



HAL
open science

A trust region Newton method for frequency domain full waveform inversion

Xavier Adriaens, Ludovic Métivier, Christophe Geuzaine

► To cite this version:

Xavier Adriaens, Ludovic Métivier, Christophe Geuzaine. A trust region Newton method for frequency domain full waveform inversion. First International Meeting for Applied Geoscience & Energy, 2021, Denver, United States. pp.757-761, <10.1190/segam2021-3581935.1>. <hal-03404514>

HAL Id: hal-03404514

<https://hal.science/hal-03404514v1>

Submitted on 26 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

A trust region Newton method for frequency domain full waveform inversion

Xavier Adriaens*, Université de Liège, F.R.S-FNRS; Ludovic Métivier, Université Grenoble Alpes, CNRS, LJK; Christophe Geuzaine, Université de Liège

SUMMARY

Exploiting Hessian information greatly improves the quality of full wave inversion. A theoretically simple way to incorporate this second order derivative is to minimize the misfit using Newton method. In practice however the pure Newton method is impossible to implement, because it requires inverting the Hessian which can actually not be constructed explicitly. Also the misfit is not necessarily quadratic, thus the exact Newton direction is not necessarily the best. Consequently it is natural to turn to truncated Newton methods, where the search direction is constructed iteratively to approximate the pure Newton direction. The bottleneck of these methods lies in the compromise to find between a direction built in few iterations, but which hardly takes the Hessian into account and a nearly exact direction which is very expensive to compute. In this work we present a truncated Newton method based on a particular trust region algorithm, in the context of full waveform inversion in the frequency domain. A numerical test is then performed to compare convergence speed with another well known truncated Newton method. This test shows that the trust region method demonstrates faster convergence, if parameters are adequately chosen.

INTRODUCTION

Full waveform inversion is a data fitting technique whose aim is to recover some model parameters by minimizing the discrepancy between recorded data and data simulated by solving wave propagation problems. By nature these data are oscillatory and consequently the misfit quantifying the discrepancy features local minima (Mulder and Plessix, 2008). Global optimization techniques should ideally be used but the typically very high dimensions of the search space prohibits their use and only local optimization tools can practically be used, with care. A straightforward direction to iteratively update the model properties is of course the (preconditioned) gradient, *i.e.* the direction of steepest decrease. However it is well-known that the inverse Hessian plays a crucial role in the reconstruction (Pratt and Shin, 1998) in addition to offering the possibility to account for coupling effects between parameter classes for multi-parameter inversions. Thus increasing attention is now paid to Newton direction, defined as the solution of a linear system involving this Hessian. Once again due to the size of the optimization problem, it is currently not bearable to store let alone invert this Hessian matrix. Hessian-free methods such as the conjugate gradients algorithm are therefore the adequate tool to solve this Newton system without constructing explicitly nor the Hessian, nor its inverse. Once a direction is chosen, the outer iteration is usually completed by finding the optimal length of the step that should be performed in that direction through a line search procedure. Among the non lin-

ear optimization community, it is sometimes argued however that line search is not well suited with Newton directions, especially when the Hessian is nearly singular. Indeed when the Hessian is nearly singular, Newton direction becomes excessively long such that the quadratic approximation implicitly made when computing it ceases to hold. Much computational effort must then be made by the line search procedure to reduce the step size (Nocedal et al., 2006). Stopping the iterative solution of the Newton system earlier appears as a solution to this problem. For example Métivier et al. (2013) proposes to relax the convergence requirements of the conjugate gradient method such that it reflects the accuracy of the quadratic approximation.

In this contribution we propose instead to limit the length of the Newton direction, also depending on the quadratic approximation accuracy. Such techniques are commonly called trust region methods.

