Dual Regularization and Coherent Noise Suppression in Velocity Analysis

William. W. Symes *

ABSTRACT

Dual regularization updates data along with velocity in a velocity analysis procedure based on differential semblance optimization. Under certain conditions the data update so produced suppresses coherent noise, and the corresponding velocity update flattens primary reflection energy. The chief of these conditions are: (1) space and time sampling must be adequate to avoid aliasing; (2) data frequency content must be high enough to resolve residual moveout between primaries and noise; (3) the primary reflection series must be the most energetic event family which can be flattened by a velocity in the admissible class. In some cases reflection seismograms appear to conform closely enough to these conditions that dual regularization is effective, essentially without any preprocessing; in other cases they do not, and preprocessing such as multiple suppression is prerequisite to a successful dual regularization velocity analysis.

Introduction

The differential semblance ("DS") approach to velocity analysis, introduced in (Symes, 1986), determines wave velocity by minimizing a functional of velocity and data, which may be written as

$$\min_v J_0[v,d], \;\; J_0[v,d] = \frac{1}{2}\|\partial G[v]d\|^2 \tag{1}$$

in which $G[v]$ is a migration or asymptotic inversion operator, producing a prestack reflectivity volume $r = G[v]d$, binned in some fashion (common offset, shot, scattering angle). The operator ∂ is a derivative in the bin direction, or an approximation to it. The velocity v is supposed to range over a set of admissible velocities; these are in most applications splines or the like, constrained by upper and lower bounds and perhaps slope bounds -

*The Rice Inversion Project, Department of Computational and Applied Mathematics, Rice University, Houston TX 77251-1892 USA, email `symes@caam.rice.edu`

the details are unimportant for the present discussion. The norm $\|\cdot\|$ is a variant of L^2 (RMS energy).

This basic definition of DS may be enhanced in various ways to deal with multicomponent data, multiple bin parameters, multiple wave modes, etc. - see (Symes, 2001a) for an overview of these and other formulation and implementation issues.

When the data consists predominantly of primary reflections (i.e. is approximately consistent with a linearized model of scattering, with a slowly varying background and oscillatory perturbations), DS appears to produce reliable velocity estimates, within the limitations imposed by its underlying theory (Symes and Carazzone, 1991; Symes and Versteeg, 1993; Kern and Symes, 1994; Araya et al., 1996; Symes, 1997; Minkoff and Symes, 1997; Chauris et al., 1998; Chauris and Noble, 2001; Mulder and ten Kroode, 2001). There is even some theoretical evidence that critical points produced by local optimization are global solutions, at least in some cases, provided that the signal-to-noise ratio is sufficiently high (Symes, 1991; Symes, 1999; Symes, 2001b). Coherent noise is known to strongly affect the output of DS, however: the algorithm finds moveout, and has no means to distinguish primary reflection moveout from coherent noise moveout (Mulder and ten Kroode, 2001; Gockenbach and Symes, 1999).

Mark Gockenbach's PhD thesis (Gockenbach, 1994; Gockenbach et al., 1995) introduced the concept of "dual regularization", as a means of analyzing the influence of noise on DS optimization. Dual regularization replaces (1) by a family of constrained optimization problems parametrized by $\sigma, 0 \le \sigma \le 1$:

$$\min_{v,e} J_0[v, d+e], \text{ subject to } \|e\| \le \sigma \|d\| \tag{2}$$

That is, the data is perturbed as well, by an amount regulated by the noise level parameter σ.

Another way of writing the problem (2) is: minimize over v the functional $J_\sigma[v, d]$ defined by

$$J_\sigma[v, d] = \min_e J_0[v, d+e], \text{ subject to } \|e\| \le \sigma \|d\| \tag{3}$$

Clearly the notation is consistent: when $\sigma = 0$ we obtain J_0 as defined in (1). On the other hand, for $\sigma = 1$ it's equally clear that $J_1[v, d] \equiv 0$, as you are permitted to perturb the data right away.

Define

$$h(\sigma) = \min_v J_\sigma[v, d]$$

and

$$\sigma^* = \inf\{\sigma : h(\sigma) = 0\}$$

The basic significance of σ^* is clear: $\sigma^* = 0$ characterizes noise-free data, whereas $\sigma^* = 1$ characterizes "pure" or incoherent noise.

The function h is analyzed in (Gockenbach and Symes, 1997), which proposes a root-finding strategy to solve the dual regularization problem: start with $\sigma = 0$, use Newton's method to locate σ^*. Since h is convex, each iterate is an underestimate, and σ^* is approached from below. The analysis of (Gockenbach et al., 1995) shows that under some conditions this process converges to the global solution of an output least squares formulation of the inverse problem; some of this analysis will be reviewed below.

The least squares formulation of velocity inversion is well known to have small domains of convergence to the global solution (eg. (Symes and Gockenbach, 1995)), but the global solution is also believed to be relatively insensitive to any sort of noise, coherent or not. Thus it is natural to speculate that the dual regularization algorithms sketched in the preceding paragraphs might function as a coherent noise remover, eliminating at least in some cases the sensitivity of DS estimates to multiple reflections, mode conversions, etc. This idea is explored in (Gockenbach and Symes, 1999), where it meets some success. This reference also explains that dual regularization is unlikely to produce useful results if coherent noise is dominant in some sense.

The purpose of this paper is to give a simple analysis to clarify the scope of dual regularization as a coherent noise suppressor. The bottom line: when the data is (1) unaliased, (2) sufficiently oscillatory, and (3) dominated by primary energy, dual regularization will succeed. The sense in which primary energy must be dominant is rather weak: the total coherent noise energy can be greater than that of the primary energy, but no *single coherent family* of noise events can have greater total energy. A coherent family is a group of events that can be flattened by a velocity model in the admissible set. The degree of oscillation, i.e. the separation from 0 Hz of the data passband, must be sufficient that the conflicting moveouts in the data are clearly distinguished on the scale of a wavelength.

Of course, some data meet these criteria, others do not. To illustrate the capabilities and limitations of dual regularization, we apply it to a number of synthetic and field CMPs using a version based on asymptotic layered modeling. The conclusions of the theory appear to describe the results well: field data appears to be sufficiently oscillatory and well-enough sampled that primary energy dominance determines the success or failure of dual regularization. In one case, dual regularization fails when applied to raw field CMPs dominated by water-column reverberations, but succeeds when applied to the same data after a standard multiple-suppression preprocess (parabolic Radon filtering). In another case with less extensive coherent noise, it succeeds without preprocessing.

Analysis

Ideal (noise-free) infinite bandwidth data, i.e. the range of the *forward map* or modeling operator $F[v]$, is the kernel of the differential semblance operator:

$$\partial G[v]d = 0 \;\Rightarrow\; d = F[v]r \tag{4}$$

for a (bin-independent) reflectivity map r. Finite bandwidth makes this relationship approximate - this is the phenomenon of NMO stretch - but it is asymptotically correct in the high frequency limit, in common with the other assumptions and conclusions in this report.

To provide a basis for analysis, assume that r, d range over Hilbert spaces R, D respectively, v over a subset V of $\mathbf{R}^n$, and that $v, r \mapsto F[v]r$ is continuous.

Theorem 1. Under the assumption (4),

$$\sigma^* = \inf_{v,r} \|F[v]r - d\|$$

Proof. From the definition, if $\sigma > \sigma^*$ then there exists $e \in D, \|e\| \leq \sigma\|d\|$ and $v \in V$ for which $\partial G[v](d+e) = 0$. According to (4), this means that the distance from d to the range of F is at most σ; this is so for every $\sigma > \sigma^*$. **QED**

That is, σ^* is truly the noise level in the data, as it measures the distance to the range of F.

Coherent noise is simply a data perturbation in the range of F. Thus the simplest model of coherently noisy data is

$$d = F[v_0]r_0 + F[v_1]r_1 \tag{5}$$

in which the first term is regarded as signal, and the second as noise. Clearly

$$\sigma^* \leq \min(\|F[v_0]r_0\|, \|F[v_1]r_1\|) \tag{6}$$

The key to understanding velocity analysis in the presence of coherent noise is the *near-orthogonality* of data and noise. We shall establish below in a special case - and it is true in general - that the ranges of $F[v_0]$ and $F[v_1]$ are nearly orthogonal if v_0, v_1 are substantially different *and* if the data are sufficiently oscillatory. The required velocity difference and degree of oscillation are linked: the residual moveout gererated by migration in the two velocities must be greater than a wavelength, roughly speaking.

Near-orthogonality is expressed by estimates of the form

$$\langle F[v_0]r_0, F[v_1]r_1\rangle \leq \epsilon\|F[v_0]r_0\|\|F[v_1]r_1\| \text{ if } \|v_0 - v_1\| \geq \delta \tag{7}$$

for suitable $\epsilon, \delta > 0$. Here $\langle\cdot,\cdot\rangle$ stands for the inner product in data space.

Theorem 2. Suppose that (7) holds with $\epsilon < \frac{1}{2}$, that the velocities appearing in the data definition (5) satisfy $\|v_1 - v_2\| \geq 2\delta$, and that

$$\|F[v_1]r_1\| \leq \sqrt{\frac{1-2\epsilon}{1+2\epsilon}}\|F[v_0]r_0\|.$$

Then

$$(1-\epsilon)\|F[v_1]r_1\|^2 - \epsilon\|F[v_0]r_0\|^2 \le (\sigma^*)^2 \le \|F[v_1]r_1\|^2$$

Also if v, r is a least squares solution, i.e. $\|F[v]r - d\| = \sigma^*$, then $\|v - v_0\| \le \delta$.

