Non-linear partial derivative and its De Wolf approximation for non-linear seismic inversion

Ru-Shan Wu and Yingcai Zheng

1Modeling and Imaging Laboratory, Earth and Planetary Sciences, University of California, Santa Cruz, CA 95064, USA. E-mail: wrs@pmc.ucsc.edu
2Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

SUMMARY

We demonstrate that higher order Fréchet derivatives are not negligible and the linear Fréchet derivative may not be appropriate in many cases, especially when forward scattering is involved for large-scale or strong perturbations (heterogeneities). We then introduce and derive the non-linear partial derivative including all the higher order Fréchet derivatives for the acoustic wave equation. We prove that the higher order Fréchet derivatives can be realized by consecutive applications of the scattering operator and a zero-order propagator to the source. The full non-linear partial derivative is directly related to the full scattering series. The formulation of the full non-linear derivative can be used to non-linearly update the model. It also provides a new way for deriving better approximations beyond the linear Fréchet derivative (Born approximation). In the second part of the paper, we derive the De Wolf approximation (DWA; multiple forescattering and single backscattering approximation) for the non-linear partial derivative. We split the linear derivative operator (i.e. the scattering operator) into forward and backward derivatives, and then reorder and renormalize the multiple scattering series before making the approximation of dropping the multiple backscattering terms. This approximation can be useful for both theoretical derivation and numerical calculation. Through both theoretical analyses and numerical simulations, we show that for large-scale perturbations, the errors of the linear Fréchet derivative (Born approximation) are significant and unacceptable. In contrast, the DWA non-linear partial derivative (NLPD), can give fairly accurate waveforms. Application of the NLPD to the least-square inversion leads to a different inversion algorithm than the standard gradient method.

Key words: Inverse theory; Body waves; Seismic tomography; Theoretical seismology; Wave propagation.

1 INTRODUCTION

The Fréchet derivative plays a key role in geophysical inverse problems. A non-linear problem is typically solved iteratively and at each step the general non-linear problem is linearized through the use of a functional derivative, that is the Fréchet derivative, which loosely speaking relates an infinitesimal change in the model to an infinitesimal change in data. The Fréchet derivative is usually referred to the first-order derivative. However, sometimes the second-order derivative, called the Hessian, is also used in the full Newton method. Higher order terms of functional derivative are assumed to be insignificant or of no importance, and therefore are generally neglected. In seismic traveltime tomography, explicit formulae for the Fréchet derivative have been derived, and numerical calculations have been carried out (e.g. Marquering et al. 1999; Dahlen et al. 2000; Hung et al. 2000; Zhao et al. 2000; Dahlen 2004, 2005; Zhou et al. 2004, 2011; de Hoop & van der Hilst 2005; Dahlen & Nolet 2006). This spatial distribution of the traveltime changes caused by the velocity perturbations at the corresponding points is called the sensitivity kernel for traveltimes. The kernels is usually derived through the first-order Born approximation based on the formulation of linearized Fréchet derivative. Similar sensitivity kernel formulation has been used in reflection seismics (de Hoop et al. 2006; Xie & Yang 2008). For the full waveform inversion, the Fréchet derivative is also used to relate model parameter perturbations to changes in seismic waveforms (Tarantola 1984, 1986, 2005; for a review see Virieux & Operto 2009).

Linearized wave equation, such as in Born modelling, assumes a linear independence between the scattered wavefields. The total scattered field is simply a superposition of the scattered fields from all individual scattering points. In terms of the Fréchet derivative, the theory predicts the resulting data change by summing up the data changes caused by all the point perturbations without any mutual interactions. In order to
mitigate the problem of huge storage for the point scattering kernels, Chevrot & Zhao (2007) proposed to use wavelet transform applied to the model space; Loris et al. (2010) and Simons et al. (2011) apply the wavelet transform to the model for the purpose of L-1 norm regularization in the inversion. However, in both cases the linearized Fréchet derivatives are used for the sensitivity kernel calculations.

For the real Earth, the wave equation is strongly non-linear with respect to the medium changes, except for some weakly perturbed media. Even though it is a big step forward from the ray-theory-based kernel to the linear finite frequency kernel for wave equation tomography, the latter has severe limitations in applying to transmission tomography and waveform inversion, especially for multiscale and strong heterogeneities. For large-scale velocity perturbations, the phase accumulation of forward scattering renders the Born approximation unacceptable in many cases. However, to our best knowledge, the kernel calculations for finite-frequency tomography and waveform inversion are almost exclusively based on the linear Fréchet derivative in the literature. We noticed that in order to mitigate the limitation of the Born approximation and take into account of multiple forward scattering for large-scale structure, in global tomography an asymptotic non-linear term has been introduced (so called ‘non-linear asymptotic coupling theory’, see Li & Romanowicz 1995, 1996; Romanowicz et al. 2008; Panning et al. 2009; Lekic & Romanowicz 2011). Later the approximation is relaxed and named ‘non-linear Born approximation’, see Romanowicz et al. 2008; Panning et al. 2012). Because of the nature of the asymptotic approximation of the non-linear term, there are cases where both the Born and asymptotic approximations are inaccurate as demonstrated in Romanowicz et al. (2008). Therefore, a theory of non-linear partial derivative (NLPD) is desirable for general non-linear seismic inversion. In other fields, NLPDs have been introduced, such as for resistivity inversion (McGillivray & Oldenburg 1990), and optical diffuse imaging (Kwon & Yazici 2010). Relatively little work has been done to address the influence of higher order Fréchet derivatives in the context of seismic wave scattering and inversion. Can these higher order terms be neglected altogether? Is there a way to handle the full NLPD including all the higher orders of Fréchet derivative in tomography or inversion? In this paper, we first discuss the Fréchet derivative for the case of acoustic wave equation (Section 2). Then, in Section 3 we derive the full NLPD as a series with all the higher order terms, and discuss its relation to the Born scattering series. In Section 4, we introduce the De Wolf approximation (DWA) of the NLPD. Based on the relation between derivative operator and scattering operator, we introduce the split of the scattering operator into forward- and backward-operator. The DWA is a multiple foreshattering and single backscattering approximation after reordering and renormalization of the multiple scattering series. In Section 5, we do some preliminary investigation on the application of the NLPD to the least-squares (LS) inversion, resulting in a different inversion algorithm than the standard gradient method. In Section 6, we show the profound differences between the linear Fréchet derivative and the non-linear derivative through numerical examples for a Gaussian ball of different sizes and a cube model. We show that for smooth perturbations (heterogeneities), the DWA of the non-linear kernel gives similar results as the FD (finite difference) calculations, significantly different from the linear prediction (Born model). Finally, we give the conclusions and discussions.

