Released to public domain under Creative Commons license type BY (https://creativecommons.org/licenses/by/4.0).
Copyright (c) 2018 SINBAD consortium - SLIM group @ The University of British Columbia.

Automatic salt delineation — Wavefield Reconstruction Inversion with convex constraints

Ernie Esser, Lluis Guash*, and Felix J. Herrmann

*Sub Salt Solutions Ltd

SLIM
University of British Columbia

John “Ernie” Esser (May 19, 1980 – March 8, 2015)

Motivation

Wave-equation based inversions suffer from local minima

- especially detrimental for high-contrast & high-velocity unconformities (salt & basalt)
- often able to find top of the unconformities but not the bottom
- gradients tend to point in the wrong direction

Borrow

- edge-preserving regularization from image processing & compressive sensing
- hinge-loss functions from machine learning
- recent developments from constrained optimization

Contributions

New computationally feasible wave-equation based inversion scheme

- leverages special structure Gauss-Newton Hessian of WRI
- relaxes convex constraints to steer away from local minima
- yields sharp delineations of top and bottom of high-velocity inclusions

Use relaxing convex constraints to

- impose bounds on the inverted velocity/slowness
- introduce sharp reflectors to delineate the top of the unconformities via TV-norms
- **software release:** Total Variation Regularized Wavefield Reconstruction Inversion

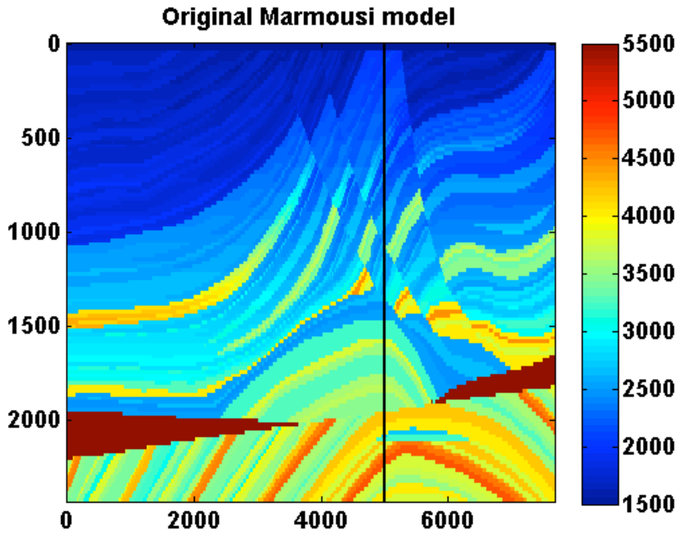
In collaboration w/ Sub Salt Solutions Ltd

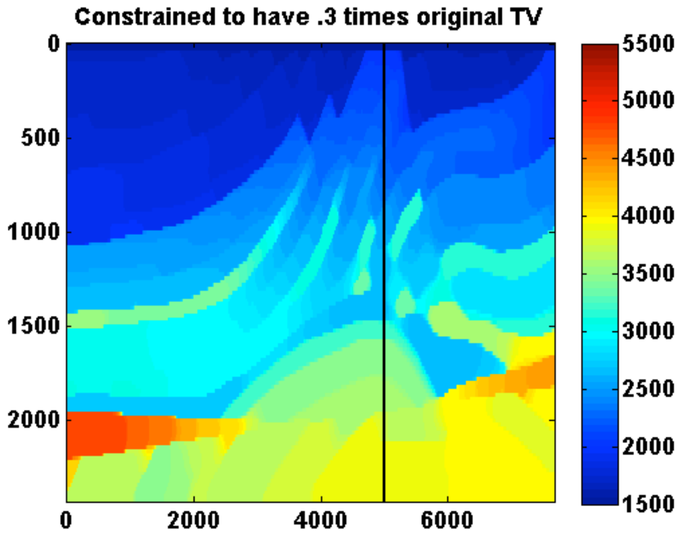
- control the sign of the model updates in certain directions via hinge-loss functions
- demonstration on BP Benchmark with both WRI & adjoint-state FWI

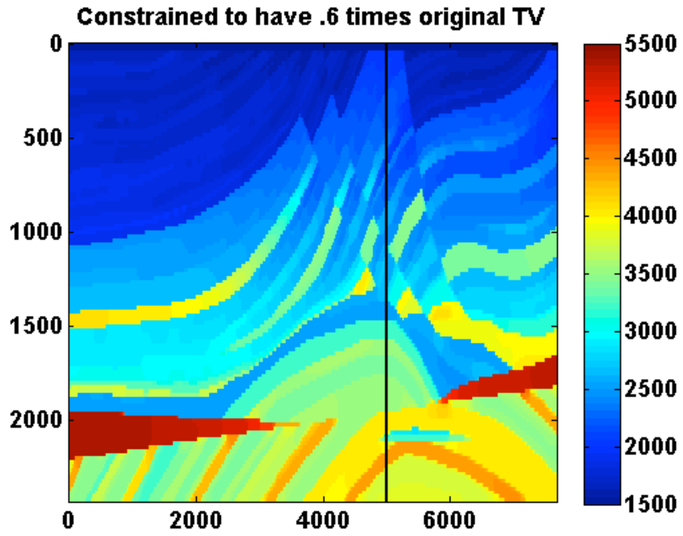
Projections on the convex sets

$v_{\min} = 1500$, $v_{\max} = 5500$, and $\tau = \{0.3\tau_0, 0.6\tau_0\}$

$$\Pi_C(\mathbf{m}_0) = \arg\min_{\mathbf{m}} \frac{1}{2}\|\mathbf{m} - \mathbf{m}_0\|^2 \quad \text{subject to} \quad \mathbf{m}_i \in [B_i^l, B_i^u] \text{ and } \|\mathbf{m}\|_{TV} \leq \tau$$

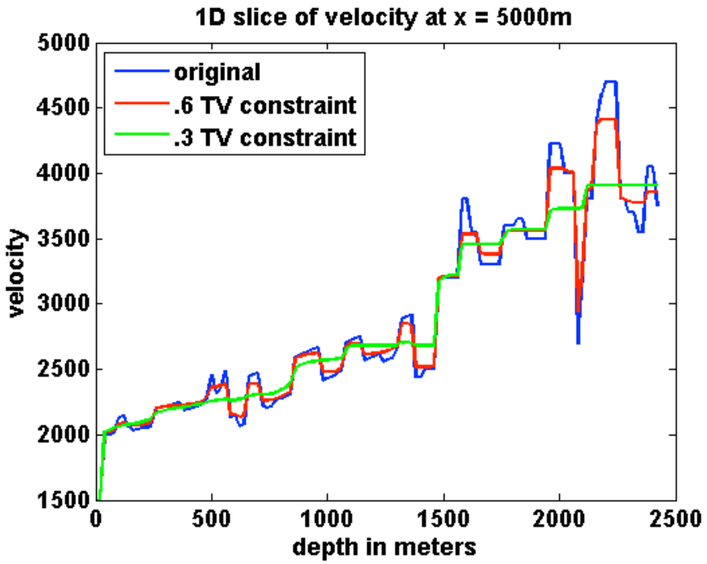


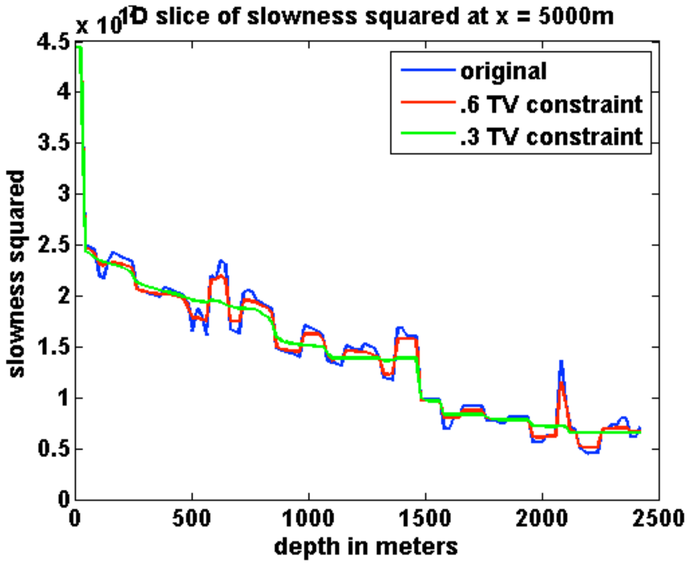




Projections on the convex sets

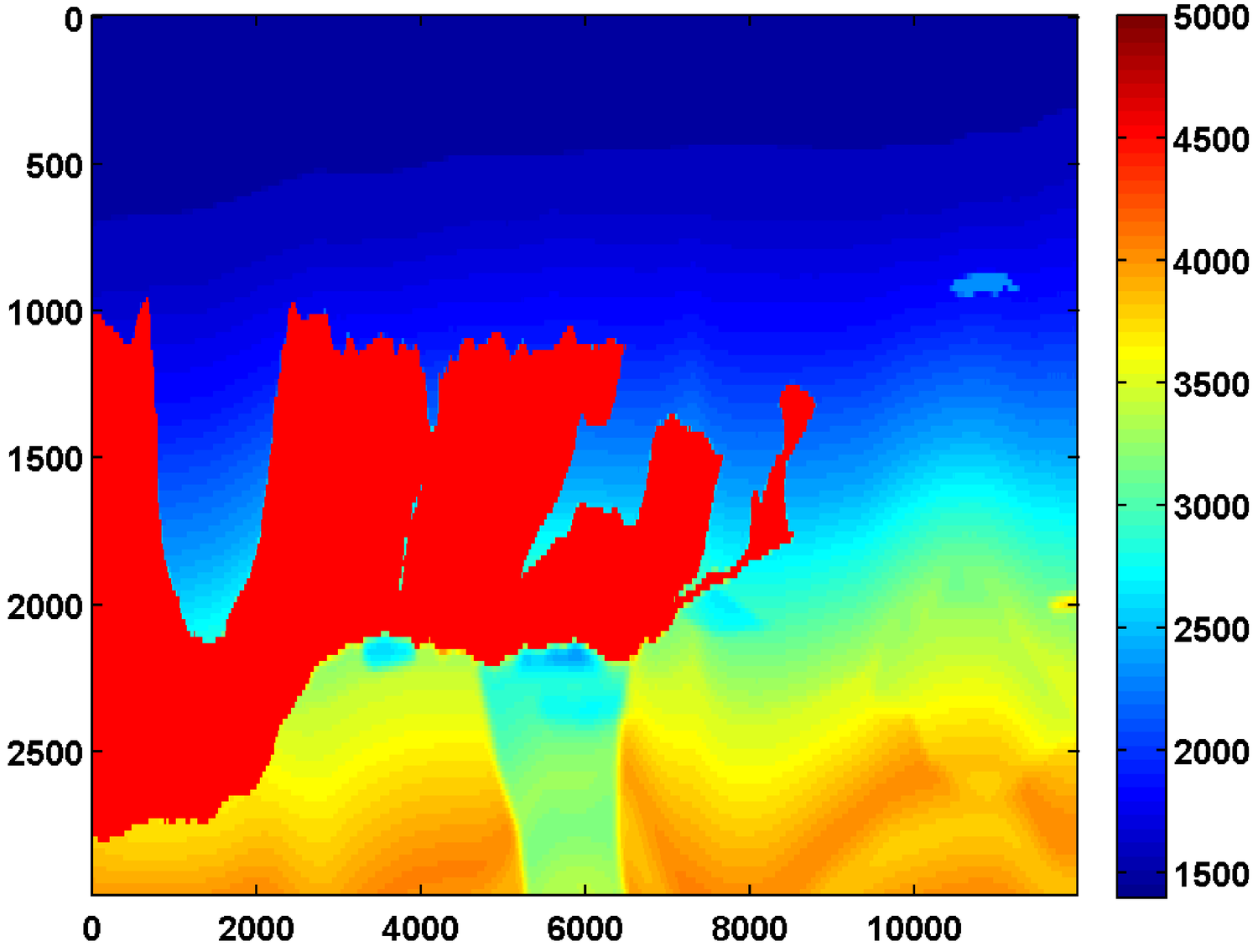
$v_{\min} = 1500$, $v_{\max} = 5500$, and $\tau = \{0.3\tau_0, 0.6\tau_0\}$



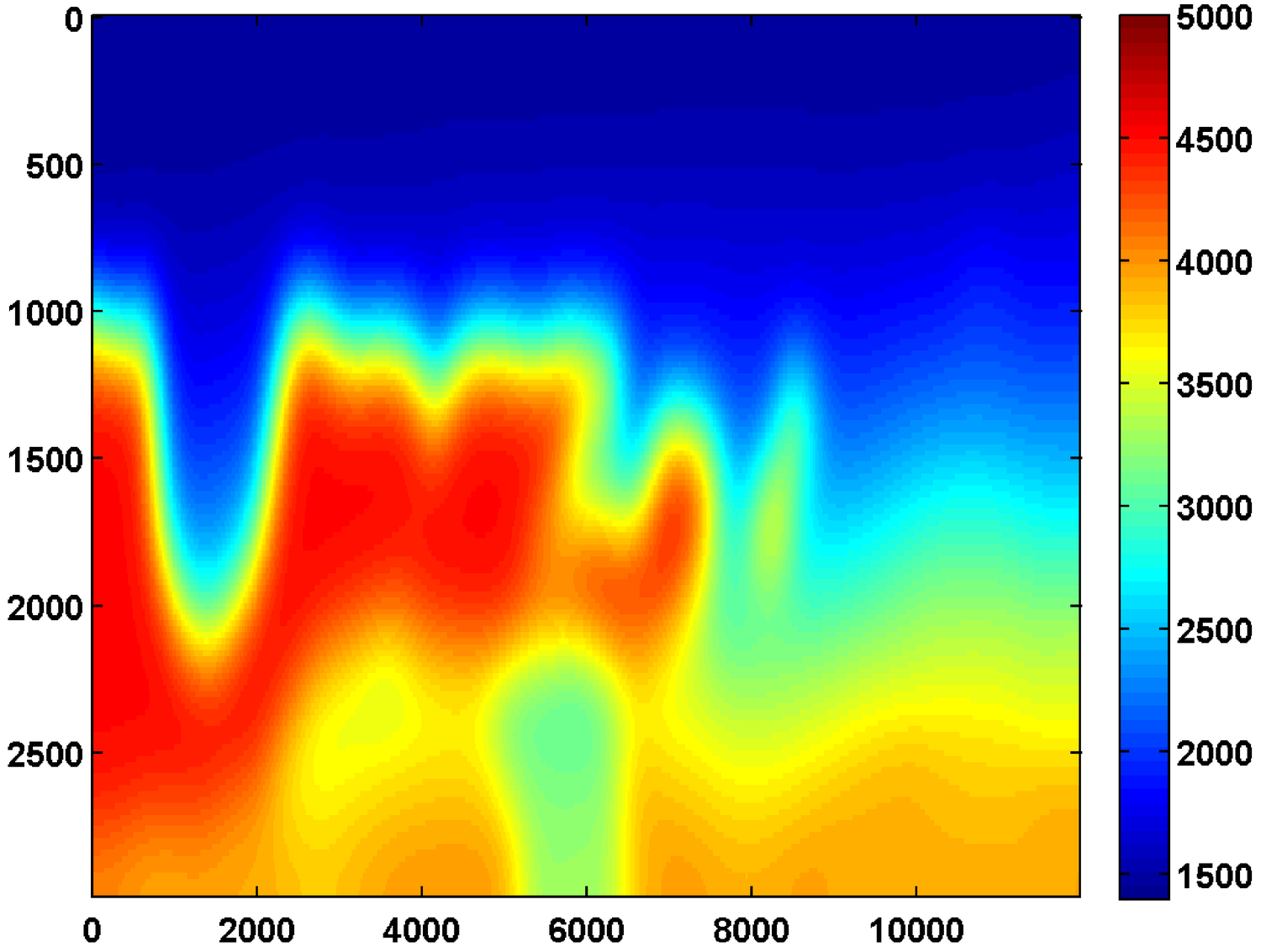


- preserves the reflectors (blocky)
- relaxing the constraint allows for more detail

