2021 EAGE ANNUAL
82ND CONFERENCE & EXHIBITION
AMSTERDAM | THE NETHERLANDS

# A MATRIX-FREE REFORMULATION OF THE MULTI-PARAMETER DESCENT AND CONJUGATE-GRADIENT METHOD FOR ISOTROPIC ELASTIC ITERATIVE REVERSE-TIME MIGRATION

W. Mulder[1,2]

[1] Shell Global Solutions International BV; [2] Delft University of Technology

## Summary

Multi-parameter inversion of linear systems appears in many problems. The focus here is on isotropic elastic iterative reverse-time migration for three position-dependent subsurface model parameters, which amounts to data fitting of processed seismic data with synthetics from the Born approximation of the elastic wave equation. In that case, the matrix of the linear system is the hessian. As it is impractical to form, a matrix-free formulation is needed, which is readily derived for the gradient descent method. For single-parameter inversion, the conjugate-gradient (CG) method is generally more efficient than simple descent. However, the multiple-parameter CG method has a significantly higher cost than the descent method. Here, first a matrix-free data-domain reformulation is derived. Then, its performance is compared to the simple descent method to see of its faster convergence justifies the higher cost. A comparison on a marine 2-D toy problem with a salt body and sea-bottom receivers shows that the multiple-parameter descent method wins in terms of efficiency if the number of iterations is limited and that the single-parameter CG method is even faster.

# A matrix-free reformulation of the multi-parameter descent and conjugate-gradient method for isotropic elastic iterative reverse-time migration

## Introduction

In multi-parameter inversion, the unknown model parameters are divided into subsets. Examples are impedance and density for the inversion of seismic data with the acoustic wave equation or P- and S-impedance and density for the isotropic elastic case. Here, the focus is on iterative reverse-time migration, which amounts to inversion for three sets of subsurface parameters, each dependent on subsurface position, given seismic data and a sufficiently accurate background model. The governing equations are the Born approximation of the isotropic elastic system of wave equations. Issues such as the best choice of parametrization and the limited validity of the Born approximation will be ignored. Instead, two methods for linear inversion are considered: multi-parameter descent and its conjugate-gradient (CG) version. For the latter, Brezinski (1999) showed that the latter requires multiple search directions, equal to the number of parameter sets. This increases the cost of the method.

In the single-parameter case, the CG method is usually more efficient than simple descent. To see if this is also true in the multi-parameter case, a test on a 2-D marine toy problem is performed. Because it is usually impractical to compute the hessian of the inverse problem, a matrix-free reformulation of both methods is derived in the next section.

## Theory

Data fitting with least-squares migration amounts to minimization of a cost functional $J = \frac{1}{2}\|d\|_2^2$, where the $L_2$-norm measures the data error $d = u^{\text{obs}} - u$ between observed data $u^{\text{obs}}$ and modelled data $u$. The last depends on $n_c$ model components $m^k$ through $u = \sum_{k=1}^{n_c} F^k m^k$. The linear operator $F$ describes forward modelling with the Born approximation of the wave equation and can be split into the contributions of the $n_c$ position-dependent model parameters $m^k$. In the acoustic case, for instance, the two model components can be chosen as perturbations of the impedance and density. In the elastic case, one may opt for perturbations in P- and S-impedance and density or other combinations thereof.

Migration of the data error produces an update of the model components: $r = F^\mathsf{T} d$, where the transpose of the Born modelling operator is involved. The result can be formally split into contributions to the model components according to $r^k = (F^k)^\mathsf{T} d$. In practice, these are computed simultaneously via the correlation of forward and reverse-time wavefields, without forming the matrix $F$ or $F^\mathsf{T}$.