THEORY

Full waveform inversion consists in finding the optimal model parameter m^* whose corresponding wave field u , defined through a wave propagation operator \mathcal{F} , matches the recorded data set d after a projection \mathcal{R} onto receivers

$$\mathcal{R}(u) = d \text{ with } \mathcal{F}(u, m^*) = f \quad (1)$$

through the optimization problem

$$m^* = \arg \min_m \mathcal{J}(m) \text{ with } \mathcal{J}(m) := \text{dist}(\mathcal{R}(u(m)), d). \quad (2)$$

Local optimization techniques are based on a local expansion of the misfit \mathcal{J} around the current model estimate

$$\begin{aligned} \mathcal{J}(m + \delta m) &\approx \\ \mathcal{J}(m) + \{D_m \mathcal{J}\}(\delta m) + \frac{1}{2} \{D_{mm}^2 \mathcal{J}\}(\delta m, \delta m). \end{aligned} \quad (3)$$

This expansion can also be written in terms of the gradient j' and the Hessian operator \mathcal{H} once an inner product $\langle \cdot, \cdot \rangle_M$ is chosen for the model space M

$$\begin{aligned} \mathcal{J}(m + \delta m) &\approx \\ \mathcal{J}(m) + \langle j', \delta m \rangle_M + \frac{1}{2} \langle \mathcal{H}(\delta m), \delta m \rangle_M. \end{aligned} \quad (4)$$

Resulting from this expansion, the pure Newton direction p_N is defined as the solution of the linear system

$$\mathcal{H}(p_N) = -j'. \quad (5)$$

The large-scale nature of this linear system requires the use of Hessian-free iterative methods. The Hessian operator being symmetric, the conjugate gradient method is the ideal candidate. An additional safeguard is however added to exit prematurely when directions of negative curvature are encountered.

Trust region Newton method

Such directions exist because the full Hessian is not necessarily positive definite, especially far from the global minimum.

The choice of the inner product plays a central role in the inversion as it defines both gradients and Hessians and is actually equivalent to preconditioning them (Zuberi and Pratt, 2017). Basically different choices of inner product do not modify the pure Newton direction but do modify the subspace constructed by the conjugate gradient method. A good choice could thus lead to better convergence or to better directions if the convergence criterion can not be met, for example because of negative curvature or trust region violation.

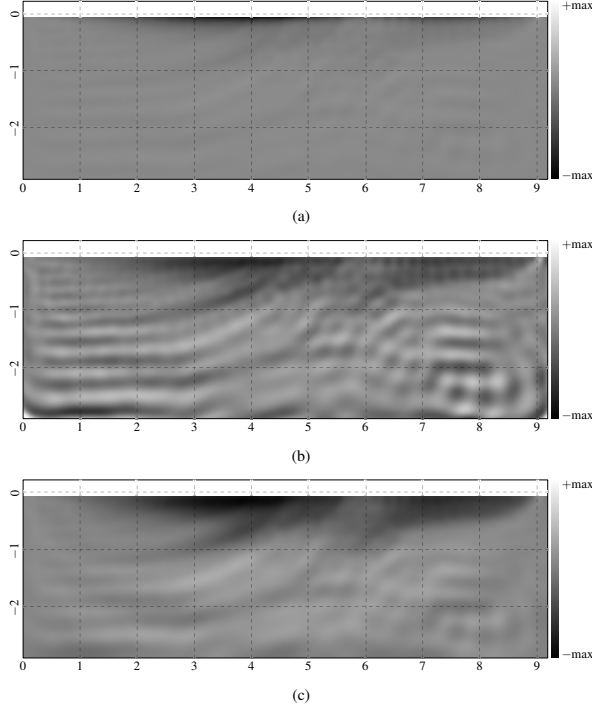


Figure 1: Gradient for the model given in Figure 2(b) for different inner product choices. (a) conventional (b) first term of Eq.(6) (c) both terms of Eq.(6).

Here we chose an inner product composed of two terms

$$\langle \cdot, \cdot \rangle_M := \left\langle \tilde{h}_{\text{diag}}^{1/2} \cdot, \tilde{h}_{\text{diag}}^{1/2} \cdot \right\rangle + \mu \cdot \left\langle \tilde{\nabla} \cdot, \tilde{\nabla} \cdot \right\rangle. \quad (6)$$

