

# Fast Solution of Time-Harmonic Wave Equation for Full-Waveform Inversion

Rafael Lago, Art Petrenko, Zhilong Fang, Felix Herrmann

Department of Earth and Ocean sciences, University of British Columbia, Vancouver, BC, Canada

January 20, 2014

## Abstract

For many full-waveform inversion techniques, the most computationally intensive step is the computation of a numerical solution for the wave equation on every iteration. In the frequency domain approach, this requires the solution of very large, complex, sparse, ill-conditioned linear systems. In this extended abstract we bring out attention specifically to CGMN method for solving PDEs, known for being flexible (i.e. it is able to treat equally acoustic data as well as visco-elastic or more complex scenarios) efficient with respect both to memory and computation time, and controllable accuracy of the final approximation. We propose an improvement for the known CGMN method by imposing a minimal residual condition, which incurs in one extra model vector storage. The resulting algorithm called CRMN enjoys several interesting properties as monotonically nonincreasing behaviour of the norm of the residual and minimal residual, guaranteeing optimal convergence for the relative residual criterion. We discuss numerical experiments both in an isolated PDE solve and also within the inversion procedure, showing that in a realistic scenario we can expect a speedup around 25% when using CRMN rather than CGMN.

## Introduction

Computing a numerical approximation for PDEs is essential in several applications in engineering and science. In seismic exploration, for some strategies of full-waveform inversion, several PDEs associated with wave propagation phenomena must be solved every iteration of an inversion procedure. Solving the wave equation for the geophysical setting is particularly challenging, as not only is the problem very large but also numerically unstable. Besides the traditional time-stepping, several methods have been proposed in the literature for frequency-domain solutions. The discretization of the wave equation in the frequency domain yields a very large, sparse unsymmetric and non-Hermitian complex linear system of the type  $Ax = b$  with thousands of right hand sides.

Amongst the methods for solving the linear system arising from the frequency domain approach, we highlight CGMN (Gordon and Gordon (2008)), a Krylov iterative solver using double Kaczmarz sweeps as a preconditioner to symmetrize the problem. This allows us to use methods such as *conjugate gradient* (CG for short) due to Hestenes and Stiefel (1952) for computing an approximate solution. The resulting algorithm is very simple, is easily parallelizable (c.f. CARPCG due to Gordon and Gordon (2010)), should be suitable for any PDE (i.e. acoustic wave equation, visco-elastic wave equation, etc) and allows a controllable accuracy, a key property to be exploited when using Frugal FWI method (c.f. Herrmann et al. (2013)).

In this extended abstract we propose an improvement to the traditional CGMN algorithm, by choosing an alternative for CG known as *conjugate residual* (CR for short) due to Stiefel (1955). It imposes a minimal residual condition resulting in a faster convergence rate and a monotonic behaviour of the approximate solutions (i.e. the approximate solution can only “improve”). We show small 3D numerical experiments illustrating that the resulting algorithm which we call CRMN can provide a significant speedup when compared to CGMN. We also analyse and compare the behaviour of the inversion when using Frugal FWI with CGMN and CRMN to solve the PDEs, and show that an equivalent solution is computed by CRMN with considerably less computational effort.

## Conjugate Residual versus Conjugate Gradient

When solving the system  $Ax = b$  for a symmetric positive definite matrix  $A$ , perhaps the most popular Krylov subspace method is the (preconditioned) CG. In addition to its simplicity, preconditioned CG is among the Krylov methods with the lowest memory requirements for symmetric positive definite problems, requiring three model vectors in addition to the approximate solution  $x$  and the right hand side  $b$ . Each iteration  $j$ , CG computes the best approximate solution  $x_j \in \{b, Ab, A^2b, \dots, A^{j-1}b\}$  for  $Ax = b$  such that the *energy norm of the error* defined as

$$\|e_j\|_A = \sqrt{(x_j - A^{-1}b)^H A (x_j - A^{-1}b)}$$

is minimal. We refer to Shewchuk (1994) for more details on the properties of CG.

