APPENDIX A
THE MAP ESTIMATE

This appendix derives objective functions for the estimation of signal from a sum of noise and transformed signal. This estimate finds the most probable values of signal and noise, provided that one knows their underlying probability distributions. Appendix B will explain how this estimate can be used as a first step in signal/noise separation.

Let us treat the data as a random process equal to the sum of noise and nonlinearly transformed signal.

$$d_i = f_i(\mathbf{s}) + n_i$$

$$\text{or } \mathbf{d} = \mathbf{f}(\mathbf{s}) + \mathbf{n} \tag{A.1}$$

s and **n** are independent, identically distributed (IID) random processes: individual random variables are statistically independent and stationary. Let us also assume that **s** and **n** have default values that give the most "predictable" outcome of individual random variables: an interpreter must feel that he can assume these default values unless he has information to the contrary. These values should show the simplest appearance and distract least from later perturbations. Definitions of the modeling transform must accommodate these assumptions.

Define geophysical noise as the untransformed component that shows no spatial coherence. If a component has enough coherence, then it should properly be defined as another variety of signal, with its own transform.

Let $p_{s_i}(\cdot)$ and $p_{n_i}(\cdot)$ be the corresponding marginal probability distribution functions (pdf's), defined by the following probability:

$$P[s_1 \leq s \leq s_2] = \int_{s_1}^{s_2} p_s(x)dx \quad .$$

x is a dummy variable. The probability of a random variable falling in a particular interval is equal to the integration of the pdf over that interval.

A maximum *a posteriori* (MAP) estimate of **s** maximizes the probability of the data **d**. Maximize the following conditional probability function.

$$\max_{\mathbf{s}} p_{\mathbf{s}\,|\,\mathbf{d}}(\mathbf{s} \mid \mathbf{d}) = \max_{\mathbf{s}} \prod_j \frac{p_{s_j}(s_j)\, p_{n_j}[d_j - f_j(\mathbf{s})]}{p_{d_j}(d_j)} \tag{A.2}$$

The denominator merely normalizes and does not affect a maximization.

Let us treat equation (A.2) as an objective function and place it in a form that can be optimized by a gradient algorithm. Because the logarithm increases monotonically for positive functions, we can equivalently minimize

$$\min_{\mathbf{s}} J(\mathbf{s}) = \min_{\mathbf{s}} \left\{ \sum_j \ln p_{s_j}(s_j) + \sum_j \ln p_{n_j}[d_j - f_j(\mathbf{s})] + constants \right\} . \qquad \text{(A.3)}$$

Specific forms of $p_s(\cdot)$ and $p_n(\cdot)$ might simplify the form of this objective function. For example, if the signal and noise pdf's are generalized Gaussians symmetric about zero (see equation [3.1]), then equation (A.3) yields a conventional objective function with L^p norms.

$$\min_{\mathbf{s}} \left\{ \beta_s^{-1} \sum_j | s_j |^{\alpha_s} + \beta_n^{-1} \sum_j | d_j - f_j(\mathbf{s}) |^{\alpha_n} \right\} \qquad \text{(A.4)}$$

Note that a least-squares norm results when a distribution is Gaussian (when α is 2).

Given an estimate $\mathbf{s}^0$ of the signal, I find the gradient of the objective function (A.3).

$$\frac{\partial J(\mathbf{s}^0)}{\partial s_i} = \frac{p_{s_i}'(s_i^0)}{p_{s_i}(s_i^0)} - \sum_j F_{ij}^0 \frac{p_{n_j}'[d_j - f_j(\mathbf{s}^0)]}{p_{n_j}[d_j - f_j(\mathbf{s}^0)]} \qquad \text{(A.5)}$$

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

Here, primes indicate differentiation.

Perturbing the model in the direction of the gradient will increase the probability of the model parameters, but not necessarily their reliability. In Appendices B and C, I shall explain which statistical assumptions are necessary for reliable estimates of signal and noise. If a non-zero perturbation is only marginally more probable than no perturbation, let us keep the original value and simplify the picture for the interpreter. By using the cross-entropy methods described in Appendix C, one can estimate $p_s(\cdot)$ and $p_n(\cdot)$ from independent sources, such as well logs. One should not allow these distributions, and thereby the objective function, to depend directly on the data being inverted; otherwise, the result might non-unique or unstable.