The first term is related to the diagonal part of the Gauss-Newton Hessian $\tilde{h}_{\text{diag}}^{1/2}$ and compensates for the geometrical spreading (Pan et al., 2017) while the second term, related to spatial derivation, increases the norm of directions that are rapidly varying and prevents the inner product from being insensitive in regions where the diagonal Hessian is close to zero. This inner product is very similar to the one introduced by Zuberi and Pratt (2017), except that the Gauss-Newton diagonal Hessian is used in addition. The stabilizing parameter μ can be expressed in terms of a characteristic length l_c

$$\mu := \tilde{h}_{\text{thres}}(l_c/2\pi)^2 \quad (7)$$

with \tilde{h}_{thres} a threshold value for the diagonal Hessian. In regions where the diagonal Hessian is close to the threshold, then

directions with details smaller than this characteristic length l_c are penalized with respect to smoother directions. From the point of view of preconditioning, this inner product generates a depth-scaling and a smoothing. The effect of this inner product is illustrated in Figure 1 on the first gradient computed during the inversion process described in the application section.

Line search

Newton method can be globalized by a line search procedure. In that case a direction p is first found by solving the Newton system approximately with the conventional conjugate gradient method (Algorithm 1). Over-solving is here avoided through the forcing term η , which is not systematically close to zero but which is instead chosen to reflect the accuracy of the second order expansion. Métivier et al. (2013) compared three famous choices for this forcing term in the context of acoustic imaging.

Algorithm 1 Conventional conjugate gradient

```

 $p \leftarrow 0, r \leftarrow j', q \leftarrow -j'$ 
loop
  if  $\langle \mathcal{H}(q), q \rangle_M \leq 0$  then return  $p$ 
   $\xi \leftarrow \langle r, r \rangle_M$ 
   $\alpha \leftarrow \frac{\xi}{\langle \mathcal{H}(q), q \rangle_M}, p \leftarrow p + \alpha q, r \leftarrow r + \alpha \mathcal{H}(q)$ 
  if  $\|r\|_M < \eta \|j'\|_M$  then return  $p$ 
   $\beta \leftarrow \frac{\langle r, r \rangle_M}{\xi}, q \leftarrow -r + \beta q$ 
end loop

```

An appropriate length γ is then given to this direction p , ideally the global minimum along the line $m + \gamma p$. In practice however less stringent satisfactory conditions are used instead (e.g. Armijo, Goldstein or Wolfe conditions (Nocedal et al., 2006)) to spare expensive wave problem resolutions. The outer loop is finally obtained by repeating these two steps until convergence.

Trust region

When the Newton method is globalized by a trust region technique, both the direction and its length are found simultaneously by solving the Newton system with an additional constraint

$$\mathcal{H}(p_N) = -j' \text{ with } \|p_N\|_M \leq \Delta. \quad (8)$$

This new problem can be solved approximately with a slightly modified version of the conjugate gradient method (Algorithm 2) due to Steihaug (1983). Basically there are only two modifications compared to Algorithm 1. First, the inner iterations are cropped to the trust region radius Δ when the unconstrained solution increases beyond it. Second, directions of negative curvature are followed up to the boundary of the trust region while these directions are never investigated in the conventional version.

The size of the trust region is actually controlled by the outer iterations. The decision of modifying the trust region is based on the accuracy of the second order expansion. When the expansion is accurate but the updates are limited by the length constraint, then the trust region radius is increased. At the opposite, when the updates are out of the range of validity of the expansion, then the trust region radius is decreased.

Trust region Newton method

Algorithm 2 Steihaug conjugate gradient

```

 $p \leftarrow 0, r \leftarrow j', q \leftarrow -j'$ 
loop
  if  $\langle \mathcal{H}(q), q \rangle_M \leq 0$  then
    Find  $\tau^* > 0$  such that  $\|p + \tau^* q\|_M = \Delta$ 
    return  $p + \tau^* q$ 
  end if
   $\xi \leftarrow \langle r, r \rangle_M$ 
   $\alpha \leftarrow \frac{\xi}{\langle \mathcal{H}(q), q \rangle_M}, p \leftarrow p + \alpha q, r \leftarrow r + \alpha \mathcal{H}(q)$ 
  if  $\|p\|_M \geq \Delta$  then
    Find  $\tau^* > 0$  such that  $\|p + \tau^* q\|_M = \Delta$ 
    return  $p + \tau^* q$ 
  end if
  if  $\|r\|_M < \eta \|j'\|_M$  then return  $p$ 
   $\beta \leftarrow \frac{\langle r, r \rangle_M}{\xi}, q \leftarrow -r + \beta q$ 
end loop

