How geophysicists' intuition helps seismic data processing

Afshin Gholamy and Vladik Kreinovich

University of Texas at El Paso

Abstract

In geophysics, signals come with noise. It is desirable to minimize the effect of this noise. If we knew the probabilities of different values of signal and noise, we could use statistical filtering techniques. In geophysics, however, we rarely know the exact values of these probabilities; instead, we have to rely on the expertise and intuition of experts. We show how fuzzy techniques can transform this expertise into precise de-noising methods, we explain that the resulting methods indeed satisfy several natural requirements, and that these methods are in good accordance with heuristic techniques successfully used by geophysicists.

Keywords: fuzzy techniques, data processing, geophysicists' intuition, de-noising, Least Squares, Tikhonov regularization, smoothing, centroid defuzzification

1. Formulation of the Problem

Need for de-noising. In geophysics, signals come with noise; this noise affects the resulting images and maps. It is therefore desirable to minimize the effect of this noise.

De-noising: ideal case. In some practical situations, we know the probabilities of different values of signal and we know the probabilities of different values of noise. In such situations, we can use statistical filtering techniques to find optimal de-noising; see, e.g., [1].

Comment. Techniques for optimal de-noising are known since the 1940s when N. Wiener designed his famous Wiener filter [16].

Specifics of geophysics. In geophysics, however, we rarely know the exact values of these probabilities. One of the reasons is that, in contrast to many other disciplines, where we can determine the true values of different quantities by measuring them directly, in geophysics, it is very difficult (or even impossible) to directly measure the physical characteristics of the rocks at large depths, and these characteristics are what we are trying to determine.

Need for experts in geophysics. Since we do not know the corresponding probabilities, it is difficult to determine a priori which de-noising methods are most adequate for geophysical data. To select an appropriate de-noising method, we therefore have to use the expertise and intuition of expert geophysicists.

How geophysical expert knowledge is used now. At present, the knowledge of geophysical experts is mainly used to select the most physically reasonable Earth model among all the models which are consistent with the observations.

If some semi-heuristic method consistently leads to models which are – according to the experts – more physically adequate than models generated by other methods, then this method is adapted by geophysicists. If some semi-heuristic method does not always result in adequate models, then geophysicists try to modify the original methods.

It is desirable to directly translate expert knowledge into data processing techniques. The current trial-and-error approach to designing geophysical data processing techniques has led to many successful applications, ranging from successful discovery of mineral deposits to description of risks in earthquake-prone areas.

However, like most trial-and-error search-in-the-dark processes, this approach is very time-consuming – and, honestly, somewhat frustrating from researchers who are designing new methods. It is desirable to come up with faster (and less frustrating) ways to translate expert knowledge into data processing techniques.

What we do in this paper. In this paper, we show that fuzzy techniques [6, 11, 17] – techniques specially designed to translate imprecise ("fuzzy") expert knowledge into precise algorithms – can help translate geophysical expert knowledge into relevant data processing techniques.

We provide two types of arguments that the resulting techniques are indeed reasonable:

- we show that the results of applying these techniques satisfy reasonable requirements, and
- we also show that these techniques are in good accordance with the empirically successful semi-heuristic geophysical methods.

2. Statistical De-Noising: Successes and Limitations

Need for de-noising: a general description. In many practical situations, we want to know how

a certain physical quantity x changes with time (or, more generally, how the values of the corresponding quantity changes with time and with a spatial location). In many such cases, we can directly measure the values $x(t)$ of this quantity at different moments of time.

For example, in meteorology, we measure the temperature, wind speed and direction, etc., at different moments of time. In geophysical seismic data processing, we measure, at different moments of time, the displacement caused by the seismic waves.

Measurements are never absolutely accurate. As a result, the measured values $\widetilde{x}(t)$ are, in general, somewhat different from the actual (unknown) values $x(t)$: $\widetilde{x}(t) = x(t) + \Delta x(t)$, where $\Delta x(t) \stackrel{\text{def}}{=} \widetilde{x}(t) - x(t) \neq 0$ is the measurement uncertainty (a.k.a. noise).

It is therefore desirable to decrease the effect of the noise, i.e., to recover the actual signal $x(t)$ as accurately as possible.

Case when we know the probability distributions. In some situations, we have many records of the actual signals, of the measured values, and/or of the values of measurement uncertainty. In such situations, we can use these records to determine the probabilities of different values of signal $x(t)$ and noise $\Delta x(t)$.

