Mathematics of Extended Inversion for Wave Propagation

William W. Symes

The Rice Inversion Project
Computational and Applied Mathematics
Rice University

SIAM 2018

Overview

Takeaway: unitary property of modeling operator $\Rightarrow$ convergent gradient computation for extended FWI

Agenda:

- Transmission FWI
- Model extension and Variable Projection
- Accurate VPM gradient computation

Agenda

Transmission FWI

Model extension and Variable Projection

Accurate VPM gradient computation

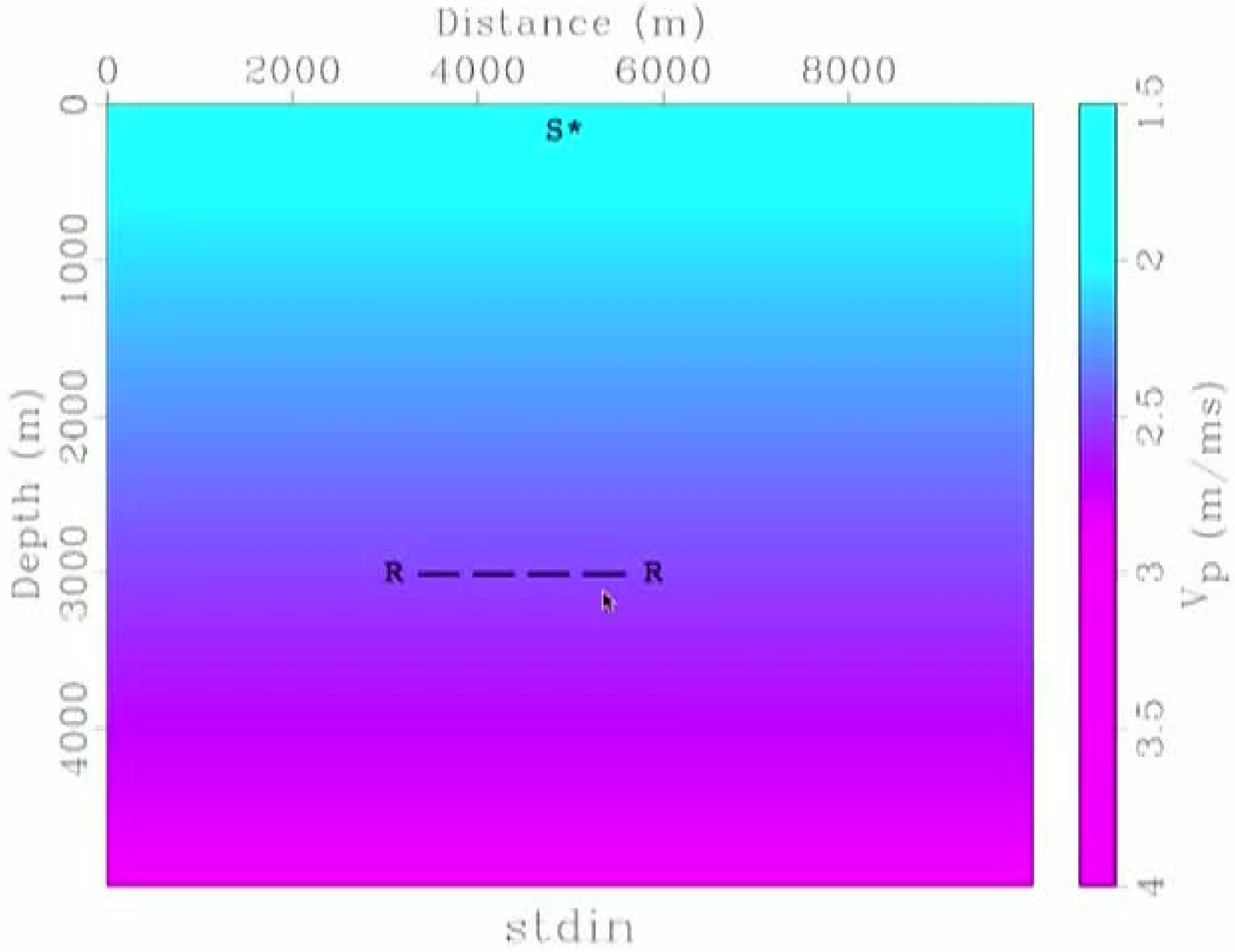
Simple example: transmission modeling, constant density acoustics with isotropic point radiator

$p(\mathbf{x}, t)$ = pressure, $v(\mathbf{x})$ = wave speed, $w(t)$ = source "wavelet"