```

The quality of the expansion is quantified by the ratio ρ between the actual decrease $\delta \mathcal{J}_a := \mathcal{J}(m+p) - \mathcal{J}(m)$ and the decrease predicted by the second order expansion $\delta \mathcal{J}_p := \langle j', p \rangle_M + 0.5 \langle \mathcal{H}(p), p \rangle_M$

$$\rho := \frac{\delta \mathcal{J}_a}{\delta \mathcal{J}_p}. \quad (9)$$

This ratio is close to one when the expansion is accurate. Standard trust region methods directly control the radius Δ . However it is an absolute quantity, in the sense that it is compared to $\|p\|_M$, which depends on the inner product. Thus, it seems more natural to control this radius relatively to the gradient norm, which provides a length reference for the Newton system. In this way, even when Newton system changes scale from one iteration to another, the trust region remains relevant. This particular variant (Algorithm 3) has been first introduced by Fan et al. (2016).

Algorithm 3 Fan trust region

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

```

 $\mu \leftarrow 1$ 
loop
   $p \leftarrow$  Algorithm 2 with  $\Delta = \mu \|j'\|_M$ 
   $\delta \mathcal{J}_a = \mathcal{J}(m+p) - \mathcal{J}(m)$ 
   $\delta \mathcal{J}_p = \langle j', p \rangle_M + 0.5 \langle \mathcal{H}(p), p \rangle_M$ 
   $\rho = \delta \mathcal{J}_a / \delta \mathcal{J}_p$ 
  if  $\rho \geq \rho_0$  then  $m \leftarrow m + p$  else  $m \leftarrow m$ 
  if  $\rho < \rho_1$  then  $\mu \leftarrow c_0 \mu$ 
  else if  $\rho \geq \rho_1$  and  $\|p\|_M > 0.5 \Delta$  then  $\mu \leftarrow c_1 \mu$ 
  else then  $\mu \leftarrow \mu$ 
end loop

```

APPLICATION

In this section we present a standard numerical test case to which both methods presented above are applied. Final details of their implementation are also introduced.

Numerical tests are performed on the 2D Marmousi acoustic model (Versteeg, 1994) (Figure 2(a)) in the frequency domain. It is here chosen that the subsurface is described by the slowness squared s^2 [s^2/m^2], thus the forward operator writes

$$\mathcal{F}(p, s^2) = \Delta p + \omega^2 s^2 p. \quad (10)$$

Three frequencies (4, 6 and 8 [Hz]) are inverted sequentially from the lowest to the highest to avoid local minima (Bunks et al., 1995). Their spacing is chosen following the guidelines from Sirgue and Pratt (2004) regarding wavenumber coverage. Three inner product characteristic scales are used (0.8 [km], 0.5 [km] and 0.4 [km]), one for each frequency. A surface acquisition system composed of 122 equally spaced (72 [m]) emitters-receivers is used and the misfit \mathcal{J} is chosen as the conventional least-square distance between simulated and recorded data

$$\mathcal{J}(s^2) = \frac{1}{2} \sum_e \sum_r \left| p_e(x_r; s^2) - d_{er} \right|^2. \quad (11)$$

A smoothed version of the exact Marmousi model is used as an initial guess. This initial model is computed by a laplacian filter $m_{\text{init}} = (1 + (l_c/2\pi)^2 \Delta)^{-1} m_{\text{exact}}$ with $l_c = 2$ [km] (Figure 2(b)).

Slowness squared field and pressure fields at the three frequencies are discretized on a square grid (36 [m]) by hierarchical finite elements, respectively of order 1 and of order 2, 3, 4. A water layer (216 [m]) is also added at the top of the model but it is kept constant during the inversion. The model is spatially truncated by Sommerfeld boundary conditions (Schot, 1992). Any gradient or Hessian vector product is computed using the (second order) adjoint state method (Métivier et al., 2013). Recorded data are generated synthetically using the same hierarchical finite elements setting than for the inversion, to reduce numerical errors.

