2009-01-01

Inverse Scattering Solutions with Applications to Electromagnetic Signal Processing

Jonathan Blackledge
Technological University Dublin, jonathan.blackledge@tudublin.ie

Timo Hamalainen
University of Jyvaskyla, Finland, timo.t.hamalainen@jyu.fi

Jyrki Joutsensalo
University of Jyvaskyla, Finland, jyrki.j.joutsensalo@jyu.fi

Recommended Citation

Blackledge, J., Hamalainen, T., Joutsensalo, J.: Inverse Scattering Solutions with Applications to Electromagnetic Signal Processing. ISAST Journal of Electronics and Signal Processing, vol: 4, issue: 1, pages: 43-60, 2009. doi:10.21427/D7N037

Inverse Scattering Solutions with Applications to Electromagnetic Signal Processing

Jonathan Blackledge[1], Timo Hämäläinen[2] and Jyrki Joutsensalo[2]

***Abstract*— When a signal is recorded that has been physically generated by some scattering process (the interaction of electromagnetic, acoustic or elastic waves with inhomogeneous materials, for example), the 'standard model' for the signal (i.e. information content convolved with a characteristic Impulse Response Function) is usually based on a single scattering approximation. An additive noise term is introduced into the model to take into account a range of non-deterministic factors including multiple scattering that, along with electronic noise and other background noise sources, is assumed to be relatively weak. Thus, the standard model is based on a 'weak field condition' and the inverse scattering problem is often reduced to the deconvolution of a signal in the presence of additive noise.**

Attempts at solving the exact inverse scattering problem for equations such as the inhomogeneous Schrödinger equation in quantum mechanics and the inhomogeneous Helmholtz equation in electromagnetism often prove to be intractable, particularly with regard to the goal of implementing algorithms that are computationally stable and/or compatible with standard signal analysis methods and Digital Signal Processing 'toolkits'. This paper provides an approach to solving the multiple scattering problem for narrow side-band systems (typically, electromagnetic signal processing systems) that is compounded in the introduction of a single extra term to the standard model. The approach is based on applying certain conditions to an exact solution of the inverse scattering problem rather than applying conditions to the forward scattering problem and then inverting the (conditional) result.

***Index Terms*— Inverse Scattering, Multiple Scattering, Electromagnetism, Side-band Signals, Autocorrelation, Digital Signal Processing.**

I. INTRODUCTION

THE 'standard model' used to describe a signal $s(t)$ is based on the equation [1]

$$s(t) = p(t) \otimes f(t) + n(t)$$

where $\otimes$ denotes the convolution integral, $p(t)$ is the Impulse Response Function (IRF), $f(t)$ is the 'input' (representing the information content of the signal associated with some physical process - a scattering process, for example) and $n(t)$ is the characteristic noise of the system whose 'output' is $s(t)$. This linear and stationary model applies to a wide range of systems and applications involving the generation, interpretation and processing of digital signals and images.

Manuscript completed in March, 2009. This work is supported by the Science Foundation Ireland.

[1]Stokes Professor of Digital Signal Processing, School of Electrical Engineering Systems (http://eleceng.dit.ie/blackledge), Faculty of Engineering, Dublin Institute of Technology Ireland, Dublin 8, Ireland (email: jonathan.blackledge@dit.ie), [2]Department of Mathematical Information Technology, University of Jyväskylä, 40014 University of Jyväskylä, Finland (emails: timo.t.hamalainen@jyu.fi, jyrki.j.joutsensalo@jyu.fi).

In any application, a fundamental signal processing problem is to recover $f(t)$ from $s(t)$ given an estimate of $p(t)$ and $\Pr[n(t)]$ (where Pr denotes the Probability Density Function of the stochastic function $n(t)$). In many applications, this model can be taken to be a relatively complete and accurate representation of the physical processes that contribute to the generation of the signal $s(t)$, the noise function $n(t)$ being taken to be a combination of electronic and background noise associated with the recording system, for example. However, in applications where the signal is based on the scattering of an incident wavefield with an inhomogeneous medium, the standard model is based on applying an approximation to the scattering equation, e.g. the Helmholtz equation. The approximation is often referred to as the Born approximation (after Max Born, who first considered the approximation with regard to scattering processes in quantum mechanics through solutions to the Schrödinger equation) and considers the scattering events that contribute to the signal $s(t)$ to be based on single scattering processes only. This requires that the 'scattering model' adheres to the 'weak field' condition in which the total scattered field is considered to be a weak perturbation of the incident field in terms of some appropriate measure. In turn, depending on the complexity of the scattering model, this condition can usually be quantified in terms of physical parameters such as the wavelength λ of the incident wavefield and the scale length L of the scatterer, a basic 'standard' being that $\lambda >> L$. However, this condition is fundamentally incompatible with a basic requirement associated with systems that are designed to recover information at a resolution compatible with the scale of the wavelength, i.e. when $\lambda \sim L$. Thus, any system that is designed and engineered to 'image' an object on the scale of the wavelength of the incident field is prone to distortion due to the effect of multiple scattering, an effect that is not incorporated within the standard model. Instead, multiple scattering processes are considered to contribute to the noise function $n(t)$.

Given the standard model, the inverse (weak) scattering problem can be reduced to the process of Fourier inversion and deconvolution in the presence of additive noise. Ideally, we consider some function $p(t)$ such that $p(t) \odot p(t) = \delta(t)$ where $\odot$ denotes the correlation integral and $\delta(t)$ is the delta function, from which it follows that

$$f(t) = p(t) \odot s(t) + n'(t)$$

where $n'(t) = -p(t) \odot n(t)$.

In this paper we consider an new approach to the inverse scattering problem that is *exact* in the sense that it includes the effects of multiple scattering [2]. Inverse solutions to the multiple scattering problem have been studied for many years and a variety of solutions developed. However, in practice, except for some special circumstances, such solutions are usually incompatible with the engineering of a system associated with the methods of signal analysis upon which it is based and the Digital Signal Processing 'toolkits' that are currently available, e.g. the MATLAB DSP 'Toolbox' [3]. In this paper, we develop a model for the signal that is compounded in the equation

$$s(t) = p(t) \otimes [f(t) + \mid s(t) \mid^2 \exp(i\omega_0 t)] + n(t)$$

where the noise function $n(t)$ is not inclusive of the multiple scattering processes that are, instead, described by the term $p(t) \otimes [\mid s(t) \mid^2 \exp(i\omega_0 t)]$. This result is based on two basic conditions: (i) the total wavefield (i.e. the sum of the incident and scattered wavefields) is a phase only field; (ii) the frequency bandwith is small compared to the carrioer wave of the scattered field. In this sense, the result is not an exact inverse solution in itself but based on an exact inverse solution to which conditions (i) and (ii) are then applied. These conditions are applicable to narrow side-band pulse-echo systems, and, for example, side-band systems that exploit (linear) frequency modulated (FM) pulses. Side-band systems are a general characteristic of inverse problems associated with electromagnetic signals where the band-width is small compared with the carrier frequency. Such systems include Real Aperture Radar and Synthetic Aperture Radar and, in the latter case, a demonstration of the technique is provided by way of an application in electromagnetic signal processing.

In order to provide the necessary background material for readers who are not familiar with the inverse scattering problem, this paper is structured into the following sections: Section II considers the derivation of a one-dimensional inhomogeneous wave equation (the Helmholtz equation) from Maxwell's (macroscopic) equations which differentiates between the effect of electromagnetic waves propagating through and scattering from a non-conductive and conductive dielectric. Section III provides a background to the forward and inverse scattering problems under the weak field (Born approximation for single scattering), weak gradient (WKB approximation) and strong field (multiple scattering) conditions. Section IV introduces an exact inverse scattering approach and includes example numerical simulations for weak and strong scattering in one- and two-dimensions. The simulations considered are based on conditions appropriate for application to side-band signal processing systems where the bandwidth of the wavefield is significantly small compared to the carrier frequency which is taken to be high and Section V adopts the same approach for modelling narrow side-band pulse-echo systems. Section VI addresses signal processing methods associated with FM pulse-echo systems which provides a background to the application of the solutions to Synthetic Aperture Radar imaging as presented in Section VII.

The material presented in this paper is based on an analysis of the problem reduced to working in one-dimension. However, the approach is directly applicable to inverse scattering problems concerned with two-dimensional systems (e.g. diffraction tomography) and three-dimensional systems associated with radio, microwave, TeraHertz, photonics and electromagnetic signal and imaging processing systems in general.

II. Electromagnetic Wave Propagation in a Linear, Isotropic and Inhomogeneous Medium

We consider forward and inverse scattering solutions to the inhomogeneous Helmholtz equation for electromagnetic waves. Use of the Helmholtz equation for this purpose is based on certain limiting conditions which need to be assessed in terms of the governing equations of electromagnetism. Thus, we revisit Maxwell's equations with regard to describing the propagation of electromagnetic waves in an inhomogeneous medium.

For a medium that is inhomogeneous, isotropic and linear, Maxwell's equations can be written in the form[1], e.g. [4], [5]

$$\nabla \cdot \epsilon \mathbf{E} = \rho, \tag{1}$$

$$\nabla \cdot \mu \mathbf{H} = 0, \tag{2}$$

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t}, \tag{3}$$

and

$$\nabla \times \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t} + \mathbf{j}. \tag{4}$$

where $\mathbf{E}(\mathbf{r}, t)$ is the electric field (volts/metre), $\mathbf{H}(\mathbf{r}, t)$ is the magnetic field (amperes/metre), $\mathbf{j}(\mathbf{r}, t)$ is the current density (amperes/metre^2), $\rho(\mathbf{r}, t)$ is the charge density (charge/metre^2), $\epsilon(\mathbf{r})$ is the permittivity(farads/metre) and $\mu(\mathbf{r})$ is the permeability (henries/metre). The values of ϵ and μ in a vacuum (denoted by ϵ_0 and μ_0, respectively) are $\epsilon_0 = 8.854 \times 10^{-12}$ farads/metre and $\mu_0 = 4\pi \times 10^{-7}$ henries/metre. Equation (1) is Coulomb's law in differential form. Equation (2) is the law (in differential form) stating that magnetic fields are generated by dipoles (no magnetic monopoles). Equations (3) and (4) are Faraday's and Ampere's laws in differential form, respectively, where, the latter equation includes Maxwell's displacement current term $\epsilon \partial_t \mathbf{E}$.

In electromagnetic signal and image processing applications there are two important physical models to consider that are based on whether a material is conductive or non-conductive. The isotropy condition used here implies that there is no directional bias to the inhomogeneous characteristics of the medium, i.e. the permittivity, permeability and conductivity are scalar functions of space only. The linearity condition implies that the medium is not affected by the propagation of electromagnetic waves - the permittivity is not a function of the electric field strength, for example [6].

A. Conductive Materials

In this case, the medium is assumed to be a good conductor. A current is induced which depends on the magnitude of the electric field and the conductivity σ (siemens/metre) of the material. To a good approximation, for a material that is linear

[1]For International Systems of Units (SI) and where $\mathbf{r}$ and t denote the three-dimensional space vector $\mathbf{r} = \hat{\mathbf{x}}x + \hat{\mathbf{y}}y + \hat{\mathbf{z}}z$ and time respectively.

and isotropic, the relationship between the electric field $\mathbf{E}$ and the current density $\mathbf{j}$ is given by Ohm's law

$$\mathbf{j} = \sigma \mathbf{E}. \tag{5}$$

A good conductor is one where σ is large. By taking the divergence of equation (4) and noting that

$$\nabla \cdot (\nabla \times \mathbf{H}) = 0$$

we obtain (using equation (1) for constant ϵ)

$$\frac{\partial \rho}{\partial t} + \frac{\sigma}{\epsilon}\rho = 0.$$

The solution to this equation is

$$\rho(t) = \rho_0 \exp(-\sigma t/\epsilon), \quad \text{where} \quad \rho_0 = \rho(t = 0)$$

which shows that the charge density decays exponentially with time. Typical values of ϵ are $\sim 10^{-12} - 10^{-10}$ farads/metre and hence, provided σ is not too small (e.g. the material is not a semi-conductor), the dissipation of charge is very rapid. It is therefore physically reasonable to set the charge density to zero and, for problems involving the interaction of electromagnetic waves with good conductors, equation (1) can be approximated by

$$\nabla \cdot \epsilon \mathbf{E} = 0 \tag{6}$$

so that equation (4) becomes

$$\nabla \times \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t} + \sigma \mathbf{E}.$$

Note that in electromagnetic applications, the medium may a dielectric or a conductive dielectric. For example, in imaging the surface of the Earth using microwave radiation (e.g. Synthetic Aperture Radar), a model may be based on a 'ground truth' that is predominantly a dielectric or, in the case of imaging the sea surface, a conductive dielectric. Further, in both bases, the medium may be populated by a distribution of good conductor, e.g. metallic objects (man-made or otherwise).