The minimum of the error functional $J$ is obtained for a zero gradient $g = \nabla_m J = -F^\mathsf{T}(u^{\text{obs}} - Fm) = Am - b$, where the hessian $A$ has blocks $A_{k,\ell} = (F^k)^\mathsf{T} F^\ell$, corresponding to the model components, and the initial residual $b^k = (F^k)^\mathsf{T} u^{\text{obs}}$ is the migration image for component $k$. For small problems, the application of a singular value decomposition, which takes care of the large null-space of $A$, provides a direct solution of $g = 0$. For larger problems, an iterative method is often more suitable. The simplest approach is a descent method. Let the residual be defined as minus the gradient for the current model, with components $r^k = (F^k)^\mathsf{T} d$, and take that equal to the search direction: $p^k = r^k$. The optimal step lengths $\lambda_k$ for each component follow from the minimization of the functional

$$\bar{J} = \tfrac{1}{2} \left\| u^{\text{obs}} - \sum_{k=1}^{n_c} F^k \left[ m^k + \lambda_k p^k \right] \right\|^2 . \tag{1}$$

Setting the derivatives w.r.t. $\lambda_k$ to zero leads to

$$C_{k,\ell}\lambda_\ell = \tau_k, \quad C_{k,\ell} = v^k \cdot v^\ell, \quad \tau_k = v^k \cdot (u^{\text{obs}} - u) = v^k \cdot d, \tag{2}$$

where $v^k = F^k p^k$ are the modelled data for a search-direction component $p^k$. The dot denotes a scalar product over all data. The functional after the model update, with $d = u^{\text{obs}} - u$ still the previous data error, becomes

$$\bar{J}_{\min} = \tfrac{1}{2}\|d\|_2^2 + \tfrac{1}{2}\boldsymbol{\lambda}^\mathsf{T} C \boldsymbol{\lambda} - \boldsymbol{\tau}^\mathsf{T}\boldsymbol{\lambda} = \tfrac{1}{2}\|d\|_2^2 - \tfrac{1}{2}\boldsymbol{\tau}^\mathsf{T}\boldsymbol{\lambda} = \tfrac{1}{2}\|d\|_2^2 - \tfrac{1}{2}\boldsymbol{\lambda}^\mathsf{T} C \boldsymbol{\lambda}. \tag{3}$$

The last term on the right-side can be expressed as $\frac{1}{2}\boldsymbol{\lambda}^\mathsf{T} C\boldsymbol{\lambda} = \frac{1}{2}\sum_{k,\ell}(\lambda_k v^k)\cdot(\lambda_\ell v^\ell) = \frac{1}{2}q\cdot q$ with $q = \sum_{k=1}^{n_c}\lambda_k v^k$, showing that $\bar{J}_{\min} \le J$, i.e., the functional will not increase. Note that the right-hand side $\boldsymbol{\tau}$ of the small $n_c \times n_c$ problem in equation (2) can be expressed in terms of the residual $r^k = b^k - \sum_{\ell=1}^{n_c} A_{k,\ell} m^\ell = (F^k)^\mathsf{T} d$ as $\tau_k = (v^k)^\mathsf{T} d = d^\mathsf{T} F^k (F^k)^\mathsf{T} d = (r^k)^\mathsf{T} r^k$. The method is summarized as Algorithm 1. The matrices $F$ and $F^\mathsf{T}$ are symbolic references to the modelling and migration operators, respectively, and are not formed in practice. Apart from the usual convergence criteria, the singularity of $C$, for instance indicated by its condition number, can signal that the method has reached the noise level and should be stopped.

**Algorithm 1:** Matrix-free descent method for $n_c$ model components.

*initialize*: $j = 1, \quad d = u^{\text{obs}} - \sum_{k=1}^{n_c} F^k m^k, \quad E = \frac{1}{2}\sum |u^{\text{obs}}|^2$
*iterate until convergence or a limit on* $j$
  $J = \frac{1}{2}\sum |d|^2$, check convergence on $J/E$
  $r = F^\mathsf{T} d$, check convergence on $r \cdot r$
  $v^\ell = F r^\ell, \quad C_{k,\ell} = v^k \cdot v^\ell, \quad \tau^k = v^k \cdot d = r^k \cdot r^k$, solve $C\boldsymbol{\lambda} = \boldsymbol{\tau}$
  $d := d - \sum_{k=1}^{n_c} \lambda^k v^k, \quad m^k := m^k + \lambda^k r^k, \quad j := j + 1$

As is well known for single-component inversion, the conjugate-gradient (CG) method is more efficient that the descent method, by imposing orthogonality of the residuals and $A$-orthogonality of the search direction (Hestenes and Stiefel, 1952). The generalization to multiple sets of parameters is not straightforward and requires $n_c$ instead of one search direction (Brezinski, 1999).

