On deducing the form of surfaces from their diffracted echoes

MVBERRY

H H Wills Physics Laboratory, University of Bristol, Bristol BS8 1TL, UK

MS received 8 July 1971

Abstract. Kirchhoff diffraction theory is set up for the echo received back at the source when a spherical outgoing pulse of radiation is reflected from a rough surface. The formulation derived here permits a clear separation between those features of the echo that depend only on the form of the incident pulse, and those involving only the properties of the surface. The surface function depends on the contours of constant distance from the source which can be drawn in the surface; these are the appropriate analogues of Fresnel zone boundaries. Geometrical optics gives the singularities and discontinuities of the surface function. The problem of determining the surface function from echo data which are incomplete in the sense that they do not contain all radiation frequencies is discussed. For the special case of one dimensional roughness, the problem of how far the surface function can be deconvoluted to obtain the shape of the surface is considered.

1. Introduction

The vast literature dealing with wave theory is principally devoted to working out how wave systems are modified when they encounter objects or structures or media with which they interact according to known laws. Once the structure acting on the wave is specified, then, while it may be difficult to calculate what happens to the wave, at least the problem is well defined. But in practice it is very commonly the case that the wave system is only a tool for the investigation of structures whose form is unknown, and the question then arises: what information about the structure can be inferred from the modification that it impresses on the wave? On this inverse problem only scattered results exist: for three dimensional, spherically symmetrical structures interacting elastically with incident particles, for example, the form of the structure may be inferred if the scattering mechanism is assumed purely classical, and there are no caustics in the pattern of particle paths, or if the interaction is quantum mechanical, and the scattering potential has no bound states (Mott and Massey 1965); general three dimensional structures may be inferred if the wave mechanism is pure shadowing or pure phase modulation (Berry and Gibbs 1970), or if the interaction is weak, as in x ray crystallography.

It is the purpose of the present paper to make some small inroads into the inverse problem for the case where the structure under investigation is a reflecting surface, and the wave system is a pulse from a localized source. The literature on the direct 'rough surface' problem is enormous (see Beckmann and Spizzichino 1963 and the references therein), but it is primarily concerned with harmonic waves rather than pulses, and much of the general theory is set up for plane, rather than expanding, waves. The immediate motivation for this study is provided by experiments currently being carried out in the Antarctic, where the echoes from pulsed radio signals sent out from

aircraft are used to investigate the form of the ice-rock interface at a glacier bed (Robin et al 1969); there are, however, many other applications of essentially the same technique, two being the study of sea beds by ultrasonic echoes and the investigation of the clutter introduced into radar echoes by the sea surface.

It is essential to realize that any inference from echo to surface must be within the context of a well defined model for the reflection process. Our model will be the Kirchhoff diffraction theory, which is the simplest model describing a wide range of qualitatively different conditions; in particular, it includes geometrical optics as a special case. The principal conditions for the validity of the Kirchhoff theory are:

- (i) there must be no shadowing of one part of the surface by any other part, as seen from the source.
- (ii) Multiple scattering of waves between different parts of the surface is neglected. Concerning the incident pulse, we assume:
 - (iii) the waves are scalar; this means, principally, that we are not going to consider the information about the surface that can be obtained from the polarization of the echo.
 - (iv) The radiation is emitted isotropically; this means that we are not going to consider the effects of the directional properties of the sources used in practice. (The finite angle of the cone of radiation illuminating the surface actually goes some way towards counteracting the inevitable violation of (i) at large angles.)
 - (v) The pulse arrives undistorted at the surface; in the glacier problem, this means that we are neglecting: (a) the refraction of the pulse at the ice surface and at layers within it; (b) the effects of the frequency dependence of the complex dielectric constant of ice, on the pulse shape which arrives at the surface; (c) scattering and absorption due to bubbles and objects in the ice; (d) reflection and absorption due to layers in the ice. All of these effects of the ice can to a certain extent be identified on the radar records, and taken into account (Robin et al 1969).

We shall be primarily concerned with the inferences that can be drawn from the temporal behaviour of the echo observed at a single point; from this one dimensional quantity it is clearly impossible to deduce the form of a surface known to vary in two dimensions. Therefore, in part of this preliminary investigation (§ 4) we shall make the further assumptions:

(vi) the surface height varies in one direction only (in other words, the surface is made up of ridges rather than hills). This commonly-employed but rather drastic assumption is somewhat justified by assumption (iv), since in practice the cones of illumination from radar aerials used in the Antarctic are wider fore-and-aft than they are sideways.

Concerning the receiver, we assume:

(vii) that it is coincident with and has the same isotropy as, the source; this is true for the glacier bed experiments, and it usually, but not invariably, applies in other applications.

Finally, concerning the surface whose topography we are examining, we assume:

(viii) that it is perfectly reflecting, so that the wavefunction is zero on it. In the case of Antarctic ice-rock interfaces, it has been estimated (Robin et al 1969) that the power reflection coefficient is only about 1%, so that this is not a very good approximation. Refraction into the underlying substrate can be taken into

account (Beckmann and Spizzichino 1963, chap 2), but this complicates matters and we shall ignore it since our main concern is with the logic of the deconvolution problem in its simplest form (in most cases the only effect of penetration into the underlying rock will be to reduce the echo strength by a constant factor).

Despite the length of this list, there remain possible simplifying approximations that we have not made. We have not assumed, for instance, that there exist characteristic lengths describing the height and lateral extent of typical roughnesses; this is because in any real landscape there are likely to be irregularities on all scales within a wide range (Nye 1970, p 385).

We begin (§ 2) by casting Kirchhoff's diffraction theory (set up in Appendix 1) into a form where the effects on the received echo due to the form of the incident pulse are clearly separated from those due to the shape of the surface. The effects of the surface are embodied in a function g(r) which is therefore a natural object of study. A general survey of the properties of g(r), together with the profound qualitative consequences of our imperfect knowledge of g(r), constitutes § 3. The general behaviour of g(r) is given by the geometrical-optics approximation, whose details are worked out in Appendix 2. To make further progress we specialize in § 4 to the case of one dimensional roughness, where a new surface function $\gamma(\rho)$, derived from g(r), is considered, whose singularities can, in principle, give information about the ridges and valleys of the rough surface.

It is not claimed that the work presented here is complete; on the contrary, at least three aspects of the deconvolution problem will be further explored in later publications:

- (1) the effect of the additional dimension(s) of information that can be obtained by moving the source.
- (2) The statistical information about the surface that can be obtained from the tail of the echo, together with the rather difficult question of how far this can be regarded as separate from the 'geographical' information about the unique topography near a particular source position.
- (3) Detailed calculation of the deconvolution integrals of § 3 and § 4, under laboratory-controlled conditions, using ultrasonic pulses and artificial surfaces scaled to correspond with the ice radar experiments.