B. Non-conductive Dielectrics

In this case, it is assumed that the conductivity of the medium is negligible and no current can flow so that

$$\mathbf{j} = 0$$

and equation (4) reduces to

$$\nabla \times \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t}.$$

Also, if the conductivity is zero then $\rho = \rho_0$ and, if $\rho_0 = 0$, then equation (1) becomes

$$\nabla \cdot \epsilon \mathbf{E} = 0.$$

The issue of when a material is a conductor or a dielectric is compounded in the relative importance of the terms $\mathbf{j}$ and $\epsilon(\partial \mathbf{E}/\partial t)$ in equation (4). If we consider the electric and magnetic fields to be monochromatic waves with a single frequency ω where $\mathbf{E}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r}, \omega) \exp(i\omega t)$ and $\mathbf{H}(\mathbf{r}, t) = \mathbf{H}(\mathbf{r}, \omega) \exp(i\omega t)$ so that equation (4) becomes (with $\mathbf{j} = \sigma \mathbf{E}$)

$$\nabla \times \mathbf{H}(\mathbf{r}, \omega) = (i\omega\epsilon + \sigma)\mathbf{E}(\mathbf{r}, \omega),$$

then the relative importance of the terms on the right hand side of equation (4) is determined by the magnitudes of σ and $\omega\epsilon$. If

$$\frac{\sigma}{\omega\epsilon} >> 1$$

then conduction currents dominate and the medium is a conductor. If

$$\frac{\sigma}{\omega\epsilon} << 1$$

then displacement currents dominate and the material behaves as a dielectric. When

$$\frac{\sigma}{\omega\epsilon} \sim 1$$

the material is a quasi-conductor; some types of semi-conductor fall into this category. Note that the ratio $\sigma/\omega\epsilon$ is frequency dependent and that, consequently, a conductor at one frequency may be a dielectric at another.

C. Electromagnetic Wave Equation

In many electromagnetic systems, the field that is measured is the electric field. It is therefore appropriate to use a wave equation which describes the behaviour of the electric field. This can be obtained by decoupling Maxwell's equations for the magnetic field $\mathbf{H}$. Starting with equation (3), we divide through by μ and take the curl of the resulting equation. This gives

$$\nabla \times \left(\frac{1}{\mu}\nabla \times \mathbf{E}\right) = -\frac{\partial}{\partial t}\nabla \times \mathbf{H}.$$

By taking the derivative with respect to time t of equation (4) and using Ohm's law - equation (5) - we obtain

$$\frac{\partial}{\partial t}(\nabla \times \mathbf{H}) = \epsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} + \sigma \frac{\partial \mathbf{E}}{\partial t}.$$

From the previous equation we can then write

$$\nabla \times \left(\frac{1}{\mu}\nabla \times \mathbf{E}\right) = -\epsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} - \sigma \frac{\partial \mathbf{E}}{\partial t}.$$

Expanding the first term, multiplying through by μ and noting that

$$\mu \nabla \left(\frac{1}{\mu}\right) = -\nabla \ln \mu$$

we get

$$\nabla \times \nabla \times \mathbf{E} + \epsilon\mu \frac{\partial^2 \mathbf{E}}{\partial t^2} + \sigma\mu \frac{\partial \mathbf{E}}{\partial t} = (\nabla \ln \mu) \times \nabla \times \mathbf{E}. \tag{7}$$

Expanding equation (6) we have

$$\epsilon \nabla \cdot \mathbf{E} + \mathbf{E} \cdot \nabla \epsilon = 0$$

or

$$\nabla \cdot \mathbf{E} = -\mathbf{E} \cdot \nabla \ln \epsilon.$$

Hence, using the vector identity

$$\nabla \times \nabla \times \mathbf{E} = -\nabla^2 \mathbf{E} + \nabla(\nabla \cdot \mathbf{E}),$$

from equation (7), we obtain the following wave equation for the electric field

$$\nabla^2 \mathbf{E} - \epsilon\mu \frac{\partial^2 \mathbf{E}}{\partial t^2} - \sigma\mu \frac{\partial \mathbf{E}}{\partial t}$$

$$= -\nabla(\mathbf{E}\cdot\nabla\ln\epsilon) - (\nabla\ln\mu)\times\nabla\times\mathbf{E}.$$

This equation is inhomogeneous in ϵ, μ and σ and its solutions provide information on the behaviour of the electric field in a fluctuating conductive dielectric environment. In electromagnetic (EM) signal processing problems, interest often focuses on the behaviour of the EM wavefield generated by variations in the material parameters ϵ, μ and σ. The problem is typically to reconstruct these parameters by measuring certain properties of the electric field. This is a three parameter inverse problem which, using formal methods of solution, is based on two steps: Step 1. solve for the electric field $\mathbf{E}$ given ϵ, μ and σ; Step 2. derive expressions for ϵ, μ and σ given the solution for $\mathbf{E}$. The difficulty of the task associated with Step 2 depends on the nature of the solution for $\mathbf{E}$, at least, in the general case.

D. Langevin Form of the Inhomogeneous Wave Equation

Forward and inverse scattering problems are often attempted through application of a Green's function solution [4]. In order to obtain a Green's function solution to the wave equation derived in the last section, it must be re-cast in the form of a Langevin equation [7]

$$\hat{D}\mathbf{E} = -\hat{L}\mathbf{E}$$

where $\hat{D}$ and $\hat{L}$ are homogeneous and inhomogeneous differential operators, respectively. By adding

$$\epsilon_0\frac{\partial^2\mathbf{E}}{\partial t^2} - \frac{1}{\mu_0}\nabla\times\nabla\times\mathbf{E}$$

to both sides of equation (7) and re-arranging the result, we can write

$$\nabla\times\nabla\times\mathbf{E} + \epsilon_0\mu_0\frac{\partial^2\mathbf{E}}{\partial t^2}$$
$$= -\epsilon_0\mu_0\gamma_\epsilon\frac{\partial^2\mathbf{E}}{\partial t^2} - \mu_0\sigma\frac{\partial\mathbf{E}}{\partial t} + \nabla\times(\gamma_\mu\nabla\times\mathbf{E})$$

where

$$\gamma_\epsilon = \frac{\epsilon-\epsilon_0}{\epsilon_0} \quad \text{and} \quad \gamma_\mu = \frac{\mu-\mu_0}{\mu}.$$

With equation (6), we can then use the result (valid for $\rho\sim 0$)

$$\nabla\times\nabla\times\mathbf{E} = -\nabla^2\mathbf{E} + \nabla(\nabla\cdot\mathbf{E})$$
$$= -\nabla^2\mathbf{E} - \nabla(\mathbf{E}\cdot\nabla\ln\epsilon)$$

so that the above wave equation can be written as

$$\nabla^2\mathbf{E} - \epsilon_0\mu_0\frac{\partial^2\mathbf{E}}{\partial t^2} = \mu_0\epsilon_0\gamma_\epsilon\frac{\partial^2\mathbf{E}}{\partial t^2}$$
$$+\mu_0\sigma\frac{\partial\mathbf{E}}{\partial t} - \nabla(\mathbf{E}\cdot\nabla\ln\epsilon) - \nabla\times(\gamma_\mu\nabla\times\mathbf{E}).$$

Finally, introducing the Fourier transform

$$\mathbf{E}(\mathbf{r},t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\widetilde{\mathbf{E}}(\mathbf{r},\omega)\exp(i\omega t)d\omega,$$

we can write the above wave equation in the time independent form

$$(\nabla^2+k^2)\widetilde{\mathbf{E}} = -k^2\gamma_\epsilon\widetilde{\mathbf{E}} + ikz_0\sigma\widetilde{\mathbf{E}}$$
$$-\nabla(\widetilde{\mathbf{E}}\cdot\nabla\ln\epsilon) - \nabla\times(\gamma_\mu\nabla\times\widetilde{\mathbf{E}}) \quad (8)$$

where

$$k = \frac{2\pi}{\lambda} = \frac{\omega}{c_0}, \quad c_0 = \frac{1}{\sqrt{\epsilon_0\mu_0}} \quad \text{and} \quad z_0 = \mu_0 c_0.$$

Here, λ is the wavelength and ω is the (angular) frequency of the electric wavefield. The parameter z_0 is the free space wave impedance and is approximately equal to 376.6 Ohms. The constant c_0 is the velocity at which electromagnetic waves propagate in a 'free space' characterised by dielectric parameters ϵ_0 and μ_0. In electromagnetic imaging systems, a fundamental problem is to obtain reconstructions of the parameters γ_ϵ, γ_μ and the conductivity σ by solving the inverse scattering problem posed by this equation.

III. Forward and Inverse Scattering Solutions in One-Dimension

We consider a model in which ϵ and σ are one-dimensional functions, $\mu=\mu_0$ and the electric field is plane polarized, i.e. $\widetilde{\mathbf{E}} = \hat{\mathbf{z}}u(x,k)$, so that equation (8) can be reduced to the form (the one-dimensional inhomogeneous Helmholtz equation)

$$\left(\frac{\partial^2}{\partial x^2}+k^2\right)u(x,k) = -k^2\gamma(x,k)u(x,k)$$

where, for a conductive dielectric,

$$\gamma(x,k) = \gamma_\epsilon(x) - i\frac{z_0\sigma(x)}{k}$$

and for a non-conductive dielectric

$$\gamma(x) = \gamma_\epsilon(x) = \epsilon_r - 1$$

with $\epsilon=\epsilon_0\epsilon_r$ where $\epsilon_r>1$ is the relative permittivity. The functions ϵ_r and σ are taken to be real whereas the wavefield u is a complex function with variations as a function of x and k in both amplitude $A_u(x,k)$ and phase $\theta_u(x,k)$, i.e.

$$u(x,k) = A_u(x,k)\exp[i\theta_u(x,k)]$$

On the basis of this model, the forward scattering problem is defined as: Given γ obtain a solution for u. The inverse scattering problem is: Given u derive a solution for γ.

We consider the case where the medium is a non-conductive dielectric and where $u(x,k)$, $x\in(-\infty,\infty)$ is given by the sum of an incident wavefield $u_i(x,k)$ and a scattered wavefield $u_s(x,k)$, u_i being given by a solution of

$$\left(\frac{\partial^2}{\partial x^2}+k^2\right)u_i(x,k) = 0.$$

Thus, with $u=u_i+u_s$, for a non-conductive dielectric,

$$\left(\frac{\partial^2}{\partial x^2}+k^2\right)u_s(x,k) = -k^2\gamma(x)[u_i(x,k)+u_s(x,k)]. \quad (9)$$

For most practically significant cases, it may be assumed that γ is of compact support, i.e. $\gamma(x)\exists\forall x\in[-X,X]$.

A. Weak Field Condition and the Born Approximation

Given equation (9), the weak field condition is based on assuming that the contribution of the scattered field on the right hand side of this equation is minimal. Under this condition, equation (9) is, to a good approximation, given by

$$\left(\frac{\partial^2}{\partial x^2}+k^2\right)u_s(x,k)=-k^2\gamma(x)u_i(x,k),$$

provided

$$\|u_s(x,k)\| << \|u_i(x,k)\|$$

where $\|\bullet\|$ denotes the norm of the function over x. This is the Born approximation and provides a Green's function solution for the scattered field given by [1]

$$u_s(x,k)=k^2 g(\mid x\mid,k)\otimes\gamma(x)u_i(x,k)$$

where $\otimes$ denotes the convolution integral, i.e.

$$g(\mid x\mid,k)\otimes\gamma(x)u_i(x,k)\equiv\int\limits_{-\infty}^{\infty}g(\mid x-y\mid,k)\gamma(y)u_i(y,k)dy.$$

Here, g is the 'outgoing' Green's function given by

$$g(\mid y-x\mid,k)=\frac{i}{2k}\exp(ik\mid y-x\mid)$$

which is the solution of

$$\left(\frac{\partial^2}{\partial x^2}+k^2\right)g(\mid y-x\mid,k)=-\delta(y-x).$$

Formally, this solution requires that u and $\partial u/\partial x$ are zero at $x=\pm\infty$.

B. Asymptotic Solution

For the case when the wavefield is detected in the far field, i.e. $\mid x\mid >>\mid X\mid$, we can consider an asymptotic solution of the form

$$u_s(x,k)=\lim_{x\to\infty}\frac{ik}{2}\int\limits_{-X}^{X}\exp(ik\mid x-y\mid)\gamma(y)u_i(y,k)dy$$

$$=\exp(ikx)\frac{ik}{2}\int\limits_{-X}^{X}\exp(-iky)\gamma(y)u_i(y,k)dy.$$

For an incident field $u_i(x,k)=\exp(-ikx)$, $u=u_i+u_s$ is now given by

$$u(x,k)=u_i(x,k)+u_s(x,k)=\exp(-ikx)+\widetilde{s}(k)\exp(ikx)$$

where

$$\widetilde{s}(k)=\frac{ik}{2}\widetilde{\gamma}(k),$$

$$\widetilde{\gamma}(k)=\int\limits_{-\infty}^{\infty}\gamma(x)\exp(-2ikx)dx.$$

This solution for u represents the right- and left-travelling components of the wavefield, the latter case being determined by the 'reflection coefficient' $\widetilde{s}(k)$. Note that in this asymptotic solution, the function $\gamma(x)$ maps to its Fourier transform $\widetilde{\gamma}(k)$ (ignoring scaling). Thus, at a fixed position in space $\mid x\mid >>\mid X\mid$, the function $\gamma(x)$ can only be recovered from information on its spectrum $\widetilde{\gamma}(k)$. In this sense, the inverse scattering problem is reduced to the problem of Fourier inversion. In practice, this requirement necessitates the application of pulse-echo methods which are discussed later.