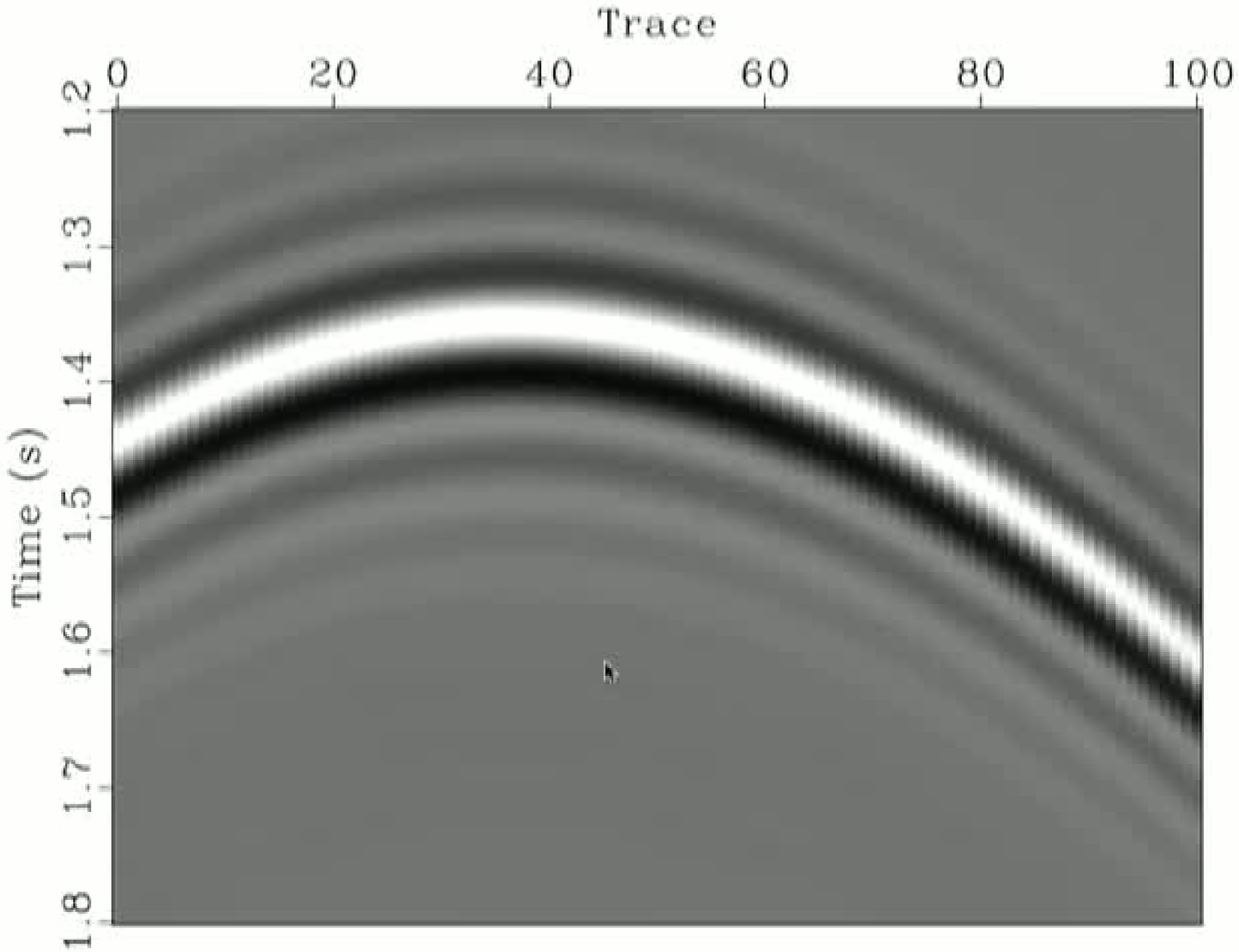
$$\left(\frac{1}{v(\mathbf{x})^2}\frac{\partial^2}{\partial t^2} - \nabla_{\mathbf{x}}^2\right) p(\mathbf{x}, t; \mathbf{x}_s) = \delta(\mathbf{x} - \mathbf{x}_s) w(t);$$

$$p \equiv 0, t << 0$$

Predicted data for source $\mathbf{x}_s$, receiver $\mathbf{x}_r = p(\mathbf{x}_r, t; \mathbf{x}_s)$ (solve wave equation, sample pressure field)



Target velocity field v for horizontal well survey (“cross-well on its side”)



"Observed" data for $\mathbf{x}_s = (5100\text{m}, 10\text{m})$:
v as in previous slide, $w = [2,5,15,20]$ Hz trapezoidal bandpass,
10 m $\times$ 10 m grid, 2-8 centered FD

Informal problem setting:

M = set of admissible models ("model space") = mechanical parameter fields (bulk modulus, density, $C_{ijkl}(\mathbf{x})$, energy sources...) - in this case, $v \in C^\infty(\mathbf{R}^3)$ and $w \in L^2(\mathbf{R})$ - causal

$(D, \langle \cdot, \cdot \rangle)$ = Hilbert space of observed data = $d \in L^2(\mathbf{R}^2_r \times \mathbf{R}_t \times \mathbf{R}^2_s$

$F : M \to D$ modeling (data prediction) operator = solve wave equations for pressure, displacement,..., sample at $\mathbf{x}_r, t$ (RHS = function of $\mathbf{x}_s$)

This map is *separable*: $M = V \times W$, F linear in W: $F[(v, w)] = S[v]w$, $S : V \to \mathcal{B}(W, D)$

$$F[(v, w)] = S[v]w = \{p(\mathbf{x}_r, t; \mathbf{x}_s) : (\mathbf{x}_r, \mathbf{x}_s) \in \Sigma, t \in [0, T]\}$$

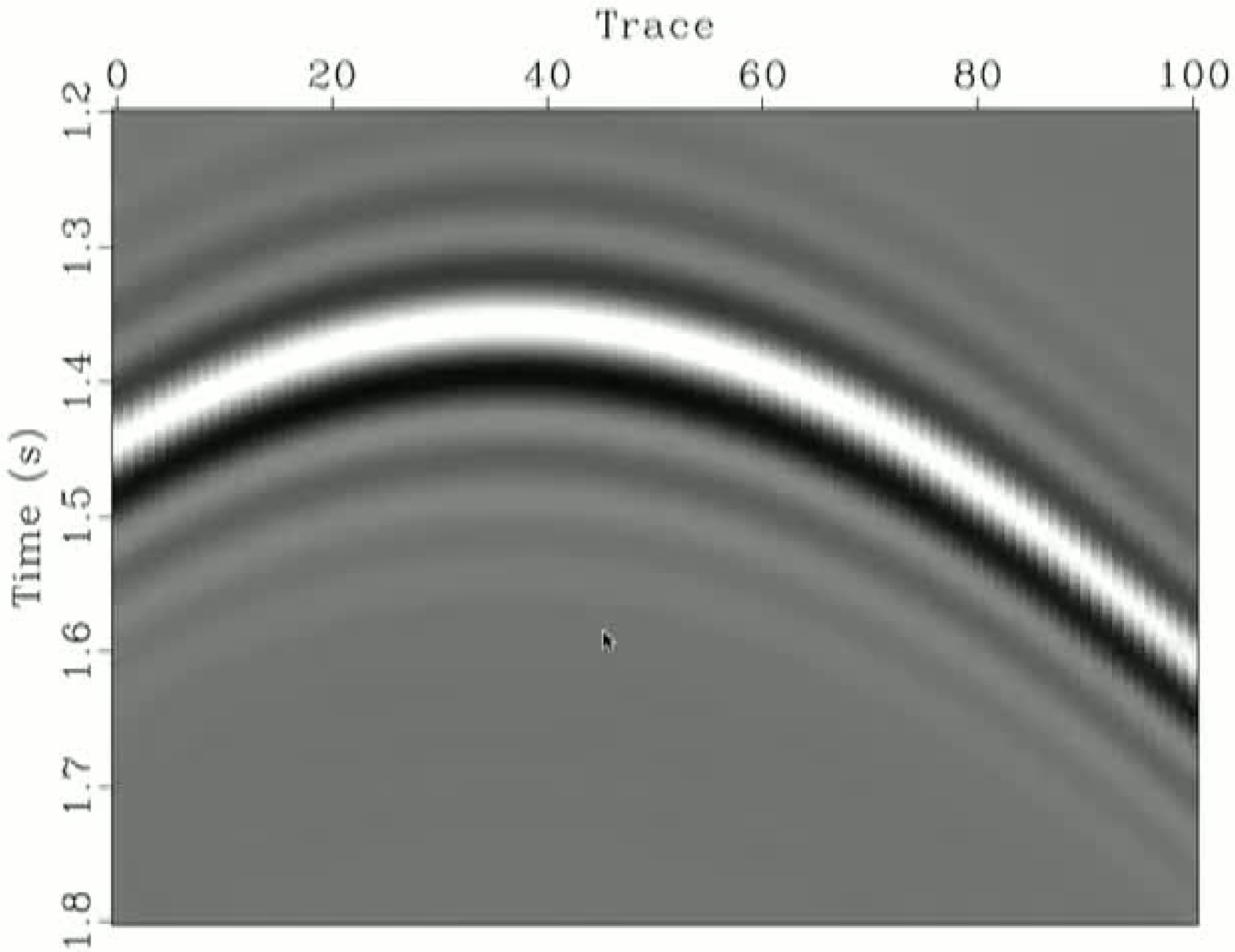
Inverse problem via Least Squares ("Full Waveform Inversion"):

