# Yin Huang's Thesis, and Computing Gradients

William Symes

RICE

# Yin Huang

PhD student in TRIP: 2010.08 - 2016.02

Thesis: *Born Waveform Inversion in Shot Coordinate Domain*

Currently: Amazon, Seattle

# Chapter 2

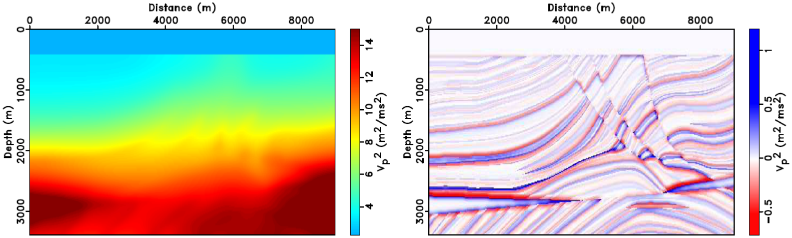
Born waveform inversion via shot record extension, variable projection, differential semblance

[SEG 2015]

RICE

# Chapter 2

## Task: estimate Marmousi from homog. initial guess




(a) (b)

**Figure 2.2:** Marmousi example: Target background model (a) target reflectivity model (b)

# Chapter 2

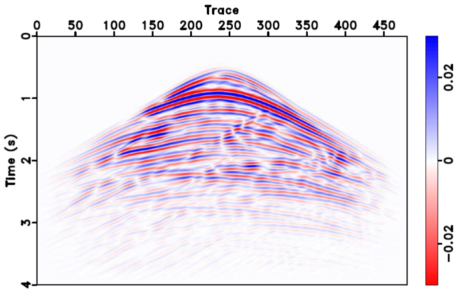



**Figure 2.3:** Marmousi example: Born shot record with index 41.

# Chapter 2

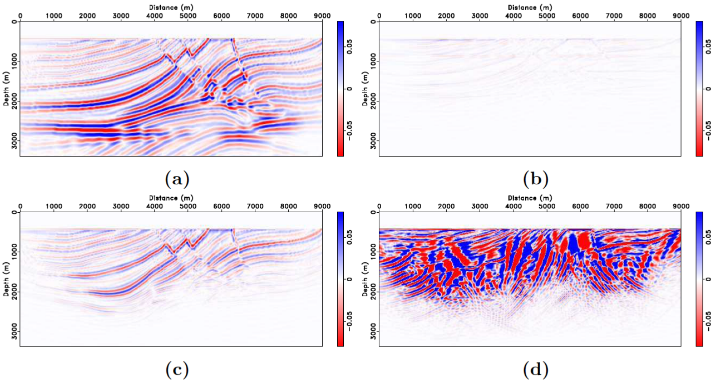



**Figure 2.11:** Marmousi example: Inverted reflectivity model at true background model (a); initial background model (b); background model with 18 steps of VPE method (c) and 18 steps of VP method (d) by solving equation 2.9.

# Chapter 2

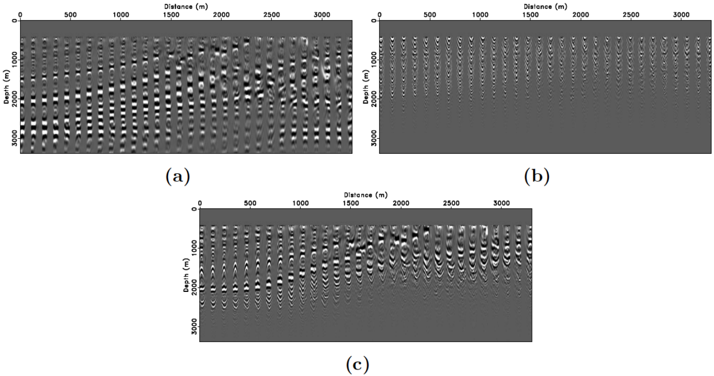



**Figure 2.12:** Marmousi example: Common image gathers at true background model (a); initial background model (b); background model with 18 steps of VPE method (c) by solving equation2.9.