Proof. Pick another $v \in V, r \in R$ and suppose first that $\|v - v_i\| \ge \delta, i = 1, 2$. Then

$$\|F[v]r - (F[v_0]r_0 + F[v_1]r_1)\|^2 = \|F[v]r\|^2 + \|F[v_0]r_0\|^2 + \|F[v_1]r_1\|^2$$

$$-2\langle F[v]r, F[v_0]r_0\rangle - 2\langle F[v]r, F[v_1]r_1\rangle + 2\langle F[v_0]r_0, F[v_1]r_1\rangle$$

$$\ge \|F[v]r\|^2 + \|F[v_0]r_0\|^2 + \|F[v_1]r_1\|^2$$

$$-\epsilon(2\|F[v]r\|\|F[v_0]r_0\| - 2\|F[v]r\|\|F[v_1]r_1\| + 2\|F[v_0]r_0\|\|F[v_1]r_1\|)$$

$$\ge (1-2\epsilon)\|F[v]r\|^2 + (1-\epsilon)(\|F[v_0]r_0\|^2 + \|F[v_1]r_1\|^2)$$

so

$$\inf_{\{v:\|v-v_i\|\ge\delta,\, i=0,1\}}\|F[v]r - d\|^2 \ge (1-\epsilon)(\|F[v_0]r_0\|^2 + \|F[v_1]r_1\|^2) \tag{8}$$

Assume on the other hand that $\|v - v_0\| \le \delta$; then necessarily $\|v - v_1\| \ge \delta$, whence

$$\|F[v]r - (F[v_0]r_0 + F[v_1]r_1)\|^2$$

$$\ge \|F[v]r - F[v_0]r_0\|^2 + \|F[v_1]r_1\|^2 - \epsilon(\|F[v]r\| + \|F[v_0]r_0\|)\|F[v_1]r_1\|$$

$$\ge \|F[v]r - F[v_0]r_0\|^2 + \|F[v_1]r_1\|^2 - \epsilon(\|F[v]r - F[v_0]r_0\| + 2\|F[v_0]r_0\|)\|F[v_1]r_1\|$$

$$\ge \left(1 - \frac{\epsilon}{2}\right)\|F[v]r - F[v_0]r_0\|^2 + (1-\epsilon)\|F[v_1]r_1\|^2 - \epsilon\|F[v_0]r_0\|^2$$

whence

$$\inf_{\{v:\|v-v_0\|\le\delta\}}\|F[v] - d\|^2 \ge (1-\epsilon)\|F[v_1]r_1\|^2 - \epsilon\|F[v_0]r_0\|^2 \tag{9}$$

Equally well,

$$\inf_{\{v:\|v-v_1\|\le\delta\}}\|F[v] - d\|^2 \ge (1-\epsilon)\|F[v_0]r_0\|^2 - \epsilon\|F[v_1]r_1\|^2 \tag{10}$$

Clearly $(\sigma^*)^2$ is at least the minimum of the RHS of (8), (9), and (10). Because of the size assumption on the two components of the data, only the RHS of (9) counts.

Similar reasoning to that leading to (9) gives for $\|v - v_0\| \le \delta$

$$\|F[v]r - d\|^2 \le (1+\epsilon)\|F[v_1]r_1\|^2 + \epsilon\|F[v_0]r_0\|^2$$

The size assumption together with (8) and (10) then assure that any least squares solution must involve v within δ of v_0. **QED**

The reasoning leading to Theorem 2 can easily be generalized to any number N of data components; the generalization is left to the reader.

Theorem 2 justifies the assertion appearing in (Gockenbach and Symes, 1999), at least in the case of a single non-primary coherent data component: “If the primary phase is dominant, then the smallest data perturbation permitting flat image gathers will be that which removes the non-primary phase” - *provided* that the near-orthogonality hypothesis (7) holds.

Layered Asymptotic Scattering

Rather that investigate the near-orthogonality hypothesis in general, I will explore its meaning for a special case of the model used to construct the examples presented below, namely the *asymptotic linearized model of scattering from a layered medium*:

$$d(t,h) = r(Z(t,h)) = F[v]r(t,h) \tag{11}$$

Here t is time, h is half-offset, $d(t,h) : 0 < t_1 \le t \le t_2, 0 < h_1 \le h \le h_2$ is a common midpoint gather, $r(z)$ is the depth-dependent reflectivity, and $Z(t,h)$ is the depth-to-time conversion function, related to the two way traveltime (time-to-depth conversion) by

$$Z(T(z,h),h) = z,\ \ v(z)|\nabla T(z,h)| = 2 \tag{12}$$

In general the reflectivity may depend on angle and amplitude factors intervene; however the end result is not seriously affected by these omissions (Symes, 2001a).

Omission of convolution with the source wavelet is harder to justify. It appears that the impulse response theory developed here successfully predicts much of the observed behaviour of dual regularization, whereas convolution with a non-impulsive source leads to as-yet unresolved theoretical difficulties.

The near-orthogonality hypothesis compares the outputs of the forward map for two velocity models, v_0 and v_1. In the remainder of this section, I will assume that these velocities are *constant*. Note that this assumption holds only for the analysis presented in this section, which could very likely be extended to cover more general models. Nonconstant velocities are used to treat the field data examples discussed later, but still within the asymptotic layered scattering model.

Two traces

Suppose first that the data set consists of two traces, $d(h_i,t), i = 1,2$. For constant v,

$$T(z,h) = \frac{2}{v}\sqrt{z^2+h^2} \tag{13}$$

Theorem 3. For a two-trace data set, the orthogonality property (7) cannot hold with $\epsilon < \frac{1}{2}$.

Proof. Pick any v_0, r_0 and set $r_1(z) = r_0(Z_0(T_1(z,h_1),h_1))$. Then

$$\langle F[v_0]r_0, F[v_1]r_1\rangle$$

$$= \int_{t_1}^{t_2} dt\, r_1(Z_1(t,h_1))r_0(Z_0(t,h_1)) + \int_{t_1}^{t_2} dt\, r_1(Z_1(t,h_2))r_0(Z_0(t,h_2))$$

$$= \int_{t_1}^{t_2} dt\, |r_0(Z_0(t,h_1))|^2 + \int_{t_1}^{t_2} dt\, r_0(Z_0(T_1(Z_1(t,h_2),h_1),h_1))r_0(Z_0(t,h_2),h_2) \tag{14}$$

Suppose that $r_0(z) = 0$ if $z \in [z_a, z_b]$ and that the support of r_0 is short relative to the residual moveout between v_0 and v_1 in the sense that

$$[z_a, z_b] \cap [Z_0(T_1(Z_1(T_0(z_a,h_2),h_2),h_1),h_1), Z_0(T_1(Z_1(T_0(z_b,h_2),h_2),h_1),h_1)] = \phi \tag{15}$$

If $v_0 \neq v_1$ and $h_1 < h_2$ then it is always possible to choose $[z_a, z_b]$ and r_0 supported there so that (15) holds.

Then the integrand in the second summand of (14) vanishes identically, so

$$\langle F[v_0]r_0, F[v_1]r_1\rangle = \int_{t_1}^{t_2} dt\, |r_0(Z_0(t,h_1))|^2 \tag{16}$$

Assume that in addition the depth-to-time conversion does not map the support of r_0 or r_1 outside of the time interval of the data. That is, if $r = r_0$ or r_1 vanishes outsize of the interval $Z(t_1,h) \leq z \leq Z(t_2,h)$ for all $h = h_1$ or h_2 and $v = v_0$ or v_1. Then for any of these choices

$$\|F[v]r\|^2 = \int_{t_1}^{t_2} dt\, (|r(Z(t,h_1))|^2 + |r(Z(t,h_2))|^2) = \int_{-\infty}^{\infty} dz\, \left(\frac{\partial T}{\partial z}(z,h_1) + \frac{\partial T}{\partial z}(z,h_2)\right) |r(z)|^2$$

$$\leq 2\int_{-\infty}^{\infty} dz\, \left(\frac{\partial T}{\partial z}(z,h_1)\right) |r(z)|^2 = 2\int_{t_1}^{t_2} dt\, |r(Z(t,h_1))|^2 \tag{17}$$

since

$$\frac{\partial T}{\partial z}(z,h) = \frac{2}{v}\frac{z}{z^2+h^2}$$

is decreasing in h.

Combining (16) and (17), you see that

$$\|F[v_0]r_0\|\|F[v_1]r_1\| \leq 2\int_{t_1}^{t_2} dt\, |r_0(Z_0(t,h_1))|^2 = 2\langle F[v_0]r_0, F[v_1]r_1\rangle$$

QED

If near-orthogonality of events with different moveouts is "normal", then you could interpret this fact as follows: depending on $h_2 - h_1$, a two-trace data set either offers too little moveout or is too spatially aliased to be "normal".

Theorem 2 therefore does not imply that dual regularization, applied to a two-trace data set with optimal $\sigma = \sigma^*$, will remove all but the dominant event family.

Dense Spatial Sampling

"Dense" here means unaliased in offset, so that sums may be replaced by integrals without committing major errors.

