Robust inversion of non-linear transformations with an application to VSP's

William S. Harlan and Patrick Lailly

Introduction: choosing and simplifying the statistical tools

An inversion of seismic data for physical parameters must suppose a statistical model, whether acknowledged or not, that limits the possible expressions of signal and noise. This model contains *a priori* information -- information not requiring the inverted result for its estimation. An unbiased model must reflect the possibilities to be found regionally in the data; a robust model must, in addition, derive directly from the data. *A priori* statistics introduce otherwise ignored information -- the frequency of events.

Signal and noise efficiently described, by the smallest number of random variables (parameters to be estimated), allow the simplest statistical tools. Joint probability functions (j.p.f.'s) allow the most arbitrary dependence between variables: the data never possess enough redundancy for their estimation. Marginal probability functions (m.p.f.'s) describe each variable independently. If a transformation has rendered all variables statistically independent (we shall say "focused"), then j.p.f.'s may be calculated from m.p.f.'s.

The data easily provide sufficient redundancy to estimate m.d.f.'s. Again, *a priori* statistics should reflect regional possibilities. Knowledge of one reliable event should increase the likelihood of finding such another event nearby. Thus, one not only expects but desires that estimated m.p.f.'s change slowly over spatial dimensions and time. Because of this stationarity, a histogram prepared from a great many samples with identical m.p.f.'s will describe the possibilities open to them all.

Because each component of the data, signal or noise, has a different focusing transformation, the corresponding m.p.f.'s measure distinctly different information. We shall see that, with such measurements, one may estimate and extract the most reliable events from a focused component. The interference of other components may be iteratively subtracted.

Inversion with known statistics

We delay until the following section the difficulty of estimating m.d.f.'s from signal and noise and assume, for the moment, that these are known. Let the data be a random process defined as a sum of noise and non-linearly transformed signal.

$$d_i = f_i(\bar{s}) + n_i \tag{1}$$

$$\text{or } \bar{d} = \bar{f}(\bar{s}) + \bar{n}$$

We define $\bar{s}$ and $\bar{n}$ as focused, stationary random processes (random vectors).

We define geophysical noise as that untransformed component showing no spatial coherence (we allow some temporal coherence). If a component possesses significant coherence, then it should be properly defined after another transformation, as a second variety of signal.

Let $p_s(x)$ and $p_n(x)$ be the corresponding m.p.f.'s. We define the MAP inverse as that $\bar{s}$ most probable for a given $\bar{d}$. (MAP abbreviates *maximum a posteriori*, so called because one assumes knowledge of the final transformed result.) We maximize the following conditional probability function

$$J_1(\bar{s}) = p_{\bar{s}\,|\,\bar{d}}(\bar{s}\,|\,\bar{d}) = \prod_i \frac{p_s(s_i)\,p_n[d_i - f_i(\bar{s})]}{p_{d_i}(d_i)} \tag{2}$$

The denominator merely normalizes, does not affect a maximization. Since the logarithm increases monotonically for positive functions, we may also maximize

$$J_2(\bar{s}) = \sum_i \ln p_s(s_i) + \sum_i \ln p_n[d_i - f_i(\bar{s})] + constants \tag{3}$$

When the signal and noise are gaussian processes, maximizing J_2 will equivalently minimize

$$J_3(\bar{s}) = \frac{||\bar{s}||^2}{\sigma_s^2} + \frac{||\bar{d} - \bar{f}(\bar{s})||^2}{\sigma_n^2} \tag{4}$$

the least-squares (l.s.) result. We have eliminated covariance matrices by choices of transformations which diagonalize and normalize them. Again, coherent noise should be defined as a second variety of signal.

In fact, a classic (deterministic), arbitrarily normed inverse such as the L^1 or L^p, corresponds to a specific m.p.f. for noise -- the signal being unconstrained by an arbitrarily broad distribution. We find

$$p_n(x) = \frac{1}{C} e^{-|x|^p} \tag{5}$$

for the L^p inverse.

