References

Barrett, B.F. and Lampard, D.G., 1955, An expansion for some second-order probability distributions and its application to noise problems: IRE Transactions-IT, March.

Box, G.E.P. and Jenkins, G.M., 1970, Time series analysis, forecasting and control, Holden-Day.

Claerbout, J.F. and Muir, F., 1973, Robust modelling with erratic data: Geophysics, v. 38, p. 826-844.

Forney, G.D. Jr., 1973, The Viterbi Algorithm: Proc. IEEE, v. 61, p. 268-276.

Gray, W., 1979, Variable norm deconvolution, Ph.D. dissertation: Dept. of Geophysics, Stanford University.

Kailath, T., 1976, Lectures on linear least-squares-estimation, Springer-Verlag.

Kemeny, J.G. and Snell, J.L., 1960, Finite markov chains, Van Nostrand Co., Inc.

Krumbein, W.C. and Dacey, M.F., 1969, Markov chains and embedded Markov chains in geology: Jour. Math. Geology, vol. 1, no. 1, p. 79-96.

Merriam, D.F., 1976, Random processes in geology, Springer-Verlag.

Miller, J.H. and Thomas, J.B., 1976, The detection of signals in impulsive noise modelled as a mixture process: IEEE Transactions in Communications Theory, v. 24, p. 559-563.

Scargle, J.D., 1977, Absolute value optimization to estimate phase properties of stochastic time series: IEEE Transactions in Information Theory, January, p. 140-143.

Schafer, R.W. and Markel, J.D., 1979, Speech analysis, IEEE Press.

Schwarzacher, W., 1972, The semi-Markov process on a general sedimentation model, Mathematical models in sedimentary processes: New York, Plenum Press, p. 247-268.

Taylor, H.L., Banks, S.C. and McCoy, J.F., 1979, Deconvolution with the L-1 norm: Geophysics, v. 44, p. 39-53.

Van Trees, H.L., 1968, Detection, estimation, and modulation theory, v. 1, John Wiley and Sons.

Wiggins, R.A., 1977, Minimum entropy deconvolution: Proceedings of the International Symposium on Computer Aided Seismic Analysis and Discrimination, June 9-10, Falmouth, Mass., IEEE Computer Society.

Appendix A
The Minimum Mean Square Error Estimator

The objective of this appendix is to prove that the estimator $\tilde{r}_2(x)$ that minimizes the mean square error, $E[R - \tilde{r}_2(X)]^2$, is identical to the conditional mean, $E[R|X=x]$. The analysis is specialized to the case of estimating R given scalar observations. The extension to vector observations is trivial. We follow Kalaith's (1976) derivation. Formally, the mean square errror (MSE) is defined in terms of the joint probability density function between the observation X and the quantity to be estimated R:

$$MSE \triangleq E\left[R - \tilde{r}_2(X)\right]^2$$

$$= \iint \left[r - \tilde{r}_2(x)\right]^2 f_{R,X}(r,x)\, dx\, dr \tag{A1}$$

Using the identity

$$f_{R,X}(r,x) = f_{R|X}(r|x)\, f_X(x)$$

in equation (A1) gives

$$MSE = \int f_X(x)\, dx \int \left[r - \tilde{r}_2(x)\right]^2 f_{R|X}(r|x)\, dr \tag{A2}$$

Because $f_X(x)dx$ represents a probability that is inherently positive, the MSE is minimized by minimizing ϵ, defined by

$$\epsilon \triangleq \int \left[r - \tilde{r}_2(x)\right]^2 f_{R|X}(r|x)\, dr$$

for every value of x. The result of this minimization can be obtained by realizing that ϵ is a moment of inertia, and hence $\tilde{r}_2(x)$ is simply the centroid:

$$\tilde{r}_2(x) = \int r f_{R|X}(r|x)\, dr = E\left[R|X=x\right] \qquad \text{(A3)}$$

An alternative way of obtaining equation (A3) is by setting $\partial\epsilon/\partial\tilde{r}_2 = 0$, and solving the resulting equation for $\tilde{r}_2(x)$.