Continue to assume layered asymptotic modeling with $v = \text{const.}$, so that the forward map is still given by (11), (12), and (13). The extent to which the reflectivity oscillates plays an important role in the next result. The measurement of oscillation is asymptotic: assume that the reflectivities r_0, r_1 appearing in (7) take the form

$$r(z) = R\left(\frac{z - z_{\text{min}}}{\epsilon^4}\right), \tag{18}$$

in which R satisfies

$$R(z) = 0,\ \ z \le 0 \tag{19}$$

$$\text{for any} z \in \mathbf{R},\ \text{there is } \delta z \text{ with } \frac{1}{2} \le \delta z \le 1 \text{ so that } \int_z^{z+\delta z} R = 0 \tag{20}$$

$$\text{sup}_{z\in\mathbf{R}} \int_z^{z+1} d\zeta\, |R(\zeta)|^2 \ \equiv \alpha(R)^2 < \infty \tag{21}$$

The shift by $z_{\text{min}} > 0$ is necessary to keep the support of r away from $z = 0$, where the asymptotic model (11) predicts nonsense. The choice of power of ϵ in the scaling law follows from the infinite extent of the reflectivity; if it is assumed that R vanishes for large z, then the power of epsilon in the denomenator can be reduced to the first.

The condition (21) implies that

$$\int_0^{z_{\max}} dz |r(z)|^2 = O(\alpha(R)^2) \tag{22}$$

uniformly in ϵ: that is, the energy in r is essentially the maximum energy in R over intervals of length 1.

From (20) it follows that for $z \ge z_{\text{min}}$,

$$\int_0^z d\zeta\, r(\zeta) = \epsilon^4 \int_0^{\frac{z-z_{\text{min}}}{\epsilon^4}} dz R(z) = \epsilon^4 \int_{z'}^{\frac{z-z_{\text{min}}}{\epsilon^4}} dz R(z) = \int_{\epsilon^4 z' + z_{\text{min}}}^z d\zeta\, r(\zeta)$$

for some z' with

$$\frac{z - z_{\text{min}}}{\epsilon^4} - 1 \le z' \le \frac{z - z_{\text{min}}}{\epsilon^4}$$

whence a standard application of the Cauchy-Schwarz inequality implies

$$\left|\int_0^z d\zeta\, r(\zeta)\right| \le \epsilon^2 \|r\| \tag{23}$$

Thus (20) requires that r be oscillatory in the sense that the indefinite integral of r decays like ϵ^2 as $\epsilon \to 0$.

To avoid pathology at zero offset, it will be assumed in what follows that the data is restricted to a nonzero offset range $0 < h_{\min} \le h \le h_{\max}$.

Theorem 4. Suppose that r_0, r_1 satisfy (18), with the corresponding R_0, R_1 satisfying (19), (20), and (21). Then for a constant $K > 0$ depending on $z_{\min}, z_{\max}, h_{\min}, h_{\max}$, and $\max(v_0, v_1)$

$$\langle F[v_0]r_0, F[v_1]r_1\rangle \le \epsilon \|F[v_0]r_0\| \|F[v_1]r_1\|$$

provided that $|v_0 - v_1| \ge K\epsilon$.

Proof: Without loss of generality, assume that $v_1 > v_0$.

$$\langle F[v_0]r_0, F[v_1]r_1\rangle = \int_0^{t_{\max}} dt \int_{h_{\min}}^{h_{\max}} dh\, r_0(Z_0(t,h)) r_1(Z_1(t,h))$$

$$= \int_0^{z_{\max}} dz\, r_0(z) \frac{2z}{v_0} \int_{h_{\min}}^{\bar{h}_{\max}(z)} dh \frac{1}{\sqrt{z^2+h^2}} r_1(Z_1(T_0(z,h),h)) \tag{24}$$

in which $h_{\min} \le \bar{h}_{\max}(z) \le h_{\max}$ and we will set from now on $z_{\max} = Z_0(t_{\max}, h_{\min})$.

Define

$$x(z,h) = Z_1(T_0(z,h),h) = \sqrt{\frac{v_1^2}{v_0^2}z^2 + \left(\frac{v_1^2}{v_0^2} - 1\right) h^2}$$

Then

$$\int_{h_{\min}}^{\bar{h}_{\max}(z)} dh \frac{1}{\sqrt{z^2+h^2}} r_1(Z_1(T_0(z,h),h))$$

$$= \int_{x(z,h_{\min})}^{x(z,\bar{h}_{\max}(z))} dx\, x \left[(x^2 - z^2)\left(x^2 - \frac{v_1^2}{v_0^2}\right)\right]^{-\frac{1}{2}} r_1(x)$$

$$= x \left[(x^2 - z^2)\left(x^2 - \frac{v_1^2}{v_0^2}z^2\right)\right]^{-\frac{1}{2}} s_1(x)\Bigg|_{x=x(z,h_{\min})}^{x(z,\bar{h}_{\max}(z))}$$

$$- \int_{x(z,h_{\min})}^{x(z,\bar{h}_{\max}(z))} dx \frac{d}{dx}\left(x\left[(x^2 - z^2)\left(x^2 - \frac{v_1^2}{v_0^2}z^2\right)\right]^{-\frac{1}{2}}\right) s_1(x) \tag{25}$$

in which

$$s_1(x) \equiv \int_0^x d\xi\, r_1(\xi)$$

Since

$$\frac{d}{dx}\left(x\left[(x^2 - z^2)\left(x^2 - \frac{v_1^2}{v_0^2}z^2\right)\right]^{-\frac{1}{2}}\right) < 0,$$

$$\left|\int_{x(z,h_{\min})}^{x(z,\bar{h}_{\max}(z))} dx \frac{d}{dx}\left(x\left[(x^2 - z^2)\left(x^2 - \frac{v_1^2}{v_0^2}z^2\right)\right]^{-\frac{1}{2}}\right) s_1(x)\right|$$

$$\leq [\sup_{x(z,h_{\min})\leq x\leq x(z,\bar{h}_{\max}(z))} |s_1(x)|] \left| \left\{ x\left[(x^2-z^2)\left(x^2-\frac{v_1^2}{v_0^2}z^2\right)\right]^{-\frac{1}{2}} \right\}_{x=x(z,h_{\min})}^{x(z,\bar{h}_{\max}(z))} \right|$$

From (23) and the definition of $x(z,h)$, this is

$$\leq \left(\frac{v_1^2}{v_0^2}-1\right)^{-1} \left| \left[\frac{x(z,h)}{(z^2+h^2)^{\frac{1}{2}}h} \right]_{h=h_{\min}}^{\bar{h}_{\max}(z)} \right| \epsilon^2 \|r_1\|$$

The boundary term in (25) is bounded by essentially the same quantity. Therefore

$$\int_{h_{\min}}^{\bar{h}_{\max}(z)} dh \, \frac{1}{\sqrt{z^2+h^2}} \, r_1(Z_1(T_0(z,h),h)) \leq C|v_1-v_0|^{-1}\epsilon^2\|r_1\|$$

C is a positive constant depending on the same quanitities as K in the statement of the theorem.

Comparison with (24) and straightforward use of the Cauchy-Schwarz inequality give

$$|\langle F[v_0]r_0, F[v_1]r_1\rangle| \leq K|v_1-v_0|^{-1}\epsilon^2\|r_0\|\|r_1\|$$

with K as described. **QED**

Algorithm

Several examples presented in the next section involve the solution of the minimization problem defining $h(\sigma)$, which involves computing the value and gradient of the function $J_\sigma[v,d]$. This in turn requires solution of the constrained optimization problem (3), plus a few additional calculations.

By inspection, (3) is a quadratically constrained quadratic minimization problem:

$$\min_f \|Rf\|^2 \text{ subject to } \|f-d\|^2 \leq \sigma^2\|d\|^2 \tag{26}$$

Here $f = d+e$ is the perturbed data, and $R = \partial G[v]$ is the differential semblance operator.

This problem is very similar to the so called trust region problem of numerical optimization theory - see (Dennis and Schnabel, 1983; Hebden, 1973; Moré, 1977; Rojas et al., 2002). The solution outlined here is motivated by the *Moré-Hebden algorithm* (Hebden, 1973; Moré, 1977) for the trust region problem.

The first order necessary conditions of optimality state that the solution satisfies

$$\lambda(f-d) + R^*Rf = 0 \tag{27}$$

$$\lambda(\|f-d)\|^2 - \sigma^2\|d\|^2) = 0 \tag{28}$$

The first condition states parallelism of the gradients of the constraint and objective functions. The second implies that either the constraint is satisfied as an equality - i.e. the solution is on the boundary of the set of constraint-satisfying models - or else the Lagrange multiplier λ vanishes, which means that the most regular solution actually has a smaller residual than assumed - i.e. σ is larger than the actual noise level.

From now on, we assume that $\sigma < \sigma^*$, so that the Lagrange multiplier λ is positive. For $\lambda > 0$, denote by $f(\lambda)$ the solution of the normal equation (27). Set

$$\phi(\lambda) = \|f(\lambda) - d\|$$

Computing the value of ϕ involves solving the normal equations (27). The implementation used to create the examples in the next section uses conjugate gradient iteration to solve (27). Since each iteration of this algorithm requires migration and adjoint migration (demigration), it is by far the most expensive step in this approach to the dual regularization problem.