**Algorithm 2:** Preconditioned multi-parameter CG method.

*initialize*
  $r = b - Am, \quad R = \mathcal{P} r, \quad P = Z = MR, \quad T = Z^\mathsf{T} R$
*iterate until convergence*
  $C = P^\mathsf{T} A P$, solve $C\Lambda = T$
  $Q = AP\Lambda, \quad R := R - Q, \quad Z := Z - MQ, \quad r = Re, \quad m := m + P\Lambda e$
  $T^{\text{old}} = T, \quad T = Z^\mathsf{T} R, \quad$ solve $T^{\text{old}} B = T, \quad P := Z + PB$

**Algorithm 3:** Matrix-free preconditioned multi-parameter CG method.

*initialize*: $j = 1, \quad d = u^{\text{obs}} - \sum_{k=1}^{n_c} F^k m^k, \quad E = \frac{1}{2}\sum |u^{\text{obs}}|^2$
*iterate until convergence or a limit on* $j$
  $J = \frac{1}{2}\sum |d|^2$, check convergence on $J/E$
  if $j = 1$
    $b = F^\mathsf{T} d, \quad R = \mathcal{P} b, \quad P = Z = MR, \quad T = Z^\mathsf{T} R$
  else
    $Q^\ell = F^\mathsf{T} v^\ell, \quad R := R - Q, \quad Z := Z - MQ$
    $T^{\text{old}} = T, \quad T = Z^\mathsf{T} R, \quad B = (T^{\text{old}})^{-1} T, \quad P := Z + PB$
  check convergence on $r = Re$ or $T$
  $u^\ell = \sum_{k=1}^{n_c} F^k P_k^\ell, \quad C_{k,\ell} = u^k \cdot u^\ell$, solve $C\Lambda = T, \quad v^k = \sum_{\ell=1}^{n_c} u^\ell \Lambda_{\ell,k}$
  $d := d - \sum_{k=1}^{n_c} v^k, \quad m := m + P\Lambda e, \quad j := j + 1$

A partitioning operator $\mathcal{P}$ applied to $r$ scatters the $n_c$ components of $r$ over $n_c$ columns, filling the rest with zeros:

$$R = \mathcal{P} r = \begin{pmatrix} r^1 & 0 & \dots \\ 0 & r^2 & \dots \\ \vdots & & \end{pmatrix} \tag{4}$$

The zeros should be read as column vectors with a length equal to the total number of points in the

migration volume. Its effect can be undone by $r = Re$ with $e = (1, 1, \ldots)^{\mathsf{T}}$, which amounts to taking row sums of $\mathcal{P}r$. Note that the partitioning (4) into one component per column is not unique if only $r = Re$ is required. Algorithm 2 starts with a search direction equal to the partitioned residuals $R$ or to $MR$, given a symmetric preconditioner $M$ that is a low-cost approximation of the inverse of the hessian $A$, for instance, the estimated block diagonal of the hessian. Note that, in general, $M\mathcal{P}R \neq \mathcal{P}Mr$. Without preconditioner, $M = I$ and $Z = P$. The reformulation of Algorithm 2 without reference to the matrix $A$ is aided by the matrix-free version of the descent method listed as Algorithm 1. The result is Algorithm 3. $Q^\ell$ should be read as column $\ell$ and $P_k^\ell$ as the row-subset of column $\ell$ of $P$ that corresponds to model component $k$.

## Results

To set the scene, Figure 1 displays convergence results for the single- and multi-component descent and CG methods when applied to a linear problem with 18 data values and 3 solution components, each having 15 unknowns. The solution consists of random numbers, as does the modelling operator. The latter has a different scale factor for each solution component and the preconditioner is a diagonal matrix that divides by the squares of those three scale factors. The multi-component methods require less iterations than the the single-component versions. The same holds for the conjugate-gradient methods compared to the descent methods. In the early stages, however, the CG and descent methods are close. Also, the multi-component CG method requires more operations than the single-component version, making the latter more efficient for the current problem if the functional does not have to decrease that much.