Unfortunately the value of  $A^{-1}b$  is unknown, and therefore  $\|e_j\|_A$  cannot be effectively computed, thus requiring another criterion to estimate “how good”  $x_j$  is as an approximation to the linear system. Some recommended criteria are the *backward error* and the *componentwise backward error* (c.f. Higham (1996) for more details). Another popular criterion, due to its simplicity is the so called relative residual which we define as

$$\|r_j\|_b = \frac{\|b - Ax_j\|_2}{\|b\|_2}$$

which is cheaply available at every iteration. We say then that  $x_j$  is a “good approximation” whenever  $\|r_j\|_b < \varepsilon$ , for some constant tolerance  $\varepsilon$ . One drawback of CG is that  $\|r_j\|_b$  does not possess a nonincreasing behaviour, that is, it may happen that  $\|r_j\|_b < \|r_{j+1}\|_b$ . This is undesirable, especially

when we are interested in performing a fixed number of CG iterations, rather than using a convergence tolerance  $\epsilon$ .

To solve this issue, we propose the use of CR in place of CG. As shown by Eiermann and Ernst (2001), CR can be derived from CG by simply performing a change in the inner product being used. The resulting method does not present the property of minimizing the energy norm of the error, but instead, CR computes  $x_j \in \text{span}\{b, Ab, A^2b, \dots, A^{j-1}b\}$  such that the norm of the residual  $b - Ax_j$  is minimized instead. This property turns out to be very convenient as it guarantees that the approximate solution can only “improve” (for the relative residual as convergence criterion), but it comes with the cost of one extra model vector, totaling four model vector storage required by CR.

The choice of using CR not only induces a nonincreasing behaviour of the residual, but in fact, it can speed up the convergence for the relative residual criterion considerably. As shown by Eiermann and Ernst (2001), the norm of the relative residual of CR is always smaller than or equal to that of CG when applied for the same problem for the same number of iterations. We recommend Fong and Saunders (2012) for a deep comparison of CG, CR and MINRES using several convergence criteria.

## Numerical Experiments

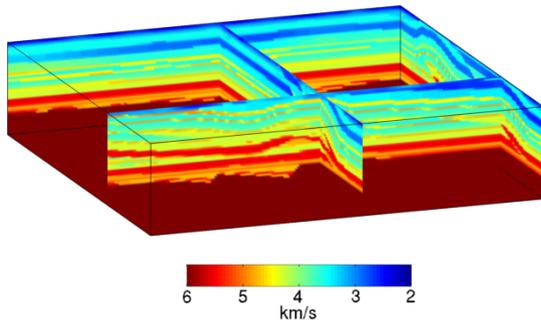
In this section we compare the behaviour of CGMN and the new method we propose which we call here CRMN, consisting of the CR algorithm preconditioned by double Kaczmarz sweep in the same fashion as in CGMN.

**True error in CRMN:** CG is known to usually produce a smaller norm of true error (that is,  $\|x_j - A^{-1}b\|_2$ ) than CR for the same number of iterations. Since CR is supposed to converge in fewer iterations, the difference between the final error of CG and the final error of CR could be quite significant. To verify that, we run a small 3D toy problem associated with the forward modelling in full-waveform inversion in the frequency domain. It consists of solving the Helmholtz equation for a subsample of the SEG/EAGE Overthrust velocity model, with a grid spacing of 100 m (depicted in Figure 1(a)). We discretize the Helmholtz equation using the staggered grid 27 points stencil due to Operto et al. (2007) using PML boundary conditions, obtaining thus the coefficients matrix  $A$ . Next we impose a known (unphysical) solution  $x_*$ , setting all its entries to one, and forming  $b$  by multiplying  $Ax_*$ . The objective is to use CGMN and CRMN to compute an approximated wavefield  $x_j$  and compare it with  $x_*$ .

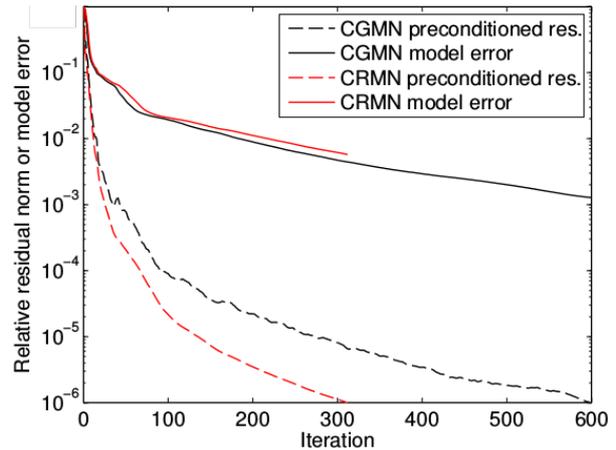
Figure 1(b) shows the norm of the true error for every iteration (scaled by  $\|x_*\|_2$ ), as well as the norm of the preconditioned relative residual for every iteration. To be able to estimate the behaviour after several iterations, we set the convergence criterion to  $\|r_j\|_b < 10^{-6}$  where  $r_j$  is the preconditioned residual.<sup>1</sup>