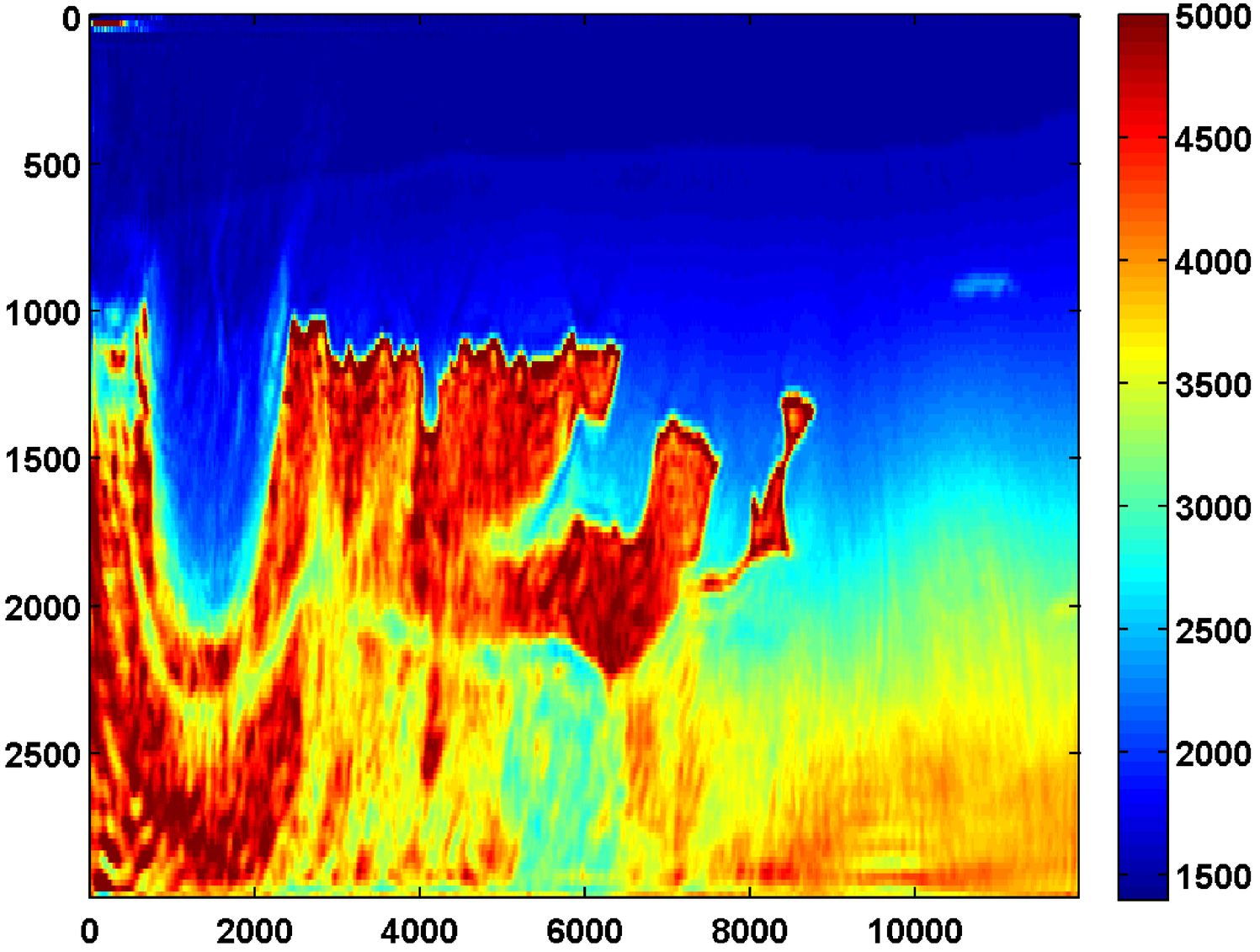
Top Left Portion of BP 2004 Velocity Benchmark



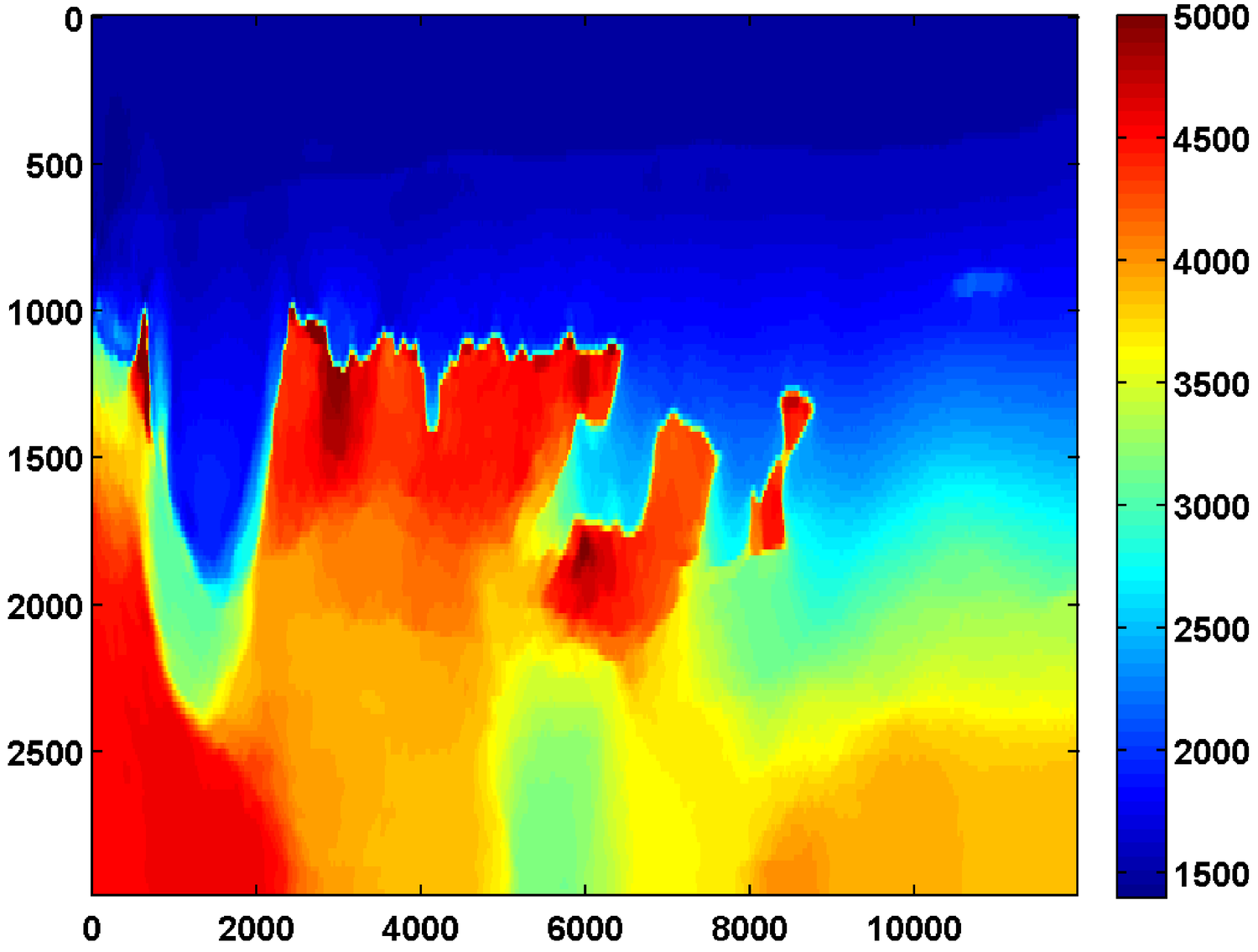
Smooth Initial Velocity Model



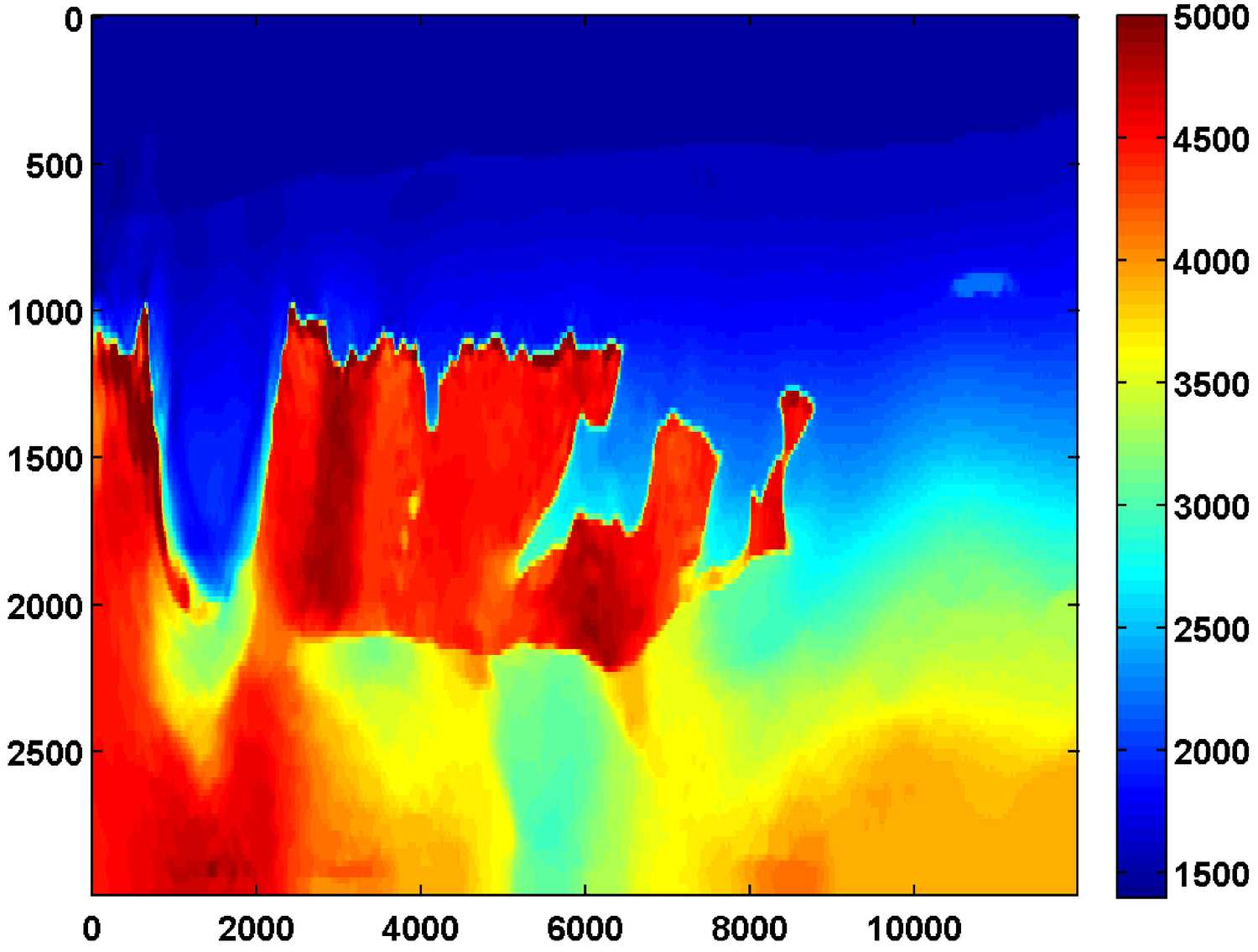
WRI without TV



TV Constrained WRI



Second Pass of TV Constrained WRI with Relaxed Constraint



Outline

Brief introduction Wavefield Reconstruction Inversion

Introduction of convex constraints

- bound constraints
- total-variation & bounds constraints
- hinge-loss & total-variation & bounds constraints

Discuss the iterative algorithm

Application to BP benchmark

[Heinkenschloss, '98 , Haber, '00]

PDE-constrained optimization
all-at-once full-space approach

simulated data

simulated wavefield

$$\min_{\mathbf{m},\mathbf{u}} \sum_{i=1}^{M} \|P_i\mathbf{u}_i - \mathbf{d}_i\|_2^2 \quad \text{s.t.} \quad A_i(\mathbf{m})\mathbf{u}_i = \mathbf{q}_i$$

observed data

Helmholtz equation

source

- avoids having to solve the PDE explicitly
- sparse (GN) Hessian
- requires storing all variables (**m,u**)
- does **not** scale to industry-scale seismic problems

[Tarantola ’84; Pratt, ’98; Plessix, ’06]

Adjoint-state/reduced-space formulation

Elimination of the constraint leads for all sources to

$$\min_{\mathbf{m}} \phi_{\text{red}}(\mathbf{m}) = \sum_{i=1}^{M} \| P_i A_i(\mathbf{m})^{-1} \mathbf{q}_i - \mathbf{d}_i \|_2^2$$

- no need to store all wavefields (block-elimination)
- suitable for black-box optimization (e.g., l-BFGS)
- need to solve forward & adjoint PDEs
- very non-linear dependence on earth model (**m**)
- dense (GN) Hessian, involves additional PDE solves
- **reliance on accurate starting models to avoid cycle skipping**

[Bertsekas, '96; Wright, '00; van Leeuwen & FJH, '13]

Penalty formulation
– Wavefield-Reconstruction Inversion (WRI)

Instead of eliminating, we add constraints as penalties–i.e.,

$$\min_{\mathbf{m},\mathbf{u}} \phi_\lambda(\mathbf{m},\mathbf{u}) = \sum_{i=1}^{M} \|P\mathbf{u}_i - \mathbf{d}_i\|_2^2 + \lambda^2 ||A_i(\mathbf{m})\mathbf{u}_i - \mathbf{q}_i\|_2^2$$

coincides with original problem when $\lambda \uparrow \infty$

- **no** need to store all the fields (**u**)
- **no** adjoint solves
- sparse approximation of Gauss-Newton Hessian for small λ
- less non-linear in **m**
- need to solve data-augmented wave equation

[Aravkin & van Leeuwen, '12; van Leeuwen & FJH, '13]

Variable projection

Solve data-augmented wave equation for each source

$$\begin{pmatrix} P_i \\ \lambda A_i(\mathbf{m}) \end{pmatrix} \mathbf{u}_{i,\lambda} \approx \begin{pmatrix} \mathbf{d}_i \\ \lambda \mathbf{q}_i \end{pmatrix}$$

Define reduced objective with proxy wavefields

$$\phi_\lambda(\mathbf{m}) = \phi_\lambda(\mathbf{m}, \bar{\mathbf{u}}_\lambda) = \|P\bar{\mathbf{u}}_\lambda - \mathbf{d}\|_2^2 + \lambda^2 ||A(\mathbf{m})\bar{\mathbf{u}}_\lambda - \mathbf{q}\|_2^2$$

Tristan van Leeuwen and Felix J. Herrmann, "**Mitigating local minima in full-waveform inversion by expanding the search space**", *Geophysical Journal International*, vol. 195, p. 661-667, 2013.

