INSTITUTO DE FÍSICA

preprint

IFUSP/P-227

CAUSALITY AND ANALYTICITY IN OPTICS

by

H.M. Nussenzveig

Instituto de Física, Universidade de São Paulo,
São Paulo, Brazil

Invited paper presented at the
International Comission for Optics Conference
"Optics in Four Dimensions"

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA
Caixa Postal - 20.516
Cidade Universitária
São Paulo - BRASIL

CAUSALITY AND ANALYTICITY IN OPTICS

H. M. Nussenzveig
Instituto de Física, Universidade de São Paulo, Brazil

ABSTRACT

In order to provide an overall picture of the broad range of optical phenomena that are directly linked with the concepts of causality and analyticity, the following topics are briefly reviewed, emphasizing recent developments: 1) Derivation of dispersion relations for the optical constants of general linear media from causality. Application to the theory of natural optical activity. 2) Derivation of sum rules for the optical constants from causality and from the short-time response function (asymptotic high-frequency behavior). Average spectral behavior of optical media. Applications. 3) Role of spectral conditions. Analytic properties of coherence functions in quantum optics. Reconstruction theorem. 4) Phase retrieval problems. 5) Inverse scattering problems. 6) Solution of nonlinear evolution equations in optics by inverse scattering methods. Application to self-induced transparency. Causality in nonlinear wave propagation. 7) Analytic continuation in frequency and angular momentum. Complex singularities. Resonances and natural-mode expansions. Regge poles. 8) Wigner's causal inequality. Time delay. Spatial displacements in total reflection. 9) Analyticity in diffraction theory. Complex angular momentum theory of Mie scattering. Diffraction as a barrier tunnelling effect. Complex trajectories in optics.

INTRODUCTION

"On revient toujours à ses premiers amours"

It is often useful and illuminating to survey a broad field of research from the vantage point of some general unifying principle, recognizing common features in apparently very diverse subjects. The present work is devoted to the task of discussing the relevance to optics of the principle of causality and of the closely related concept of analyticity. This is done by very briefly reviewing a number of illustrative examples, both in classical and in quantum optics.

1. DERIVATION OF DISPERSION RELATIONS

It was in optics that dispersion relations were first applied to physics and that their connection with causality was first realized[1]. Typically, in the propagation of light through a linear medium, the medium response $r(t)$ (we omit all tensor indices) is linked with the excitation $e(t')$ through a response function $g(t,t')$, so that, for a time-invariant medium,

$$r(t) = \int_{-\infty}^{\infty} g(t-t')\, e(t')\, dt' \,, \qquad (1.1)$$

corresponding to a constitutive relation for the Fourier transforms,

$$R(\omega) = G(\omega)\, E(\omega) \,, \qquad (1.2)$$

where the material constant $G(\omega)$ is a function of the circular frequency ω (dispersion).

According to the "primitive causality condition", the response at time t cannot depend on the excitation at later times, i.e.,

$$\delta r(t)/\delta e(t') = 0 \,, \quad t' > t \,, \qquad (1.3)$$

where $\delta r/\delta e$ denotes the functional derivative. Thus,

$$g(\tau) = 0 \,, \quad \tau < 0 \,, \qquad (1.4)$$

so that

$$G(\omega) = \int_{0}^{\infty} g(\tau) \exp(i\omega\tau)\, d\tau \,. \qquad (1.5)$$

Since this is a half-range Fourier transform, $G(\omega)$ is the boundary value of an analytic function, regular in the half-plane $\mathrm{Im}\,\omega > 0$.

The derivation of dispersion relations for $G(\omega)$ is usually based on additional information concerning high-frequency behavior. Thus, if $G(\omega)$ is the complex dielectric susceptibility $\chi'(\omega)$, the expected behavior at frequencies well above optical absorption bands is free electron-like,

$$\chi'(\omega) \approx -\mathcal{N} e^2/m\omega^2 \,, \quad \omega \to \infty \,, \qquad (1.6)$$

where $\mathcal{N}$ is the electron density. If, e.g., $G(\omega)$ is square integrable on the real axis (as far as the behavior at infinity is concerned, (1.6) is more than enough to guarantee this for $\chi'(\omega)$), it follows[2] that

$$G(\omega) = \frac{1}{\pi i}\, P\int_{-\infty}^{\infty} \frac{G(\omega')}{\omega'-\omega}\, d\omega' \,, \qquad (1.7)$$

where P denotes the Cauchy principal value. Thus, the real (dispersive) and imaginary (absorptive) parts of $G(\omega)$ obey the Kramers-Kronig (dispersion) relations, i.e., they are each other's Hilbert transforms.

