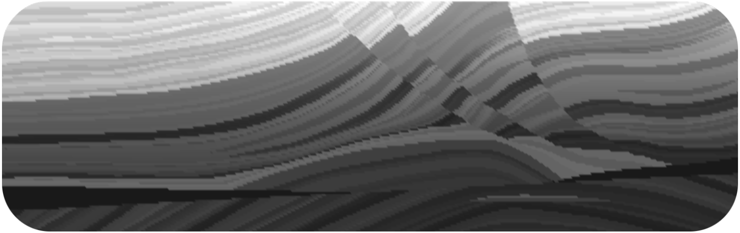
Full waveform inversion in the frequency domain

A trust-region Newton method



X . Adriaens[1], L. Métivier[2] and C. Geuzaine[1]

[1]Université de Liège

[2]Université Grenoble Alpes

September 2021

Problem statement

Consider

- a model m
- a wave propagation operator F
- a wavefield u
- a mesurement operator R
- a dataset d

Full wave inversion consists in finding m^* such that

$$R(u) = d \text{ with } F(m^*)u = f$$

through the **optimization problem**

$$m^* = \arg\min_m J(m) \triangleq \arg\min_m \text{dist}(R(u(m)), d)$$

Newton methods

Derivation

Newton methods are **local optimization techniques** that originate from successive **second order expansions** of the misfit

$$\begin{aligned} J(m+\delta m) &\approx J(m) + \{D_m J\}(\delta m) + \frac{1}{2}\{D^2_{mm} J\}(\delta m, \delta m) \\ &\approx J(m) + \langle j', \delta m \rangle_M \quad + \frac{1}{2}\langle H\delta m, \delta m \rangle_M \end{aligned}$$

The optimal search direction w.r.t. this expansion is called the **pure Newton direction** p_N. It is defined as the **solution of a linear system**

$$H p_N = -j'$$

According to Newton methods, the model is **updated iteratively** along an **approximation of this pure direction**.

Preconditioner

Equivalence bewteen both expansions is granted by the gradient j' and the Hessian operator H definitions

$$\langle j', \delta m \rangle_M \triangleq \{D_m J\}(\delta m), \; \forall \delta m$$

and

$$\langle H \delta m_1, \delta m_2 \rangle_M \triangleq \{D^2_{mm} J\}(\delta m_1, \delta m_2), \; \forall \delta m_1 \, \forall \delta m_2$$

that **strongly depend on the chosen inner product** $\langle \cdot, \cdot \rangle_M$.

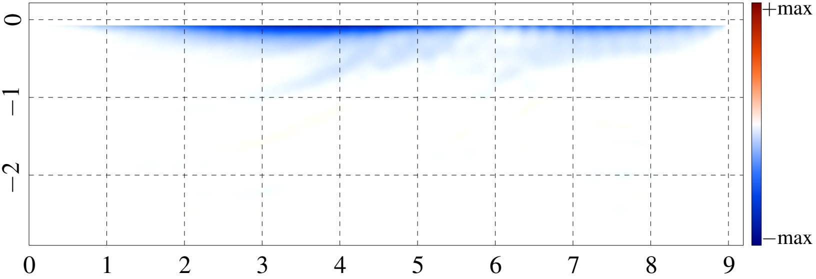
Changing the inner product is equivalent to **applying a preconditioner** to both kernels[1]. Preconditioning does not change the pure Newton direction but does change approximate solutions.

[1] Zuberi and Pratt, "Mitigating nonlinearity in full waveform inversion using scaled-Sobolev pre-conditioning".

Example I: conventional

The standard choice is a **least squares inner product**, which yields the conventional gradient and Hessian operator.

$$\langle \cdot, \cdot \rangle_M = \langle \cdot, \cdot \rangle_{L_2}$$

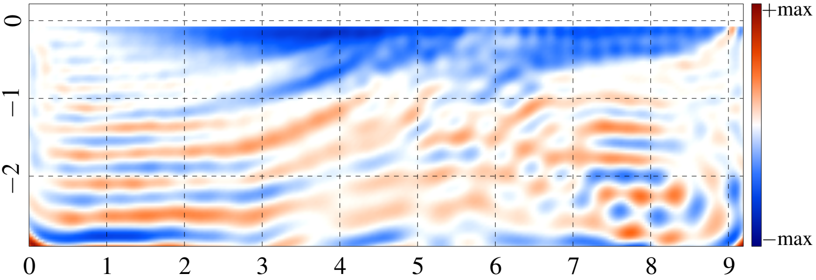


Balance between shallow and deep contributions is broken.