2. The basic diffraction formulae

We define the surface Σ under investigation by its height $f(\mathbf{R})$ above a horizontal reference plane xy, where \mathbf{R} is the two dimensional vector leading to xy from the origin O of the plane. We are not generally going to assume that the surface is in any sense 'close' to the reference plane. The source S, and the receiver, are situated at a height h vertically above O. The distance $r(\mathbf{R})$ from S to the point P on the surface above the point \mathbf{R} on the reference plane, will be important throughout our argument; it is given by

$$r(\mathbf{R}) = \{ (h - f(\mathbf{R}))^2 + R^2 \}^{1/2}. \tag{1}$$

This geometry is summarized in figure 1; the only restriction on the magnitudes involved is that h must exceed the surface height $f(\mathbf{R})$ for all \mathbf{R} .

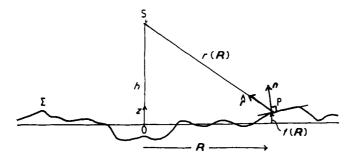


Figure 1. Basic geometry.

The pulse emitted by the source is specified by its time behaviour F(t), or by the Fourier transform of this, namely

$$\bar{F}(\omega) \equiv \frac{1}{2\pi} \int_{-\infty}^{\infty} dt F(t) \exp(i\omega t).$$
 (2)

Characteristically, F(t) is a quasisinusoidal pulse with basic frequency ω_0 , consisting of a small number m of waves (about ten in the ice radar experiments), so that $\overline{F}(\omega)$ consists of two peaks at $\omega = \pm \omega_0$ whose breadths are about $2\omega_0/m$ (there are two peaks because F(t) is real, implying that $\overline{F}(-\omega) = F^*(\omega)$). In accordance with our assumptions (4) and (5), the wave incident upon the surface at the point r at time t can be written as

$$\phi(r,t) = \frac{F(t - (r/c))}{r} \tag{3}$$

where c is the appropriate wave velocity; this represents an expanding spherical wave train whose mean wavelength is $\lambda_0 = 2\pi c/\omega_0$.

Because we are using the Kirchhoff approach to diffraction theory, we regard the reflected wave $\psi(t)$, that is received back at S at time t (the echo), as being made up of wavelets from all points of the rough surface. The wavelet received at time t from the point P on the surface, depends on the part of the pulse that was emitted at time t-2r(R)/c since the wave traverses the distance SP twice. The received wavelet is reversed in sign relative to the emitted part of the pulse, because of the reflection; furthermore, it depends not on the value F of the pulse at the instant of emission, but on the rate of change F'; this is always necessary in Kirchhoff's theory (where it usually appears as a factor $i\omega$ since that theory is usually set up for monochromatic waves) in order for the original expanding wave to be obtained in the case where there is no diffracting surface. It is shown in Appendix 1 that the received wave $\psi(t)$ which correctly takes these effects into account is given by

$$\psi(t) = \frac{-1}{2\pi c} \int \int d\mathbf{R} \frac{F'(t - (2r(\mathbf{R})/c))}{(h - f(\mathbf{R}))r(\mathbf{R})}.$$
 (4)

The form of the denominator in this formula takes account of the transformation from the integral over the actual surface to that over the flat reference surface, and also includes the obliquity factor in a modified form; the equation (4) is valid (Appendix 1) if

$$\lambda_0 \ll h - f \tag{5}$$

a condition satisfied for the pulses and source heights employed in practice. Equation (4)

defines the model on which our theory is based, and we shall use it even for expressing the reflected waves in the case of delta function pulses, where equation (5) is not obviously satisfied; the diffraction occurring in these cases should still be correctly described qualitatively.

A simple test of equation (4) is provided by the flat surface $f(\mathbf{R}) = 0$: use of (two dimensional) polar coordinates for the integration over \mathbf{R} then yields

$$\psi(t) = -\frac{F(t - (2h/c))}{2h} \tag{6}$$

which is exactly the form in which we would expect the wave given by equation (3) to be reflected from a flat surface.

As it stands, equation (4) is not in a convenient form for extracting information about the surface; to transform it, we notice that the R dependence of the integrand occurs primarily in the function r(R), and that the spheres of constant r, centred on S, cut the rough surface in a series of contours. These contours of constant r are the generalization to the case of pulses and rough surfaces of the circular Fresnel-zone boundaries which are useful when considering harmonic waves and flat surfaces. Thus we can perform the integration over R by first integrating around each contour (projected onto the R plane) at constant r, and then integrating over r from zero to infinity. If the total length round all the (possibly disconnected) parts of the contour at r is L(r), then we can integrate around the contour by using a variable l running from O to L(r) (figure 2), so that we have to find the transformation of the area elements

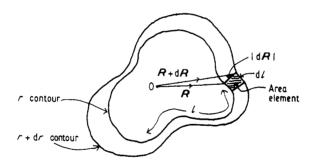


Figure 2. Contours in the R plane.

from dx dy to dr dl. Using equation (1), it is easy to see that the perpendicular distance $|d\mathbf{R}|$ between the contours r and r+dr at the point l is

$$|d\mathbf{R}| = \frac{r dr}{|\mathbf{R} - (h - f(\mathbf{R}))\nabla_{\mathbf{R}} f(\mathbf{R})|}.$$
 (7)

Thus our basic diffraction formula (4) becomes (figure 2)

$$\psi(t) = -\frac{1}{2\pi c} \int_0^\infty dr F' \left(t - \frac{2r}{c} \right) g(r)$$
 (8)

where

$$g(r) \equiv \int_0^{L(r)} \frac{\mathrm{d}l}{(h - f(\boldsymbol{R}(r, l)))|\boldsymbol{R}(r, l) - (h - f(\boldsymbol{R}(r, l)))\nabla f(\boldsymbol{R}(r, l))|}.$$
 (9)

The function g(r) represents the modification impressed by the surface on that part of the wave received at t which was emitted at $\{t - (2r/c)\}$. For a flat surface, there are no contours until r = h, after which the contours are circles, so that

$$L(r) = 2\pi (r^2 - h^2)^{1/2} \theta(r - h)$$
(10)

and

$$g(r) = \frac{2\pi}{h}\theta(r-h)$$

where $\theta(x)$ is the unit step function. Substitution of this into (8) yields (6) immediately. The formulation (8) and (9) has the advantage of separating those contributions to the echo which depend only on the form of the incident pulse from those which depend only on the form of the surface. The latter contributions are embodied in the function g(r) which clearly represents the totality of information about the surface that is obtainable by analysing the reflected signal received back at S.