In such situations, when we have the measured values $\widetilde{x}(t)$, we can estimate the probability of different values of the actual signal $x(t)$. Based on this information, we would like to find an estimate $X(t)$ which is the closest to the actual value, i.e., for which the difference $e(t) \stackrel{\text{def}}{=} X(t) - x(t)$ between the reconstructed and actual signals is close to 0 for all t.

Both the original signal and the reconstructed signal can be naturally represented as points in multi-D space, for which the corresponding value $X(t)$ or $x(t)$ is the t-th coordinate. We want $X(t)$ to be close to $x(t)$; in geometric terms, this means that we want the corresponding points $X(t)$ and $x(t)$ to be close.

How do we describe closeness. In general, the distance between the two points $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in n-dimensional space is equal to $d(x, y) = \sqrt{(x_1 - y_1)^2 + \ldots + (x_n - y_n)^2}$. Here, strictly speaking, we have infinitely many coordinates, so we need to take the limit of the corresponding sun $\sum\limits_t (X(t) - x(t))^2$, i.e., the integral:

$$d(X, x) = \sqrt{\int (X(t) - x(t))^2 \, dt}.$$

In effect, this distance is what we would like to minimize.

From minimizing distance to the Least Squares approach. Strictly speaking, it is possible to start minimizing the distance directly. However, from the computational viewpoint, this direct minimization has a problem: most traditional optimization techniques are based on computing the derivatives. For example, a straightforward calculus-based minimization of a function means that we take the partial derivatives of the minimized function with respect to all the unknowns, and equate all these partial derivatives to 0 (since, according to calculus, at minimum and maximum points, the function's derivative is equal to 0). The problem is that for the square root $f(x) = \sqrt{x}$, the derivative at $x = 0$ (where we want to be) is infinite.

So, from the computational viewpoint, it is desirable to avoid the square root when formulating the objective function. In minimization, the square root is easy to eliminate. Indeed, since the square root is an increasing function, minimizing the square root of some expression is equivalent to minimizing the expression itself. Thus, instead of minimizing the square root, we can minimize the integral $\int (X(t) - x(t))^2 \, dt$. Minimization of the sum of the squares of approximation errors – known as the *Least Squares* method – is indeed one of the most widely used ideas in signal processing; see, e.g., [14].

The Least Squares method can also be justified by the fact that the purpose of the signal reconstruction is to minimize the loss $L(X(t) - x(t))$ caused by the inaccuracy in reconstructing the signal. Since the inaccuracy is usually small, we can expand the loss function into Taylor series and keep only the first few terms in this expansion:

$$L(\Delta x(t)) = c_0 + \sum_t c(t) \cdot \Delta x(t) +$$

$$\sum_t \sum_{t'} c(t, t') \cdot \Delta x(t) \cdot \Delta x(t').$$

This loss is equal to 0 when $\Delta x(t) = 0$; thus, we should have $c_0 = 0$ and $c(t) = 0$ for all t. Thus,

$$L(\Delta x(t)) = \sum_t \sum_{t'} c(t, t') \cdot \Delta x(t) \cdot \Delta x(t').$$

If $c(t, t') \neq 0$ for some $t \neq t'$, this would mean that for some combination of signs of the inaccuracies $\Delta x(t)$ and $\Delta x(t')$ (same sign if $c(t, t') < 0$ and different signs if $c(t, t') > 0$), the presence of both inaccuracies would decrease the overall loss. This may be the case in some specific situations, but in general, this seems counter-intuitive. It is therefore reasonable to require that $c(t, t') = 0$ when $t = t'$. Thus, the loss formula takes the simplified form

$$L(\Delta x(t)) = \sum_t c(t, t) \cdot (\Delta x(t))^2.$$

In general, we have no prior reason to believe that the loss at one moment of time is more damaging than a similar loss at another moment of time. Thus, it makes sense to assume that the coefficient

$c(t,t)$ does not depend on time t, i.e., that the loss is simply proportional to the sum

$$\sum_t (\Delta x(t))^2,$$

or, in the continuous case, to the corresponding integral $\int (X(t)-x(t))^2\,dt$.

Need to minimize the expected value. According to decision making theory (see, e.g., [7, 10, 12]), an appropriate decision is the one that minimizes the expected value of the utility – or, equivalently, minimizes the expected value of the loss. Thus, it is reasonable to look for an estimate $X(t)$ for which the expected value $E[\int (X(t)-x(t))^2\,dt]$ of the above integral attains the smallest possible value.