RICE

# Chapter 2

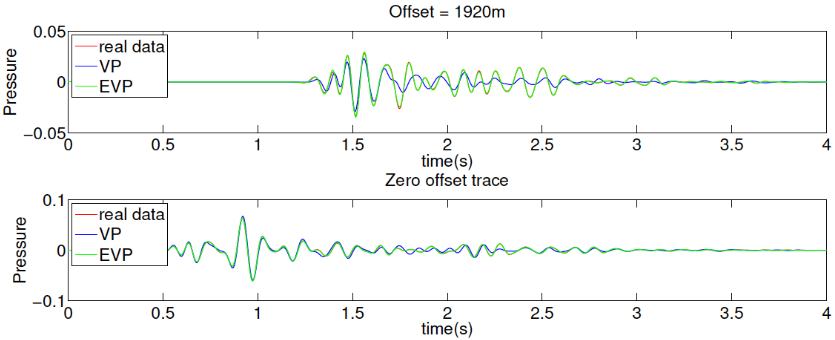



**Figure 2.13:** Marmousi example: Trace comaprsion of real data (red), predicted data by VPE (gree) and VP (blue) methods for far (top) and near (bottom) offsets.

RICE

# Chapter 2

Bottom line: works, but slow

18 VP its $\times$ 50 CG iterations - *way* too much

# Chapter 3

*Flexibly Preconditioned Extended Least Squares Migration in Shot Record Domain*

Joint with Rami Nammour - in review @ *Geophysics*

RICE

# Chapter 3

Task: use $\Psi$DO scaling to precondition inner problem

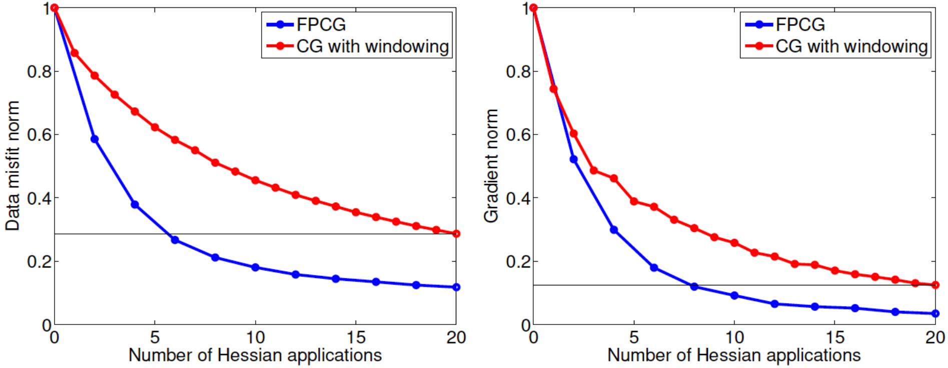
# Chapter 3

$\Psi$DO scaling - Nammour 09, uses Bao-S. 96

Estimate amplitude by 2 Hessian ops

Flexibly Preconditioned CG

# Chapter 3




(a) (b)

**Figure 3.15:** Marmousi example: convergence curves of numerical methods, (a) normalized data misfit and (b) normalized gradient length.

# Chapter 3

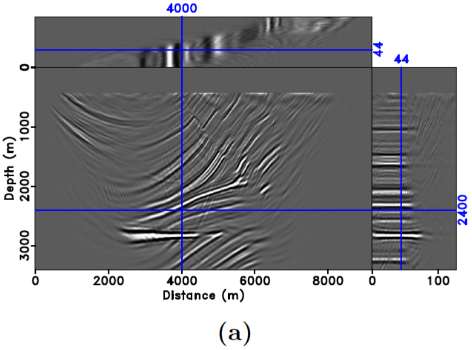



**Figure 3.16:** Marmousi example: inverted model perturbation cube after 20 Hessian applications using FPCG (a) and using CG with windowing (b).

RICE

# Chapter 3

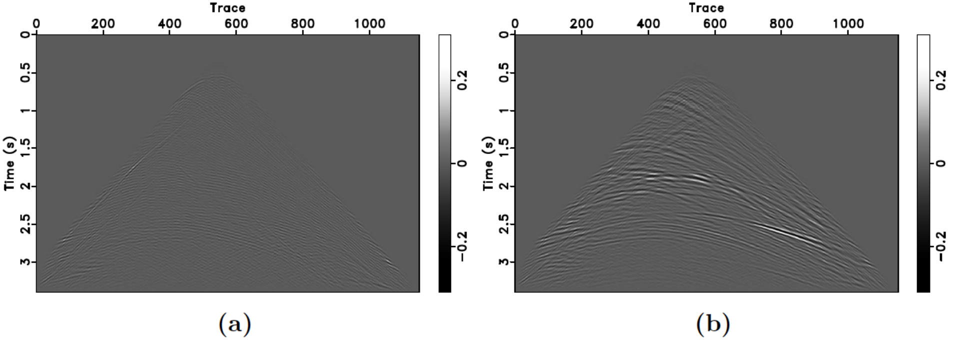



**Figure 3.20:** Marmousi example: data residual, same shot record as in Figure 3.14b, after 20 Hessian applications using FPCG (a), and using CG with windowing (b).