Line search

On the one hand we choose a line search algorithm that satisfies strong Wolfe conditions (Nocedal et al., 2006) with a parameter set such that a step is very easily accepted ($c_1 = 1e^{-4}$, $c_2 = 0.9$, $\gamma_0 = 1$). On the other hand the forcing sequence is chosen according to the guidelines of Métivier et al. (2013) (choice (a)).

Trust region

Three parameters sets (a,b and c) have been tested. The first one (a) is very similar to what was originally proposed by Fan et al. (2016). The other two (b,c) are more cautious because they modify the radius more rarely and when they do, it increases by a smaller factor. The forcing term is constant for all trust region methods ($\eta = 0.4$).

	ρ_0	ρ_1	c_0	c_1
(a)	$1e^{-4}$	0.25	0.2	5.
(b)	$1e^{-4}$	0.75	0.25	2.
(c)	$1e^{-4}$	0.9	0.5	2.

Table 1: Parameter sets for Fan trust region algorithm

Trust region Newton method

RESULTS

The squared slowness estimated with the trust region (c) method is shown in Figure 2(c). From a relatively low resolution initial guess (Figure 2(b)), full wave inversion indeed provides a high resolution estimation of the exact model (Figure 2(a)). The image obtained with the other methods does not differ significantly.

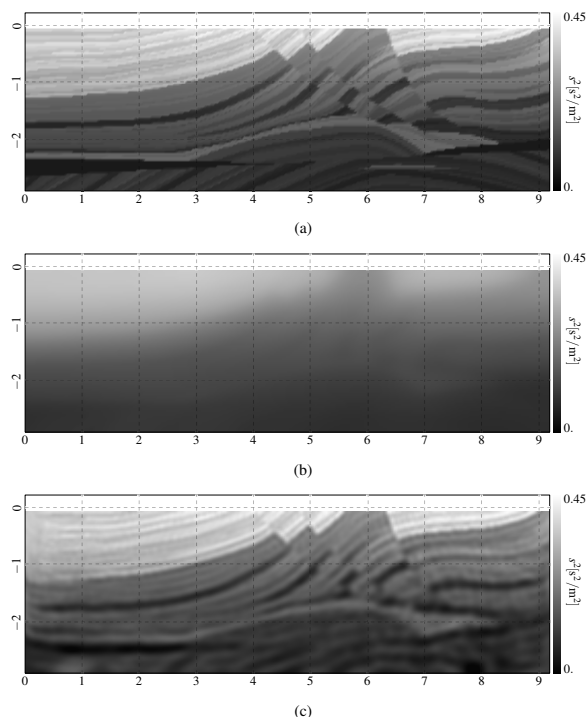


Figure 2: (a) Marmousi model (b) Initial guess (c) Inversion results (trust region (c))

Some interesting quantities concerning optimization algorithms are given in Table 2. For the comparison a preconditioned gradient descent has been added.

	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

Table 2: Statistics related to the different inversion algorithms.

Among these indicators lies the percentage of direction refusal (rejected). Such a denial yields an additional cost because either other step length must be tried (line search) or the entire external iteration must be restarted with a smaller radius Δ (trust region). The convergence analysis can not be done on the number of outer iterations (outer it.) because the computational cost of these iterations is highly dependant on the number of conjugate gradient iterations (inner it.). Instead each

outer iteration is quantified by the number of wave propagation problems it requires to solve (wave sol.), as detailed in Métivier et al. (2013). The misfit is plotted against this measure of computational complexity in Figure 3.

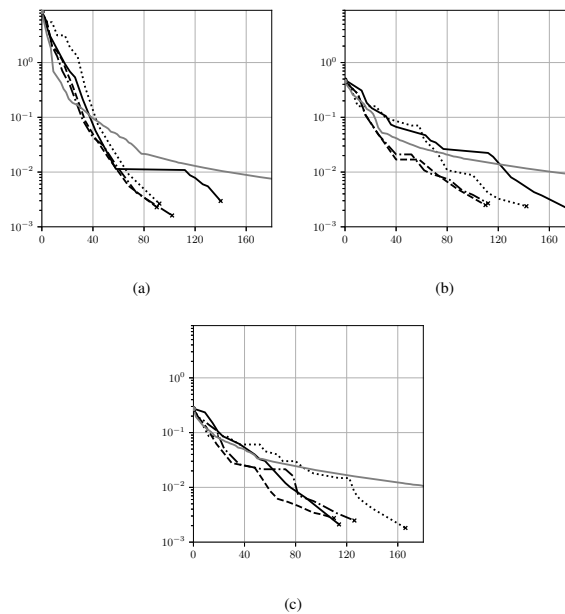


Figure 3: Misfit against the number of wave problem solved. (a) 4 [Hz] (b) 6 [Hz] (c) 8 [Hz]; gradient descent (—) line search (---) trust region (a)(··) (b) (- -) (c) (-·)