2 SENSITIVITY KERNEL BASED ON THE WAVE EQUATION AND THE LINEAR FRÉCHET DERIVATIVE

In this paper, we will treat the acoustic wave problem (see, Stolt & Benson 1986; Berkhourt 1987; Wu 1996; Tarantola 2005; Wu et al. 2006) and leave the generalization to elastic waves for future research and publications.

For a linear isotropic acoustic medium, the wave equation in the frequency \( \omega \) domain is

\[
\nabla \cdot \left( \frac{1}{\rho} \nabla p + \frac{\omega^2}{\kappa} p \right) = 0,
\]

(1)

where \( p \) is the pressure field, \( \rho \) and \( \kappa \) are the density and bulk modulus of the medium, respectively. Assuming \( \rho_0 \) and \( \kappa_0 \) as the parameters of the background medium, eq. (1) can be written as

\[
\frac{1}{\rho_0} \nabla^2 p + \frac{\omega^2}{\kappa_0} p = -\left[ \omega^2 \left( \frac{1}{\kappa} - \frac{1}{\kappa_0} \right) p + \nabla \cdot \left( \frac{1}{\rho} - \frac{1}{\rho_0} \right) \nabla p \right],
\]

(2)

or

\[
(\nabla^2 + k^2) p(x) = -k^2 \varepsilon(x) p(x),
\]

(3)

where

\[
k = \frac{\omega}{v_0}, \quad v_0 = \sqrt{\frac{\kappa_0}{\rho_0}}.
\]

(4)

The right-hand side of (3) is an equivalent force term with

\[
\varepsilon(x) = \varepsilon_s(x) + \frac{1}{k^2} \nabla \cdot \varepsilon_p \nabla
\]

(5)

as the scattering potential. Note that \( \varepsilon(x) \) is an operator instead of a scalar function, with

\[
\varepsilon_p(x) = \frac{\rho_0}{\rho}(x) - 1 = -\frac{\delta\rho}{\rho},
\]

(6)
Figure 1. Directional patterns of the scattering kernels for different parametrizations. In the figure, positive $z = x_3$ is in the up direction. (a) Bulk modulus $\kappa$ and (b) density $\rho$; (c) velocity $v$ and (d) impedance $\zeta$. Incident plane wave is along the $z$-direction ($x_3$).

and

$$\varepsilon(x) = \frac{\kappa_0}{\kappa}(x) - 1 = -\frac{\delta\kappa}{\kappa},$$

(7)

We can parametrize the medium using velocity $v = \sqrt{\kappa/\rho}$ and impedance $\zeta = v\rho = \sqrt{\kappa\rho}$, so that

$$\frac{\delta v}{v} = \frac{1}{2} \frac{\delta\kappa}{\kappa}, \quad \frac{1}{2} \frac{\delta\rho}{\rho},$$

$$\frac{\delta\zeta}{\zeta} = \frac{1}{2} \frac{\delta\kappa}{\kappa} + \frac{1}{2} \frac{\delta\rho}{\rho}.$$  

(8)

Therefore, the equivalent force term (5) can be also expressed as

$$\varepsilon(x) = -\left[\frac{\delta v}{v}(x) - \frac{1}{k^2} \nabla \cdot \frac{\delta v}{v}(x) \nabla\right] - \left[\frac{\delta\zeta}{\zeta}(x) + \frac{1}{k^2} \nabla \cdot \frac{\delta\zeta}{\zeta}(x) \nabla\right].$$

(9)

We see that different parameters have different scattering patterns. We will come back to this issue in Section 4 when dealing with the split of scattering operator (see Fig. 1).

If $\rho$ is kept constant ($\rho = \rho_0$), then $\varepsilon_s = c_0^2/c^2 - 1$, resulting in the scalar medium case, where the perturbation function (scattering potential) is a scalar quantity,

$$\varepsilon_s(x) = \frac{c_0^2}{c^2(x)} - 1 = \frac{s^2(x) - s_0^2}{s_0^2}.$$  

(10)

Eq. (3) can be written into an equivalent integral equation form (Lippmann-Schwinger equation),

$$p(x) = p^0(x) + k^2 \int_V d^3x' g_0(x;x') \varepsilon(x') p(x').$$  

(11)

where $g_0(x;x')$ is the background Green’s function satisfying the equation

$$(\nabla^2 + k^2) g_0(x;x') = -4\pi \delta(x-x').$$

(12)

Eq. (11) is an integral equation, since $p(x')$ inside the volume integral is unknown. In operator form, eq. (11) can be written as

$$p = p^0 + G_0 \varepsilon p,$$  

(13)

where $G_0$ is the Green’s operator, a non-diagonal integral operator. For a given reference medium, $G_0$ is a volume integral operator with Green’s function $g_0(x;x')$ as the kernel, defined through,

$$(G_0u)(x) = k^2 \int_V d^3x' g_0(x;x') u(x').$$  

(14)
where \( u \) is a wavefield (e.g. the scattered wavefield). Eq. (13) is expressed as a summation of the incident field and the perturbed field (scattered field). Substituting this sum into the unknown \( p \) in the right-hand side iteratively, results in a formal solution of the wave equation as a scattering series (the Born series)

\[
p = p^0 + G_\delta p^0 + G_\delta G_\delta p^0 + \cdots \]

\[
= \sum_{n=0}^{\infty} [G_\delta]^n p^0
\]

and the first-order approximation is called the Born approximation

\[
p_{\text{Born}} = p^0 + G_\delta p^0
\]

\[
= p^0(x) + k^2 \int_V d^3 x' g_0(x; x') \varepsilon(x') p^0(x').
\]

Now we consider the Fréchet derivative for the forward modelling operator. In accordance with Tarantola’s (2005, ch. 5) format and terminology, we write the forward problem in an operator form

\[
d = A(m).
\]

where \( d \) is the data vector (pressure field generated by the modelling), \( m \) is the model vector and \( A \) is the forward modelling operator. Assume an initial model \( m_0 \), we want to quantify the sensitivity of the data change \( \delta d \) to the model change \( \delta m \). For a linear modelling operator, or a quasi-linear operator, we can calculate the data change using the linear Fréchet derivative (Tarantola 2005, eq. 5.93)

\[
A(m_0 + \delta m) = A(m_0) + A'(m_0)\delta m + O(\|\delta m\|^2),
\]

where \( A' \) is the first Fréchet derivative operator (‘\( G_\delta \)’ in Tarantola’s notation). By comparing (18) with (16), we obtain

\[
A(m_0) = p^0 = g_0(x; x_0)
\]

\[
A'(m_0) = G_\delta p^0 = k^2 \int_V d^3 x' g_0(x; x') p^0(x').
\]

We see that the linear Fréchet derivative operator is equivalent to a Born modelling (scattering) operator. In this case, the model perturbation is \( \delta m = \varepsilon(x') \).