Appendix B

Properties of the Conditional Mean

In this appendix, we show that the non-linearity $\tilde{r}_1(x)$ resulting from solving equation (2.6) in the text can be approximated by a threshold device:

$$\tilde{r}_1(x) = 0 \qquad |x| < x_c$$

$$= \frac{\sigma_r^2}{\sigma^2} x \qquad |x| \geq x_c$$

where x_c is the device threshold. This non-linearity was implemented and results were virtually identical to those obtained using $\tilde{r}_2(x)$.

To analyze the shape of $\tilde{r}_1(x)$, equation (2.6) must be interpreted correctly:

$$\int_{-\infty}^{\tilde{r}_1(x)} f_{R|X}(r|x)\, dr = 0.5$$

Using the pdfs for R and N as given by equations (2.7b) and (2.12b), the unique nature of the above equation becomes apparent:

$$\lambda \int_{-\infty}^{\tilde{r}_1(x)} \delta(r) f_N(x-r)\, dr +$$

$$(1-\lambda) \int_{-\infty}^{\tilde{r}_1(x)} G(\sigma_r^2, r^2) f_N(x-r)\, dr = \frac{1}{2} f_X(x) \qquad \text{(B1)}$$

If we assume $\tilde{r}_1(x) >> 0$, the first integral in equation (B1) causes no difficulty, and the delta function sifts out the value at r=0, i.e.

$$\lambda f_N(x) + (1-\lambda) \int_{-\infty}^{\tilde{r}_1(x)} G(\sigma_r^2, r^2) f_N(x-r)\, dr = \frac{1}{2} f_X(x)$$

It is straightforward to simplify the above equation, the result being

$$\frac{1}{2} - \frac{1}{\sqrt{\pi}} \int_{-\infty}^{q(x)} e^{-t^2} dt = \frac{\lambda}{2(1-\lambda)} \frac{G(\sigma_N^2, x^2)}{G(\sigma^2, x^2)} \tag{B2}$$

$$q(x) = \frac{\tilde{r}_1(x) - px}{\sqrt{2\beta}} ; \quad \sigma^2 = \sigma_r^2 + \sigma_N^2$$

$$p = \frac{\sigma_r^2}{\sigma^2} ; \quad \beta^2 = \frac{\sigma_r^2 \sigma_N^2}{\sigma^2}$$

If $x \to \infty$, the RHS of equation (B2) $\to 0$ and therefore the integral must have a value of $\frac{\sqrt{\pi}}{2}$ or

$$q(x) = 0 \to \tilde{r}_1(x) = px \tag{B3}$$

From equation (B3) we conclude that the asymptotic behavior of $\tilde{r}_1(x)$ is identical to $\tilde{r}_2(x)$.

The delta function conceptually has a width $2\epsilon(-\epsilon < r < \epsilon)$. When $\tilde{r}_1(x) < -\epsilon$, the first integral of equation (B1) has no contribution; when $\tilde{r}_1(x) = 0$, the integral equals $(\lambda/2)f_N(x)$, and finally a full contribution of $\lambda f_N(x)$ is reached when $\tilde{r}_1(x) = \epsilon$. Denoting the x coordinate at which $\tilde{r}_1(x) = \epsilon$ by x_c, equation (B1) gives a complicated relation for x_c in terms of λ, σ_r^2 and σ_N^2. The clutter can be reduced to the following:

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{q_c} e^{-t^2} dt = \frac{\lambda}{1-\lambda} \frac{G(\sigma_N^2, x_c^2)}{G(\sigma^2, x_c^2)} \tag{B4}$$

$$q_c = \frac{px_c - \epsilon}{\sqrt{2\beta}}$$

Further simplification requires restrictions on the range of λ and $\sigma_r^2/\sigma_N^2 = S$. First, assume q_c is such that the integral in equation (B4) can be approximated by the first term in its asymptotic expansion (the expansion is good for $q_c > 1$):

$$\frac{1}{\sqrt{\pi}} \int_0^{q_c} e^{-t^2} \simeq \frac{1}{2}\left[1 - \frac{\exp(-q_c^2)}{\sqrt{\pi} q_c}\right] \tag{B5}$$

Substituting equation (B5) into (B4) and letting $\epsilon \to 0$ gives

$$e^{q_c^2} = \sqrt{1+S}\left[\frac{\lambda}{1-\lambda} + \sqrt{\frac{2}{\pi S}} \frac{\sigma_N}{x_c}\right] \tag{B6}$$