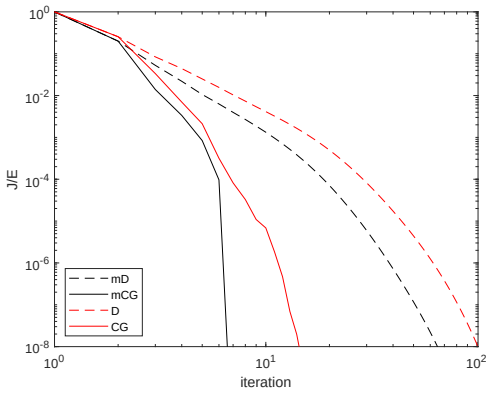
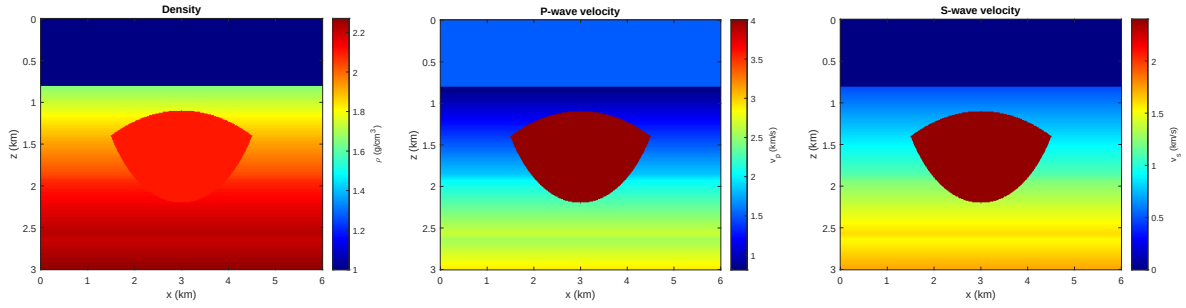
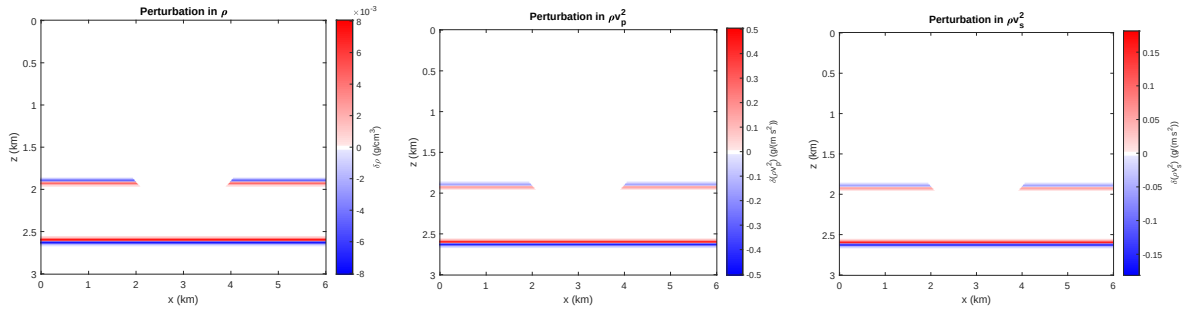



***Figure 1*** *Convergence on a log-log scale for the single- (D) and multi-component descent (mD) and the single- (CG) and multi-component conjugate-gradient (mCG) methods.*




(a) (b) (c)

***Figure 2*** *Background model for (a) density, (b) P-wave velocity and (c) S-wave velocity.*

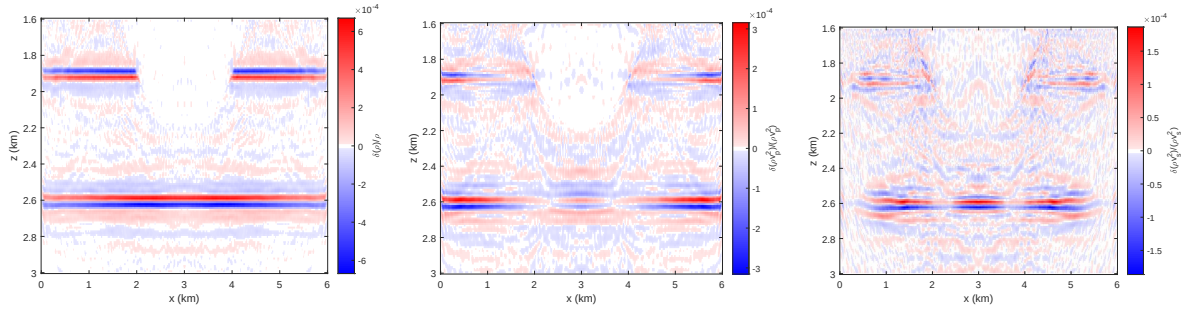



(a) (b) (c)