Once we know all the probability distributions, this minimization is a well-defined optimization problem, so we can find the corresponding estimate $X(t)$.

In many reasonable cases, we have an explicit formula for de-noising. In the general case, optimization may be complicated, but in many cases, we have explicit formulas for optimal de-noising; see, e.g., [1].

These formulas are made possible by the fact that the measurement uncertainty usually comes from many different independent factors. It is known that the distribution of the joint effect of many independent factors is close to Gaussian (normal); this is known as the Central Limit Theorem (see, e.g., [14]). Thus, it is reasonable to assume that the measurement uncertainty $\Delta x(t)$ is normally distributed.

Similarly, in many situations, the signal $x(t)$ is caused by the joint effect of several different factors. For example, the seismic signal comes from the movement of different rocks which are reasonably independent. Thus, it is also reasonable to assume that the distribution of the signal is Gaussian.

It also makes sense to assume that the signal and the measurement uncertainty are independent – since they are caused by different causes.

In many practical applications, the probability distributions for $x(t)$ and $\Delta x(t)$ are *stationary* – i.e., the corresponding probabilities do not change with time. It is known that stationary process are convenient to describe in terms of their Fourier transforms

$$\hat{x}(\omega) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi}} \cdot \int x(t) \cdot \exp(\mathrm{i}\cdot t\cdot\omega)\,dt \text{ and}$$

$$\widehat{\Delta x}(\omega) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi}} \cdot \int \Delta x(t) \cdot \exp(\mathrm{i}\cdot t\cdot\omega)\,dt:$$

namely, the Gaussian stationary process with 0 mean is uniquely determined by the corresponding *spectra* $S_{x,x}(\omega) \stackrel{\text{def}}{=} E\left[|\hat{x}(\omega)|^2\right]$ and $S_{\Delta x,\Delta x}(\omega) \stackrel{\text{def}}{=} E\left[\left|\hat{\Delta x}(\omega)\right|^2\right]$. In terms of the spectra, the Fourier transform $\hat{X}(\omega)$ of the optimal estimate $X(t)$ has the form

$$\hat{X}(\omega) = \frac{\hat{\widetilde{x}}(\omega)}{1+\dfrac{S_{\Delta x,\Delta x}(\omega)}{S_{x,x}(\omega)}}.$$

Comment. This formula was first derived by N. Wiener, the father of cybernetics (see, e.g., [16]) and is, therefore, known as the *Wiener filter*.

Wiener filter and similar formulas have many successful applications. As shown in [1], optimal statistical de-noising formulas are ubiquitous, they are efficiently used in all areas of signal processing.

Limitations of the optimal statistical de-noising. The formulas for the optimal de-noising assume that we know the corresponding probabilities. In many practical situations, however, we do not know these probabilities.

Geophysics is one of such areas, since, as we have mentioned earlier, in contrast to many other disciplines where, in principle, we can directly and accurately measure the actual values of the physical quantities of interest, in geophysics, it is often impossible to directly measure the desired physical characteristics.

To handle such situations, it is therefore important to use methods of de-noising under uncertainty, when we do not have full information about the corresponding probabilities.

3. De-Noising under Uncertainty, Case of Smooth Signals: Existing Algorithms, Successes, and Limitations

Case of smooth signals: description. In many practical situations, while we do not know the corresponding probabilities, we do know that the actual signal $x(t)$ is smooth.

In contrast, the measurement errors $\Delta x(t)$ corresponding to different moments of time t are usually independent random variables, as a result of which the dependence of $\Delta x(t)$ on time t is not smooth.

De-noising: idea. The signal is smooth $x(t)$ and close to the observed signal $\widetilde{x}(t)$. There may be other smooth functions $y(t)$ which are close to $\widetilde{x}(t)$. However:

- For the actual smooth signal $x(t)$, there is only one reason why this signal is different from the measured signal $\widetilde{x}(t)$: the presence of the noise.
- In contrast, for a smooth function $y(t) \neq x(t)$, there are two reasons why $y(t) \neq \widetilde{x}(t)$: the presence of the noise and the difference between $x(t)$ and $y(t)$.

We therefore expect that the distance $d(\widetilde{x}, x)$ between the observed signal $\widetilde{x}(t)$ and the actual signal $x(t)$ is smaller than the distance $d(\widetilde{x}, y)$ between the

observed signal $\widetilde{x}(t)$ and any different smooth function $y(t) \neq x(t)$.