Example II: spatially scaled

An appropriate **spatial weight** is often applied, *e.g.* the diagonal of the Gauss-Newton Hessian[2] $h_a(\boldsymbol{x})$.

$$\langle \cdot, \cdot \rangle_M = \left\langle \cdot \sqrt{h_a(\boldsymbol{x})}, \sqrt{h_a(\boldsymbol{x})} \cdot \right\rangle_{L_2}$$



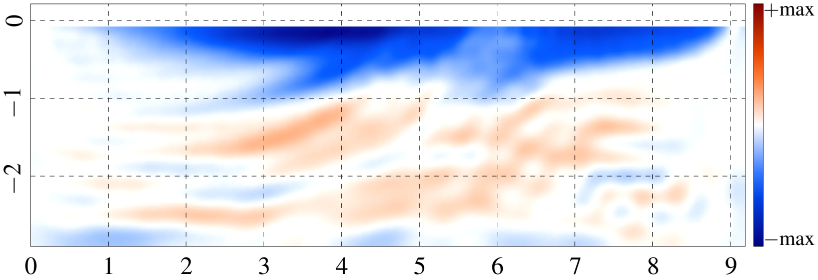
This inner product choice **restores balance** between gradient contributions everywhere.

[2]W. Pan, Innanen, and Liao, "Accelerating Hessian-free Gauss-Newton full-waveform inversion via l-BFGS preconditioned conjugate-gradient algorithm".

Example III: spatially smoothed

A **regularization term** penalizing rough models can also be added. Kernels w.r.t this inner product are therefore smoother.

$$\langle \cdot, \cdot \rangle_M = \left\langle \cdot \sqrt{h_a(\boldsymbol{x})}, \sqrt{h_a(\boldsymbol{x})} \cdot \right\rangle_{L_2} + \mu \langle \boldsymbol{\nabla} \cdot, \boldsymbol{\nabla} \cdot \rangle_{L_2}$$



Encouraging **smooth updates** early in the inversion process is a strategy to avoid local minima trapping[3].

[3] Zuberi and Pratt, "Mitigating nonlinearity in full waveform inversion using scaled-Sobolev pre-conditioning".

Truncated Newton methods

Approximate solution

Approximate solution are here obtained by **solving the Newton system iteratively** with an **accuracy reflecting the expansion quality**.
Hessian-free iterative methods such as the **conjugate gradient algorithm** are preferred because the linear system is typically large.

Conventional conjugate gradient (Alg. 1a)

$p \leftarrow 0$, $r \leftarrow j'$, $q \leftarrow -j'$
loop
  if $\langle Hq, q \rangle_M \leq 0$ **then return** p (Safeguard for negative curvature)
  $\xi \leftarrow \langle r, r \rangle_M$
  $\alpha \leftarrow \frac{\xi}{\langle Hq,q \rangle_M}, \quad p \leftarrow p + \alpha q$, $r \leftarrow r + \alpha H q$

  if $\|r\|_M < \eta \|j'\|_M$ **then return** p (Convergence criterion)
  $\beta \leftarrow \frac{\langle r,r \rangle_M}{\xi}, \quad q \leftarrow -r + \beta q$
end loop

Line search

Over-solving is avoided by **relaxing the convergence criterion** depending on the expansion quality[4,5]

$$\eta = \frac{\|j'(m_n) - j'(m_{n-1}) - \gamma_{n-1} H(m_{n-1}) p_{n-1}\|_M}{\|j'(m_{n-1})\|_M} \quad (\diamond)$$

An appropriate length γ is given to this direction p, *e.g.* satisfying Wolfe conditions[6]. The outer loop is finally obtained by repeating these two steps until convergence.

Eisenstat line search (Alg. 2a)

loop
  $p \leftarrow$ Alg. 1 with $\eta = (\diamond)$
  $m \leftarrow m + \gamma p$
end loop

[4] Métivier et al., "Full Waveform Inversion and the Truncated Newton Method".
[5] Eisenstat and Walker, "Choosing the forcing terms in an inexact Newton method".
[6] Nocedal, Wright, and Robinson, *Numerical Optimization*.