### 3 Higher Order Fréchet Derivatives for the Acoustic Wave Equation and the Born Series

Now, we discuss the higher order Fréchet derivatives and the full NLPD that includes all the higher order terms.

The linear Fréchet derivative operator \( A'(m_0) \) is defined as the ratio between the data change and the model perturbation when the model perturbation is approaching infinitely small values (Tarantola 2005; Zhang 2005). \( A' \) is a bounded linear operator, which maps the model space \( \mathcal{M} \) to the data space \( \mathcal{D} \). Its mathematical definition reads as

\[
\lim_{\|\delta m\|_\mathcal{M} \to 0} \frac{\|A(m_0 + \delta m) - A(m_0) - A'(m_0)\delta m\|_\mathcal{D}}{\|\delta m\|_\mathcal{M}} = 0,
\]

(20)

where \( \|\cdot\|_\mathcal{M} \) and \( \|\cdot\|_\mathcal{D} \) are norms defined on \( \mathcal{M} \) and \( \mathcal{D} \), respectively. The Fréchet derivative associates a model \( m_0 \in \mathcal{M} \) to a linear bounded operator \( A'(m_0) \). We define the space of all such bounded operators as

\[
L(\mathcal{M}, \mathcal{D}) = \{A'(m_0) | m_0 \in \mathcal{M}\}.
\]

(21)

In the context of the wave equation, \( m_0 \) is the starting model for the inversion. Higher order Fréchet derivatives can be derived iteratively from the first-order Fréchet derivative. Let’s first consider the second-order derivative. The first-order functional difference from the perturbation is

\[
A(m_0 + \delta m) - A(m_0) = A'(m_0)\delta m + O(\|\delta m\|^2).
\]

(22)

Second-order functional difference from the perturbation can be added to improve the accuracy (Teschl 1998; Zhang 2005),

\[
A(m_0 + \delta m) - A(m_0) = A'(m_0)\delta m + \frac{1}{2} A''(m_0)\delta m^2 + O(\|\delta m\|^3).
\]

(23)

\( A''(m_0) \) can be derived from the first-order derivative, defined in a similar way as in (20) (derivative of the derivative field),

\[
\lim_{\|\delta m_2\|_\mathcal{M} \to 0} \frac{\|A''(m_0 + \delta m_2) - A'(m_0) - A''(m_0)\delta m_2\|_{(\mathcal{M}, \mathcal{D})}}{\|\delta m_2\|_\mathcal{M}} = 0,
\]

(24)
where the norm on $L(M, D)$ is defined as
\[
\|Q\|_{L(M, D)} = \sup \frac{\|Qx\|_D}{\|x\|_M}, \quad x \in M \setminus \{0\}.
\]  

(25)

Following the same procedure, we can derive the $n$th-order Fréchet derivative,
\[
\lim_{\delta m_n \to 0} \frac{\|A^{(n)}(m_0 + \delta m_n) - A^{(n)}(m_0)\|_D}{\|\delta m_n\|_M} = 0.
\]

(26)

The higher order Fréchet derivatives are multilinear mapping operators with properly defined norms (for details, see Teschl 1998; Zhang 2005). The total functional difference can be expressed by a Taylor series including all orders of derivatives
\[
A(m_0 + \delta m) - A(m_0) = A'(m_0) \delta m + \frac{1}{2!} A''(m_0) (\delta m)^2 + \cdots + \frac{1}{n!} A^{(n)}(m_0) (\delta m)^n + \cdots,
\]

(27)

where $A'$, $A''$ and $A^{(n)}$ are the first-, second-, and $n$th-order Fréchet derivatives. Series (27) is closely related to the Born series. We will discuss their relation in the following.

In seismic inversion, normally higher order Fréchet derivatives are neglected. However, the neglect of higher order terms leads to the Born approximation. We know that the Born approximation is a poor approximation for the forward-scattering problems and is not acceptable in many applications. We also understand that the physical basis of waveform tomography using traveltimes is the forward scattering. Therefore, it is necessary to develop a theory of partial derivative and sensitivity kernel beyond the Born approximation.

With the NLPD including all higher order terms, data residual (measured scattered field) can be related to model perturbations more accurately. We define
\[
A^{(NLPD)}(m_0, \delta m) \delta m = A(m_0 + \delta m) - A(m_0),
\]

(28)

where $A^{(NLPD)}$ is the NLPD at the current model (background model) $m_0$. Then, we have
\[
A^{(NLPD)}(m_0, \delta m) = A'(m_0) \delta m + \frac{1}{2!} A''(m_0) (\delta m)^2 + \cdots + \frac{1}{n!} A^{(n)}(m_0) (\delta m)^n + \cdots.
\]

(29)

Note that $A^{(NLPD)}$ is $\delta m(x)$ dependent because of the non-linear mutual interactions (multiple scattering).

Under the linear approximation, $A'$ is a linear operator and $A' \delta m$ is a matrix multiplication on the perturbation vector in the discrete form. Therefore, the influences of parameter perturbations at different points are independent from each other. In the case of the wave equation, this linearization is only valid for small-scale weak heterogeneities (it will be discussed in later sections). On the other hand, the NLPD, if the series converges, can precisely predict the data change $\delta d$ due to the model perturbation $\delta m$. Higher order terms account for the interactions between different parameter perturbations (multiple scattering). Eq. (27) and can be written as
\[
d = d_0 + \delta d
\]
\[
d_0 = A(m_0)
\]
\[
\delta d = A(m_0 + \delta m) - A(m_0) = d_1 + d_2 + \cdots + d_n + \cdots,
\]

(30)

where
\[
d_1 = A'(m_0) \delta m
\]
\[
d_2 = \frac{1}{2!} A''(m_0) (\delta m)^2.
\]

(31)

The zero-order term is the wave solution in the background medium. In the case of a point source at $x_s$, it can be expressed as the background Green’s function:
\[
d_0 = p_0(x_s) = A(m_0) = g_0(x_s; x_s),
\]

(32)

where $x_s$ and $x_s$ are the receiver and source position vectors, respectively (this notation is consistent with Dahlen et al. 2000).

