Lecture 4 (or its last half, depending on how you want to look at it): Smooth objectives for velocity estimation

- Duality between Stack Power and Least Squares

- Why stack power behaves as it does

- Characterization of smooth quadratic forms - differential semblance

- Examples - first installment

Since in general (for nonlayered models) stack power and/or semblance cannot be diagonalized (i.e. localized in t_0), must treat as a *global* optimization problem in a suitable space of velocity fields.

Toldi (1985), Fowler (1988), al Yahya (1989) approached velocity estimation as an optimization problem *via* semblance. Kolb et al. (1986), Tarantola (1986, 1987), Sen and Stoffa (1991), and many others have approached velocity estimation as an optimization problem *via* least squares.

- descent methods often fail
- global opt (SA, GA,...) very costly
- why?

Why stack power, least squares have similar properties:

[To simplify notation, incorporate mute Φ into definition of forward map: $F[c] \leftarrow \Phi F[c], S^{\mathrm{obs}} \leftarrow \Phi S^{\mathrm{obs}}$]

Reduced least squares objective:

$$J^{\mathrm{red}}[c; S^{\mathrm{obs}}] = \frac{1}{2}\|F[c]R[c] - S^{\mathrm{obs}}\|^2$$

where

$$R[c] \equiv (F[c]^* F[c])^{-1} F[c]^* S^{\mathrm{obs}}$$

(note the slight lie: the muted $F[c]$ does not have an invertible normal operator, so regularization is necessary!)

Stuff this in, simplify:

$$J^{\mathrm{red}}[c; S^{\mathrm{obs}}] = \frac{1}{2}\|(F[c](F[c]^* F[c])^{-1} F[c]^* - I)S^{\mathrm{obs}}\|^2$$

$$= \frac{1}{2}\|F[c](F[c]^*F[c])^{-1}F[c]^*S^{\mathrm{obs}}\|^2$$

$$- < F[c](F[c]^*F[c])^{-1}F[c]^*S^{\mathrm{obs}}, S^{\mathrm{obs}} > + \frac{1}{2}\|S^{\mathrm{obs}}\|^2$$

$$= \frac{1}{2}\|S^{\mathrm{obs}}\|^2 - \frac{1}{2} < F[c]^*S^{\mathrm{obs}}, (F[c]^*F[c])^{-1}F[c]^*S^{\mathrm{obs}} >$$

Recall that

$$< F[c]^*S^{\mathrm{obs}}, (F[c]^*F[c])^{-1}F[c]^*S^{\mathrm{obs}} > =$$

$$\int dz \, \left(\sum_h 2\Phi^2(h, 2\bar{\tau}(z,h))A(h,z)^2\frac{\partial\bar{\tau}}{\partial z}(h,z)\right)^{-1} \times$$

$$\left(\sum_h 2A(h,z)\frac{\partial\bar{\tau}}{\partial z}(h,z)S^{\mathrm{obs}}(h, 2\bar{\tau}(h,z))\right)^2$$

$$= \int dz \, \left(\sum_h B(h,z)S^{\mathrm{obs}}(h, 2\bar{\tau}(h,z))\right)^2$$

i.e. a weighted version of stack power, with smooth weight $B(h,z)$.

What we have learned from this:

- reduced LS is a linear function of a weighted stack power

- minimizing reduced LS is equivalent to maximizing (appropriately) weighted stack power

- both amount to *quadratic forms in the data, parametrized by velocity, of the form*

  $$< B[c]\mathcal{F}[c]^* S^{\mathrm{obs}}, W B[c]\mathcal{F}[c]^* S^{\mathrm{obs}} >$$

  where $B[c]$ is a c-dependent weight, and W is a (c-independent) operator - in this case, $W = \Sigma^* \Sigma$

Another look at stack power: why it is so unloveable as an objective

Make a couple of simplifying assumptions -

- c = const. (but we don't know which one!): then all of the different types of velocity are the same, i.e. $c = V_{\text{int}} = V_{RMS}$

- the data is consistent with some (unknown) model (c_*, R_*):

$$S^{\text{obs}}(h,t) = A(h, \bar{\zeta}_*(h,t))R_*(\bar{\zeta}_*(h,t))$$

- R_* is random and mean zero (rapidly oscillating) but with bounded energy: $R_*(z) = \frac{1}{\sqrt{\epsilon}}\bar{R}_*(\frac{2\bar{\tau}(0,z)}{\epsilon})$

- assume continuous sampling in h, replace sums by integrals

The any version of stack power looks like

$$P[c] = \frac{1}{\epsilon} \int dt_0$$

$$\left(\int_0^\infty dh\, B(h, t_0; c) \bar{R}_* \left(\frac{1}{\epsilon} \sqrt{t_0^2 + h^2 \left(\frac{1}{c_*^2} - \frac{1}{c^2} \right)} \right) \right)^2$$

where B is smooth, generally (smoothly) c-dependent. Assume also that it vanishes at $h = 0$ (incorporates a cutoff for small offsets).