Figure 1(b) shows that during the computation of the approximated wavefield, the norm of the error of CRMN is marginally larger than that of CGMN, whereas the norm of the preconditioned residual decays considerably faster for CRMN, allowing CRMN to converge in 311 iterations against 599 iterations for CGMN. However this particular example shows that since CGMN performs more iterations it decreases the error in the approximate wavefield further; the norm of the scaled true error is  $1.3 \times 10^{-3}$  for CGMN and  $5.8 \times 10^{-3}$  for CRMN. Although the the final norm of the error of CRMN is less than one order of magnitude larger than that of CGMN, it raises the question on how this larger error affects the gradient computation when using quasi-Newton methods in a full-waveform inversion context. The computation of the gradient is an essential step for these methods, as it is used to compute the descent direction, or in other words, how to update the model estimate for the next inversion iteration. If CRMN allows too large an error when computing the gradient of the objective function we might be choosing a very poor direction.

<sup>1</sup>When the left preconditioner is used as in CGMN, the quantity we control is in fact  $M^{-1}(b - Ax_j)$  rather than the true residual



(a) SEG/EAGE Overthrust velocity model used in this experiment, subsampled to  $47 \times 201 \times 201$ , with a gridspacing  $h = 100$  m



(b) Norm of the true error and preconditioned relative residual for both CGMN and CRMN

**Figure 1** Behaviour of CRMN and CGMN for a small 3D toy problem where the exact solution is imposed. The wavefield is computed for 3 Hz, with a preconditioned relative residual tolerance of  $\epsilon = 10^{-6}$

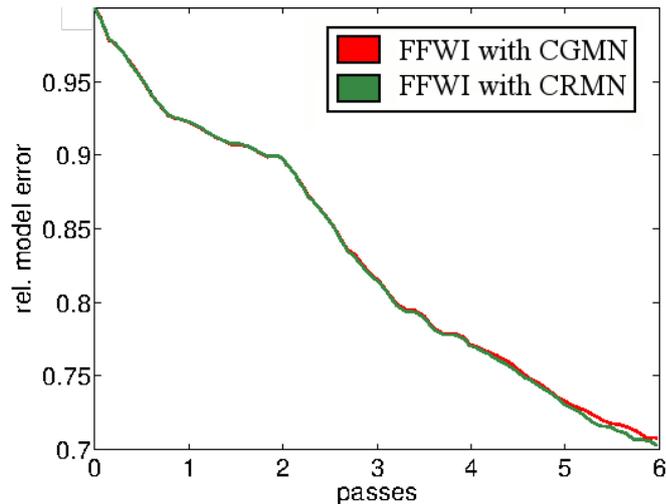
**Frugal FWI inversion:** We next show a numerical experiment in which this is not the case and that CRMN does compute a satisfactory gradient direction in considerably fewer iterations. We attempt to invert the data generated with the iWave software package (Terentyev et al. (2014)) for a  $5 \times 5 \times 2.5$  km<sup>3</sup> central part of the SEG/EAGE Overthrust synthetic velocity model. The data generated contains 121 sources and 2601 receivers regularly spaced, and we invert for three frequency slices, 4Hz, 6Hz and 8Hz.

We use frugal FWI (FFWI for short) with adaptive sample-size and simulation accuracy. CGMNIt consists of a l-BFGS based method, in which each PDE solve<sup>2</sup> is computed with a chosen (loose) tolerance  $\epsilon$  which slowly decreases along the inversion procedure. In addition to that, we use a variable sample size, meaning that we do not evaluate all the 121 shots every iteration, but instead consider a small subsample of shots and allow this subsample size to increase as the iversion progresses. We refer the reader to Herrmann et al. (2013) for a more insightful discussion on these techniques and a comparison with other existing methods.