WRI method

for each source i

$$\text{solve} \quad \begin{pmatrix} P_i \\ \lambda A_i(\mathbf{m}) \end{pmatrix} \mathbf{u}_{\lambda,i} \approx \begin{pmatrix} \mathbf{d}_i \\ \lambda \mathbf{q}_i \end{pmatrix}$$

$$\mathbf{g} = \mathbf{g} + \lambda^2 \omega^2 \text{diag}(\bar{\mathbf{u}}_{i,\lambda})^* (A(\mathbf{m})\bar{\mathbf{u}}_{i,\lambda} - \mathbf{q}_i)$$

end

$$\mathbf{m} = \mathbf{m} - \alpha \mathbf{g}$$

correlation proxy wavefield & PDE residual

Conventional method

for each source i

$$\text{solve} \quad A(\mathbf{m})\mathbf{u}_i = \mathbf{q}_i$$

$$\text{solve} \quad A(\mathbf{m})^* \mathbf{v}_i = P_i^* (P_i \mathbf{u}_i - \mathbf{d}_i)$$

$$\mathbf{g} = \mathbf{g} + \omega^2 \text{diag}(\mathbf{u}_i)^* \mathbf{v}_i$$

end

$$\mathbf{m} = \mathbf{m} - \alpha \mathbf{g}$$

correlation wavefield & data residual

WRI method

for each source i

$$\text{solve} \quad \begin{pmatrix} P_i \\ \lambda A_i(\mathbf{m}) \end{pmatrix} \mathbf{u}_{\lambda,i} \approx \begin{pmatrix} \mathbf{d}_i \\ \lambda \mathbf{q}_i \end{pmatrix}$$

$$\mathbf{g} = \mathbf{g} + \lambda^2 \omega^2 \text{diag}(\bar{\mathbf{u}}_{i,\lambda})^* (A(\mathbf{m})\bar{\mathbf{u}}_{i,\lambda} - \mathbf{q}_i)$$

$$H_{GN} = H_{GN} + \lambda^2 \omega^4 \text{diag}(\mathbf{u}_i)^* \text{diag}(\mathbf{u}_i)$$

end

$$\mathbf{m} = \mathbf{m} - \alpha H_{GN}^{-1} \mathbf{g}$$

diagonal Hessian = pseudo Hessian

Conventional method

for each source i

solve $A(\mathbf{m})\mathbf{u}_i = \mathbf{q}_i$

solve $A(\mathbf{m})^* \mathbf{v}_i = P_i^* (P_i \mathbf{u}_i - \mathbf{d}_i)$

$$\mathbf{g} = \mathbf{g} + \omega^2 \text{diag}(\mathbf{u}_i)^* \mathbf{v}_i$$

end

$$\mathbf{m} = \mathbf{m} - \alpha \mathbf{g}$$

dense Hessian & too expensive

WRI

Replace PDE-constrained formulation for FWI:

simulated data

simulated wavefield

$$\min_{\mathbf{m},\mathbf{u}} \sum_{sv} \frac{1}{2} \|P\mathbf{u}_{sv} - \mathbf{d}_{sv}\|^2 \quad \text{such that} \quad A_v(\mathbf{m})\mathbf{u}_{sv} = \mathbf{q}_{sv}$$

observed data

Helmholtz equation

source

WRI

by the penalty formulation

$$\min_{\mathbf{m},\mathbf{u}} \sum_{sv} \frac{1}{2}\|P\mathbf{u}_{sv} - \mathbf{d}_{sv}\|^2 + \frac{\lambda^2}{2}\|A_v(\mathbf{m})\mathbf{u}_{sv} - \mathbf{q}_{sv}\|^2$$

and solve for the nth iteration

$$\bar{\mathbf{u}}_{sv} = \arg\min_{\mathbf{u}_{sv}} \frac{1}{2}\|P\mathbf{u}_{sv} - \mathbf{d}_{sv}\|^2 + \frac{\lambda^2}{2}\|A_v(\mathbf{m}^n)\mathbf{u}_{sv} - \mathbf{q}_{sv}\|^2$$

followed by computing the gradient

$$\mathbf{g}^n = \sum_{sv} \mathrm{Re}\left\{\lambda^2\omega_v^2\,\mathrm{diag}(\bar{\mathbf{u}}_{sv})^*\left(A_v(\mathbf{m}^n)\bar{\mathbf{u}}_{sv} - \mathbf{q}_{sv}\right)\right\}$$

WRI

and reduced diagonal Gauss-Newton Hessian

$$H_{sv}^n \approx \sum_{sv} \mathrm{Re}\left\{\lambda^2 \omega_v^4 \,\mathrm{diag}(\bar{\mathbf{u}}_{sv}(\mathbf{m}^n))^* \,\mathrm{diag}(\bar{\mathbf{u}}_{sv}(\mathbf{m}^n)\right\}$$

to minimize the reduced objective

$$\Phi(\mathbf{m}) = \sum_{sv} \frac{1}{2}\|P\bar{\mathbf{u}}_{sv}(\mathbf{m}) - \mathbf{d}_{sv}\|^2 + \frac{\lambda^2}{2}\|A_v(\mathbf{m})\bar{\mathbf{u}}_{sv}(\mathbf{m}) - \mathbf{q}_{sv}\|^2$$

via a scaled gradient descents [Bertsekas '99]

$$\Delta\mathbf{m} = \underset{\Delta\mathbf{m}\in\mathbb{R}^N}{\arg\min}\, \Delta\mathbf{m}^T\mathbf{g}^n + \frac{1}{2}\Delta\mathbf{m}^T H^n \Delta\mathbf{m} + c_n \Delta\mathbf{m}^T\Delta\mathbf{m}$$

$$\mathbf{m}^{n+1} = \mathbf{m}^n + \Delta\mathbf{m} \text{ with } c_n \geq 0$$

Including convex constraints

Wave-equation based inversions call for regularization, e.g. via convex constraints

$$\Delta\mathbf{m} = \underset{\Delta\mathbf{m}\in R^N}{\arg\min}\, \Delta\mathbf{m}^T\mathbf{g}^n + \frac{1}{2}\Delta\mathbf{m}^T H^n \Delta\mathbf{m} + c_n \Delta\mathbf{m}^T\Delta\mathbf{m}$$

$$\text{such that} \quad \mathbf{m}^n + \Delta\mathbf{m} \in C$$

- guarantees $\mathbf{m}^{n+1} \in C$
- more difficult to compute
- feasible if it iseasy to project onto
- naive projections $\mathbf{m}^{m+1} = \Pi_C\left(\mathbf{m}^n - (H^n)^{-1}\mathbf{g}^n\right)$ are not guaranteed to converge [Bertsekas '99]

Bound constraints
– via scaled gradient projections

For positive diagonal Gauss-Newton Hessians

$$\Delta\mathbf{m} = \arg\min_{\Delta\mathbf{m}} \Delta\mathbf{m}^T\mathbf{g}^n + \frac{1}{2}\Delta\mathbf{m}^T(H^n + c_n\mathrm{I})\Delta\mathbf{m}$$

$$\text{subject to } \mathbf{m}_i^n + \Delta\mathbf{m}_i \in [B_i^l, B_i^u],\ i = 1\cdots N$$

for which there exists a closed form solution

$$\Delta\mathbf{m}_i = \max\left(B_i^l - \mathbf{m}_i^n, \min\left(B_i^u - \mathbf{m}_i^n, -[(H^n + c_n I)^{-1}\mathbf{g}^n]_i\right)\right)$$

that is computationally affordable.

Total-variation regularization
– w/ bound constraints

Promote models w/ sharp boundaries via

$$\mathbf{m}^{n+1} = \mathbf{m}^n + \Delta\mathbf{m} \quad \text{subject to} \quad \mathbf{m}^{n+1} \in C_{\text{box}} \cap C_{\text{TV}}$$

where $C_{\text{TV}} = \{\mathbf{m} : \|\mathbf{m}\|_{\text{TV}} \leq \tau\}$ and

$$\begin{aligned}
\|\mathbf{m}\|_{TV} &= \frac{1}{h}\sum_{ij}\sqrt{(m_{i+1,j}-m_{i,j})^2 + (m_{i,j+1}-m_{i,j})^2} \\
&= \sum_{ij}\frac{1}{h}\left\|\begin{bmatrix}(m_{i,j+1}-m_{i,j})\\(m_{i+1,j}-m_{i,j})\end{bmatrix}\right\| \\
&= \|D\mathbf{m}\|_{1,2} := \sum_{l=1}^{N}\|(D\mathbf{m})_l\| \ .
\end{aligned}$$

Proposed algorithm

Solve

$$\underset{\mathbf{m}}{\text{minimize}}\, \Phi(\mathbf{m}) \quad \text{subject to} \quad \mathbf{m}^{n+1} \in C_{\text{box}} \cap C_{\text{TV}}$$

by iterating

$$\Delta\mathbf{m} = \arg\min_{\Delta\mathbf{m}} \Delta\mathbf{m}^T\mathbf{g}^n + \frac{1}{2}\Delta\mathbf{m}^T(H^n + c_n\mathrm{I})\Delta\mathbf{m}$$

$$\text{subject to} \quad \mathbf{m}_i^n + \Delta\mathbf{m}_i \in [B_i^l, B_i^u] \text{ and } \|\mathbf{m}^n\Delta\mathbf{m}\|_{TV} \leq \tau$$

$$\mathbf{m}^{n+1} = \mathbf{m}^n + \Delta\mathbf{m}$$

Solving the convex subproblems

Find saddle point of

$$\mathcal{L}(\Delta\mathbf{m},\mathbf{p}) = \Delta\mathbf{m}^T\mathbf{g}^n + \frac{1}{2}\Delta\mathbf{m}^T(H^n + c_n\mathrm{I})\Delta\mathbf{m} + g_B(\mathbf{m}^n + \Delta\mathbf{m}) + \mathbf{p}^T D(\mathbf{m}^n + \Delta\mathbf{m}) - \tau\|\mathbf{p}\|_{\infty,2}$$

with

bound constraint

$$g_B(\mathbf{m}) = \begin{cases} 0 & \text{if } m_i \in [B_i^l, B_i^u] \\ \infty & \text{otherwise} \end{cases}$$

TV-norm constraint

$$\sup_{\mathbf{p}} +\mathbf{p}^T D(\mathbf{m}^n + \Delta\mathbf{m}) - \tau\|\mathbf{p}\|_{\infty,2} = \begin{cases} 0 & \text{if } \|D(\mathbf{m}^n + \Delta\mathbf{m})\|_{1,2} \leq \tau \\ \infty & \text{otherwise} \end{cases}$$

Ernie Esser, Xiaoqun Zhang, and Tony F. Chan. A General Frame- work for a Class of First Order Primal-Dual Algorithms for Convex Optimization in Imaging Science. SIAM Journal on Imaging Sciences, 3(4):1015–1046, 2010.

Iterations
– primal dual hybrid gradient (PDHG)

projection onto TV ball

$$\mathbf{p}^{k+1} = \mathbf{p}^k + \delta D(\mathbf{m}^n + \Delta\mathbf{m}^k) - \Pi_{\|\cdot\|_{1,2}\leq\tau\delta}(\mathbf{p}^k + \delta D(\mathbf{m}^n + \Delta\mathbf{m}^k))$$

$$\Delta\mathbf{m}_i^{k+1} = \max\left((B_i^l - \mathbf{m}_i^n), B_i\right)$$

$$B_i = \min\left((B_i^u - \mathbf{m}_i^n), [(H^n + (c_n + \frac{1}{\alpha})\mathrm{I})^{-1}(-\mathbf{g}^n + \frac{\Delta\mathbf{m}^k}{\alpha} - D^T(2\mathbf{p}^{k+1} - \mathbf{p}^k)]_i\right)$$

for steplengths $\alpha\delta \leq \frac{1}{\|D^T D\|}$ and $\alpha = \frac{1}{\max(H^n + c_n \mathrm{I})}$

- do not involve solutions of (data-augmented) wave equations
- allows for data-dependent stepsizes

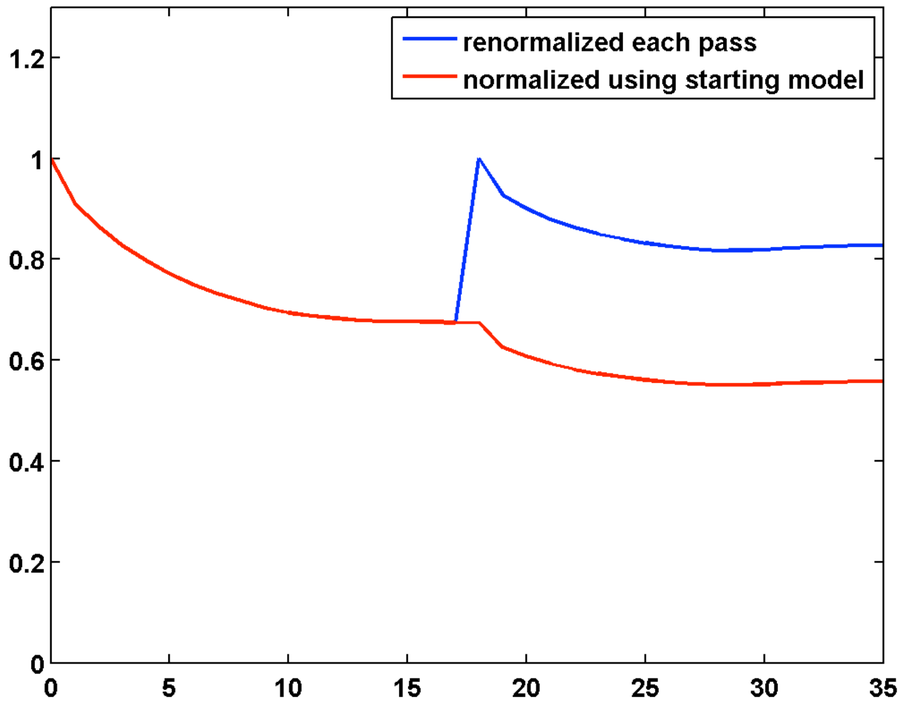
Solution strategy

Do multiple frequency sweeps

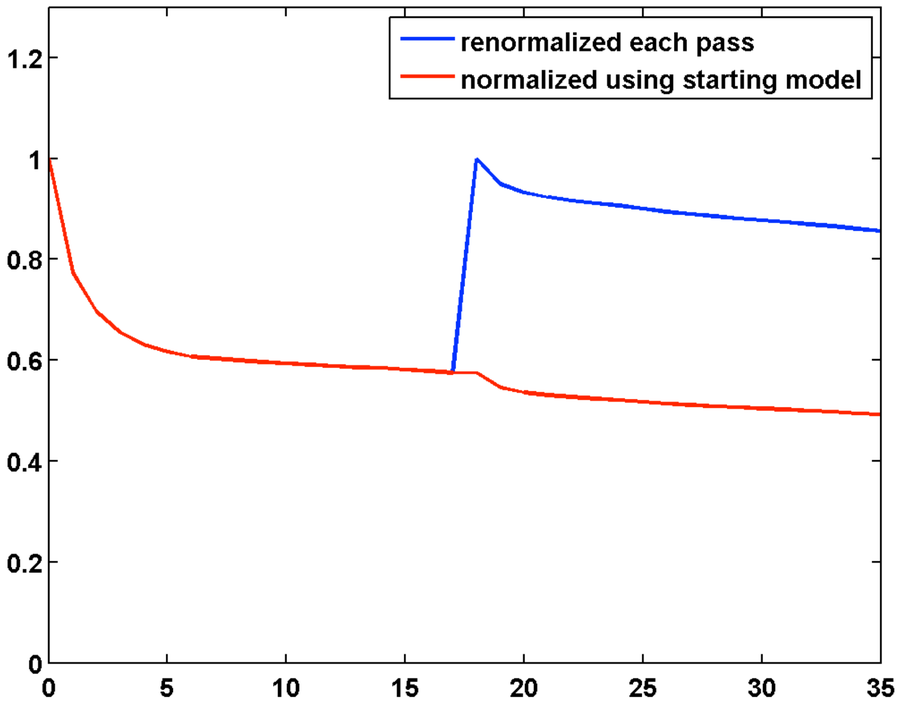
- w/ warm starts
- while relaxing the TV constraint