Assuming that $\lambda > 0$ secular equation (28) is equivalent to

$$\phi(\lambda) = \sigma\|d\|. \tag{29}$$

Taking the reciprocal of both sides in (29) actually makes an easier problem. To see why, consider the simplest possible case: f and d are one-dimensional, and R is multiplication by a scalar. In that very special case,

$$f(\lambda) = \frac{\lambda d}{\lambda + R^2}$$

hence

$$\phi(\lambda) = \frac{R^2 d}{\lambda + R^2}$$

i.e. the reciprocal of ϕ is a linear function of λ. This suggests that Newtons's method is more likely to converge quickly when applied to

$$\psi(\lambda) \equiv \frac{1}{\phi(\lambda)} = \frac{1}{\sigma\|f(\lambda) - d\|} \tag{30}$$

and that is exactly what the Moré-Hebden algorithm does.

The algorithm used to create the examples explained in the next section actually uses Newton's method to solve (30). The iteration proceeds as follows:

- initialize λ somehow
- until convergence do: replace λ by

$$\lambda - \frac{\psi(\lambda) - \frac{1}{\sigma\|d\|}}{\frac{d\psi}{d\lambda}(\lambda)}$$

This algorithm requires the computation of the derivative $d\psi/d\lambda$. This computatation in turn requires solution of another linear system similar to the normal equations (27), again by conjugate gradient iteration, so roughly doubles the already considerable expense of this algorithm. Nonetheless it is quite effective when it can be carried out, as the examples in the next section show.

A better implementation of this approach *via* the secular equation would solve (30) by use of an interpolation-based root finder such as Brent's method (Forsythe et al., 1977). Nonetheless the elimination of the perturbed data f is still a source of much expense, as a fairly accurate solution must be achieved in order for the root finder to proceed successfully.

An alternative approach would solve the system composed of the normal equation (27) and the modified secular equation (30) *via* a quasi-Newton method. The difficulty in this approach is the assignment of relative weights to the residuals in (27) and (30). With proper assignment of weights, such an algorithm would in effect automatically adjust the precision with which the perturbed data f is estimated as the Lagrange multiplier estimate λ converges. This approach is potentially much less expensive than the Moré-Hebden approach for large problems like those occuring in this application, but has not yet been tested.

Yet another alternative was pioneered by Santos and Sorensen (Rojas et al., 2002). This solution proceeds *via* the equivalence of (26) with an eigenvalue problem, which is solved by means of the Lanczos algorithm. This alternative has also not been tested for the problem (26).

Whatever method one chooses to solve (26), i.e. to compute $J_\sigma[v,d]$, one must compute the gradient $\nabla_v J_\sigma[v,d]$ in order to update the velocity using some optimization algorithm or another. Fortunately, the gradient is essentially a by-product of the solution of (27), (30). Denote by $f[v], \lambda[v]$ the solutions of these equations. From (28),

$$\langle Df[v]\delta v, f[v]-d\rangle = 0 \tag{31}$$

whence

$$\begin{aligned} D\|\partial G[v]f[v]\|^2(\delta v) &= 2\langle DG[v](\delta v, f[v]), G[v]f[v]\rangle \\ &\quad +2\langle Df[v]\delta v, G[v]^*G[v]f[v]\rangle \\ &= 2\langle DG[v](\delta v, f[v]), G[v]f[v]\rangle + 2\langle DF[v]\delta v, -\lambda[v](f[v]-d)\rangle \\ &= 2\langle DG[v](\delta v, f[v]), G[v]f[v]\rangle \end{aligned}$$

in view of (31). Define the linear operator $W[v]$ by

$$W[v,d]\delta v = DG[v](\delta v, f[v])$$

Then

$$\nabla_v J_\sigma[v,d] = 2W[v,d]^*G[v]f[v]$$

Thus the calculation of ∇J_σ involves only the application of an operator, whose definition involves $f[v]$, to $G[v]f[v]$, and these latter quantities are by-products of the computation of J_σ itself.

Examples

To illustrate the scope of dual regularization, I will use a synthetic CMP and CMPs extracted from two North Sea marine streamer surveys. In all cases the underlying model is the layered asymptotic model explained above. The Moré-Hebden algorithm described in the preceding section was used to compute $J_\sigma[v,d]$ and its gradient, and the L-BFGS quasi-Newton iteration (Nocedal and Wright, 1999) to estimate an optimal v. I did not implement the σ updating strategy proposed in (Gockenbach and Symes, 1997); instead I first calculated the optimal v for $\sigma = 0$ (i.e. ordinary differential semblance optimization), estimated σ^* (in most cases by guessing), and minimized $J_{\sigma^*}[v,d]$ using the differential semblance v as starting guess. In the examples presented here this procedure appeared to be effective, but in general the process proposed in (Gockenbach and Symes, 1997) is likely to be more robust.

Computation of the of a step in the Moré-Hebden iteration requires (i) calculation of the migrated image volume $G[v]d$, (ii) application of the bin derivative ∂, and (iii) for the gradient, formation of the operator W^* described at the end of the **Algorithm** section. For CMPs and layered modeling, migration amounts to NMO correction. This change of variables is carried out numerically, using local cubic interpolation to produce a result depending sufficiently smoothly on velocity. The traveltime function defining the time-to-depth transformation was computed *via* the hyperbolic moveout approximation, which seemed adequate for the degree of data fidelity achieved in this exercise. The differential semblance operator was simply the first order divided difference in half-offset. The operator W^* was constructed from the same components as $G[v]$.

Synthetic Example

Figure 1 shows a linear combination of a single "primary" reflection ($v_0 = 1$ km/s), and "multiple" reflection ($v_0 = 0.8$ km/s). The "primary" amplitude is twice the "noise" energy. The two events have the same zero offset intercept, but diverge at large offsets. Thus this data is of exactly the form studied in the last two subsections.

The "common image gather" for this type of data, which permits assessment of velocity quality, is simply the migrated (i.e. normally moved out) data

$$G[v]d(z,h) = d(T(z,h),h)$$

A perfect choice of velocity for noise free data yields a perfectly flat CIG, i.e. $G[v]d$ independent of h. This is clearly impossible without data perturbation for the data in the figure.

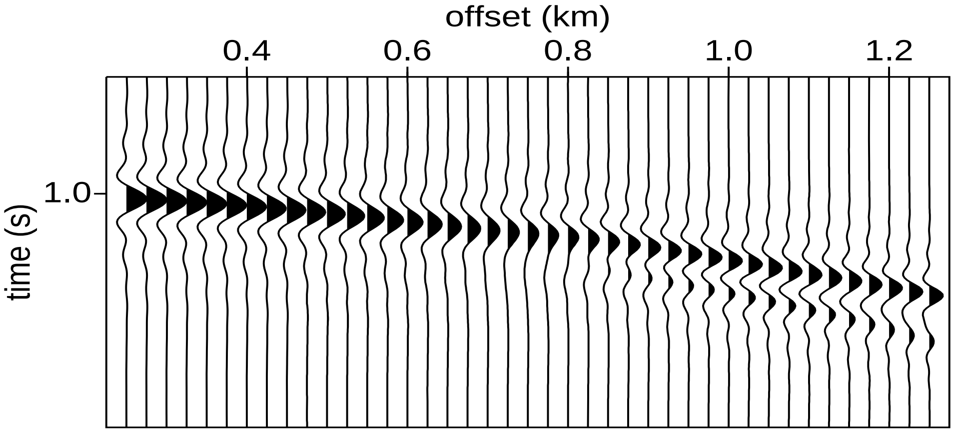

FIG. 1. Linear combination of events with different moveouts, same zero offset intercept

DS velocity inversion for $\sigma = 0.0$ gives a "best guess" velocity between "primary" and "noise", velocities - as is typical, DS "splits the difference", hence undercorrects the faster ("primary") event and overcorrects the slower ("multiple"). This effect is clearly evident in the CIG (Figure 2).

In this case we know exactly the noise level $\sigma^* = 0.5$. Dual regularization velocity inversion for $\sigma = 0.5$ permits most of the "multiple" to be discarded as part of e, produces flat "primary" CIG $G[v](d + e)$ (Figure 3). The ability of dual regularization to ignore the coherent noise in this case is even more evident in the standard CIG $G[v]d$ (Figure 4).

A North Sea CMP

The CMP displayed in Figure 5 is extracted from a North Sea survey. The event at around 1.48 s (zero offset) is interpreted as a primary from a major unconformity. The event following it by about 40 ms is nearly parallel, and has the correct timing for a seabed multiple (the water depth is approximately 30 m). While the moveout difference is not great, it is enough to be noticeable in the CIG based on DS inversion (Figure 6). Since a considerable part of the data is estimated to be coherent noise, even after the mute evident in Figure 5, a rough estimate of $\sigma = 0.6$ did not seem unreasonable (this time, we don't know the answer!). Dual regularization with this choice of σ produced the CIG displayed in Figure 7.

