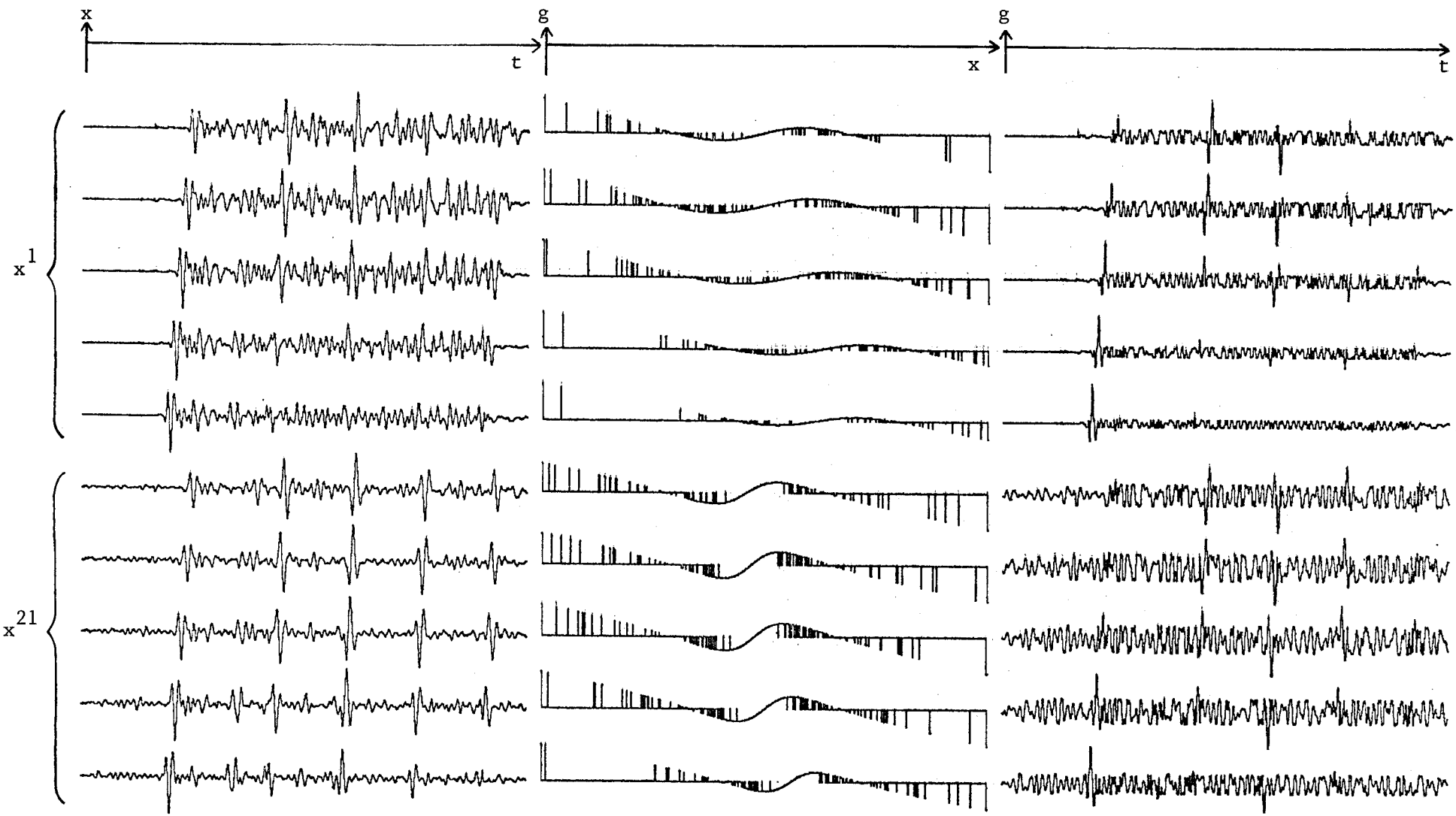


FIGURE 1.--Top five traces are raw data, bottom five traces are deconvolved. Middle traces are gradient-plotted as function of x (amplitude). Right-hand traces are gradient-plotted as a function of time.

igth= 1
b set to spike

iter	value of m for channel 1	2	3	4	5
1	2.665	2.828	2.955	2.650	2.885
2	2.290	2.384	2.471	2.317	2.350
3	2.115	2.168	2.249	2.162	2.126
4	2.028	2.054	2.114	2.065	2.031
5	1.982	1.989	2.030	2.006	1.975
6	1.958	1.946	1.979	1.966	1.940
7	1.943	1.916	1.945	1.937	1.917
8	1.936	1.896	1.919	1.914	1.902
9	1.933	1.881	1.902	1.898	1.891
10	1.932	1.871	1.888	1.886	1.883
11	1.932	1.864	1.879	1.879	1.877
12	1.931	1.859	1.871	1.873	1.872
13	1.932	1.854	1.866	1.869	1.868
14	1.931	1.851	1.861	1.866	1.866
15	1.931	1.847	1.857	1.864	1.864
16	1.930	1.844	1.853	1.863	1.862
17	1.929	1.842	1.851	1.862	1.861
18	1.928	1.840	1.848	1.861	1.859
19	1.927	1.838	1.846	1.860	1.858
20	1.927	1.837	1.844	1.860	1.857
21	1.926	1.835	1.843	1.859	1.855

EXP METHOD DONE

TABLE 1.--Value of "m" is tabulated for each channel as iteration proceeds.