In order to derive the explicit form in terms of parameter perturbations in acoustic media, we rewrite the linear Fréchet derivative (the first-order Born solution) into another form. Substitute (5) into the volume integral in (16) yielding
\[
d_1 = A'(m_0) \delta m
\]
\[
= k^2 \int_D d^3 x \, g_0(x_s; x) \, \varepsilon_s(x) \, g_0(x_s; x) + \int_D d^3 x \, g_0(x_s; x) \, \nabla \cdot \left[ \varepsilon_{ss} \, (x) \, \nabla g_0(x_s; x) \right],
\]

(33)
where the gradient operator $\nabla$ is with respect to $x$. Assume the integral volume is large enough so that the contribution from its boundary can be neglected, and using integration by parts, the second term at the right-hand side of (33) can be written as

$$
\int_{V} d^{3}x_{1}g_{0}(x_{1};x_{i}) \nabla \cdot \left[ \epsilon_{\rho}(x) \nabla g_{0}(x_{1};x_{j}) \right] = -\int_{V} d^{3}x_{1} \epsilon_{\rho}(x) \nabla g_{0}(x_{1};x_{i}) \cdot \nabla g_{0}(x_{1};x_{j}).
$$

(34)

In the derivation, it is assumed that the medium perturbations do not have sharp boundaries (relative to the wavelength) so that the gradient operation does not produce singular integrals. Then (33) becomes

$$
d_{1} = A'(m_{0}) \delta m
= k^{2} \int_{V} d^{3}x_{0}(x_{1};x) \epsilon_{\rho}(x) g_{0}(x_{1};x_{j}) - \int_{V} d^{3}x_{1} \epsilon_{\rho}(x) \nabla g_{0}(x_{1};x_{i}) \cdot \nabla g_{0}(x_{1};x_{j}).
$$

(35)

In operator form, we can write the above equation as

$$
A'(m_{0}) \delta m = G_{0} \left( S_{s} \right)^{T} \left( \epsilon_{\rho} \right) g_{0} = G_{0} S^{T} \epsilon g_{0},
$$

(36)

where

$$
g_{0} = A^{0} = A(m_{0}) = g_{0}(x_{1};x_{j})
$$

(37)

is the background Green’s function and

$$
G_{0} = \int_{V} d^{3}x_{0}(x_{1};x)
$$

(38)

is the background Green’ operator, $S_{s}$, $S_{p}$ are the scattering operators defined as

$$
S_{s} = k^{2} \sqrt{\nabla_{s} \cdot \nabla_{in}},
$$

(39)

$$
\nabla_{in} \equiv \frac{1}{k} \nabla \text{ is the frequency-normalized gradient operator, } \nabla_{in} \text{ and } \nabla_{out} \text{ are the gradient operators for the incident and outgoing (scattered) waves (related to the corresponding Green’s functions), respectively. So, the first Fréchet derivative is an operator}

A'(m_{0}) = G_{0} S^{T} A^{0} = G_{0} S^{T} g_{0}
$$

(40)

and the perturbation vector is

$$
\delta m = \epsilon(x) = 
\begin{pmatrix}
\epsilon_{s}(x) \\
\epsilon_{p}(x)
\end{pmatrix}.
$$

(41)

In a similar way, we can get the second-order term and the second Fréchet derivative

$$
d_{2} = \frac{1}{2!} A''(m_{0})(\delta m)^{2} = (G_{0} S)^{T} g_{0} \epsilon^{2} = \left( G_{0} S_{s}, G_{0} S_{p} \right)^{2} \left( \begin{array}{c}
\epsilon_{s} \\
\epsilon_{p}
\end{array} \right)^{2} g_{0}
$$

(42)

where

$$
(G_{0} S_{s}, G_{0} S_{p}) \left( \begin{array}{c}
\epsilon_{s} \epsilon_{s} \\
\epsilon_{p} \epsilon_{p}
\end{array} \right) g_{0} =
\begin{pmatrix}
S_{s} G_{0} S_{s} & S_{s} G_{0} S_{p} \\
S_{p} G_{0} S_{s} & S_{p} G_{0} S_{p}
\end{pmatrix}
\begin{pmatrix}
\epsilon_{s} \epsilon_{s} \\
\epsilon_{p} \epsilon_{p}
\end{pmatrix} g_{0};
$$

(43)

where

$$
(G_{0} S_{s}, G_{0} S_{p}) \left( \begin{array}{c}
\epsilon_{s} \epsilon_{s} \\
\epsilon_{p} \epsilon_{p}
\end{array} \right) g_{0} = k^{4} \int_{V_{1}} d^{3}x_{1} g_{0}(x_{1};x_{i}) \epsilon_{s}(x_{1}) g_{0}(x_{1};x_{j}),
$$

$$
(G_{0} S_{s}, G_{0} S_{p}) \left( \begin{array}{c}
\epsilon_{s} \epsilon_{p} \\
\epsilon_{p} \epsilon_{p}
\end{array} \right) g_{0} = k^{4} \int_{V_{1}} d^{3}x_{1} g_{0}(x_{1};x_{i}) \epsilon_{p}(x_{1}) \nabla_{1} g_{0}(x_{1};x_{i}) \cdot \nabla_{1} g_{0}(x_{1};x_{j}),
$$

$$
(G_{0} S_{s}, G_{0} S_{p}) \left( \begin{array}{c}
\epsilon_{s} \epsilon_{s} \\
\epsilon_{p} \epsilon_{p}
\end{array} \right) g_{0} = k^{4} \int_{V_{1}} d^{3}x_{1} \epsilon_{p}(x_{1}) \nabla_{1} g_{0}(x_{1};x_{i}) \cdot \nabla_{1} g_{0}(x_{1};x_{j}).
$$
4 NON-LINEAR PARTIAL DERIVATIVE AND ITS DWA

First, we discuss the differences between the non-linear and linear Fréchet derivatives, and show why we should concern ourselves with the NLPD.

The full NLPD accounts for all the non-linear interactions between the perturbations at different points with different parameters, so that all the multiple scattering effects are included. For strong and large volume perturbations, the interactions between perturbations are not negligible, so that the linear Fréchet derivative will produce large errors in predicting the model perturbations. This may cause slow convergence or even divergence of the iteration process. Even for weak perturbations, if the model is perturbed on a large-scale (large volume perturbations), the non-linear effects are also proven non-negligible (see our numerical tests in Section 6).

We see that the full NLPD can be expanded around the given model as a series in terms of higher order Fréchet derivatives corresponding to multiple scattering terms in the Born series. To compute the full Born series is very time consuming or intractable. In addition, the convergence of the Born series is a serious problem. In principle, we can use any full-wave modelers, such as FD, finite element, or spectral element methods to calculate accurately the full wave solution in complex media. However, these solutions mix all waves (different types and scattering orders) together in space and time and it is hard to single out the appropriate waves for inverting velocity models. Therefore, these methods cannot provide insight for the development of inversion theory based on the NLPD. In recent decades, much progress and improvements have been made on constructing better forward propagators for surface waves in global seismology (Friederich et al. 1993; Li & Tanimoto 1993; Li & Romanowicz 1995; Friederich 1999; Romanowicz et al. 2008; Panning et al. 2009, 2012). In exploration seismology, various one-way and one-return (single backscattering) propagators have been developed and used in modelling and imaging (for a review, see, e.g. Wu et al. 2006, 2012). In order to gain the benefit provided by the Fréchet derivative series but keep the computation manageable, we turn to the DWA (multiple foreseeing single backscattering approximation) for the full series computation. We will prove that through reordering the Born series and summing up the forward-scattering subseries (renormalization), the DWA can have the same accuracy as the full Born series for weak heterogeneities, but has superior convergence property than the full Born series summation in the case of strong heterogeneities. The explicit relation between the data and the model in the De Wolf series and DWA will facilitate the development of new inversion theory. As the first attempt, we derive the NLPD for transmission tomography in smooth media.