given $d \in D$, find $m = (v, w) \in M$ so that

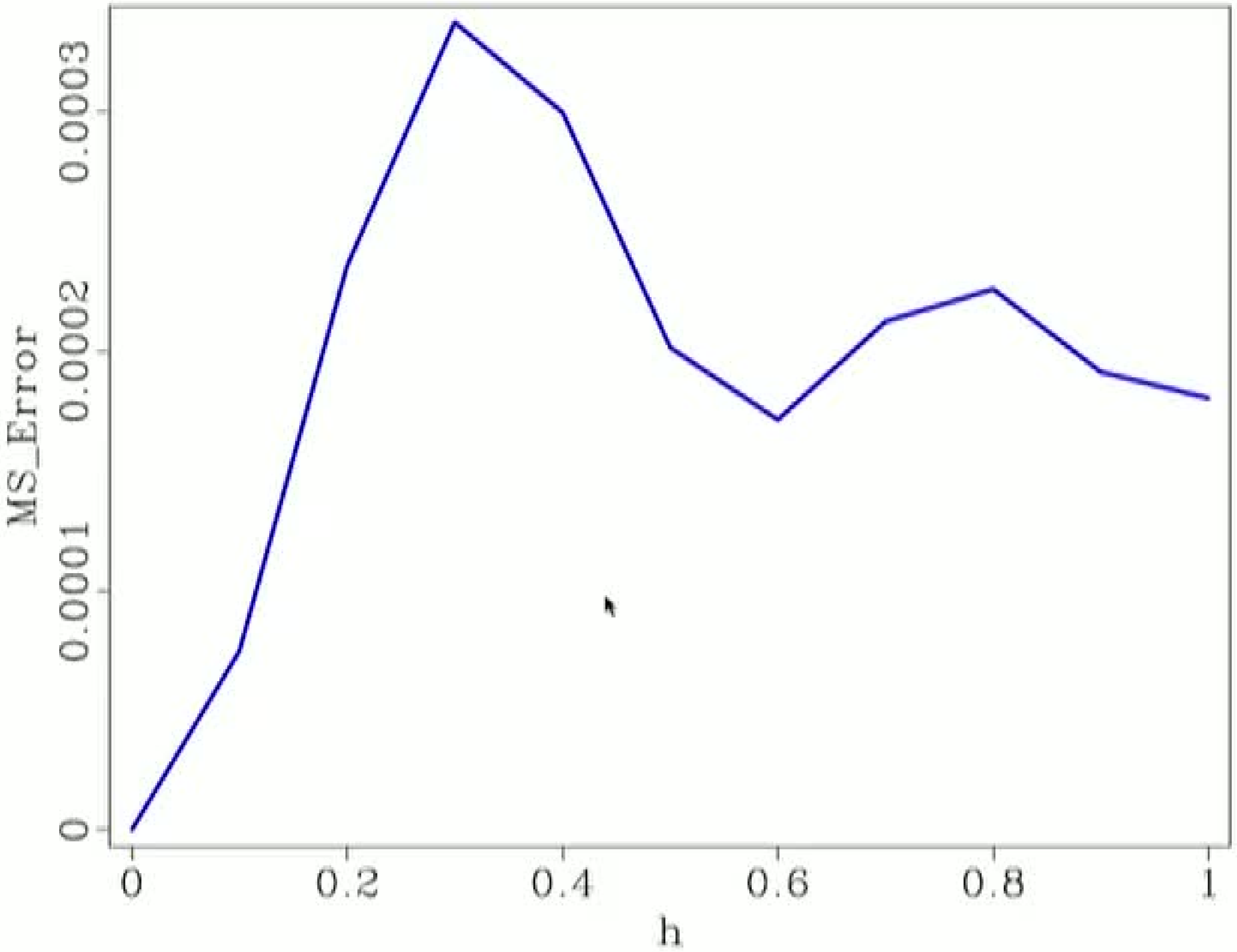
$$S[v]w \simeq d$$

by minimizing over v, w

$$J_{\mathrm{FWI}}[v, w; d] = \frac{1}{2}\|S[v]w - d\|^2 [+ \text{ regularizing terms}]$$



"Observed" data for $\mathbf{x}_s = (5100\text{m}, 10\text{m})$: target ("exact") v



$$J_{\mathrm{FWI}}[(1-h)*v+h*v_0,w;d]$$

Agenda

Transmission FWI

Model extension and Variable Projection

Accurate VPM gradient computation

Inference: local optimization of J_{FWI} (descent, Newton-like) unlikely to succeed in finding v, starting at v_0