***Figure 3*** *Scatterers represented as perturbations of (a) density, (b) $\lambda + 2\mu = \rho v_p^2$ and (c) $\mu = \rho v_s^2$.*

To assess the performance of the methods on a less abstract example, an isotropic elastic 2-D marine

toy problem with a salt body is considered on a domain 6 km wide and 3 km deep. The 57 receivers are placed 100 m apart on the sea bottom, at a depth of 800 m between $x = 200$ and 5800 m and only the vertical component is used. The 59 explosive sources are fired at 10-m depth between $x = 100$ and 5900 m at a 100-m interval. Figure 2 shows the background model. A free-surface boundary condition is imposed, contrary to the common approach in migration. The sediment part of the model is smoothed and two reflectors are added. The salt body is not smoothed, nor is the sea bottom. Figure 3 shows the perturbations that act as scatterers in the Born approximation. The chosen representation is as a perturbation in density $\rho$, $\lambda + 2\mu = \rho v_p^2$ and $\mu = \rho v_s^2$, where $v_p$ and $v_s$ are the P- and S-wave velocities, respectively, and $\lambda$ and $\mu$ the Lamé parameters. Other parametrizations may be better, but these are the natural ones from an implementation point of view. A time-domain finite-difference code generated 3.5 seconds of 'observed' data for 10-Hz Ricker wavelet, although in 2D, a frequency-domain code would have been more efficient (Mulder and Plessix, 2002). The preconditioner for the inversion consisted in a spatial low-cut filter.




(a) (b) (c)

***Figure 4*** *Reconstructed scattering model for (a) $\delta(\rho)/\rho$, (b) $\delta(\rho v_p^2)/(\rho v_p^2)$ and (c) $\delta(\rho v_s^2)/(\rho v_s^2)$.*

Figure 4 displays the reconstructed scattering model. Figure 5 compares the convergence of the descent and the CG methods. The CG methods produces smaller residuals than the descent methods after a few iterations and the same is seen for the multi-component versions compared to the single-component versions. In terms of compute cost to reach a given level of convergence $J/E$, however, the two descent methods come close, whereas the single-component CG method performs better than the others. The multi-component CG method is initially the slowest and only starts to become more efficient than the descent methods after about 20 iterations in this example.

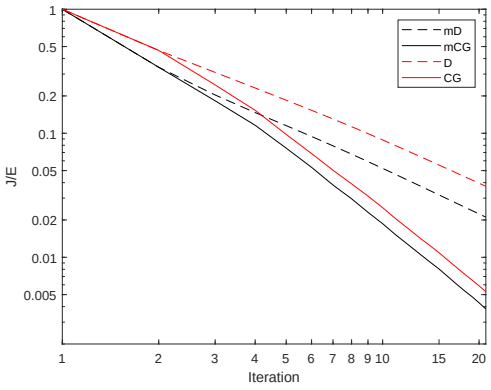



***Figure 5*** *Convergence on a log-log scale for the single- and multicomponent descent (D, mD) and CG (CG, mCG) methods.*

## Conclusions

Application of the multi-parameter descent and conjugate-gradient methods to elastic reverse-time migration shows that latter convergence faster. In a 2-D toy problem, the gain in convergence speed only started to compensate for its higher cost after 10 iterations and the single-component conjugate-gradient method was the most efficient.

## References

Brezinski, C. [1999] Multiparameter descent methods. *Linear Algebra and its Applications*, **296**(1), 113–141.

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.

Mulder, W.A. and Plessix, R.É. [2002] Time- versus frequency-domain modelling of seismic wave propagation. In: *Conference Proceedings, 64th EAGE Conference & Exhibition*.