4.1 DWA of the Fréchet series

Based on (27) and (29), the NLPD can be expanded into an infinite series. The higher order terms are formed by consecutive application of the scattering operator plus a zero-order operator. If we split the scattering operator into forward-scattering and backscattering parts

\[ \mathbf{S} = \mathbf{S}^f + \mathbf{S}^b \]  

and substitute it into the Fréchet series, we can have all combinations of higher order forward and backward derivatives. These higher order derivatives are similar to the higher order scattering terms in the Born series. The DWA in scattering series corresponds to neglecting multiple backscattering (reverberations), that is dropping all the terms containing two or more backscattering operators but keeping all the forward-scattering terms untouched. The final result is a multiple forescattering single backscattering approximation: the DWA (De Wolf 1971, 1985; Wu 1994, 1996; For a summary and review, see Wu et al. 2006, 2012). For the Fréchet derivative, let us first discuss the meaning of the operator split.
From (5) and (39), we see that the scattering operator is a local differential operator. We know that the first Fréchet derivative is the Born solution for a unit point acoustic scatterer (a point perturbation in an otherwise homogeneous medium). When the scattering operator is applied to the perturbations of media parameters, it behaves quite differently in the forward and backward directions. The scattering pattern of $\delta x$ and $\delta p$ are plotted in Figs 1(a) and (b), shown as an isotropic monopole and a two-lobed dipole, respectively. However, if we parametrize the perturbations as velocity and impedance, the corresponding scattering patterns are shown in Figs 1(c) and (d). Different parametrizations result in different scattering patterns for the perturbations (Wu & Aki 1985; Tarantola 1986). We see that velocity perturbation has only forward-scattering lobe, while impedance perturbation has only backscattering lobe. For scattering operator split, the parametrization by acoustic velocity (wave speed) and acoustic impedance is more convenient. The goal of transmission tomography is to determine the velocity perturbations of the target structure and the physical process in play is the forward scattering. Therefore, the correct handling of the multiple forward scattering is very important for transmission tomography, especially for large-scale perturbations.

For the linear Fréchet derivative, the operator split yields a forward Fréchet derivative (transmission problem)

$$\mathbf{A}'(\mathbf{m}_0) = G_0 S' g_0$$

and a backward Fréchet derivative (reflection problem)

$$\mathbf{A}''(\mathbf{m}_0) = G_0 S'' g_0.$$ 

To demonstrate the principle, we first treat the simple problem of transmission tomography in smooth media. In this case, there is no reflection and we have only forward scattering due to velocity perturbations. In the higher order Fréchet derivative (46) and the corresponding Taylor expansion of NLFD, only forescattering operator $S'$ is involved.

$$\mathbf{A}''(\mathbf{m}_0) = n! G_0 \left( S'_0 G_0 S'_{-1}, ..., G_0 S'_{l} \right) g_0.$$ 

NLFD can be calculated by summing up all the terms in (27). In this non-linear treatment, model perturbation set $\delta \mathbf{m}$ is defined as a spatial function. There are several ways to implement the calculation. One can decompose $\delta \mathbf{m}(x)$ into orthogonal bases in the space domain. We can also use the point perturbation model, but here the difference from the linear Fréchet derivative is that all the non-linear interactions are taking into account. Let us consider the kernel of NLFD acting on the perturbation $\delta \mathbf{m}(x)$ at point $x$. The $n$th-order term is

$$\frac{1}{n!} \mathbf{A}''(\mathbf{m}_0) \delta \mathbf{m} = G_0 \left[ S'_0 \delta \mathbf{m} G_0 S'_{-1} \delta \mathbf{m} ... G_0 S'_{l} \delta \mathbf{m}(x) G_0 S'_{l-2} \delta \mathbf{m} ... G_0 S'_{1} \delta \mathbf{m} \right] \delta \mathbf{m}_0 = G_0 (S' \delta \mathbf{m})^{n+1} S' G_0 (S' \delta \mathbf{m})^{-1} g_0.$$ 

Following the renormalization procedure in the DWA, we sum up all the higher order terms in the Taylor series (27) first for the multiple forescattering operators on the left-hand side of $\delta \mathbf{m}(x)$ (receiver path) and then for that on the right-hand side of $\delta \mathbf{m}(x)$ (source path) in the above equation. Since the active parameter for the forescattering renormalization is only the velocity perturbation, so we have $\delta \mathbf{m}(x) \approx \delta v(x)/v_0$ in Green’s function renormalization:

$$G_{f'} = \sum_{l=0}^{n+1} \left[ G_0 (S' \delta \mathbf{m})^l \right] \delta \mathbf{m}_0.$$ 

When $n \rightarrow \infty$ and the step length becomes infinitely small, we have

$$G_{f'} = \sum_{l=0}^{\infty} \left[ G_0 (S' \delta \mathbf{m})^l \right] \delta \mathbf{m}_0.$$ 

Under this approximation, the Taylor expansion of NLFD is changed to

$$A^{NLFD}(\mathbf{m}_0, \delta \mathbf{m}) \delta \mathbf{m}(x) = A' (\mathbf{m}_0) \delta \mathbf{m} + \frac{1}{2!} A'' (\mathbf{m}_0) (\delta \mathbf{m})^2 + \cdots + \frac{1}{n!} A^{(n)} (\mathbf{m}_0) (\delta \mathbf{m})^n + \cdots,$$

$$\approx A'*NLFD(\mathbf{m}_0, \delta \mathbf{m}) \delta \mathbf{m}(x) = G_{f'} S' \delta \mathbf{m}(x) g_f,$$

where $g_f$ and $G_{f'}$ are the forescattering renormalized Green’s function and Green’s operator, respectively. We see that even though the application of NLFD to a perturbation function under the DWA (here the forward-scattering approximation) has a simple form, but the all non-linear interactions between velocity perturbations are incorporated into $g_f$ and $G_{f'}$.