Reason: data fit error $\approx$ 100% unless v nearly correct

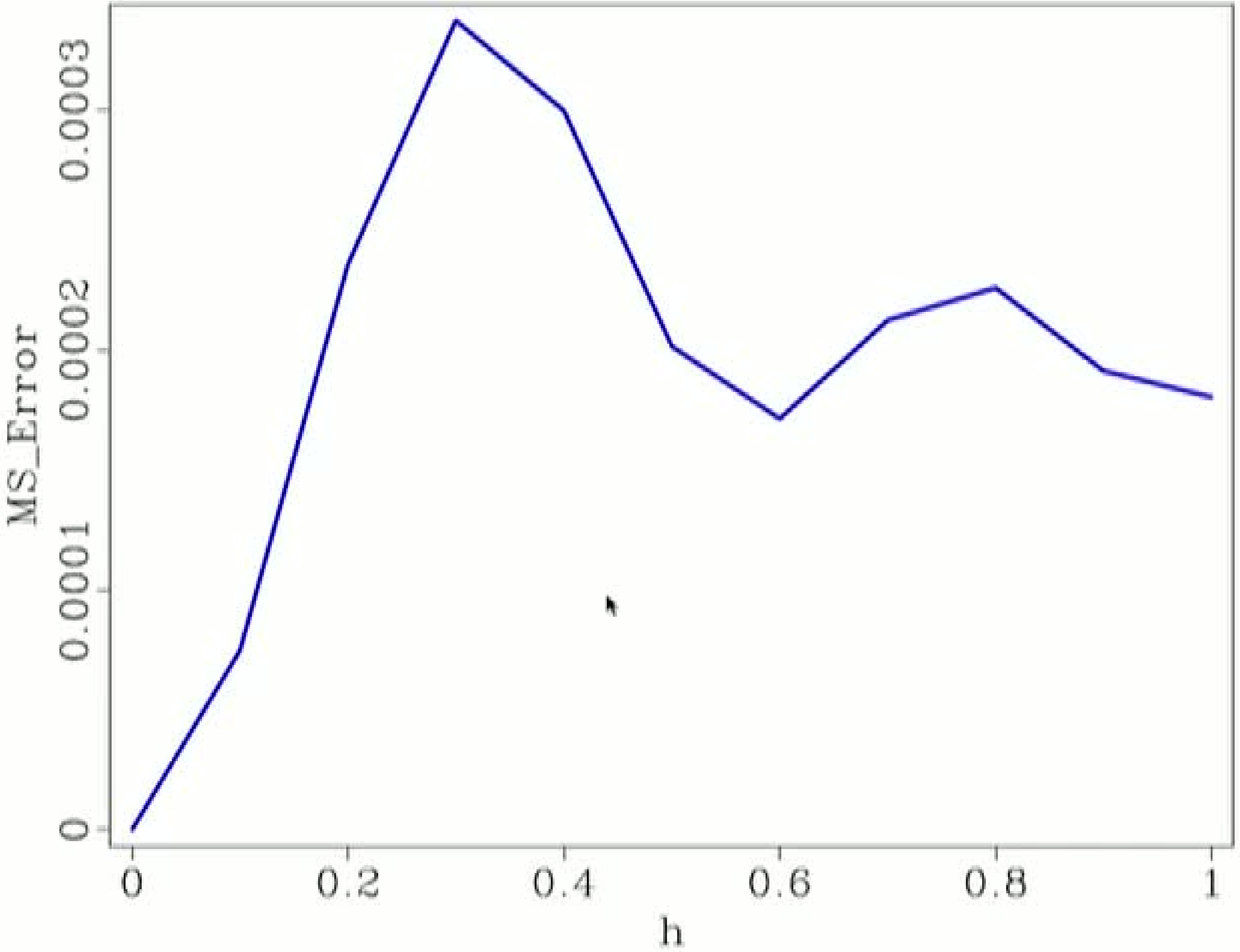
One (of many) possible fixes: enlarge set of models (model extension) to permit data fit for *any* v, penalize additional degrees of freedom via *annihilator* = operator vanishing on original ("physical") models

Agenda

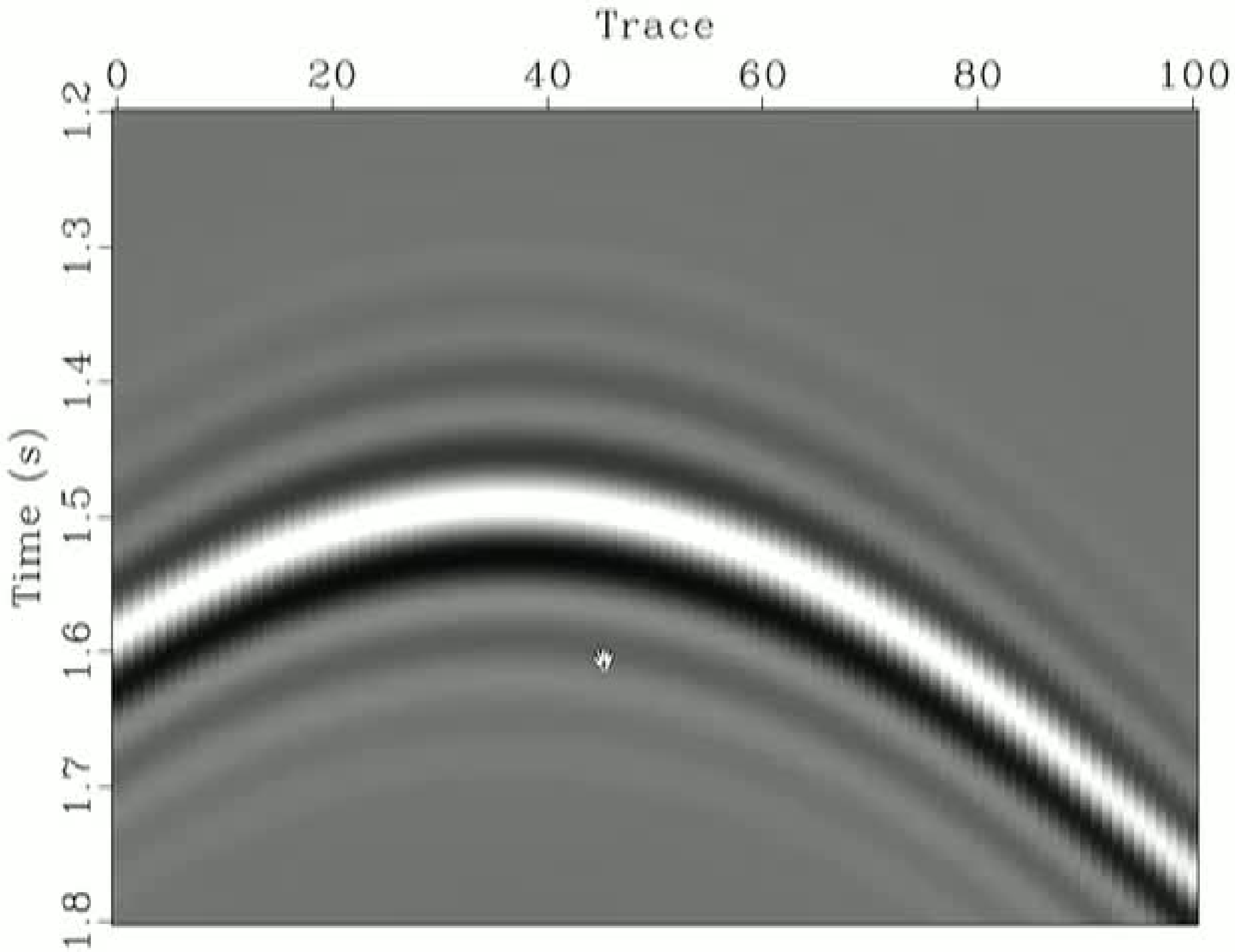
Transmission FWI

Model extension and Variable Projection

Accurate VPM gradient computation



$$J_{\mathrm{FWI}}[(1-h)*v+h*v_0, w; d]$$



"Predicted" data for $\mathbf{x}_s = (5100\text{m}, 10\text{m})$:
$v = v_0 = 2$ km/s - possible initial velocity estimate