C. Small Wavelength Approximation

If the wavelength of the wavefield u is small compared to characteristic variations in γ, then a suitable approximation can be introduced which provides an appropriate solution. This is known as the Wentzel-Kramers-Brillouin (WKB) method which is based on the use of an exponential type or 'Eikonal' transformation where a solution of the form $A(x,k)\exp[\pm s(x,k)]$ is considered with amplitude function $A(x,k)$ and phase function $s(x,k)$ [8]. This is analogous to a plane wave solution of the type $A\exp(\pm ikx)$.

We again consider a general solution to the one-dimensional inhomogeneous Helmholtz equation

$$\left(\frac{\partial^2}{\partial x^2}+k^2\right)u(x,k)=-k^2\gamma(x)u(x,k). \tag{10}$$

However, instead of considering the solution to be the sum of two wavefields $u=u_i+u_s$, suppose we introduce the Eikonal transform

$$u(x,k)=u_i(x,k)\exp[s(x,k)].$$

Substituting this result into equation (10) and differentiating, we obtain

$$\frac{\partial^2 u_i}{\partial x^2}+2\frac{\partial s}{\partial x}\frac{\partial u_i}{\partial x}+u_i\left(\frac{\partial s}{\partial x}\right)^2+u_i\frac{\partial^2 s}{\partial x^2}+k^2u_i=-k^2\gamma u_i.$$

If we now consider u_i to be a solution to $\partial^2 u_i/\partial x^2+k^2u_i=0$ then, after differentiating u_i and rearranging, we have

$$2ik\frac{\partial s}{\partial x}+\left(\frac{\partial s}{\partial x}\right)^2+\frac{\partial^2 s}{\partial x^2}=-k^2\gamma. \tag{11}$$

This is a nonlinear Riccatian equation for s which at first sight, appears to be more complicated than the original. However, if we introduce the condition that the wavelength $\lambda=2\pi/k$ is significantly smaller than the spatial extent over which s varies, then the nonlinear term and the second derivative can be ignored and we can write

$$2ik\frac{ds}{dx}=-k^2\gamma$$

whose general solution is (ignoring the constant of integration)

$$s(x)=\frac{ik}{2}\int^{x}\gamma(x)dx.$$

The solution for u is therefore given by

$$u(x,k)=u_i(x,k)\exp\left(\frac{ik}{2}\int^{x}\gamma(x)dx\right)$$

$$=\exp\left[ik\left(x+\frac{1}{2}\int^{x}\gamma(x)dx\right)\right].$$

This is an example of the WKB approximation. It is based on the idea that if k is large compared to the magnitudes of the terms $(\partial s/\partial x)^2$ and $\partial^2 s/\partial x^2$ then the only terms in equation (11) that matter are $2ik(\partial s/\partial x)$ and $-k^2\gamma$. In other words, if L is the characteristic scale length over which s varies, then

$$\frac{\lambda}{L} << 1.$$

The solution describes a plane wavefield whose phase kx is modified by $\frac{k}{2}\int \gamma dx$, the inverse scattering solution being given by (ignoring scaling) [9]

$$\gamma(x) \sim \frac{d}{dx}\ln\left[\frac{u(x,k)}{u_i(x,k)}\right].$$

D. Solution for Multiple Scattering

The Born approximation can be considered to be a first solution u_1 to the iterative series (for $n = 1, 2, 3, ...$)

$$u_{n+1}(x,k) = u_i(x,k) + k^2 g(\mid x \mid, k) \otimes \gamma(x) u_n(x,k)$$

where $u_0 = u_i$. The scattered field, which can be written in the form (Born series solution)

$$u_s(x,k) = k^2 g(\mid x \mid, k) \otimes \gamma(x) u_i(x,k)$$

$$+ k^4 g(\mid x \mid, k) \otimes \gamma(x)[g(\mid x \mid, k) \otimes \gamma(x) u_i(x,k)] + ...$$

can be interpreted as follows:

$u(x,k) =$incident wavefield
+
wavefield generated by single scattering events
+
wavefield generated by double scattering events
+
wavefield generated by triple scattering events
+
⋮

Each term in this series expresses the effects due to single, double and triple etc. scattering, i.e. the wavefields generated by an increasing number of interactions.

In principle, if this series converges, then it must converge to the solution. Using operator notation and writing

$$u_{n+1} = u_i + \hat{I} u_n$$

where

$$\hat{I} = k^2 \int dx g\gamma,$$

at each iteration n we consider the solution to be given by

$$u_n = u + \epsilon_n$$

where ϵ_n is the error associated with the solution at iteration n and u is the exact solution. A necessary condition for convergence is then $\epsilon_n \to 0$ as $n \to \infty$. Since

$$u + \epsilon_{n+1} = u_i + \hat{I}(u + \epsilon_n) = u_i + \hat{I}u + \hat{I}\epsilon_n$$

we can write

$$\epsilon_{n+1} = \hat{I}\epsilon_n$$

given that $u = u_i + \hat{I}u$. Thus

$$\epsilon_1 = \hat{I}\epsilon_0; \quad \epsilon_2 = \hat{I}\epsilon_1 = \hat{I}(\hat{I}\epsilon_0); \quad \epsilon_3 = \hat{I}\epsilon_2 = \hat{I}[\hat{I}(\hat{I}\epsilon_0)]; \quad ...$$

or

$$\epsilon_n = \hat{I}^n \epsilon_0$$

from which it follows that

$$\|\epsilon_n\| = \|\hat{I}^n \epsilon_0\| \leq \|\hat{I}^n\| \times \|\epsilon_0\| \leq \|\hat{I}\|^n \|\epsilon_0\|.$$

The condition for convergence therefore becomes

$$\lim_{n\to\infty} \|\hat{I}\|^n = 0.$$

This is only possible if

$$\|\hat{I}\| < 1$$

or

$$k^2 \|g(\mid x \mid, k) \otimes \gamma(x)\| < 1.$$

Defining the Euclidean norm of a complex function $f(x)$ to be

$$\|f(x)\|_2 = \left(\int \mid f(x) \mid^2 dx\right)^{\frac{1}{2}}$$

we have

$$k^2 \|g(\mid x \mid, k) \otimes \gamma(x)\|_2 \leq k^2 \|g(\mid x \mid, k)\|_2 \|\gamma(x)\|_2$$

and noting that $x \in [-X, X]$, we can write

$$kX\langle\gamma\rangle < 1$$

where

$$\langle\gamma\rangle \equiv \left(\int_{-X}^{X} \mid \gamma(x) \mid^2 dx\right)^{\frac{1}{2}}.$$

This is the condition required for the Born series to converge, the Born approximation being dependent on the weak field condition

$$\langle\gamma\rangle << \frac{\lambda}{X}.$$

E. Inverse Solution for Multiple Scattering Processes

Using operator notation, the Born series can be written as

$$u(x,k) = u_i(x,k) + \hat{I}_i\gamma(x,k) + \hat{I}_i(\gamma(x,k)\hat{I}\gamma(x,k))$$

$$+ I_i[\gamma(x,k)\hat{I}(\gamma(x,k)\hat{I}\gamma(x,k))] + ...$$

where $\gamma(x,k) = k^2\gamma(x)$ and

$$\hat{I}_i = \int dx u_i g, \quad \hat{I} = \int dx g.$$

Let $\epsilon U = u - u_i$ and

$$\gamma = \sum_{j=1}^{\infty} \epsilon^j \gamma_j.$$

Then

$$\epsilon U = \hat{I}_i[\epsilon\gamma_1 + \epsilon^2\gamma_2 + \epsilon^3\gamma_3 + ...]$$

$$+ \hat{I}_i[(\epsilon\gamma_1 + \epsilon^2\gamma_2 + \epsilon^3\gamma_3 + ...)\hat{I}(\epsilon\gamma_1 + \epsilon^2\gamma_2 + \epsilon^3\gamma_3 + ...)]$$

$$+ \hat{I}_i\{(\epsilon\gamma_1 + \epsilon^2\gamma_2 + \epsilon^3\gamma_3 + ...)\hat{I}[(\epsilon\gamma_1 + \epsilon^2\gamma_2 + \epsilon^3\gamma_3 + ...)$$

$$\hat{I}(\epsilon\gamma_1 + \epsilon^2\gamma_2 + \epsilon^3\gamma_3 + ...)]\} + ...$$

Equating terms with common coefficients ϵ, ϵ^2 etc. we have For $j = 1$:

$$U = \hat{I}_i\gamma_1; \quad \gamma_1 = \hat{I}_i^{-1}U.$$

For $j = 2$:

$$0 = \hat{I}_i\gamma_2 + \hat{I}_i(\gamma_1\hat{I}\gamma_1); \quad \gamma_2 = -\hat{I}_i^{-1}[\hat{I}_i(\gamma_1\hat{I}\gamma_1)]$$

and so on. By computing the functions γ_j using this iterative method, the scattering function γ is obtained by summing γ_j for $\epsilon = 1$. This approach provides a formal exact inverse scattering solution [10], [11] but it is not unconditional, i.e. the inverse solution is only applicable when the Born series converges to the exact scattering solution and thus when

$$\|g(\mid x \mid, k) \otimes \gamma(x)\| < 1.$$

Note that for $j = 1$, the solution for γ_1 is that obtained under the Born approximation.

IV. Exact Inverse Scattering Solutions

Given equation (10), an exact inverse scattering solution can be formulated based on defining γ as

$$\gamma(x) = \frac{u^*(x,k)}{\mid u(x,k) \mid^2}\frac{\partial^2}{\partial x^2}\left[R(x) \otimes u_s(x,k) - \frac{1}{k^2}u_s(x,k)\right].$$

where (c_1 and c_2 being arbitrary constants)

$$R(x) = \begin{cases} (c_1 - 1)x + c_2, & x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

This result is derived in Appendix I and is an extension of the trivial solution

$$\gamma(x) = -\frac{1}{k^2 u(x,k)}\left(k^2 + \frac{\partial^2}{\partial x^2}\right)u(x,k)$$

which, noting that $\gamma = \epsilon_r - 1$ can be written in the form

$$\epsilon_r(x) = -\frac{u^*(x,k)}{k^2 \mid u(x,k) \mid^2}\frac{\partial^2}{\partial x^2}u(x,k).$$

However, since $u = u_i + u_s$ where u_i is a solution of

$$\left(\frac{\partial^2}{\partial x^2} + k^2\right)u_i(x,k) = 0 \tag{12}$$

we can write

$$\gamma(x) = -\frac{u_i^*(x,k) + u_s^*(x,k)}{k^2 \mid u_i(x,k) + u_s(x,k) \mid^2}\left(k^2 + \frac{\partial^2}{\partial x^2}\right)u_s(x,k).$$

Further, we note that under the weak field condition - as discussed in Section III(A) - where we consider the equation

$$\left(\frac{\partial^2}{\partial x^2} + k^2\right)u_s(x,k) = -k^2\gamma(x)u_i(x,k),$$

the equivalent expression for γ (under the Born approximation) is

$$\gamma(x) = -\frac{u_i^*(x,k)}{k^2 \mid u_i(x,k) \mid^2}\left(k^2 + \frac{\partial^2}{\partial x^2}\right)u_s(x,k)$$
$$= -\frac{u_i^*(x,k)}{k^2}\left(k^2 + \frac{\partial^2}{\partial x^2}\right)u_s(x,k)$$

given that $u_i(x,k) = \exp(ikx)$ is a solution of equation (12).

A. Narrow Side-band Condition

In general, and in comparison to the solution under the Born approximation, the exact inverse scattering solution for γ includes $\mid u(x,k) \mid^{-2}$ which describes the (inverse square) amplitude envelop of the sum of the incident and scattered fields, i.e. with $u(x,k) = A_u(x,k)\exp[i\theta_u(x,k)]$, it is clear that $\mid u(x,k) \mid^2 = [A_u(x,k)]^2$. If u is taken to be a narrow side-band signal that is dominated by a high frequency carrier wave determined by k_0, which, in turn, is determined by a narrow band incident wave u_i, then, we can consider the condition where $\mid u(x,k) \mid^2 \sim 1 \forall x$ (in general, a constant). This condition is equivalent to considering the case where u is a phase only field $u(x,k) = \exp[i\theta_u(x,k)]$ say, and is imposed in respect of the fact that, for narrow side-band signals, the contribution of $\mid u \mid^{-2}$ to the reconstruction of γ from u is insignificant when compared to $u^*(k_0^2 + \partial_x^2)u_s$.