We provide the solver with a budget of 484 PDE-solves for each frequency slice, after which the algorithm is forced to stop, and we set the initial tolerance for PDE solves to  $\epsilon = 10^{-2}$ . We run the inversion once using CGMN for computing the approximate PDE solves, and once for CRMN computing the approximate PDE solves. Figure 2 shows the evolution of the model error along the inversion procedure for both CGMN and CRMN. We can see that the model error for CRMN is as a matter of fact equivalent to that of CGMN, except for a neglectable discrepancy at the very few PDE evaluations. This result suggests that the more accurate solution of CGMN with respect to true error bears no significant benefits for the inversion.

Table 1 shows the number of CGMN iterations used during the whole inversion process for each frequency slice, the same quantity for CRMN and the respective speedup obtained by using CRMN. Notice that here the gain is considerably inferior to the one illustrated in Figure 1(b) since the initial tolerance is set to  $\epsilon = 10^{-2}$  (versus  $\epsilon = 10^{-6}$  for Figure 1(b)). We expect that as more accurate PDE solves are required, the discrepancy between the two methods tend to become more and more evident. Nonetheless, especially for 8 Hz the speedup is considerable.

<sup>2</sup>That comprises the misfit, adjoint wavefield and gradient computation



**Figure 2** Model error after each pass of FFWI for all frequencies. One pass equals 242 PDE solves.

	FFWI with CGMN	FFWI with CRMN	Speedup
4 Hz	23,403	19,846	18%
6 Hz	30,189	24,387	24%
8 Hz	34,724	26,265	32%
Total	88,316	70,498	25%

**Table 1** Total number of iterations of CGMN and CRMN during the inversion for each frequency slice

## Conclusion

We proposed an improvement for the known CGMN method, by replacing the traditional conjugate gradients method by the conjugate residual method. The new proposed algorithm called CRMN presents minimal residual properties and is able to compute an approximate solution satisfying the relative residual criterion for each PDE solve in about 25% fewer iterations than the traditional CGMN, at the expense of one extra model vector storage. We observed that CGMN also allows a smaller true error in the proposed experiments, but that in the context of inversion, this smaller error produces no benefit in the final model approximation. Experiments using frugal FWI show that the speedup provided by CRMN is substantial, encouraging further investigation for larger problems, higher frequency, more realistic settings (e.g. viscoelastic larger dataset) and massively parallel experiments.

## References

- Eiermann, M. and Ernst, O.G. [2001] Geometric aspects in the theory of Krylov subspace methods. *Acta Numerica*, **10**(10), 251–312.
- Fong, D. and Saunders, M. [2012] CG versus MINRES: An empirical comparison. *SQU Journal for Science*, **17**:1, 44–62.
- Gordon, D. and Gordon, R. [2008] CGMN revisited: Robust and efficient solution of stiff linear systems derived from elliptic partial differential equations. *ACM Trans. Math. Softw.*, **35**(3), 18:1–18:27, ISSN 0098-3500, doi: 10.1145/1391989.1391991.
- Gordon, D. and Gordon, R. [2010] CARP-CG: A robust and efficient parallel solver for linear systems, applied to strongly convection dominated pdes. *Parallel Computing*, **36**(9), 495–515.
- Herrmann, F.J. et al. [2013] Frugal full-waveform inversion: from theory to a practical algorithm. *The Leading Edge*, **32**(9), 1082–1092, doi:http://dx.doi.org/10.1190/tle32091082.1.
- Hestenes, M.R. and Stiefel, E. [1952] Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, **49**(6), 409–436.
- Higham, N.J. [1996] *Accuracy and Stability of Numerical Algorithms*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, ISBN 0898713552.
- Operto, S., Virieux, J., Amestoy, P., L'Excellent, J.Y., Giraud, L. and Ali, H.B.H. [2007] 3d finite-difference frequency-domain modeling of visco-acoustic wave propagation using a massively parallel direct solver: A feasibility study. *Geophysics*, **72**(5), SM195–SM211, doi:10.1190/1.2759835.
- Shewchuk, J.R. [1994] An introduction to the conjugate gradient method without the agonizing pain. Tech. rep.
- Stiefel, E. [1955] Relaxationsmethoden bester strategie zur losung linearer gleichungssystem. *Commentarii Mathematici Helvetici*, **29**(1), 157–179.
- Terentyev, I.S., Vdovina, T., Symes, W.W., Wang, X. and Sun., D. [2014] iWave: a framework for wave simulation. <http://www.trip.caam.rice.edu/software/iwave/doc/html/index.html>, [Online; accessed 14-January-2014].