The l.s. inverse of many linear transformations possesses a simple closed form. Other statistics require a method of descent.

For a given estimate $\bar{s}_0$ of the signal, the gradient of J_2 is

$$g^i = \frac{\partial J_2}{\partial s_i}\Big|_{\bar{s}=\bar{s}_0} \tag{6}$$

$$= \frac{p_s'(s_i^0)}{p_s(s_i^0)} - \sum_j F_{ij}^0 \frac{p_n'[d_j - f_j(\bar{s}_0)]}{p_n[d_j - f_j(\bar{s}_0)]}$$

$$where \ \ F_{ij}^0 \equiv \frac{\partial f_j(\bar{s}_0)}{\partial s_i}$$

Primes indicate definite derivatives. Successive gradients become a linear function of the data only for gaussian m.p.f.'s and a linear $\bar{f}$:

$$g_i = \frac{s_i^0}{\sigma_s^2} - \frac{1}{\sigma_n^2}\sum_j F_{ij}^0[d_j - f_j(\bar{s}_0)] \tag{7}$$

For a linear, or linearized, $\bar{f}$, the projection $\sum_j F_{ij}^0$ becomes the adjoint operation. Most descent methods would require that $p_s(x)$ and $p_n(x)$ have no local maxima to obscure the unique global maximum of J_2. Smoothing the estimates of these distributions will considerably speed the descent: local gradients then better indicate the global maximum.

The simplest method of steepest descent would update

$$s_i = s_i^0 - \alpha g_i \tag{8}$$

where α minimizes $J_2(\bar{s}_0 - \alpha\bar{g})$. With the easily calculated derivative $dJ_2/d\alpha$, a line search for the best α converges quickly. For a linear $\bar{f}$ calculate α directly from appropriate scalar products (the classical l.s. method). Certainly conjugate gradient methods, such as the Fletcher-Reeves, should converge more quickly.

The approach of this section gives simplicity and robustness to the inversion of highly non-gaussian noise and non-linearly transformed signal. Specific m.p.f.'s for noise produce all other robust normed inverses as a subset of this formulation. But for an unbiased robust inverse, *a priori* statistics for signal and noise must derive directly from the data.

Inversion with estimated statistics

Distinguishing signal and noise m.d.f.'s must precede a separation (inversion) of their corresponding components. A lack of knowledge of one m.d.f. weakens an estimate of the other. We take the following strategy: estimate the m.p.f. of one component given the greatest possible over-estimate of the other. Such statistics allow the most pessimistic possible estimate of the first component. From such an estimate, extract those events that, for a chosen reliability, are minimally corrupted by other components.

For each iterative re-estimate of the signal we shall choose to linearize the nonlinear transformation $\bar{f}$ assuming the new estimate to be a small perturbation of the previous one, $\bar{s}_0$. We shall see that his linearization greatly simplifies the transformed statistics and their estimation. Because of the central-limit theorem, components with gaussian m.p.f.'s yield gaussian m.p.f.'s after linear transformation. Thus, gaussian signal and noise remain indistinguishable as a third useless component, hereafter called gaussian noise. In the previous section we found that, for gaussian noise, and for a linear, or linearized, $\bar{f}$, the MAP inverse becomes the l.s. inverse, made as a series of linear AP inverse If we iteratively extract (subtract) both nongaussian signal and noise from the data, then the l.s. inverse will approach the optimum MAP inverse. A linearized $\bar{f}$ will become increasingly accurate as signal perturbations decrease in magnitude.