3. Inferences about the surface: general two dimensional case

The basic surface function g(r) is defined by the integral (9), from which it is clear that important contributions are likely to arise at those values of r for which the denominator in the integrand vanishes. From (7), these are the points where

$$\nabla_{\mathbf{R}} r(\mathbf{R}) = 0. \tag{11}$$

These are the 'reflection points' of geometrical optics, where the surface is oriented perpendicular to the line joining it to the source (an equivalent way of expressing this is to say that at these points the r sphere is tangential to the surface). The way in which these reflection points contribute to g(r) is worked out in Appendix 2, where it is shown that hilltops and floors of hollows of r(R) produce step discontinuities in g(r) (of which equation (10) is the simplest example) while saddle points (where a contour crosses itself) are characterized by logarithmic singularities. More severe singularities of g(r) are to be expected at those reflection points where focusing occurs, so that as well as (11) we also have the condition (Appendix 2)

$$\det \left\| \frac{\partial^2 r}{\partial R_k \, \partial R_l} \right\| = 0 \tag{12}$$

 $(R_k \text{ or } R_l \text{ can represent } x \text{ or } y)$, and at least one great circle in the r sphere osculates the surface; this case rarely occurs, and has not been examined in detail. The smooth variation of g(r) between these various types of reflection point depends on the details of the contour system, and can only be found by a full evaluation of (9).

We shall not discuss here the question of what detailed information about the surface $f(\mathbf{R})$ can be obtained from a knowledge of g(r), because in § 4 we shall treat this problem for the simpler case of a surface which varies in one direction only. Instead we shall now examine what can be found out about g(r) itself by deconvoluting the measured data. If we define $\overline{\psi}(\omega)$ as the Fourier transform of the received echo, by analogy with (2),

it is easy to derive a formal expression for g(r) in terms of measured quantities, by using equation (8); the expression is

$$g(r) = -2i \int_{-\infty}^{\infty} d\omega \frac{\overline{\psi}(\omega)}{\omega \overline{F}(\omega)} \exp\left(-\frac{2i\omega r}{c}\right). \tag{13}$$

This is a formal expression only, because the integral can only be evaluated if the incident signal F(t) contains components of all frequencies ω . The simplest such signal is the delta function

$$F(t) = \delta(t)$$

which gives, from (13) or (8)

$$g(r) = -4\pi \int_{-\infty}^{2r/c} dt \psi(t).$$

In practice, of course, the incident signal has Fourier components only in some range of frequencies surrounding ω_0 . This means that we can never know the complete behaviour of g(r), but only those of its wavenumber Fourier components, given by

$$\bar{\mathbf{g}}(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}r \mathbf{g}(r) \exp(\mathrm{i}kr) = \frac{2\mathrm{i}\bar{\psi}(kc/2)}{k\bar{F}(kc/2)}$$
(14)

for which the frequency kc/2 is present significantly in the outgoing pulse (in other words, a wavelength λ in the incident pulse gives information about $k = 4\pi/\lambda$ in $\bar{g}(k)$).

For quasimonochromatic pulses, these considerations imply that there are limits to the fineness with which we can examine g(r): if the average wavelength in the pulse is λ_0 , then we cannot know those components of $\overline{g}(k)$ for which k substantially exceeds $4\pi/\lambda_0$. If we adopt the criterion that knowledge of $\overline{g}(k)$ is equivalent to knowledge of the details of g(r) on a scale π/k (half a wavelength), then we have shown that we cannot discern details of g(r) that are finer than $\lambda_0/4$, a result which is almost obvious a priori. (Of course we also have only limited information about the low frequency, coarse, variation of g(r), but in this case equation (14) can be supplemented by the knowledge that g(r) is never negative, and by the first and last returns in the echo, which delimit the spatial extent of the region where g(r) is nonzero. The numerical aspects of the deconvolution will, it is hoped, be examined in later work, in connection with laboratory ultrasonic pulse reflection experiments.)

The finest details in g(r) are the discontinuities and singularities arising from the reflection points of geometrical optics, and we know now that these cannot be resolved in all their detail, because we can at best reconstruct a smoothed version of g(r), averaged over a distance $\lambda_0/4$. If, however, successive reflection points r_i are separated by distances which, on average, considerably exceed $\lambda_0/4$, then their effects on g(r) will still be discernible—of course the step discontinuities from hills and hollows will be smoothed into shoulders, and the logarithmic singularities from saddle points will appear as finite peaks, but it will still be possible to discern the topological features of the function r(R) by examining the smoothed function g(r).

We may regard these circumstances, where the observed echo is dominated by the geometrical-optical features of g(r), as defining a smooth surface $f(\mathbf{R})$. In the opposite case of a rough surface, the successive reflection points occur at separations Δr_i which are comparable with or smaller than $\lambda_0/4$ so that their effects on g(r) are not separately discernible after the smoothing. This criterion for roughness involves distance r from

the source, and not position R on the surface; if we say that two features can be resolved if the difference between their distances to the source exceeds $\lambda_0/2$, then in the case of, for instance, two hills of approximately equal height, their separation in radial distance $\Delta |R|$ must satisfy

$$\Delta |\mathbf{R}| > \frac{\lambda_0 h}{2r} = \frac{\lambda_0}{2 \sin \alpha}$$

where α is the angle away from the vertical of the direction from the source to the hills. Thus two hills with the same value of |R| cannot be resolved, even if their azimuth angles are very different, and their separation $|\Delta R|$ is many wavelengths.

Surfaces occurring in nature can generally be expected to rough according to our definition, since they contain irregularities on all scales (Nye 1970, p 385ff). If we were to follow the convolutions of the contours of constant r around boulders, pebbles, sand, dust and, ultimately, atoms, it is clear that the connective topology of these contours would change infinitely often over a small finite range of r, so that g(r) would contain discontinuities and singularities, due to reflection points, on an infinitely fine scale—it would in fact be a mathematically pathological function. This type of behaviour is well known in geomorphology (Scheidegger 1970), where it is recognized that any integral around a contour (eg its length, or our function g(r) defined by (9)) has a value dependent upon the fineness of scale on which one is prepared to examine the surface. In our case the natural scale is provided by the wavelength λ_0 which predominates in the incident pulse, and we may sum up our argument by saying that if we intend drawing a map of the surface, there is no advantage in the thickness of the ink lines used for the contours of constant r being less than $\lambda_0/4$.