Thus, we can find the actual signal as the smooth function $X(t)$ which is the closest to the observed signal $\widetilde{x}(t)$.

How this idea is usually formalized: Tikhonov regularization. We already know how to describe that the functions $x(t)$ and $X(t)$ are close: by using the integral $\int (X(t) - \widetilde{x}(t))^2\,dt$. So, to describe the above idea in precise terms, it is sufficient to come up with an exact definition of a smooth function.

Intuitively, a function $X(t)$ is smooth if all the values of its derivative $\dot{X}(t)$ are reasonable small. In other words, we can say that a function $X(t)$ is smooth if the vector consisting of all the values $\dot{X}(t)$ is sufficiently close to the vector consisting of all zeroes. The corresponding distance can be described as the square root of the interval $\int (\dot{X}(t))^2\,dt$; thus, a bound b on this distance is equivalent to a bound on this integral $\int (\dot{X}(t))^2\,dt \le b^2$.

In these terms, the above idea means that we find the signal $X(t)$ for which the integral

$$\int (X(t) - \widetilde{x}(t))^2\,dt$$

attains its smallest possible value under the constraint $\int (\dot{X}(t))^2\,dt \le b^2$. In general, the minimum under the inequality constraint is attained:

- either inside the constraint region, i.e., in this case, when $\int (\dot{X}(t))^2\,dt < b^2$,
- or at this region's boundary $\int (\dot{X}(t))^2\,dt = b^2$.

If the minimum was attained inside, this would mean that we have a global minimum, the constraint could be ignored. However, the global minimum of the non-negative integral $\int (X(t) - \widetilde{x}(t))^2\,dt$ (0 value) is attained when $X(t) = \widetilde{x}(t)$ $(= x(t) + \Delta x(t))$, and this function $\widetilde{x}(t)$ is not smooth.

Thus, the minimum cannot be attained inside, it has to be attained at the boundary $\int (\dot{X}(t))^2\,dt = b^2$. So, our optimization problem is equivalent to minimizing the integral $\int (X(t) - \widetilde{x}(t))^2\,dt$ under the constraint $\int (\dot{X}(t))^2\,dt = b^2$. Due to the Lagrange multiplier idea, this is equivalent to an unconstrained minimization of a functional

$$\int (X(t) - \widetilde{x}(t))^2\,dt + \lambda \cdot \int (\dot{X}(t))^2\,dt$$

for an appropriate Lagrange multiplier λ. This formulation is known as *Tikhonov regularization*; see, e.g., [15].

Tikhonov regularization: successes. Tikhonov regularization is easy to implement: for example, in terms of Fourier transforms, its solution can be described explicitly, as

$$\hat{X}(\omega) = \frac{\hat{\widetilde{x}}(\omega)}{1 + \lambda \cdot |\omega|^2}.$$

The computational efficiency of regularization is one of the reasons why regularization has been successfully used in many application areas [15].

Tikhonov regularization: first limitation. Tikhonov regularization works reasonably well when we know how smooth is the original signal, i.e., equivalently, what is the value λ. In many practical situations, however, we do not have a priori information about this smoothness. As a result, sometimes, after using Tikhonov regularization to smoothen the measured signal, we realize that the resulting signal is not smooth enough, and an additional smoothing is needed.

Intuitively, we expect that if the original smoothing was not sufficient, then we just need to apply, to an under-smoothed signal, an additional smoothing of the same type. Alas, this is not the case: if we first apply Tikhonov smoothing with some λ, and then, to the result $\hat{X}(\omega)$ of this smoothing, we apply a similar smoothing with some other parameter $\Delta\lambda$, we get a signal

$$\hat{S}(\omega) = \frac{\hat{\widetilde{x}}(\omega)}{(1 + \lambda \cdot |\omega|^2) \cdot (1 + \Delta\lambda \cdot |\omega|^2)}$$

which is *different* from the expected signal

$$\hat{S}(\omega) = \frac{\hat{\widetilde{x}}(\omega)}{1 + \lambda' \cdot |\omega|^2}$$

corresponding to some appropriate value λ': simply because the product of two quadratic expression $(1 + \lambda \cdot |\omega|^2) \cdot (1 + \Delta\lambda \cdot |\omega|^2)$ is not itself equal to a quadratic expression $1 + \lambda' \cdot |\omega|^2$, no matter what value λ' we select; see, e.g., [3, 9].