The asymptotic high-frequency behavior is not always as readily determined as for (1.6); it may happen, for instance, that the macroscopic description associated with the constitutive relation (1.2) breaks down at high frequencies. However, it may still be possible to derive dispersion relations by an alternative procedure, based on the properties of a passive linear system[3,4].

A passive system is one that can absorb or store energy, but cannot generate it. Every passive system is causal, but the converse need not be true. What is important is that the response function of a passive linear system not only has a regular analytic continuation in a half-plane, but it also has a positive semi-definite imaginary part in it, i.e., it is a Herglotz function[2]. This turns out to be a strong constraint on the high-frequency growth properties, leading to the dispersion relations in a slightly generalized form.

If one assumes that the system is not just passive, but that it is actually dissipative, i.e., that it has strictly positive absorption at <u>all</u> positive frequencies, one gets even stronger constraints on the asymptotic behavior, as well as the absence of zeros of $G(\omega)$ in $\mathrm{Im}\,\omega > 0$. This stronger assumption is usually made in the work of the Russian school[5,6].

The dielectric susceptibility is associated with a local response function of the type (1.1). This is not true, however, for the important case of the complex refractive index $N'(\omega) = n(\omega) + i\kappa(\omega)$, which is connected with the propagation factor $\exp(i\,N'\omega z/c)$ of a monochromatic plane wave. Although $N'(\omega) = \sqrt{\varepsilon'(\omega)}$ (taking $\mu = 1$), the analyticity of ε' does not entail that of N', unless one accepts[5] the strong dissipativity assumption to exclude branch points due to eventual zeros of ε' in $\mathrm{Im}\,\omega > 0$. This is unnecessary, however,[2] if one employs the "relativistic causality condition", according to which <u>no signal can propagate with velocity greater than c</u>. Even in the presence of conductivity, which gives rise to a $\omega^{-1/2}$ singularity at $\omega = 0$, this condition allows one to derive the dispersion relations for the refractive index.[7]

The derivation of dispersion relations for more general linear media, including the effects of spatial dispersion, has also been discussed.[6,8] The response function in such media goes beyond the usual electric dipole approximation, bringing in higher multipoles.

The theory of natural optical activity provides one of the simplest illustrations of these effects. The normal modes of propagation in an optically active medium are circularly polarized waves. Besides circular double refraction, such a medium also shows circular dichroism, i.e., both the phase velocity and the extinction are different for right and left circular polarization, leading to different complex refractive indices $N'_+(\omega)$ and $N'_-(\omega)$, respectively.

As a consequence of these features, the medium rotates the plane of polarization of linearly polarized light and also converts it into elliptically polarized light. The rotatory power $\rho(\omega)$ and the ellipticity $\phi(\omega)$ per unit length of the medium are related with the complex refractive indices by [9]

$$\rho(\omega) + i\phi(\omega) = \omega \left[N'_-(\omega) - N'_+(\omega)\right]/2c \,. \qquad (1.8)$$

The usual derivation of dispersion relations does not go through[10] for $N'_\pm$, because one cannot build up a real field using only one of them; we need both, in view of the crossing relation

$$\left[N'_+(\omega)\right]^* = N'_-(-\omega) \quad (\omega \text{ real}). \qquad (1.9)$$

However, one can derive[11] dispersion relations between ρ and ϕ from primitive causality, by applying it to the constitutive equation[9]

$$\vec{B}(\omega) - \vec{H}(\omega) = i\omega\, g(\omega)\, \vec{E}(\omega) \,, \qquad (1.10)$$

where $g(\omega)$ is related with ρ and ϕ by

$$\rho(\omega) - i\phi(\omega) = \omega^2 g(\omega)/c \,. \qquad (1.11)$$

The dispersion relations between ρ and ϕ follow from those between Reg and Img . Physically, (1.10) describes the induced magnetic dipole moment that arises from a time-varying electric field, due to the helical structure that is associated with the natural optical activity of the medium. Thus, this procedure captures the basic causal relationship that is responsible for the effect.