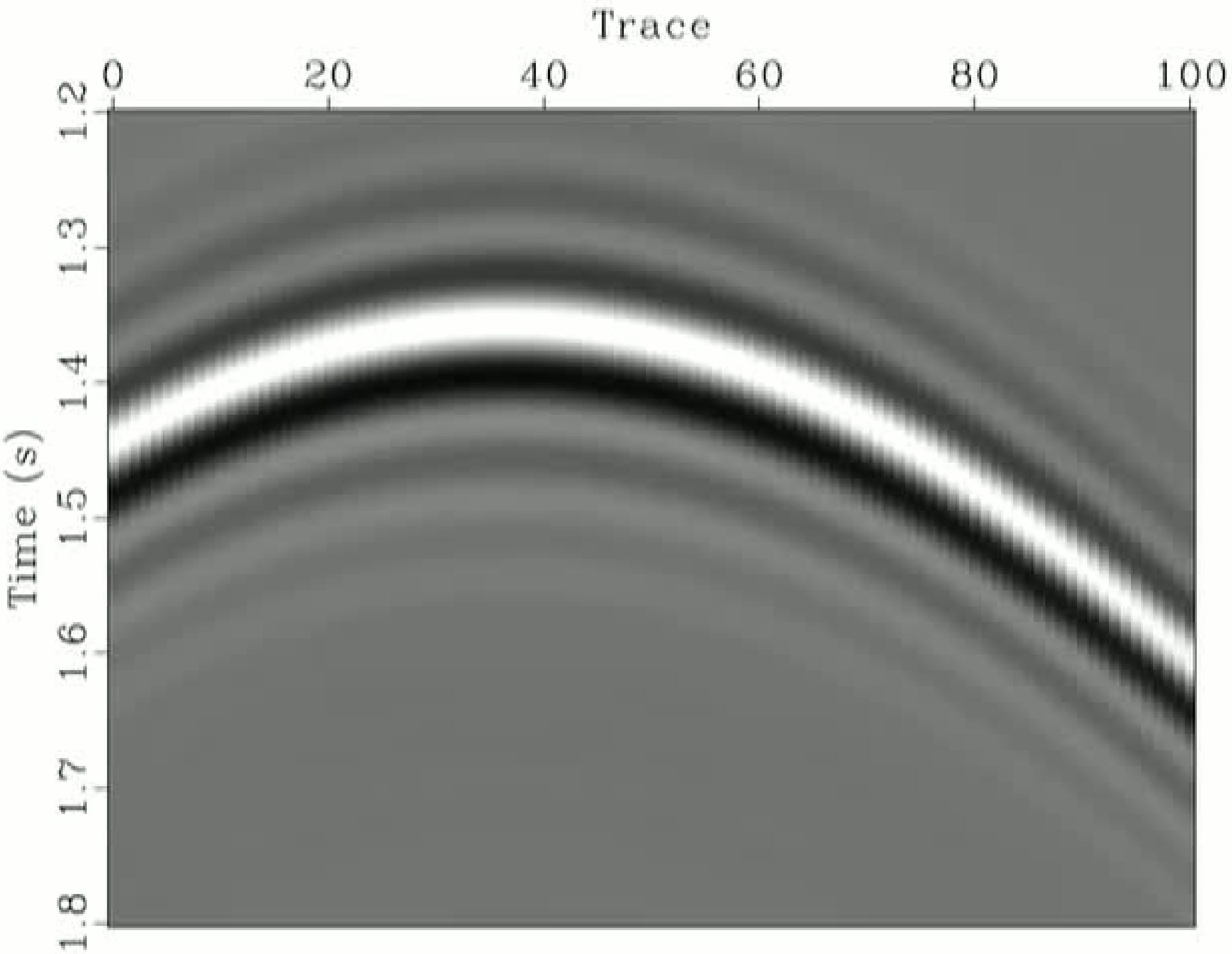
Inverse problem via Least Squares ("Full Waveform Inversion"):