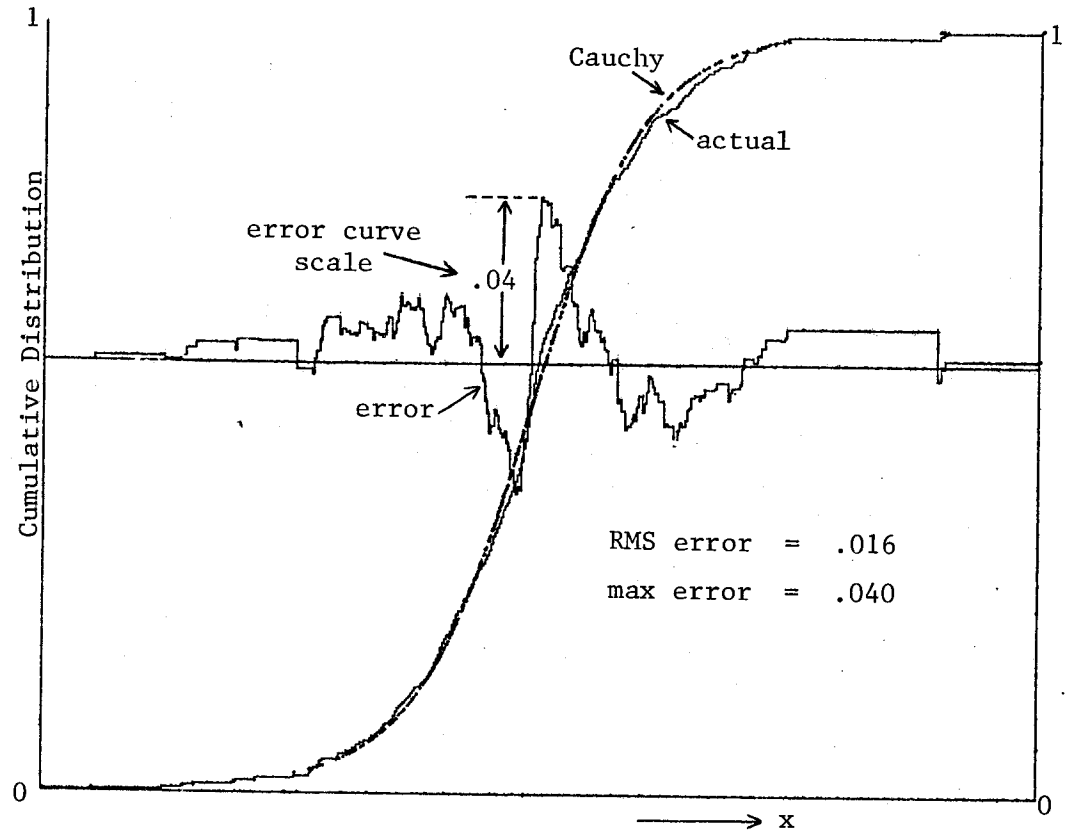
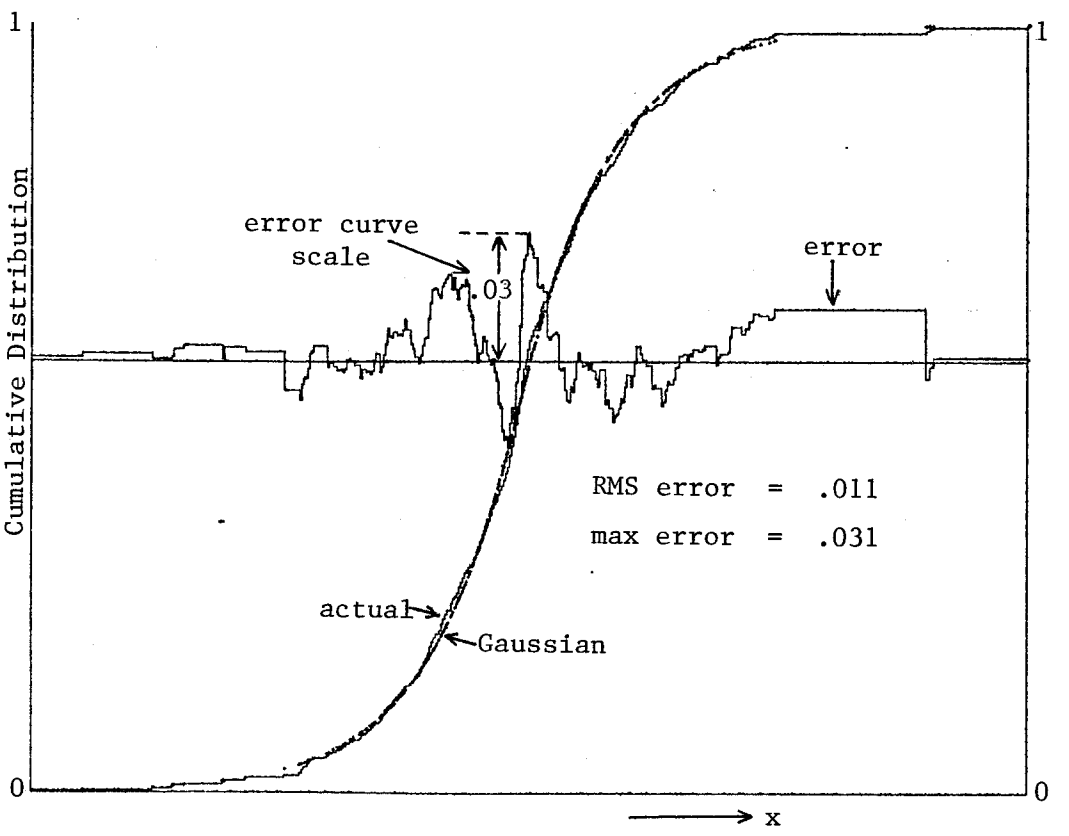


FIGURE 2.--Top figure shows Cauchy distribution vs. actual distribution, and bottom figure shows Gaussian distribution vs. actual. The middle curve in both cases is error between actual and theoretical curves.