It may happen that a naturally occurring surface contains no irregularities whose scale is approximately λ_0 . Such a surface is *smooth* according to our criterion, since the finer irregularities cannot be perceived by the exploring pulse, while the coarser variations contribute in the form of geometrical reflections which can be discerned in the reconstructed smoothed function g(r).

Now let us consider how different parts of the echo arise from different regions of the surface. If we assume the signal F(t) to begin at t=0, then the first return will be received at a time $t=t_0=2r_0/c$ where r_0 is the distance of S from the nearest point of the surface. For most surfaces this nearest point will be approximately vertically below S; if this is assumed, and r_0 measured for a number of different source positions R in the plane of height h, then a rough map of the surface f(R) can be obtained simply by calculating

$$f(\mathbf{R}) \simeq h - r_0(\mathbf{R}). \tag{15}$$

This method, utilizing the first return, is the simplest way of interpreting the echo and it is the one most commonly used (eg in sonar). Although it has proved extremely useful in ascertaining the subglacial topography of Antarctica (Robin *et al* 1969), it clearly neglects an enormous amount of information.

For a short time after the first return the echo arises from those contours r which are exploring the parts of the surface nearest to S. These will generally take the form of loops surrounding the peak at r_0 on the surface, as in figure 3(a). At the time $t = t_0 + T$, where T is the duration of the outgoing pulse, the echo will arise from all those contours between $r = r_0$ (corresponding to the trailing edge of the outgoing pulse) and $r = r_0 + cT/2$ (corresponding to the beginning of the outgoing pulse); this band of contours is shown schematically in figure 3(b) for a smooth surface whose topography

does not vary significantly over distances corresponding to a pulse length $cT = m\lambda_0$ (m is the number of 'waves' in the pulse) and in figure 3(c) for a rough surface, where the topology of the contours changes drastically over the contributing range of r.

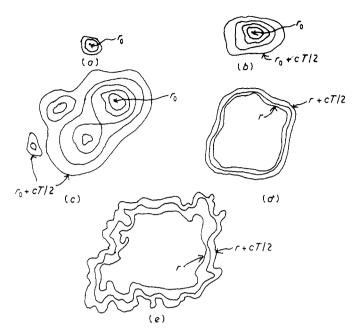


Figure 3. Contours contributing to different parts of echo. (a) Shortly after first return; (b) at time T after first return (smooth surface); (c) at time T after first return (rough surface); (d) in the tail of the echo (smooth surface); (e) in the tail of the echo (rough surface).

For times much later than $t=t_0+T$ the echo will come from a band of contours whose edges are separated by $\Delta r=cT/2$, corresponding to more distant parts of the surface. After a long enough time these contours contributing to the tail of the echo will, if the surface slopes have a finite maximum value, form single closed loops for each r (topologically equivalent to the circles from a flat surface), as shown in figure 3(d) (smooth surface) and figure 3(e) (rough surface). For a smooth surface, where the contributions come from the reflection points, the echo strength from this band of contours will be negligible since the simple closed contours for large r will never introduce new reflection points. Isolated late returns in a signal therefore indicate that the surface is smooth, with isolated distant points which have exceptionally large slope values, so that the r sphere can touch and not cut the surface there.

For a rough surface, the long-time contours shown in figure 3(e) give a diffusely-scattered contribution to the echo tail, from which statistical information can be deduced, whose nature will be explored in a later paper; this information augments the general topographical knowledge that can be obtained from the first return.

4. Inferences about the surface: one dimensional case

If the surface height f is known to vary only in one direction, then we might hope to be able to extract from g(r) sufficient information to determine the form of the surface

completely, apart from an unavoidable ignorance about the particular direction of variation (which could perhaps be resolved by moving the source and receiver to a single different position). If the direction of variation is x, then we require the function f(x); we ignore resolution difficulties for the moment.

The general 'surface function' g(r), defined by (9), takes on a simple form if we write the l integration (around the r contour) in terms of the variable

$$\rho \equiv (x^2 + (h - f(x))^2)^{1/2}. \tag{16}$$

Each value of ρ defines a circle centred on S, lying in the plane y = 0, which cuts the surface at the points $x_i(\rho)$ where i is a positive integer (see figure 4). As the r contour is

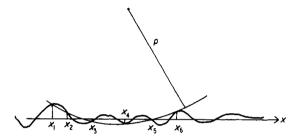


Figure 4. Intersections of the ρ circle with the surface at the points $x_i(\rho)$.

described, ρ varies between r_0 (from x = 0) and r; at each value of ρ there are contributions from the different parts of the contour, specified by the values $x_i(\rho)$. A short calculation then yields the transformation

$$\frac{\mathrm{d}l}{\mathrm{d}\rho} = \frac{2\rho}{(r^2 - \rho^2)^{1/2}} \sum_{i} \frac{|\mathbf{R}(r, x_i) - (h - f(x_i))\nabla f(x_i)|}{|x_i(\rho) - (h - f(x_i(\rho)))\,\mathrm{d}f(x_i(\rho))/\mathrm{d}x|}$$

where the factor two comes from the half of the surface for which y is negative, and the Σ_i arises because of the need to include all sections of the contours.

If this is substituted into (9), the result is

$$g(r) = 2 \int_0^r d\rho \frac{\rho}{(r^2 - \rho^2)^{1/2}} \gamma(\rho)$$
 (17)

where

$$\gamma(\rho) = \sum_{i} \frac{1}{(h - f(x_i(\rho)))\rho |d\rho(x_i(\rho))/dx|}$$

$$= \sum_{i} \frac{1}{(h - f(x_i(\rho)))|x_i(\rho) - (h - f)|df(x_i(\rho))/dx|}.$$
(18)

The lower limit of integration in (17) has set equal to zero and not r_0 , since $\gamma = 0$ if $\rho < r_0$. This new function $\gamma(\rho)$ represents the totality of the information that can be obtained about the surface from its echo; it can be found if g(r) is known, since (17) is an Abel integral equation whose solution (Bôcher 1926) is

$$\gamma(\rho) = \frac{1}{\pi \rho} \frac{d}{d\rho} \int_0^{\rho} dr \frac{rg(r)}{(\rho^2 - r^2)^{1/2}}.$$
 (19)

However, we have equation (13) for g(r), and if we substitute it into (19) we arrive after some algebra at a formal representation for $\gamma(\rho)$, which is the analogue of equation (13):

$$\gamma(\rho) = -\frac{4}{c} \int_{0}^{\infty} d\omega \left(J_{0} \left(\frac{2\rho\omega}{c} \right) \operatorname{Re} \frac{\overline{\psi}(\omega)}{\overline{F}(\omega)} + \mathcal{H}_{0} \left(\frac{2\rho\omega}{c} \right) \operatorname{Im} \frac{\overline{\psi}(\omega)}{\overline{F}(\omega)} \right). \tag{20}$$

The functions J_0 and \mathcal{H}_0 are, respectively, the Bessel functions of the first kind of order zero, and the Struve function of order zero (see Abramowitz and Stegun 1964, pp 358, 496). The real and imaginary parts occur in (19) because we have made use of the fact that the signal and echo are real functions of time, to eliminate negative frequencies.

