

A time domain derivation of the Kirchhoff migration as the gradient of a data misfit function

Guy Chavent

▶ To cite this version:

Guy Chavent. A time domain derivation of the Kirchhoff migration as the gradient of a data misfit function. [Research Report] RR-1928, INRIA. 1993. inria-00074746

HAL Id: inria-00074746 https://inria.hal.science/inria-00074746

Submitted on 24 May 2006

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.



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

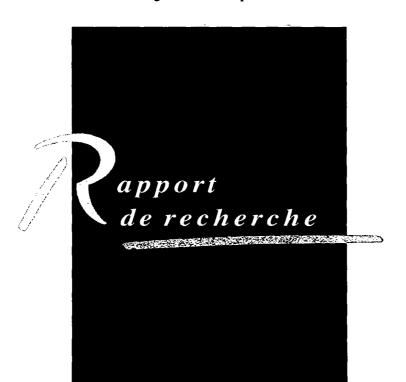
A Time Domain Derivation of the Kirchhoff Migration as the Gradient of a Data Misfit Function

Guy CHAVENT

N° 1928 Juin 1993

PROGRAMME 6 _____

Calcul scientifique, modélisation et logiciels numériques



A Time Domain Derivation of the Kirchhoff Migration as the Gradient of a Data Misfit Function

Les Images Migrées par Kirchhoff sont le Gradient d'une Fonction d'Erreur sur les Données

Guy Chavent *†

May 17, 1993

Résumé

Nous explicitons une approximation WKBJ de l'équation des ondes dans le domaine temps, telle que le gradient du critère d'erreur des moindres carrés redonne les formules classiques de migration de Kirchhoff, à un coefficient multiplicatif positif près. Comme retombées de cette présentation élémentaire de Kirchhoff, nous obtenons le filtrage optimal à effectuer sur les données, et les coefficients à utiliser pour qu'une section migrée par Kirchhoff soit le plus proche possible d'une section inversée. Nous proposons un algorithme itératif de migration de Kirchhoff coopératives, qui illustre les possibilités ouvertes par cette approche de la migration de Kirchhoff comme gradient d'un critère d'erreur.

Abstract

We explicit a forward time domain WKBJ approximation of the wave equation such that the gradient of the associated data misfit function yields the classical Kirchhoff migration formula, up to a positive migration weight. As a by product of this elementary derivation, we find the optimal prefiltering of the data, and a set of optimal migration weights such that the corresponding Kirchhoff migrated sections are as close as possible to inverted sections. We propose a cooperative iterated Kirchhoff migration algorithm which illustrate the possibilities opened by this gradient approach to Kirchhoff migration.

Mots Clés

Migration, imagerie sismique, contrôle optimal, identification, estimation de paramètres.

Keywords

Migration, seismic imaging, optimal control, identification, parameter estimation.

^{*}INRIA, domaine de Voluceau-Rocquencourt, BP 105, 78153 Le Chesnay Cédex, France

[†]CEREMADE, Université Paris Dauphine, 75775 Paris Cédex 16, France

Introduction

Relatively soon after the development of numerical simulation of the wave equation, it was recognized by Lailly (1983) that the corresponding wave equation migrations beared a strong ressemblance with the gradient of the data misfit functional he was used to compute: the adjoint equation of the control theory was nothing but the back propagation of the residual, and both the gradient and the migrated section were obtained by (slightly different) correllation of the forward and backward propagated fields. This recognition was the basis of the saying "migration is the first step of inversion", which is now widely accepted, and opened the way to new usages of the wave equation migration, in particular in iterative migration algorithm, which could then be understood as minimization algorithms.

Surprisingly, nothing similar happened with the - much more widely used - Kirchhoff migration. The reason may be that the Kirchhoff migration has been developed, since the dawn of Geophysics, as an attempt to solve an integral equation relating the diffracted field at a point interior to the earth to the diffracted field at the surface of the earth, and hence to the data. We mention in the references a few papers in this line we have been looking at, but they represent only an infinitesimal part of the geophysical literature on this subject. Approximately at the same time where Lailly linked the wave equation migration to the gradient of the data misfit function, Beylkin (1984, 1985) recognized the strong ties that the Kirchhoff migration had with the inversion of a generalized Radon transform. This allowed a precise mathematical analysis of the resolution capacity of the Kirchhoff migration, but did not give any hint on wether or not Kirchhoff migration was related to the gradient of some data misfit function.

It is this gap which we try to fill in this paper. In order to do this, we have discussed with specialists of the Kirchhoff migration in order to establish the discrete formula which where actually coded in their Kirchhoff migration codes. Then using these formula as a starting point, we have solved an "inverse inverse" problem, i.e. we have searched for the discrete forward model such that the gradient of the associated data misfit functional coïncides with the Kirchhoff formula. The resulting forward model is described in paragraph 3: not surprisingly it includes the WKBJ approximation to the solution of linearized wave equations with point sources located at the seismic source (incident wave field) and at the scattering points (scattered wave field).

Once this forward model and the corresponding data misfit function J are established, we define in paragraph 4 a migrated image as being any descent direction for J at reflectivity r=0. Hence a migrated image can be obtained by multiplying $-\nabla J(0)$ by a positive definite weight matrix (specific of the skills and tastes of the designer of the migration algorithm). Then investigating the relation with the inversion approach, which consists in minimizing J, leads to the conclusion that the best (but unaffordable...) weight matrix for the migration would be the pseudo-inverse of the Hessian of J (in which case the migrated and inverted reflectivities would coı̈ncide).

In paragraphs 5 and 6, we calculate the gradient of the data misfit function J with respect to the reflectivity vector r, and show that, for a proper choice of the (diagonal) weight matrix, the migration formula introduced in paragraph 4 coïncide with well-known Kirchhoff migration formula, like the ones of Docherty (1991), Keho and Beydoun (1988) and Bleistein (1987), provided a proper filtering of the data is chosen.

Then we take advantage, in paragraph 7, of the degree of freedom we have in the choice of the weight matrix, and determine the best diagonal matrix which approximate the pseudo-

inverse of the Hessian of J: using these weights in the migration formula will restore "at best" the amplitude of the migrated reflectors, and bring migration as close as possible to inversion.

We conclude in paragraph 8 by giving an example of application of the fact that the Kirchhoff migration is related to the gradient of the misfit function: we construct a cooperative Kirchhoff migration algorithm, aimed at enhancing the coherency panels by removing the incoherence caused by lack of illumination and/or edge effects, as suggested by Ehinger (1992) for the case of full wave-equation migration.