2. SUM RULES FOR THE OPTICAL CONSTANTS

We usually know a bit more about the response function besides its causal behavior represented by the dispersion relations. We can take advantage of this extra knowledge to get additional constraints.

Suppose, for instance, that the excitation takes the form of a step function in time, switched on at $t = 0$. Causality requires that the response must vanish for $t < 0$, but it does not prevent a jump at $t = 0$, whereas we actually expect that the response will rise continuously from zero at $t = 0$. The short-time behavior of the response is related with the high-frequency asymptotic behavior of $G(\omega)$. For the dielectric susceptibility $\chi'(\omega)$, this is given by (1.6): at high frequencies, we get just the inertial free-electron response, which prevents a sudden jump.

Instead of (1.6), a slower decrease at infinity, e.g., like $\omega^{-1-\varepsilon}$, $\varepsilon > 0$, would suffice for square integrability. Thus, in this case, the dispersion integrals converge faster than they "need" to at the high-frequency end; this is called "superconvergence". By comparing the actual high-frequency behavior with the high-frequency limit of the Kramers-Kronig relations, one gets[12,13] additional constraints in the form of sum rules.

Besides the well-known "f-sum rules", one finds a whole set of new ones, which may be called "inertial sum rules", because of their connection with the inertial response. Thus, for the dielectric constant of an isotropic nonconducting medium, one finds

$$\int_0^\infty \left[\mathrm{Re}\ \varepsilon'(\omega) - 1\right] d\omega = 0\ , \qquad (2.1)$$

which is equivalent[13] to the condition $g(0) = 0$ for the associated response function $g(\tau)$ of (1.4).

Similarly, if $n(\omega) = \mathrm{Re}\ N'(\omega)$ is the real refractive index of an isotropic medium, conducting or not, one has

$$\int_0^\infty \left[n(\omega) - 1\right] d\omega = 0\ . \qquad (2.2)$$

For the optical rotatory power of a medium with natural optical activity, one finds[11,14]

$$\int_0^\infty \rho(\omega) d\omega = 0\ . \qquad (2.3)$$

The inertial sum rules (2.2) and (2.3) may be interpreted in terms of the average behavior of optical media over the whole frequency spectrum. According to (2.2), frequency ranges where $n(\omega) > 1$ must be compensated by other ones where $n(\omega) < 1$, in such a way that the total area under the curve of $n(\omega) - 1$ vanishes. Thus, the phase velocity is sometimes greater than c and sometimes smaller, but, on the average over the whole spectrum, in the sense of (2.2), it is just equal to c.

Similarly, according to (2.3), optically active media are "middle-of-the-road": "leftist" and "rightist" frequency ranges compensate each other on the average.

These results might be summed up by saying that, on the average over the whole frequency spectrum, optical media tend to behave like the vacuum!

From (1.6) follows the well-known fact that $n(\omega) < 1$ in the asymptotic high-frequency domain, and it might be thought that the whole negative contribution to (2.2) arises from this domain. That this is not so follows from the fact[12] that $n(\omega) - 1$ satisfies a sum rule analogous to (2.2), but with the extra weighting factor $\omega\kappa(\omega)$ (where $\kappa(\omega) = \mathrm{Im}\, N'(\omega)$), which reduces the high-frequency contribution and enhances that of the absorption bands. Similarly,[11] $\rho(\omega)$ satisfies sum rules analogous to (2.3), but with extra weighting factors $\omega\phi(\omega)$ and $\phi(\omega)/\omega$.

In terms of a Lorentz-like dispersion formula, the compensation of positive and negative contributions in the inertial sum rules can be physically related with the change from in-phase to out-of-phase character of the response of a forced damped oscillator as it goes across a resonance.