Extract the most reliable signal as follows. Find the l.s. inverse of signal for the linearized $\bar{f}$, constraining signal sufficiently for a stable treatment of noise. Next, over-estimating the distribution of noise after transformation, determine which events in this linear inverse contain a sufficiently small percentage of noise, with sufficient reliability. Preserve these events as a reliable perturbation of the signal and disregard others. Iterate with updated linear transforms and m.d.f.'s until the presence of noise prevents any further reliable improvements in the signal. Subtract the most reliable signal perturbations (containing negligible noise) from the data and extract the most reliable noise by similar methods. With this noise absent, iteratively improve the signal once again, and re-extract the noise. The amount of corruption allowed in the extraction of a given component will determine the speed of the algorithm and the number of events considered invertible. At any iteration, however, one may feel confident of having made the most reliable and thereby the most important improvements in the estimates.

Define the following random variables

$$\bar{s}' = \bar{f}(\bar{s}) - \bar{f}(\bar{s}_0) \ ; \quad \bar{n}' = \bar{n} - \bar{n}_0 \ ; \quad \bar{d}' = \bar{s}' + \bar{n}' \tag{9}$$

where $\bar{n}_0$ is previously extracted noise. Consider a series of linear transformations (8) by the l.s. gradient (7) as a single linear transform L_{ij} of the data, giving the optimum l.s.

inverse.

$$d_i'' \equiv \sum_j L_{ij}\, d_i' \; ; \quad \bar{d}'' \equiv L\bar{d}' \tag{10}$$

and the corresponding transformed components as

$$\bar{s}'' \equiv L\bar{s}' \; ; \quad \bar{n}'' \equiv L\bar{n}' \tag{11}$$

Each prime designates a transformation of a component away from the definition of equation (1). Because we assume $\bar{n}$ to be focused (samples statistically independent), we may write

$$p_{n''}(x) = \prod_j{}^* \left[\frac{1}{L_{ij}} p_{n'}\left(\frac{x}{L_{ij}}\right)\right] \tag{12}$$

The asterisk here indicates convolution, and the $\prod^*$ multiple convolutions. In order to suppress the subscripts on m.p.f.'s, hereafter assume that the linear transformation preserves local stationarity for our *a priori* statistics. (Such an assumption adds robustness to the signal distribution and should to its perturbation.)

Define an exaggerated estimate of $p_{n''}(x)$ by assuming all events are noise, by ignoring the coherence of any signal.

$$\hat{p}_{n''}(x) \equiv \prod_j{}^* \left[\frac{1}{L_{ij}} p_{d'}\left(\frac{x}{L_{ij}}\right)\right] = p_{n''}(x) * \prod_j{}^* \left[\frac{1}{L_{ij}} p_{s'}\left(\frac{x}{L_{ij}}\right)\right] \tag{13}$$

This m.p.f. must overestimate the transformed noise and all positive moments. If the data contain no signal, then the estimate is perfect (the signal m.p.f. becomes a delta function). Estimate (13) easily by generating a random, focused array with the same m.p.f.'s as the data, transforming the L, and taking local histograms. Because the signal and noise remain statistically independent and additive after transformation, choices of their m.p.f.'s determine that for the data:

$$p_{d''}(x) = p_{s''}(x) * p_{n''}(x) \tag{14}$$

For consistency, always use this data m.p.f. with the other two. With the assumption of local stationarity, estimate $p_{d''}(x)$ from local histograms of the transformed data. The divergence of the estimate from the *a priori* m.p.f. in equation (14) should be minimal. Measure this divergence with the directed divergence (cross entropy) of Kullback. Minimizing $\int p_1(x)\ln[p_1(x)/p_2(x)]dx$ minimizes the "unpredictability" of $p_1(x)$ assuming $p_2(x)$ to be the most predictable. Iteratively discover the best estimate of $p_{s''}(x)$, given $p_{d''}(x)$ and $p_{n''}(x)$, by minimizing the following (suppressing primes)

$$J_4[p_s(x)] = \int p_d(x)\ln\left[p_d(x) / \int p_s(x-y)p_n(y)dy\right]dx \tag{15}$$

$$+ \frac{\lambda_1}{2}[\int p_s(x)dx - 1]^2 + \frac{\lambda_2}{4}\int [p_s(x) - |p_s(x)|]^2 dx$$