Setting of the problem

The reflectivity of the earth is represented by a family of scattering points M located at the nodes of a rectangular grid (see figure 1). We denote by M the collection of the nodes M of the grid. Given a background velocity, we want to migrate the common shot gather data d recorded at a collection G of geophones G for a shot at point source S (see figure 1).

Let us denote by:

$$n = 1/v \tag{1}$$

the slowness associated to the given smooth background velocity v. The incident field $u_I(x,t)$ generated by the source in the smooth medium satisfies:

$$n^2 \frac{\partial^2 u_I}{\partial t^2} - \Delta u_I = f(t)\delta(x - x_S), \tag{2}$$

where:

$$f(t)$$
 is the band-limited source function (3)

The scattered field $u_S(x,t)$ corresponding to a slowness perturbation δn satisfies:

$$n^2 \frac{\partial^2 u_S}{\partial t^2} - \Delta u_S = -n^2 \cdot \frac{2\delta n}{n} \frac{\partial^2 u_I}{\partial t^2} \tag{4}$$

When δn corresponds to a family of point scatterers as in our case, it is of the form :

$$\frac{2\delta n}{n} = \sum_{M \in M} r_M \delta(x - x_M) \tag{5}$$

so that our scattered field u_S is given by :

$$n^2 \frac{\partial^2 u_S}{\partial t^2} - \Delta u_S = -n^2 \sum_{M \in M} r_M \frac{\partial^2 u_I}{\partial t^2} \delta(x - x_M), \tag{6}$$

where:

$$r_M = \text{reflectivity of point scatterer } M.$$
 (7)

Equations (2) and (6) (plus adequate initial and boundary conditions) determine the scattered field u_S which is recorded at the geophones G.

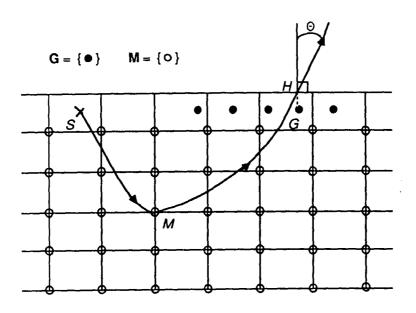


Figure 1. The geometry of the problem.

The forward WKBJ map

We describe in this paragraph the approximation of (2) and (6) which have to be used if one wants that the gradient of the corresponding data misfit function corresponds (up to positive weights) to a Kirchhoff migration.

Given the background slowness n, one first calculate (by ray tracing or finite difference solution of the eikonal equation for example):

For every grid point
$$M$$
 and every geophone G :

$$\tau_M^S = \text{ travel time from source } S \text{ to } M$$

$$\tau_M^G = \text{ travel time from } M \text{ to } H(\text{ not } G!)$$

$$\theta_M^G = \text{ emergence angle at } H(\text{ not } G!) \text{ of the ray coming from } M$$

$$n_M = \text{ slowness at point } M.$$
(8)

For every geophone
$$G$$

$$\Delta \tau^G = \text{travel time from geophone } G$$
 to the point H of the free surface located just above G .

These quantities will be considered as fixed in all the sequel. Then the forward map ϕ^{ta} associates, to any reflectivity section $r=(r_M,M\in \mathbf{M})$ a "true amplitude" synthetic section $c^{ta}=(c_G^{ta}(t),G\in \mathbf{G},0\leq t\leq T)$:

$$\phi^{ta}: r \to c^{ta} = \phi^{ta} r \tag{10}$$

which is defined as follows:

• the incident field u_I , solution of (2), is approximated, at any point $M \in \mathbf{M}$, using the WKBJ approximation:

$$u_I(M,t) = A_M^S S(t - \tau_M^S) \tag{11}$$

where:

$$\begin{cases} S(t) = f(t) & \text{in } 3 - D \\ S(t)^{(1/2)} = f(t) & \text{in } 2 - D \end{cases}$$
 (12)

the describes the shape of the wavelet signal radiated by the source which propagates in the smooth medium, and:

$$A_M^S$$
 is the WKBJ attenuation factor for the propagation from S to M. (13)

Notice that when the slowness n is constant, then (11) yields the exact solution to (2) in 3-D, and a solution which is asymptotically exact for large t in 2-D.

We shall suppose in the sequel that:

$$\begin{cases} S \text{ is three times continuously derivable on } \mathbf{R}, \text{ with } S(t) = 0 \text{ for } t \leq 0 \\ \text{and for } t \geq T_S, \text{ where } T_S > 0 \text{ is duration of the source.} \end{cases}$$
 (14)

• the scattered field u_S , solution of (6), is approximated, at the point H of the free surface just above the geophone G, by the same WKBJ approximation:

$$u_S(H,t) = -\sum_{M \in \mathbf{M}} A_H^G n_M^2 r_M \frac{\partial^2 u_I}{\partial t^2} (M, t - \tau_M^G)$$
 (15)

• the pressure field $c_G^{ta}(t)$ at geophone G is then given, using the fact that, when the scattering points M are far enough from the geophone, the scattered field u_S given by (15) is a superposition of (almost) plane waves, and that u_S is zero on the free surface just above G, by (see the Appendix):

$$c_G^{ta}(t) = -2\cos\theta_M^G \Delta \tau^G \frac{\partial u_S}{\partial t}(H, t)$$
 (16)

(the upperscript ta stands for "true amplitude", emphasizing the fact that this synthetic try to imitate the true amplitude data).

Putting together (11), (15) and (16), we see that:

$$\begin{cases} \text{ at each geophone } G \in \mathbf{G}, \text{ at each time } t \in [0, T] : \\ c_G^{ta}(t) = 2\Delta \tau^G \sum_{M \in \mathbf{M}} S^{(3)}(t - \tau_M^S - \tau_M^G) A_M^S n_M^2 r_M A_M^G \cos \theta_M^G \end{cases}$$
(17)

With the notations:

$$\begin{cases} \tau_M^{SG} = \tau_M^S + \tau_M^G = \text{ travel time from } S \to M \to H \\ A_M^{SG} = 2\Delta \tau^G \cos \theta_M^G A_M^S A_M^G n_M^2 \end{cases}$$
(18)

the WKBJ forward map can be rewritten

$$\begin{cases}
\phi^{ta} : r = (r_M, M \in \mathbf{M}) \to c^{ta} = (c_G^{ta}(t), G \in \mathbf{G}, o \leq t \leq T) \\
\text{defined by} \\
c_G^{ta}(t) = \sum_{M \in \mathbf{M}} S^{(3)}(t - \tau_M^{S,G}) A_M^{S,G} r_M
\end{cases}$$
(19)