Another way to visualize the quality of velocity estimates is to plot the estimated *RMS* velocity over the standard velocity spectrum (Taner and Koehler, 1969). Figures 8 and

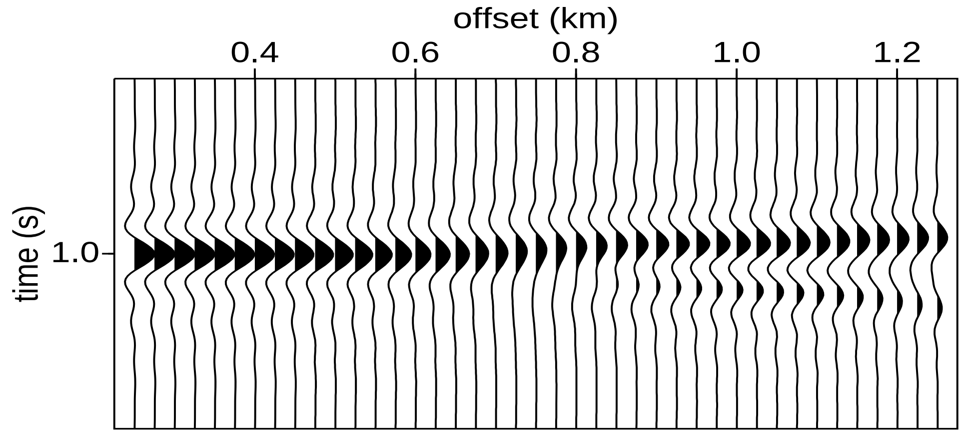

FIG. 2. CIG for differential semblance velocity estimate ($\sigma = 0$, no data perturbation).

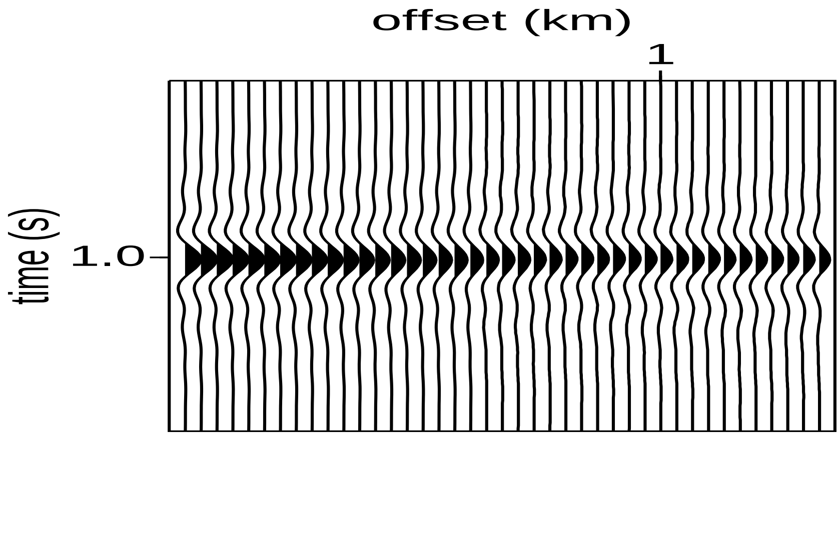

FIG. 3. "Primary" CIG, dual regularization with $\sigma = 0.5$.

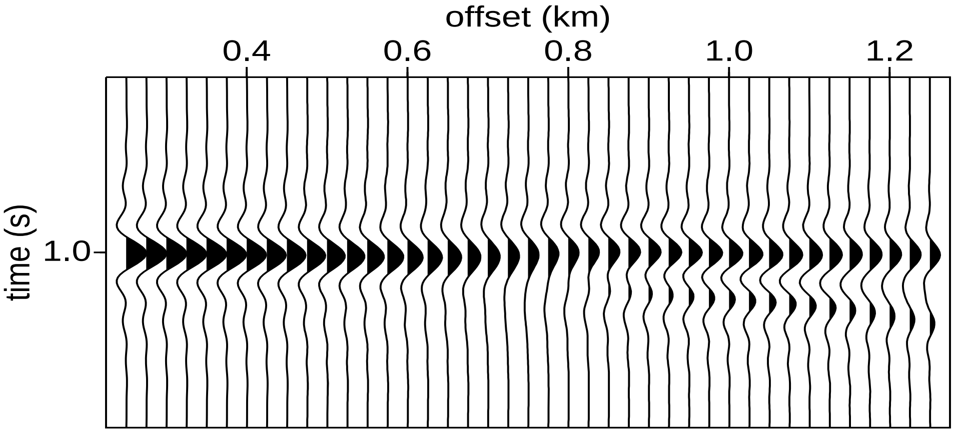

FIG. 4. CIG, dual regularization with $\sigma = 0.5$.

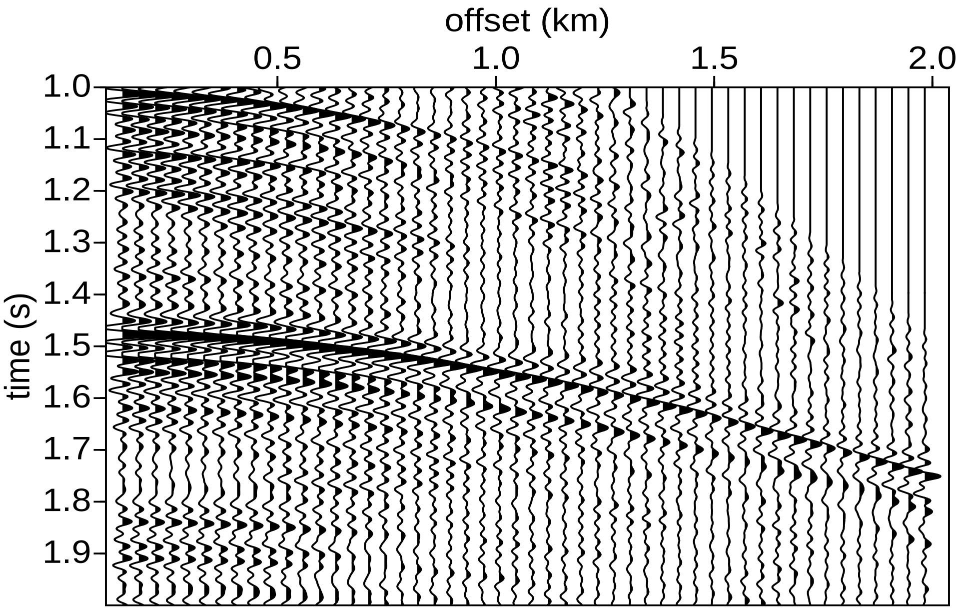

FIG. 5. North Sea CMP 1

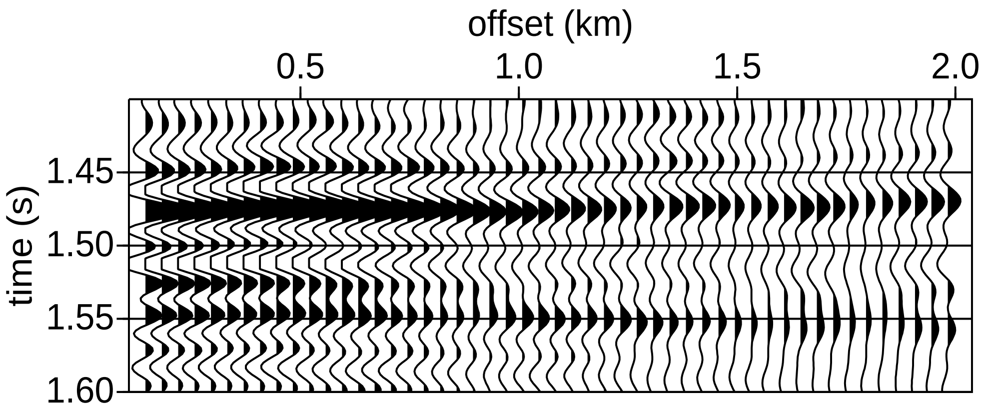

FIG. 6. CIG, DS ($\sigma = 0.0$).

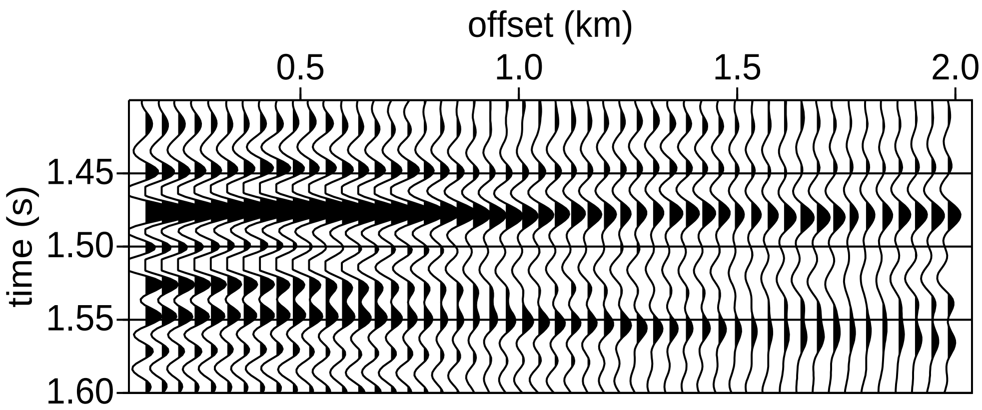

FIG. 7. CIG, dual regularization with $\sigma = 0.6$.