Strictly speaking, the condition being imposed implies that, with $u_i = \exp(\pm ik_0x)$,

$$u_i^*u_s + (u_i^*u_s)^* + \mid u_s \mid^2 = 0.$$

The principal difference between the reconstruction of γ using the Born approximation and the inverse solution considered here is compounded in the addition of a single term, i.e. for a weak field where $\|u_s\| << \|u_i\|$

$$\gamma(x) = -u_i^*(x,k_0)\left(1 + \frac{1}{k_0^2}\frac{\partial^2}{\partial x^2}\right)u_s(x,k_0) \tag{13}$$

and for a strong field where $\|u_s\| \sim \|u_i\|$

$$\gamma(x) = -u_i^*(x,k_0)\left(1 + \frac{1}{k_0^2}\frac{\partial^2}{\partial x^2}\right)u_s(x,k_0)$$
$$-u_s^*(x,k_0)\left(1 + \frac{1}{k_0^2}\frac{\partial^2}{\partial x^2}\right)u_s(x,k_0). \tag{14}$$

Note that under the strict definition of a phase only field, for $\|u_s\| \sim \|u_i\|$

$$\gamma(x) = u_i(x,k_0)u_s^*(x,k_0)$$
$$-\frac{[u_i(x,k_0) + u_s(x,k_0)]^*}{k_0^2}\frac{\partial^2}{\partial x^2}u_s(x,k_0)$$
$$= -u_i(x,k_0)u_s^*(x,k_0), \quad k_0 \to \infty$$

which yield the weak field condition. In this paper, we relax this condition and utilze equations (13) and (14) given that $\mid u_i + u_s \mid^2 \sim 1$. This is the 'key' to the results that follow.

B. Numerical Simulation I: One-dimensional Model

We consider a numerical simulation based on computing the scattered field using a forward differencing approach to equation (10). For a fixed wavenumber k_0 (which defines a continuous wave mode), the vector u_n computed over a uniformly sampled discrete array composed of N line elements, each of length Δ, is given by

$$\frac{u_{n+1} - 2u_n + u_{n-1}}{\Delta^2} + k_0^2u_n = -k_0^2\gamma_nu_n, \quad n \in [1,N]$$

which has the solution

$$u_n = \frac{u_{n+1} + u_{n-1}}{2 - \Delta^2k_0^2(1 + \gamma_n)} = \frac{u_{n+1} + u_{n-1}}{2 - \Delta^2k_0^2\epsilon_{r,n}}. \tag{15}$$

This solution requires the following iteration to be implemented

$$u_n^{m+1} = \frac{u_{n+1}^m + u_{n-1}^m}{2 - \Delta^2 k_0^2 \epsilon_{r,n}} \tag{16}$$

where u_n^1 is taken to be a discrete representation of the (unit amplitude) incident field which we define with a fixed number of periods p over $n \in [1, N]$, i.e.

$$u_n^1 = \exp\left(\frac{2\pi i p(n-1)}{(N-1)}\right).$$

A necessery condition that must be applied to equation (16) is that

$$\Delta k_0 < \sqrt{\frac{2}{\|\epsilon_{r,n}\|_\infty}}$$

where $\| \bullet \|_\infty$ defines the 'uniform norm' and

$$k_0 = \frac{2\pi p}{N}.$$

Since a solution to equation (10) is not necesserily conditional on the amplitude of the wavefield (i.e. the incident field can be $A\exp(\pm i k_0 x)$ where A is an arbitrary value), the amplitude of the array u_n^m, after M iterations is normalised on output, i.e. $u_n^M \rightarrow u_n^M / \|u_n^M\|_\infty$.

For the discrete case, equations (13) and (14) transform to (for M iterations used to compute the scattered field)

$$\epsilon_{r,n} = 1 - u_{i,n}^*[u_{s,n} + (\Delta k_0)^{-2} u_{s,n} \otimes (1, -2, 1)]$$

$$= 1 - (u_n^1)^*[(u_n^M - u_n^1) + (\Delta k_0)^{-2}(u_n^M - u_n^1) \otimes (1, -2, 1)] \tag{17}$$

and

$$\epsilon_{r,n} = 1 - u_n^*[u_{s,n} + (\Delta k_0)^{-2} u_{s,n} \otimes (1, -2, 1)]$$

$$= 1 - (u_n^M)^*[(u_n^M - u_n^1) + (\Delta k_0)^{-2}(u_n^M - u_n^1) \otimes (1, -2, 1)] \tag{18}$$

respectively where $\otimes$ now denotes the discrete convolution sum. Figures 1-3 show comparisons between the numerical results given by equations (17) and (18), illustrating the superiority of the inverse scattering solution considered over that given by the weak field solution for the case when $\Delta k_0 = 0.01$. These results are based on applying the initial solution $u_n^1 = \sin[2\pi i p(n-1)/(N-1)]$ and Hilbert transforming the output arrays before computation of equations (17) and (18) through use of the MATLAB function *hilbert*. The profile of the reconstruction for $\epsilon_{r,n}$ based on equation (18) is preserved inclusive of characteristic 'ringing' due to the discontinuities associated with the model for $\epsilon_{r,n}$. In comparison, the weak field solution given by equation (17) has a relatively narrow dynamic range and provides a poor reconstruction particularly with regard to resolving the discontinuities associated with $\epsilon_{r,n}$.

C. Numerical Simulation II: Two-dimensional Model

Further appreciation of the difference between these weak and strong field solutions is realised in Figure 4. Here, we have considered an iterative forward scattering solution to the equation

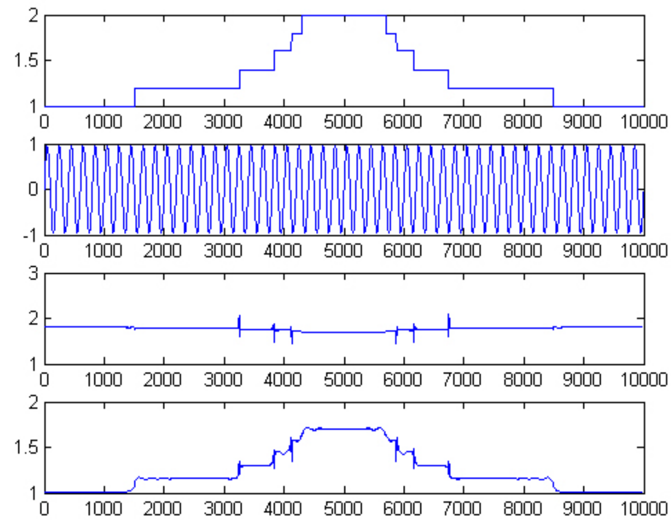

Fig. 1. Comparison between the weak and strong field inverse scattering solutions for the case when $N = 10000$, $M = 100$, $\Delta k_0 = 0.01$ with $p = 50$. From top to bottom: Relative permittivity function model $\epsilon_{r,n}$; real part of wavefield computed via equation (16); inverse solution (real part) computed using equation (17); inverse scattering solution (real part) computed using equation (18).

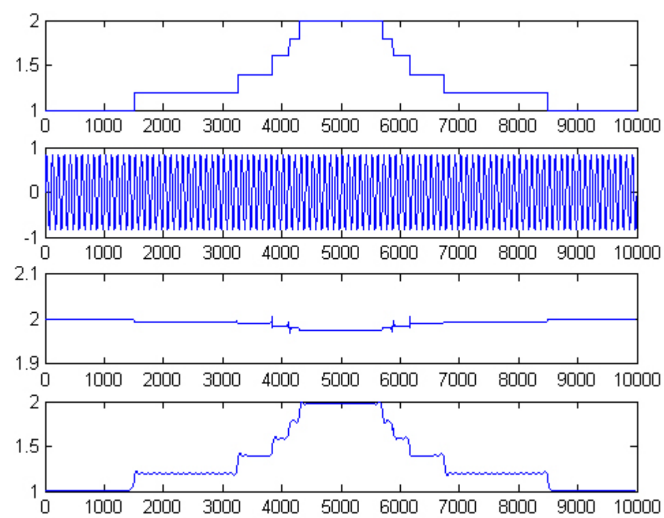

Fig. 2. Comparison between the weak and strong field solutions for the case when $N = 10000$, $M = 100$, $\Delta k_0 = 0.01$ with $p = 100$. The descriptions of each plot follow those as given in Figure 1.

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + k_0^2\right) u(x, y, k_0) = -k_0^2 \gamma(x) u(x, y, k_0)$$

based on the application of a regular (square) grid of size N^2, i.e.

$$\frac{u_{n+1,m} + u_{n-1,m} + u_{n,m+1} + u_{n,m-1} - 4u_{n,m}}{\Delta^2} + k_0^2 u_{n,m}$$

$$= -k_0^2 \gamma_{n,m} u_{n,m}, \quad n \in [1, N], \quad m \in [1, N]$$

with iterative solution

$$u_{n,m}^{q+1} = \frac{u_{n+1,m}^q + u_{n-1,m}^q + u_{n,m+1}^q + u_{n,m-1}^q}{4 - \Delta^2 k_0^2 \epsilon_{r,n,m}}$$

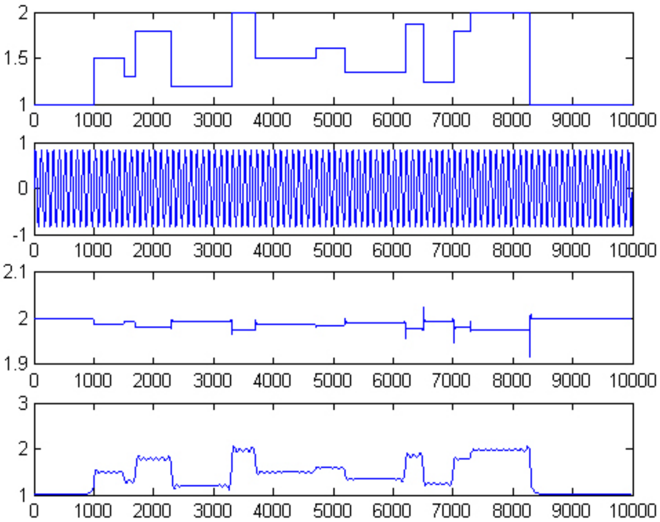

Fig. 3. Comparison between the weak and strong field inverse scattering solutions for the case when $N = 10000$, $M = 100$, $\Delta k_0 = 0.01$ with $p = 50$ and a non-symmetric model of the relative permittivity $\epsilon_{r,n}$. The descriptions of each plot follow the same as those given in Figure 1.

under the condition that

$$\Delta k_0 < \frac{2}{\sqrt{\|\epsilon_{r,n,m}\|_\infty}}.$$

with initial condition

$$u^1_{n,m} = \sin[2\pi ipn(m-1)/(N-1)]$$

After M iteration, $u^M_{n,m}$ is normalised and the Hilbert transforms are taken of $u^M_{n,m}$, $u^1_{n,m}$ and $u^M_{n,m} - u^1_{n,m}$ over m for all values of n (to compute analytic signals $\forall n$) prior to the computation of the weak field reconstruction

$$\epsilon_{r,n,m} = 1 - (u^1_{n,m})^*(u^M_{n,m} - u^1_{n,m}) - (\Delta k_0)^{-2}(u^1_{n,m})^* \times \left[(u^M_{n,m} - u^1_{n,m}) \otimes_2 \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix} \right] \tag{19}$$

and the strong field reconstruction

$$\epsilon_{r,n,m} = 1 - (u^M_{n,m})^*(u^M_{n,m} - u^1_{n,m}) - (\Delta k_0)^{-2}(u^M_{n,m})^* \times \left[(u^M_{n,m} - u^1_{n,m}) \otimes_2 \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix} \right] \tag{20}$$

where $\otimes_2$ denotes the two-dimensional convolution sum.

V. Pulse-Echo Mode Signals

The example numerical results presented in the previous section have been introduced to illustrate the characteristic differences between the weak and strong field solutions given the exact inverse scattering appraoch that we have considered. However, in practice, these inverse solutions have little practical value with regard to engineering a system designed to recover $\epsilon_r(x)$ from information on the scattered field measured 'outside' the scatterer, i.e. when $\mid x \mid > \mid X \mid$. This is because the solutions considered so far require that the scattered field

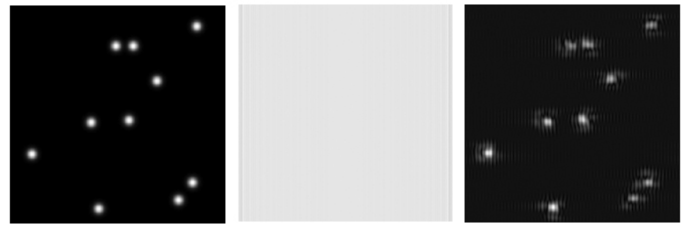

Fig. 4. Comparison between the two-dimensional weak field and strong field inverse scattering solutions for the case when $N = 500$, $M = 100$, $\Delta k_0 = 0.01$ with $p = 64$ for the (left-to-right) CW case. Left: graded point scattering model for permittivity function $1 < \epsilon_{r,n,m} < 2$; centre: inverse solution (absolute value) computed using equation (19); right: inverse scattering solution (absolute value) computed using equation (20). Note that in each case, the numerical fields have been normalised for the purpose of generating grey level image displays.

is known for $x \in [-X, X]$. Practically applicable systems typically measure the scattered field at a fixed point x_0 in the far field (based on an asymptotic solution where $x \to \infty$) by recording the spectrum $u_s(x_0, k)$ at x_0 over a range of values of k. Theoretically, this requires that the inverse problem is based on an asymptotic solution of the type derived in Section III(B). In practice, this requires the application of pulse-echo mode methods.