Error function, descent imaging principle, and inversion

Suppose now that we are given a true amplitude shot-gather section $d^{ta} = (d_G^{ta}(t), G \in G, 0 \le t \le T)$ which we want to migrate. In order to compare this section with the synthetic section $c^{ta} = \phi.r$ defined by (19), we perform on both sections a preliminary treatment. We define at all geophones G:

$$d_G(t) = W_G(t)h * d_G^{ta}(t), \quad c_G(t) = W_G(t)h * c_G^{ta}(t)$$
(20)

where:

$$\begin{cases} W_G(t) &= \text{ amplitude correction factor (for the compensation of } \\ & \text{ the geometrical speading : } W_G(t) = t^{1/2} \text{ for } \\ & 2 - D data, \text{ or } W_G(t) = t^1 \text{ for } 3 - D \text{ data)} \\ h &= \text{ misfit filter (for a partial deconvolution of the data for example).} \end{cases} \tag{21}$$

We shall call:

$$d = (d_G, G \in \mathbf{G}), \quad c = (c_G, G \in \mathbf{G})$$
(22)

the data and the synthetics after this preliminary treatment, and:

$$\phi: r \to c \tag{23}$$

the forward map obtained by composing the WKBJ map (19) with the preliminary treatment (20).

A natural way to judge the ability of any reflectivity section $r = (r_M, M \in M)$ to explain the shot gather section d is to evaluate the least square error function:

$$J(r) = \frac{1}{2} \| \phi r - d \|^2 = \frac{1}{2} \sum_{G \in G} \int_0^T |c_G(t) - d_G(t)|^2 dt$$
 (24)

The smaller J, the better r!

The error function J allows two formulate two approaches (of increasing difficulty) to the problem of the interpretation of the data d:

• The imaging approach. The objective here is to find images $(m = m_M, M \in \mathbf{M})$ in the offset-depth domain which satisfy the:

Definition 1 Descent imaging principle

 $m = (m_M, M \in \mathbf{M})$ is a migrated image of the section $d = (d_G(t), G \in \mathbf{G}, 0 \le t \le T)$ iff:

$$J(0 + \lambda m) < J(0)$$
 for λ small enough or equivalently iff (25)

$$\langle m, \nabla J(0) \rangle_{\mathbf{M}} < 0 \tag{26}$$

The motivation for this definition is clear: if m satisfies (25), this means that λm generate a synthetic c which contain events which will substract partially, in (24), from d in order to decrease J: this means that the reflectors in m are located at the right place to generate some of the events of d; but there is no guarantee concerning the amplitudes (absolute and relative) of the events in m.

Of course, this imaging principle is quite weak, and there are an infinity of migrated sections m of a given time section d which satisfy (25) or (26)!

Definition 2 Migration algorithm:

A migration algorithm defines a way of associating, to any time section d, a depth section m (the migrated section) satisfying (25) or (26).

But all the m which satisfy (26) are clearly of the form:

$$m = -HM\nabla J(0) \tag{27}$$

where:

$$HM =$$
symetric positive definite matrix (28)

From (27) we see that a large class of migration algorithms is obtained, once a forward map and hence J- has been chosen, by computing $-\nabla J(0)$ and applying a given positive definite matrix HM:

<u>Definition</u> 3 The migration operator \mathcal{M} a ssociated to the forward map ϕ and the symetric positive definite migration-weight matrix HM is:

$$\begin{cases} \mathcal{M}: d \to m = -HM\nabla J(0) \\ where \\ J(r) = \frac{1}{2} \parallel \phi r - d \parallel^2 \end{cases}$$
 (29)

We shall see in paragraph 4 how to compute $\nabla J(0)$ and discuss the possible choices for the migration weight matrix HM.

• The inversion approach. The objective here is of course to find a reflectivity section \hat{r} which minimizes J:

Definition 4 Inversion principle

 $\hat{r}=(\hat{r}_M, M\in \mathbf{M})$ is an inverted image of the section $d=(d_G(t), G\in \mathbf{G}, 0\leq t\leq T)$ iff:

$$J(\hat{r}) \le J(r) \quad \text{for any } r \tag{30}$$

An inverted image \hat{r} is a solution of the normal equation:

$$\phi^T \phi \hat{r} = \phi^T d. \tag{31}$$

i.e.

$$HJ\hat{r} = -\nabla J(0) \tag{32}$$

where:

$$\begin{cases} HJ = \phi^T \phi &= \text{ Hessian of } J\\ -\nabla J(0) = \phi^T d &= \text{ Gradient of } J \text{ at } r = 0. \end{cases}$$
 (33)

Usually, HJ is not invertible, as not all the points of the grid are illuminated, so that the inversion problem has still many solutions - but which differ mainly outside the illumination zone (usually, the \hat{r} with minimum-norm is selected). The actual minimization of J can be performed by a gradient algorithm:

<u>Definition</u> 5 A gradient inversion algorithm is of the form:

$$\hat{r} = \lim_{k \to \infty} r^k \tag{34}$$

where r^k is defined by:

$$r^{0} = 0, \quad r^{k+1} = r^{k} - H^{k} \nabla J(r^{k})$$
 (35)

where $H^0, H^1, \ldots, H^k, \ldots$ are symetric positive definite matrices specific of the algorithm.

• Relations between the imaging and inversion approaches :

i) from (34) we see that:

$$r^1 = -H^0 \nabla J(0) \tag{36}$$

i.e. using (29) that:

$$\begin{cases} r^1 = \mathcal{M}d \\ \text{where } \mathcal{M} \text{ is the migration operator with the weights } H^0 \end{cases}$$
 (37)

which is the well-known fact that "migration is the first step of inversion"

ii) more interestingly, comparing the normal equation (32) for the inverted image \hat{r} and the formula (29) for the migrated image m, we see that if we choose as migration weight matrix:

$$HM = HJ^{\dagger} = \text{pseudo-inverse of } HJ$$
 (38)

then the migrated section m satisfy:

$$m = \text{minimum norm inverted image } \hat{r}$$
 (39)

(this choice amounts to perform one Newton step for the minimization of J starting from the initial guess r=0). This remark gives us a guideline for the choice of the migration weight matrix HM: the better the migration weight matrix HM approaches the pseudo inverse HJ^{\dagger} of the Hessian of J, the closer the migrated section m will be from the minimum norm solution \hat{r} , i.e. the closer the migration operator will approach the minimum norm inversion operator.