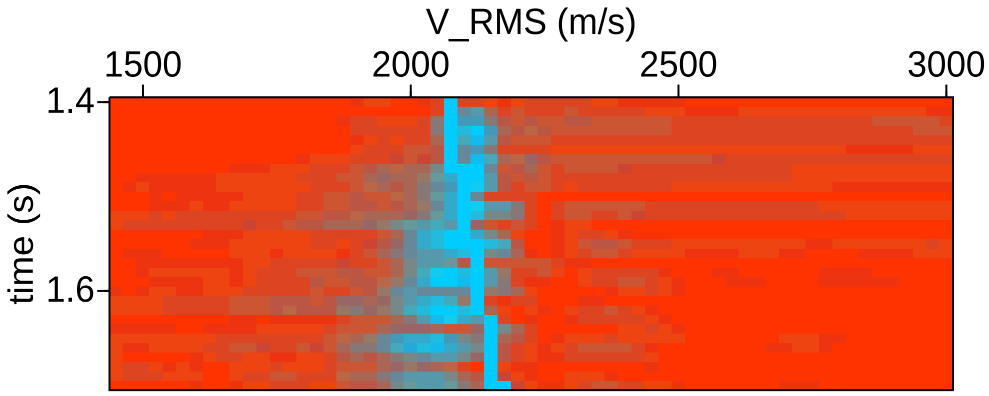

FIG. 8. DS RMS velocity estimate without dual regularization ($\sigma = 0.0$) overplotted on velocity spectrum.

9 show the velocity spectrum of Figure 5, with the velocities produced from DS without (Figure 8) and with (Figure 9) dual regularization with $\sigma = 0.6$.

Mobil AVO Data

Several years ago, Mobil research staff distributed a 2D marine streamer survey for research use worldwide (Keys and Foster, 1996). I have extracted three CMPs, numbered 1250, 1300, and 1350 in the numbering scheme produced by Seismic Unix processing (Stockwell, 2001). Figure 10 shows CMP 1300 after low-pass filter and mute. This data evidently consists mostly of multiple energy. Application of dual regularization with various values of σ ranging from 0.0 to 0.9 produced roughly the same result, also depicted in Figure 10. The moveout of the various multiple series exhibits an increase in apparent velocity with depth, but is at all depths well below the sedimentary compressional velocity.

The authors of the study (Keys and Foster, 1996) also distributed a version of this data to which parabolic Radon filtering multiple suppression had been applied. CMP 1300 after multiple suppression is displayed in Figure 11, along with DS ($\sigma = 0$) estimated RMS velocity overplotted on the spectrum and migrated (i.e. NMO corrected) data. The results are much better, but some apparent primary events are still not quite flattened - they seem undercorrected, whereas some slow events, still present at significant amplitude, are undercorrected. This is the classic signature of DS applied to data exhibiting conflicting moveout, as demonstrated in the synthetic example above. The RMS velocity plot navigates between the corridor of fast semblance peaks - the primaries,

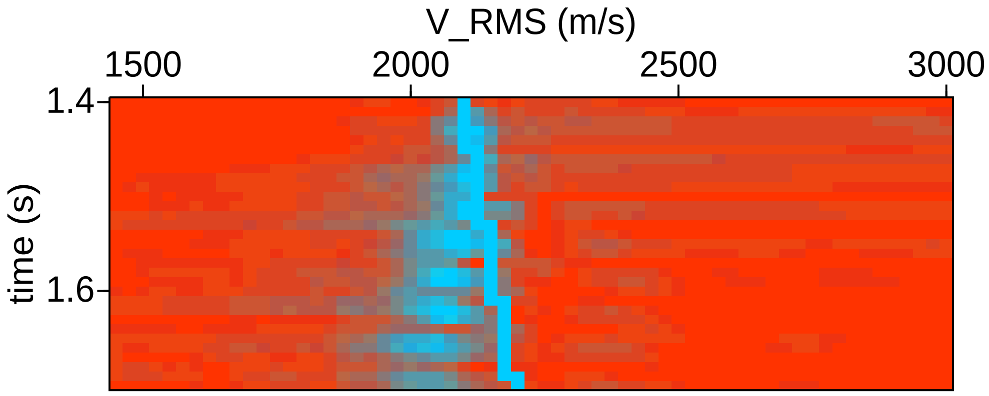

Fig. 9. DS RMS velocity estimate with dual regularization ($\sigma = 0.6$) overplotted on velocity spectrum.

presumeably - and the slower peaks, presumeably generated by multiples whose moveout is too close to that of the assumed primary velocity used in the parabolic Radon filter to be altogether suppressed.

A guess of 50% noise level seemed realistic for this data, based simply on inspection. Application of dual regularization with $\sigma = 0.5$ produced the RMS velocity and migrated data displayed in Figure 12. Now the apparent primary events are essentially flat while the slower events are more dramatically overcorrected. The dual regularization generated RMS velocity now picks the fast peaks on the spectrum.

I applied the same procedure to a number of other CMPs in this part of the line, with similar results: in all cases, dual regularization successfully suppressed the influence of residual coherent noise and flattened the apparent primary event series, picking the fastest corridor of spectral peaks rather precisely. A close look at the 2.3 s event in the migrated (NMO corrected) CMP 1250 is instructive: this is displayed as Figure 13. Evidently the simple layered medium time-depth model used in this exercise is not really adequate; a depth migration algorithm capable of modeling the effects of lateral velocity heterogeneity is necessary to account for this data at the 1/4 wavelength scale, even in this largely layered part of the sedimentary column.

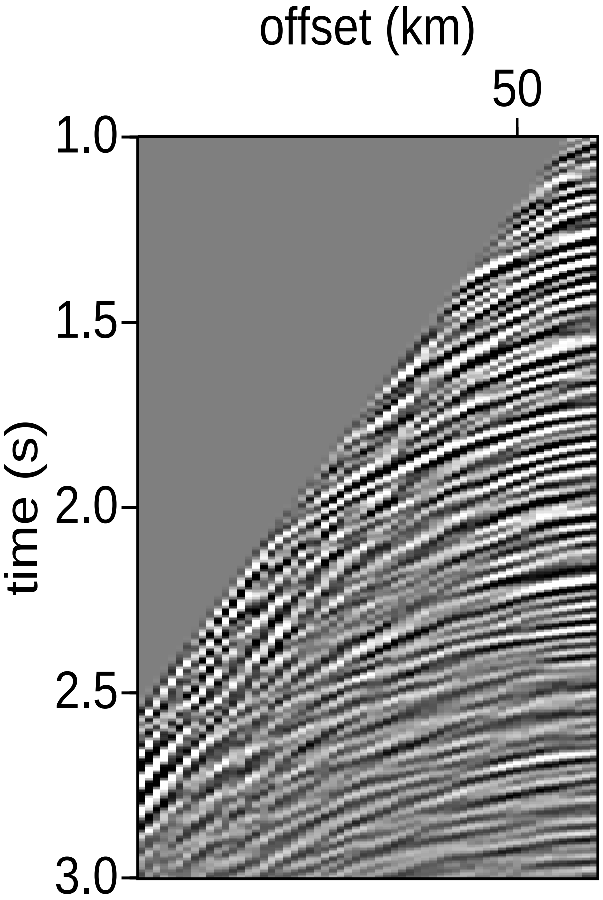

Fig. 10. Mobil AVO CMP 1300, after low pass filter and mute.

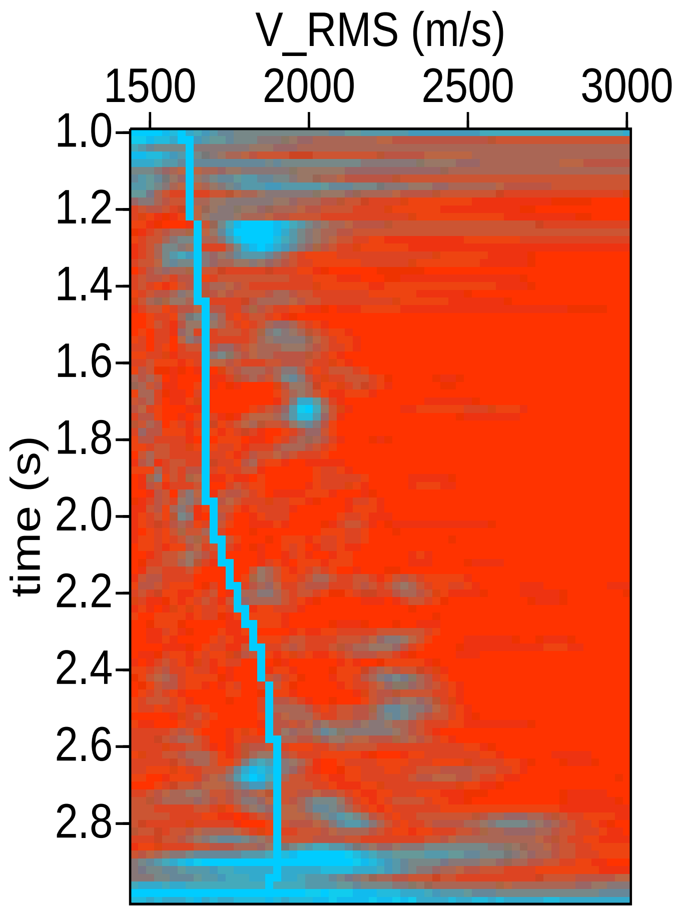

Fig. 11. Dual regularization velocity analysis of Mobil AVO CMP 1300 (any σ!) overplotted on velocity spectrum. Evidently the vast bulk of energy in this CMP is low apparent velocity coherent noise (multiple reflections) and thus noise velocity is dominant. Dual regularization flattens the noise rather than the primaries.