Change variables by setting

$$r = \frac{1}{\epsilon} \sqrt{t_0^2 + h^2 \left(\frac{1}{c_*^2} - \frac{1}{c^2} \right)}$$

Then

$$P[c] = \epsilon^3 \int dt_0 \left(\frac{1}{c_*^2} - \frac{1}{c^2} \right)^{-2} \times$$

$$\left(\int_{\frac{t_0}{\epsilon}}^\infty dr\, \frac{r}{h(\epsilon r, t_0)} B(h(\epsilon r, t_0), t_0) \bar{R}_*(r) \right)^2$$

where

$$h(r,t_0) = \sqrt{\frac{r^2 - t_0^2}{\frac{1}{c_*^2} - \frac{1}{c^2}}}$$

The support of the first factor in the integrand grows like $1/\epsilon$ and R is assumed to be random $\Rightarrow$ its mean is $\sqrt{\epsilon} \Rightarrow$ its integral over an interval of length $1/\epsilon$ is $O(1/\sqrt{\epsilon})$ Therefore for $c \neq c_*$, the stack power decreases like ϵ^2 as $\epsilon \to 0$. Since the max power (obtained when $c = c_*$) is $O(1)$, the power attains a proportion $\delta < 1$ of its peak value when

$$P[c] > \delta P_{\max} \Rightarrow \frac{1}{c_*^2} - \frac{1}{c^2} < O(\delta^{-\frac{1}{2}}\epsilon)$$

That is, *the width of the peak is proportional to the dominant wavelength*.

Conclude:

- Stack power does not converge as $\epsilon \to 0$ to a smooth limiting function of c

- Derivatives wrt c are not bounded uniformly in terms of data energy
- Least squares has the same properties, since it is a linear function of stack power!
- Generic existence of spurious local mins (synth. expl., also Herve Chauris, 1998)

These properties lead to the poor results reported for attempts at velocity estimation using local (gradient-based) optimization of either stack power or least squares approaches. To converge, must know the velocity to “within a wavelength” *a priori*!

Q: Do objectives of the same form (quadratic forms in the data, parametrized by c) exist which have derivatives uniformly bounded over data of bounded energy?

A: Yes, but not many...

(show velocity scans from Yilmaz)

What varies slowly: *slope* of NMO-corrected reflections.

How to measure slope: you learned this in calculus! Differentiate!

$\Rightarrow$ choose $W = \frac{\partial}{\partial h}$ in

$$DS[c; S^{\mathrm{obs}}] = < B[c]\mathcal{F}[c]^* S^{\mathrm{obs}}, W B[c]\mathcal{F}[c]^* S^{\mathrm{obs}} >$$

$$= < S^{\mathrm{obs}}, \mathcal{F}[c] B[c] W B[c]\mathcal{F}[c]^* S^{\mathrm{obs}} >$$

To compute this, use the chain rule:

$$W B[c]\mathcal{F}[c]^* S^{\mathrm{obs}}(h,z)$$

$$= \frac{\partial}{\partial h} B_1[c](h,z) S^{\mathrm{obs}}(h, 2\bar{\tau}(h,z))$$

$$= B_1[c](h,z) \left(\frac{\partial}{\partial h} + 2\frac{\partial \bar{\tau}}{\partial h}\frac{\partial}{\partial t}\right) S^{\mathrm{obs}}(h,t)\Big|_{t=2\bar{\tau}(h,z)} + ...$$

where "..." represents lower order terms, not involving derivatives of S^{obs} and so not leading order in frequency.

Applying the other half, get

$$DS[c; S^{\text{obs}}] \simeq$$

$$\int dt \int dh\, b[c](h,t) \left(\frac{\partial S^{\text{obs}}}{\partial h} + p(h,t) \frac{\partial S^{\text{obs}}}{\partial t} \right)(h,t)$$

where $b[c]$ is a positive c-dependent weight and

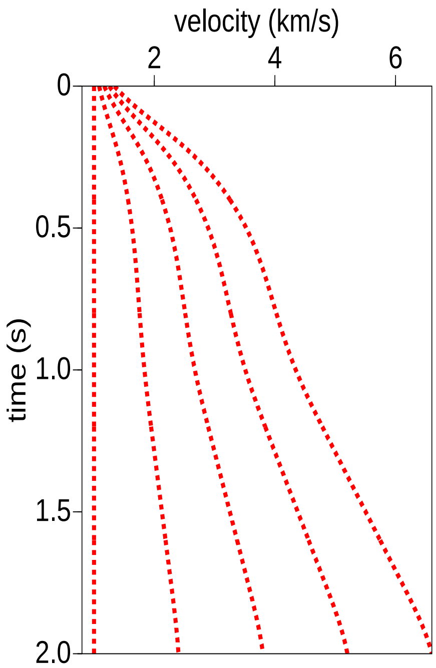
$$p(h,t) = 2\frac{\partial \bar{\tau}}{\partial h}(h, \bar{\zeta}(h,t))$$