The gradient of the WKBJ error function J

We suppose from now on that the observation time T, the diffracting array M, and the geophone array G are chosen such that:

$$T \ge Max_{M \in \mathbf{M}, G \in \mathbf{G}} \{ \tau_M^{SG} \} + T_S \tag{40}$$

where T_S is the duration of the propagating wavelet S(t) (14). This will ensure that we have at each geophone G a complete information on the waves diffracted by each point M. Differentiating J(r) defined by (24) with respect to r yields:

$$\delta J = \sum_{G \in \mathbf{G}} \int_0^T (c_G(t) - d_G(t)) \delta c_G(t) dt$$
 (41)

If we define:

$$e = d - c =$$
residual section = part
of the data which is not explained by the reflectivity r (42)

and use the definition (19) (20) of c_G , formula (41) rewrites as:

$$\delta J = -\sum_{G \in G} \sum_{M \in M} A_M^{S,G} \delta r_M \int_0^T e_G(t) W_G(t) h * S^{(3)}(t - \tau_M^{S,G}) dt.$$
 (43)

Picking up the coefficient of δr_M in this formula yields $\partial J/\partial r_M(r) = (\nabla J(r))_M$:

$$[-\nabla J(r)]_M = \sum_{G \in \mathbf{G}} A_M^{S,G} \int_0^T e_G(t) W_G(t) h * S^{(3)}(t - \tau_M^{S,G}) dt$$
 (44)

When r = 0, then c = 0 and the residual e coïncides with the data d, so that:

$$[-\nabla J(0)]_M = \sum_{G \in G} A_M^{S,G} \int_0^T d_G(t) W_G(t) h * S^{(3)}(t - \tau_M^{S,G}) dt.$$
 (45)

But using (40) we can extend the integral to $(-\infty, +\infty)$ and use convolution notations:

$$[-\nabla J(0)]_M = \sum_{G \in G} A_M^{S,G}(W_G d_G) * h * S^{(3)}(\tau_M^{S,G})$$
 (46)

which can be rewritten as:

$$[-\nabla J(0)]_{M} = \sum_{G \in G} A_{M}^{SG} E_{G}(\tau_{M}^{S,G})$$
 (47)

where:

$$E_G = h * (W_G d_G) * S^{(3)} = h * (W_G^2 h * d_G^{ta}) * S^{(3)}.$$
(48)

Using the relation (12) between the propagating wavelet S(t) and the source function f(t), we can rewrite (48) as:

$$E_G = \left(f^{(2)} * h * [W_G^2 h * d_G^{ta}]\right)^{(s)}, \ s = \begin{cases} 1 & (3-D) \\ 1/2 & (2-D) \end{cases}$$
(49)

We shall call E_G the migration-filtered data, by opposition to $d_G = W_G h * d_G^{ta}$ which we called the misfit-filtered data.

When the weights $W_G(t)$ vary slowly with respect to the signal d_G^{ta} , one can approximate $E_G(t)$ by :

$$E_G = W_G^2 h * h * S^{(3)} * d_G^{ta} (50)$$

or equivalently:

$$E_G = W_G^2 h * h * f^{(2)} * d_G^{ta(s)}, \ s = \begin{cases} 1 & (3-D) \\ 1/2 & (2-D) \end{cases}$$
 (51)

Looking at the formula (47) for the gradient, we see that its right-hand side has the same structure as the Kirchhoff migration algorithm: the gradient of J at M is obtained by migrating at M, with a proper attenuation A_M^{SG} , the migration filtered data E_G recorded at each geophone G at a time equal from the travel time: source $S \to \text{scatterer } M \to \text{geophone } G$, and by stacking all these quantities at each point M. However, in the current usage of Kirchhoff migration, the true amplitude data are filtered according to:

$$E_G = W_G^K g^K * d_G^{ta(s)}, \ s = \begin{cases} 1 & (3 - D) \\ 1/2 & (2 - D) \end{cases}$$
 (52)

where:

$$\begin{cases} g^K = \text{Kirchhoff filter used to partially deconvolve the data ("spiking")} \\ W_G^K = \text{Kirchhoff amplitude correction factor.} \end{cases}$$
 (53)

One sees that the E_G given by (51) (gradient approach) and (52) (Kirchhoff migration) will coïncide as soon as:

$$W_G^K(t) = W_G(t)^2 (54)$$

$$g^K = h * h * f^{(2)} (55)$$

Equation (54) suggests that the amplitude correction $W_G^K(t)$ to be used when preparing data for the Kirchhoff migration should be t^2 (for 3-D data) or t (for 2-D data).

Concerning the preliminary filtering of the data, (55) suggests to use a filter g^K related to the source function f. In the current practice of Kirchhoff migration g^K is chosen such that it performs a "spiking deconvolution" of the signal generated by reflectors. In order to explicit how this can be achieved by a proper choice of h, we first compute the shape of the signal generated in our model by a plane wave reflector: as we have seen in (11), S(t) gives the shape of the incident signal arriving on the reflector; by the method of images, the reflected signal (i.e. the sum of all diffracted signals coming from all the slowness perturbations modelizing the reflector) has the same shape S(t); hence the signal recorded at a given geophone G near the free surface has a shape $S^{(1)}(t)$, i.e., using (12):

$$d_G^{ta}(t) \simeq \sum_{\text{reflectors } i} \alpha_j f^{(s)}(t) \text{ with } s = \begin{cases} 1 & (3-D) \\ 1/2 & (2-D). \end{cases}$$
 (56)

Let us now choose h such that it performs a "spiking deconvolution" of $d_G^{ta^{(1)}}$: from (56) this implies that:

$$h * f^{(1+s)} = \delta_B \tag{57}$$

where:

$$\delta_B = \text{band-limited Dirac function.}$$
 (58)

Then the corresponding migration filter g^K is given by:

$$g^K = h * h * f^{(2)} = h^{(1-s)} * \delta_B$$
 (59)

so that:

$$g^K * f^{(2s)} = \delta_B * h^{(1-s)} * f^{(2s)} = \delta_B * \delta_B$$
 (60)

which shows that g^K corresponds also to a (less demanding) spiking deconvolution for $(d_G^{ta})^{(s)}$.