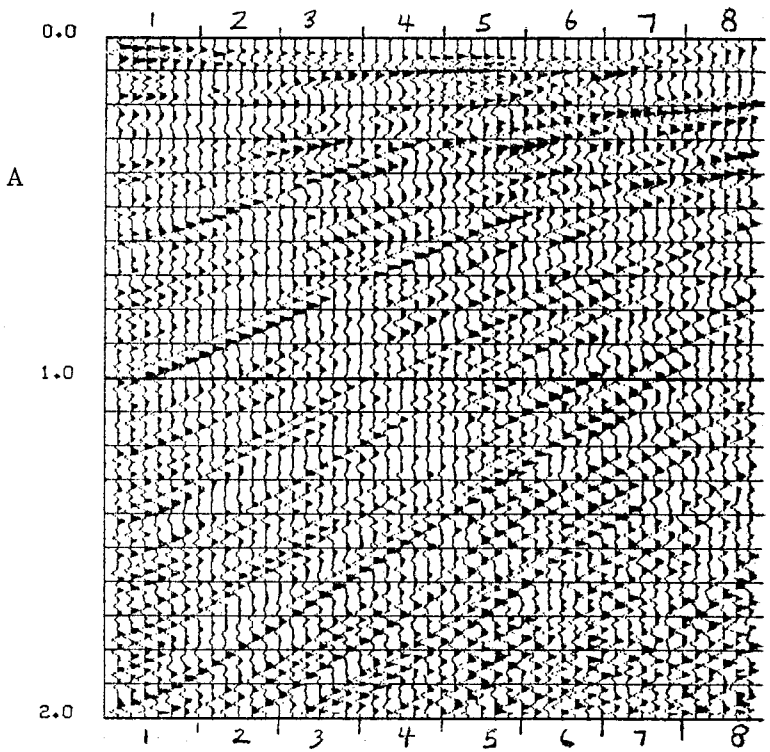


FIGURE 3a.--Raw data consisting of eight groups of six channels, each group having a different exponential gain.

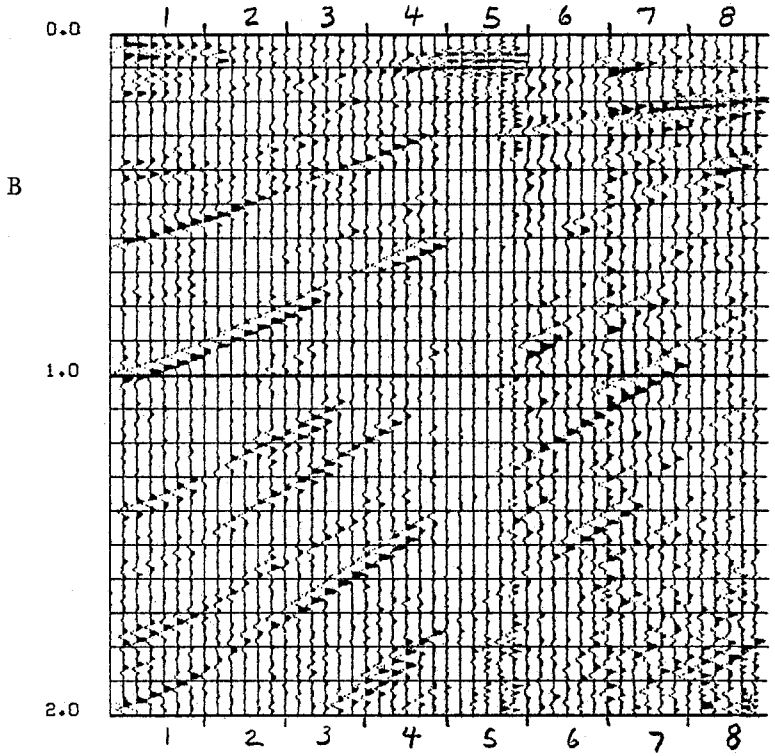
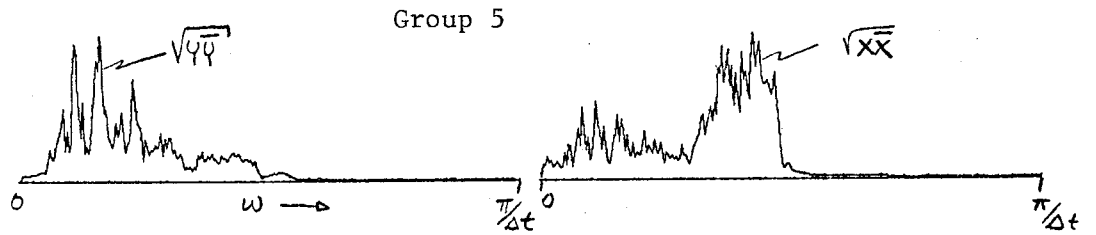
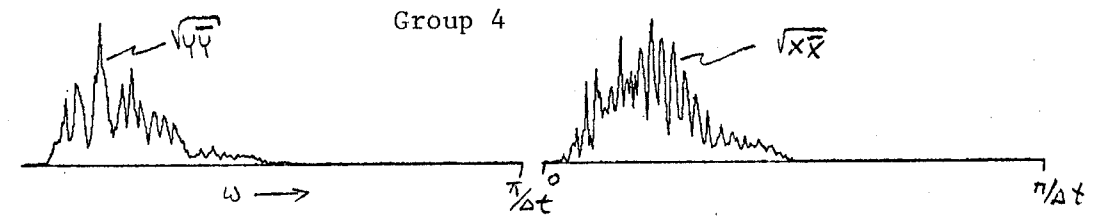


FIGURE 3b.--Deconvolved data. The average amplitude spectrum of groups 4 and 5 are shown below in Figures 3d and 3c.