Tikhonov regularization: second limitation. The second limitation is even more fundamental: Tikhonov regularization assumes that the actual signal is smooth. In many practical applications, this assumption makes sense.

However, in geophysics, there are real abrupt transitions between different layers of rocks. As a result, recorded geophysical signals are often not smooth. Smoothing the signal smoothes the actually abrupt transition – i.e., makes the resulting model less realistic.

What we plan to do. In this paper, we show that expert knowledge can help us overcome both limitations of Tikhonov regularization.

4. From Semi-Heuristic Ideas of Tikhonov Regularization to Smoothing Motivated by Expert Knowledge

What is smoothing? Let us describe this problem in intuitive terms. Instead of trying to formulate smoothing as a complex optimization problem, let us instead try to describe simple rules describing smoothing.

We want to come up with a smooth function $X(t)$ which is close to measured values $\widetilde{x}(t)$. Again, instead of trying to describe closeness in complex mathematical terms, let us simply describe this requirement as an imprecise (fuzzy) relation

$$X(t) \approx \widetilde{x}(t).$$

We have not yet described what smoothness mean here; let us now try to describe smoothness in similar simple terms. Intuitively, a smooth function is the one for which the values at nearby points are close to each other: if t is close to t', then $X(t) \approx X(t')$. This is simple to describe, but not as simple to solve, since, in contrast to the relation $X(t) \approx \widetilde{x}(t)$ that relates the unknown $X(t)$ directly to the measured value $\widetilde{x}(t)$, the property $X(t) \approx X(t')$ relates two unknown values $X(t)$ and $X(t')$ to each other. So, to make it easier to use this constraint when solving a problem, it is desirable to reformulate these constraints in a more direct terms. This can be done if we take into account that from $X(t) \approx X(t')$ and $X(t') \approx \widetilde{x}(t')$, we can conclude that $X(t) \approx \widetilde{x}(t')$.

Thus, the intuitive meaning of smoothing is captured by the following general rule:

$$\text{if } x' \text{ is close to } x, \text{ then } X(t) \approx \widetilde{x}(t').$$

From the intuitive description to exact formulas. For each moment t, once we fix a membership function $\mu(t, t')$ that describes closeness, the above rules leads to different values $X(t) = \widetilde{x}(t')$ with different degrees $\mu(t, t')$. In other words, we get a membership function that describes possible values of $X(t)$. We would like to select a single estimate from this membership function. For this "defuzzification", let us use *centroid defuzzification*, the most widely used defuzzification procedure [6, 11], according to which we select the value

$$X(t) = \frac{\int \mu(t, t') \cdot \widetilde{x}(t)\, dt'}{\int \mu(t, t')\, dt'}.$$

How do we describe closeness? In [8], we have shown that, under reasonable assumptions, a natural way to describe closeness is by using a Gaussian membership function, which, in the isotropic case, has the form $\mu(t, t') = \exp\left(-\dfrac{(t - t')^2}{2\sigma^2}\right)$ for some σ. For this function, the integral $\int \mu(t, t')\, dt'$ does not depend on t and is, thus, equal to a constant. Thus, after the corresponding smoothing, the value $X(t)$ is equal to $X(t) = C \cdot \int \widetilde{x}(t') \cdot \exp\left(-\dfrac{(t - t')^2}{2\sigma^2}\right)$.

In terms of Fourier transforms, this smoothing takes the form $\hat{X}(\omega) = \hat{\widetilde{x}}(\omega) \cdot \exp\left(\frac{1}{2} \cdot \sigma^2 \cdot \omega^2\right)$.

The resulting expert-motivated smoothing overcomes the first limitation of Tikhonov regularization. The first limitation of Tikhonov's regularization is that when we apply an additional smoothing to an already smoothed signal, we do not get the same result as when we performed more smoothing from the very beginning.

For the new smoothing procedure, there is no such limitation: if we apply several such smoothings one after another, the joint effect of all these smoothings is equivalent to a single smoothing of the same type. This is the easiest to show in terms of Fourier transforms: every smoothing is equivalent to multiplying a signal by an expression $\exp(-a \cdot \omega^2)$ for some constant a (which is equal to $\sigma^2/2$). If we apply several smoothings, corresponding to values $a_1, \ldots, a_n$, then as a result of these smoothings, the original Fourier transform is multiplied by the product $\prod_{i=1}^{n} \exp(-a_i \cdot \omega^2)$ of such terms, and one can easily see that this product is equal to a $\exp(-a \cdot \omega^2)$, where $a \stackrel{\text{def}}{=} \sum_{i=1}^{n} a_i$, i.e., to a single smoothing of the same type.