In conclusion, we see that when the data d_G^{ta} have been generated by reflectors, chosing an h which performs a deconvolution of $d_G^{ta(1)}$ is equivalent to closing g^K which performs a deconvolution of $d_G^{ta(s)}$. This allows to replace for the determination of E_G , formula (51) (which would require the knowledge of the source function f) by formula (52) (where g^K can be determined approximately from the data themselves).

A family of Kirchhoff migration operators

We can now apply the definition (29) of a migration to the case where the forward map ϕ is defined by (20) - (23). Using formula (47) (49) for $-\nabla J(0)$ and restricting us to the case of diagonal migration weight matrices $HM = diag(WM_M, M \in M)$, we obtain a whole family of Kirchhoff migration operators:

Definition 6 A Kirchhoff migration operators is the mapping

$$\mathcal{M}: d = (d_G(t), G \in \mathbf{G}, 0 \le t \le T) \to m = (m_M, M \in \mathbf{M})$$

$$\tag{61}$$

defined by :

$$m_M = W M_M \sum_{G \in G} A_M^{SG} E_G(\tau_M^{SG}) \tag{62}$$

where:

$$WM_M = migration weights, specific of each migration operator$$
 (63)

$$A_M^{SG} = 2\Delta \tau^G \cos \theta_M^G A_M^S A_M^G n_M^2 \tag{64}$$

$$\tau_M^{SG} = travel\ time\ from\ S\ to\ H\ (just\ above\ G,\ see\ figure\ 1)$$
 (65)

$$E_G = \left(f^{(2)} * h * [W_G^2 h * d_G^{ta}]\right)^{(s)} \text{ with } s = \begin{cases} 1 & (3-D) \\ 1/2 & (2-D) \end{cases}$$
 (66)

$$d_G^{ta} = true \ amplitude \ seismic \ trace \ at \ geophone \ G.$$
 (67)

The above migration operators contain many (probably all) existing Kirchhoff migration formula, for an adequate choice of the misfit filter h, the amplitude factor $W_G(t)$ and the migration weights WM_M . For example if one chooses h as in (57), $W_G(t) = 1$ and $WM_M = 1/(A_M^S n_M)^2$, formula (62) rewrites, in 3-D:

$$m_M = \frac{2}{A_M^S} \sum_{G \in \mathbf{G}} \Delta \tau^G \cos \theta_M^G A_M^G (g^K * d_G^{ta})'(\tau_M^{S,G})$$
 (68)

 $(g^K$ given by (55)), which is a discrete version of formula (16) of Docherty, 1991, which itself is equivalent to equation (5) of Keho and Beydoun (1988), and to formula β of Bleistein (1987).

However, our derivation of these Kirchhoff migration formula has, beside the fact that it is simple and elementary, three advantages:

- 1. It indicates the precise relationship between the migration filter g^K used in conventional migration formula like (68), and the shape of the source function f(t) (cf. the discussion at the end of paragraph 5).
- 2. As noted at the end of paragraph 4, it provides a guideline for the choice of the migration coefficients WM: the corresponding diagonal matrix has to be chosen as close as possible to the pseudoinverse HJ^{\dagger} of the Hessian of the data misfit function. This will be investigated in paragraph ().

3. It shows that migrated sections are descent directions for the data misfit function:

$$\begin{cases} \text{ for any } r = (r_M, M \in \mathbf{M}) \text{ one has} \\ < \mathcal{M}(d-c), \nabla J(r) >_{\mathbf{M}} < 0 \\ \text{ where } d = \text{ is the amplitude corrected and filtered data time section,} \\ \text{ and } c = \phi r \text{ the amplitude corrected and filtered synthetic time section} \end{cases}$$

$$(69)$$

For example, performing iterative migration:

$$\begin{cases} r^{k+1} = r^k + \rho^k \mathcal{M}(d^k - c^k), & r^0 = 0 \\ \text{where} \\ c^k = \phi r^k \\ \rho^k > 0 \text{ is a descent step} \end{cases}$$
(70)

is nothing but computing a minimizing sequence for the least square error function J. Hence when the ρ^k are well chosen, the iterative migration (70) will converge to the solution \hat{r} of the inverse problem:

$$r^k \longrightarrow_{k \to \infty} \hat{r}$$
 which minimizes J . (71)

We will see in paragraph () an other way of taking advantage of this descent property.

Chosing optimal migration weights

The Hessian HJ of the least square error function J(r) defined in (24) is, using the formula (19) (20) for $c_G(t)$:

$$\begin{cases}
\text{ for any grid points } M \text{ and } P, \\
H J_{M,P} = \sum_{G \in \mathbf{G}} \int_0^T W_G(t)^2 h * S^{(3)}(t - \tau_M^{S,G}) h * S^{(3)}(t - \tau_P^{S,G}) A_M^{S,G} A_P^{S,G} dt
\end{cases}$$
(72)

As the weight $W_G(t)$ varies slowly compared $h * S^{(3)}(t)$, this rewrites, using (14), as:

$$HJ_{MP} \simeq \sum_{G \in \mathbf{G}} A_M^{S,G} A_P^{S,G} W_G(\tau_M^{S,G}) W_G(\tau_P^{S,G}) w * w(\tau_M^{S,G} - \tau_P^{S,G})$$
 (73)

where:

$$w = h * S^{(3)} (74)$$

In order to obtain a physical interpretation of this Hessian, we consider noiseless data d generated by a true reflectivity section r_{true} :

$$d = \phi r_{\mathsf{true}} \tag{75}$$

The application to this data of the Kirchhoff migration operator \mathcal{M} defined in (29) can be conceptually split into two steps:

Step 1: Compute a gradient image.

$$-\nabla J(0) = HJr_{\text{true}} \tag{76}$$

Step 2: Restore, as far as possible, the true amplitue of migrated reflectors by multiplying the gradient image by a migration weight matrix:

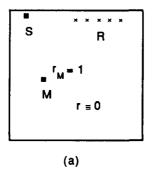
$$m = HM(-\nabla J(0)) \tag{77}$$

In practice, this matrix HM is always chosen to be diagonal:

$$HM = diag(WM_M, M \in M). \tag{78}$$

We obtain immediately from Step 1 a simple interpretation of the Hessian HJ (see figure 2): if the true reflectivity $r_{\rm true}$ contrains only one diffracting point located at node M with coefficient $r_M=1$, then the $M^{\rm th}$ line of HJ is nothing else than the gradient image $-\nabla J(0)$ reconstructed from the one-pixel reflectivity $r_{\rm true}$! It is also clear from formula (72) that, when the frequency content of the signal increases (i.e. the support of S(t) becomes smaller and smaller) and the width of the recording array increases, then the gradient image of figure 2(b) becomes more and more concentrated around M.