If λ and S are restricted so that

$$\frac{\lambda}{1-\lambda} \gg \sqrt{\frac{2}{\pi S}} \frac{\sigma_N}{x_c} \tag{B7}$$

$$S \gg 1$$

Equation (B6) can be simplified to give a closed form solution for x_c in terms of S, λ, and σ_N:

$$x_c = \sqrt{2}\sigma_N \left\{\ln\left(\frac{\sqrt{S}\lambda}{1-\lambda}\right)\right\}^{\frac{1}{2}} \tag{B8}$$

The term σ_N is related to σ_x [see equation (2.16)] and so equation (B8) can be solved for the ratio x_c/σ_x. After viewing Figure 2.2, in which equation (B1) was solved computationally without approximations, it is evident that if S and λ are restricted as in equation (B7), a valid model for $\tilde{r}_1(x)$ is the one given at the beginning of the appendix. This certainly appears to be the simplest non-linearity possible.

Appendix C
Penalty Function Factorization

The purpose of this appendix is to show that the mean square error (MSE) can be factored into a penalty function $P(\tilde{\lambda},\tilde{S};\lambda,S)$ and the noise variance σ_N^2.

Recall the definition of the MSE:

$$(MSE)^k = E\left[\tilde{r}(X^k) - R\right]^2 \tag{C1}$$

Henceforth, superscripts are dropped with the understanding that all variables refer to the k-th iteration. Specializing the analysis to the case $\tilde{r} = \tilde{r}_2$, equation (C1) becomes

$$\begin{aligned} MSE &= \iint \left[\tilde{r}_2(x) - r\right]^2 f_{R,X}(r,x)\, dr\, dx \\ &= \iint \left[\tilde{r}_2(r+n) - r\right]^2 f_R(r)\, f_N(n)\, dr\, dn \end{aligned} \tag{C2}$$

The probability density functions (pdf) for both R and N are given by equations (2.7b) and (2.12b) respectively:

$$f_R(r) = \lambda\delta(r) + (1-\lambda)G(\sigma_r^2,r^2)$$

$$f_N(n) = G(\sigma_N^2,n^2)$$

Using these pdfs in equation (C2) and integrating out the delta function gives

$$\begin{aligned} MSE = \lambda \int \tilde{r}_2^2(n) f_N(n)\, dn + (1-\lambda) \iint \left[\tilde{r}_2(r+n) - r\right]^2 \\ G(\sigma_r^2,r^2) f_N(n)\, dr\, dn \end{aligned} \tag{C3}$$

The reflection coefficient estimator $\tilde{r}_2(x)$ is given by equation (2.14) in the text. Using the variables $(\tilde{S},\tilde{\lambda})$, this equation can be rewritten as

$$
\begin{aligned}
\tilde{r}_2(x) &= x h_2(x) \\
h_2(x) &= \frac{\tilde{S}}{1+\tilde{S}}\left[1 + \tilde{c}\exp\left(\frac{-x^2}{2\sigma_h^2}\right)\right] \\
\tilde{c} &= \frac{\tilde{\lambda}}{1-\tilde{\lambda}}\sqrt{1+\tilde{S}} \qquad\qquad (C4) \\
\sigma_h^2 &= \left(\frac{1+\tilde{S}}{\tilde{S}}\right)\tilde{a}\sigma_x^2 \\
\tilde{a} &= (1-\tilde{\lambda})\tilde{S} + 1
\end{aligned}
$$

Note that $\tilde{r}_2(x)$ is now a function of $\tilde{S}$, $\tilde{\lambda}$ and σ_x. A similar transformation can be made on the functions $G(\sigma_r^2,r^2)$ and $f_N(n)$ to give

$$
G(\sigma_r^2,r^2) = \left(\frac{a}{2\pi S}\right)^{1/2}\frac{1}{\sigma_x}\exp\left(\frac{-ar^2}{2S\sigma_x^2}\right)
$$

$$
f_N(n) = \left(\frac{a}{2\pi}\right)^{1/2}\frac{1}{\sigma_x}\exp\left(\frac{-an^2}{2\sigma_x^2}\right) \qquad (C5)
$$

$$
a = (1-\lambda)S + 1
$$

Both $G(\sigma_r^2,r^2)$ and $f_N(n)$ are now functions of S, λ and σ_x. All that remains is to transform the variables in equation (C3) into dimensionless quantities, using