Pulse-echo methods involve the emission of a pulse and a recording of the back-scattered wavefield (echo). This approach is consistent with the physical nature a system if: (i) the scattering function is a one-dimensional function; (ii) the incident wavefield is a 'pencil-line beam'. However, since all physical systems are intrinsically three-dimensional, this model is idealised. Nevertheless, a variety of electromagnetic information and imaging systems can be 'cast' in terms of problems involving layered materials (e.g. the response of light, radio and microwaves to layered dielectric materials including the propagation of electromagnetic waves along transmission lines such as an optical fiber).

Pulse-echo mode signal analysis systems typically involve the utilization a pulse that is emitted from a source in which the 'time history' of the back-scattered field is recorded by a receiver which is placed in the vicinity of the location of the source. By moving both the source and receiver and repeating this type of experiment, an image can be built up based on the nature of the reflected pulse at different source locations. The resolution that can be obtained with pulse-echo experiments of this type is normally determined by the length of the pulse that is used and the width of the beam. To obtain high resolution, a short pulse and narrow 'pencil beam' are required. In some cases, the lateral resolution can be synthesized using synthetic aperture methods. Also, in some cases (e.g. Real and Aynthetic Aperture Radar), it is possible to modulate the frequency of the pulse thereby providing a method of reconstruction in which the resolution improves with pulse length (as discussed in Section VI).

In a pulse-echo experiment, the receiver monitors the time history of the reflected waves (the echo). After a short delay (which depends on the distance of the source from the scatterer and the speed at which the pulse propagates), the first reflections are received followed by a series or 'train' of other reflections from the surface or the interior of the material. This process continues until all the energy of the pulse has been dissipated. In electromagnetic systems (coherent time-history resolved), the scattered electric field is typically measured by the way in which it induces a time varying voltage in the antenna.

A. Base-band and Side-band Systems

Pulse-echo systems are based on using wavefields at frequencies where the time variations of the wavefield can be recorded to produce a set of signals. With electromagnetic systems, the frequency range is from kHz - GHz. Apart from synthetic aperture imaging systems, most pulse-echo based systems provide partially coherent (in time) data. There is one important difference between them, however, which is concerned with whether or not the pulse is a side-band of base-band wavefield. Baseband pulses are multi-frequency wavefields with a frequency range from 0-Ω Hz where Ω is the bandwidth of the pulse. Sideband pulses are fields with a bandwidth of Ω but with a central frequency of ω_0 (the carrier frequency of the pulse) where $\omega_0 >> \Omega$. In side-band systems, it is usual to demodulate back to base-band and then digitize the resulting signal(s). Sideband systems are a natural consequence of utilizing high frequency radiation sources where the pulse length is much longer than the wavelength. Thus, suppose a pulse of radiation denoted by $p(t)$ has a spectrum $\widetilde{p}(\omega)$ where $\mid \omega \mid \leq \Omega$. Then, for a base-band system we have

$$p(t) \leftrightarrow \widetilde{p}(\omega)$$

but for a side-band system

$$p(t)\exp(-i\omega_0 t) \leftrightarrow \widetilde{p}(\omega) \otimes \delta(\omega+\omega_0) = \widetilde{p}(\omega+\omega_0)$$

where $\leftrightarrow$ denotes the transformation from time to frequency space. In the latter case, there are many oscillations of the field over the duration of the pulse and hence we have $p(t)\exp(-i\omega_0 t)$ rather than just $p(t)$. Under the Born approximation, noting that $\omega = k/c_0$, for an incident field $u_i(t,\omega) = \exp(-i\omega t)$ the reflection coefficient - as derived in Section III(B) - is

$$\widetilde{s}(\omega) = \frac{i\omega}{2c_0}\int_{-\infty}^{\infty}\gamma(t)\exp(-2i\omega t)dt.$$

where

$$\widetilde{\gamma}(\omega) = \int_{-\infty}^{\infty}\gamma(t/2)\exp(-i\omega t)dt.$$

Thus, for an incident field with a spectrum given by $\widetilde{p}(\omega)$, i.e. an incident field given by $u_i(t,\omega) = \widetilde{p}(\omega)\exp(-i\omega t)$, it follows that the reflection coefficient is given by

$$\widetilde{s}(\omega) = \frac{i\omega}{4c_0}\widetilde{p}(\omega)\widetilde{\gamma}(\omega)$$

or using the convolution theorem,

$$s(t) = p(t) \otimes f(t)$$

where

$$p(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\widetilde{p}(\omega)\exp(i\omega t)d\omega$$

and $f(t)$ is the Impulse Response Function (IRF) given by (where t is the two-way travel time and we ignore scaling by $1/4c_0$)

$$f(t) = \frac{d}{dt}\gamma(t).$$

However, for a side-band band system

$$\widetilde{s}(\omega) = i\omega\widetilde{p}(\omega-\omega_0)\widetilde{\gamma}(\omega) \simeq i\omega_0\widetilde{p}(\omega)\widetilde{\gamma}(\omega+\omega_0), \quad \Omega << \omega_0$$

so that

$$s(t) = p(t) \otimes f(t)$$

where

$$f(t) = i\omega_0\gamma(t)\exp(-i\omega_0 t).$$

For a conductive dielectric, the (equivalent) IRFs are given by

$$f(t) = \frac{d}{dt}\gamma(t) + z_0\sigma(t)$$

and

$$f(t) = [i\omega_0\gamma(t) + z_0\sigma(t)]\exp(-i\omega_0 t)$$

for a base-band and side-band system respectively. Irrespective of whether a base-band or side-band pulse is used, the signal model

$$s(t) = p(t) \otimes f(t)$$

is based on the weak field condition. The 'standard' approach is to extended this model to the form

$$s(t) = p(t) \otimes f(t) + n(t)$$

where the noise term $n(t)$ is taken to include all effects that do not conform to the model used, including multiple scattering processes. The inverse scattering problem is thus reduced to the problem of deconvolution in the presence of additive noise. This is a fundamental problem in both signal and image processing.

B. Inverse Solution for Side-Band Pulse-Echo Systems

For a narrow side-band system with carrier frequency $k_0 >>\mid k \mid$, we consider the equation

$$\gamma(x) = -\frac{u^*(x,k)}{k_0^2}\left(k_0^2 + \frac{\partial^2}{\partial x^2}\right)u_s(x,k)$$

which, for a weak field where $\|u_s\| << \|u_i\|$ reduces to

$$\gamma(x) = -\frac{u_i^*(x,k)}{k_0^2}\left(k_0^2 + \frac{\partial^2}{\partial x^2}\right)u_s(x,k).$$

These results apply for all values of $x \in (-\infty,\infty)$ but γ is, as usual, considered to be of compact support $\gamma(x)\exists\forall x \in [-X,X]$. Under the Born approximation, for $x \rightarrow \infty$, a mapping is obtained between the scattered field u_s and γ that is based on the Fourier transform $\widetilde{\gamma}(k)$ of $\gamma(x)$, i.e.

$$u_s(x,k) = \frac{ik_0}{2}\exp(ik_0 x)\widetilde{\gamma}(k), \quad \mid k \mid << k_0$$

which is a solution of

$$\gamma(x) = -\frac{u_i^*(x,k)}{k_0^2}\left(k_0^2 + \frac{\partial^2}{\partial x^2}\right)u_s(x,k).$$

This result suggests taking the Fourier transform of γ where

$$\gamma(x) =$$

$$-[u_i^*(x,k_0) + u_s^*(x,k_0)]\left(1 + \frac{1}{k_0^2}\frac{\partial^2}{\partial x^2}\right)u_s(x,k_0)$$

giving

$$\widetilde{\gamma}(k) =$$

$$-[\widetilde{u}_i^*(k,k_0) + \widetilde{u}_s^*(k,k_0)] \odot \left(1 - \frac{|\,k\,|^2}{k_0^2}\right)\widetilde{u}_s(k,k_0)$$

$$= -[\widetilde{u}_i^*(k,k_0) + \widetilde{u}_s^*(k,k_0)] \odot \widetilde{u}_s(k), \quad |\,k\,| << k_0$$

where $\widetilde{\gamma}$, $\widetilde{u}_i^*$, $\widetilde{u}_s^*$ and $\widetilde{u}_s$ are the Fourier transforms of γ, u_i^*, u_s^* and u_s respectively and $\odot$ denotes the correlation integral. Noting that, for $u_i(x,k_0) = \exp(-ik_0x)$, $\widetilde{u}_i^*(k,k_0) = 2\pi\delta(k_0 - k)$ we have

$$\widetilde{\gamma}(k) = 2\pi\widetilde{u}_s(k - k_0) - \widetilde{u}_s^*(k) \odot \widetilde{u}_s(k), \quad |\,k\,| << k_0$$

or (ignoring scaing by 2π)

$$\widetilde{u}_s(k) =$$

$$\widetilde{p}(k)\widetilde{\gamma}(k + k_0) + \widetilde{p}(k)[\widetilde{u}_s^*(k + k_0) \odot \widetilde{u}_s(k + k_0)], \quad \forall k$$

where $\widetilde{p}(k)$ is some lowpass filter with a bandwidth significantly less than k_0. Thus, upon takling the inverse Fourier transform, we obtain an expression for the (demodulated) scattered field u_s given by

$$u_s(x,k_0) = p(x) \otimes [\gamma(x)\exp(-ik_0x)]$$

$$+p(x) \otimes [|\,u_s(x,k_0)\,|^2 \exp(-ik_0x)]$$

where $p(x)$ is the inverse Fourier transform of $\widetilde{p}(k)$. Here, $p(x)$ described the bandlimited pulse that is incident on a layered dielectric described by $\gamma(x)$. The first term is a description for the weak scattered field and the second term describes the effects of multiple scattering. In terms of the 'standard model' for a stationary signal $s(t)$ as a function of time t where

$$s(t) = p(t) \otimes f(t) + n(t) \tag{21}$$

it is now clear that the Impulse Response Function (IRF) f is given by

$$f(t) = \gamma(t)\exp(-i\omega_0 t)$$

and the noise $n(t)$ is given by

$$n(t) = p(t) \otimes [|\,s(t)\,|^2 \exp(-i\omega_0 t)]$$

where ω_0 is the angular (carrier) frequency. Note that in practice, $n(t)$ will include additional background noise and in this sense, we have extract a component of the noise term that is attributed to multiple scattering effects, albeit under the conditions that: (i) $|\,u_i + u_s\,|^{-2} \sim 1$; (ii) $|\,k\,| << k_0$.

VI. Linear FM Signal Processing

A linear FM 'chirp' signal, which is taken to be of compact support $t \in [-T/2, T/2]$, is given by (in complex form and for a unit amplitude)

$$p(t) = \exp(i\alpha t^2), \quad |\,t\,| \leq \frac{T}{2}$$

where α is a constant which defines the 'chirp rate' and T is the length of the signal. The phase of this signal is given by αt^2 (i.e. it has a quadratic phase factor) and its instantaneous frequency is therefore given by

$$\frac{d}{dt}(\alpha t^2) = 2\alpha t$$

which varies linearly with time t. The frequency modulations are linear, the signal therefore being referred to as a 'linear' FM pulse.