Because w*w has always maximum amplitude at 0, the gradient image is maximum at the true pixel location, so that reflectors will always be imaged at zero crossings of the migrated section when a gradient approach is used.



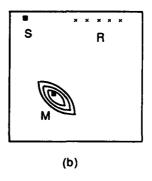


Figure 2. Interpretation of the M-th line of the Hessian HJ of the least-square error function J:

(a) one-pixel reflectivity (r_p = δ MP, P∈ M)

(b) M-th line of HJ = gradient image obtained from the one-pixel reflectivity (a)

We turn now to the choice of the migration weights $HM = diag(WM_M, M \in M)$ in step 2 of the migration. In sight of (76) (77) we want that the migrated image:

$$m = HM.HJr_{\rm true} \tag{79}$$

is as close as possible to the true reflectivity section r_{true} . This can be done by chosing HM such that:

$$\begin{cases} \|HM.HJ - I\|_{\infty} \text{ is minimum} \\ \text{over all diagonal matrices } HM, \end{cases}$$
 (80)

where $\| \|_{\infty}$ is the matrix norm associated to the $\| \|_{\infty}$ vector norm :

$$\begin{cases}
 \| A \|_{\infty} = Sup_{\|x\|_{\infty}=1} \| Ax \|_{\infty} \\
 \| x \|_{\infty} = Max_{i=1-n} | x_i |
\end{cases}$$
(81)

This norm has been chosen in order to ensure an easy solution of the problem (80), which rewrites, with generic matrix notations $A = n \times n$ matrix and $\Lambda =$ diagonal matrix:

$$\| \Lambda A - I \|_{\infty}$$
 is minimum. (82)

But:

$$\| \Lambda A - I \|_{\infty} = Max_{i=1...n} \sum_{j=1...n} | \lambda_i a_{ij} - \delta_{ij} |$$

i.e.

$$\| \Lambda A - I \|_{\infty} = Max_{i=1...n} \left\{ | \lambda_i | \sum_{j \neq i} | a_{ij} | + | \lambda_i a_{ii} - 1 | \right\}$$
 (83)

The minimum of $\|\Lambda A - I\|_{\infty}$ will be obtained by chosing, for each i = 1...n, $\lambda_i \in \mathbb{R}$ which minimizes $|\lambda_i| \sum_{j \neq i} |a_{ij}| + |\lambda_i a_{ii} - 1|$. As it can be seen on figure 3, the solution is given by:

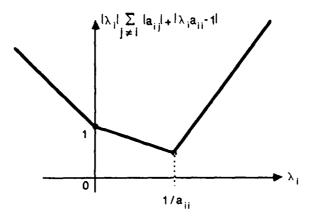
$$\begin{cases} \lambda_i = 1/a_{ii} & \text{if} \quad |a_{ii}| > \sum_{j \neq i} |a_{ij}| \\ \lambda_i = 0 & \text{if} \quad |a_{ii}| < \sum_{j \neq i} |a_{ij}|, \end{cases}$$
(84)

and the corresponding value of $\| \Lambda A - I \|_{\infty}$ is given by :

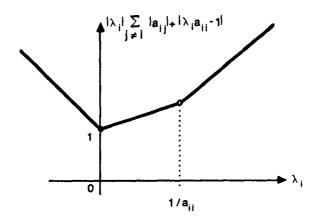
$$\| \Lambda A - I \|_{\infty} = Min \left\{ 1, Max_{i \mid s.t. \mid a_{ii} \mid > \sum_{j \neq i} \mid a_{ij} \mid \sum_{j \neq i} \mid a_{ij} \mid / \mid a_{ii} \mid \right\}$$
(85)

which satisfies:

$$0 \le || \Lambda A - I ||_{\infty} \le 1. \tag{86}$$



Case (a): $|a_{ij}| > \sum_{j \neq i} |a_{ij}|$



Case (b): $|a_{ij}| < \sum_{j \neq i} |a_{ij}|$

Figure 3. Graphical resolution of the minimization of $|A-I|_{\infty}$

Coming back to our migration problem, we see that the optimal (in the sense of (80)) migration weight matrix $HM^* = diag(WM_M^*, M \in M)$ is given by:

i) at points M where the Hessian HJ is "diagonal dominant"

(in the sense that $\mid HJ_{MM}\mid > \sum_{P\neq M}\mid HJ_{MP}\mid)$:

$$WM_{M}^{*} = HJ_{MM}^{-1} \tag{87}$$

ii) at points M where the Hessian HJ is "not diagonal dominant" (in the sense that $|HJ_{MM}| < \sum_{P \neq M} |HJ_{MP}|$):

$$WM_M^* = 0 (88)$$

The corresponding migrated section m satisfies (see (79) and (86):

$$\parallel m - \hat{r} \parallel_{\infty} \le 1. \tag{89}$$

Such section m would be called, in the usual terminology of migration, an "inverted section" because some effort has been done to recover the true amplitude of the reflector. However, from a mathematical point of view, m does not deserve that name, as it does by no way minimize the data misfit function J.

Remark 1 When the approximations (50) for $E_G(t)$ and (73) for HJ_{MP} are used, the Kirchhoff migration (62) with optimal migration weights (87) (88) rewrites:

i) at nodes M where the Hessian HJ is diagonal dominant:

$$m_{M} = \frac{\sum_{G \in G} A_{M}^{SG} W_{G}(\tau_{M}^{S,G})^{2} w * h * d_{G}^{ta}(\tau_{M}^{SG})}{\sum_{G \in G} (A_{M}^{SG})^{2} W_{G}(\tau_{M}^{S,G})^{2} w * w(0)}$$
(90)

ii) at all other nodes:

$$m_M = 0 (91)$$

Remark 2 The Hessian HJ may have no or too few diagonal dominant lines M. In that case one has to replace the data misfit function J(r) by a regularized version:

$$J_{\epsilon}(r) = J(r) + \frac{1}{2} \sum_{M \in M} \varepsilon_M r_M^2$$
 (92)

where:

$$\varepsilon = (\varepsilon_M, M \in \mathbf{M}) \tag{93}$$

is the collection of regularisation parameters.