4.2 Convergence property of the Fréchet derivative series (NLFD) under the DWA

The NLFD with the DWA takes into account all the non-linear interactions of the point perturbations on the wave path under the forward-scattering approximation, so that the phase accumulation and amplitude changes from the perturbations at other points are included. Compared with the Taylor expansion (Born series) it has the stability and efficiency advantages. We first look at the stability (series convergence) problem. The original Taylor series (Born series), derived by applying the Born-Neumann iterative procedure to the Lippmann-Schwinger equation, is a Fredholm type, and has the well-known problem of limited region of convergence and slow convergence, even divergence. The iterative
procedure based on the series using gradient or Newton method will have no guarantee of converging to a correct solution. In contrast, the DWA of the NLPD $A^\text{NLPD}(m_0, \delta m)$ in (54) changes the Fredholm series into a Volterra series which is guaranteed with an absolute and uniform convergence (Schetzen 1980; Tricomi 1985). To see this, we write out explicitly the $n$th term in the infinite series (54)

$$g_f = \sum_{i=0}^{\infty} \left( G_{m} S^i \delta m \right)^f g_0 = g_0 + g_1 + \cdots + g_n + \cdots$$

where

$$g_0 = g_0(x, x_s)$$

$$\cdots$$

$$g_n = \left( G_{m} S^i \delta m \right)^f g_0$$

$$= G_{m}(x, x_s)S^i \delta m(x_{n-i}) \cdots G_{m} S \delta m(x_1)g_0(x_1, x_s)$$

$$= \int_z^r dz_1 dy_1 dx_1 g_0(x_1, x_s) S \delta m(x_1) \cdots$$

$$\times \int_z^r dz_2 dy_2 dx_2 g_0(x_2, x_s) S \delta m(x_2) \int_z^r dz_1 dy_1 dx_1 g_0(x_2, x_s) S \delta m(x_1)g_0(x_1, x_s),$$

where $z$ is taken as the forward marching direction.

We see that each term of the series in (55) is a Volterra-type integral (Schetzen 1980; Tricomi 1985). Therefore, the series in (55) is a Volterra series that converges absolutely and uniformly (Schetzen 1980; Tricomi 1985, see also Boyd et al. 1984). The same conclusion can be reached for the series of $g_f$ in (54). Therefore, the series for $A^\text{NLPD}(m_0, \delta m)$ has a guaranteed convergence.

The other advantage of NLPD in the form of (54) is the computational efficiency. Although it is in the form of an infinite series, it can be implemented efficiently by thin-slab propagator or GSP (generalized screen propagator; Wu 1994, 1996, 2003; de Hoop et al. 2012). Instead of globally summing up the scattering series, thin-slab propagator sums up the series locally step-by-step in the marching direction. In this way, it not only solves the efficiency problem, but also avoids the strong oscillation (sometimes even singular integrals) of the global summation.

5 PRELIMINARY INVESTIGATION OF NLPD APPLYING TO LS INVERSION

The misfit function (error functional) $E$ for LS inversion is defined as

$$E(m) = \frac{1}{2} \sum_i r^2_i(m) = \frac{1}{2} \mathbf{r}^T \mathbf{r}_s,$$  

(56)

where the residual function (data residual) is defined as

$$\mathbf{r} = \mathbf{d}_s - \mathbf{d}_o = (\mathbf{d}_0 + \delta \mathbf{d}) - \mathbf{d}_o = (\mathbf{d}_0 - \mathbf{d}_o) + \delta \mathbf{d},$$  

(57)

which is the scattered field in the case of waveform inversion, ‘$*$’ denotes complex-conjugated. The linear partial derivative operator (Linear Fréchet derivative) is the Jacobian

$$J_r(m_0) = \left. \frac{\partial \mathbf{r}}{\partial \mathbf{m}} \right|_{m_0} = A(m_0).$$  

(58)

For a linear perturbation model, we have

$$\mathbf{r} = (\mathbf{d}_0 - \mathbf{d}_o) + \delta \mathbf{d}$$

$$= (A(m_0) - \mathbf{d}_o) + A(m_0) \delta \mathbf{m} = \mathbf{r}(m_0) + J_r(m_0) \delta \mathbf{m}.$$  

(59)

So, the misfit function becomes

$$E = \frac{1}{2} \mathbf{r}^T \mathbf{r} = [\mathbf{r}(m_0) + J_r(m_0) \delta \mathbf{m}]^T [\mathbf{r}(m_0) + J_r(m_0) \delta \mathbf{m}].$$  

(60)

According to the optimum principle (see, e.g. Bonnans et al. 2006), the necessary condition for a local minimum of $E$ is to have a stationary point for parameter perturbation $\delta \mathbf{m}$. Taking the functional derivative of $E$ in the above equation with respect to $\delta \mathbf{m}$ and equating it to zero, we get

$$\frac{\partial E}{\partial \delta \mathbf{m}} = \frac{\partial}{\partial \delta \mathbf{m}} [\mathbf{r}(m_0) + J_r(m_0) \delta \mathbf{m}]^T [\mathbf{r}(m_0) + J_r(m_0) \delta \mathbf{m}] = J_r(m_0)^T [\mathbf{r}(m_0) + J_r(m_0) \delta \mathbf{m}]^* = 0.$$  

(61)
This leads to the familiar LS solution

\[ \delta m = \left[ J^T J \right]^{-1} J^T (r(m_0) + \delta d) . \]  

(62)

In a similar way, if we have a quadratic perturbation model

\[ r = r(m_0) + \delta (m_0) + \frac{1}{2} \delta^2 m \cdot \delta m^T . \]  

(63)

Repeating the process of finding the stationary point of the misfit function \( E \) with respect to \( \delta m \), we can have a solution of the full Newton method (Bonnans et al. 2006).

In the same principle, if we use the non-linear perturbation model with infinite orders as in the form of Taylor series, then the data residual becomes

\[ r_{NL} = (d_0 - d_{obs}) + \delta d_{NL}. \]

(64)

Note that

\[ \delta m = J_r \cdot \delta m \]

(65)

Following (61) we can find a local minimum by the stationary point method

\[ \frac{\partial E}{\partial \delta m} = \left[ \frac{\partial}{\partial \delta m} A_{NLPD}(\delta m) \delta m + A_{NLPD}(\delta m) \delta m \right]^T \left[ r(m_0) + A_{NLPD}(\delta m) \delta m \right]^* = 0 . \]  

(66)

Since \( A_{NLPD}(\delta m) \) is \( \delta m \) dependent, so the partial functional derivative with respect to \( \delta m \) has two terms, different from the linear case which has only one term. Then (67) becomes

\[ \frac{\partial E}{\partial \delta m} = \left[ \frac{\partial}{\partial \delta m} A_{NLPD}(\delta m) + A_{NLPD}(\delta m) \right]^T \left[ r(m_0) + A_{NLPD}(\delta m) \delta m \right]^* = 0 . \]  

(68)