Consider a signal generated by the reflection of a linear FM pulse. Application of the weak field condition yields the following single scattering model:

$$s(t) = \exp(i\alpha t^2) \otimes f(t), \quad |\,t\,| \leq \frac{T}{2}.$$

Using a pulse of this type provides a simple but effective way of recovering the IRF by correlating the signal with the complex conjugate of the pulse to give the reconstruction

$$\hat{f}(t) = \exp(-i\alpha t^2) \odot \exp(i\alpha t^2) \otimes f(t), \quad |\,t\,| \leq \frac{T}{2}.$$

Evaluating the correlation integral, we have

$$\exp(-i\alpha t^2)\odot\exp(i\alpha t^2) = \int_{-T/2}^{T/2} \exp[-i\alpha(\tau+t)^2]\exp(i\alpha\tau^2)d\tau$$

$$= \exp(-i\alpha t^2)\int_{-T/2}^{T/2}\exp(-2i\alpha t\tau)d\tau$$

$$= T\exp(-i\alpha t^2)\text{sinc}(\alpha T t)$$

and hence

$$\hat{f}(t) = T\exp(-i\alpha t^2)\text{sinc}(\alpha T t) \otimes f(t).$$

A further simplification can now be made to the result for $\hat{f}$ which allows the exponential term to be ignored. In particular, if we consider a 'long pulse' in which $T >> 1$, then

$$\cos(\alpha t^2)\text{sinc}(\alpha T t) \simeq \text{sinc}(\alpha T t)$$

and

$$\sin(\alpha t^2)\text{sinc}(\alpha T t) \simeq 0$$

so that

$$\hat{f}(t) \simeq T\text{sinc}(\alpha T t) \otimes f(t).$$

This simplification, under a condition that is not only practically applicable, but systems desirable, especially with regard to electromagnetic pulse generation[2] allows the result for $\hat{f}$

[2] A 'long pulse' provides greater signal energy and thus, the recorded data has a higher Signal-to-Noise Ratio.

to be easily analysed in Fourier space. Using the convolution theorem we can write (ignoring scaling by π/α)

$$\hat{F}(\omega) = \begin{cases} F(\omega), & \mid \omega \mid \leq \alpha T; \\ 0, & \mid \omega \mid > \alpha T. \end{cases}$$

which describes $\hat{f}$ as being a band-limited version of f (given that f is not a band-limited function) where the bandwidth is determined by αT.

In the presence of additive noise, the result is

$$\hat{f}(t) \simeq T\text{sinc}(\alpha T t) \otimes f(t) + \exp(-i\alpha t^2) \odot n(t).$$

The correlation function produced by the correlation of $\exp(-i\alpha t)$ with a noise function $n(t)$ is taken to be relatively low in amplitude on the grounds that $n(t)$ will not normally have features that match those of a (complex) chirp so that

$$\|T\text{sinc}(\alpha T t) \otimes f(t)\| >> \| \exp(-i\alpha t^2) \odot n(t)\|$$

The function $\hat{f}$ is then interpreted as being a band-limited reconstruction of f with high Signal-to-Noise Ratio (SNR).

For a non-conductive dielectric, and for a side-band FM system, the results discussed above imply that, under the single scattering approximation

$$s(t) = p(t) \otimes f(t)$$

with inverse scattering estimate

$$\hat{f}(t) = p^*(t) \odot s(t)$$

for a side-band system with $p(t) = \exp(i\alpha t^2)$. Here, $s(t)$ is taken to be the (complex) analytic signal obtained by taking the Hilbert transfrom of the real signal $\text{Re}[s(t)]$. For side-band systems this is typically undertaken by quadrature detection during demodulation from side-band to base-band [1]. The strong field inverse scattering solution derived in Section V(B) and compunded in equation (21), implies that for the same (side-band) system

$$\hat{f}(t) = T\text{sinc}(\alpha T t) \otimes f(t)$$
$$-T\text{sinc}(\alpha T t) \otimes [\mid s(t) \mid^2 \exp(-i\omega_0 t)]$$

VII. Applications to Synthetic Aperture Radar

Synthetic Aperture Radar (SAR) is a side-band pulse-echo system which utilizes the response of a scatterer as it passes through the radar beam to synthesize the lateral (azimuth) resolution [13], [14]. This allows relatively high resolution images to be obtained at a long range. The antenna emits a pulse of microwave radiation toward the ground and the return signal is recorded at fixed time intervals along the flight path.

SAR is a peak power limited system. In other words it operates at the maximum power available. The energy of the system is therefore given by Energy = Peak Power $\times$ Time. In order to transmit a microwave field with enough energy to establish a measurable return, the duration of the pulse must be made relatively long. The length of this pulse is large compared to the wavelength and, hence the system is based on application of a side-band spectrum. If a simple on/off pulse is emitted then the characteristic spectrum is a narrow-band sinc function, the frequency content of this type of pulse not being broad enough to obtain adequate range resolution. For this reason, a frequency sweep or 'chirp' is applied over the duration of the pulse usually based on linear frequency modulation as discussed in Section VI. Even with a frequency sweep applied to it, the pulse has a very narrow frequency band. In other words, the energy of the pulse is concentrated near to the carrier frequency and can be modelled as (in complex form)

$$p(\tau) = \exp(ik_0\tau)\exp(i\alpha\tau^2), \quad -T/2 \leq \tau \leq T/2$$

where T is the pulse length, $\tau = tc_0$, α is the quadratic chirp rate ($\times c_0^{-2}$) and k_0 is the carrier wave number (angular carrier frequency = $c_0 k_0$). Note that in reality the pulse is of course not a complex but a real valued function of time. It is one of a number of different types of coded pulses that can, in principle, be used and correlated with the return signal. However, it is used extensively in radar systems for generating range resolution and it can be implemented comparatively easily. SARs are side-band systems that utilize values of k_0 and α where $k_0 >> 1$ and $\alpha << 1$.

The spectrum of a side-band FM pulse is given by

$$P(k) = \int_{-T/2}^{T/2} \exp(ik_0\tau)\exp(-i\alpha\tau^2)\exp(-ik\tau)d\tau$$

$$= \sqrt{\frac{\pi}{2\alpha}}\left[K\left(\frac{\alpha T + u}{\sqrt{2\pi\alpha}}\right) + K\left(\frac{\alpha T - u}{\sqrt{2\pi\alpha}}\right)\right]\exp(-iu^2/4\alpha)$$

where $u = k + k_0$ and

$$K(x) = \int_0^x \exp(i\pi x^2/2)dx = C(x) + iS(x)$$

with real and imaginary parts being defined in terms of the Fresnel integrals

$$C(x) = \int_0^x \cos\frac{\pi}{2}x^2 dx$$

and

$$S(x) = \int_0^x \sin\frac{\pi}{2}x^2 dx$$

respectively. Note that the bandwidth of the pulse is determined by the value of αT. With microwave systems this is typically two to three orders of magnitude smaller than the carrier wavenumber k_0.

A. Range Processing

Consider a single point scatterer which reflects a replica of the emitted pulse. At the receiver the return signal is coherently mixed down to base-band (i.e. frequency demodulated). In practice, the field that is actually measured is not a complex but a real valued signal. The imaginary part of this signal is obtained using a quadrature filter which demodulates the return signal using $\sin(k_0\tau)$ instead of $\cos(k_0\tau)$. This is equivalent to computing the Hilbert transform of the signal after demodulation to base-band. The ideal complex or analytic signal that is obtained after demodulation is then given by

$$\exp(i\alpha\tau^2), \quad -T/2 \leq \tau \leq T/2.$$

At this stage, the range resolution is determined by the pulse length T. By correlating the signal with its complex reference function $\exp(-i\alpha\tau^2)$ - a process known in SAR as range compression - we obtain[3]

$$R(\tau) \simeq T\text{sinc}(\alpha T\tau), \quad T >> 1.$$

By defining the range resolution to be the distance between the first two zeros of the sinc function which occur when $\alpha T\tau = \pm\pi$ the range resolution is given by $2\pi/\alpha T$ metres. Note that as the value of αT increases, the range resolution improves. For example, with a 20 μs pulse, T = 6 km and, with α = 7 $\times$ 10^{-4} m^{-2}, the range resolution is approximately 1.5 metres [12]

B. Azimuth Processing

As the radar travels along its flight path (repeatedly emitting a linear FM pulse and recording the back-scattered electric field scattered by the ground), the radar beam ‘illuminates’ an area of the ground which depends upon the grazing angle, its angle of divergence and the range at which the radar operates. The width of the beam in azimuth is given by $R\tan(\beta/2)$ where R is the range and β is the angle of divergence of the beam. For a SAR system, this value corresponds to the maximum length of the synthetic aperture.

In practice, $\beta \sim 1^0$ and so the width of the beam is approximately given by $R\beta/2$. This value determines the resolution in azimuth of a Real Aperture Radar (RAR). At a range of of say 50 km with $\beta = 1^0$, this resolution is just under a kilometre which is very poor. Hence real aperture radar images are only useful when short ranges are involved. The whole point of SAR is to obtain high resolution at long ranges which is accomplished by studying the response of the radar in azimuth as it passes by a scatterer and synthesizing the azimuth resolution.

At relatively large distances from the location of the source, the wavefield has a curvature which is parabolic. This is characteristic of a wavefield in the intermediate or Fresnel zone as illustrated in Figure 5. Using the geometry shown in Figure 5, we have

$$(R+\delta R)^2 + y^2 = R^2$$

or

$$2R\delta R + (\delta R)^2 + y^2 = 0.$$

If the angle of divergence of the beam is small, then δR is much less than 1. We can then ignore the nonlinear term $(\delta R)^2$ leaving the equation

$$2\delta R = -\frac{y^2}{R}.$$

[3]As discussed in Section VI

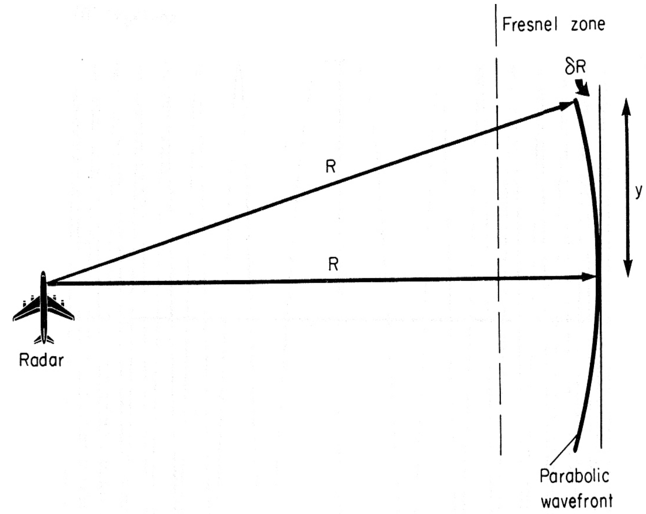

Fig. 5. By the time the wavefield emitted by the radar has reached a point scatterer, the curvature of the wavefront is parabolic. Scattering occurs in the Fresnel zone. This gives a phase history that is proportional to the square of the distance moved in azimuth.

A simple plane wave travelling along the two-way path length $2(R+\delta R)$ can therefore be written as

$$\exp[(-2ik_0(R+\delta R)] = \exp(-2ik_0R)\exp(-2ik_0\delta R)$$

where k_0 is the wavenumber. This wave has two phase factors. The first phase $2k_0R$ is constant but the second phase $2k_0\delta R$ is a function of y and is given by k_0y^2/R. Hence, as the radar moves past the scatterer a quadratic phase shift takes place. If we denote the width of the beam at R by L, then the complex azimuth response of the radar can be written as

$$\exp(ik_0y^2/R), \quad -L/2 \leq y \leq L/2$$

where $-L/2$ is the point where the scatterer enters the beam and $L/2$ is the position where the scatterer leaves the beam. This response can be clearly observed with real data when the radar passes by a strong scatterer with a large radar cross-section which can be used to calibrate the system in practice. Note that if the beamwidth is small, then this effect is not significant and that only if k_0 is sufficiently large will the effect be observed. In other words, the wavelength of the wavefield must be small compared with the range which necessitates the use of a narrow side-band pulse-echo system.

The azimuth response of the radar is the same as the response in range to a linear FM pulse. Hence, by utilizing the principles of range compression, azimuth resolution can by ‘synthesized’. This is known as azimuth compression and, like range compression, is based on correlating the complex function $\exp(ik_0y^2/R)$ with its complex reference function $\exp(-ik_0y^2/R)$ over the beam width L. Hence, the azimuth compressed signal is given by

$$A(y) = \int\limits_{-L/2}^{L/2} \exp[-ik_0(y+u)^2/R]\exp(ik_0u^2/R)du.$$

$$\simeq L\text{sinc}(k_0Ly/R), \quad L >> 1.$$

For both azimuth compression and range compression, the correlation between the return signal and its reference may be computed in Fourier space using the correlation theorem and a Fast Fourier Transform (FFT).

By defining the azimuth resolution to be the distance between the first zeros of the sinc function which occur when $k_0Ly/R = \pm\pi$, the azimuth resolution is given by $2\pi R/k_0L = 2\pi/\beta k_0$. However, the microwave antenna (i.e. essentially the horn at the end of the microwave transmission line) acts like a rectangular aperture which diffracts an otherwise collimated beam of microwaves (generated by the transmission line). The diffraction pattern can be determined from the Kirchhoff diffraction integral for fixed $k = k_0$, that for an aperture of width w, say, and an incident plane wave propagating in the z-direction, is given by (ignoring scaling)

$$\int_{-w/2}^{w/2}\int_{-w/2}^{w/2} \exp(-ik_0x_0x/z_0)\exp(-ik_0y_0y/z_0)dxdy$$

$$= 4\frac{\sin(k_0x_0w/2z_0)}{k_0x_0/z_0}\frac{\sin(k_0y_0w/2z_0)}{k_0y_0/z_0}.$$

The first zeros of this diffraction pattern in azimuth can be taken to determine the width of the radar beam (i.e. the first lobe). These zeros occur when

$$k_0\frac{w}{2}\sin\frac{\beta}{2} = \pm\pi$$

where $\beta/2 = y_0/z_0$. Hence, for small values of β,

$$\beta \simeq \frac{4\pi}{k_0w}$$

and the azimuth resolution is therefore linearly proportional to w, i.e. the azimuth resolution of the SAR is independent of the wavelength.