The new sum rules for the refractive index and dielectric constant have been verified experimentally for a variety of substances, including both metals and insulators.[13,15-18] For metals, the refractive index sum rule (2.2) is particularly sensitive to the low-frequency behavior, because of the $\omega^{-1/2}$ singularity of $n(\omega)$ at $\omega = 0$; this property has been employed[13] to choose between different sets of optical constants, based on different extrapolation procedures in the far infrared. The sum rules with extra weighting factors play a similar role, emphasizing the contributions of different frequency ranges.

3. ANALYTIC PROPERTIES OF COHERENCE FUNCTIONS

Coherence functions in quantum optics are generally defined by[19]

$$G^{(n,m)}(x_1,\ldots,x_n;\, x_{n+1},\ldots x_{n+m})$$
$$= \mathrm{Tr}\left[\rho E^-(x_1)\ldots E^-(x_n)E^+(x_{n+1})\ldots E^+(x_{n+m})\right] , \qquad (3.1)$$

where x_i are spacetime points, ρ is the density operator, and E^- and E^+ are, respectively, creation and annihilation operators associated with the quantized radiation field (tensor indices are omitted). For $m = n$ and $x_{n+i} = x_i$, the function $G^{(n,n)}$ can be interpreted in terms of an n-fold delayed coincidence counting rate involving n photodetectors. For $m = n = 1$, (3.1) corresponds to the mutual coherence function between the spacetime points x_1 and x_2 .

The normal ordering in (3.1) arises from the fact that ordinary photon detectors operate by absorption.[20] There are certain parallels[19] between the set of coherence functions (3.1) in quantum optics and the set of Wightman functions in quantum field theory. In particular, the analogue of the "spectral condition" is the fact that the Fourier transforms of $G^{(n,m)}$ with respect to time contain only positive frequencies in the m annihilation operators, and only negative frequencies in the n creation operators.

It follows, by analogy with (1.5) (but note that the roles of frequency and time are reversed!), that $G^{(n,m)}$ is the boundary value of an analytic function of the time variables, regular in $\mathrm{Im}(x_i^0) > 0$ for $1 \leqslant i \leqslant n$ and in $\mathrm{Im}(x_i^0) < 0$ for $n+1 \leqslant i \leqslant n+m$, where $x_i^0 = ct_i$ is the time component of the 4-vector x_i. In the particular case of a time-stationary field, $G^{(1,1)}(\vec{x}_1,t_1;\ \vec{x}_2,t_2)$ depends only on $\tau = t_2-t_1$, and

$$G^{(1,1)}(\vec{x}_1,\vec{x}_2,\ \tau = t_2-t_1) = \int_0^\infty W(\vec{x}_1,\vec{x}_2,\omega)e^{-i\omega\tau}d\omega, \quad (3.2)$$

where W is the cross-spectral density, is regular in $\mathrm{Im}\ \tau < 0$. Note that this analyticity in time does not follow from causality, but rather from the normal ordering in (3.1), which expresses the privileged role of absorption as opposed to emission in photodetection. This may be likened to the passivity condition in Sect. 1.

Since this may look like a purely accidental feature of the detection process[20], one may wonder whether, by employing (3.1) to define the set of coherence functions, we do not lose some statistical information about the field. Given the set of coherence functions (3.1) for all values of m and n, can one reconstruct from it the density operator ρ associated with the field? This is analogous to reconstructing a field theory from the set of Wightman functions.

A <u>sufficient</u> condition for the reconstruction to be possible has been given[21,19] making use of the coherent state representation.[22] This representation, which emphasizes the correspondence with classical coherence theory and has played a central role in quantum optics, can be regarded as a one-to-one mapping between Hilbert space vectors and entire analytic functions[23], thus establishing another close link between analyticity and optics.