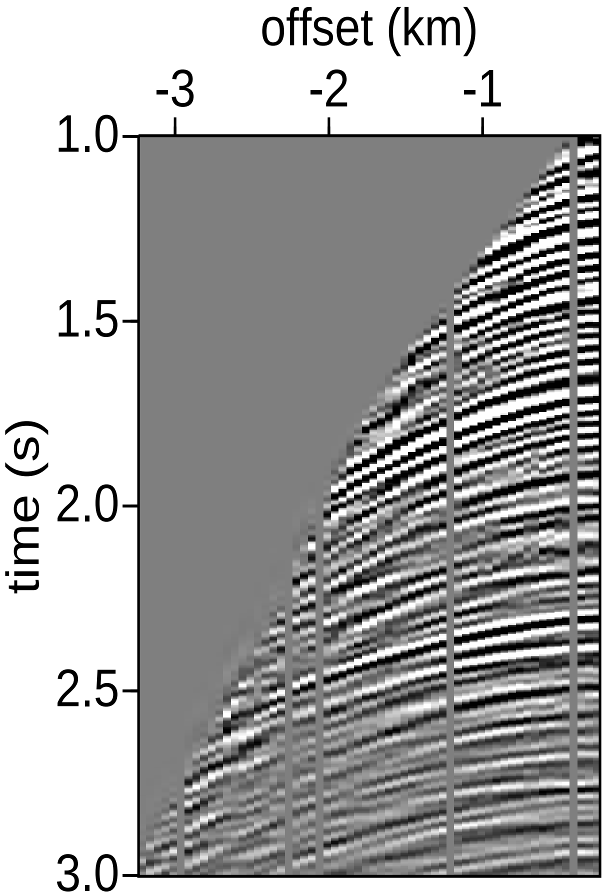

Fig. 12. Mobil AVO CMP 1300 after application of parabolic Radon filter multiple suppression, low pass filter, mute.

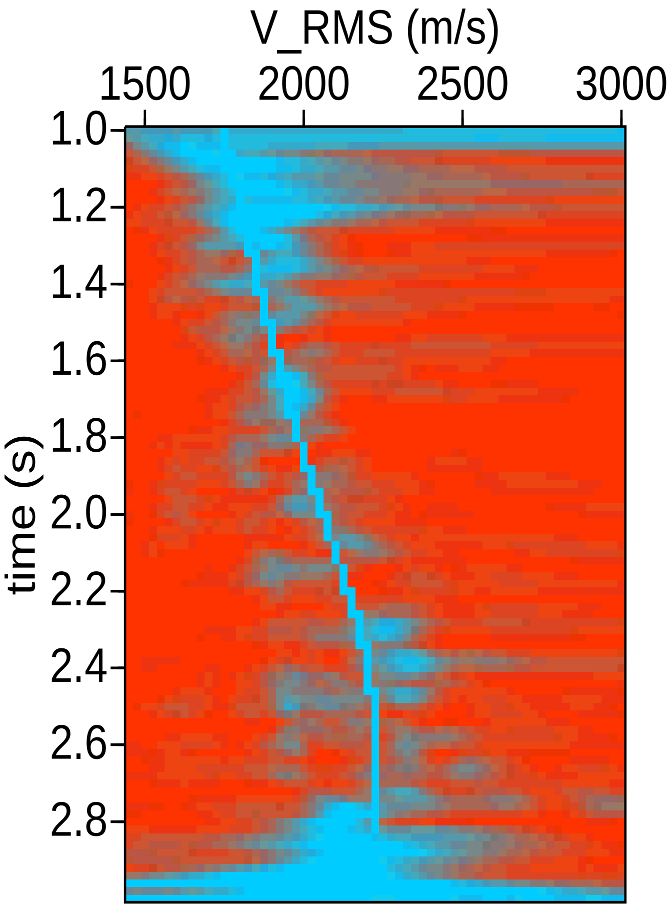

Fig. 13. DS RMS velocity estimate for multiple suppressed Mobil AVO 1300 (Figure 11), overplotted on velocity spectrum. Note that DS RMS velocity navigates between fast and slower peaks.

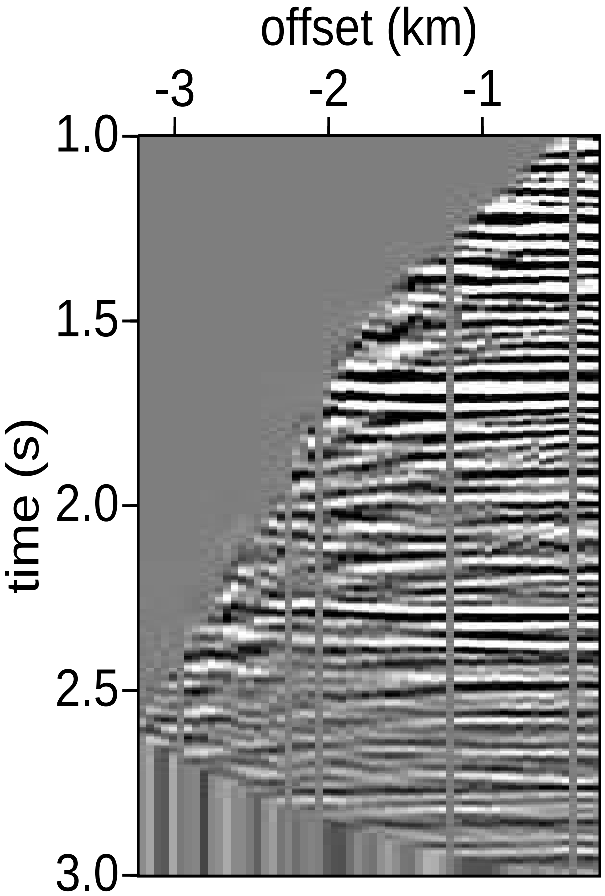

FIG. 14. NMO corrected data of Figure 12 using DS velocity. Apparent primary events slightly undercorrected, slower events overcorrected.

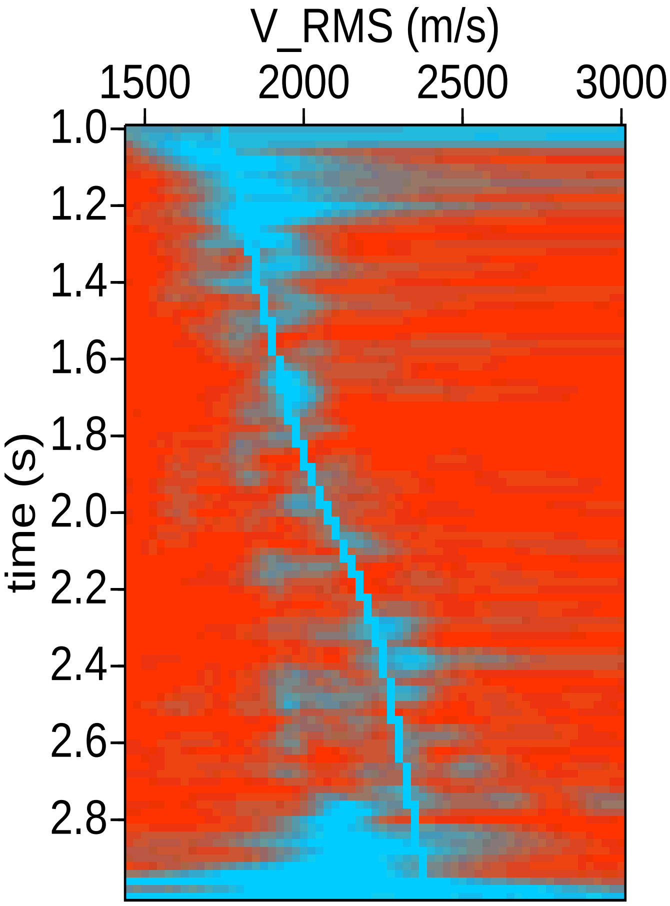

FIG. 15. Dual regularization RMS velocity estimate with $\sigma = 0.5$ for multiple suppressed Mobil AVO 1300 (Figure 12), overplotted on velocity spectrum. Dual regularization velocity picks the fast peaks and ignores the slower peaks.

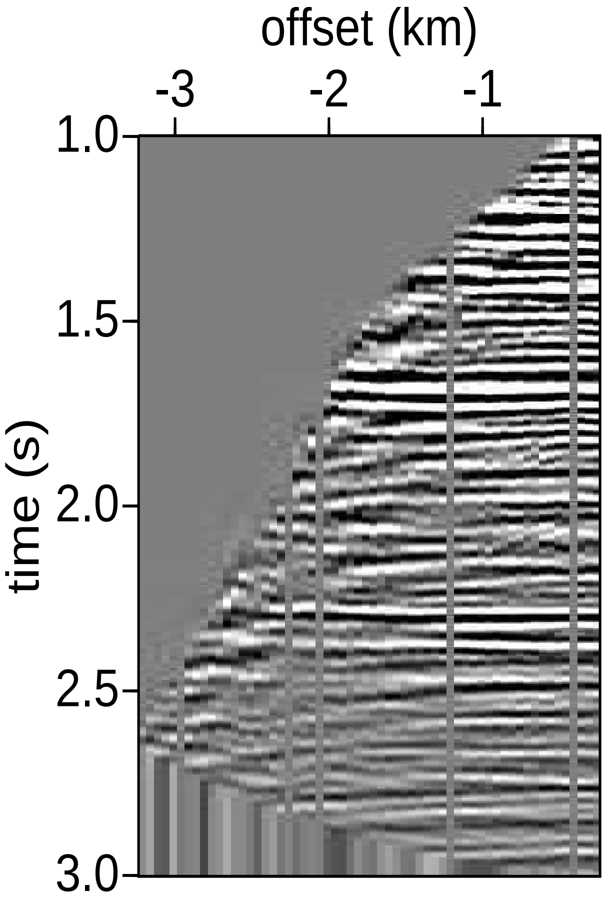

FIG. 16. NMO corrected data of Figure 12 using dual regularization velocity, $\sigma = 0.5$. Apparent primary events are now flat.