FIGURES 3c,3d.--The six channels of each group were summed before and after deconvolution to give $\sqrt{Y\bar{Y}}$ and $\sqrt{X\bar{X}}$ Group 5 is obviously whiter than Group 4.





```
      subroutine gradmp(x,g,a,nt,nd,ist,rm,a2,a0,an,n,gam)
c
c     Input: x(1...nt),a,nt,nd,ist,a0,an,n,gam
c     Output: g(1...nt),rm,a2
c     x=data vector(nt points)
c     g=gradient vector(nt points)
c     nd=number of points used in computing u1,u2
c     ist=first data point of x used in computing u1,u2
c     a=vector needed to form table for t(=u1*u1/u2)-m graph.
c     rm=shape parameter "m"-returned from sub'r ESTM
c     a2=scale parameter
c     a0,an= first and last point for absicca of t-m plot.
c     n=number of points computed for absicca of t-m plot.
c     gam=2. ==>prior density is gaussian.
c           ==>other values are possible but in practice only gam=2.
c              was used
c
      real x(1),a(1),g(1)
      rnd=nd
      call mu1(x(ist),u1,u2,nd)
      t=u1*u1/u2
      call estm(a,n,a0,an,rm,t)
      a2=2.*u2*(rm-1.5)
      b2=-1./(2.*rm)
      b0=a2*b2
      sum1=0.
      sum2=0.
      do 20 i=1,nt
      ax=abs(x(i))
      ax2=ax*ax
      sum1=sum1+ax**gam
20    sum2=sum2+(ax2/(b0+b2*ax2))
      gam1=gam-1.
      top=sum2/sum1
      do 30 i=1,nt
      ax=abs(x(i))
      term=x(i)/(b0+b2*(x(i)*x(i)))
30    g(i)=term-((ax**gam1)*(sgn(x(i)))*top)
      return
      end
      subroutine estm(a,n,a0,an,rm,t)
      real a(1)
      data ib/0/
      rn=n
      pi=3.141592
      if(ib.ne.0)go to 20
      ib=1
      con=(an-a0)/(rn-1.)
      do 10 i=1,n
      rm=a0+con*float(i-1)
      t1=gamma(rm-1.)
      t2=gamma(rm-0.5)
      t3=rm-1.5
10    a(i)=(2.*t1*t1*t3)/(pi*t2*t2)
20    continue
      k=n
```

```
      do 30 i=1,n
      if(t.gt.a(i))go to 30
      k=i
      go to 40
30    continue
40    rma=a0+(con*float(k-2))
      rmb=((t-a(k-1))/(a(k)-a(k-1)))*con
      rm=rma+rmb
      return
      end
      subroutine mu1(x,u1,u2,n)
      real x(1)
      double precision du1,du2,ax
      rn=n
      du1=0.
      du2=0.
      do 10 i=1,n
      ax=abs(x(i))
      du1=du1+ax
      du2=du2+ax*ax
10    continue
      du1=du1/rn
      du2=du2/rn
      u1=du1
      u2=du2
      return
      end
```

APPENDIX

A general method of parameterizing probability density functions

Rather than parameterizing pdfs by two-parameter Cauchy or generalized Gaussian families, a general multi-parameter technique that is valid irrespective of the type of distribution from which the sample is taken, is derived. It is conjectured that the technique will be particularly useful when multi-modal data are encountered.

The multichannel algorithm outlined previously needs a subroutine for computing g_t, given x_t. Essentially, Equation (18b) is the appropriate subroutine when the pdf is parameterized by a Cauchy. The derivation of a new gradient, when the pdf is characterized by many parameters, is the subject of the remaining part of the paper.

First, rewrite Equation (2):

$$K = \int f(x) \ln [f(x)] \, dx - \int f(x) \ln [h(x)] \, dx \tag{A1}$$

A discrete representation of the last term in Equation (A1) is readily obtained if the prior pdf $h(x)$ is assumed to be uniform:

$$\int f(x) \ln [h(x)] \, dx = \ln h = -\ln [x_{(max)} - x_{(min)}] \tag{A2}$$

The first term on the RHS of Equation (A1) can be discretized as follows. Denote an estimate of $f(x)$ by $\hat{f}(x)$. Approximate $\hat{f}(x)$ by a rectangular function $\hat{\hat{f}}(x)$ and denote the x coordinates of the rectangles by $x_{(q)}$. Require that all rectangles have equal areas. Figure A1 illustrates the relation between $\hat{f}(x)$ and $\hat{\hat{f}}(x)$.