C. Multiple Scattering Model

By studying the response of a SAR to a point scatterer in range, and then in azimuth, the point spread function of the system is established which is given by

$$p(x, y) = LT\text{sinc}(\alpha Tx)\text{sinc}(\beta k_0y)$$

and is identical to the diffraction pattern produced by a rectangular aperture. Thus, the (post-processed) SAR image data $\hat{f}(x, y)$ generated by scattering from the ground is given by the convolution of the object function for the ground $f(x, y)$ with the appropriate point spread function, i.e.

$$\hat{f}(x, y) = p(x, y) \otimes_2 f(x, y)$$

where $\otimes_2$ denotes the two-dimensional convolution integral and $\hat{f}$ is taken to be complex data generated by f. A SAR image is usually generated by displaying the amplitude modulations of the data, i.e.

$$I_{SAR}(x, y) = \mid \hat{f}(x, y) \mid .$$

The object function describes the imaged properties of the ground surface. A conventional model for this function is the point scattering model that is typically used in SAR simulations [15]. This is where the object function is taken to be a distribution of point scatterers each of which reflects a replica of the emitted pulse and responds identically in azimuth. Here, nothing is said about the true physical nature of the ground surface such as its shape and material (dielectric) properties. Moreover, this standard convolution model for the data is based on application of the Born approximation where multiple scattering effects are taken to be part of the noise function $n(x, y)$ [16]. SAR images are coherent images (i.e. based on complex data containing magnitude and phase information) and consequently, contain noise that is characteristic of speckle patterns (coherent noise).

Based on results presented in Section V, we consider a model for the 'ground truth' estimate $\hat{f}(x, y)$ given by

$$\hat{f}(x, y) = p(x, y) \otimes_2 [f(x, y)+ \mid s(x, y) \mid^2 \exp(-ik_0x)]$$

where the second term (on the right hand side) is taken to be the multiple scattering term[4]. Figure 6 shows the effect of applying a Gaussian lowpass filter to the complex data $\hat{f}$ before generation of the image $\mid \hat{f}(x, y) \mid$ using data obtained from the Sandia National Laboratories SAR database [17]. Filtering the complex data is undertaken in order to attempt to suppress the term $p(x, y) \otimes_2 [\mid s(x, y) \mid^2 \exp(-ik_0x)]$. Applying a lowpass filter prior to generating an amplitude image of the complex data illiminates the cross terms generated by computing the image $\mid \hat{f}(x, y) \mid$, thereby reducing speckle, the convential approach to speckle reduction being to apply a filter to the amplitude image.

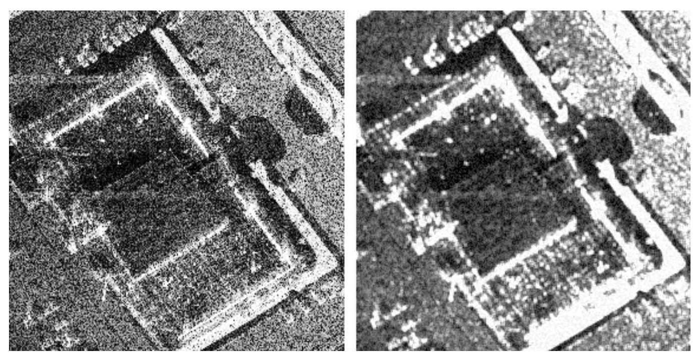

Fig. 6. Example of an original SAR image (left) and the same image after applying a lowpass filter to the conmplex data (right). In both cases, the images have been histogram equalized utilizing the MATLAB function *histeq*.

VIII. Conclusion

The extension of the standard model $s(t) = p(t) \otimes f(t) + n(t)$ to include multiple scattering effects that are compounded in a single term such that

$$s(t) = p(t) \otimes f(t) + p(t) \otimes [\mid s(t) \mid^2 \exp(-i\omega_0t)] + n'(t)$$

provides a method of processing signals that is compatible with a standard signal processing 'toolkit'. Here the term $n'(t)$ include all forms of noise that does not include multiple scattering effects.

[4]After application of conventional SAR signal processing.

The inverse scattering problem is usually formulated by first solving the forward scattering problem as discussed in Section III. Once the relationship between the scattered field and the scattering function has been established by solving the Helmholtz equation, an inverse solution is then attempted. The approach taken in this paper has been to work directly with the Helmholtz equation to produce an expression for the scattering function (as derived in Appendix I). The example numerical simulations derived in IV(C) - Figures 1-3 - provides evidence of the expected superiority of this solution over the weak scattering solution. However, it should be noted that this result is based on the application of the narrow side-band condition $\mid k \mid << k_0$ and that $\mid u_i(x,k) + u_s(x,k) \mid^2 \sim 1$. In this sense, the solutions developed are not strictly based on an 'exact' inverse scattering solution, but rather a modified version of the exact solution derived in Appendix I tailored for applicability to narrow side-band systems. Such systems are consistent with the applications of electromagnetic signal processing which is the focus of this paper.

The approach taken in this paper has been to consider a theoretically 'exact solution' to the inverse scattering problem (as compounded in Appendix I) and then modify this result to accommodate conditions that reduce the solution to a form that is practically realisable in terms of existing signal processing models. This is a different approach to that which is traditionally taken where a forward scattering solution is developed based on physical conditions to obtain a (forward scattering) transform which can then be inverted. With regard to pulse-echo side-band systems, our approach provides a model for strong scattering that is compounded in a single additional term. The form of this term indicates the use of lowpass filtering applied to the complex data of a SAR rather than to the image itself which is the more usual practice with regard to speckle reduction.

In general, the approach reported in this paper may provide a framework for developing 'strong' scattering solutions that are of practical value to signal processing systems. For example, a further development of the approach used can be undertaken by relaxing the condition $u = \exp(i\theta_u)$ and considering an expansion of the term

$$\frac{1}{\mid u_i + u_s \mid^2} = 1 - u_i u_s^* - u_i^* u_s - \mid u_s \mid^2 + ...$$

Finally, this paper is primarily based on a one-dimensional analysis of the problem. However, the same method can be applied in two- and three-dimensions (at least for the Schrödinger and Helmholtz equations) and for completeness, the underlying results are given in Appendix II.

APPENDIX I
EXACT INVERSE SCATTERING SOLUTION IN ONE-DIMENSION

Consider the 1D inhomogeneous Helmholtz equation for a scalar (complex) wavefield $u(x,k)$ given by

$$\left(\frac{\partial^2}{\partial x^2} + k^2\right) u(x,k) = -k^2 \gamma(x) u(x,k), \qquad x \in (-\infty, \infty)$$

where $k > 0$ is the wavenumber (taken to be a constant) and the scattering function may be of compact support, i.e.

$$\gamma(x) \exists \forall x \in [-X, X].$$

The Forward Scattering Problem is defined as follows: Given $\gamma(x) \forall x$ find an exact solution for $u(x,k)$ The Inverse Scattering Problem is defined as follows: Given $u(x,k) \forall x$ find an exact solution for $\gamma(x)$.

A. Theorem

Given that the (forward scattering Green's function) solution to the Helmholtz equation (as defined above) is

$$u(x,k) = u_i(x,k) + u_s(x,k) \tag{I.1}$$

where u_i is a solution of

$$\left(\frac{\partial^2}{\partial x^2} + k^2\right) u_i(x,k) = 0,$$

$$u_s(x,k) = k^2 g(\mid x \mid, k) \otimes \gamma(x) u(x,k)$$
$$\equiv k^2 \int_{-\infty}^{\infty} g(\mid x - y \mid, k) \gamma(y) u(y,k) dy \tag{I.2}$$

and

$$g(\mid x - y \mid, k) = \frac{i}{2k} \exp(ik \mid x - y \mid)$$

which is the (outgoing Green's function) solution of

$$\left(\frac{\partial^2}{\partial x^2} + k^2\right) g(\mid x - y \mid, k) = -\delta(x - y),$$

then

$$\gamma(x) = \frac{u^*(x,k)}{\mid u(x,k) \mid^2} \frac{\partial^2}{\partial x^2} \left[R(x) \otimes u_s(x,k) - \frac{1}{k^2} u_s(x,k) \right].$$

where (c_1 and c_2 being arbitrary constants)

$$R(x) = \begin{cases} (c_1 - 1)x + c_2, & x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

B. Proof

From equations (I.1) and (I.2), we can write

$$(u - u_i) = k^2 g \otimes \gamma u.$$

Consider a piecewise continuous function q that is twice differentiable, such that

$$q \otimes (u - u_i) = k^2 q \otimes g \otimes \gamma u.$$

Differentiating twice, we have

$$\frac{\partial^2}{\partial x^2}[q \otimes (u - u_i)] = k^2 \frac{\partial^2}{\partial x^2}(q \otimes g \otimes \gamma u)$$

$$= k^2 \frac{\partial^2}{\partial x^2}(q \otimes g) \otimes \gamma u = -k^2 \delta \otimes \gamma u = -k^2 \gamma u$$

provided

$$\frac{\partial^2}{\partial x^2}(q \otimes g) = -\delta.$$

But

$$\frac{\partial^2}{\partial x^2}(q \otimes g) = q \otimes \frac{\partial^2}{\partial x^2} g$$

$$= q \otimes (-k^2 g - \delta) = -k^2 q \otimes g - q = -\delta$$

and hence

$$q = \delta - k^2 q \otimes g$$

so that

$$\frac{\partial^2}{\partial x^2}[q \otimes (u - u_i)] = \frac{\partial^2}{\partial x^2}[\delta \otimes (u - u_i) - k^2 q \otimes g \otimes (u - u_i)]$$

$$= \frac{\partial^2}{\partial x^2}[(u - u_i) - k^2 q \otimes g \otimes (u - u_i)] = -k^2 \gamma u.$$

Thus,

$$\gamma = \frac{1}{u} \frac{\partial^2}{\partial x^2} \left[q \otimes g \otimes (u - u_i) - \frac{1}{k^2}(u - u_i) \right]$$

The function q is determined by the solution of

$$\frac{\partial^2}{\partial x^2}(q \otimes g) = -\delta$$

$$\Longrightarrow \frac{\partial}{\partial x}(q \otimes g) = -H(x) + c_1 \tag{I.3}$$

where

$$H(x) = \begin{cases} 1, & x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

and c_1 is a constant. But the solution of equation (I.3) is

$$q \otimes g = R(x)$$

where (c_2 being a constant of integration)

$$R(x) = -\int_{-\infty}^{x} H(x)dx + c_1 x + c_2$$

$$= \begin{cases} (c_1 - 1)x + c_2, & x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

so that

$$\gamma = \frac{1}{u} \frac{\partial^2}{\partial x^2} \left[R \otimes (u - u_i) - \frac{1}{k^2}(u - u_i) \right].$$

Finally, since $u = u_i + u_s$, we can write

$$\gamma = \frac{1}{u} \frac{\partial^2}{\partial x^2} \left[R \otimes u_s - \frac{1}{k^2} u_s \right]$$

$$= \frac{u^*}{| u |^2} \frac{\partial^2}{\partial x^2} \left[R \otimes u_s - \frac{1}{k^2} u_s \right]. \tag{I.4}$$

C. Corollary

Since

$$\left(\frac{\partial^2}{\partial x^2} + k^2 \right) u_i = 0.$$

it follows that (for $c_1 < 1$)

$$\begin{aligned}
\gamma &= \frac{1}{u} \frac{\partial^2}{\partial x^2} \left[R \otimes (u - u_i) - \frac{1}{k^2}(u - u_i) \right] \\
&= \frac{1}{u} \left[-\delta \otimes (u - u_i) - \frac{1}{k^2} \left(\frac{\partial^2}{\partial x^2} u - \frac{\partial^2}{\partial x^2} u_i \right) \right] \\
&= \frac{1}{u} \left[-(u - u_i) - \frac{1}{k^2} \frac{\partial^2}{\partial x^2} u + \frac{1}{k^2} \frac{\partial^2}{\partial x^2} u_i \right] \\
&= \frac{1}{k^2 u} \left[- \left(\frac{\partial^2}{\partial x^2} u + k^2 u \right) + \frac{\partial^2}{\partial x^2} u_i + k^2 u_i \right] \\
&= -\frac{1}{k^2 u} \left(\frac{\partial^2}{\partial x^2} + k^2 \right) u.
\end{aligned}$$

which recovers the Helmholtz equation

$$\left(\frac{\partial^2}{\partial x^2} + k^2 \right) u = -k^2 \gamma u.$$

D. Remark I.1

The equivalent inverse solution for the Schrödinger equation

$$\left(\frac{\partial^2}{\partial x^2} + k^2 \right) u(x, k) = \gamma(r) u(x, k)$$

is

$$\gamma = \frac{u^*}{| u |^2} \frac{\partial^2}{\partial x^2} \left[u_s - k^2 R \otimes u_s \right]. \tag{I.5}$$

E. Remark I.2

The inverse solutions given by equations (I.4) and (I.5) rely on the condition:

$$| u(x, k) | = | u_i(x, k) + u_s(x, k) | > 0 \forall x, k.$$

Thus, for a fixed wavenumber $k > 0$, the incident u_i and scattered u_s wavefields must be 'out-of-phase' $\forall x$.