The associated regularized Kirchhoff migration formula with optimal weights is:

$$m_{M}^{\epsilon} = \frac{\sum_{G \in \mathbf{G}} A_{M}^{SG} W_{G}(\tau_{M}^{S,G})^{2} w * h * d_{G}^{ta}(\tau_{M}^{SG})}{\varepsilon_{M} + \sum_{G \in \mathbf{G}} (A_{M}^{SG})^{2} W_{G}(\tau_{M}^{S,G})^{2} w * w(0)}$$

$$\dots if \quad \varepsilon_{M} \geq Max \left\{ 0, -HJ_{MM} + \sum_{P \neq M} HJ_{MP} \right\}$$

$$(94)$$

$$m_M^{\epsilon} = 0$$

... if $\epsilon_M < Max \left\{ 0, -HJ_{MM} + \sum_{P \neq M} HJ_{MP} \right\}$ (95)

The size of the "blank regions" in the migrated section m^{ε} will diminish when the $\varepsilon's$ are increasing; they will disappear if:

at each
$$M$$
, $\varepsilon_M \ge Max \left\{ 0, -HJ_{MM} + \sum_{P \ne M} HJ_{MP} \right\}.$ (96)

For example, if one chooses:

at each
$$M$$
, $\varepsilon_M = \sum_{P \neq M} |HJ_{MP}|$ (97)

then the optimal migration weight is:

at each
$$M$$
, $WM_M^{e*} = 1/\sum_{P \in \mathbf{M}} |HJ_{MP}|$ (98)

Notice that this weight could have been obtained by first performing mass condensation on the Hessian HJ, and then inverting it!

Cooperative Iterated Migrations

If we shot NSHOT sources $S_1 ldots S_{NSHOT}$, it is well known that the corresponding migrated sections will differ from one shot to the next (lack of illumination, edge effects ...) even if the exact velocity model is used. This is a difficulty for the picking of events on the migrated sections, which is currently achieved on the coherency panels: one would like that a default in the migration velocity result only in more or less dip of these events in the coherency panel, i.e. eliminate from the coherency panels the discoherence caused by lack of illumination and edge effects. Following the idea of Ehinger (1992), we consider a cooperative error function.

$$J_{WC}(r_1 \dots r_{NSHOT}) = \sum_{N=1}^{NSHOT} J_N(r_N) + \frac{1}{2} WC \sum_{N=2}^{NSHOT} || r_N - r_{N-1} ||^2$$
 (99)

where:

$$J_N(r_N)$$
 is the least square error function (24) for the N^{th} shot (100)

$$WC$$
 is a coherency weight of the reflectivity sections $r_1 \dots r_{NSHOT}$ (101)

Then the cooperatively inverted sections $\hat{r}_1 \dots \hat{r}_{NSHOT}$ are defined by :

$$\hat{r}_1 \dots \hat{r}_{NSHOT}$$
 minimizes $J_{WC}(r_1 \dots r_{NSHOT})$ (102)

But the gradient of J_{WC} is given by:

$$\nabla_{r_{N}} J_{WC}(r_{1} \dots r_{NSHOT}) = \begin{cases} r_{N} - r_{N-1} & N = NSHOT \\ -r_{N+1} + 2r_{N} - r_{N-1} & N = NSHOT - 1 \dots 2 \\ r_{1} - r_{2} & N = 1 \end{cases}$$
(103)

Multiplying (103) by the migration weights WM_N associated to a given migration operator \mathcal{M}_N for the N^{th} shot (cf.(29)) and changing the signs, we see that:

$$\begin{cases} \mathcal{M}_{N}(d_{N}-c_{N}) + WC \ WM_{N} \begin{cases} r_{N-1}-r_{N} & N = NSHOT \\ r_{N+1}-2r_{N}+r_{N-1} & N = NSHOT-1, \dots, 2 \\ r_{2}-r_{1} & N = 1 \end{cases} \\ \underline{\text{is a descent direction for } J_{WC} \text{ with respect to } r_{N}}. \end{cases}$$

$$(104)$$

Hence the cooperative iterated migration algorithm:

$$\begin{cases}
\bullet r_{N}^{0} = 0, & N = 1 \dots NSHOT \\
\bullet r_{N}^{k+1} = r_{N}^{k} + \rho_{k} \mathcal{M}_{N}(d_{N} - c_{N}^{k}) + WC \ WM_{N} \begin{cases}
r_{N-1}^{k} - r_{N}^{k} \\
r_{N+1}^{k} - 2r_{N}^{k} + r_{N-1}^{k} \\
r_{2}^{k} - r_{1}^{k}
\end{cases} \\
c_{N}^{k} = \phi_{N} r_{N}^{k}, & N = 1 \dots NSHOT \\
\text{where } \rho_{k} \text{ is a descent step}
\end{cases}$$
(105)

yields a minimizing sequence for the cooperative error function J_{WC} , i.e.:

$$r_N^k \to \hat{r}_N \quad N = 1, 2 \dots NSHOT$$
 (106)

In practice, this algorithm should not be much more difficult to implement than a classical Kirchhoff migration (computing r^k amounts to perform on each shot a separate migration). The additional burden comes from the necessity to forward simulate the model, in order to compute c_N^k and to choose a descent step ρ^k before performing the next set of migrations. Remark 3 When the migration weights WM are chosen to be optimal as indicated in paragraph (), it becomes possible to use a fixed $\rho^k \equiv 1$ for all shots and all iterations.

Remark 4 The same approach can be extended to common offset migrations or plane wave migrations in a straightforward way.

Conclusion

We have derived the Kirchhoff migration formula using the gradient of the data misfit function associated to a forward model of the WKBJ type. Beside the fact that it is simple and elementary, this approach has the following advantages:

- it makes explicit the forward time domain WKBJ model (11) (15) (16) (20) which is underlying the Kirchhoff migration;
- it leads to a precise definition of the preliminary treatment which has to be performed on the true amplitude data d_G^{ta} before the migration itself is performed:

$$E_G = \left(f^{(2)} * h * [W_G^2 h * d_G^{ta}]\right)^{(s)}$$
 (107)

where:

$$s = 1(3 - D \text{ data}) \text{ or } 1/2 (2 - D \text{ data})$$
 $W_G = t^s$
 $f = \text{ source function (Ricker)}$
 $h = \text{ data misfit filter (for deconvolution purpose)}$

When the data have been generated by reflectors, E_G can be approximated by:

$$E_G^K = W_G^2 g^K * (d_G^{ta})^{(s)} (108)$$

where g^K is chosen to perform a "spiking deconvolution" of $(d_G^{ta})^{(s)}$, which is the current practice for Kirchhoff migration.