To check the consistency of our work, and also to help understand these complicated expressions, let us evaluate both sides of (20) for a flat surface. Equation (6) gives

$$\frac{\overline{\psi}(\omega)}{\overline{F}(\omega)} = -\frac{\exp(2\mathrm{i}\omega h/c)}{2h}.$$

The right hand side of (20) can now be evaluated in terms of tabulated sine and cosine transforms (Erdèlyi 1954), to give

$$\gamma(\rho) = \frac{2\theta(\rho - h)}{h(\rho^2 - h^2)^{1/2}} \tag{21}$$

where θ again represents the unit step function. But this result can also be obtained by evaluating $\gamma(\rho)$ directly from its definition (18), since there are no intersections x_i of the ρ circle with the flat surface whilst $\rho < h$ after which there are two, specified by

$$x_{\frac{1}{2}} = \mp (\rho^2 - h^2)^{1/2}$$

hence equation (21).

Let us now discuss the function $\gamma(\rho)$ in detail, along lines similar to Appendix 2 where the corresponding (more complicated) problem for g(r) is worked out. Its behaviour is dominated by those values of ρ denoted by ρ_c for which the denominator of (18) is zero; these are the geometrical reflection points x_c where (equation (16))

$$\frac{\mathrm{d}\rho(x = x_{\mathrm{c}}(\rho_{\mathrm{c}}))}{\mathrm{d}x} = 0. \tag{22}$$

By contrast with the general function g(r), which is discontinuous, or at worst logarithmically singular, at these points, the present function $\gamma(\rho)$ actually has poles there, and is therefore potentially a more sensitive indicator of the form of the surface. If we ignore the rare points where higher order reflections occur (characterized by

$$\frac{\mathrm{d}^2 \rho(x = x_{\rm c}(\rho_{\rm c}))}{\mathrm{d}x^2} = 0$$

as well as (22)), then there are just two types of reflection point, corresponding to the first $(d^2\rho/dx^2 > 0)$ and the last $(d^2\rho/dx^2 < 0)$ contact of the ρ circles with parts of the surface (in figure 4, x_6 is a point of the first type, and x_4 a point of the second type). The form of $\gamma(\rho)$ near to the singularity can be obtained by expanding (18) and (16) about ρ_c and x_c , with the result

$$\gamma(\rho) \simeq \text{Re} \frac{2}{h - f(x_c)} \left[\left\{ \left(1 + \frac{df(x_c)}{dx} \right)^2 + (h - f(x_c)) \frac{d^2 f(x_c)}{dx^2} \right\} (\rho^2 - \rho_c^2) \right]^{-1/2}$$
(23)

an expression that describes reflections of both types, and reduces to (21) for a flat

horizontal surface, which is a reflection of the first type with $x_c = 0$, $\rho_c = h$. Equation (23) shows that only weak singularities in $\gamma(\rho)$ are produced by highly curved surfaces, for which

$$\left| h \frac{\mathrm{d}^2 f(x_{\mathrm{c}})}{\mathrm{d}x^2} \right| \gg 1.$$

A moment's thought about the possible topologies of intersection of the ρ circles with surfaces which are continuous in slope soon shows that any reflection point of type two must be associated with two reflection points of type one, occurring for smaller ρ . The simplest surface for which this occurs is shown in figure 5(a): reflection occurs at ρ_1 (type one), ρ_2 (type one) and ρ_3 (type two). The two points of type one each give rise to two intersections; the intersection originating from ρ_1 which moves off to infinity towards the right causes a singularity shown in figure 5(b), while the intersection from ρ_2 which moves off to infinity towards the left causes a singularity shown in figure 5(c). The other two intersections from ρ_1 and ρ_2 approach and coalesce at the singularity occurring at ρ_3 which is the reflection point of type two, as shown in figure 5(d) and e. The total function $\gamma(\rho)$ is the sum of all these singular contributions, and is shown in figure 5(f).

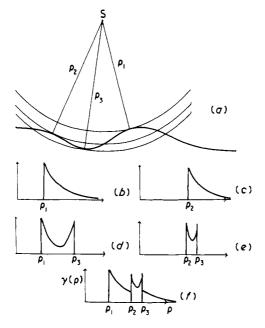


Figure 5. Contributions to $\gamma(\rho)$ from a surface with three reflection points. (a) Identification of reflection points on surface; (b) contribution from right hand intersection due to ρ_1 ; (c) contribution from left hand intersection due to ρ_2 ; (d) contribution from left hand intersection due to ρ_1 ; (e) contribution from right hand intersection due to ρ_2 ; (f) form of the total function $\gamma(\rho)$.

This association between the ridges and valleys of the landscape as seen from S, and the singularities of $\gamma(\rho)$, suggests that it might be possible to deconvolute the observations and determine f(x) (subject of course to the obvious left-right ambiguity) merely from the form and type of the singularities. To examine this, consider the

function $\gamma(\rho)$ shown in figure 6(a): there are clearly five reflection points, the three nearest the source (ρ_1, ρ_2, ρ_3) being of type one, and the remaining two (ρ_4, ρ_5) being of type two. Thus there are three ridges and two valleys in the landscape. However, we cannot decide from this alone whether for instance, the deeper valley (ρ_5) lies between the two highest ridges $(\rho_1 \text{ and } \rho_2)$ as in figure 6(b), or between the two lowest ridges, as in figure 6(c). Nor is it possible to resolve this and similar ambiguities by appealing to equation (23), which relates the curvature at a reflection point to the strength of the resulting singularity in $\gamma(\rho)$ since it is possible to adjust the curvatures at ρ_1, \ldots, ρ_5 in figure 6(b and c) independently. It would, of course, be possible to resolve the ambiguities by moving the source along the x axis and using the extra dimension of information thus obtained, possibly in the form of a plot of the critical reflection radii ρ_c against x; this question will be examined in a later paper.