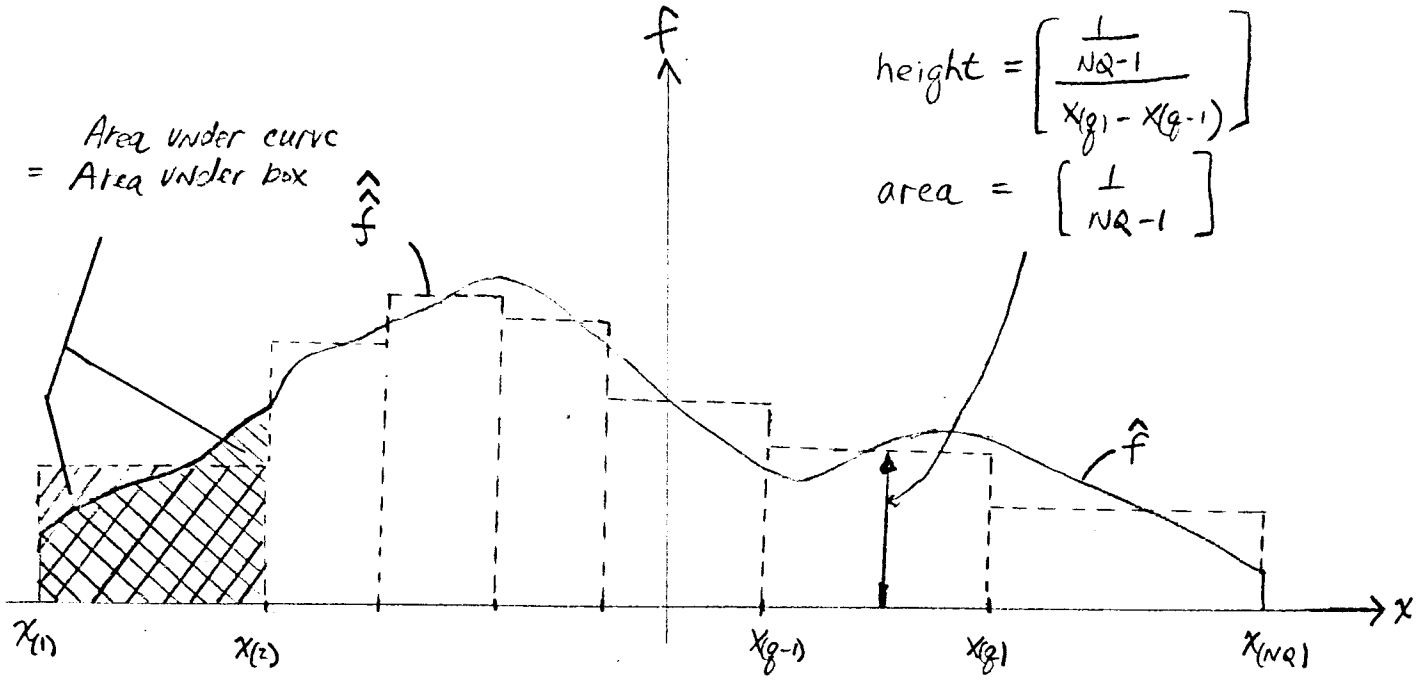


FIGURE A1.--The rectangular approximation $\hat{\hat{f}}(x)$ fitted to the estimated density $\hat{f}(x)$. NQ = number of quantiles used to discretize the x-axis.

From Figure A1, it follows that

$$\int f(x) \ln [f(x)] \, dx = \int \hat{f}(x) \ln [\hat{f}(x)] \, dx$$

$$= \frac{1}{NQ - 1} \sum_{p=2}^{NQ} \ln \left[\frac{\frac{1}{NQ - 1}}{x_{(p)} - x_{(p-1)}} \right] \qquad \text{(A3)}$$

Substituting (A3) and (A2) into Equation (A1) gives the Kullback measure as a function of $x_{(q)}$:

$$K = \frac{1}{NQ - 1} \left\{ \sum_{p=2}^{NQ} \ln \left[\frac{\frac{1}{NQ - 1}}{x_{(p)} - x_{(p-1)}} \right] \right\} + \ln [x_{(NQ)} - x_{(1)}] \qquad \text{(A4)}$$

where

$$x_{(max)} = x_{(NQ)}$$

$$x_{(min)} = x_{(1)}$$

As computed by Equation (A4), K measures how far the current pdf is from being uniform. Maximizing K drives the data away from uniformity. Assuming that reasonable estimates of $x_{(q)}$ can be found, it is straightforward to derive an expression for $g_{(q)}$:

$$g_{(q)} = \frac{\partial K}{\partial x_{(q)}} = \frac{-1}{NQ - 1}\left[\frac{1}{x_{(q)} - x_{(q-1)}} - \frac{1}{x_{(q+1)} - x_{(q)}}\right] + \frac{1}{x_{(NQ)} - x_{(1)}}\left[\delta_{q,NQ} - \delta_{q,1}\right] \quad \text{(A5)}$$

$$\delta_{i,j} = 1 \quad \text{if and only if} \quad i=j$$

$$= 0 \quad \text{otherwise}$$

Equation (A5) gives the gradient only at x coordinates given by $x_{(q)}$. To obtain g_t, the linear interpolation scheme outlined in Figure A2 is used.