- it allows the determination of optimal migration weights, which ensure that the migrated section will be as close as possible to the inverted section (which minimizes the data misfit function). These optimal weights, combined with an ounce of regularization, will ensure the best possible restitution of the true amplitude of the migrated events, for data with a given frequency content.
- it shows that Kirchhoff migrated sections are descent directions for the data misfit function:

```
< \mathcal{M}(d-c), \nabla J(r)>_{\mathbf{M}} < 0 \ \forall r=(r_{\mathbf{M}}, M\in \mathbf{M}) where d=W_Gh*d^{ta} is the amplitude corrected and filtered data time section and c=\phi x the synthetic time section.
```

Hence it becomes possible to build up iterative processes which include the Kirchhoff migration as a brick and correspond to the minimization of some objective function, as for example the cooperative iterated migration.

Acknowledgements

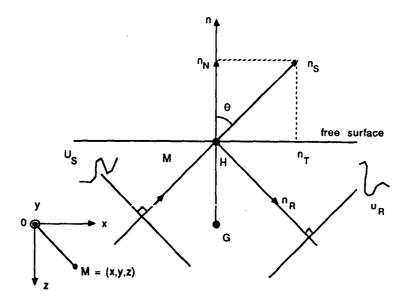
The author is indebted to Sam GRAY (Amoco Research Center, Tulsa) who helped him to find a path in the jungle of the literature on migration, and explained the grass-root nature of the Kirchhoff Migration, and to Francis COLLINO (INRIA, Rocquencourt) for fruitful discussions on the WKBJ approximation.

References

- [1] G. Beylkin. The Inversion Problem and Applications of the Generalized Radon Transform. XXXVII:579-599, 1984. Comm. Pure and Applied Math.
- [2] G. Beylkin. Imaging of Discontinuities in the Inverse Scattering Problem by Inversion of a Causal Generalized Radon Transform. J. Math. Phys, 26((1)), January 1985.
- [3] N. Bleistein. On the Imaging of Reflectors in the Earth. Geophysics, 52(7):931-942, July 1987.
- [4] J.A. Carter and L.N. Frazer. Accommodating Lateral Velocity Changes in Kirchhoff Migration by Means of Fermat's Principle. *Geophysics*, 49(1):46-53, January 1984.
- [5] P. Docherty. A Brief Comparison of Some Kirchhoff Integral Formula for Migration and Inversion. Geophysics, 56(8):1164-1169, August 1991.
- [6] A. Ehinger. PSI-report, 1992.
- [7] S.H. Gray, J.K. Cohen, and N. Bleisten. Velocity Inversion in a Stratified Medium with Separated Source and Receiver. J. Acoustical Soc. Am., 68((1)):234-240, July 1980.
- [8] T.H. Keho and W.B. Beydoun. Paraxial Ray Kirchhoff Migration. Geophysics, 53:1540-1546.

Appendix

Calculation of the signal recorded at a geophone located close to a plane free surface for a given incident plane wave.



The wave field in the vicinity of the geophone G is splitted into the sum of an incident wave u_S (which is plane, as the geophone is far enough from the diffracting point) and reflected wave u_R :

$$\begin{aligned} u_S(M,t) &= S\left(t - \frac{OM.n_S}{c}\right) & \parallel n_S \parallel = 1 \\ u_R(M,t) &= -S\left(t - \frac{OH.n_S + HM.n_R}{c}\right) & \parallel n_R \parallel = 1 \\ u(M,t) &= u_S + u_R = S\left(t - \frac{OM.n_S}{c}\right) - S\left(t - \frac{OH.n_S + HM.n_R}{c}\right) \end{aligned}$$

If the incident and reflected wave satisfy:

$$n_S = n_T + n_N, \quad n_R = n_T - n_N,$$

then for all M of the surface, i.e. such that $HM.n_N=0$, one has:

$$u(M,t) = S\left(t - \frac{OH.n_S + HM.n_T}{c}\right) - S\left(t - \frac{OH.n_S + HM.n_T}{c}\right) = 0$$

Hence u(M,t) is the solution near the free-surface boundary. At a geophone $G=(x_G,y_G,z_G)$, the signal is:

$$c_G(t) = u(G, t)$$

If the geophone depth z_G is a fraction of the spatial wavelength of the signal, one can write:

$$\frac{\partial u}{\partial z}(H,t) \simeq \frac{u(G,t) - u(H,t)}{z_G - z_H}$$

so that, as $u(H,t) \equiv 0$:

$$c_G(t) \simeq (z_G - z_H) \frac{\partial u}{\partial z}(H, t)$$

But:

$$\frac{\partial u}{\partial z}(M,t) = -\frac{n_{Sz}}{c}S'\left(t - \frac{OM.n_I}{c}\right) + \frac{n_{Rz}}{c}S'\left(t - \frac{OH.n_S + HM.n_R}{c}\right)$$

Hence, as $n_{Sz}=\cos\theta=-n_{Rz}$ and $n_{Sx}=n_{Rx}=n_x$; $n_{Sy}=n_{Ry}=n_y$

$$\frac{\partial u}{\partial z}(H,t) = -\frac{2\cos\theta}{c}S'\left(t - \frac{OH.n_I}{c}\right) = -\frac{2\cos\theta}{c}\frac{\partial u_S}{\partial t}(H,t)$$

An approximate expression for the signal at geophone G is hence

$$c_G(t) \simeq -2\cos\theta \Delta \tau^G \frac{\partial u_S}{\partial t}(H,t)$$

where:

$$\Delta \tau^G = \frac{z_G - z_H}{c}$$
 = travel time from geophone G to free surface

Remark:

Notice that, in the above formula, the emergence angle θ and the time delay $\frac{OH.n_S}{c}$ are evaluated at the point H of the free surface located above the geophone G, and not at G itself.



Unité de Recherche INRIA Rocquencourt Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)

Unité de Recherche INRIA Lorraine Technopôle de Nancy-Brabois - Campus Scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 VILLERS LES NANCY Cedex (France)
Unité de Recherche INRIA Rennes IRISA, Campus Universitaire de Beaulieu 35042 RENNES Cedex (France)
Unité de Recherche INRIA Rhône-Alpes 46, avenue Félix Viallet - 38031 GRENOBLE Cedex (France)
Unité de Recherche INRIA Sophia Antipolis 2004, route des Lucioles - B.P. 93 - 06902 SOPHIA ANTIPOLIS Cedex (France)

EDITEUR

INRIA - Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)

ISSN 0249 - 6399