$$
p = \frac{r}{\sigma_x} \qquad\qquad q = \frac{n}{\sigma_x} \qquad (C6)
$$

Substituting equation (C6) into (C3) gives

$$
MSE = \sigma_x^2\, P'(\tilde{\lambda},\tilde{S};\lambda,S)
$$

$$
P'(\tilde{\lambda},\tilde{S};\lambda,S) = \lambda \int \tilde{r}_2^2(q) f_N(q)\, dq +
$$

$$(1-\lambda) \quad \iint \left[\tilde{r}_2(p+q) - p\right]^2 G(\sigma_r^2, p^2) f_N(q) \, dp \, dq \qquad \text{(C7)}$$

In equation (C7), the functions $\tilde{r}_2$, f_N and G are as in equations (C4) and (C5) with σ_x^2 replaced by unity. Finally, converting the σ_x^2 multiplying P' into σ_N^2 gives the result

$$MSE = P(\tilde{\lambda},\tilde{S};\lambda,S)\sigma_N^2$$

$$P(\tilde{\lambda},\tilde{S};\lambda,S) = a \, P'(\tilde{\lambda},\tilde{S};\lambda,S)$$

Neither the single nor the double integral in equation (C7) can be evaluated analytically, and both were computed for each set of variables $\tilde{\lambda},\tilde{S}$ to obtain the plots of Figure 2.1.

Appendix D
Bussgang Theorems

The purpose of this appendix is to prove two theorems referred to in Chapter II. We start by proving yet another theorem that defines the class of all Bussgang processes.

As a preliminary, recall the definition for a process $\{X_t\}$ to be Bussgang. Letting z( ) refer to any ZNL, we have:

$$\text{If} \quad \frac{E\left[X_i z(X_{i+\tau})\right]}{E\left[X_i X_{i+\tau}\right]} = \text{constant, any } \tau, \tag{D1}$$

$$\text{Then} \quad \{X_t\} \quad \text{is} \quad \text{Bussgang.}$$

Two common classes of Bussgang processes are (i) all independent processes and (ii) all colored Gaussian processes. Less common Bussgang processes are telegraph waves and their derivatives [if z( ) is odd].

Theorem One: A second order, stationary stochastic process is Bussgang if and only if

$$E\left[X_{i+\tau} | X_i = x\right] = p(\tau)x$$

$p(\tau)$ = normalized autocorrelation (A/C) function

Proof:

(i) For $\{X_t\}$ to be Bussgang, we require equation (D1) to be valid:

$$E\left[X_1 z(X_2)\right] = \text{constant } E\left[X_1 X_2\right] \tag{D2}$$

but

$$E\left[X_1 z(X_2)\right] \triangleq \iint x_1 z(x_2) f_{X_1,X_2}(x_1,x_2)\,dx_1 dx_2$$

$$= \int z(x_2)\left[\int x_1 f_{X_1|X_2}(x_1|x_2)dx_1\right] f_{X_2}(x_2)dx_2$$

The term in brackets is the conditional expectation, hence

$$E\left[X_1 z(X_2)\right] = \int z(x_2) E\left[X_1|X_2=x_2\right] f_{X_2}(x_2)dx_2 \tag{D3}$$

Also, define $E\left[X_1X_2\right]$ via

$$E\left[X_1X_2\right] \triangleq r(\tau) \tag{D4}$$

(ii) Substituting equations (D3) and (D4) into equation (D2), the constant is independent of τ if and only if

$$E\left[X_1|X_2=x_2\right] = r(\tau)h(x_2) \tag{D5}$$

$h(\ )$ = unspecified function

(iii) When $\tau=0$, equation (D5) becomes

$$E\left[X_1|X_1=x_2\right] = r(0)h(x_2)$$

$$x_2 = r(0)h(x_2)$$

$$h(x_2) = \frac{x_2}{r(0)} \tag{D6}$$

Substituting equation (D6) into (D5) gives the stated theorem.

Theorem One can be used to readily determine whether a process is Bussgang (e.g. it is well known that the conditional mean of a Gaussian distribution is linear - the same holds for telegraph waves). The remainder of the appendix is concerned with establishing the results quoted in Chapter II.