given $d \in D$, find $m = (v, w) \in M$ so that

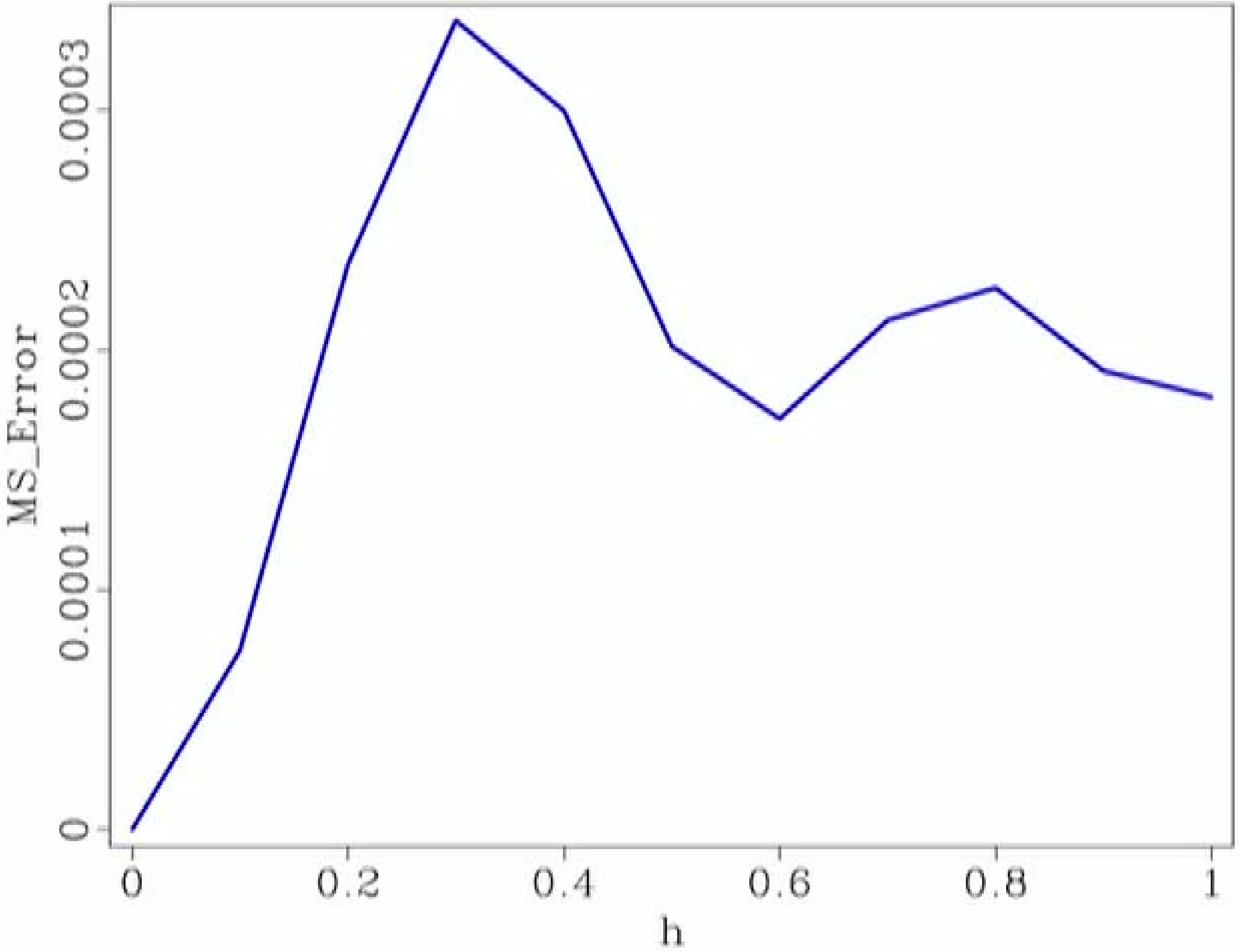
$$S[v]w \simeq d$$

by minimizing over v, w

$$J_{\mathrm{FWI}}[v, w; d] = \frac{1}{2}\|S[v]w - d\|^2 [+ \text{ regularizing terms}]$$



"Observed" data for $\mathbf{x}_s = (5100\text{m}, 10\text{m})$:
target ("exact") v



$$J_{\mathrm{FWI}}[(1-h)*v+h*v_0, w; d]$$

Inference: local optimization of J_{FWI} (descent, Newton-like) unlikely to succeed in finding v, starting at v_0

Reason: data fit error $\approx 100\%$ unless v nearly correct

One (of many) possible fixes: enlarge set of models (model extension) to permit data fit for *any* v, penalize additional degrees of freedom via *annihilator* = operator vanishing on original ("physical") models

Example of model extension: source-receiver extension - independent source wavelet $w(\mathbf{x}_r, t; \mathbf{x}_s)$ for each data trace

$\bar{M}$ = set of extended models ("model space") = $v \in C^\infty(\mathbf{R}^3)$ and $\bar{w} \in L^2(\mathbf{R}_r^2 \times \mathbf{R}_t \times \times \mathbf{R}_s^2)$ (causal)

$(D, \langle \cdot, \cdot \rangle)$ = Hilbert space of observed data = $d \in L^2(\mathbf{R}_r^2 \times \mathbf{R}_t \times \mathbf{R}_s^2$ (same!)

$\bar{F} : \bar{M} \to D$ modeling (data prediction) operator = solve wave equations for pressure, displacement,..., sample at $\mathbf{x}_r, t$ (RHS = function of $\mathbf{x}_r, \mathbf{x}_s$)

$$\bar{F}[(v,\bar{w})] = \bar{S}[v]\bar{w} = \{p(\mathbf{x}_r, t; \mathbf{x}_s) : (\mathbf{x}_r, \mathbf{x}_s) \in \Sigma, t \in [0, T]\}$$

Asymptotics (Hadamard, 3D): absent conjugate points,

$$\bar{S}[v]\bar{w}(\mathbf{x}_r, t; \mathbf{x}_s) \approx a(\mathbf{x}_r, \mathbf{x}_s)\bar{w}(\mathbf{x}_r, t - \tau[v](\mathbf{x}_r, \mathbf{x}_s); \mathbf{x}_s)$$

τ = travel time, a = geometric amplitude

$\Rightarrow$ can fit any data with small error - take
$\bar{w}(\mathbf{x}_r, t; \mathbf{x}_s) = d(\mathbf{x}_r, t + \tau[v](\mathbf{x}_r, \mathbf{x}_s); \mathbf{x}_s)/a(\mathbf{x}_r, \mathbf{x}_s)$

Example of model extension: source-receiver extension - independent source wavelet $w(\mathbf{x}_r, t; \mathbf{x}_s)$ for each data trace

$\bar{M}$ = set of extended models ("model space") = $v \in C^\infty(\mathbf{R}^3)$ and $\bar{w} \in L^2(\mathbf{R}_r^2 \times \mathbf{R}_t \times \times \mathbf{R}_s^2)$ (causal)

$(D, \langle \cdot, \cdot \rangle)$ = Hilbert space of observed data = $d \in L^2(\mathbf{R}_r^2 \times \mathbf{R}_t \times \mathbf{R}_s^2$ (same!)