# Chapter 3

Bottom line: speedup by factor of 3-4

Much better inner solve with same effort

# Chapter 4

Task: evaluate effect of FPCG/CG inner solve on gradient accuracy

# Chapter 4

Fast lens over flat reflector

Computed gradient at const background model

# Chapter 4

Relative error in

$$\frac{J[m + h\delta m] - J[m - h\delta m]}{2h}$$

as approx to $\langle \nabla J[m], \delta m \rangle$

# Chapter 4

| Numerical methods | dm1 | | | dm2 | | |
|---|---|---|---|---|---|---|
| h | 0.1 | 0.05 | 0.025 | 0.1 | 0.05 | 0.025 |
| FPCG | 0.1108 | 0.0525 | 0.0643 | 0.1715 | 0.0814 | 0.1346 |
| CG with $I_t^2$ | 0.1131 | 0.1147 | 0.1105 | 0.0467 | 0.0579 | 0.0633 |
| CG with windowing | 0.9890 | 1.0088 | 1.0673 | 1.3502 | 1.1384 | 0.6748 |

**Table 4.1:** Gradient test at constant background model $m = (2\text{km/s})^2$ for different $dm$ and different numerical methods with 20 applications of LSM Hessian

RICE

# Chapter 4

Bottom line: not so hot

Why? Look to nature of tomo op

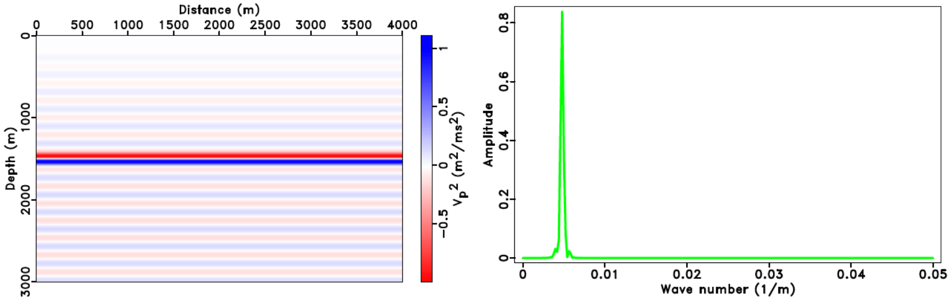
# Chapter 4

$$J[m] = min_r \frac{1}{2}\|F[m[r] - d\|^2 + \alpha^2\|Ar\|^2$$

$$\nabla J[m] = DF[m](F[m]r - d, r)$$

fact: $DF[m]$ is badly scaled (unbounded)

# Chapter 4




(a) (b)

**Figure 4.3:** Model $\delta m + r_k$ (a) and spectrum of model $r_k$ (b) with $k = 30$.

# Chapter 4

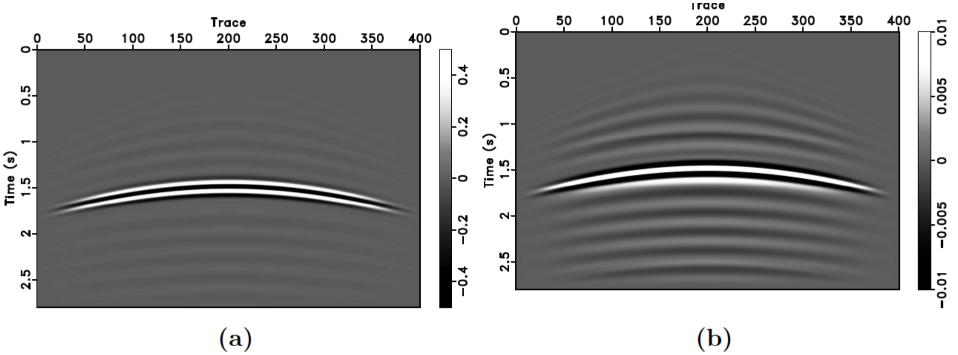



**Figure 4.27:** Output of $DF[m][dm, \delta m + r_k]$ with $k = 30$, $m = (2\text{km/s})^2$ and $\delta m$ shown in Figure 4.10b with bandpass filtered source wavelet (a) and source wavelet integrated over time twice (b).