Relative model errors

w/o TV



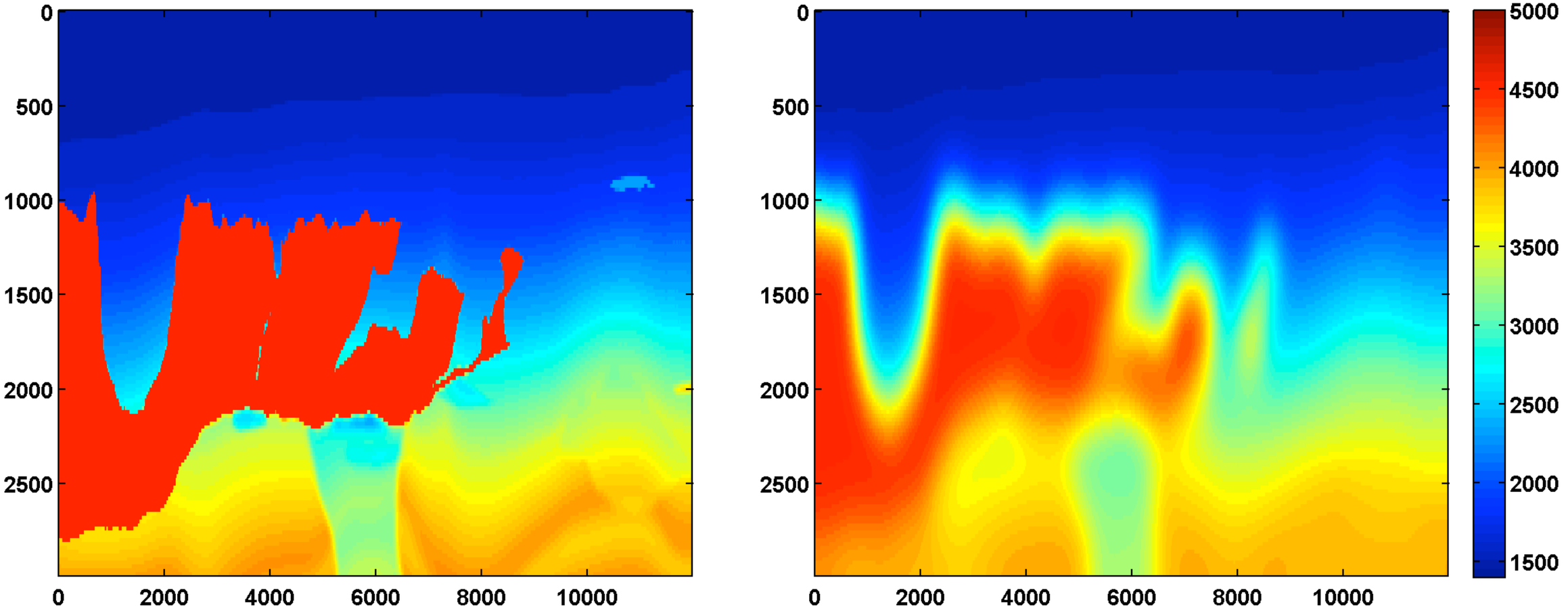
w/ TV



BP model

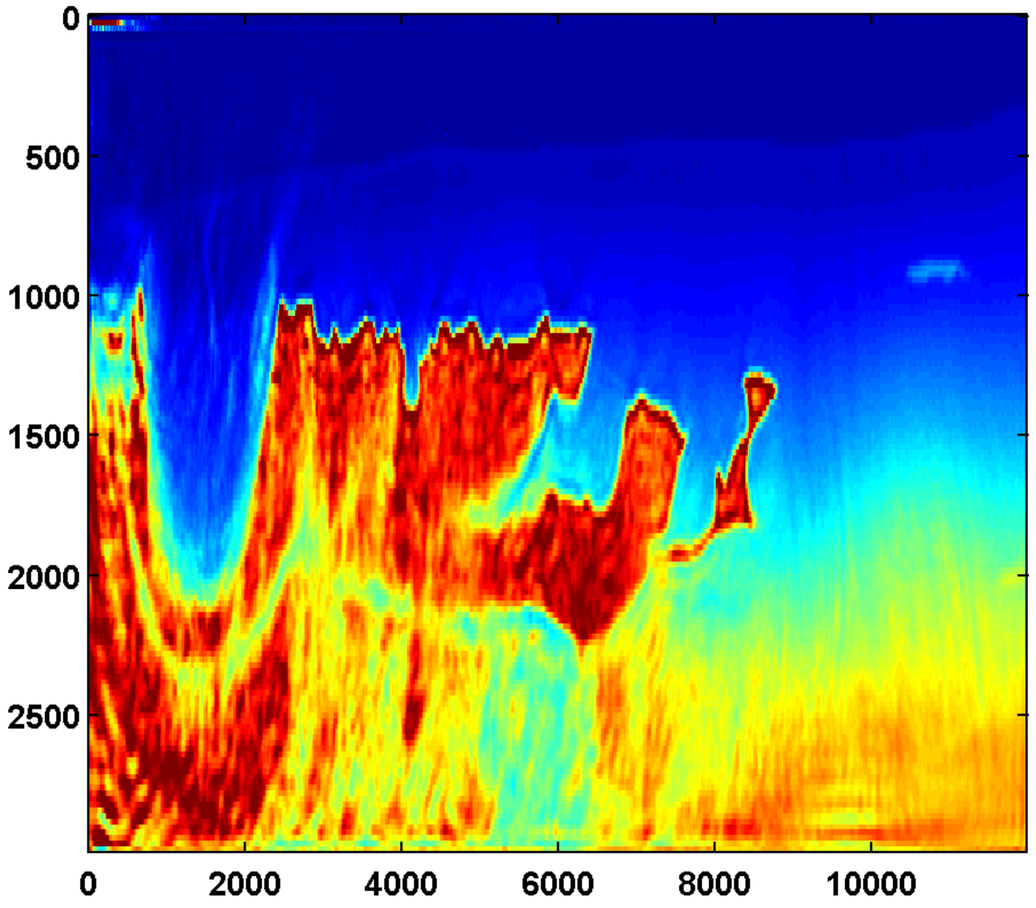
- number of sources: 126
- number of receivers: 299
- frequency continuation over 3-20Hz in overlapping batches of 2
- maximum number of outer iterations per frequency batch: 25
- maximum number of inner iterations for convex subproblems: 2000
- known Ricker wavelet sources with 15Hz peak frequency
- two simultaneous shots with Gaussian weights w/ redraws
- no added noise