F. Remark I.3

The theorem provides a result that is compatible with the trivial inverse solution

$$\gamma = -\frac{u^*}{| u |^2} \left(1 + \frac{1}{k^2} \frac{\partial^2}{\partial x^2} \right) u.$$

However, unlike this trivial solution, the theorem provides an expression for the scattering function γ which is, at least, consistent with the (exact) forward scattering (Green's function) solution $u = u_i + u_s$ and is determined by the scattered wavefield $u_s = k^2 g \otimes \gamma u$ that, like the incident wavefield u_i, is assumed to be a measurable quantity.

G. Remark I.4

In order to use this inverse solution, the wavefield u must be known $\forall x$. For a scatterer of compact support, the field may only be measurable beyond this support and thus, the data on u may be incomplete.

APPENDIX II
EXACT INVERSE SCATTERING SOLUTION IN THREE-DIMENSIONS

A. Theorem

Consider the 3D inhomogeneous Helmholtz equation for a scalar (complex) wavefield $u(\mathbf{r}, k)$ given by

$$(\nabla^2 + k^2)u(\mathbf{r}, k) = -k^2\gamma(\mathbf{r})u(\mathbf{r}, k)$$

where k is the wavenumber and the scattering function is of compact support, i.e.

$$\gamma(\mathbf{r})\exists\forall\mathbf{r} \in V.$$

Given that the forward (Green's function) solution to this equation is

$$u(\mathbf{r}, k) = u_i(\mathbf{r}, k) + u_s(\mathbf{r}, k) \tag{II.1}$$

where u_i is a solution of

$$(\nabla^2 + k^2)u_i(\mathbf{r}, k) = 0,$$

$$u_s(\mathbf{r}, k) = k^2 g(\mid \mathbf{r} \mid, k) \otimes_3 \gamma(\mathbf{r})u(\mathbf{r}, k)$$

$$\equiv k^2 \int_V g(\mid \mathbf{r} - \mathbf{r}' \mid, k)\gamma(\mathbf{r}')u(\mathbf{r}', k)d^3\mathbf{r}' \tag{II.2}$$

and

$$g(\mid \mathbf{r} - \mathbf{r}' \mid, k) = \frac{\exp(ik \mid \mathbf{r} - \mathbf{r}' \mid)}{4\pi \mid \mathbf{r} - \mathbf{r}' \mid}$$

which is the solution of

$$(\nabla^2 + k^2)g(\mid \mathbf{r} - \mathbf{r}' \mid, k) = -\delta^3(\mathbf{r} - \mathbf{r}'),$$

then, with $r \equiv \mid \mathbf{r} \mid$,

$$\gamma(\mathbf{r}) = \frac{u^*(\mathbf{r}, k)}{\mid u(\mathbf{r}, k) \mid^2}\nabla^2 \left[\frac{1}{4\pi r} \otimes_3 u_s(\mathbf{r}, k) - \frac{1}{k^2}u_s(\mathbf{r}, k)\right].$$

B. Proof

From equations (II.1) and (II.2), we can write

$$(u - u_i) = k^2 g \otimes_3 \gamma u.$$

Consider a function q such that

$$q \otimes_3 (u - u_i) = k^2 q \otimes_3 (g \otimes_3 \gamma u).$$

Taking the Laplacian of this equation, we have

$$\nabla^2[q \otimes_3 (u - u_i)] = k^2\nabla^2(q \otimes_3 g \otimes_3 \gamma u)$$

$$= k^2\nabla^2(q \otimes_3 g) \otimes_3 \gamma u = -k^2\delta^3 \otimes_3 \gamma u = -k^2\gamma u$$

provided

$$\nabla^2(q \otimes_3 g) = -\delta^3.$$

But

$$\nabla^2(q \otimes_3 g) = q \otimes_3 \nabla^2 g = q \otimes_3 (-k^2 g - \delta^3)$$

$$= -k^2 q \otimes_3 g - q = -\delta^3$$

and hence

$$q = \delta^3 - k^2 q \otimes_3 g$$

so that

$$\nabla^2[q \otimes_3 (u - u_i)] = \nabla^2[\delta^3 \otimes_3 (u - u_i) - k^2 q \otimes_3 g \otimes_3 (u - u_i)]$$

$$= \nabla^2[(u - u_i) - k^2 q \otimes_3 g \otimes_3 (u - u_i)] = -k^2\gamma u.$$

Thus,

$$\gamma = \frac{1}{u}\nabla^2 \left[q \otimes_3 g \otimes_3 (u - u_i) - \frac{1}{k^2}(u - u_i)\right]$$

where q is determined by the solution of

$$\nabla^2(q \otimes_3 g) = -\delta^3. \tag{II.3}$$

But the solution of equation (II.3) is

$$q \otimes_3 g = \frac{1}{4\pi r}$$

so that

$$\gamma = \frac{1}{u}\nabla^2 \left[\frac{1}{4\pi r} \otimes_3 (u - u_i) - \frac{1}{k^2}(u - u_i)\right].$$

Finally, since $u = u_i + u_s$, we can write

$$\gamma = \frac{1}{u}\nabla^2 \left[\frac{1}{4\pi r} \otimes_3 u_s - \frac{1}{k^2}u_s\right]$$

$$= \frac{u^*}{\mid u \mid^2}\nabla^2 \left[\frac{1}{4\pi r} \otimes_3 u_s - \frac{1}{k^2}u_s\right]. \tag{II.4}$$

C. Corollary

Since

$$(\nabla^2 + k^2)u_i = 0.$$

it follows that

$$\begin{aligned}
\gamma &= \frac{1}{u}\nabla^2 \left[\frac{1}{4\pi r} \otimes_3 (u - u_i) - \frac{1}{k^2}(u - u_i)\right] \\
&= \frac{1}{u}\left[-\delta^3 \otimes_3 (u - u_i) - \frac{1}{k^2}(\nabla^2 u - \nabla^2 u_i)\right] \\
&= \frac{1}{u}\left[-(u - u_i) - \frac{1}{k^2}\nabla^2 u + \frac{1}{k^2}\nabla^2 u_i\right] \\
&= \frac{1}{k^2 u}\left[-(\nabla^2 u + k^2 u) + \nabla^2 u_i + k^2 u_i\right] \\
&= -\frac{1}{k^2 u}(\nabla^2 + k^2)u.
\end{aligned}$$

which recovers the Helmholtz equation

$$(\nabla^2 + k^2)u = -k^2\gamma u.$$

D. Remark II.1

Remarks I.1-I.4 apply to this three dimensional derivation.

E. Remark II.2

In the 2D case, equation (II.3) becomes

$$\nabla^2(q \otimes_2 g) = -\delta^2$$

and has the solution

$$q \otimes_2 g = \frac{1}{2\pi}\ln r$$

where $\otimes_2$ denotes the two-dimensional convolution integral. The equivalent 2D inverse solution is then given by

$$\gamma(\mathbf{r}) = \frac{u^*(\mathbf{r}, k)}{\mid u(\mathbf{r}, k) \mid^2}\nabla^2 \left[\frac{1}{2\pi}\ln r \otimes_2 u_s(\mathbf{r}, k) - \frac{1}{k^2}u_s(\mathbf{r}, k)\right].$$

F. Remark II.3

Equation (II.1) relies on the boundary condition $u(\mathbf{r},,k) = u_i(\mathbf{r},k) \quad \forall \mathbf{r} \in S$ where S defines the surface of $\gamma(\mathbf{r})$ which is taken to be of compact support. The Green's function solution to the three dimensional inhomogeneous Helmholtz equation is

$$u(\mathbf{r},k) = k^2 \int\limits_V g\gamma u d^3\mathbf{r}' + \oint\limits_S (g\nabla u - u\nabla g)\cdot \hat{\mathbf{n}} d^2\mathbf{r}'.$$

To compute the surface integral, a condition for the behaviour of u on the surface S of γ must be chosen. If we consider the case where the incident wavefield u_i is a simple plane wave of unit amplitude

$$\exp(i\mathbf{k}\cdot\mathbf{r})$$

satisfying the homogeneous wave equation

$$(\nabla^2 + k^2)u_i(\mathbf{r},k) = 0,$$

then

$$u(\mathbf{r},k) = k^2 \int\limits_V g\gamma u d^3\mathbf{r}' + \oint\limits_S (g\nabla u_i - u_i\nabla g)\cdot \hat{\mathbf{n}} d^2\mathbf{r}'.$$

Using Green's theorem to convert the surface integral back into a volume integral, we have

$$\oint\limits_S (g\nabla u_i - u_i\nabla g)\cdot \hat{\mathbf{n}} d^2\mathbf{r}' = \int\limits_V (g\nabla^2 u_i - u_i\nabla^2 g) d^3\mathbf{r}'.$$

Noting that

$$\nabla^2 u_i = -k^2 u_i$$

and that

$$\nabla^2 g = -\delta^3 - k^2 g$$

we obtain

$$\int\limits_V (g\nabla^2 u_i - u_i\nabla^2 g) d^3\mathbf{r}' = \int \delta^3 u_i d^3\mathbf{r}' = u_i.$$

Hence, by choosing the field u to be equal to the incident wavefield u_i on the surface of γ, we obtain a solution of the form

$$u = u_i + u_s$$

where

$$u_s(\mathbf{r},k) = k^2 g(\mid \mathbf{r} \mid, k) \otimes_3 \gamma(\mathbf{r}) u(\mathbf{r},k).$$

ACKNOWLEDGMENTS

J M Blackledge is supported by the Science Foundation Ireland (Stokes Professorship Programme).

REFERENCES

[1] J. M. Blackledge, *Digital Signal Processing* Second Edition, Horwood Publications, 2006.

[2] P. A. Martin, *Multiple Scattering: Interaction of Time-Harmonic Waves with N Obstacles*, Encyclopedia of Mathematics and its Applications 107, Cambridge University Press, 2006.

[3] http://www.mathworks.com/applications/dsp_comm/

[4] P. M. Morse and H. feshbach, *Methods of Theoretical Physics*, McGraw-Hill, 1953

[5] M. Fujimoto, *Physics of Classical Electromagnetism*, Springer, 2007.

[6] S. Cloude, *An Introduction to Electromagnetic Wave Propagation and Antennas*, Springer, 1996.

[7] W. Coffey, Y. P. Kalmykov and J. T. Waldron, *The Langevin Equation: With Applications in Physics, Chemistry, and Electrical Engineering*, World Scientific, 1996.

[8] D. J. Griffiths, *Introduction to Quantum Mechanics* (Second Edition), Prentice Hall, 2004.

[9] S. Winitzki, *Cosmological Particle Production and the Precision of the WKB Approximation*, Physical Review D 72: 104011, 2005

[10] R. Jost and W. Kohn, *Construction of a Potential from a Phase Shift*, Phys. Rev. **37**, 977-992, 1952.

[11] R. T. Prosser, *Formal Solutions of Inverse Scattering Problems*, J. Math. Phys. **17**, 1175-1779, 1976

[12] A. W. Rihaczek, *Principles of High Resolution Radar*, McGraw-Hill, 1969.

[13] R. O. Harger, *Synthetic Aperture Radar Systems*, Academic Press, 1970.

[14] J. J. Kovaly, *Synthetic Aperture Radar*, Artech, 1976.

[15] R. L. Mitchell, *Radar Signal Simulation*, MARK Resources, 1985.

[16] J. M. Blackledge, *Digital Image Processing*, Horwood, 2006.

[17] http://www.sandia.gov/RADAR/sar-data.html

Jonathan Blackledge graduated in physics from Imperial College in 1980. He gained a PhD in theoretical physics from London University in 1984 and was then appointed a Research Fellow of Physics at Kings College, London, from 1984 to 1988, specializing in inverse problems in electromagnetism and acoustics. He currently holds the Stokes Professorship in Digital Signal Processing at Dublin Institute of Technology.

Timo Hämäläinen received his B.Sc in automation engineering from the Jyväskylä University in Finland on 1991 and MSc and PhD degrees in Telecommunications from Tampere University of Technology and from Jyväskylä University, Finland in 1996 and 2002, respectively. Currently, he is Professor of Telecommunications at Jyväskylä University and has research interests that include traffic engineering and Quality of Service in wired and wireless networks.

Jyrki Joutsensalo was born in Kiukainen, Finland, in July 1966. He received a Diploma Engineer, Licentiate of Technology, and Doctor of Technology degrees from Helsinki University of Technology, Finland, in 1992, 1993 and 1994, respectively. Currently, he is Professor of Telecommunications at the University of Jyväskylä. His research interests include DSP for telecommunications, networks and communications engineering.