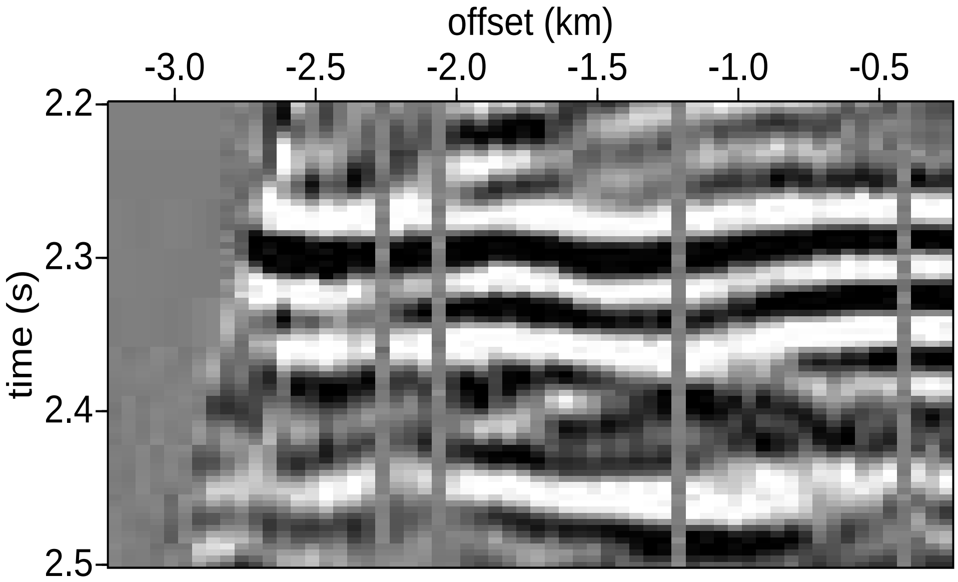

FIG. 17. Detail of 2.3 s event after NMO correction on Mobil AVO CMP 1250. Clearly migration is needed!

Discussion

The utility of dual regularization seems sharply delimited by the theory and supporting examples presented in this paper. The source of the method's effectiveness is revealed to be the near-orthogonality of synthetic seismograms for essentially different velocities. This near-orthogonality property holds for data which is well enough sampled, sufficiently oscillatory, and in which the primary reflection series is the largest single coherent event family. The theory developed here is qualitative, but the examples tend to suggest that in general exploration seismograms have the first two properties. The third may be satisfied by the raw data, or may not be; if not, it may be possible to eliminate enough of the coherent noise by conventional preprocessing that the preprocessed data has the primaries-dominant property.

Dual regularization is not in any sense a complete solution to the problem of coherent noise in automated velocity analysis. However dual regularization does extend the scope of differential semblance velocity analysis somewhat beyond the domain of ideal primaries-only data: it makes differential semblance robust against small but significant amounts of coherent noise.

Acknowledgements

This work was supported in part by National Science Foundation grants DMS-9973423, DMS-9973308, and EAR-9977697, by the Los Alamos National Laboratory Computer Science Institute (LACSI) through LANL contract number 03891-99-23, by the Department of Energy EMSP grant DE-FG07-97 ER14827, and by The Rice Inversion Project (TRIP). TRIP sponsors for 2001 were Amerada Hess Corp., Conoco Inc., ExxonMobil Upstream Research Co., Landmark Graphics Corp., Shell International Research, and Western Geco.

REFERENCES

Araya, K., Kim, S., Nolan, C. J., and Symes, W. W. (1996). Velocity inversion and high frequency asymptotics. In *Mathematical Methods in Geophysical Imaging*. Society of Photooptical Instrumentation Engineers.

Chauris, H. and Noble, M. (2001). Two-dimensional velocity macro model estimation from seismic reflection data by local differential semblance optimization: applications synthetic and real data sets. *Geophys. J. Int.*, pages 14–26.

Chauris, H., Noble, M., and Podvin, P. (1998). Testing the behaviour of differential semblance for velocity estimation. In *68th Annual International SEG Meeting*, New Orleans. Society of Exploration Geophysicists. Expanded Abstract.

Dennis, Jr., J. and Schnabel, R. (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Englewood Cliffs.

Forsythe, G., Malcolm, M., and Moler, C. (1977). *Computer Methods for Mathematical Computations*. Prentice-Hall, Englewood Cliffs, NJ.

Gockenbach, M. (1994). *An abstract analysis of differential semblance optimization*. PhD thesis, Computational and Applied Mathematics Department, Rice University, Houston, Texas, U.S.A.

Gockenbach, M. and Symes, W. (1997). Duality for inverse problems in wave propagation. In Biegler, L., Coleman, T., Santosa, F., and Conn, A., editors, *Large Scale Optimization*, New York. Institute for Mathematics and its Applications, Springer Verlag.

Gockenbach, M. and Symes, W. (1999). Coherent noise suppression in velocity inversion. In *Expanded Abstracts, Society of Exploration Geophysicists, 69th Annual International Meeting*, Tulsa. SEG.

Gockenbach, M., Symes, W., and Tapia, R. (1995). The dual regularization approach to seismic velocity inversion. *Inverse Problems*, 11(3):501–531.

Hebden, M. (1973). An algorithm ofr minimization using exact second derivatives. Technical Report TP 515, A. E. R. E. Harwell.

Kern, M. and Symes, W. (1994). Inversion of reflection seismograms by differential semblance analysis: Algorithm structure and synthetic examples. *Geophysical Prospecting*, 99:565–614.

Keys, R. G. and Foster, D. J. (1996). A data set for evalutating and comparing seismic inversion methods. *Geophysics*. in press.

Minkoff, S. E. and Symes, W. W. (1997). Full waveform inversion of marine reflection data in the plane-wave domain. *Geophysics*, 62(2):540–553.

Moré, J. (1977). The Levenburg-Marquardt algorithm: implementation and theory. In *Numerical Analysis*. Springer-Verlag. Lecture Notes in Mathematics 630.

Mulder, W. and ten Kroode, A. (2001). Automatic velocity analysis by differential semblance optimization. In *Expanded Abstracts, Society of Exploration Geophysicists, 71st Annual International Meeting*, pages xx–xx, Tulsa. SEG.

Nocedal, J. and Wright, S. (1999). *Numerical Optimization*. Springer Verlag, New York.

Rojas, M., Santos, S., and Sorensen, D. (2002). A matrix free algorithm for the large scale trus region subproblem. *SIAM J. on Optimization*. (in press; available electronically from `www.caam.rice.edu` as TR 99-19).

Stockwell, J. (2001). Seismic Unix home page. `www.cwp.mines.edu/cwpcodes`.

Symes, W. (1986). Stability and instability results for inverse problems in several-dimensional wave propagation. In Glowinski, R. and Lions, J., editors, *Proc. 7th International Conference on Computing Methods in Applied Science and Engineering*. North-Holland.

Symes, W. (1991). Layered velocity inversion: a model problem from reflection seismology. *SIAM J. Math. Anal.*, 22:680–716.

Symes, W. and Carazzone, J. (1991). Velocity inversion by differential semblance optimization. *Geophysics*, 56(5):654–663.

Symes, W. and Gockenbach, M. (1995). Waveform inversion for velocity: Where have all the minima gone? In *Proc. 65th Annual International Meeting of the Society of Exploration geophysicists*, Houston, TX. Extended abstract.

Symes, W. and Versteeg, R. (1993). Velocity model determination using differential semblance optimization. In *Proc. 63rd Annual International Meeting*, pages 696–699, Washington, D.C. Society of Exploration Geophysicists. Expanded abstract.

Symes, W. W. (1997). High frequency asymptotics and velocity estimation. In *Depth Imaging Symposium*, Tulsa, Oklahoma, U.S.A. Geophysical Society of Tulsa.

Symes, W. W. (1999). All stationary points of differential semblance are asymptotic global minimizers: Layered acoustics. In *The Rice Inversion Project: Annual Report 1999*, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005-1892. also appeared in SEP-95.

Symes, W. W. (2001a). Differential semblance: Formulation and implementation. In *The Rice Inversion Project: Annual Report 2001*, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005-1892.

Symes, W. W. (2001b). Global solution of reflection seismic velocity analysis problem, layered acoustic case. In *The Rice Inversion Project: Annual Report 2001*, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005-1892.

Taner, M. and Koehler, F. (1969). Velocity spectra: digital computer derivation and application of velocity functions. *Geophysics*, 34:859–881.