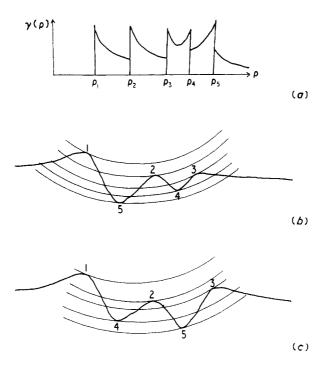


Figure 6. (a) Typical form for $\gamma(\rho)$. (b) and (c): two of the possible surfaces that could give rise to the same $\gamma(\rho)$.

In these illustrative examples we have assumed for simplicity that the surface is smooth on a small scale; but any real surface contains irregularities on all scales of fineness, and similar remarks can be made about $\gamma(\rho)$ as were made in § 3 about the more general function g(r), except that $\gamma(\rho)$, because of its greater sensitivity to the form of the surface, is even more mathematically pathological than g(r); in fact it is singular at every point! But the singularities due to dust, atoms etc are weak (cf (23) and the succeeding discussion), so that the average value of $\gamma(\rho)$, when smoothed over small ranges, is well defined, and graphs showing $\gamma(\rho)$ with separated singularities (eg figures 5 and 6) refer in practice to the case of smooth surfaces, where the reflection can be described by geometrical optics, since the smoothing distance that must be employed

in practice depends on the incident signal F(t) being of order $\lambda_0/4$ as we have already discussed in connection with g(r).

It is important to remember from the discussion in § 3 concerning g(r) that this smoothing must be carried out directly on $\gamma(\rho)$ and not on the surface itself; its effect on inferences about the surface depends on the height h of the source. To see this, consider figure 7(a, b, and c) which shows the same smooth surface viewed from various distances;

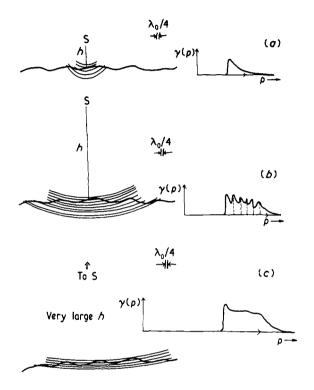


Figure 7. The same surface viewed from three distances, and the resulting smoothed functions $\gamma(\rho)$. (a) h small, only one reflection point; (b) h medium, several reflection points, separated by distances exceeding $\lambda_0/4$; (c) h large, many reflection points, separated by distances less than $\lambda_0/4$.

it is clear that even for the most gently undulating surface the nearly-straight ρ circles for large h will intersect at a large number of points, giving a function $\gamma(\rho)$ containing a large number of reflection singularities. Eventually these singularities will occur at values of ρ which are closer together than $\lambda_0/4$, and it will no longer be possible to discern them (figure 7(c)) by processing the received signal according to equation (20) or any similar prescription; in another terminology we might say that the *Fraunhofer limit* has been reached.

5. Conclusions

In this preliminary survey of the rough-surface deconvolution problem, we have drawn attention to the functions g(r) and $\gamma(\rho)$ as being the most natural and direct embodiments

of the information obtainable about the surface by analysis of the echo and pulse involving a single source and receiver point. Neither of these functions are mathematically well behaved for real surfaces (their derivatives do not exist at any point), but vary over infinitely small distances, a fact which corresponds to the fine detail that can be probed if the exploring pulse contains high frequencies. In practice the upper limit of frequencies which are significant in the incident pulse spectrum ($\simeq \omega_0$) means that we must deal, not with g(r) and $\gamma(\rho)$ directly, but with the smoothed forms of these functions averaged over a distance of about $\lambda_0/4$.

The two surface functions depend on the undulations of the surface via the topology of the contours of constant distance from the source. The points where this topology changes are the geometrical reflection points, which, if they are much further apart than the minimum distance $(\lambda_0/4)$ that can be resolved, dominate the form of g(r) and $\gamma(\rho)$; this explains the geometrical optics limit for pulses.

In future work we propose to examine the relation between the statistical properties of g(r) and $\gamma(\rho)$ and the statistics of the surface. Further, we intend actually to calculate numerically the integrals (13) and (20) for ultrasonic pulses and echoes involving artificial rough surfaces. Finally, we shall examine the way in which further information about the surface can be extracted from the variations of g(r) and $\gamma(\rho)$ with source position.

6. Acknowledgments

I would like to thank Professor J Nye for introducing me to this problem, and for his keen criticisms which led to substantial alterations to this paper. I would also like to thank Dr D Greenwood for a helpful suggestion.

Appendix 1

It is possible to derive our basic formula (4) from the results given by Beckmann and Spizzichino (1963, p 27), by writing the incident spherical pulse (3) as a spectrum of harmonic plane waves of various frequencies and directions. It is more illuminating, however, to proceed directly from the familiar Kirchhoff diffraction integral. The physical assumption is that the wavefunction on our perfectly reflecting surface is zero, and the Kirchhoff approximation is that this can be achieved by making the outgoing wave on the surface equal and opposite to the incoming wave there. The wave received back at the source is then given by an integral over the rough surface Σ . Making use of the identity of the source and receiver positions enables us to write the contribution to the echo from the element $d\Sigma$ of the surface distant r from S in the following form (see eg Longhurst 1957, p 193), which is valid if $r \gg \lambda_0$:

$$d\psi(t) = -\frac{d\Sigma F'(t - (2r/c))\mathbf{n} \cdot \hat{\mathbf{r}}}{2\pi c r^2}.$$
(A.1)

In this expression n is the unit normal to the surface at the point considered (pointing into the illuminated half-space) and \hat{r} is the unit vector along the line from $d\Sigma$ to S (see figure 1); $n \cdot \hat{r}$ is the obliquity factor for this case.

If z is a coordinate, measured vertically upwards from O, then the transformation required in order to integrate over the reference plane instead of Σ is

$$d\mathbf{R} = n_z d\Sigma. \tag{A.2}$$

The explicit expressions for n and f in (R, z) coordinates are easily derived to be

$$n(R) = \left(\frac{-\nabla_R f(R)}{(1+|\nabla_R f(R)|^2)^{1/2}}, \frac{1}{(1+|\nabla_R f(R)|^2)^{1/2}}\right)$$

and

$$\hat{\mathbf{r}} = \left(-\frac{\mathbf{R}}{r}, \frac{h - f(\mathbf{R})}{r}\right).$$

These explicit forms, and the transformation (A.2), can be used with the formula (A.1) for the wave radiated from an element of the surface, to yield the following expression for the received signal:

$$\psi(t) = -\frac{1}{2\pi c} \int \int d\mathbf{R} \frac{F'(t - (2r(\mathbf{R})/c))}{r^3(\mathbf{R})} \{ \mathbf{R} \cdot \nabla_{\mathbf{R}} f(\mathbf{R}) + (h - f(\mathbf{R})) \}. \tag{A.3}$$