$\bar{F} : \bar{M} \to D$ modeling (data prediction) operator = solve wave equations for pressure, displacement,..., sample at $\mathbf{x}_r, t$ (RHS = function of $\mathbf{x}_r, \mathbf{x}_s$)

$$\bar{F}[(v,\bar{w})] = \bar{S}[v]\bar{w} = \{p(\mathbf{x}_r, t; \mathbf{x}_s) : (\mathbf{x}_r, \mathbf{x}_s) \in \Sigma, t \in [0, T]\}$$

Asymptotics (Hadamard, 3D): absent conjugate points,

$$\bar{S}[v]\bar{w}(\mathbf{x}_r, t; \mathbf{x}_s) \approx a(\mathbf{x}_r, \mathbf{x}_s)\bar{w}(\mathbf{x}_r, t - \tau[v](\mathbf{x}_r, \mathbf{x}_s); \mathbf{x}_s)$$

τ = travel time, a = geometric amplitude

$\Rightarrow$ can fit any data with small error - take
$\bar{w}(\mathbf{x}_r, t; \mathbf{x}_s) = d(\mathbf{x}_r, t + \tau[v](\mathbf{x}_r, \mathbf{x}_s); \mathbf{x}_s)/a(\mathbf{x}_r, \mathbf{x}_s)$

$\Rightarrow$ fit error does not constrain v - restore constraint by penalizing extended degrees of freedom (dependence of $\bar{w}$ on $\mathbf{x}_r, \mathbf{x}_s$)

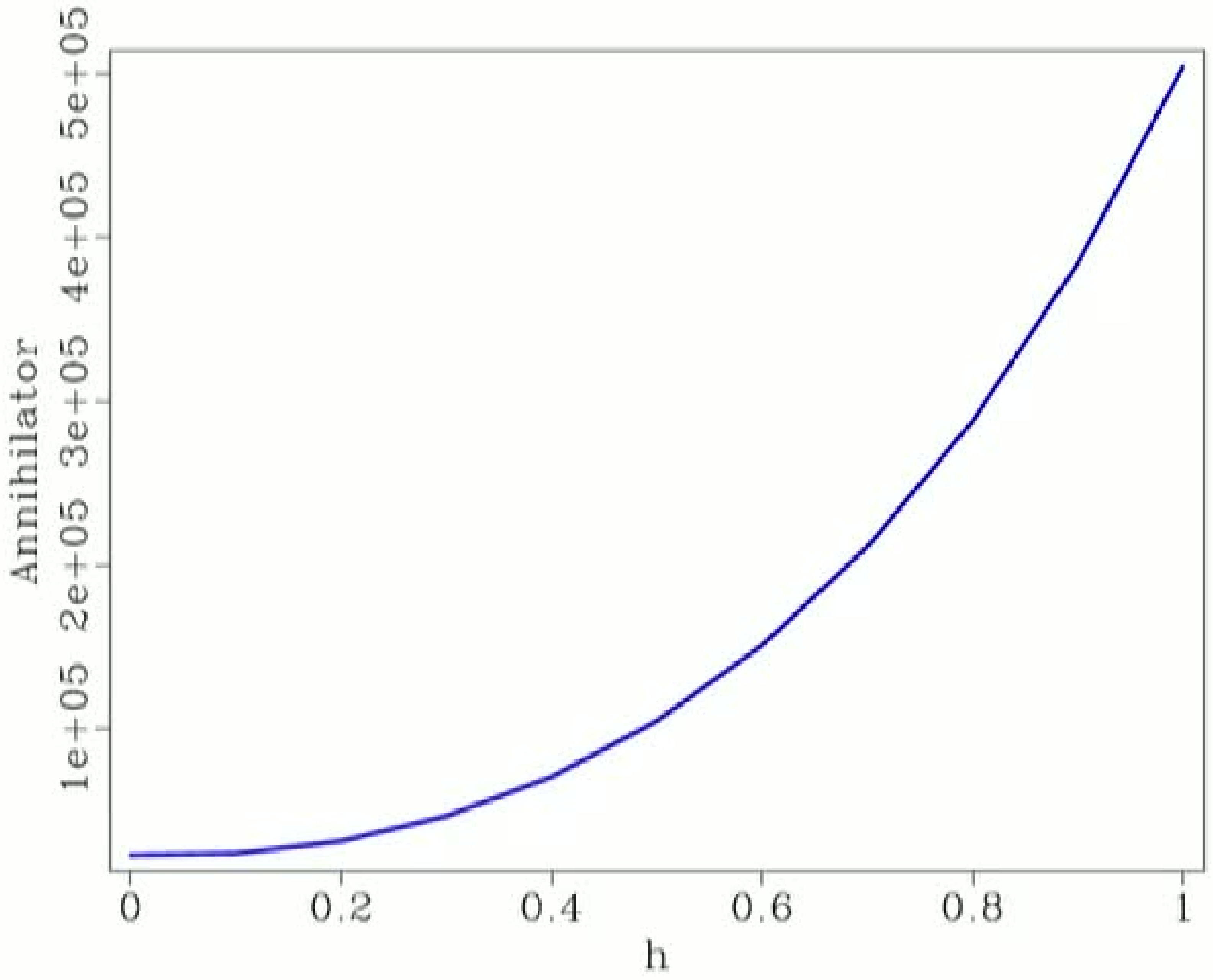
Many choices - eg. presume correct w known, pre-process (deconvolve) so that $w(t) \approx \delta(t)$. Popular choice of annihilator (Plessix et al. 00, Warner & Guatsch 14): $A\bar{w}(\mathbf{x}_r, t; \mathbf{x}_s) \equiv t\bar{w}(\mathbf{x}_r, t; \mathbf{x}_s) \approx 0$ if $\bar{w}(..., t) \approx \delta(t)$

Penalized least squares: minimize

$$J_{\mathrm{ELS}}[v, \bar{w}] = \frac{1}{2}(\|\bar{S}[v]\bar{w} - d\|^2 + \alpha^2\|A\bar{w}\|^2)$$

Reduced objective (Variable Projection Method, Golub & Pereyra 73, 03)

$$\bar{w}[v] = \mathrm{argmin}_{\bar{w}} J_{\mathrm{ELS}}[v, \bar{w}]; \;\; J_{\mathrm{ELS}}^{\mathrm{red}}[v] = J_{\mathrm{ELS}}[v, \bar{w}[v]]$$



$$J_{\mathrm{ELS}}^{\mathrm{red}}[(1-h)*v+h*v_0;d]$$

Agenda

Transmission FWI

Model extension and Variable Projection

Accurate VPM gradient computation

The Problem:

(1) nice VPM gradient formula

$$\nabla_v J_{\mathrm{ELS}}^{\mathrm{red}}[v;d] = D_v(\bar{S}[v]\bar{w})^T(\bar{S}[v]\bar{w} - d))_{\bar{w}=\bar{w}[v]}$$

requires $D_v\bar{S}$ - one order less regular than $\bar{S}$:

$$D_v\bar{S}[v]\bar{w}(...,t) \approx D_v(a[v]\bar{w}(...,t-\tau[v]) + ...)$$

$$= -(a[v]D_v\tau[v])\frac{\partial \bar{w}}{\partial t} + ...$$

(2) $\bar{w}[v]$ estimated iteratively $= \lim_{a\to\infty} \bar{w}_a$ in L^2

$\Rightarrow$ obvious gradient approximation

$$\nabla_v J_{\mathrm{ELS}}^{\mathrm{red}}[v;d] \approx D_v(\bar{S}[v]\bar{w})^T(\bar{S}[v]\bar{w} - d))_{\bar{w}=\bar{w}_a}$$

may not converge (Kern & S. 94, Yin Zhang thesis 16)