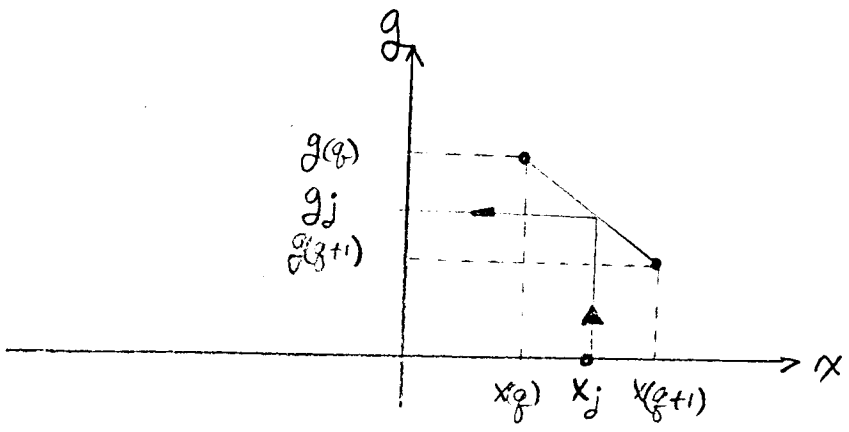


FIGURE A2.--Given x_t. Assume $[x_{(q)}, g_{(q)}]$ have already been computed. To compute g_j, first find x_j then perform linear interpolation to get g_j.

The subroutine for computing g_t contains three steps:

(i) Compute $x_{(q)}, q = 1,NQ.$

(ii) Compute $g_{(q)}$ from $x_{(q)}$ by Equation (A5).

(iii) Compute g_t by Figure A2.

Computation of $x_{(q)}$

Recall that $x_{(q)}$ are the x-coordinates of equal area rectangles used to approximate the estimated pdf, $\hat{f}(x)$. Wegman (1972) reviews techniques used in the statistical estimation of density functions. An obvious, sub-optimal, estimate of $f(x)$ is

$$\hat{f}(x) = \frac{F_N(x+h) - F_N(x-h)}{2h} \tag{A6}$$

where $F_N(u)$ is the empirical cumulative distribution function. The parameter "h" controls how smooth $\hat{f}(x)$ is. Assuming some value of h, Equation (A6) allows $\hat{f}(x)$ to be constructed. Finally, $x_{(q)}$, $q = 1,NQ$ are determined by solving

$$\frac{q - 1}{NQ - 1} = \int_{x_{(1)}}^{x_{(q)}} \hat{f}(s)\, ds, \quad \text{where} \quad q = 1,2,...NQ \tag{A7}$$

Equation (A7) selects NQ quantiles from data having a pdf $\hat{f}(x)$. Since $\hat{f}(x)$ is known only for a finite number of x values, some sort of integration rule must be used to compute the integration in (A7). In practice, the trapezoidal rule was implemented.

The optimal mean integrated squared error value for h, corresponding to Equation (A6), can be derived by solving:

$$\min_h \left\{ E \int [\hat{f}(x) - f(x)]^2 \, dx \right\} \tag{A8}$$

The solution (Wegman outlines a derivation) of Equation (A8) is

$$h^5 = \frac{9}{2N \int_{-\infty}^{\infty} [f''(x)]^2 \, dx}$$

The appearance of the true population density is inevitable. Different choices of $f(x)$ result in $\hat{f}(x)$ attaining differing degrees of smoothness. An estimate of the optimal h is obtained by selecting some particular $f(x)$. Choosing $f(x)$ to be Gaussian gives

$$h = \left[\frac{9\sqrt{\pi}}{2N}\right]^{1/5} \sqrt{\mu_2}$$

This value was used in Equation (A6) in order to compute $f(x)$.

Examples

Figure A3 shows the same input data as used in Figure 1, deconvolved using the quantile method. The number of quantiles, NQ, used was 20. The output after 11 and 21 iterations is given in Figure A3 and Table 1 (below) tabulates the Kullback measure, K, [Equation (A4)] at each iteration. Also tabulated is KS = K summed over the five channels. After 11 iterations, KS has increased 48 percent over its initial value. An additional 11 iterations increases KS only two percent over the iter = 11 value. This is reflected in Figure A3: the x^{11} frame is almost identical to the x^{21} frame. Even though KS is converging throughout 21 iterations, Table A1 shows that channels 4 and 5 start to diverge after 11 iterations.

igth= 1

b set to spike

iter	Kullback measure 1	2	3	4	5	KS		Cauchy shape parameter 1	2	3	4	5
1	0.750	0.629	0.662	0.857	1.002	3.900	m=	2.663	2.827	2.960	2.660	2.859
2	0.872	0.763	0.767	0.964	1.135	4.501	m=	2.329	2.450	2.488	2.388	2.406
3	0.968	0.893	0.840	1.044	1.242	4.987	m=	2.140	2.234	2.266	2.254	2.169
4	1.018	0.970	0.876	1.095	1.300	5.259	m=	2.045	2.110	2.138	2.131	2.075
5	1.048	1.014	0.907	1.121	1.348	5.438	m=	1.992	2.034	2.058	2.071	2.014
6	1.067	1.041	0.942	1.136	1.379	5.563	m=	1.963	1.982	1.999	2.021	1.977
7	1.075	1.055	0.965	1.144	1.396	5.636	m=	1.942	1.948	1.964	1.983	1.950
8	1.078	1.064	0.980	1.147	1.401	5.669	m=	1.932	1.922	1.932	1.954	1.930
9	1.079	1.081	0.997	1.150	1.404	5.710	m=	1.928	1.904	1.912	1.932	1.916
10	1.080	1.096	1.015	1.151	1.403	5.745	m=	1.925	1.890	1.897	1.918	1.908
11	1.082	1.110	1.035	1.149	1.398	5.775	m=	1.924	1.880	1.883	1.907	1.902
12	1.084	1.120	1.049	1.150	1.396	5.799	m=	1.924	1.870	1.872	1.899	1.898
13	1.087	1.127	1.061	1.151	1.393	5.819	m=	1.925	1.865	1.863	1.892	1.897
14	1.089	1.134	1.073	1.147	1.390	5.833	m=	1.927	1.860	1.857	1.884	1.897
15	1.089	1.140	1.079	1.146	1.388	5.842	m=	1.930	1.856	1.850	1.880	1.894
16	1.089	1.148	1.087	1.147	1.380	5.851	m=	1.933	1.853	1.844	1.876	1.892
17	1.087	1.152	1.095	1.147	1.376	5.857	m=	1.936	1.851	1.839	1.872	1.891
18	1.087	1.156	1.106	1.144	1.372	5.865	m=	1.939	1.848	1.835	1.870	1.892
19	1.086	1.160	1.113	1.143	1.368	5.870	m=	1.942	1.845	1.830	1.869	1.893
20	1.086	1.164	1.119	1.143	1.363	5.875	m=	1.944	1.842	1.827	1.868	1.893
21	1.087	1.168	1.124	1.143	1.354	5.876	m=	1.945	1.840	1.825	1.867	1.893