By writing $\nabla_{\mathbf{R}} f$ in terms of $\nabla_{\mathbf{R}} r$ from equation (1), this can be written as

$$\psi(t) = -\frac{1}{2\pi c} \int \int d\mathbf{R} \frac{F'(t - (2r(\mathbf{R})/c))}{r(\mathbf{R})(h - f(\mathbf{R}))} \left(1 + \frac{\mathbf{R} \cdot \nabla_{\mathbf{R}} r(\mathbf{R})}{r(\mathbf{R})} \right). \tag{A.4}$$

Thus we have arrived at a Kirchhoff integral which is identical with (4) except for the extra term $R \cdot \nabla r/r$, whose effect on $\psi(t)$ we must now assess. Integrating the second term by parts, we may write, symbolically

$$\psi = -\frac{1}{2\pi c} \int d\mathbf{R} \frac{F'}{rh} - \frac{1}{4\pi} \int d\mathbf{R} \frac{\mathbf{R} \cdot \nabla_{\mathbf{R}} F}{r^2 h}$$

$$= -\frac{1}{2\pi c} \int d\mathbf{R} \frac{F'}{rh} + \frac{1}{4\pi} \int d\mathbf{R} F \nabla \cdot \left(\frac{\mathbf{R}}{r^2 h}\right)$$

$$\simeq -\frac{1}{2\pi} \int d\mathbf{R} \frac{F'}{crh} + \frac{1}{2\pi} \int d\mathbf{R} \frac{F}{r^2 h}$$

where we have assumed $h \gg f$ and $r \simeq h$ as order of magnitude estimates. Now if the incident signal has a predominant frequency ω_0 we may write

$$|F'| \simeq \omega_0 |F|$$

so that the second integral is negligible in comparison with the first if

$$\frac{1}{r} \ll \frac{\omega_0}{c}$$

that is, $\lambda_0 \ll h$, which is precisely the condition (5) as well as the condition for the Kirchhoff theory (A.1) to be valid.

Our convenient basic equation (4) is thus completely equivalent to the Kirchhoff integral (A.4) with the much more complicated obliquity factor. In fact for theoretical purposes (4) is much better than (A.4) since it gives the flat-surface limit (6) exactly, whereas if we use (A.4) we only get this limit after a messy approximation procedure.

Appendix 2

In order to understand the geometrical-optics limit for the function g(r), it is not convenient to work with the direct representation (9) involving the r contours. An equivalent expression can be obtained by writing, in equation (4)

$$F'\left(t - \frac{2r(\mathbf{R})}{c}\right) = \int_0^\infty dr \ \delta(r - r(\mathbf{R}))F'\left(t - \frac{2r}{c}\right)$$

which yields, on comparison with (8) and replacement of the delta function by its Fourier transform, the following results:

$$g(r) = \frac{1}{r} \iint d\mathbf{R} \frac{\delta(r - r(\mathbf{R}))}{h - f(\mathbf{R})}$$

$$= \frac{1}{2\pi r} \int_{-\infty}^{\infty} dk \exp(-ikr) \iint d\mathbf{R} \frac{\exp(ikr(\mathbf{R}))}{h - f(\mathbf{R})}.$$
(A.5)

Considered for fixed k as a function of position R on the reference surface, the integrand is oscillatory except at the geometrical reflection points where equation (11) holds. There will always be at least one such point (even for a flat surface), and we label their positions by R_i and the radii of the corresponding r spheres by r_i . The geometrical-optics approximation results from assuming that all the contributions to g(r) come from the points R_i and that these contributions can be found by evaluating the second integral in (A.5) by the method of stationary phase, which amounts to replacing $\exp(ikr(R))$ by its (two dimensional) osculating gaussian functions near the points R_i . We ignore for the moment questions as to the validity of this procedure, and merely quote the result

$$g(r) \simeq \frac{1}{r} \int_{-\infty}^{\infty} dk \frac{\exp(-ikr)}{|k|} \sum_{i} \frac{\exp(ikr_{i})\epsilon_{x}\epsilon_{y}}{(h - f(\mathbf{R}_{i}))(|\det||\partial^{2}r(\mathbf{R})/\partial\mathbf{R}_{i}|\partial\mathbf{R}_{k}|||)^{1/2}}$$
(A.6)

where the phase factors ϵ_x and ϵ_y are given by

$$\epsilon_j^{(i)} = \exp\left(\pm \frac{i\pi}{4}\right) \quad \text{if} \quad k \frac{\partial^2 r(\mathbf{R}_i)}{\partial \mathbf{R}_i^2} \ge 0.$$
(A.7)

The denominator of expression (A.6) vanishes, and the method of stationary phase cannot be applied in the simple form used here, if any of the reflection points involves focusing, for which equation (12) holds; such cases arise only rarely, and we do not consider them here.

At a reflection point \mathbf{R}_i the topology of the r contours changes, and the different phase factors (A.7) correspond to the various ways in which this can occur; there are three cases:

Case (i): hilltops (minima of $r(\mathbf{R})$, elliptical contours degenerating to a point at \mathbf{R}_i). Here $\partial^2 r/\partial x^2$ and $\partial^2 r/\partial y^2$ are both positive, and application of (A.7) gives, for the combination of phases occurring in (A.6)

$$\frac{\epsilon_x \epsilon_y}{|k|} = \frac{\mathrm{i}}{k}$$

so that it is necessary to evaluate the integral

$$I(r) \equiv i \int_{-\infty}^{\infty} \frac{dk}{k} \exp\{ik(r_i - r)\}$$
 (A.8)

in order to find how the hilltop at R_i contributes to g(r). This integral is defined by the deportment of the path of integration in the complex k plane in relation to the pole at k = 0, and this can be established uniquely by using the condition that g(r) must always be positive (cf (11) or the first of equations (A.5)). The result of the integration is a step function

$$I(r) = 2\pi\theta(r-r_i)$$

so that the contribution to g(r) from all hilltops (htp) is

$$g^{(\text{htp})}(r) = \frac{2\pi}{r} \sum_{\substack{i \text{(hills)} \\ \text{(hills)}}} \frac{\theta(r - r_i)}{(h - f(\mathbf{R}_i))(|\det|| \partial^2 r(\mathbf{R}_i) / \partial R_j \partial R_k || |)^{1/2}}.$$
 (A.9)

A special case of this formula is the flat surface, for which there is only one hilltop, at $\mathbf{R}_i = 0$, $r_i = h$, and g(r) is given by equation (10).