Not surprisingly second order methods converge orders of magnitude faster than the preconditioned gradient, while the average number of internal iterations is not much higher than one. (As a reminder a single internal iteration corresponds to a preconditioned gradient descent). Line search and trust region (a) methods give comparable results because both actually reject directions equally often. A rejected direction is potentially a heavy efficiency loss if lots of internal iterations were necessary to compute it. Plateaus appearing in the convergence curves are a consequence of these refusals. The trust region (b) shows the best convergence, closely followed by trust region (c). Their good performance is due to the fact that both almost do not reject any step and therefore do not waste computing time. Actually convergence slope looks the same for all second order methods if plateaus are omitted and the advantage of trust regions method lies in reducing the number rejected directions and thus the number of plateaus.

CONCLUSION

We introduced a trust region truncated Newton method and compared its computational performance with a line search truncated Newton method, in the context of full wave inversion in the frequency domain. In particular we showed that trust region method significantly reduces over-solving and thus yields faster convergence.

Trust region Newton method

ACKNOWLEDGMENTS

The authors would like to thank Anthony Royer for his help on the finite element solver used in this work (Royer et al., 2020). This research was funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) and the ARC grant for Concerted Research Actions (ARC WAVES 15/19-03), financed by the Wallonia-Brussels Federation of Belgium. The authors acknowledge the use of the computational resources provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) and by the Walloon Region.

Trust region Newton method

APPENDIX A

THE SOURCE OF THE BIBLIOGRAPHY

```
@article{Pan2017,
  title = {{Accelerating Hessian-free Gauss-Newton full-waveform inversion via l-BFGS preconditioned conjugate-gradient algorithm}},
  year = {2017},
  journal = {GEOPHYSICS},
  author = {Pan, Wenyong and Innanen, Kristopher A and Liao, Wenyuan},
  number = {2},
  month = {3},
  pages = {R49-R64},
  volume = {82}
}

@article{Sirgue2004,
  title = {{Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies}},
  year = {2004},
  journal = {GEOPHYSICS},
  author = {Sirgue, Laurent and Pratt, R. Gerhard},
  number = {1},
  month = {1},
  pages = {231--248},
  volume = {69}
}

@article{Mulder2008,
  title = {{Exploring some issues in acoustic full waveform inversion}},
  year = {2008},
  journal = {Geophysical Prospecting},
  author = {Mulder, W.A. and Plessix, R.-E.},
  number = {6},
  month = {11},
  pages = {827--841},
  volume = {56}
}

@article{Metivier2013,
  title = {{Full Waveform Inversion and the Truncated Newton Method}},
  year = {2013},
  journal = {SIAM Journal on Scientific Computing},
  author = {M{\'}{e}tivier, Ludovic and Brossier, Romain and Virieux, Jean and Operto, Stéphane},
  number = {2},
  month = {1},
  pages = {B401-B437},
  volume = {35}
}

@article{Pratt1998,
  title = {{Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion}},
  year = {1998},
  journal = {Geophysical Journal International},
  author = {Pratt, R. Gerhard and Shin, Changsoo},
  number = {2},
  pages = {341--362},
```