For the reconstruction problem, it suffices to consider a single mode of the radiation field (the general case corresponds to a direct product involving all the modes). Let $|z>$ be a coherent state associated with this mode (z is a complex eigenvalue of the annihilation operator). The above-mentioned sufficient condition then is

$$< z|\,\rho\,|z > = \mathcal{O}\left[\exp(-\,\varepsilon|z|^2)\right]\ ,\ |z| \to \infty \qquad (\varepsilon > 0)\ . \qquad (3.3)$$

Within limits set by the uncertainty principle, the left-hand side of (3.3) can be interpreted as a sort of "probability of finding the field in a coherent state $|z>$". According to (3.3), this "probability" should decrease at least exponentially with the average number of photons $\bar{n} = |z|^2$ in $|z>$, as $\bar{n} \to \infty$. The proof of the "reconstruction theorem"[21,19] makes use of the analytic properties of the norr ''y-ordered characteristic function in the coherent-state presentation.

The result is expected to remain valid under considerably more general conditions, but (3.3) is satisfied by all of the most widely employed models in quantum optics. The reconstruction theorem implies that the set of coherence functions (3.1) indeed contains all the statistical information about the field.

4. PHASE RETRIEVAL PROBLEMS

Applying (3.2) for $\vec{x}_1 = \vec{x}_2$, we see that the complex degree of temporal coherence $\gamma(\tau)$ and the spectral density are Fourier transforms of each other. Michelson's method of interference spectroscopy is based on this relationship. In practice, however, while $|\gamma(\tau)|$ (which is proportional to the visibility of the interference fringes) is readily measurable, it is very difficult to measure $\arg\gamma(\tau) = \phi(\tau)$, the phase of the degree of coherence. The problem of determining ϕ is known as the phase problem of coherence theory. Similar phase retrieval problems occur in x-ray crystallography and in image reconstruction.

According to (3.2), $\gamma(\tau)$ has a regular analytic continuation in $\mathrm{Im}\,\tau < 0$. Is this analyticity helpful in connection with the phase problem? We have

$$\ell n\,\gamma(\tau) = \ell n|\gamma(\tau)| + i\phi(\tau)\ , \qquad (4.1)$$

so that, if $\ell n\,\gamma$ has similar analytic properties, yielding a dispersion relation between ϕ and $\ell n|\gamma|$, we would have the solution of the phase problem. Such "logarithmic Hilbert transforms" have been discussed by Toll;[24] an important practical application is the determination of optical constants by reflectivity measurements.[25]

If there are zeros of $\gamma(\tau)$ in $\mathrm{Im}\,\tau < 0$, they are branch points for $\ln \gamma(\tau)$, so that the phase ϕ contains, besides the contribution from the "minimal phase", a Hilbert transform involving $\ln|\gamma(\tau)|$, an additional contribution from the zeros, known as the "Blaschke phase". In some cases, including that of blackbody radiation, the zeros are absent, and the solution of the phase problem is given by the minimal phase.[26]

In general, however, this is no longer true[27]: for a variety of practically important spectra, including natural (Lorentzian) and Doppler-broadened line shapes, as well as for band-limited spectra, one finds not only that $\gamma(\tau)$ has a large number of zeros, but also that their contribution to the phase is essential; the minimal phase is a very poor approximation.

Therefore, in order to solve the phase problem, one needs additional information. It is interesting to note that holography, the most successful practical method of image reconstruction, can be regarded as a device to solve the phase problem by generating a zero-free half-plane.[28]

Indeed, if we consider a band-limited function (one-dimensional case)

$$f(z) = \int_a^b F(k)\, \exp(ikz)\, dk \,, \qquad (4.2)$$

where k is a spatial frequency, then $f(z)$ is [29] an entire function of exponential type. By Rouché's theorem,[29] the function

$$f(z) + A \exp(icz) \,, \qquad (4.3)$$

where $|A| > |\, f(x) \,|$ (x real) and c lies outside the support of $F(k)$, i.e.,

$$c \notin [a, b], \qquad (4.4)$$

is still entire but has a zero-free half-plane, so that its phase may be reconstructed from its modulus by a logarithmic Hilbert transform.

The added plane wave term in (4.3) is a "reference function"[28] that plays exactly the same role as the reference beam in off-axis holography.[30] The off-axis condition corresponds to (4.4); this condition can also be enforced by spatial filtering, as in single-sideband holography.[31]

The proposal of adding a reference source to solve the phase problem of coherence theory[32,33] has been experimentally demonstrated[34,35] as a practically feasible method of solution.

