SEMBLANCE AND OTHER SIMILARITY MEASUREMENTS

By: M. Turhan (Tury) Taner , Rock Solid Images

November, 1996

INTRODUCTION:

Similarity measurements are one of the most important tools in the seismic processing tool box. They are used to determine all kinds of velocities, track reflected events, measure signal to noise ratios and etc. I will describe similarity measurements as a simple least squares solution, then cover correlation and semblance coefficients. Finally I will cover the complex trace correlations and show the group and phase relationships.

CORRELATION COEFFICIENT (SCALAR):

The conventional computation of the similarity is by the correlation coefficient. The computation is made so that the trace amplitudes do not influence the results. This suggests that we are measuring the phase component of the similarity. Let f(t) and h(t) are two traces for which we wish to compute the correlation coefficient c(t). (Where t is the time lag between two traces). Then;

$$c(t) = \sum_{t=T_1}^{t=T_2} \{ f(t).h(t+t) \} \Big/ \sqrt{\sum_{t=T_1}^{t=T_2} f^2(t)} \sqrt{\sum_{t=T_1}^{t=T_2} h^2(t)} \qquad (1)$$

Denominator in Equation 1 causes both f(t) and h(t) to have an RMS amplitude [sum of squares of mean amplitudes of both of traces be equal to 1 (one)], thus removing the effects of the difference in amplitude scaling. c(t) varies between -1.0 and 1.0. The value of -1.0 means that two traces are identical with they are of opposite polarity. The value zero means that they are orthogonal, and zero similarity. 1.0 correlation coefficient means identical traces with exception of possible scaling difference.
If we pick the lag (the t value) for largest positive peak of the cross-correlation coefficient values, we have determined the time delay between two traces. In other words, if we shift one trace with respect to the other by the amount of t delays, they would be most alike. We can express this in a least mean square problem. That is, we wish to determine the time delay that would make the two traces most alike (or, minimize the sum difference squares).

$$e^2 = \sum_{t=T_1}^{t=T_2} [f(t) - h(t+t)]^2 \qquad (2)$$

We can expand this equation to obtain;

$$e^2 = \sum_{t=T_1}^{t=T_2} f^2(t) + \sum_{t=T_1}^{t=T_2} h^2(t+t) - 2\sum_{t=T_1}^{t=T_2} [f(t).h(t+t)] \qquad (3)$$

We can see from equation 3 that, sum of squares of f(t) and h(t) will be constant over long time gathers. The only way e(t) is minimized is when we pick the largest positive value of the sum of the cross products of the two traces. Thus picking the lag corresponding to the largest positive peak of the cross-correlation function gives us the lag that the two traces are most alike. Then, the value of the cross-correlation function is the percent similarity of two traces at that lag.

This type of correlation is used to determine the time delays between the pilot trace and the individual traces within the CDP gather in a residual statics computation. We should note here that this lag correspond to the phase alignment only. Since only the scalar seismic traces are involved, we could call it the scalar cross-correlation coefficient.

CORRELATION COEFFICIENT (COMPLEX):

Let us now consider two complex traces F(t) and H(t), and compute complex cross-correlation function;

$$C(t) = \sum_{t=T_1}^{t=T_2} [F(t).H^*(t+t)] \qquad (4)$$

where F(t) = f(t) + i g(t) and H(t) = a(t) +i b(t) and i = $\sqrt{-1}$.
Since both the functions are complex, then the cross-correlation function will also be complex;

$$C(t) = \{\mathbf{Real}\}[C(t)] + i\ \{\mathbf{Imaginary}\}[\mathbf{C(}t\mathbf{)}] \qquad (5)$$

We can compute real and imaginary part separately,

{Real} [C(t)] = f(t).a(t+ t) + g(t).b(t $+t$) and,
{Imaginary} [C(t) =f(t).b(t+ t) **-** g(t).a(t+ t) (6)