True velocity & good starting model

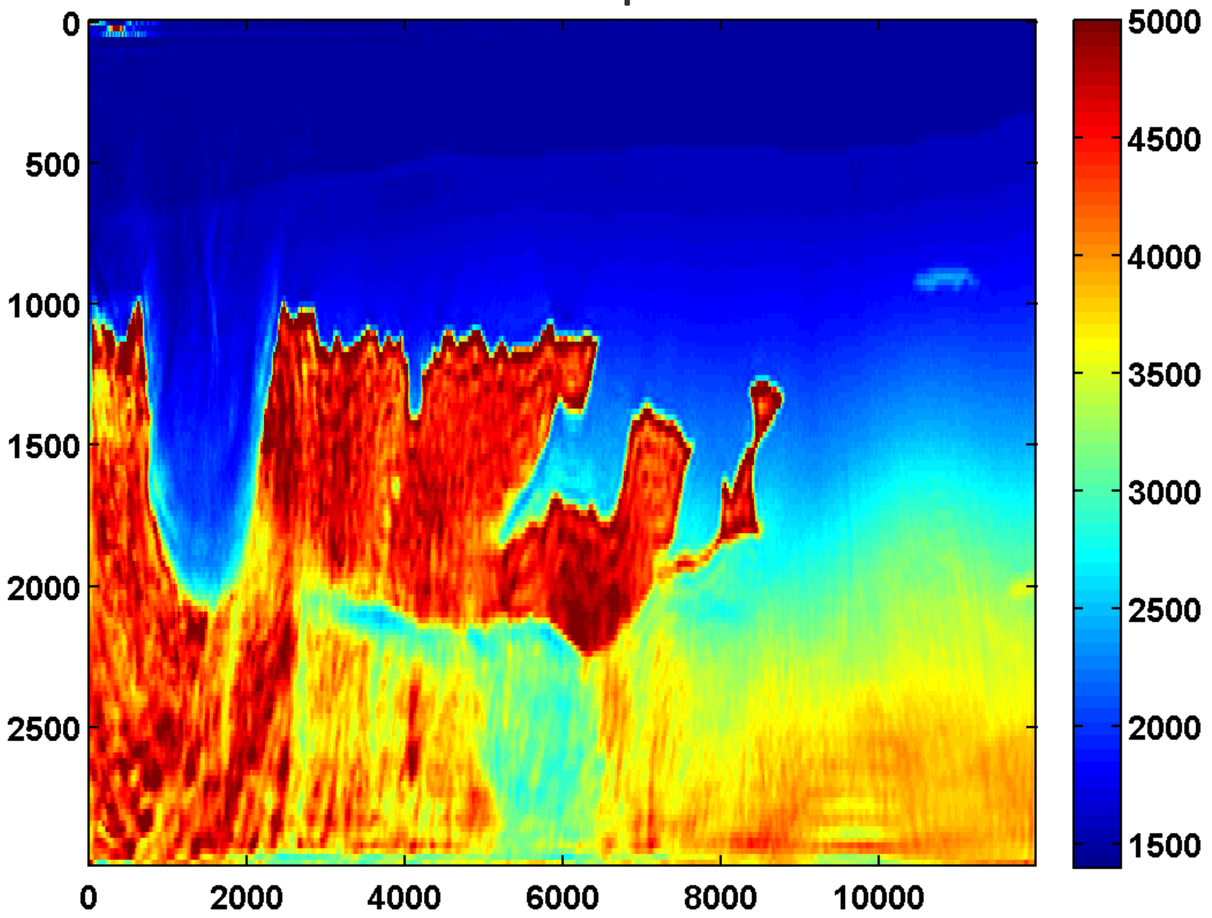


Results w/o TV

After one cycle through the frequencies

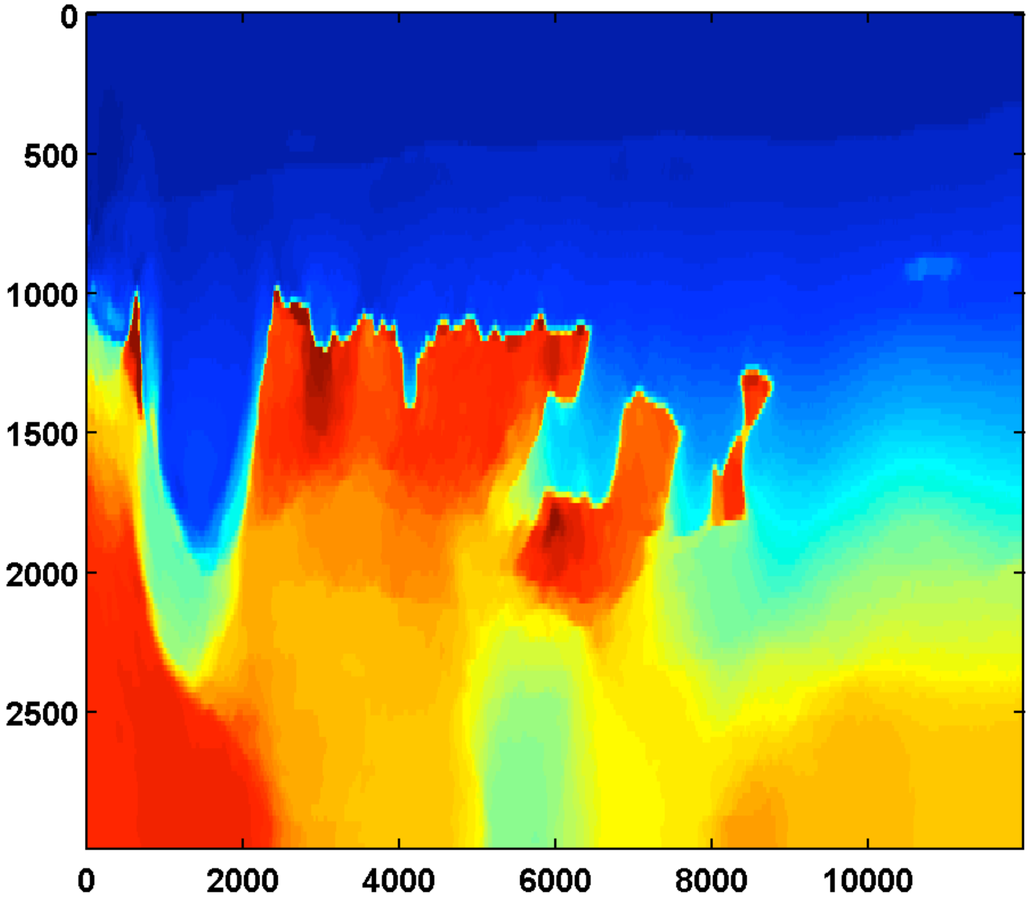


After two cycles through the frequencies

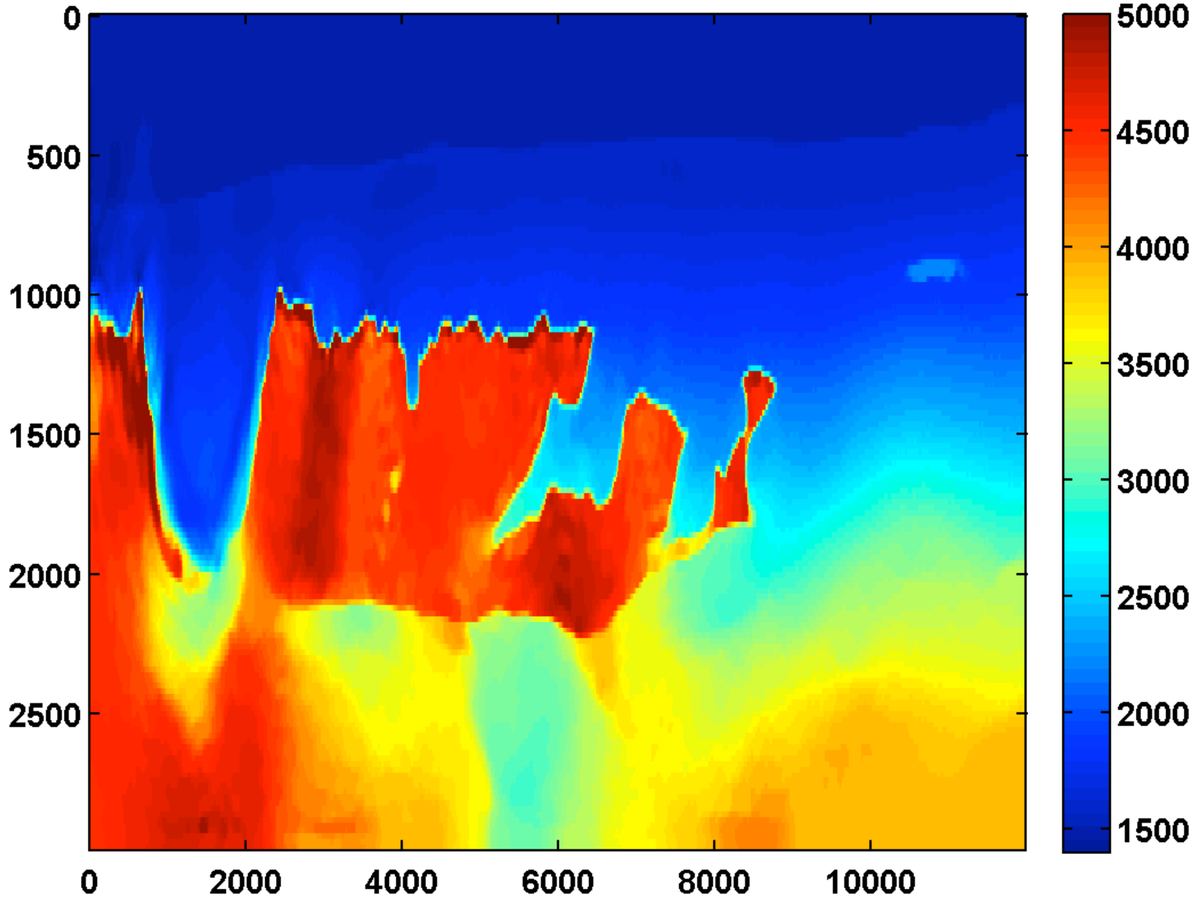


Results w/ TV

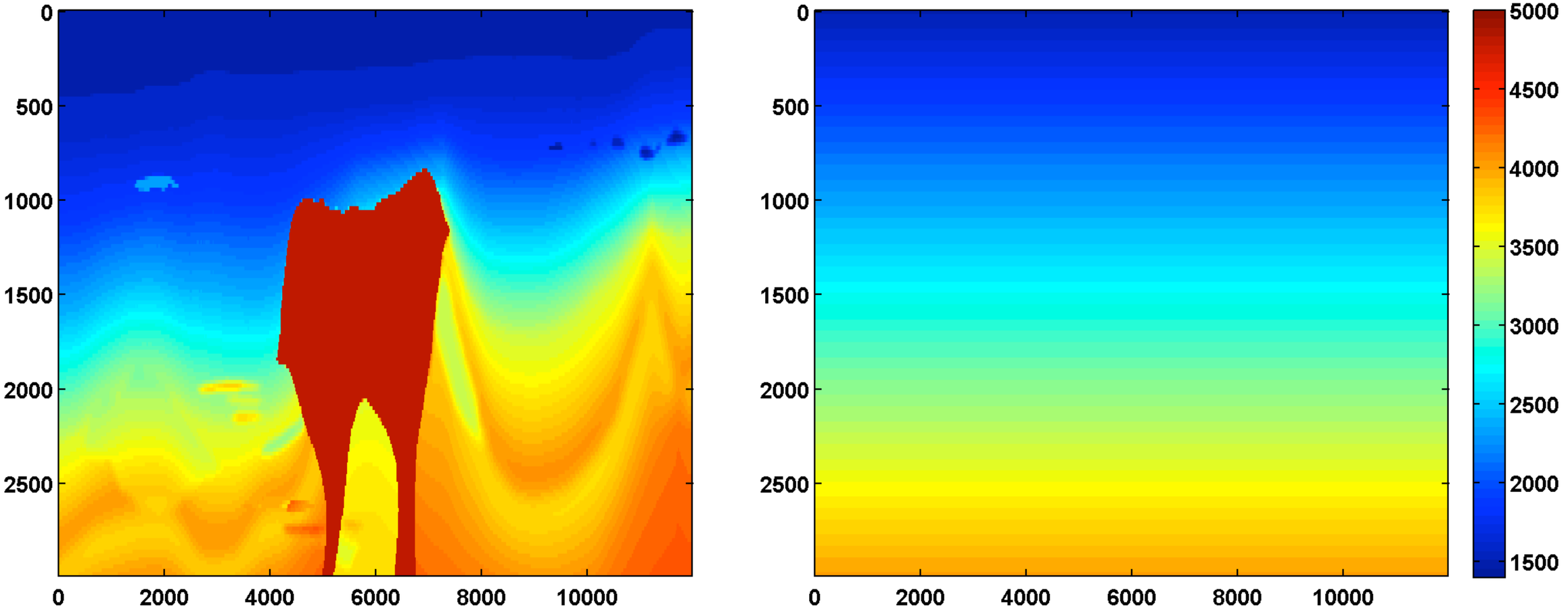
After one cycle through the frequencies



After two cycles through the frequencies

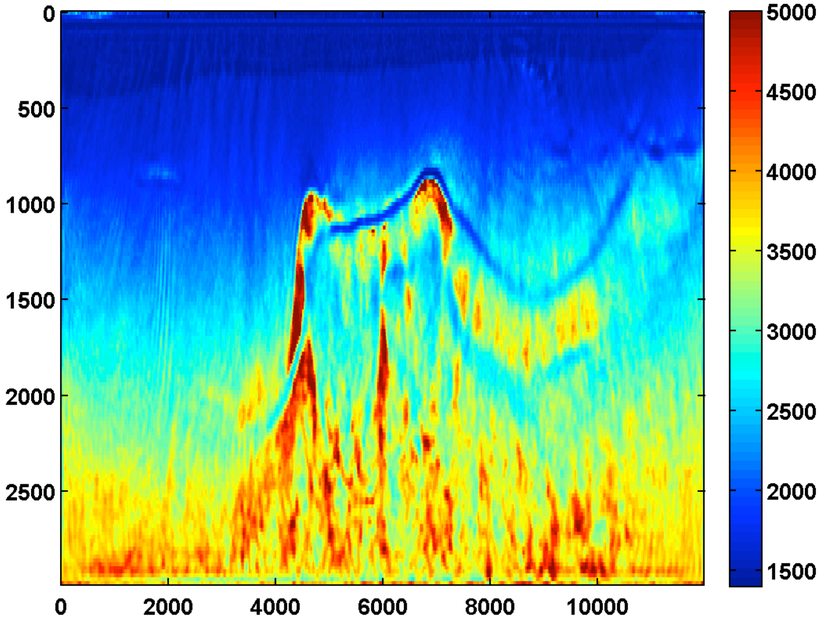


True velocity & poor starting model

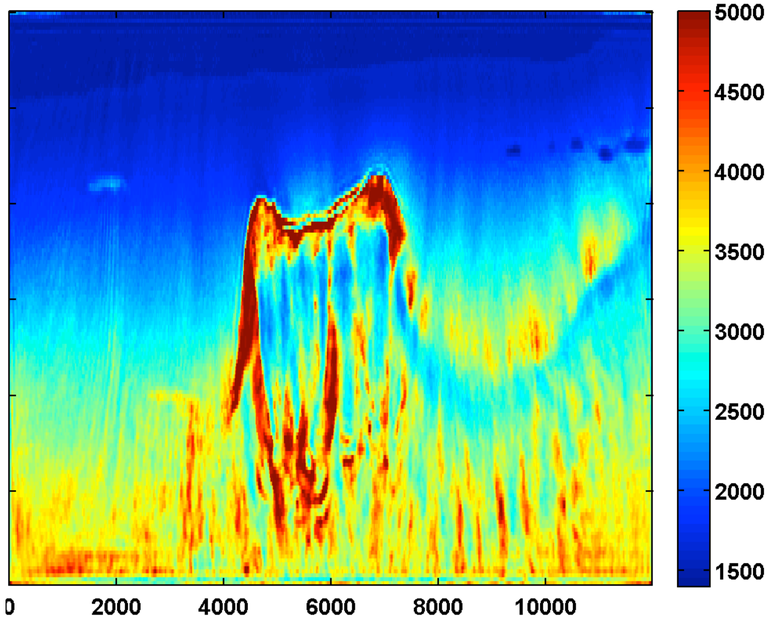


Results w/o TV

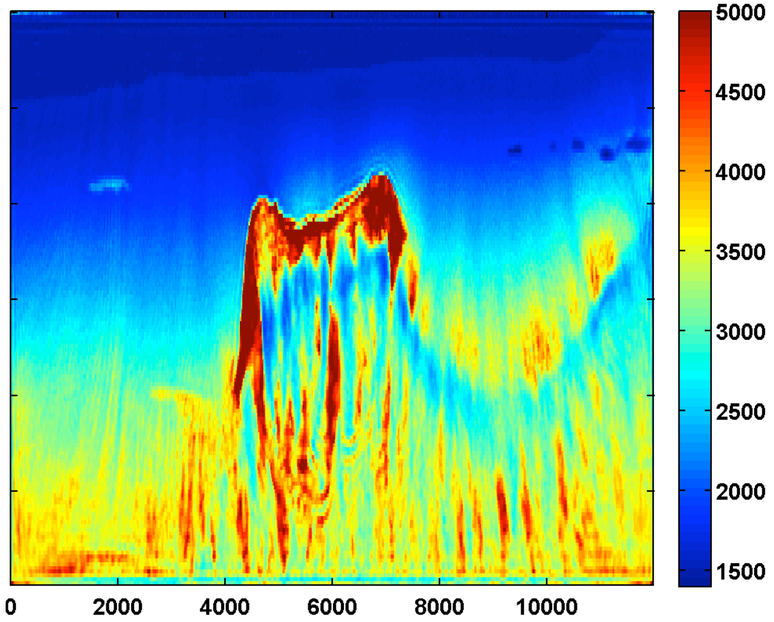
after one cycle through the frequencies



after two cycles through the frequencies

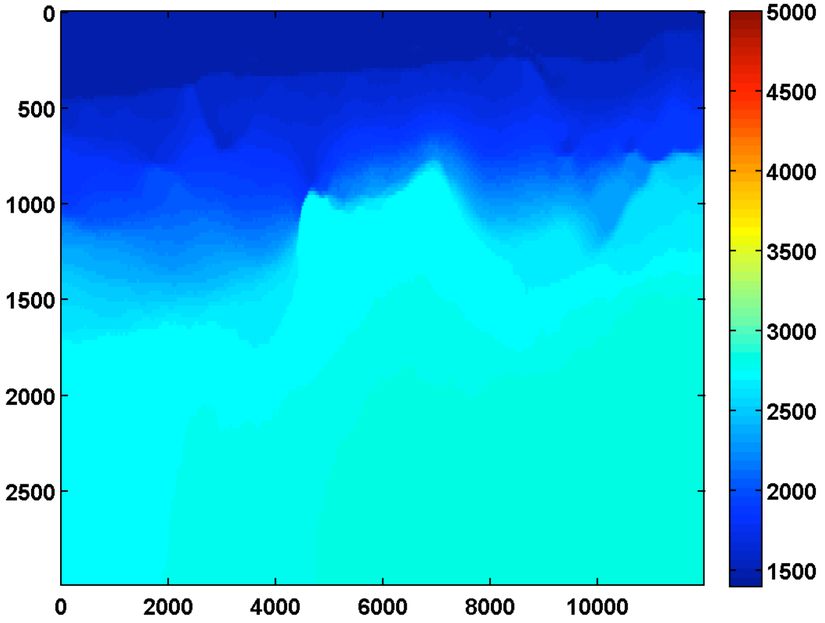


after three cycles through the frequencies

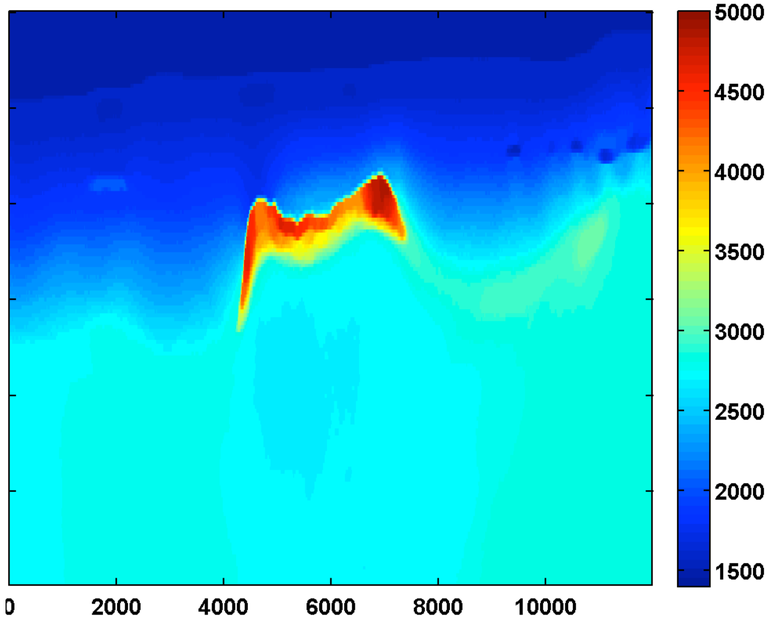


Results w/ TV

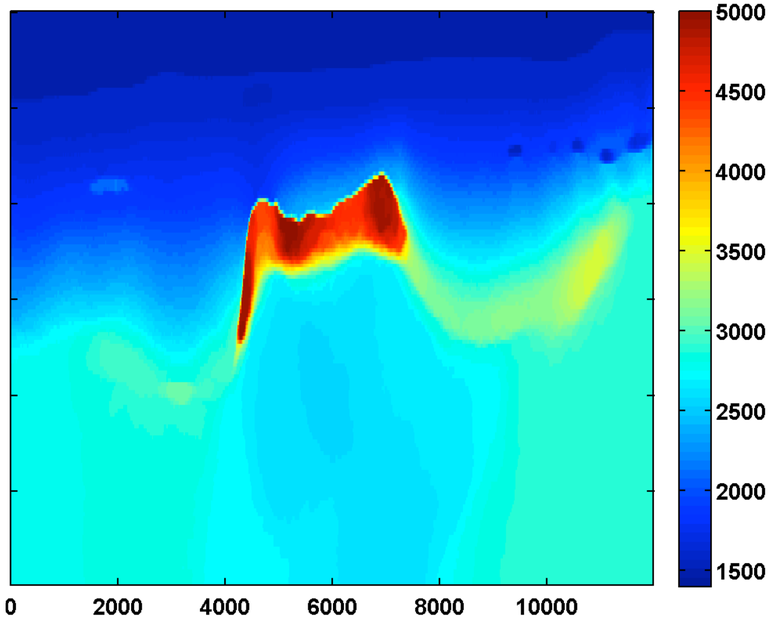
after one cycle through the frequencies



after two cycles through the frequencies



after three cycles through the frequencies



Observations

Convex constraints

- remove randomized source sampling artifacts
- sharpen top and bottom of high-contrast high-velocity unconformities for good starting models
- sharpen top only for poor starting models

Included in software release “Total Variation Regularized Wavefield Reconstruction Inversion”

So convex constraints may not be enough…

Joint work w/ Sub Salt Solutions Ltd. outside of SINBAD
Patent filed GB1509337.0

Ernie Esser, Lluis Guash*, and Felix J. Herrmann

*Sub Salt Solutions Ltd

SLIM
University of British Columbia

Hinge loss
– one-sided TV constraint

Mitigate erroneous velocity model updates by using the fact that

- vertical slowness profiles tend to decrease w/ depth
- makes it less probable that velocities jump down along the vertical

Mathematically expressed as the one-norm of the hinge function

$$\| \max(0, D_z \mathbf{m}) \|_1 \leq \xi$$

- for ξ small slowness is unlikely to step up
- extended to a weighted directional gradient
- combined w/ omni-directional TV and bound constraints

Joint work w/ Subsalt Solutions Ltd. Patent filed GB1509337.0

Scaled-gradient projections
– w/ convex total-variation, box, and hinge-loss constraints

