Geophys. J. Int. (1990) **103**, 755–756

Comment on 'Comparison of iterative back-projection inversion and generalized inversion without blocks: case studies in attenuation tomography' by P. Ho-Liu, J.-P. Montagner and H. Kanamori

Jeannot Trampert
Institut de Physique du Globe, 5 rue René Descartes, F-67084 Strasbourg Cédex, France

Accepted 1990 July 13. Received 1990 March 23

In a recent paper, Ho-Liu, Montagner & Kanamori (1989) compared two inversion techniques: the iterative back-projection inversion, also known as 'Simultaneous Iterative Reconstruction Technique', or SIRT for short (e.g. Van der Sluis and Van der Vorst 1987) and the generalized least-square inversion 'without blocks' (Tarantola & Valette 1982). Ho-Liu *et al.* (1989) were the first to mention a possible connection between both algorithms, but their paper presents two points which need clarification.

(1) They derive SIRT from the generalized least-square inversion under certain assumptions. However, their approach presents a conceptual mistake which has a detrimental effect on the interpretation of the parameter μ in the SIRT algorithm. A correct demonstration showing the relationship between SIRT and the generalized least-square solution is given by Trampert & Lévêque (1990).

(2) Then, they argue that no formal error and resolution estimation is possible with SIRT. Trampert & Lévêque (1990) showed that any linear inversion algorithm can be used to compute a resolution matrix in the sense of Wiggins (1972). If furthermore, the algorithm used is a special case of the generalized least-square inversion (which most algorithms are), the error estimation is very easy.

Let us now analyse these two points in more detail.

To derive SIRT from the generalized least-square inversion, Ho-Liu *et al.* (1989) start from equation (25) of Tarantola and Valette (1982) which says that

$$\mathbf{p}_{k+1} = \mathbf{p}_k + (\mathbf{L}_k^T\mathbf{C}_{d0}^{-1}\mathbf{L}_k + \mathbf{C}_{p0}^{-1})^{-1} \times \{\mathbf{L}_k^T\mathbf{C}_{d0}^{-1}[\mathbf{d}0 - L(\mathbf{p}_k)] - \mathbf{C}_{p0}^{-1}(\mathbf{p}_k - \mathbf{p}_0)\} \quad (1)$$

where $\mathbf{L}$ is the matrix of partial derivatives of a function L, $\mathbf{d}_0$ the data vector, $\mathbf{p}_k$ the model estimation at iteration k and $\mathbf{p}_0$ the starting model. $\mathbf{C}_{d0}$ and $\mathbf{C}_{p0}$ are the *a priori* covariance matrices for the data and the starting model respectively.

With certain number of assumptions they obtain

$$\mathbf{p}_{k+1} \sim \mathbf{p}_k + \mathbf{A}^{-1}\mathbf{L}^T\mathbf{B}^{-1}(\mathbf{d}_0 - \mathbf{L}\mathbf{p}_k) \quad (2)$$

where $\mathbf{A} = \mathrm{diag}\,(\mu + \sum_i |L_{ij}|)$, the cumulated ray length per block and $\mathbf{B} = \mathrm{diag}\,(\sum_j |L_{ij}|)$ the total ray length. We will discuss the interpretation of the constant μ below. Expression (2) represents the classical SIRT algorithm.

As they derive equation (2) from equation (1), the number k counting the iterations is the same in both expressions. This is precisely where the problem occurs. Equation (1) solves a non-linear problem $\mathbf{d}0 = L(\mathbf{p})$, when the function L is differentiable. This generalized least-square algorithm minimizes the misfit function $\mathbf{d}0 - L(\mathbf{p}_k)$ using a step-by-step linearization technique. Each step represents one iteration. If the problem is linear, this Gauss–Newton method converges in one iteration (Tarantola 1987). On the other hand, (2) solves iteratively a linear problem (Van der Sluis & Van der Vorst 1987). This means that a linear problem, $\mathbf{d}_0 = \mathbf{L}\mathbf{p}$ is solved in one iteration by equation (1) and hence the SIRT algorithm has to converge in one iteration as well. This however is generally impossible. The SIRT solution can be expressed as a geometrical series expansion and requires a certain number of iterations to converge depending on the eigenvalue associated with the model parameter.

The authors thus mixed two different concepts belonging to linear and non-linear iterative problems. But how can they obtain the linear SIRT algorithm from a non-linear approach? The answer is found in their assumptions. We do not want to analyse in detail all assumptions made by Ho-Liu *et al.* (1989); nevertheless, one needs to be mentioned: they assume that the ray paths are short relative to the block size and say that $l_{ij} \sim L_i$ (l_{ij} is the length of ray i in block j and L_i is the total length of ray i). This means that a ray starting in one block never leaves this block, or that the original problem $\mathbf{d}_0 = \mathbf{L}\mathbf{p}$ is already diagonal and its solution is obvious. This assumption is completely unrealistic in tomographic problems.