# Chapter 4

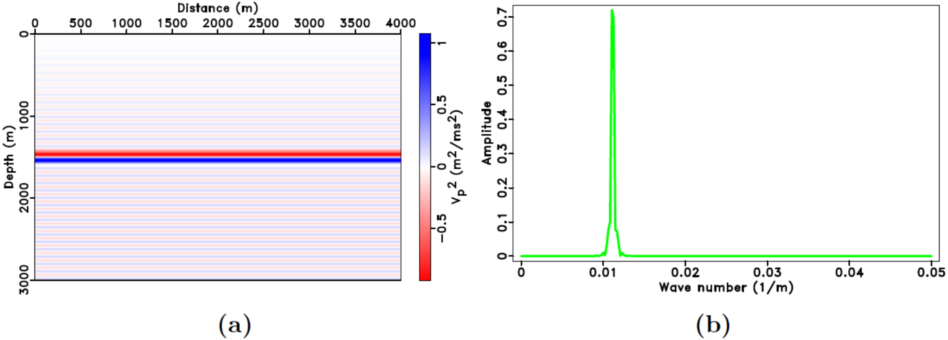
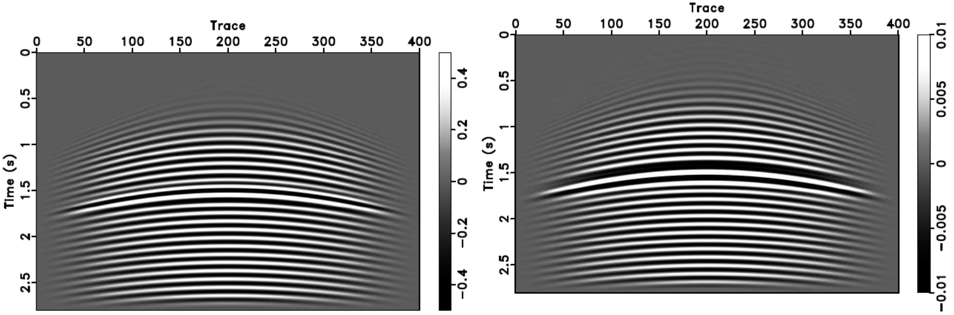



**Figure 4.5:** Model $\delta m + r_k$ (a) and spectrum of model $r_k$ (b) with $k = 70$.

# Chapter 4




**(a)** **(b)**

**Figure 4.28:** Output of $DF[m][dm, \delta m + r_k]$ with $k = 70$, $m = (2\text{km/s})^2$ and $\delta m$ shown in Figure 4.10b with bandpass filtered source wavelet (a) and source wavelet integrated over time twice (b).

RICE

# Chapter 4

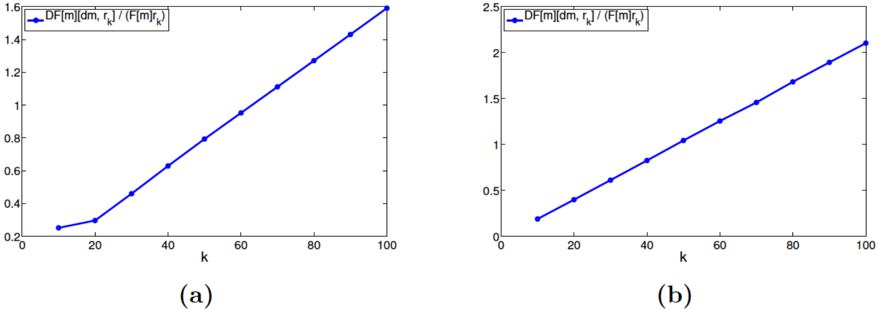



**Figure 4.30:** Quotient of $L_2$ norms of $DF[m][dm, r_k]$ and $F[m]r_k$, when using (a) bandpass filtered source wavelet (Figure 4.1b) and (b) twice integral of bandpass filtered source wavelet (Figure 4.8).