Solve for given $\bar{\mathbf{u}}_\lambda$

$$\min_{\mathbf{m}} \phi(\mathbf{m}, \bar{\mathbf{u}}_\lambda) \quad \text{subject to} \quad \begin{cases} m_i \in [B_1, B_2] \\ \|\mathbf{m}\|_{\text{TV}} \leq \tau \\ \|\mathbf{m}\|_{\text{Hinge}} \leq \xi \end{cases}$$

with

$$\|\mathbf{m}\|_{TV} = \sum_{ij} \frac{1}{h} \left\| \begin{bmatrix} (m_{i,j+1} - m_{i,j}) \\ (m_{i+1,j} - m_{i,j}) \end{bmatrix} \right\|$$

and

$$\|\mathbf{m}\|_{\text{Hinge}} = \| \max(0, D_z \mathbf{m}) \|_1$$

Proposed algorithm

Solve

$$\underset{\mathbf{m}}{\text{minimize}}\, \Phi(\mathbf{m}) \quad \text{subject to} \quad \mathbf{m}^{n+1} \in C_{\text{box}} \cap C_{\text{TV}} \cap C_{\text{Hinge}}$$

by iterating

$$\mathbf{p}_1^{k+1} = \mathbf{p}_1^k + \delta D(\mathbf{m}^n + \Delta\mathbf{m}^k) - \Pi_{\|\cdot\|_{1,2}\leq\tau\delta}(\mathbf{p}_1^k + \delta D(\mathbf{m}^n + \Delta\mathbf{m}^k))$$

$$\mathbf{p}_2^{k+1} = \mathbf{p}_2^k + \delta D_z(\mathbf{m}^n + \Delta\mathbf{m}^k) - \Pi_{\|\max(0,\cdot)\|_1\leq\xi\delta}(\mathbf{p}_2^k + \delta D_z(\mathbf{m}^n + \Delta\mathbf{m}^k))$$

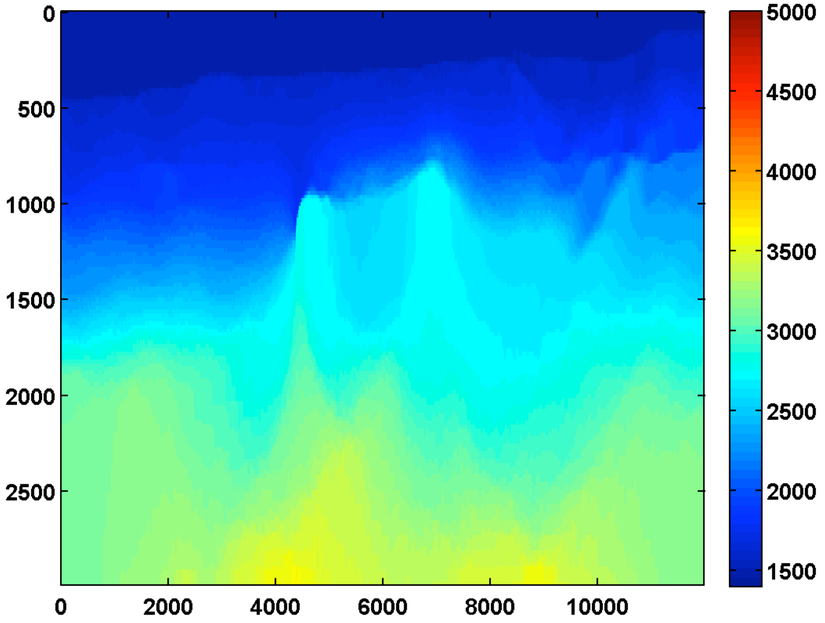
$$B_i = \min\left((B_i^u - \mathbf{m}_i^n), [(H^n + (c_n + \frac{1}{\alpha})\mathrm{I})^{-1}(-\mathbf{g}^n + \frac{\Delta\mathbf{m}^k}{\alpha} - D^T(2\mathbf{p}_1^{k+1} - \mathbf{p}_1^k) - D_z^T(2\mathbf{p}_2^{k+1} - \mathbf{p}_2^k))]_i\right)$$

$$\Delta\mathbf{m}_i^{k+1} = \max\left((B_i^l - \mathbf{m}_i^n), B_i\right)$$

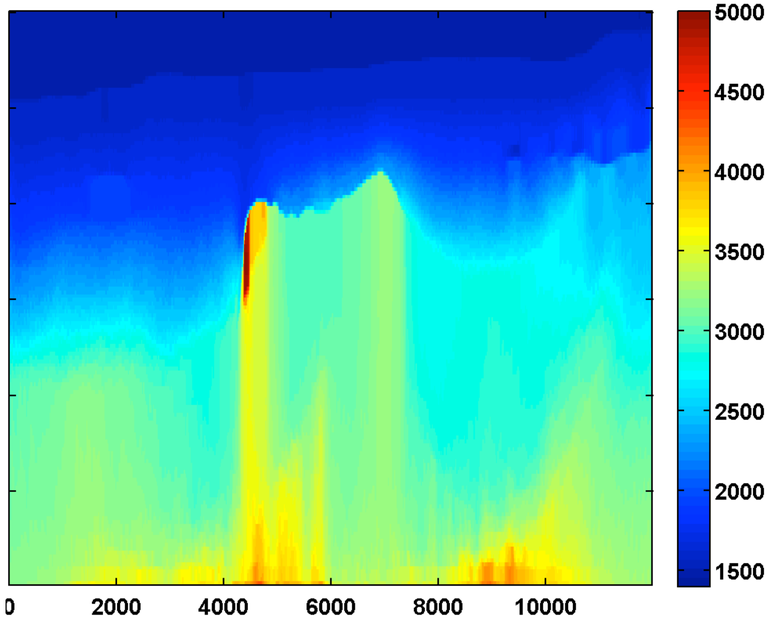
Results w/ hinge loss continuation

$$\frac{\xi}{\xi_{\text{true}}} = \{.01, .05, .10\}$$

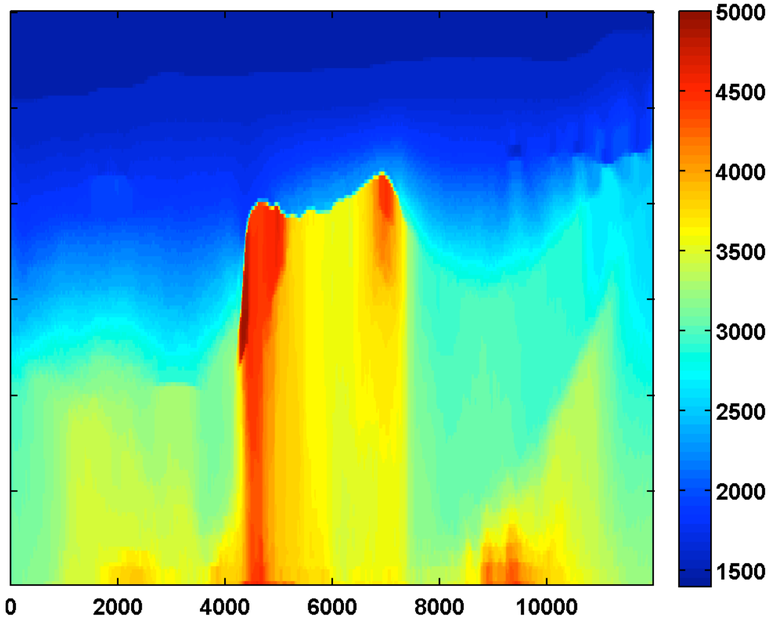
after one cycle through the frequencies



after two cycles through the frequencies



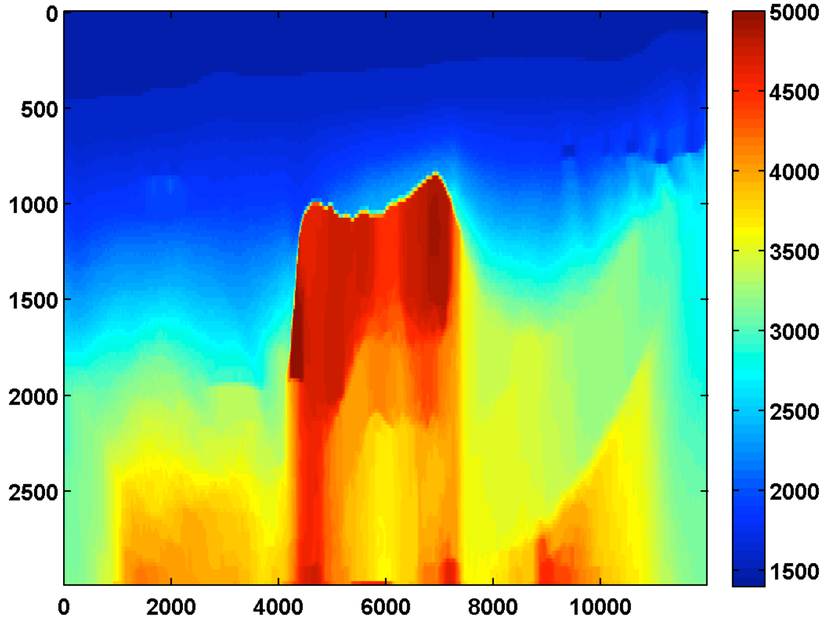
after three cycles through the frequencies



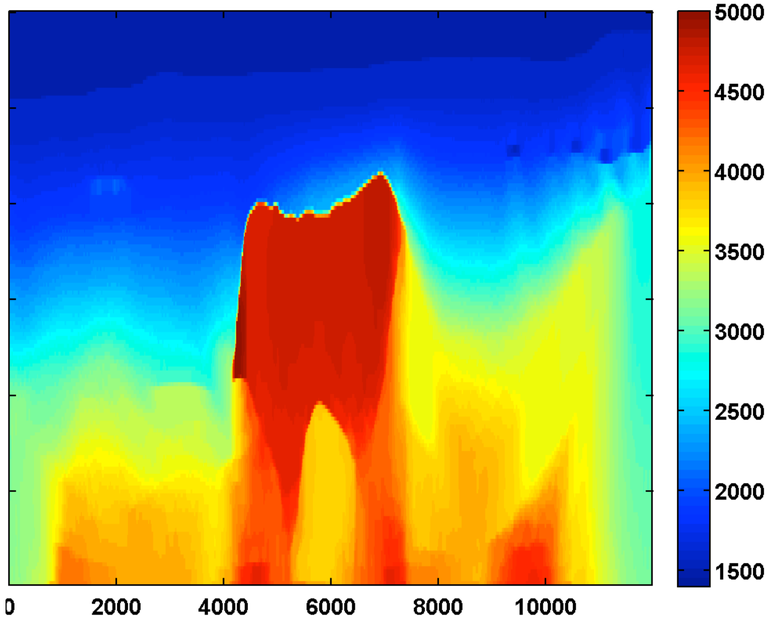
Results w/ hinge loss continuation

$$\frac{\xi}{\xi_{\text{true}}} = \{.15, .20, .25\}$$

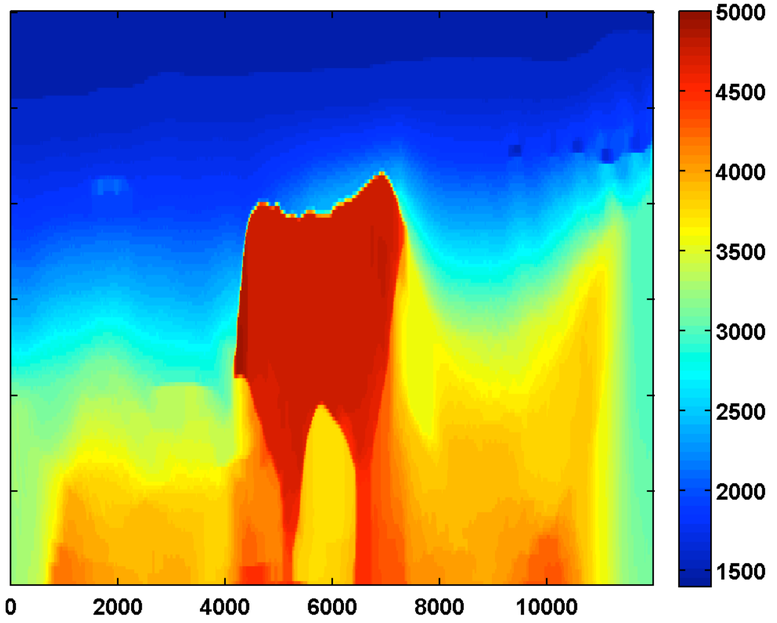
after four cycles through the frequencies



after five cycles through the frequencies

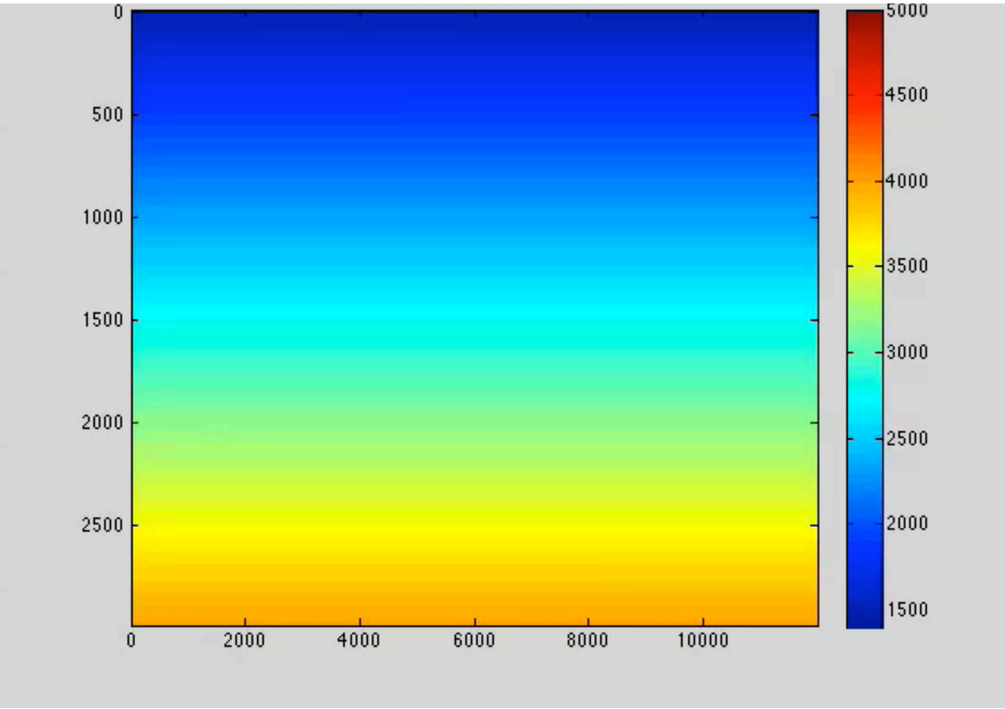


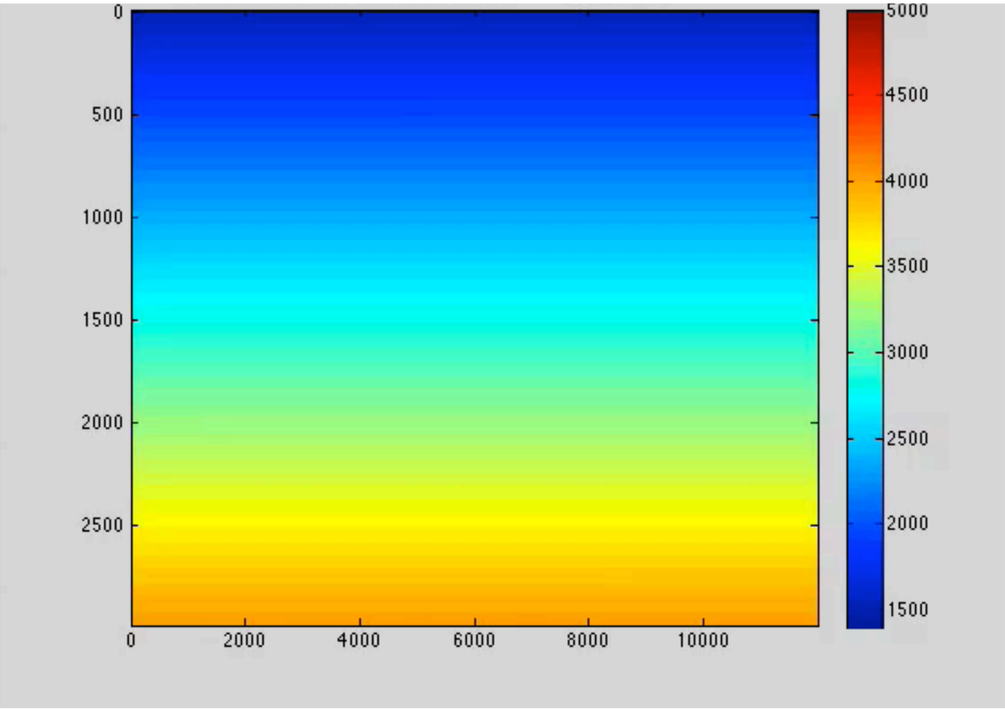
after six cycles through the frequencies