EXP METHOD DONE

TABLE A1.--Kullback measure (K) tabulated as iteration proceeds. The sixth column, KS, is the sum of K for the five channels. The last five columns show the Cauchy shape parameter for each channel. Note that whereas KS increases continually, K for channels 4 and 5 decrease after 13 and 9 iterations respectively.

To investigate the effect of changing NQ, a number of deconvolutions were performed with NQ ranging between 10 and 20. The results were almost identical. This is encouraging; otherwise NQ would be a critical parameter in the algorithm. It means that the approximate density, $\hat{\hat{f}}(x)$, does not change appreciably as a function of NQ. To illustrate the point, Figure A4a is a plot of $\hat{\hat{f}}(x)$, using the first channel of Figure A3, before deconvolution where NQ = 20. For comparison, Figure A4c used NQ = 10. Both plots were concatenated together to give Figure A4d, and the only mismatch occurs in the tails. As an aside, Figure A4b shows $\hat{\hat{f}}(x)$ after deconvolution where NQ = 20. The spikier appearance of Figure A4b vs. Figure A4a is as expected.

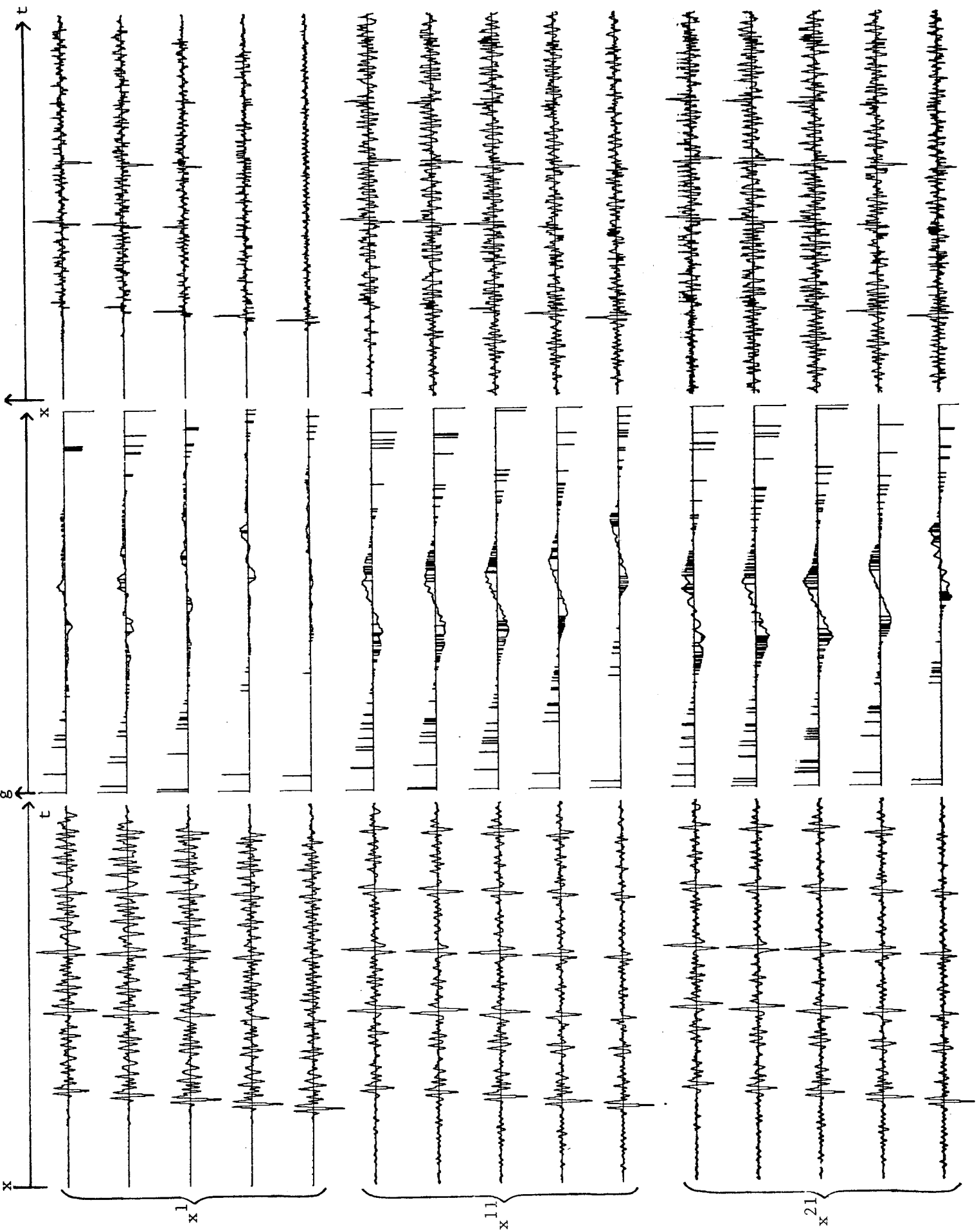


FIGURE A3.--Same dataset as in Figure 1, deconvolved using quantile method. NQ = 20.