Trust region Newton method

```
    volume = {133}
}

@article{Zuberi2017,
  title = {{Mitigating nonlinearity in full waveform inversion using scaled-Sobolev
pre-conditioning}},
  year = {2017},
  journal = {Geophysical Journal International},
  author = {Zuberi, M AH and Pratt, R G},
  number = {1},
  pages = {706--725},
  volume = {213}
}

@article{Bunks1995,
  title = {{Multiscale seismic waveform inversion}},
  year = {1995},
  journal = {GEOPHYSICS},
  author = {Bunks, Carey and Saleck, Fatimetou M. and Zaleski, S. and Chavent, G.},
  number = {5},
  month = {9},
  pages = {1457--1473},
  volume = {60}
}

@book{Nocedal2006,
  title = {{Numerical Optimization}},
  year = {2006},
  author = {Nocedal, Jorge and Wright, Stephen J and Robinson, Stephen M},
  pages = {1--651},
  series = {Springer Series in Operations Research and Financial Engineering},
  publisher = {Springer New York}
}

@article{Marmousi2D,
  title = {{The Marmousi experience: Velocity model determination on a synthetic complex
data set}},
  year = {1994},
  journal = {The Leading Edge},
  author = {Versteeg, Roelof},
  number = {9},
  month = {9},
  pages = {927--936},
  volume = {13}
}

@article{Steihaug1983a,
  title = {{The Conjugate Gradient Method and Trust Regions in Large Scale
Optimization}},
  year = {1983},
  journal = {SIAM Journal on Numerical Analysis},
  author = {Steihaug, Trond},
  number = {3},
  pages = {626--637},
  volume = {20}
}

@article{Fan2016,
  title = {{A Retrospective Trust Region Algorithm with Trust Region Converging to Zero}},
```

Trust region Newton method

```
year = {2016},
journal = {Journal of Computational Mathematics},
author = {Fan, Jinyan and Pan, Jianyu and Song, Hongyan},
number = {4},
month = {6},
pages = {421--436},
volume = {34}
}

@article{Schot1992,
  title = {{Eighty years of Sommerfeld's radiation condition}},
  year = {1992},
  journal = {Historia Mathematica},
  author = {Schot, Steven H.},
  number = {4},
  pages = {385--401},
  volume = {19}
}

@inproceedings{Royer2020,
  title = {{GmshFEM: an efficient finite element library based on Gmsh}},
  year = {2020},
  booktitle = {14th World Congress on Computational Mechanics (WCCM)},
  author = {Royer, Anthony and Eric, B and Geuzaine, Christophe},
  number = {July},
  pages = {19--24}
}
```

Trust region Newton method

REFERENCES

- Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: *GEOPHYSICS*, **60**, 1457–1473.
- Fan, J., J. Pan, and H. Song, 2016, A Retrospective Trust Region Algorithm with Trust Region Converging to Zero: *Journal of Computational Mathematics*, **34**, 421–436.
- Métivier, L., R. Brossier, J. Virieux, and S. Operto, 2013, Full Waveform Inversion and the Truncated Newton Method: *SIAM Journal on Scientific Computing*, **35**, B401–B437.
- Mulder, W., and R.-E. Plessix, 2008, Exploring some issues in acoustic full waveform inversion: *Geophysical Prospecting*, **56**, 827–841.
- Nocedal, J., S. J. Wright, and S. M. Robinson, 2006, *Numerical Optimization*: Springer New York. Springer Series in Operations Research and Financial Engineering.
- Pan, W., K. A. Innanen, and W. Liao, 2017, Accelerating Hessian-free Gauss-Newton full-waveform inversion via l-BFGS preconditioned conjugate-gradient algorithm: *GEOPHYSICS*, **82**, R49–R64.
- Pratt, R. G., and C. Shin, 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: *Geophysical Journal International*, **133**, 341–362.
- Royer, A., B. Eric, and C. Geuzaine, 2020, GmshFEM: an efficient finite element library based on Gmsh: 14th World Congress on Computational Mechanics (WCCM), 19–24.
- Schot, S. H., 1992, Eighty years of Sommerfeld’s radiation condition: *Historia Mathematica*, **19**, 385–401.
- Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: *GEOPHYSICS*, **69**, 231–248.
- Steihaug, T., 1983, The Conjugate Gradient Method and Trust Regions in Large Scale Optimization: *SIAM Journal on Numerical Analysis*, **20**, 626–637.
- Versteeg, R., 1994, The Marmousi experience: Velocity model determination on a synthetic complex data set: *The Leading Edge*, **13**, 927–936.
- Zuberi, M. A., and R. G. Pratt, 2017, Mitigating nonlinearity in full waveform inversion using scaled-Sobolev preconditioning: *Geophysical Journal International*, **213**, 706–725.