Add two Lagrange multipliers for the constraints of unit area and of positivity. To calculate the gradient of J_4 with respect to each point of the function $p_s(x)$, perturb a previous estimate with a delta function: $p_s(x) + \varepsilon\delta(x - x_0)$ and differentiate.

$$\frac{\partial}{\partial\varepsilon} J_4[p_s(x) + \varepsilon\delta(x - x_0)] \tag{16}$$

$$= -\int \frac{p_d(x)}{\int p_s(y)p_n(x-y)dy} p_n(x - x_0)dx$$

$$+ \lambda_1[\int p_s(x) - 1] + \lambda_2[p_s(x_0) - |p_s(x_0)|]$$

Iteratively perturb $p_s(x)$ with the negative of this gradient; an inexpensive line search finds the correct magnitude. The constraints easily determine the proper values of λ_1 and λ_2 for any magnitude of perturbation. The second term equally raises or lowers all points of $p_s(x)$ until the constraint of unit area is satisfied. The third term moves each point a sufficient positive distance to remove any negative excursions. The first term divides the estimate $p_d(x)$ by the *a priori* value and cross correlates with a shifted noise distribution, contributed by the perturbation of $p_s(x_0)$. The cross correlation thus identifies where the divergence is not uniform and compensates with appropriate perturbations.

The search for $\bar{s}$ as a focused (independent) random process allows us to estimate its perturbation $\bar{s}''$ sample by sample from $\bar{d}''$. Consider a sample d'' to be a reliable estimate of s'' (assume zero-mean noise, suppress subscripts), if its percentage error is less than c with greater than $1-e$ probability. Choose c and e as small fractions sufficient for one's purposes. Larger values will speed the inversion and allow more but less reliable events. A reliable estimate should satisfy (for positive d)

$$1 - e \le p[-cd \le s - d \le cd \mid d'' = d] \tag{17}$$

$$= \frac{\int_{-cd}^{cd} p_s(d - x)p_n(x)dx}{\int_{-\infty}^{\infty} p_s(d - x)p_n(x)dx}$$

Calculate this function once, then locate those samples of $\bar{d}''$ with satisfactory amplitudes. Zero all others as unreliable perturbations for $\bar{s}''$.

Approach the extraction of noise similarly. First subtract the most reliable signal and find the l.s. inverse of all remaining events. Assuming this inverse to contain only signal, over-estimate the m.p.f. for $\bar{s}'$ in the data domain. If one expects some coherence of noise over time, then smooth the extraction accordingly. Else redefine the noise as the convolution of two functions, as we shall illustrate for source waveform inversion.

Inversion of VSP's

We illustrate our methods for differential systems by defining the transformations appropriate to the 1D inversion of vertical seismic profiles (VSP's). Assume planar, compressional, elastic waves travel vertically in the earth, perpendicularly to the well. For geophones that measure the velocity of vertical displacement, we use the following differential equations.

$$\rho \frac{\partial^2 y}{\partial t^2} - \frac{\partial}{\partial z}[K \frac{\partial y}{\partial z}] = 0 \tag{18a}$$

$$K \frac{\partial y}{\partial z}|_{z=0} = g(t) \tag{18b}$$

$$y\,|_{t=0} = \frac{\partial y}{\partial t}|_{t=0} = 0 \tag{18c}$$

$y(z,t)$ gives the displacement as a function of depth and time; $\rho(z)$, $K(z)$, the density and bulk modulus; and $g(t)$, the source wave form at $z=0$. With the Neumann boundary conditions (18c), an explicit finite-difference scheme provides y from the signal parameters ρ, K, and g.

Now transform signal parameters into focused random processes. Adjacent depths give statistically dependent values of ρ and K because geologic formations tend to homogeneous packages. The derivatives

$$\rho' = \frac{\partial \rho}{\partial z} \;; \quad K' = \frac{\partial K}{\partial z}$$

however, recognize only transitions, which appear independently. Assume zero values (homogeneous earth) as a first estimate.