We have just seen that Ho-Liu *et al.* (1989) made a conceptual mistake in demonstrating that SIRT is a special case of the generalized least-square inversion. As a consequence, their interpretation of μ as the ratio between the data variance and the *a priori* model variance (which would correspond to the damping constant in the classical damped least-square solution) is incorrect. This could considerably affect the choice of μ and the solution in general. To make a correct interpretation of μ, we have to emphasize some general features of SIRT. A detailed discussion of the following is given by Trampert & Lévêque (1990). The convergence speed of SIRT is not uniform, but depends on the eigenvalue associated with the model parameter. The smaller this eigenvalue, the slower the convergence. The big drawback of SIRT, however, is that it uses implicit non-physical *a priori* information to converge. In this context, μ plays a double part: μ is decreasing the eigenvalues and thus the convergence speed, and μ may correct the non-physical *a priori* information, if we take care to introduce an additional constant θ^2. Equation (2) then becomes

$$\mathbf{p}_{k+1} = (1-\theta^2)\mathbf{p}_k + \mathbf{A}^{-1}\mathbf{L}^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{d}_0 - \mathbf{L}\mathbf{p}_k). \tag{3}$$

It is easy to show that the *a priori* model covariance is proportional to $(1/\theta^2)\mathbf{A}^{-1}$. We see that this *a priori* model covariance depends on the theory matrix $\mathbf{L}$ via $\mathbf{A}$. If θ^2 is different from zero, a large value of μ, compared to $\max(\sum_i |L_{ij}|)$, can make this covariance theory independent to the first order. Ho-Liu *et al.* (1989) described the SIRT algorithm for $\theta^2 = 0$, which means that they allow an infinite *a priori* variation to all model parameters, independently of μ. Then, the only effect they produce with μ is to slow down the convergence speed. If they stop the iterations before complete convergence, they think that they are damping the solution, but in fact, they are further away from the final solution. For an infinite number of iterations, μ has no effect whatsoever on the final solution with their algorithm.

The second point we want to clarify concerns the resolution and error analysis.

Following Trampert & Lévêque (1990), we assume that a linear problem $\mathbf{d}_0 = \mathbf{L}\mathbf{p}$ can be solved by an algorithm $\mathbf{H}$. The resolution matrix is defined as (Wiggins 1972)

$$\mathbf{R} = \mathbf{HL}. \tag{4}$$

It is clear that the jth column of $\mathbf{R}$ is $\mathbf{R}_j = \mathbf{HL}_j$, where $\mathbf{L}_j$ is the jth column of the known theory matrix $\mathbf{L}$. The vector $\mathbf{R}_j$ can thus be obtained by applying the inversion algorithm $\mathbf{H}$ to the 'data vector' $\mathbf{L}_j$.

The resolution of SIRT can be derived the same way, where the SIRT algorithm $\mathbf{H}$ is applied as many times as we have unknowns in our system. Furthermore, we know that SIRT is a special case of the generalized least-square solution (Trampert & Lévêque 1990), which means that the error estimation is quite simple. In this case, the *a posteriori* model covariance is given by

$$\mathbf{C}_p = (\mathbf{I} - \mathbf{R})\mathbf{C}_{p0}. \tag{5}$$

Therefore, it is possible to make a formal resolution and error analysis with SIRT, contrary to what has been stated by Ho-Liu *et al.* (1989).

REFERENCES

Ho-Liu, P., Montagner, J.-P. & Kanamori, H., 1989. Comparison of iterative back-projection inversion and generalized inversion without blocks: case studies in attenuation tomography, *Geophys. J.*, **97**, 19–29.

Tarantola, A., 1987. *Inverse Problem Theory*, Elsevier, Amsterdam.

Tarantola, A. & Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. Geophys. Space Phys.*, **20**, 219–232.

Trampert, J. & Lévêque, J.-J., 1990. Simultaneous Iterative Reconstruction Technique: Physical interpretation based on the generalized least squares solution, *J. geophys. Res.*, **95**, 12 553–12 559.

Van der Sluis, A. & Van der Vorst, H. A., 1987. Numerical solution of large, sparse linear algebraic systems arising from tomographic problems, in *Seismic Tomography*, pp. 49–83, ed. Nolet, G. D., Reidel, Dordrecht.

Wiggins, R. A., 1972. The general inverse problem: Implications of surface waves and free oscillations for Earth Structure, *Rev. Geophys. Space Phys.*, **10**, 251–285.