5. INVERSE SCATTERING

The inverse scattering problem, in general terms, is that of reconstructing the scatterer given a set of scattering data. Depending on the choice of this set, there is a variety of different inverse problems, not only in optics, but also in many other fields.[36-38]

Analyticity plays an important role in the solution of most such problems. The physical reasons for this can be illustrated by a couple of examples.

Consider first the problem of diffraction by an aperture, employing the representation in terms of an angular spectrum of plane waves.[39-40] The angular distribution in Fraunhofer diffraction (scattering data) and the aperture distribution are related to each other by Fourier transformation. However, only the homogeneous plane wave components in the angular spectrum, which carry information about details larger than the wavelength, will survive in the Fraunhofer region. In order to recover the aperture distribution by an inverse Fourier transform, one needs also the inhomogeneous (evanescent) wave components, and one may try to relate them to the homogeneous ones by analytic continuation.

Consider next scattering by a three-dimensional scatterer. In the first Born approximation, the scattering amplitude $A(\vec{k}, \vec{k}')$ depends only on the difference $\vec{K} = \vec{k} - \vec{k}'$ between the incident and scattered wave vectors, and it is in fact just the Fourier transform of the scattering distribution with respect to $\vec{K}$. It then follows from an extension of the Paley-Wiener theorem that A is an entire function in the components of $\vec{K}$. For a nonspherically symmetric scatterer, already in this approximation, the analyticity of A suffices to show that the inverse scattering problem cannot be uniquely solved from scattering data obtained in a single experiment. This remains true for the exact amplitude.[41]

The most successful approach to the solution of the inverse scattering problem is the Gelfand-Levitan-Marchenko method[42-45] in nonrelativistic quantum scattering by a spherically symmetric potential $V(r)$, which has also been applied to optics, as will be seen in the next Section. The set of scattering data consists of the phase shift for a given partial wave, assumed known for all energies, plus the energies and normalization constants of the bound states associated with this partial wave. In the Gelfand-Levitan version (the Marchenko method is similar), the solution is given by

$$V(r) = 2 \frac{d}{dr} K(r,r) , \tag{5.1}$$

where $K(r, r')$, for each r , is the solution of the Gelfand-Levitan integral equation, a linear integral equation of the Fredholm type,

$$K(r,r') + G(r,r') + \int_0^r K(r,r'') \, G(r'',r')dr'' = 0, \quad (5.2)$$

whose kernel G is constructed entirely from the set of scattering data.

The essential feature of K for the success of the method is that it is a triangular kernel, i.e., that

$$K(r,r') = 0 \quad \text{for} \quad r' > r . \quad (5.3)$$

This looks like a causal property. Indeed, in a classical application to the plasma inverse problem, this property of the Gelfand-Levitan kernel has been shown[46] to follow from causality. For a cutoff potential in quantum scattering, it can also be traced back to primitive causality.[2]

The quantum inverse scattering problem has also been solved in one dimension, i.e., for scattering on the line[47-49]. However, for the three-dimensional problem , where the scattering amplitude $A(\vec{k}, \vec{k}')$ is given, the proposed solutions[50-51] are still incomplete.

6. NONLINEAR OPTICS

One of the most remarkable theoretical advances in the treatment of nonlinear waves was the discovery[52] of a totally unexpected connection between the inverse scattering problem and a number of nonlinear evolution equations of physical interest, allowing them to be exactly solved[53]. This new method of solution represents, in a sense, a nonlinear generalization of the Fourier transform.[54]

The physical basis of the new method appears to be best understood in its application to the phenomenon of self-induced transparency[55,19] in nonlinear optics, so that we shall illustrate it in this connection.