Moreover, the expert-motivated smoothing procedure is the only one that is free from this limitation. Let us show that not only the expert-motivated smoothing is free from the above first limitation, it is also *the only* possible smoothing procedure which is free from this limitation.

Indeed, if a smoothing procedure is free from this limitation, this means that any smoothing can be represented as a composition of many small smoothings. Each shift-invariant smoothing has the form $x(t) \to \int k(t' - t) \cdot x(t)\, dt$ for an appropriate function $k(t' - t)$. When we apply two smoothings one after another, this is equivalent to applying a smoothing $x(t) \to \int k_2(t' - t) \cdot x(t)\, dt$, where $k_2(t) = \int k(s) \cdot k(t - s)\, ds$ is a convolution of two copies of the function $k(t)$ corresponding to a single smoothing. Similarly, n consequent smoothings are equivalent to applying a function $k_n(t)$ which is a convolution of n copies of the function $k(t)$.

Interestingly, the same convolution also describes how the probability density function (pdf) $k(t)$ of a single random variable ζ get transformed into a pdf for the sum $\zeta_1 + \ldots + \zeta_n$ of several independent variables distributed according to the same distribution with pdf $k(t)$. We have already mentioned that, according to the Central Limit Theorem, the distribution of the sum of many independent random variables is close to Gaussian – and the more variables we add, the closer to the Gaussian distribution we get.

The pdf of the Gaussian distribution has the form $C \cdot \exp(-a \cdot t^2)$. Thus, the convolution $k_n(t)$ of many functions $k_n(t)$ has the same Gaussian form – and the convolution with a Gaussian function is exactly our expert-based smoothing.

Comment. From the mathematical viewpoint, the above argument is similar to the one given in [13] for a somewhat different problem related to geophysical data processing.

The resulting expert-based smoothing operation is in line with heuristic geophysical smoothing techniques. It is worth mentioning that our expert-based smoothing is close to the smoothing successfully used in geophysics under the name of *shape regularization* [3, 9]. For example, the paper [3] shows, on the example of the Gulf of Mexico data, that the new smoothing leads to a more adequate geophysical description than the more traditional Tikhonov smoothing.

5. What If Signals Are Sometimes Not Smooth?

Need to consider possibly non-smooth signals: reminder. Many geophysical structures have abrupt boundaries. As a result, the corresponding signals are not always smooth: they have abrupt transitions corresponding to these boundaries.

How such signals are processed now: splines. In mathematical terms, the corresponding signals are not always smooth, they are sometimes piece-wise smooth. To describe such signals, it is thus natural to use piece-wise smooth models (*splines*).

Splines have indeed been efficiently used in data processing, in particular, in seismic data processing.

Limitations of the splines approach. While splines are often successful, in some situations, splines do not fully reflect our understanding of the geophysical phenomena.

Indeed, general splines simply assume that there is a non-smooth transition, without explaining what type of a transition this is. A more adequate description should take into account not only the specifics of the smooth part of the signal, but also specifics of the corresponding transitions.

How to take geophysical specifics into account: finite-parametric models. In general, specifics means that instead of considering all possible smooth (or piece-wise smooth) dependencies, we have a family of models $m(t, \vec{c})$ depending on some parameter(s) $\vec{c} = (c_1, \ldots, c_k)$, models that describe the corresponding phenomenon.

For example, a linear (and, more generally, smooth dependence can be locally well described by a linear model $x(t) \approx m(t) = c_1 + c_2 \cdot t$, a wave can be locally described by a sinusoid $x(t) \approx m(t) = c_1 \cdot \cos(c_2 \cdot t + c_3)$. There are also finite-parametric models for such non-smooth phenomena as phase transitions, etc.

The corresponding models are usually local. In all the above cases, a model with a finite number of parameters is only an approximation; the longer the period of time that we need to cover, the less accurate the model becomes. Thus, to provide a more adequate description, a reasonable idea is to use each such model only locally, i.e., to have several (n) models each of which covers some short time period.

In this approach, instead of k parameters, we need $k \cdot n$ parameters.

Traditional approach to combining local models: use of splines. The more parameters we need, the more computational time and computational resources we need. As a result, in the past, when these computer resources were limited – so that routine geophysical computations required computers much more powerful that a usual PC – practitioners tried to limit the number of models n, i.e., in effect, go back to piece-wise models.