For $z=0$ at the surface, $g(t)$ should be a very short wavelet, unpredictable in frequency content, but with an indeterminate time shift, due to non-unique estimates of near surface rock parameters. For $z=0$ at the depth of the first geophone, z_0, $g(t)$ becomes this short wavelet convolved with all near surface multiples. For this latter case, define a source by the focused functions $w(s)$ and $h(t)$ where

$$g(t) = \int_0^{\Delta t} \left[\int w(s) e^{i 2\pi s t'} ds \right] h(t - t') dt' \tag{19}$$

For a sufficiently short Δt, the Fourier transform of the wavelet $w(s)$ becomes highly focused. A sufficiently large value focuses the multiples $h(t)$. (Definition (19) allows solution of the "spiking" deconvolution problem). To add slow time adaptability to the above, replace $w(s)$ by

$$\int_{-\Delta r}^{\Delta r} w(s,r) e^{i2\pi rt} dr$$

A good first estimate of $w(s)$ is the minimum phase wavelet for the average amplitude spectrum of the data traces.

We now require a linearization of the signal transformation

$$\delta\rho \frac{\partial^2 y}{\partial t^2} - \frac{\partial}{\partial z}[\frac{\partial y}{\partial z}\delta K] = -\rho \frac{\partial^2}{\partial t^2}\delta y + \frac{\partial}{\partial z}[K\frac{\partial}{\partial z}\delta y] \tag{20a}$$

$$-\delta K\frac{\partial y}{\partial z} + \delta g = K\frac{\partial}{\partial z}\delta y \tag{20b}$$

$$\delta y \,|_{t=0} = \frac{\partial}{\partial t}\delta y \,|_{t=0} = 0 \tag{20c}$$

All unperturbed functions must be taken as their previous estimates.

To descend to the l.s. inverse of this transformation by means of the gradient (7), we require the adjoint transformation of the error of previous estimates with the data,

$$\delta d(z_i,t) = d(z_i,t) - \frac{\partial}{\partial t}y(z_i,t) \tag{21}$$

Define this error to be zero elsewhere. The data, $\bar{d}$, are measured at roughly equal, but limited depth intervals. Use the calculus of variations and integration by parts. Solve for q as an intermediate result. We suppose the data to end at time T. Invert down to the lowest depth to produce a recorded reflection.

$$\rho \frac{\partial^2 q}{\partial t^2} - \frac{\partial}{\partial z}[K\frac{\partial q}{\partial z}] = \sum_{z_i}(\frac{\partial}{\partial t}\delta d)\,\delta(z - z_i) \tag{22a}$$

$$q \,|_{t=T} = 0 \;; \quad \rho \frac{\partial q}{\partial t}|_{t=T} = \sum_{z_i} (\delta d)\delta(z - z_i)\,|_{t=T} \tag{22b}$$

$$K \frac{\partial q}{\partial z}|_{z=0} = (\frac{\partial}{\partial t}\delta d)\,|_{z=0} \tag{22c}$$

$$\delta\rho = \frac{-1}{R^2}\int_0^T \frac{\partial^2 y}{\partial t^2} q\,dt \tag{23a}$$

$$\delta K = \frac{-1}{C^2}\int_0^T \frac{\partial y}{\partial z}\frac{\partial q}{\partial z}dt \tag{23b}$$