Complex cross-correlation functions largest amplitude lag represent the best alignment of the total energy (group lag), and the instantaneous phase at this location represent the phase difference between the two traces. The group lag shows the amount of time shift and instantaneous phase shows the amount of phase rotation to best align the two traces. We can compute complex cross-correlation coefficient similar to the scalar one by dividing the cross-correlation function by the sum of the squares of the total energy. Real part represent the dot product of two vectors. This, normalized by the total power, represent the cosine of the angle between two vectors. Thus, the absolute value of the normalized real part will vary between zero and one. Zero value will show that the vectors are orthogonal, and value of one represent the two vectors are parallel. Similarly, imaginary part corresponds to the sine of the angle between the vectors, thus the ratio of imaginary to real part is equal to the tangent of the angle between the two vectors. Thus, the arctangent of the ratio is the instantaneous phase, which is equal to the phase difference between two traces, for example, between the pilot trace and a new trace.

SIMILARITY MEASUREMENT OF A NUMBER OF TRACES IN A GATHER:

The cross-correlation coefficients discussed above are good only to measure the similarity between two traces. To expand this to ensembles of many traces, we need to change our measuring tools. For example , we can consider stacking as a filter. Filter performance can be measured by computing the output to input power ratio. (This is generally known as the semblance coefficient, but I will give a different definition of the semblance coefficient later). We can compute pair-wise cross-correlation coefficients between all of the traces in the gather and look at the simple or weighted average of the cross-correlation coefficients and the variance of the coefficients. These will allow us the state of the similarity between traces and also isolate the outliers, so we can reject them from the gather.
Let us now look at some of these measurements and investigate their interrelations.

INPUT TO OUTPUT POWER RATIO:

If we consider stacking as a multi-channel filter, individual traces being the input and the stack trace as the output, then we can measure the stack performance by the ratio of input power to the output power. It is obvious that if output power is same as the input power, we can conclude that stack passed all of the energy without any loss. The ratio will be one. Since we are measuring the power , the output can not be negative, it could be at least zero, corresponding to the zero output to input power ratio. Let f(n,t) be the n'th trace of a gather consisting of N traces, and s(t) represent the stack trace as;

$$s(t) = \sum_{n=1}^{N} f(n,t) \tag{7}$$

then, the output to input power ratio is given by; (summing over a time window of $T_1 \le t \le T_2$)

$$r(t) = \sum_{t=T_1}^{t=T_2} s^2(t) \, / \, N \sum_{n=1}^{N} \sum_{t=T_1}^{t=T_2} [f^2(n,t)] \tag{8}$$

where the ratio varies between 0 and 1.
This ratio can be computed during the stacking process, in a running fashion using a small time window (rectangular, Butterworth or Gaussian type) by accumulating the sum of traces and trace squares. Then the square of sum trace can be divided by the sum of squares to obtain the desired ratio. This ratio will show the similarity variation in time direction, which could be used to identify relative coherence of events. The power ratio will not be influenced by the event amplitudes. However ambient background noise level will make lower amplitude primary (aligned) events appear less coherent.

SEMBLANCE COEFFICIENT:

If we look at the power ratio computation, the numerator contains the squares of the sum trace. This, in essence, contains the squares of individual traces and all of their pair-wise cross products. Let me demonstrate this by a simple example:
Let a, b, c be the input traces and their sum square will the as;

$$(a+b+c)^2 = a^2 + b^2 + c^2 + 2a.b + 2a.c + 2b.c \tag{9}$$

As we see from example above square of s(t), will likewise contain squares of all input traces, which are non-zero and do not contribute to differentiation between noise and signal elements. On the other hand cross-products represent the zero lag of cross-correlation between the traces and contain similarity information. If we subtract the sum of input trace squares (which is already computed) from the sum trace square, we will be left with the sum of all of the cross-products. Since there are N.(N-1) such products, we can compute average cross-product as a measure of similarity. By proper scaling, we can set the coefficients vary between zero and one. I call this coefficient the 'semblance' coefficient.