For notation simplicity, we drop the \( \delta m \) dependence of \( A_{NLPD}(\delta m) \). Therefore,

\[ \left[ \frac{\partial}{\partial \delta m} A_{NLPD}(\delta m) + A_{NLPD}(\delta m) \right]^T A_{NLPD}^* \delta m = - \left[ \frac{\partial}{\partial \delta m} A_{NLPD}(\delta m) + A_{NLPD}(\delta m) \right]^T r^*(m_0) . \]  

(69)

We can define

\[ \frac{\partial}{\partial \delta m} A_{NLPD}(\delta m) = \delta A_{NLPD}(\delta m) \]  

(70)

as the NLDP operator variation due to medium perturbation \( \delta m \), and (69) can be written as

\[ \left[ \delta A_{NLPD} + A_{NLPD} \right]^T A_{NLPD}^* \delta m = - \left[ \frac{\partial}{\partial \delta m} A_{NLPD}(\delta m) + A_{NLPD}(\delta m) \right]^T r^*(m_0) . \]  

(71)

It is a formal solution equation and different approaches can be applied to solve it. On the right-hand side of the above equation, one term is the application of the adjoint operator of \( A_{NLPD}(\delta m) \) to the residual field, resulting in a backprojection to the model change; the second term is the application of a differential operator of \( \delta A_{NLPD} \) which accounts for the effect of the propagator change due to the medium perturbation caused by the first operation. On the left-hand side, the operator inside the braces is similar to a modified Hessian matrix.

From the series expansion (29) of \( A_{NLPD}(\delta m) \), we derive

\[ \delta A_{NLPD}(\delta m) = \frac{\partial}{\partial \delta m} A_{NLPD}(\delta m) \delta m . \]  

(72)

\[ = \frac{1}{2!} A''(m_0) \delta m + \cdots + \frac{1}{n!} A^{(n)}(m_0) \delta m^{n-1} + \cdots . \]

If we apply the DWA of \( A_{NLPD}(\delta m) \) (54)

\[ A_{NLPD} \approx A'_{NLPD} = G_f S'_{f} , \]  

(73)
then
\[ \delta A^{(\text{NLPD})} = \delta A_f^{(\text{NLPD})} = (\delta G_f) S' g_f + G_f S' (\delta g_f). \] (74)

It can be seen clearly that the updates of Green’s functions for both the source side and receiver side are included in the LS minimization procedure using the non-linear kernel. The alternating optimization scheme (e.g. Šroubek & Milanfar 2012) or the joint misfit function approach (e.g. Clément 1991; Clément et al. 2001) for optimizing iteratively both the medium parameter perturbations and the propagators can be used to implement the method.

6 NUMERICAL EXAMPLES OF NLPD AND ITS DWA

In order to show the limitations of the linear Fréchet derivative and the merits of NLPD with the DWA, we conduct a set of forward-scattering experiments using Gaussian shape velocity perturbations. The experiment configuration is similar to the case of transmission tomography. We will compare the full waveform, but the traveltimes and amplitude information can be readily extracted in these simple cases.

The source is the red star on the top (Fig. 2). The receivers are white triangles near the bottom. The velocity model is a fast Gaussian anomaly embedded in a constant background. The background velocity is 10 km s\(^{-1}\). The scales of the Gaussian anomaly are \( a = 1, 10 \) and 50 km, respectively, in the form
\[ dv(x, z) = \varepsilon v_0 \exp \left( -\frac{r^2}{2a^2} \right), \quad r = \sqrt{(x - 500)^2 + (z - 300)^2}, \quad v_0 = 10 \text{ km s}^{-1} \] (75)
where \( \varepsilon \) is the percentage perturbation, \( a \) is the scale parameter and \( v_0 \) is the background velocity. The source is a point source at \((500, 0)\). The source wavelet is a Ricker of 1.0 Hz central frequency. A line of receivers are placed at \( z = 500 \) km. For Born modelling, a simple summation integral is used (Aki & Richards 1980; Wu & Aki 1985); the FD calculations are done by a regular FD algorithm with fourth order in space and second order in time; the De Wolf modelling is realized by a GSP one-way propagator (see Wu 2003; Wu et al. 2006), because there is no reflection involved. In the calculations, we include the evanescent wave components (inhomogeneous plane waves; Aki & Richards 1980) to reduce the wavenumber truncation artifacts.

The above experimental setup can be scaled down to the length scales in exploration seismology: \( f_0 = 20 \text{ Hz}, \ v_0 = 2 \text{ km s}^{-1}, \ z = 5 \text{ km}, \ a = 10 \text{ m} (a = 0.1\lambda_0), \ 100 \text{ m} (a = 1\lambda_0) \) and \( 500 \text{ m} (a = 5\lambda_0) \), respectively. So the conclusion and discussions of the numerical experiments can be applied to both global seismology and exploration seismology.

Fig. 3 shows comparisons of synthetic seismograms calculated by different kernels [left: linear kernel (Born approximation); right: DWA of non-linear kernel] for a Gaussian ball of 20 percent perturbation in the centre, with \( a = 1.0 \text{ km} (a = 0.1\lambda_0); \top \text{ panel} \), \( 10 \text{ km} (a = 1\lambda_0; \text{ mid} \text{ panel} \) and \( 50 \text{ km} (a = 5\lambda_0; \text{ bottom} \text{ panel}) \). For comparison, we show the FD results in the mid column. We see that for the small-scale perturbation, the linear kernel (Born approximation) is valid, so good agreement can be seen between all three methods. However, for large-scale perturbations, such as \( a = 10.0 \text{ km} (a = 1\lambda_0) \) and \( a = 10.0 \text{ km} \), the deviations from the linear kernel prediction are significant. Especially in the case of \( a = 50.0 \text{ km} (a = 5\lambda_0; \text{ bottom} \text{ panel}) \), the linear kernel predictions become unacceptable. In contrast, the forward scattering renormalized non-linear kernel can give fairly accurate waveforms. In the case of transmission tomography, the Born kernel (linear Fréchet derivative) gives erroneous predictions for both the traveltimes and amplitudes. The striking large amplitude in the centre shows the famous forward-scattering catastrophe of the Born approximation. Fig. 4 shows the corresponding result for a slow (~20 per cent) Gaussian ball of \( a = 50.0 \text{ km} \). Similar conclusions can be drawn.

Figs 5 and 6 show the results for cube (square box) models of length scale 50 and 100 km, respectively. We can see similar features. The linear kernel predictions has significant errors, while the DWA of the non-linear kernel can predict well the main feature of the transmitted waveform, except for some errors for the scattered waves from the tips of the cube.