The Mc Call-Hahn equations of self-induced transparency, for an inhomogeneously broadened medium of two-level atoms, are a set of coupled Maxwell-Bloch equations. By suitable choice of variables, they may be reduced to the form[56]

$$\partial\Lambda/\partial\tau - i\gamma\Lambda = - eW , \quad (6.1a)$$

$$\partial W/\partial\tau = \frac{1}{2} (e^*\Lambda + e\,\Lambda^*) , \quad (6.1b)$$

$$\partial e/\partial\zeta = - \langle \Lambda \rangle , \quad (6.2)$$

where $e(\tau,\zeta)$ is a slowly-varying complex electric field amplitude propagating in the z direction, ζ is proportional to z, and $\tau = t - z/U$, with U representing the phase velocity in the host medium; $\Lambda(\gamma,\tau,\zeta)$ is proportional to the slowly-varying complex polarization amplitude at the detuning $\gamma = \omega - \omega_0$ (ω_0 = resonance frequency), and $W(\gamma,\tau,\zeta)$ is proportional to the excitation energy density in the active medium; finally, $\langle \Lambda \rangle$ denotes the average of Λ over the inhomogeneously broadened line shape.

The Bloch equations correspond to (6.1), with the Bloch vector normalized by $|\Lambda|^2 + W^2 = 1$. Maxwell's equations reduce to (6.2), describing $\langle \Lambda \rangle$ as a source for e. Asymptotically, for $\tau \to -\infty$, the field and the polarization vanish. The problem is to determine $e(\tau,\zeta)$, given $e(\tau, 0)$.

The solution by the inverse scattering method in the presence of inhomogeneous broadening was given by Ablowitz, Kaup and Newell[57] and, under less restrictive assumptions, by de Castro.[58] A physical interpretation of the inverse scattering method in this problem was proposed by Haus.[59] However, a different interpretation,[58] in which the roles of space and time variables are interchanged, turns out to be more appropriate, and it will be adopted here.

The linear "scattering" problem associated with (6.1) is simply the time evolution of a two-level atom driven by the electric field, as described by the Schrödinger equation. "Scattering" in time means asymptotic excitation, and the "scattering data" correspond to the asymptotic excitation state in which the atom is left by the passage of the electric field pulse (there is no spontaneous emission in this semiclassical description).

Given $e(\tau,0)$, the determination of the associated set of "scattering data" by solving the direct scattering problem is the analogue of the Fourier transform. The evolution in ζ of this set is obtained so as to make it compatible with the McCall-Hahn equations. Finally, the analogue of the inverse Fourier transform is the solution of the inverse scattering problem from the set of "scattering data" at ζ, yielding $e(\tau,\zeta)$.

The inverse problem is solved by a generalization of the Marchenko method due to Zakharov and Shabat.[60] The analytic properties of scattering solutions play a basic role in the solution; in particular, the reduction to an integral equation of the Marchenko type depends crucially on causality[58]: the state of a two-level atom at a given time can depend only on the values taken by the electric field at earlier times. Self-transparent pulses correspond to solitons.[53]

In the unrealistic (but semiclassically allowed) case where the medium is initially fully inverted, there exist "fast" self-transparent solutions which may travel faster than c . However, this does not constitute a violation of causality[61]: the solution at a given spacetime point still depends only on data contained within the backward light cone with vertex at that point. Though there may exist a preserved pulse shape travelling faster than c , the energy does not travel within the pulse: rather, the excitation energy stored in the active medium ahead of the pulse is triggered by its weak leading edge, through stimulated emission. This is an interesting example of how causality still applies to the case of an active system.

The derivation of dispersion relations from causality can also be extended to nonlinear response.[62] E.g., for quadratic nonlinearity, (1.1) is replaced by

$$r(t) = \int_{-\infty}^{\infty} dt' \int_{-\infty}^{\infty} dt'' \; g(t-t', t-t'')\, e(t') e(t''), \qquad (6.3)$$

where, from causality (cf.(1.4)),

$$g(\tau_1, \tau_2) = 0 \quad \text{for} \quad \tau_1 < 0 \quad \text{or} \quad \tau_2 < 0 \,. \qquad (6.4)$$

It follows that the double Fourier transform $G(\omega_1, \omega_2)$ of $g(\tau_1, \tau_2)$ satisfies dispersion relations in each frequency variable when the other one is kept fixed, or, what is more appropriate, dispersion relations in terms of the sum and difference frequencies $\omega_1 \pm \omega_2$. These relations connect susceptibilities for sum and difference frequency generation, showing that they represent different aspects of the same analytic function. However, in contrast with the linear case, the real and imaginary parts of the nonlinear susceptibilities do not correspond to different physical processes like dispersion and absorption, so that the Kramers-Kronig relations are apparently less useful here.