$$R(t) = \{\sum_{t=T_1}^{t=T_2} s^2(t) - \sum_{n=1}^{N} \sum_{t=T_1}^{t=T_2} [f^2(n,t)]\} \, / \, \{(N-1)\sum_{n=1}^{N} \sum_{t=T_1}^{t=T_2} [f^2(n,t)]\} \tag{10}$$

Scaling can be checked easily by setting all of the input traces identically equal to f(t), then R(t) will be at most equal to 1. At worst case is sum trace equal to zero (where traces may have been identical but half of them opposite polarity to the others) the R(t) will be equal to -1/(N-1). Therefore (including this very special case) the semblance coefficient varies between -1/(N-1) and 1.
Semblance of traces with varying RMS amplitudes will be lower than the ones with balanced amplitudes. Thus, before semblance computation, traces should be RMS amplitude balanced.

TRACE GATHER COHERENCY IN COMPLEX FORM:

Complex number can represent vectors in 2 dimensional domain. Real part representing the projection on the real axis and the imaginary part representing the projection on the imaginary axis. In seismic data complex numbers can represent the kinetic and potential energy components. In fact, we use this representation in computation of seismic attributes. Here we will investigate their use in trace alignments in a gather and their degree of similarity. Complex cross correlation function characteristics have been discussed above. The location of the envelope peak shows the group delay and the instantaneous phase at that location represent the phase difference between the two traces.
Since each trace sample is represented as a vector, we can measure alignment of these vectors as a similarity measurement. This corresponds to the phase component. We can also measure the envelop alignment, which corresponds to the group alignment. In CDP gathers these will represent phase and group velocities.

INPUT TO OUTPUT POWER RATIO IN COMPLEX FORM:

Input to output power ration is computed similar to the scalar ratio computation. Complex trace output is a vectorial sum, thus the sum trace is formed separately summing the real and the imaginary parts.

$$S(t) = \sum_{n=1}^{N} f(n,t) + i.\sum_{j=1}^{N} g(n,t) \tag{11}$$

Then, the total power is computed by multiplying S(t) by its complex conjugate. Therefore the complex output to input power ratio is given by;

$$R(t) = \sum_{t=T_1}^{t=T_2} S(t).S^*(t) \,/\, \sum_{t=T_1}^{t=T_2} \{\sum_{n=1}^{N} F(n,t).F^*(n,t)\} \tag{12}$$

COMPLEX SEMBLANCE COEFFICIENT:

Since the total power is a non-negative value, then the multiplication of a complex number with its complex conjugate must be a non-zero scalar value. Sum of complex trace power can be computed similarly as in the example given by equation 9. However, this time we will have to form the product by the complex conjugate of the sum trace;

$$(a+b+c).(a+b+c)^* = a.a^* + b.b^* + c.c^* + a.b^* + a^*.b + a.c^* + a^*.c + b.c^* + b^*.c \tag{13}$$

where $a.a^*$ is the power of vector a and $a.b^* + a^*.b$ is equal to twice of the dot product between vectors a and b. Let subscripts r and im represent the real and imaginary parts of the complex numbers, respectively, then;

$$a.b^* + a^*.b = [a(r) + ia(im)].[b(r) - ib(im)] + [a(r) - ia(im)].[b(r) + ib(im)] \quad ,$$

which reduces to;

$$a.b^* + a^*.b = 2.\{a(r).b(r) + a(im).b(im)\} = 2[a.b] \tag{14}$$

This similar to the results shown by equation 9, with addition of the cross products of the imaginary components. Therefore the complex semblance coefficient is given by;

$$R(t)=[\sum_{t=T_1}^{t=T_2} S(t).S^*(t)-\sum_{t=T_1}^{t=T_2}\{\sum_{n=1}^{N} F(n,t).F^*(n,t)\}]/[(N-1)\sum_{t=T_1}^{t=T_2}\{\sum_{n=1}^{N} F(n,t).F^*(n,t)\}\} \quad (15)$$

Complex trace semblance is a measure of trace similarity from the point of view of vector alignment. We could get similar values by computing instantaneous phase values and look at their statistics.