Case (ii): floors of hollows (maxima of $r(\mathbf{R})$, elliptical contours degenerating to a point at r_i). There $\partial^2 r/\partial x^2$ and $\partial^2 r/\partial y^2$ are both negative, and an argument precisely similar to that given for case (i) leads to the result, for floors of hollows (foh)

$$g^{(\text{foh})}(r) = \frac{2\pi}{r} \sum_{\substack{i \text{(hollows)}}} \frac{\theta(r_i - r)}{(h - f(\mathbf{R}_i))(|\det||\hat{\sigma}^2 r(\mathbf{R}_i)/\hat{\sigma} R_j \hat{\sigma} R_k|||)^{1/2}}.$$
 (A.10)

Case (iii): saddle points (contour for $r = r_i$ crosses itself at R_i). In this case $\partial^2 r/\partial x^2$ and $\partial^2 r/\partial y^2$ are of opposite sign, and the integral to be evaluated for g(r) is

$$I \equiv \int_{-\infty}^{\infty} \frac{\mathrm{d}k}{|k|} \exp\{ik(r_i - r)\}$$
 (A.11)

which is infinite no matter how it is interpreted! To understand this behaviour it is necessary to consider the osculating surface which is implied by the application of the method of stationary phase to this case; in a suitable coordinate system, the approximating function $r(\mathbf{R})$ has the form

$$r(\mathbf{R}) = r_i + ax^2 - by^2 (A.12)$$

where a and b are positive. The contours corresponding to this expression are shown in figure 8; they are infinitely long, and that, basically, is why the integral in (A.11) diverges, since according to (9) the function g(r) depends on an integral around the r contour.

For any real surface, of course, the r contours are of finite length, the commonest shape for a contour going through a saddle point being the 'figure of eight' (cf figure 3(c)). The contours corresponding to all r near to r_i will have approximately the same length, and this fact alone will result in a large contribution to g(r). But we are interested in the qualitative feature that arises from the neighbourhood of the saddle point, irrespective of what the contour does elsewhere. To extract this contribution, we consider the surface defined by (A.12) as being cut off at some radius R = L, and work out g(r) in the limit of large L; the required contribution will be that part of the result which is independent of L.

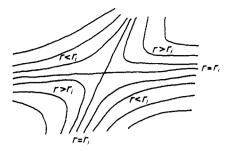


Figure 8. Contours near a saddle point.

From (A.5) and (A.12), the integral to be evaluated is, using polar coordinates R. ϕ for R

$$g^{(i)}(r) = \frac{1}{2\pi r(h - f(\mathbf{R}_i))} \int_{-\infty}^{\infty} dk \exp\{-ik(r - r_i)\} \int_{0}^{L} dRR$$
$$\times \int_{0}^{2\pi} d\phi \exp\{ikR^2(a\cos^2\phi - b\sin^2\phi)\}.$$

The integration over R is trivial, and the resulting integral involving k is similar to (A.8), which leads to

$$g^{(i)}(r) = \frac{1}{r(h - f(\mathbf{R}_i))} \int_0^{\pi} d\phi \frac{\theta(r - r_i) - \theta(r - r_i - L^2(a\cos^2\phi - b\sin^2\phi))}{a\cos^2\phi - b\sin^2\phi}.$$

A tedious inspection shows that the step functions prevent the pole contributing to the range of integration, and also shows that $g^{(i)}(r)$ is positive. The integral over ϕ is tabulated (Gradshteyn and Ryzhik 1965, p 152), and in the limit of large L, when the zero of the argument of the second step function nearly coincides with the pole, the result can be written explicitly as

$$g^{(i)}(r) = \frac{1}{r(h - f(\mathbf{R}_i))(ab)^{1/2}} \ln \left(\frac{2L^2 ab}{|r - r_i|(a + b)} \right).$$

The logarithm may be split into the sum of a term depending on L but not r, which corresponds to the contribution from the distant parts of the contours, and a term independent of L which gives the characteristic form of g(r) as r passes through a saddle point (sad). This second contribution is the one we are seeking; it can be expressed in invariant form as

$$g^{(\text{sad})}(r) = \frac{2}{r} \sum_{i} \frac{-\ln(|r - r_i|(|\det||\partial^2 r(\mathbf{R}_i)/\partial R_j \partial R_k|||)^{1/2})}{(h - f(\mathbf{R}_i))(|\det||\partial^2 r(\mathbf{R}_i)/\partial R_j \partial R_k|||)^{1/2}}.$$
 (A.13)

We can sum up our conclusions about the geometrical-optics approximation to g(r) by saying that the function is dominated by step discontinuities which appear (with increasing r) at hilltops and disappear at the floors of hollows, and by logarithmic singularities which appear whenever the r sphere touches a saddle point. In order to ascertain how far the true function g(r) will resemble the approximation we have derived, it is necessary to examine the validity of the stationary-phase method that we have used. If the surface f(R) is smooth, then our approximation (A.6) is justified for those k which

are so large that the 'wavelength' $2\pi/k$ is small compared with the variation of r(R) over the surface. This will always be true for sufficiently large k, so that the fine structure of g(r)—the discontinuities and singularities—is given correctly by our approach. But for small k the method breaks down, and the slow variation of g(r) between adjacent reflection points is not given correctly by any of our three formulae (A.9), (A.10) or (A.13); this part of the behaviour of g(r) depends in detail on the shape of the r contours, and must be calculated directly from (9).

For real surfaces, which are not smooth but vary on all scales, the function g(r) is pathological, and must be 'interpreted' in a way that depends on the form of the incident pulse, as discussed in § 3 of the paper.

References

Abramowitz M and Stegun I A 1964 Handbook of Mathematical Functions (Washington: National Bureau of Standards)

Beckmann P and Spizzichino A 1963 The Scattering of Electromagnetic Waves from Rough Surfaces (London: Pergamon)

Berry M V and Gibbs D F 1970 Proc. R. Soc. A 314 143-52

Bôcher M 1926 An Introduction to the Study of Integral Equations (London: Cambridge University Press)

Erdèlyi A 1954 Tables of Integral Transforms (New York: McGraw-Hill)

Gradshteyn I S and Ryzhik I M 1965 Table of Integrals, Series and Products (New York and London: Academic Press).

Longhurst R S 1957 Geometrical and Physical Optics (London: Longmans)

Mott N F and Massey H S W 1965 The Theory of Atomic Collisions (Oxford: Clarendon) chap VII § 8

Nye J F 1970 Proc. R. Soc. A 315 381-403

Robin G de Q, Evans S and Bailey J T 1969 Phil. Trans. R. Soc. 265 437-505

Scheidegger A E 1970 Theoretical Geomorphology (London: Allen and Unwin)