The motivation for proposing the next theorem occurred while studying the deconvolution of the synthetic data. During the course of iterating, it was noted that a tiny echo was successively eliminated from the deconvolution x_t. This indicated to us that the algorithm was trying to force the deconvolved time series to be Bussgang by removing the echo. A simple way to simulate an echo is to input a reflection sequence into a two-term filter.

Theorem Two: If a stationary independent process $\{C_t\}$, characterized by a normal mixture, is input into a two-term filter, $(1,a)$, the resulting process $\{X_t\}$ is not Bussgang.

Proof:

(i) If we can prove that $E\left[X_{N+\tau}|X_N=x\right] \neq p(\tau)x$, then the above theorem implies $\{X_t\}$ is not Bussgang.

(ii) The process $\{X_t\}$ is related to $\{C_t\}$ via

$$X_k = C_k + aC_{k+1} \tag{D7}$$

Using equation (D7), the normalized A/C of $\{X_t\}$ can be derived:

$$\begin{aligned} p(0) &= 1 \\ p(1) &= \frac{a}{1+a^2} \\ p(\tau) &= 0, \ |\tau|\geq 2 \end{aligned} \tag{D8}$$

(iii) Next, expand the conditional expectation:

$$E\left[X_{N+\tau}|X_N=x\right] = E\left[C_{N+\tau} + aC_{N+\tau+1}|C_N + aC_{N+1}=x\right] \tag{D9}$$

For $|\tau|\neq 1$, it is evident from equation (D9) that

$$E\left[X_{N+\tau}|X_N=x\right] = p(\tau)x, \ |\tau|\neq 1$$

If the same holds for $|\tau|=1$, $\{X_t\}$ would be Bussgang. It is simple, yet tedious, to prove otherwise. For $\tau=+1$, equation (D9) becomes (simplifying the notation):

$$E\left[X_{N+1}|X_N=x\right] = E\left[C_2 + aC_3|C_1 + aC_2=x\right]$$

$$= E\left[C_2|C_1 + aC_2=x\right]$$

$$= \frac{\int c_2 f_{C_2,C_1+aC_2}(c_2,x)dc_2}{f_{C_1+aC_2}(x)}$$

$$= \frac{\int c_2 f_C(c_2) f_C(x - ac_2)dc_2}{f_{C_1+aC_2}(x)} \qquad \text{(D10)}$$

(iv) The pdf of $\{C_t\}$ is characterized by a normal mixture:

$$f_C(c) = \lambda G(1,c^2) + (1-\lambda)G(S^2,c^2) \qquad \text{(D11)}$$

In equation (D11), the variance of the first Gaussian has been normalized to one.

(v) Equation (D11) can be used to calculate both numerator and denominator of equation (D10). Noting that the pdf in the denominator is simply a linear combination of normal RVs, equation (D10) becomes (omitting algebra):

$$E\left[X_{N+1}|X_N=x\right] = ax\frac{N^+(\lambda,S,x)}{D(\lambda,S,x)}$$

$$N^+(\lambda,S,x) = \frac{\lambda^2 G(1+a^2) + (1-\lambda)^2 G\left[S^2(1+a^2)\right]}{1+a^2}$$

$$+ \lambda(1-\lambda)\left[\frac{G(S^2+a^2)}{S^2+a^2} + \frac{S^2 G(1+S^2a^2)}{1+S^2a^2}\right]$$

$$D(\lambda,S,x) = \lambda^2 G(1+a^2) + (1-\lambda)^2 G\left[S^2(1+a^2)\right]$$

$$+ \lambda(1-\lambda)\left[G(1+a^2S^2) + G(S^2+a^2)\right]$$

The variable x^2 has been omitted from the function G for notational convenience. At this point it is evident that $\{X_t\}$ is not Bussgang unless $\lambda=0$, $=1$ or $S=1$, i.e. unless $\{C_t\}$ is Gaussian. It is instructive, however, to compare the negative lag with the above result.