A Solution:

Part of VPM gradient/deriv construction is harmless:

$$D_v J^{\mathrm{red}}_{\mathrm{ELS}}[v;d] = \langle D\bar{S}[v]\bar{w}, \bar{S}[v]\bar{w} - d\rangle_{\bar{w}=\bar{w}[v]}$$

$$= \mathbf{1} + \mathbf{2}$$

$$\mathbf{1} = \langle D_v\bar{S}[v]\bar{w}, \bar{S}[v]\bar{w}\rangle_{\bar{w}=\bar{w}[v]} = \left(D_v\frac{1}{2}\langle\bar{w}, \bar{S}[v]^T\bar{S}[v]\bar{w}\rangle\right)_{\bar{w}=\bar{w}[v]}$$

$\bar{S}[v]^T\bar{S}[v]$ is ΨDO w symbol smooth in $v \Rightarrow$

$$\mathbf{1} = \lim_{a\to\infty}\langle D_v\bar{S}[v]\bar{w}, \bar{S}[v]\bar{w}\rangle_{\bar{w}=\bar{w}a}$$

$\bar{S}[v]$ is approx. unitary in appropriate weighted L^2 norm:

$$\bar{S}[v]^T \bar{S}[v] = a[v]^2 I \text{ modulo } OPS^{-1}$$

Define

$$\langle \bar{w}_1, \bar{w}_2 \rangle_v = \langle \bar{w}_1, a[v]^2 \bar{w}_2 \rangle_{L^2}$$

denote adjoint of $\bar{S}[v]$ wrt $\langle \cdot, \cdot \rangle_v$ by $\bar{S}[v]^\dagger = a[v]^{-2} \bar{S}[v]^T$

$$\Rightarrow \bar{S}[v]^\dagger \bar{S}[v] = I + K[v],\ K[v] \in OPS^{-1}$$

Weighted normal residual for $\bar{w}_a$:

$$g_a = (\bar{S}[v]^\dagger \bar{S}[v] + \alpha^2 A^\dagger A)(\bar{w}_a - \bar{w}[v])$$

$$= (I + \alpha^2 A^\dagger A)(\bar{w}_a - \bar{w}[v]) + K[v](\bar{w}_a - \bar{w}[v]), \;\; K[v] \in OPS^{-1}$$

$\Rightarrow$

$$\bar{w}_a - \bar{w}[v] = (I + \alpha^2 A^\dagger A)^{-1}(K[v](\bar{w}_a - \bar{w}[v]) - g_a)$$

$\Rightarrow$

$$\tilde{w}_a \equiv \bar{w}_a + (I + \alpha^2 A^\dagger A)^{-1} g_a = \bar{w}[v] \text{ modulo error } \rightarrow 0 \text{ in } H^1_t$$

$\Rightarrow$

$$\mathbf{2} = -\lim_{a \to \infty} \langle D\bar{S}[v]\bar{w}, d \rangle_{\bar{w}=\tilde{w}_a}$$

$\bar{S}[v]$ is approx. unitary in appropriate weighted L^2 norm:

$$\bar{S}[v]^T \bar{S}[v] = a[v]^2 I \text{ modulo } OPS^{-1}$$

Define

$$\langle \bar{w}_1, \bar{w}_2 \rangle_v = \langle \bar{w}_1, a[v]^2 \bar{w}_2 \rangle_{L^2}$$

denote adjoint of $\bar{S}[v]$ wrt $\langle \cdot, \cdot \rangle_v$ by $\bar{S}[v]^\dagger = a[v]^{-2} \bar{S}[v]^T$

$$\Rightarrow \bar{S}[v]^\dagger \bar{S}[v] = I + K[v], \; K[v] \in OPS^{-1}$$

Weighted normal residual for $\bar{w}_a$:

$$g_a = (\bar{S}[v]^\dagger \bar{S}[v] + \alpha^2 A^\dagger A)(\bar{w}_a - \bar{w}[v])$$

$$= (I + \alpha^2 A^\dagger A)(\bar{w}_a - \bar{w}[v]) + K[v](\bar{w}_a - \bar{w}[v]),\ K[v] \in OPS^{-1}$$

$\Rightarrow$

$$\bar{w}_a - \bar{w}[v] = (I + \alpha^2 A^\dagger A)^{-1}(K[v](\bar{w}_a - \bar{w}[v]) - g_a)$$

$\Rightarrow$

$$\tilde{w}_a \equiv \bar{w}_a + (I + \alpha^2 A^\dagger A)^{-1} g_a = \bar{w}[v] \text{ modulo error } \to 0 \text{ in } H_t^1$$

$\Rightarrow$

$$\mathbf{2} = -\lim_{a\to\infty} \langle D\bar{S}[v]\bar{w}, d\rangle_{\bar{w}=\tilde{w}_a}$$

Upshot: $\bar{w}_a \rightarrow \bar{w}[v] \Rightarrow$

$$\nabla_v J_{\mathrm{ELS}}^{\mathrm{red}}[v; d] = \lim_{a \rightarrow \infty} D_v(\bar{S}[v]\bar{w})^T(\bar{S}[v]\bar{w} - d))_{\bar{w}=\tilde{w}_a}$$

where

$$\tilde{w}_a \equiv \bar{w}_a + (I + \alpha^2 A^\dagger A)^{-1} g_a$$

Exercise for audience: accounting for v-dep of weighted norms does not change conclusion

Computability:

- g_a, $\tilde{w}_a$ computable *in principle* - in practice, feasible computation $\Rightarrow$ additional $O(\lambda)$ error
- doesn't matter because $\|D\bar{S}\|_{L^2}$ actually $= O(\lambda^{-1})$

And so on...

Conjugate points $\Rightarrow$ source-receiver extension loses unitary property (Song & S. 94, Huang & S. 17)

Surface source extension (Huang & S. 18) recovers unitary property - but only microlocally

- reliable gradient comp requires (computable) ΨDO cutoff

Reflection inversion - subsurface offset extension (Stolk & de Hoop 01, Shen et al. 03,...) microlocally unitary at physical subspace (ten Kroode 12, Hou & S. 15)

- additional error absorbed in frequency continuation scheme (Fu & S. 17)