Limitations of the splines approach. The resulting spline-type models were a good fit, but at the transitions between the n regions, we have deviations from smoothness which were not corresponding to anything geophysically meaningful.

Let us explain this problem on an example of approximation a smooth function by linear models. Each smooth function $x(t)$, in the vicinity of each point t_0, can be well approximated by a linear function: $x(t) \approx m(t) = x(t_0) + \dot{x}(t_0) \cdot (t - t_0)$. The further t from t_0, the less accurate the corresponding approximation. To come up with an approximation which is accurate for all moments t, we thus need to use different linear approximations on different time intervals: first $m(t) = x(t_0) + \dot{x}(t_0) \cdot (t - t_0)$, then $m(t) = x(t_1) + \dot{x}(t_1) \cdot (t - t_1)$, etc. This is OK, but at the border between two different intervals, the derivative of the approximating function $m(t)$ changes abruptly – e.g., from $\dot{x}(t_0)$ to $\dot{x}(t_1)$, while the original signal $x(t)$ may have been very smooth, with a differentiable first derivative $\dot{x}(t)$.

New approach: use of local attributes. Nowadays, the computers are more powerful, so it has become possible to use numerous local models. It has even been computationally possible to consider, for each moment t_0, a model $x(t, \vec{c}(t_0))$ which is the best fit for some neighborhood of t_0 [2, 5, 4].

This approach is consistent with the geophysicists' intuition: for example, a geophysicist can meaningfully talk about a local frequency and amplitude of a wave, and how these quantities change as a wave changes.

This approach has led to interesting practical results. For example, in [2, 5] this approach helped to enhance oil recovery by providing a clearer picture of how oil migrated in the (natural) reservoir rock (where oil accumulates before going into the production well).

How to translate this general idea into a precise algorithm? The papers [2, 5] use a semi-heuristic approach to process the corresponding models. Let us show that – similarly to smoothing – these semi-heuristic techniques can be explained as a natural formalization of the expert rules.

Let us therefore describe the natural expert's ideas. In situations when a signal is not necessarily smooth, we cannot claim that the dependence $x(t)$ is smooth. However, it is still reasonable to claim that when we have two local models $m(t, \vec{c}(t_0))$ and $m(t, \vec{c}(t_0'))$ corresponding to two nearby moments of time $t_0' \approx t_0$, then the corresponding parameters $\vec{c}(t_0)$ and $\vec{c}(t_0')$ should be close. In other words, the dependence $\vec{c}(t_0)$ of the corresponding parameters on the time t_0 should be smooth.

Thus, a natural idea is to apply the above-described expert-based smoothing to the dependence $\vec{c}(t_0)$. To be more precise:

- first, we determine, for each t_0, the values $\vec{c}(t_0)$ that provide the best fit for the observed signal $\widetilde{x}(t)$ in the appropriate neighborhood of the point t_0;
- then, we apply the expert-based (Gaussian) smoothing to the resulting dependence $\vec{c}(t_0)$.

This is, in effect, how the successful practical applications were obtained in [2, 5].

Additional comment. Let us show that not only the above idea naturally comes from the expert knowledge, but that this idea is in good accordance with the general Least Squares approach.

For simplicity, let us first consider the case when the corresponding model $m(t, c)$ has only one parameter c. In this case, for each moment t_0, we find the value $c(t_0)$ that provides, on average, the best fit for all the values from the neighborhood, i.e., the value for which the integral $\int (\widetilde{x}(t) - m(t, c(t_0)))^2\, dt$ is the smallest possible, where integration is over all the points t which are neighbors to t_0. We need to solve such a minimization problem for all t_0, i.e., we need to solve many optimization problems $\int (\widetilde{x}(t) - m(t, c(t_0)))^2\, dt \to \min\limits_{c(t_0)}$ corresponding to different value t_0.

The integral minimizes in each of these problems depends only on the value $c(t_0)$ and does not depend on the values $c(t_0')$ corresponding to all other moments t_0'. Thus, instead of considering many independent minimization problems, we can simply minimize the sum

$$J \stackrel{\text{def}}{=} \int dt_0 \cdot \int (\widetilde{x}(t) - m(t, c(t_0)))^2\, dt.$$

This sum is the smallest if and only each of its terms $\int (\widetilde{x}(t) - m(t, c(t_0)))^2\, dt$ is the smallest.