WRI

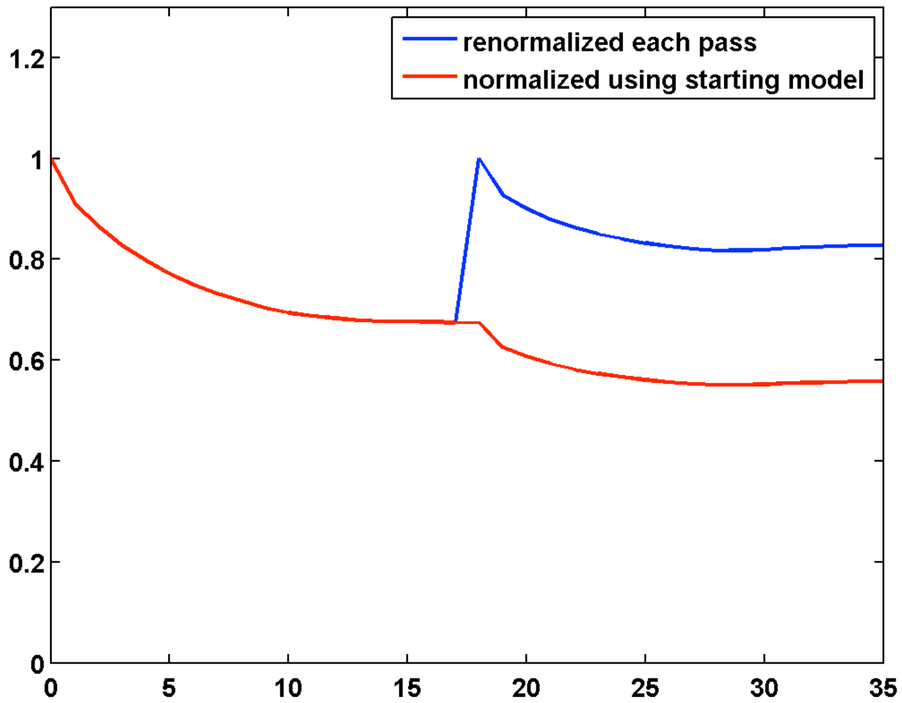
w/ or w/o TV-norm projections & poor starting model



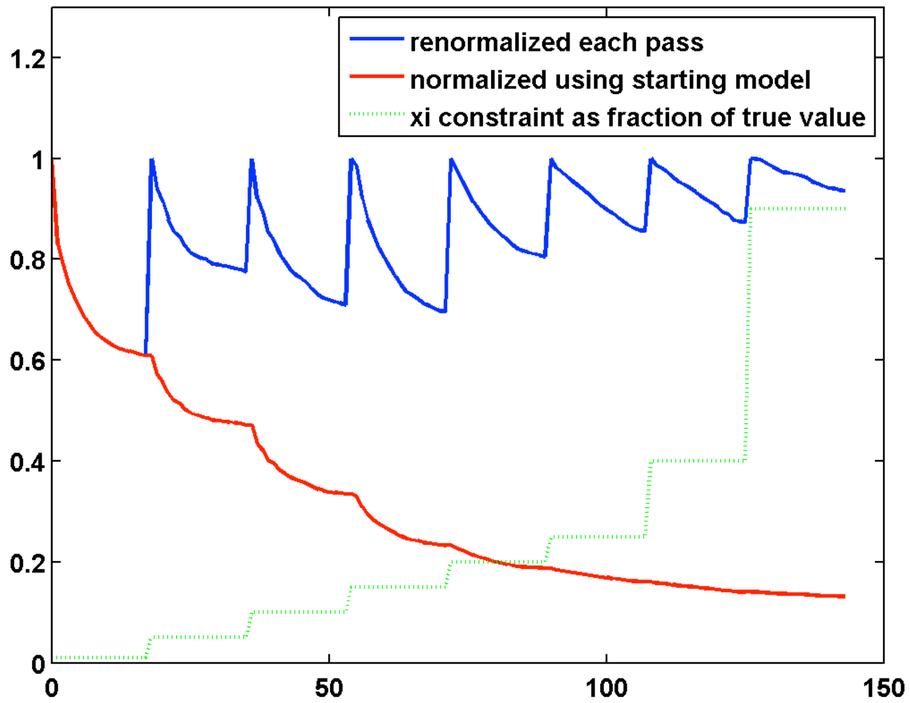


Relative model errors

w/o TV

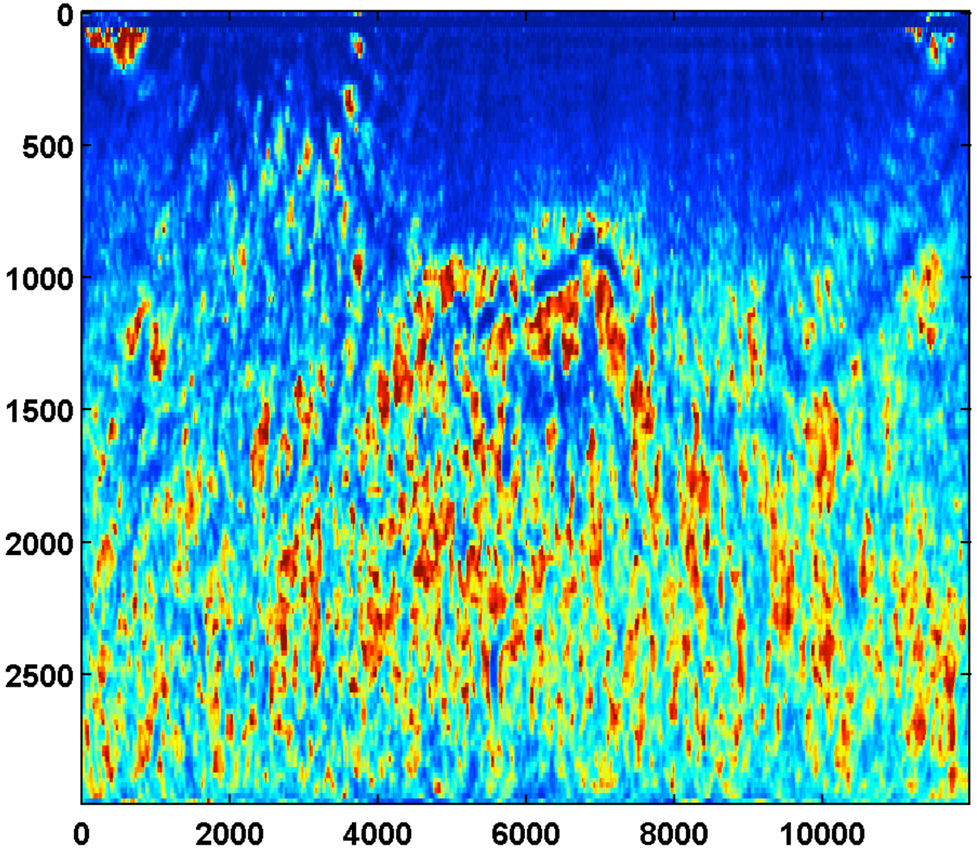


w/ TV & hinge

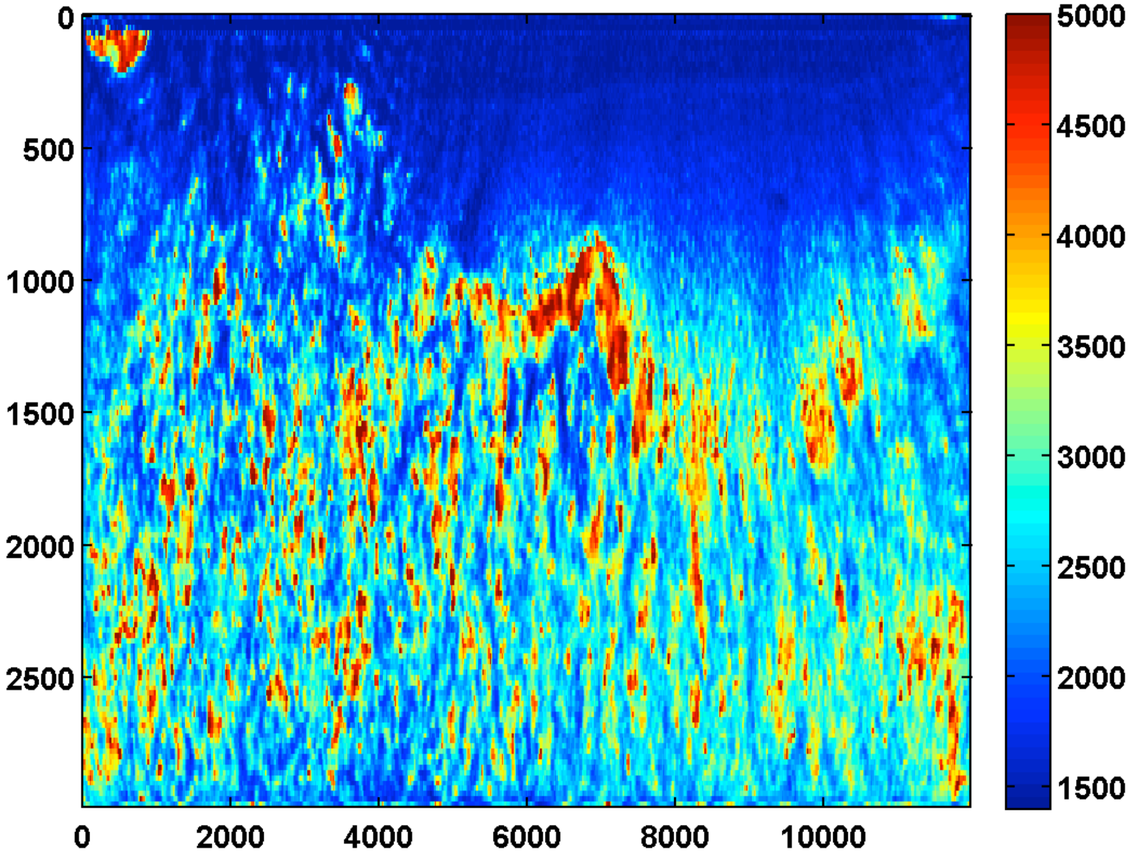


Adjoint-state w/o TV

After one cycle through the frequencies



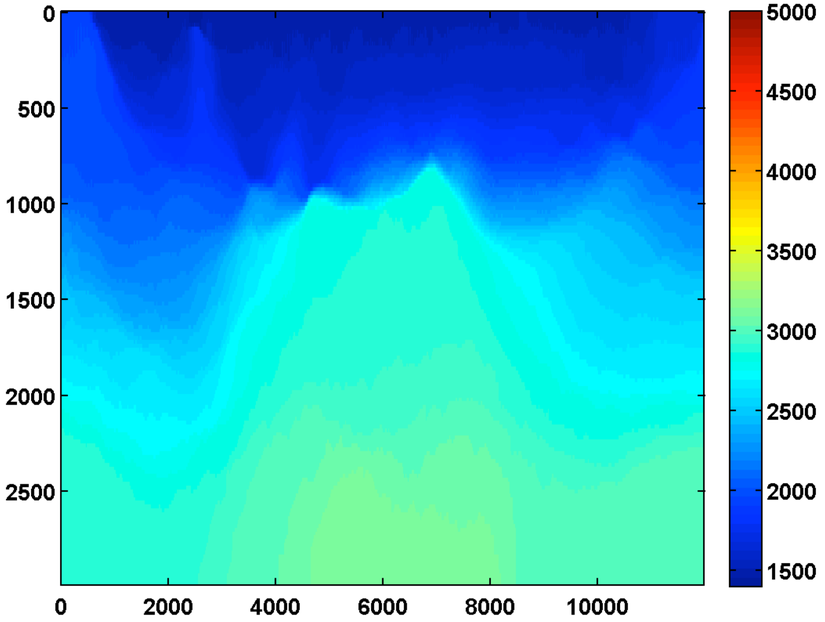
After two cycles through the frequencies



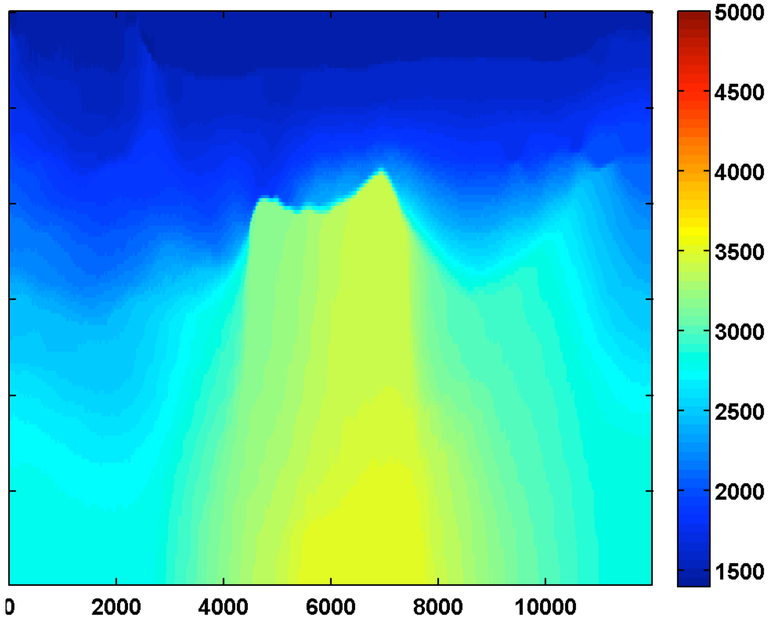
Adjoint-state w/ hinge loss continuation