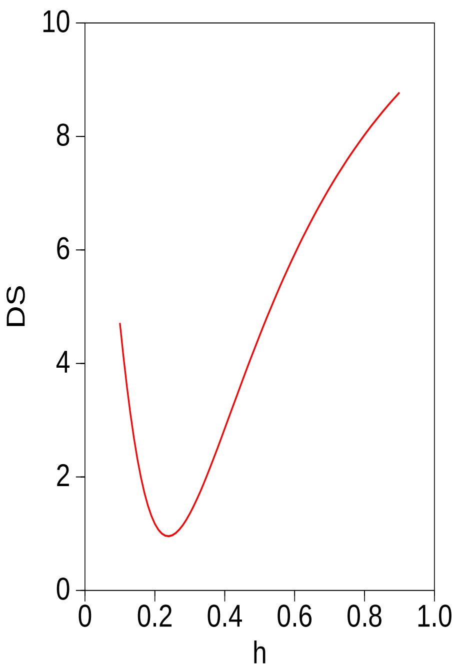
is the *event slowness* field for velocity c.

This function is smooth in c, because the traveltime field and all derived quantities (b, p) are smooth in c as follows from geometric optics.

Consequence $DS[c, S]$ will have a smooth, stable graph, regardless of the frequency content of S: it is smooth jointly in c, S over suitable bounded sets in (say) $C^2 \times L^2$.

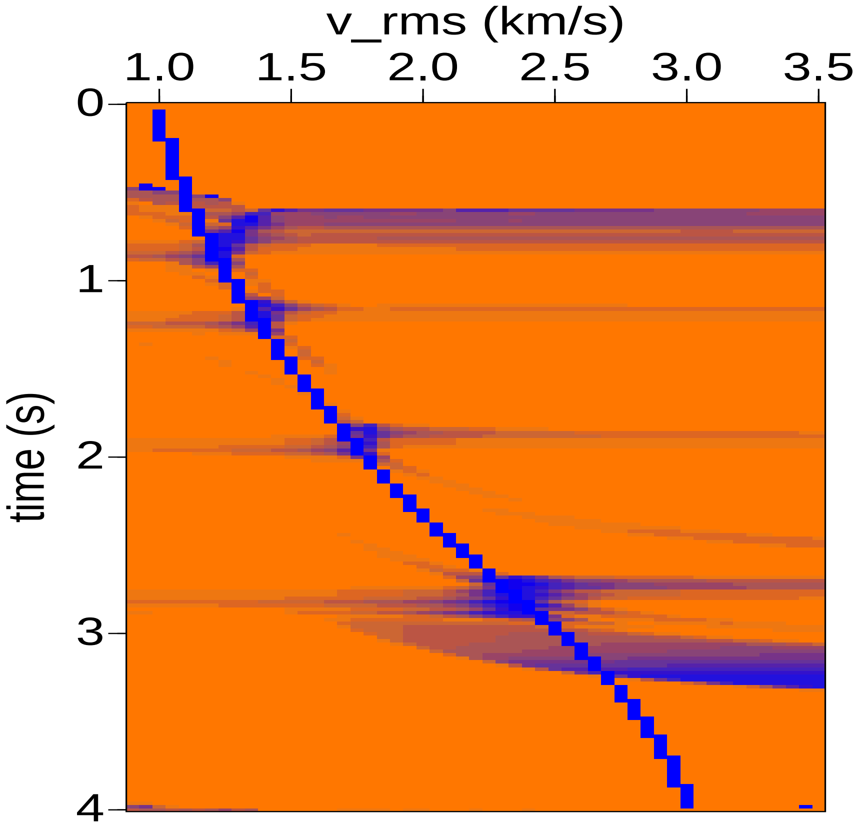
Example: synthetic split spread





Left: samples along line segment in velocity space (*not* including target!)
Right: differential semblance

Note apparent unimodality - is this really true?



RMS velocity derived from optimizing DS over 6-diml space of spline interval velocities, using limited memory BFGS.