(vi) The conditional mean for $\tau=-1$ is derived in an analagous manner, resulting in

$$E\left[X_N | X_{N+1}=x\right] = ax\frac{N^-(\lambda,S,x)}{D(\lambda,S,x)}$$

$$N^-(\lambda,S,x) = \frac{\lambda^2 G(1+a^2) + (1-\lambda)^2 G\left[S^2(1+a^2)\right]}{1+a^2}$$

$$+ \lambda(1-\lambda)\left[\frac{G(1+S^2a^2)}{1+S^2a^2} + \frac{S^2G(S^2+a^2)}{S^2+a^2}\right]$$

Note that when $\{C_t\}$ is Gaussian, $N^+=N^-$. If $\{C_t\}$ is a normal mixture, however, $N^+\neq N^-$, and this means the mixture is sensitive to phases - a property not shared by Gaussian processes.

The extension of Theorem Two for (i) arbitrary length wavelets and (ii) arbitrary non-Gaussian pdfs is not possible. In our analysis, we restrict the wavelet to being "delta-like." The restriction is necessary to avoid the implications of the Central Limit Theorem.

Theorem Three: An independent, stochastic process $\{C_t\}$, convolved with a delta-like wavelet w_t, is Bussgang if and only if the process $\{C_t\}$ is Gaussian.

Proof:

(i) If - All Gaussian processes are Bussgang.

(ii) Only if - Let:

$$X_i = \Sigma\, C_{i-\tau_1} w_{\tau_1} \tag{D12}$$

$$= C_i + \sum_{\tau_1 \neq 0} C_{i-\tau_1} w_{\tau_1}$$

w_t is delta-like: $w_0 \triangleq 1$, $w_t << w_0$, $t \neq 0$

$\{C_t\}$ is a zero mean, independent process with pdf $f_C(c)$

(iii) If $\{X_t\}$ is Bussgang, the crosscorrelation (C/C) between $\{X_t\}$ and $\{z(X_t)\}$ must be symmetric. Our strategy is to show that the C/C is symmetric if and only if $\{C_t\}$ is Gaussian. Letting z( ) be a ZNL, we can approximate it, according to the delta-like character of w_t:

$$z(X_i) \simeq z(C_i) + z'(C_i) \sum_{\tau_2 \neq 0} C_{i-\tau_2} w_{\tau_2} \tag{D13}$$

(iv) The C/C can be calculated using equations (D12) and (D13), giving the result:

$$E\left[X_i z(X_j)\right] = A+B+C+D \tag{D14}$$

$$j = i + \tau$$

$$A = Z\delta_{ij} \qquad B = w_{-\tau} Z,\ \tau \neq 0 \qquad C = w_\tau \sigma^2 Z',\ \tau \neq 0 \qquad D = Z'\sigma^2 \sum_{k \neq 0, \tau} w_k w_{k-\tau}$$

$$Z = E\left[C_i z(C_i)\right] \qquad Z' = E\left[z'(C_i)\right] \qquad \sigma^2 = E\left[C_i^2\right]$$

(v) Assuming $w_\tau \neq w_{-\tau}$, we require the coefficients of w_τ and $w_{-\tau}$ to be identical for the C/C to be symmetric:

$$Z = \sigma^2 Z'$$

$$\int c z(c) f_C(c) dc = \sigma^2 \int \frac{dz}{dc} f_C(c) dc \tag{D15}$$

After integrating equation (D15) by parts, we obtain

$$\int z(c)\left[cf_C(c) - \sigma^2 \frac{df}{dc}\right]dc = 0, \text{ any } z$$

The arbitrariness of the ZNL means the integrand in the above equation must be identically zero:

$$cf_C(c) = \sigma^2 \frac{df}{dc}$$

$$f_C(c) = G(\sigma^2, c^2)$$

(vi) The remaining part of the proof is to consider the case when $w_\tau = w_{-\tau}$, i.e. a symmetric wavelet is present. For $\{X_t\}$ to be Bussgang, the ratio of the C/C to the A/C at lag τ must be identical to the ratio at lag 0, independent of the ZNL:

$$\frac{E\left[X_i z(X_{i+\tau})\right]}{E\left[X_i X_{i+\tau}\right]} = \frac{E\left[X_i z(X_i)\right]}{E\left[X_i^2\right]} \tag{D16}$$

The A/C is obtained by setting $Z=\sigma^2$ and setting $Z'=1$ in the C/C formula. If equation (D14) is substituted into (D16) and all terms of $O(w_k^2)$ are dropped, eventually this result is obtained:

$$\sigma^2 Z' = Z$$

This is identical to equation (D15), and the proof is complete.