$$\frac{\xi}{\xi_{\text{true}}} = \{.01, .05, .10\}$$

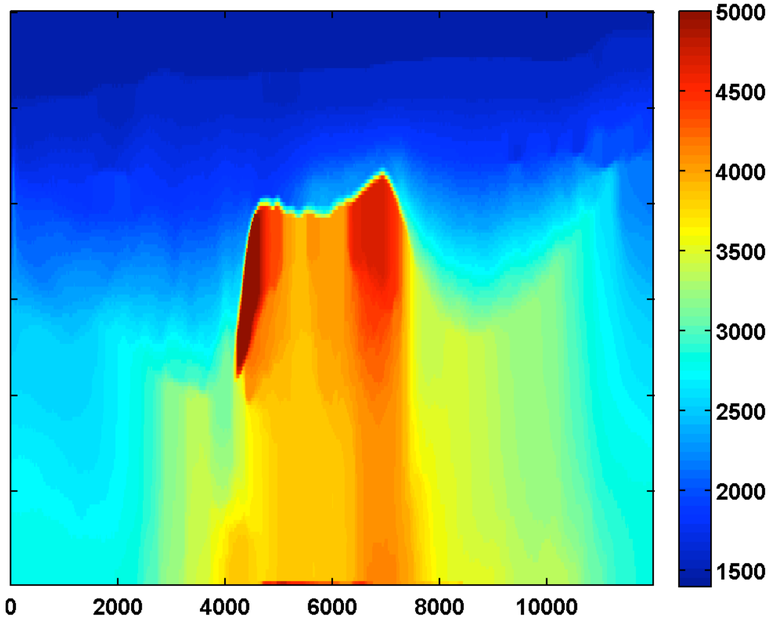
after one cycle through the frequencies



after two cycles through the frequencies



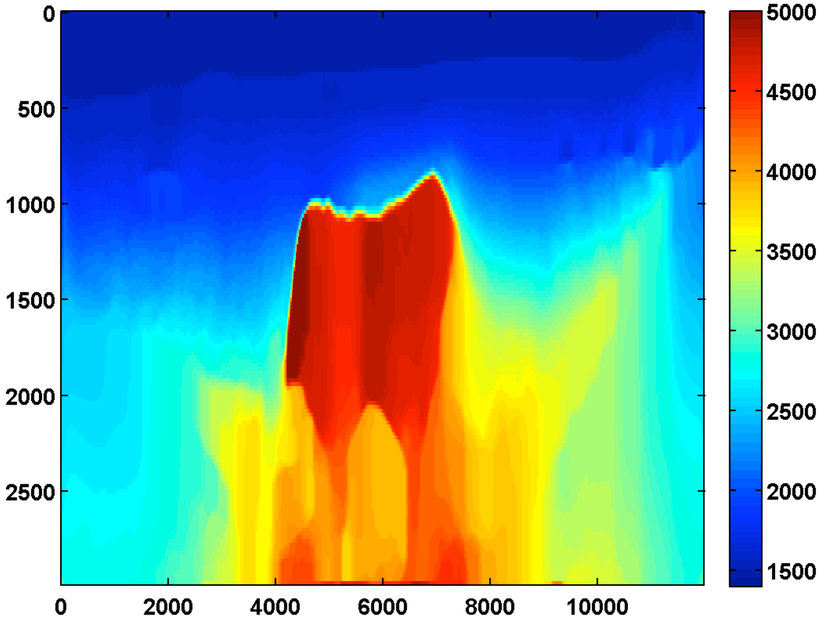
after three cycles through the frequencies



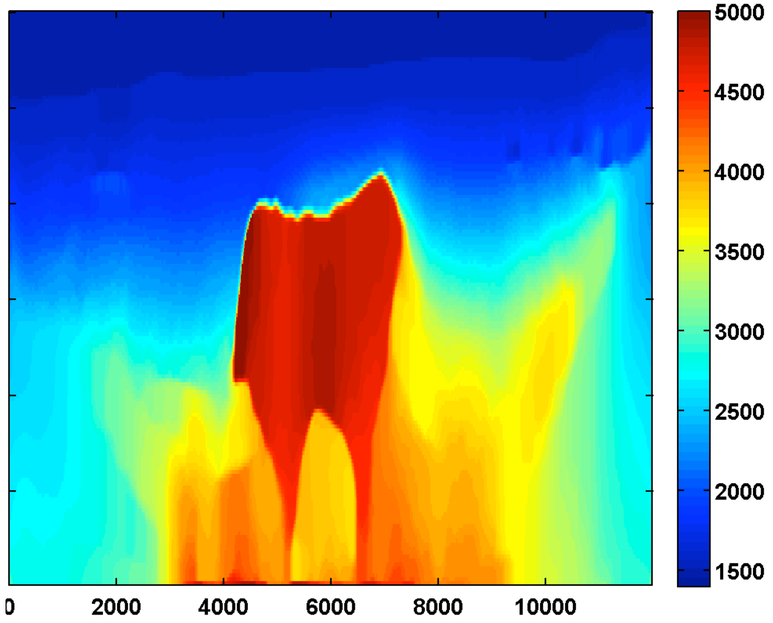
Adjoint-state w/ hinge loss continuation

$$\frac{\xi}{\xi_{\text{true}}} = \{.15, .20, .25\}$$

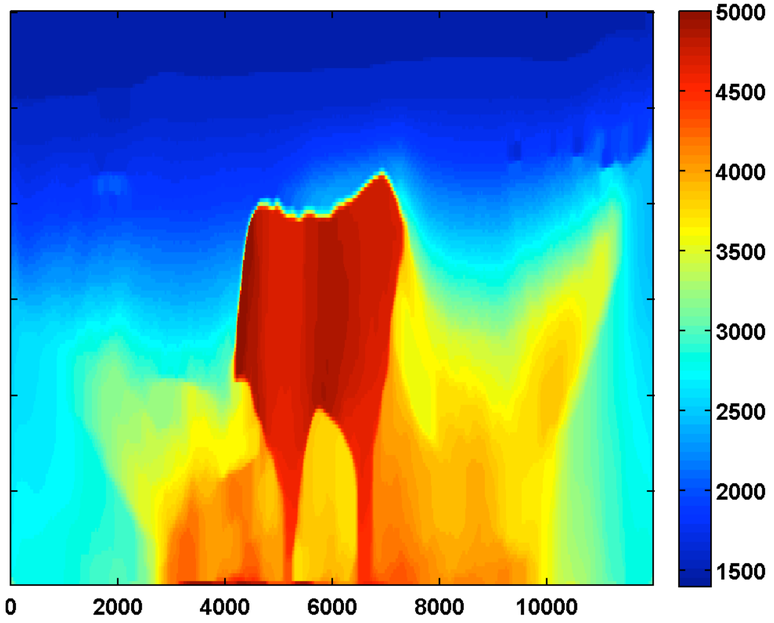
after four cycles through the frequencies



after five cycles through the frequencies

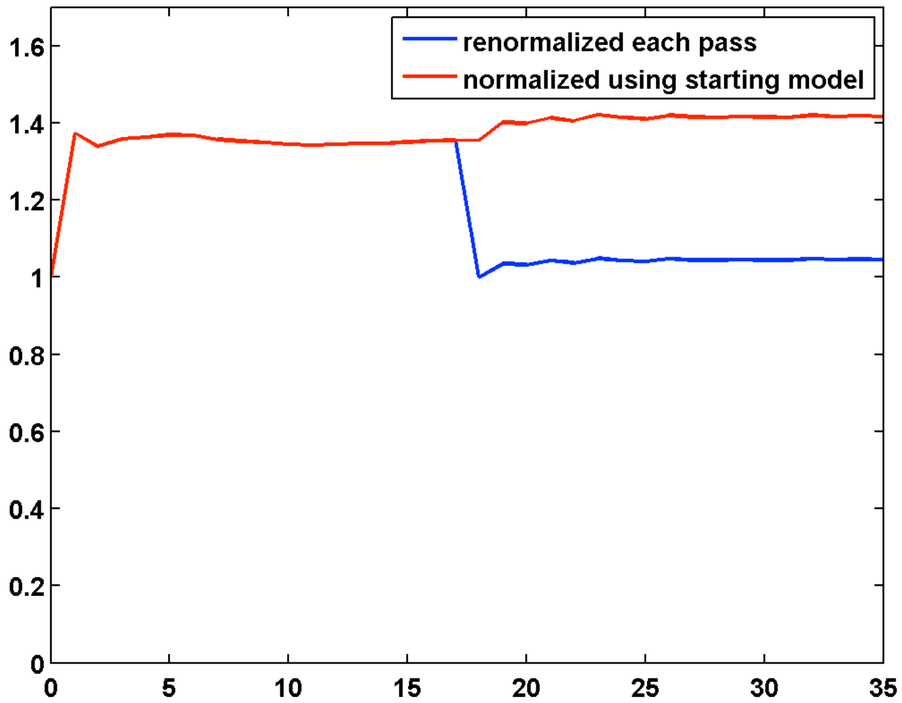


after six cycles through the frequencies

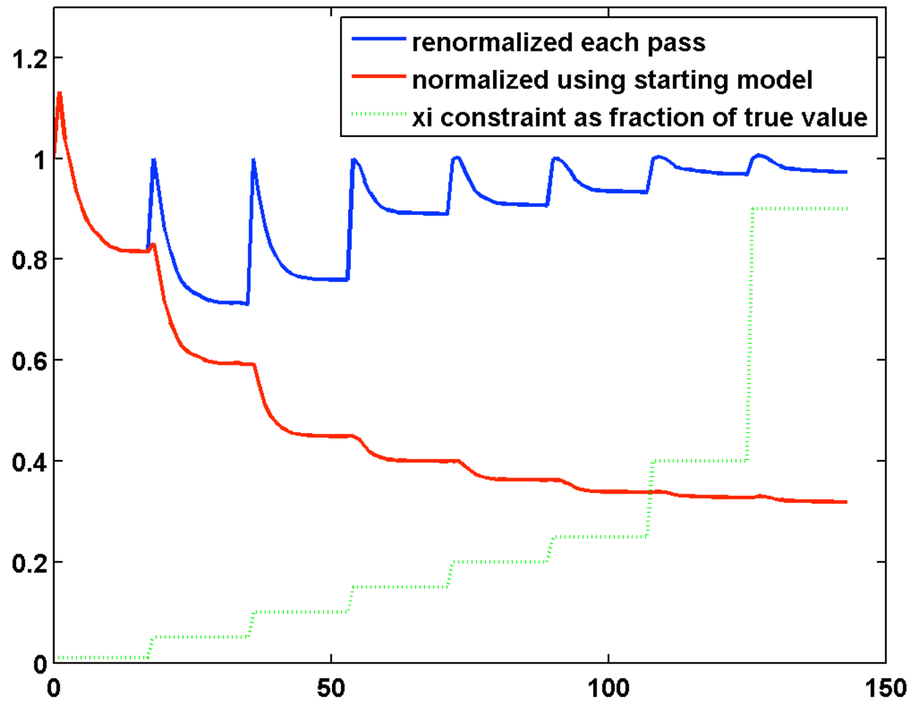


Relative model errors

w/o TV

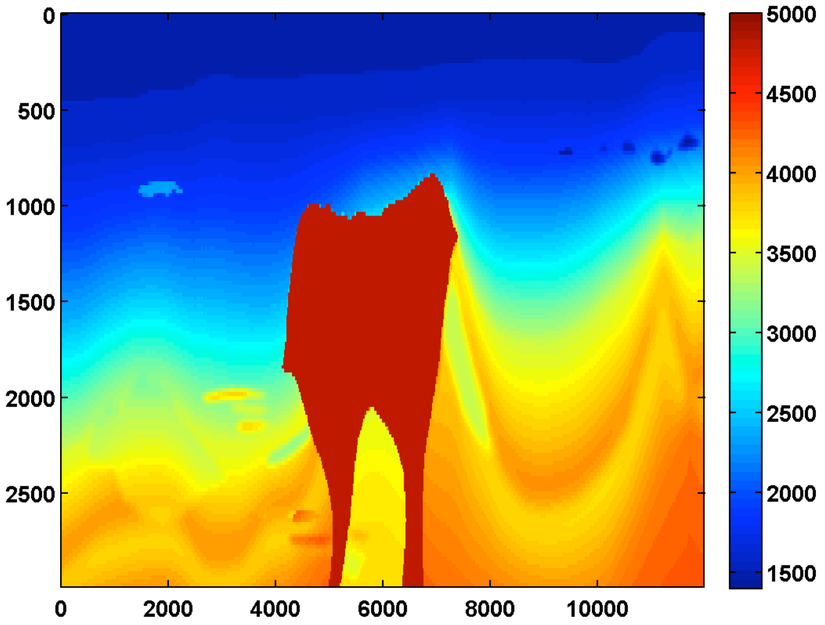


w/ TV & hinge

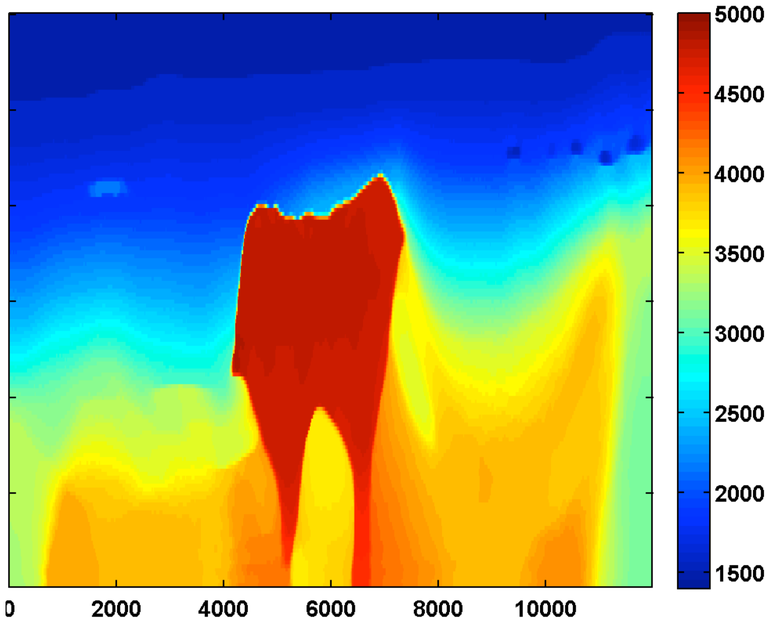


WRI vs adjoint-state

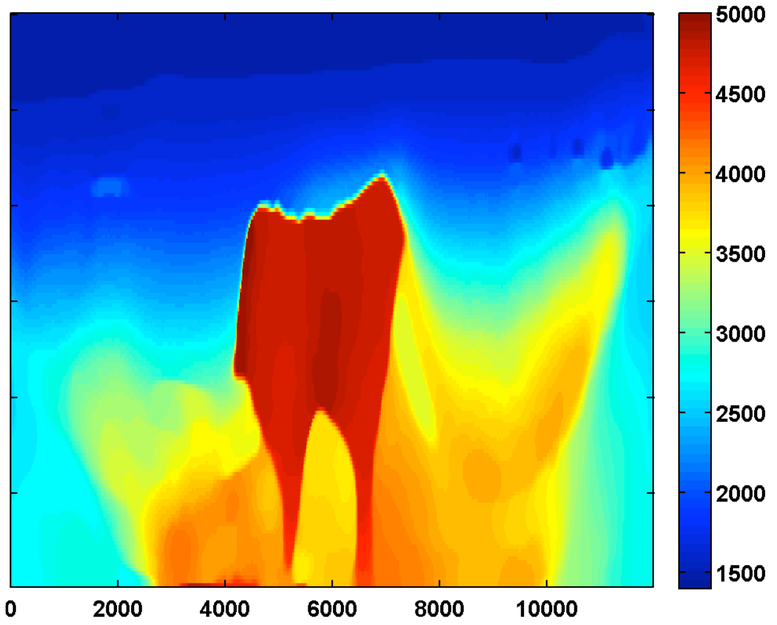
initial model



WRI



adjoint-state



Conclusions

New method for regularizing wave-equation based inversion benefits from

- combining of convex constraints
- multiple frequency sweeps w/ warm starts & relaxing of the constraints
- hinge-loss function plays a critical role
- works for both WRI & adjoint-state FWI

Development of automatic continuation strategies for relaxing the constraints is ongoing.

Candidate for “automatic” salt flooding…