Trust region I

Over-solving is avoided by **adding a length constraint** on the Newton system approximate solution

$$\|p\|_M \leq \Delta$$

This new problem can be solved approximately with a slightly modified version of the conjugate gradient method[7].

[7] Steihaug, "The Conjugate Gradient Method and Trust Regions in Large Scale Optimization".

Trust region II

Steihaug conjugate gradient (Alg. 1b)

$p \leftarrow 0$, $r \leftarrow j'$, $q \leftarrow -j'$
loop
  if $\langle Hq, q \rangle_M \leq 0$ **then** (Safeguard for negative curvature)
    Find $\tau^* > 0$ such that $\|p + \tau^* q\|_M = \Delta$
    return $p + \tau^* q$
  end if
  $\xi \leftarrow \langle r, r \rangle_M$, $\alpha \leftarrow \frac{\langle r,r \rangle_M}{\langle Hq,q \rangle_M}$
  if $\|p + \alpha q\|_M \geq \Delta$ **then** (Trust region constraint)
    Find $\tau^* > 0$ such that $\|p + \tau^* q\|_M = \Delta$
    return $p + \tau^* q$
  end if
  $p \leftarrow p + \alpha q$, $r \leftarrow r + \alpha H q$
  if $\|r\|_M < \eta \|j'\|_M$ **then return** p (Convergence criterion)
  $\beta \leftarrow \frac{\langle r,r \rangle_M}{\xi}$, $q \leftarrow -r + \beta q$
end loop

Trust region III

The trust region radius is controlled by the outer loop, again depending on the expansion quality[8].
Radius evolution depends on the **ratio between the actual decrease and the predicted decrease**

$$\rho := \frac{\delta J_{\mathrm{a}}}{\delta J_{\mathrm{p}}}$$

with

$$\begin{cases} \delta J_{\mathrm{a}} & = J(m+p) - J(m) \\ \delta J_{\mathrm{p}} & = \left\langle j', p \right\rangle_M + 0.5 \left\langle Hp, p \right\rangle_M \end{cases}$$

[8] Fan, J. Pan, and Song, "A Retrospective Trust Region Algorithm with Trust Region Converging to Zero".

Trust region IV

Fan trust region (Alg. 2b)

Require: $0 \leq \rho_0 < \rho_1 < 1$ and $0 < c_0 < 1 < c_1$

loop

$\mu \leftarrow 1$

$p \leftarrow$ Alg. 1b with $\Delta = \mu \, \|j'\|_M$

$\delta J_\mathrm{a} = J(m+p) - J(m)$

$\delta J_\mathrm{p} = \langle j', p\rangle_M + 0.5 \, \langle Hp, p\rangle_M$ (Quality ratio)

$\rho = \delta J_\mathrm{a} / \delta J_\mathrm{p}$

if $\rho \geq \rho_0$ **then** $m \leftarrow m + p$ **else** $m \leftarrow m$ (Step acceptance)

if $\rho < \rho_1$ **then** $\mu \leftarrow c_0 \mu$

else if $\rho \geq \rho_1$ **and** $\|p\|_M > 0.5\Delta$ **then** $\mu \leftarrow c_1 \mu$ (Radius update)

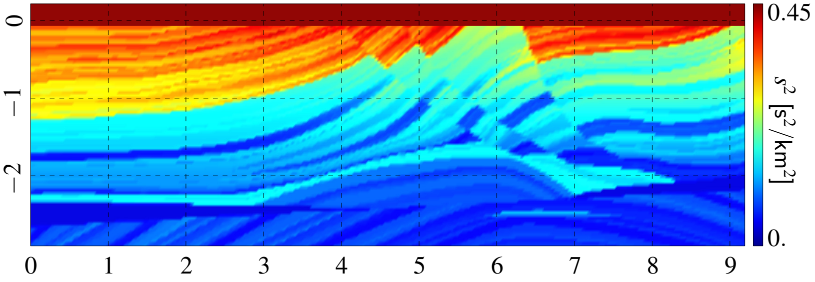
else **then** $\mu \leftarrow \mu$

end loop

Application

Marmousi 2D (acoustics)

Algorithms are compared on the **Marmousi 2D acoustic** case[9]



with

- 122 emitters and 122 receivers on the surface
- 3 frequencies $(4, 6, 8$ [Hz]$)$ inverted sequentially
- data distances measured by the least square norm

[9] Versteeg, "The Marmousi experience: Velocity model determination on a synthetic complex data set".