To clarify what J is equal to, let us re-arrange the integration in the expression J. In the original expression, J is the sum of expressions corresponding to different values t_0. For each such value, we consider the values $\widetilde{x}(t)$ corresponding to all neighboring moments t. Instead, let us group together all the terms containing the given value $\widetilde{x}(t)$. As a result, we get the following equivalent expression for the minimized expression J:

$$J = \int dt \cdot \int (\widetilde{x}(t) - m(t, c(t_0)))^2\, dt_0,$$

where the internal integration is over all moment t_0 which are close to t.

We expect the local model $m(t, c(t_0))$ to be smoothly changing when t_0 changes. Thus, when t_0 is close to t, we have

$$m(t, c(t_0)) \approx m(t, c(t)) + \frac{\partial m}{\partial c} \cdot \dot{c}(t) \cdot (t_0 - t).$$

So,

$$(\widetilde{x}(t) - m(t, c(t_0)))^2 \approx (\widetilde{x}(t) - m(t, c(t)))^2 +$$

$$\langle \text{terms linear in } t - t_0 \rangle + w(t) \cdot (\dot{c}(t))^2 \cdot (t_0 - t)^2,$$

where $w(t) \stackrel{\text{def}}{=} \left(\dfrac{\partial m}{\partial c} \right)^2$. The integral of the linear term $t - t_0$ over a symmetric neighborhood is 0. The integral of the quadratic term $(t - t_0)^2$ is a constant c_0. Thus,

$$J = \int (\widetilde{x}(t) - m(t, c(t)))^2\, dt + c_0 \cdot \int w(t) \cdot (\dot{c}(t))^2\, dt.$$

In effect, this means that, in addition to the Least Squares best-fit first term, we also have a regularization-type second term describing the smoothness of the dependence of parameters $c(t)$ on time t – this is exactly the idea from [2, 5].

Acknowledgments

This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721.

The authors are thankful to Sergey Fomel for useful suggestions and for his encouragement.

References

[1] R. G. Brown and P. Y. C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*, John Wiley and Sons, New York, 1996.

[2] S. Fomel, Local seismic attributes, *Geophysics*, 72(3):A29–A33, Society of Exploration Geophysicists, 2007.

[3] S. Fomel, Shaping regularization in geophysical-estimation problems, *Geophysics*, 72(2):R29–R36, Society of Exploration Geophysicists, 2007.

[4] S. Fomel, Adaptive multiple subtraction using regularized nonstationary regression, *Geophysics*, 74(1):V25–V33, Society of Exploration Geophysicists, 2009.

[5] S. Fomel and L. Jin, Time-lapse image registration using the local similarity attribute, *Geophysics*, 74(2):A7–A11, Society of Exploration Geophysicists, 2009.

[6] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, Upper Saddle River, New Jersey, 1995.

[7] V. Kreinovich, Decision Making under Interval Uncertainty (and beyond), In: P. Guo and W. Pedrycz (eds.), *Human-Centric Decision-Making Models for Social Sciences*, pages 163–193, Springer Verlag, Berlin, Heidelberg New York, 2014.

[8] V. Kreinovich, C. Quintana, and L. Reznik. Gaussian membership functions are most adequate in representing uncertainty in measurements, In: *Proceedings of NAFIPS'92: North American Fuzzy Information Processing Society Conference*, Puerto Vallarta, Mexico, December 15–17, 1992, NASA Johnson Space Center, Houston, TX, 1992, pages 618–625, 1992.

[9] Y. Liu and S. Fomel, Seismic data analysis using local time-frequency decomposition, *Geophysicsl Prospecting*, 61:516–525, European Association of Geoscientists and Engineers, 2013.

[10] R. D. Luce and R. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.

[11] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2006.

[12] H. Raiffa, *Decision Analysis: Introductory Lectures on Choices Under Uncertainty*, McGraw-Hill, New York, 1997.

[13] C. Servin, A. Velasco, and V. Kreinovich, How to estimate relative spatial resolution of different maps or images of the same area? In: *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics SMC'2014*, San Diego, California, October 5–8, 2014, pages 3507–3511.

[14] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC Press, Boca Raton, Florida, 2011.

[15] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, V. H. Winston & Sons, Washington, DC, 1977.

[16] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, John Wiley and Sons, New York, 1949.

[17] L. A. Zadeh, Fuzzy sets, *Information and Control*, 8:338–353, Elsevier, 1965.