![Figure 2](http://gji.oxfordjournals.org/)

**Figure 2.** Experiment configuration resembling transmission tomography. The source is the red star on the top. The receivers are white triangles near the bottom. The scale of the Gaussian anomaly is \( a = 1 \text{ km} (a = 0.1\lambda_0), \ 10 \text{ km} (a = 1\lambda_0) \) and \( 50 \text{ km} (a = 5\lambda_0) \), respectively.
Figure 3. Waveform comparison from calculations using different kernels for a fast Gaussian ball (20 per cent at the centre) with \( a = 1 \text{ km} \) (top), 10 km (\( a = 1 \lambda_0 \); mid), 50 km (\( a = 5 \lambda_0 \); bottom); left: Born (linear \( F \)-derivative); mid: FD (finite difference method) for comparison; right: DWA of non-linear kernel. The reduced traveltime is defined as the true traveltime minus the background traveltime.

7 CONCLUSION AND DISCUSSION

In this study, we have concentrated on two themes. First is the introduction of the NLPD for the acoustic wave equation. Generalization to the elastic wave equation is straightforward. We proved that the higher order Fréchet derivatives can be realized by consecutive applications of the scattering operator plus a zero-order propagator to the source. The full NLPD is directly related to the full scattering series. The formulation of the full non-linear derivative provides a new way for deriving better approximations beyond the linear Fréchet derivative (Born approximation). Secondly, we derive the DWA (multiple forescattering and single backscattering approximation) of NLPD by scattering operator split and renormalization of the multiple scattering series. This approximation can be useful for both theoretical derivation and numerical calculation.
Figure 4. Numerical experiment for a −20 per cent (slow) Gaussian ball of $a = 50\text{ km}$ ($5\lambda$; top panel); waveforms (bottom panel) are calculated by linear kernel (left), FD (middle) and non-linear kernel (GSP; right). The reduced traveltime is defined as the true traveltime minus the background traveltime.

Through both theoretical analysis and numerical simulations, we conclude that for large-scale perturbations, such as $a = 10.0$ and 50.0 km for global seismology, and 100 and 500 m for exploration seismology in our examples (corresponding to $a = 1\lambda_0$ and $a = 5\lambda_0$, respectively), the errors of the linear kernel (Fréchet derivative) are significant. In the case of a Gaussian ball with $a = 50.0\text{ km}$ ($a = 5\lambda_0$), the kernel predictions based on the linear Fréchet derivative become unacceptable for both amplitude and phase (traveltime) in transmission tomography. In the meanwhile, the DWA of NLPD can give fairly accurate waveforms.
Now, we discuss two important issues related to the NLPD. One is the meaning of NLPD and its DWA; the other is the importance of multiple forward scattering in inversion.

7.1 Why we need NLPDs and the meaning of the DWA

In order to understand the problems with the linear $F$-derivative and the meaning of the NLPD, we consider a transmission problem, in which the observation is in the forward direction, and we assume that the perturbations form a large volume, smooth velocity (e.g. fast) anomaly. A functional derivative is an operator that predicts the data change $\delta d$ caused by a model perturbation $\delta m$. If we look at the Taylor expansion of the NLPD, it should be clear that each term in the series (including the first term, the linear $F$-derivative) is divergent (approaching infinity) when the volume goes to infinity. This is because of the neglect of multiple forward-scattering correction, ending up with in-phase linear superposition of the scattered fields. However, when we sum up all the terms, the contributions from different orders of terms cancel each other by destructive interference, resulting in a finite and correct wavefield (data). That is what happens in the real physical situation. When you put all the singular terms together and let them interact with each other, they become normal (convergent) as a whole. The divergent
behaviour of the linear $F$-derivative or the forward-scattering catastrophe of the Born modelling is a heavy price for the convenience of using the linear approximation.

From the above analysis, we know also that the divergent behaviour of the Born modelling is very different for the forward and backscatterings (for a detailed discussion, see Wu et al. 2006). The most serious problem is for forward scattering. That is why we split the scattering operator and then re-order and re-sum the Taylor series to eliminate the forward-scattering divergence. After summing up the subseries of forward scattering to infinite order, we can rearrange the series of non-linear derivative in the order of multiple backscattering operators. The new series is no longer divergent. Taking only the first-order backscattering, we reached the DWA of the NLPD.

### 7.2 Importance of multiple forward scattering in waveform inversion and the non-linear sensitivity kernel

Now, we discuss the importance of forward scattering and the need for computing the non-linear sensitivity kernel (NLSK).

In eq. (17) $A$ is a modelling operator, which usually is strongly non-linear in the case of the wave equation. However, if we linearize the equation, in discretized version, it becomes a matrix operator. Since perturbations at each point are independent from each other in the
linear formulation, the sensitivity operators for the data space and the model space are just the transpose of each other. The row vector of the modelling operator can be considered as the point response (or impulse response) function, which predicts the data change (scattered field) due a point perturbation (in space–time); The column vector relates a data change to the spatial distribution of model perturbations, and is called the sensitivity function. In the terminology of integral operator, it is called the sensitivity kernel. The sensitivity kernel defined this way is in fact the linear sensitivity kernel (LSK).

From the above analysis, we see that the predictions from the linear F-derivative can be far from the correct ones. The linear F-derivative is currently used for back projection through the gradient operator, which is the adjoint operator of the Born modelling operator. Because of the linearization, the predicted model perturbations may be far from the real perturbations causing the data changes. This is true even when the perturbations are weak but extended for a large volume. This causes the strong dependence of inversion on initial models, especially on the low-wavenumber component of the model (smooth background).

To solve or mitigate the problem of strong dependence on the initial model, researchers have turned to low-frequency data for estimating the long-wavelength component in the initial model (for a review, see Virieux & Operto 2009). For low frequencies, the validity condition for the Born approximation (linear F-derivative) is easier to be satisfied. In this way, the inversion can stay in the linear realm. However, to obtain the data of very low frequencies is very costly, and might be impossible in some situations.

From the other perspective, if we can handle the forward scattering correctly, so that the backprojection operator can update the low-wavenumber components (smooth background) during the iteration, the dependence on the initial model may be much reduced. How to derive approximate NLSKs and efficient inversion methods based on the theory of NLPPDs will be the topic of future research.

ACKNOWLEDGEMENTS

We thank Thorne Lay for careful reading the manuscript and many inspiring suggestions. We thank also Babara Romanowicz for reading the manuscript, pointing out the relevant references, and many insightful discussions. We thank the editors, especially J. Virieux, for suggestions and comments. We also thank other anonymous reviewers for useful comments. The research is supported by the W.T.OPI (wavelet transform on propagation and imaging for seismic exploration) Research Consortium at the University of California, Santa Cruz. Facility support from the W.M. Keck Foundation is also acknowledged. YZ would like to thank ERL/MIT for support.

REFERENCES