DATA EXAMPLES:

Semblance computation examples will be furnished in a separate report. I wished to give theoretical background of the computation in this report.

REFERENCES:

Kimball, C. V. and Marzetta, T. L., 1983, Improved interval transit time measurement based on semblance: 53rd Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, Session:S6.8.

Leven, J. H. and Roy, C. K., 1984, A semblance-weighted slowness-filter in the time domain: 54th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, Session:S4.6.

Li, G. and Stewart, R. R., 1991, Automatic velocity analysis of crosswell seismic data: 61st Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 829-831.

Haskell, N. L., Nissen, S. E., Lopez, J. A. and Bahorich, M. S., 1995, 3-D seismic coherency and the imaging of sedimentological features: 65th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1532-1534.

Katz, S. and Henyey, T., 1990, Poststack diffraction-coherency migration: 60th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1377-1380.

Kern, M. and Symes, W. W., 1992, Calculation of the differential semblance gradient: 54th Mtg. Eur. Assoc. Expl Geophys., Abstracts, 104-105.

Keydar, S., Landa, E., Ryzhikov, G. and Sorin, V., 1995, A new approach to coherency inversion by global optimization: 65th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1081-1083.

Landa, E., Thore, P., Sorin, V. and Koren, Z., 1990, Uncertainty analysis of velocity estimates from coherency inversion: 60th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1249-1252.

Neidell, N. S. and Taner, M. T., 1990, Semblance and other coherency measures for multichannel data, in Byun, B. S., Ed., Velocity analysis on multichannel seismic data: Soc. Expl. Geophys., 263-278.

Nemeth, T., 1994, Migration and coherency stacking of the Friendswood reverse VSP data: 64th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 97-100.

Nissen, S. E., Haskell, N. L., Lopez, J. A., Donlon, T. J. and Bahorich, M. S., 1995, 3-D seismic coherency techniques applied to the identification and delineation of slump features: 65th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1535-1536.

Ortmann, K. A. and Wood, L. J., 1995, Successful application of 3-D seismic coherency models to predict stratigraphy, offshore eastern Trinidad: 65th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 101-103.

Reiter, E. C., Toksoz, M. N. and Purdy, G. M., 1993, A semblance-guided median filter: Geophys. Prosp., 41, 15-42.

Reiter, E. E. and Toksoz, M. N., 1991, A semblance-guided median filter: 53rd Mtg. Eur. Assoc. Expl Geophys., Abstracts, 4.

Reshef, M., Landa, E. and Ravid, T., 1989, Shear velocity estimation by coherency optimization: 59th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 925.

Sattlegger, J. W., Stiller, P. K. and Echterhoff, J. A., 1976, Dip selective migration velocity determination: Geophys. Prosp., 24, 650-659.

Symes, W. W. and Kern, M., 1992, Velocity discrimination by differential semblance: 54th Mtg. Eur. Assoc. Expl Geophys., Abstracts, 98-99.

Symes, W. W. and Kern, M., 1994, Inversion of reflection seismograms by differential semblance analysis: algorithm structure and synthetic examples: Geophys. Prosp., 42, 565-614.

Symes, W. W., 1988, Velocity inversion by coherency optimization: 58th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, Session:S16.3.

Sicking, C. J., 1987, Diffraction semblance for velocity and structure analysis: 57th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, Session:S1.4.

Symes, W. W. and Kern, M., 1992, Velocity inversion by differential semblance optimization for 2-D common source data: 62nd Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1210-1213.

Symes, W. W. and Versteeg, R., 1993, Velocity model determination using differential semblance optimization: 63rd Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 696-699.

Versteeg, R., Ehinger, A. and Geoltrain, S., 1991, Sensitivity of migration coherency panels to the velocity model: 61st Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1251-1254.

Versteeg, R., Geoltrain, S. and Ehinger, A., 1991, Use of migration coherency panels for velocity model determination: 61st Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1255-1258.