Minimization algorithms

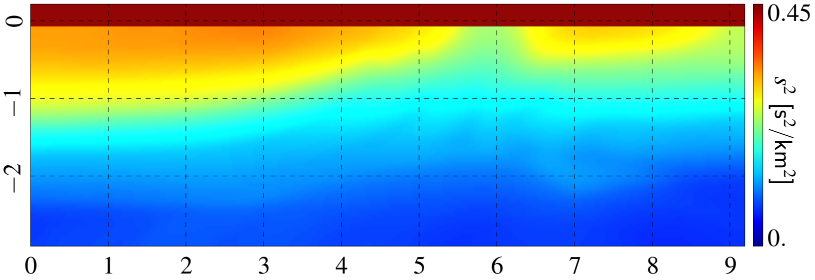
Three algorithm are compared

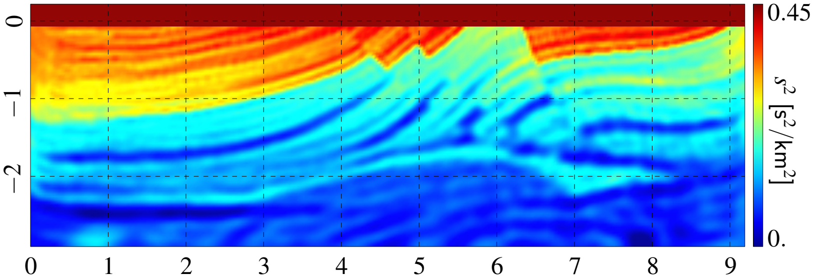
- A gradient descent method (gd)
- A line search truncated Newton method (ls)
- A trust region truncated Newton method (tr)

For the trust region implementation, three parameter sets are compared

	ρ_0	ρ_1	c_0	c_1	
(a)	10^{-4}	0.25	0.2	$5.$	Radius increases rapidly
(b)	10^{-4}	0.75	0.25	$2.$	Radius increases more cautiously
(c)	10^{-4}	0.9	0.5	$2.$	Radius increases very cautiously

Inversion results

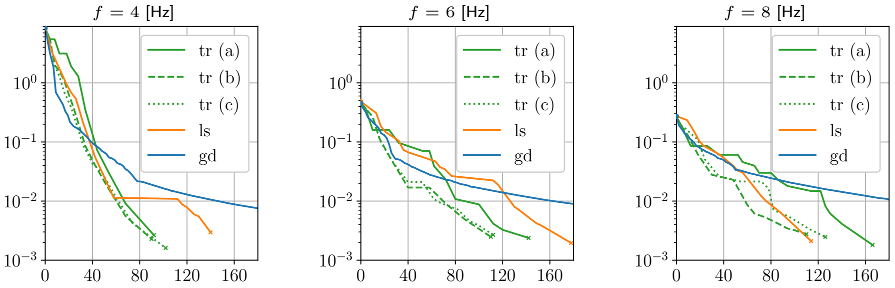




Convergence

Computational complexity is quantified by the **wave solution count**, which is the **number of forward problem solved**.

	gd	ls	tr (a)	**tr (b)**	tr (c)
Wave sol. (tot)	1303	432	400	310	340
Outer it. (tot)	630	42	42	33	41
Inner it. (avg)	(1.)	3.81	3.76	3.7	3.15
Rejected (%)	.04	.24	.21	.03	.07
Constrained (%)	-	-	.83	.85	.93



Conclusion

1. Truncated Newton methods converge faster than gradient methods.
2. Line search and trust region implementations are very similar.
3. Line search and loose trust region perform similarly.
4. Tighter trust region reduces over-solving and converge even faster.

5. Trust region are known for being robust.
   It still needs to be verified in the context of full wave inversion.

Conclusion

1. Truncated Newton methods converge faster than gradient methods.
2. Line search and trust region implementations are very similar.
3. Line search and loose trust region perform similarly.
4. Tighter trust region reduces over-solving and converge even faster.
5. Trust region are known for being robust.
   It still needs to be verified in the context of full wave inversion.

Thank you for your attention