$$\delta g = \frac{1}{G^2} q \,|_{z=0} \tag{23c}$$

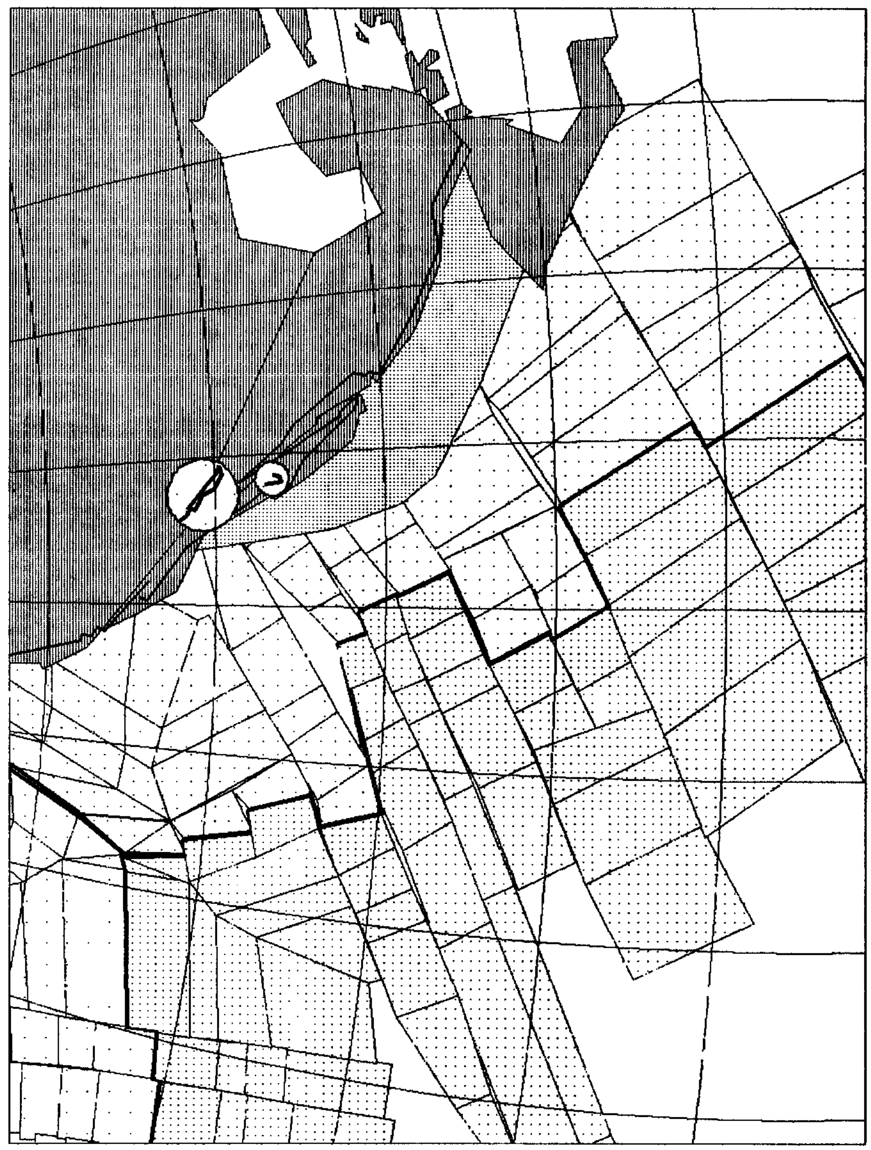
Perturb signal parameters $\delta\rho'$, $\delta K'$, δw, and δh directly from $\delta\rho$, δK, and δg. Add a broad gaussian constraint on signal (the first term in gradient (7)) to avoid wildly oscillating, unstable l.s. inversions of noise. Subsequently extracted signal perturbations may thus be more modest in amplitude, but will be more numerous because less inhibited by transformed noise. Subsequent extractions, of course, remain free to increase these perturbations.

REFERENCES

Harlan, W.S., J.F. Claerbout, and F. Rocca, 1983, Signal/noise separation and velocity estimation, SEP-37.

Lions, J.L., 1968, Controle optimal de systemes gouvernes par des equations aux derivees partielles: Paris, Dunod.

Mace, D., and P. Lailly, 1984, A solution of an inverse problem with the 1D wave equation applied to the inversion of vertical seismic profiles: submitted to the Sixth international conference on analysis and optimization of systems: Nice, Springer Verlag.



50 MIDDLE EOCENE 22