RICE

# Chapter 4

Order of DF $=$

Order of F $+ 1$

Convergent inner solve not sufficient for convergent computed gradient

# Convergent Gradients

Key ingredients:

- parametrix $=$ asymptotic inverse
- robust optimization

# Convergent Gradients

Computable parametrices exist -

- subsurface offset extn
- some source extns

# Convergent Gradients

Example: subsurface offset acoustic Born (Hou &S. *Geophys.* 15)

$F[v]$ = modeling op, velo v

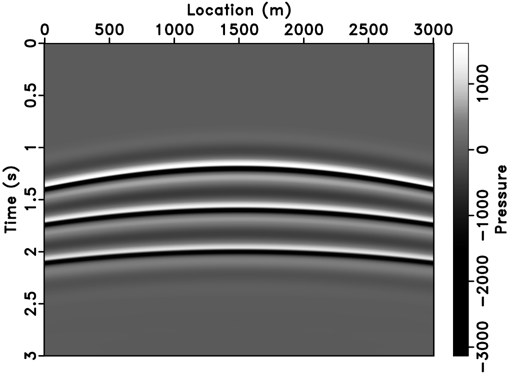
$F[v]^\dagger$ = asympt inverse = $W_{\mathrm{m}}[v]^{-1}F[v]^T W_{\mathrm{d}}[v]$

# Convergent Gradients

Weight ops $W_{\rm m}, W_{\rm d}$ are *filters* - cheap, no raytracing or PDE solves

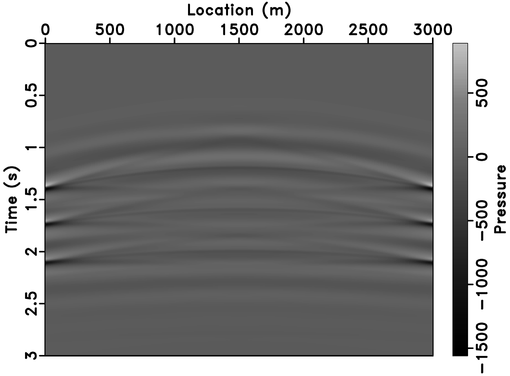
Makes $F$ *almost unitary* in weighted norms

# Convergent Gradients




Modeled data $d = F[v]r$

# Convergent Gradients




Data residual $F[v]F[v]^{\dagger}d$

# Convergent Gradients

means (roughly):

$$F[v]^{\dagger}F[v] = I + S[v],$$

S[v] is *smoothing* (suppresses HF signal) of order -1

# Convergent Gradients

Ignoring regularization, $\alpha \rightarrow 0$ limit is

$$\min \; J[v] = \frac{1}{2}\|Ar[v]\|^2 :$$

$$\text{subj } F[v]^\dagger F[v] r[v] = F[v]^\dagger d$$

some algebra (see paper in TRIP16)...

# Convergent Gradients

and ignoring second LS problem,

$$\nabla J[v] = DF[v]^*(d, A^{\dagger}Ar[v])$$

trouble: $r[v] \leftarrow r_{\mathrm{approx}}$

# Convergent Gradients

$$F[v]^{\dagger}F[v]r[v] = r[v] + S[v]r[v] = F[v]^{\dagger}d$$

so

$$r[v] = F[v]^{\dagger}d - S[v]r[v]$$

# Convergent Gradients

order $+1$/-1:

$$\nabla J[v] \approx DF[v]^*(d, A^\dagger AF[v]^\dagger d)$$

$$+DF[v]^*(d, A^\dagger AS[v]r_{\mathrm{approx}})$$

First term: Jie's appinv gradient; second term: correction for inner inversion

RICE

# Convergent Gradients

can compute $S[v] = I - F[v]^\dagger F[v]$!

[grad error] $\leq$ [error in r]

RICE

# Convergent Gradients

more huffing and puffing:

[Error in $\nabla J$] $\leq$ [$K \times$ error in normal eqn]

more trouble: no explicit control of $K$

# Convergent Gradients

Heinkenschloss-Vicente 01: variant of *trust-region* qN

step length control: short enough step is near steepest descent so always works

# Convergent Gradients

H-V01: converges with inexact grad, provided

[grad error] $\leq K \times$ max(|approx grad|, step bound)

our case: [error in normal eqn] $\leq$ max(...)

# Convergent Gradients

Upshot: assure convergence via

- parametrix $\Rightarrow$ control grad error
- couple grad error to step control

# Conclusion

Inversion: practical $\Rightarrow$ reliable, efficient

- Yin's thesis: shot record LSM accel., clarified reliability issues with EFWI
- for separable EFWI: critical requirement is computable parametrix
- couple accuracy and step control

RICE

# Thanks to...

- our sponsors
- present and former TRIPpers
- many colleagues
- developers of SU, Madagascar
- TACC & RCSG

RICE