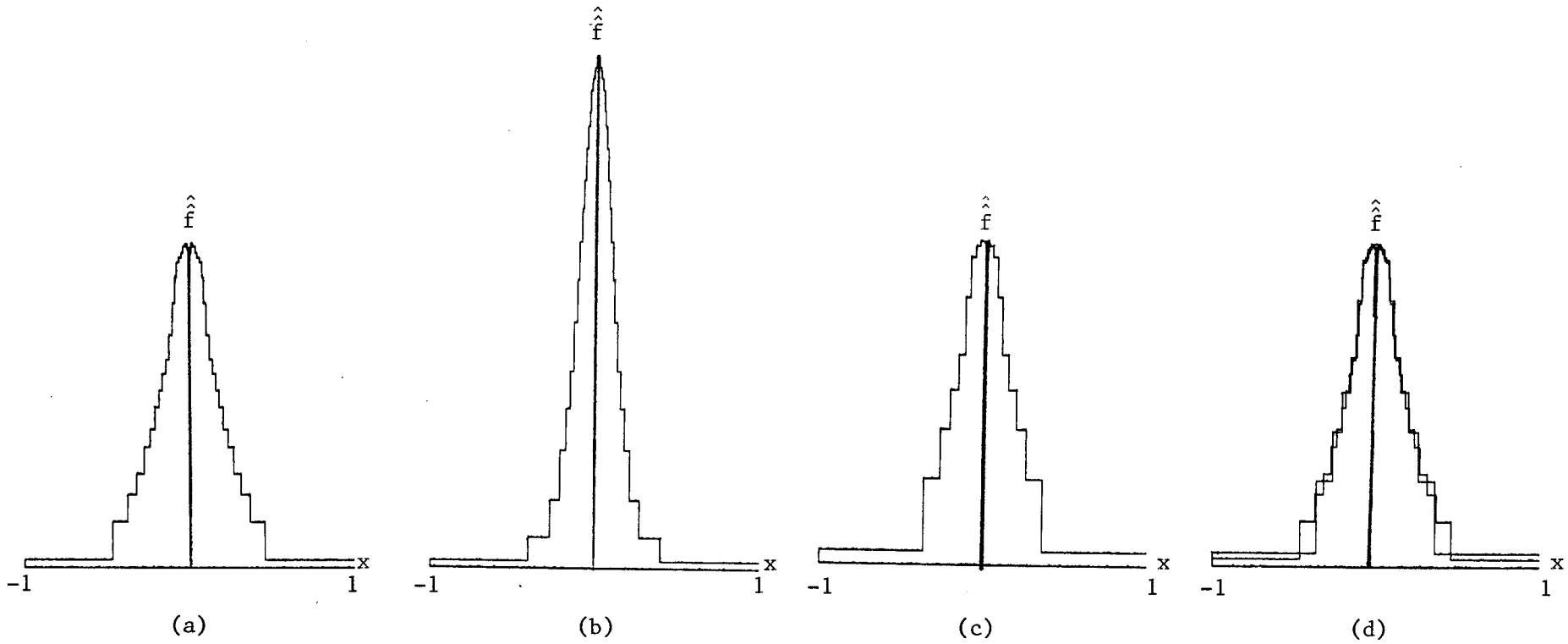


FIGURE A4.--(a) For the data of Figure A3, $\hat{\hat{f}}(x)$, before deconvolution, is plotted with NQ = 20.

(b) Same as (a) except after deconvolution.

(c) To investigate the dependence of $\hat{\hat{f}}(x)$ on NQ, $\hat{\hat{f}}(x)$, before deconvolution, is plotted with NQ = 10.

(d) Comparison between (a) and (c). The agreement between the two curves is an indication that NQ is not a critical parameter in the algorithm.

Note: For this dataset, $\hat{\hat{f}}(x)$ is assumed to be symmetric. Only one side was computed and then was folded to give a symmetric function. Figure (b) is scaled to have the same range $[x_{(max)} - x_{(min)}]$ as Figure (a).

REFERENCES

ELDERTON, W. P. and N. L. JOHNSON (1969), *Systems of Frequency Curves* (Cambridge University Press).

HOBSON, A. and B. K. CHENG (1973), "A comparison of the Shannon and Kullback information measures," *J. Stat. Phys.*, 7, 301-310.

SHANNON, C. E. and W. W. WEAVER (1949), *The Mathematical Theory of Communication* (University of Illinois Press).

WOOD, L. C., R. C. HEISER, S. TREITEL and P. L. RILEY (1978), "The debubbling of marine source signatures," *Geophysics*, 43.

WEGMAN, E. J. (1972), "Nonparametric probability density estimation: I. A summary of available methods," *Technometrics*, 14, pp. 533-546.