7. NATURAL MODES

So far, we have considered the role of analytic properties only within a regularity domain. However, one may often extend the domain of analyticity by analytic continuation, including simple types of singularities such as poles. This procedure has long been employed in electric circuit theory, where poles in the frequency domain are associated with natural modes of oscillation, that play an important role in transient response.

In scattering by a bounded scatterer, one can consider the analytic continuation of tha scattering amplitude in the frequency plane. For a spherically symmetric scatterer of finite radius, each partial-wave amplitude is meromorphic in ω or, equivalently, in the wave number k, so that its only singularities are poles. Probably the earliest example was Thomson's treatment[63] of the natural modes of oscillation of the eletromagnetic field outside a perfectly conducting sphere. The modes were defined as purely outgoing solutions of the boundary-value problem. This is equivalent to associating them with complex poles of the scattering matrix. In quantum mechanics, "complex-energy eigenfunctions" were introduced by Gamow,[64] in his treatment of alpha decay.

The imaginary part of a complex pole in the frequency plane yields exponential decay in time. By the same token, it leads to exponential growth in space, since distant waves were emitted a long time ago. It follows that "complex-frequency eigenfunctions" cannot be legitimate solutions for all time, i.e., a satisfactory formulation of the decay process also involves considering the excitation process. This can be done[2] by looking for the general solution of the initial-value problem in terms of natural modes.

For the initial-value problem in the exterior of an arbitrary spherically symmetric scatterer of finite radius, the natural mode expansion is obtained[2] through a partial-fraction decomposition of the scattering matrix in terms of its poles. In view of the infinite number of poles, this is a Mittag-Leffler or Cauchy expansion. In the cross section, the contribution from a complex pole close to the real axis and well separated from other poles has a typical resonance shape, with a width given by the imaginary part of the pole. In this sense, natural modes correspond to resonances.

The contribution from such a pole to Green's function is a "transient-mode" term, corresponding to a "complex-frequency eigenfunction", but with a temporal cutoff due to the excitation at the initial time. This removes the exponential growth problem at sufficiently large distances. Besides the pole contributions, the Green's function also contains terms associated with direct reflection from the surface of the scatterer, arising from the entire function contribution in the Cauchy expansion.[2]

Natural mode expansions have also been investigated for the interior problem, both in quantum mechanics[65] and in optics[66], where they have been related with the Ewald-Oseen extinction theorem.[67]

For a spherically symmetric scatterer, the elements of the scattering matrix $S_\ell(k)$ are functions of k and of the multipole or partial-wave order ℓ, which is associated with angular momentum. For a scatterer of finite range, one can introduce an interpolating function $S(\lambda,k)$ of a continuous variable λ, which reduces to $S_\ell(k)$ at the discrete integer values ℓ, and one can also extend the definition of S to complex values of λ ("complex angular momentum"). This was done at the beginning of the century by Poincaré and Watson[68], in connection with the propagation of radio waves around the Earth, giving rise to the well-known Watson transform method.

Poles of $S(\lambda,k)$ in the λ plane have become known as Regge poles, in view of their application by Regge to quantum scattering.[2] Poles in the λ plane for fixed k and poles in the k plane for fixed λ (in particular, for the physical values ℓ) are different aspects of the same singularity surface in both variables.

The counterpart[2] of the conjugate variable pair frequency and time is the pair angular momentum and angle. Thus, the imaginary part of a Regge pole leads to angular damping, and one can associate it with a "life-angle". Narrow resonances correspond to large "life-angles", i.e., to a "capture" situation involving many turns around the scattering center. For a scatterer with finite radius, Regge poles can also describe surface waves travelling around the surface of the scatterer, with an angular damping due to radiation. These are the well-known "creeping modes" found in the Watson transformation.

8. TIME DELAY AND SPATIAL DISPLACEMENTS

Another implication of causality was first pointed out by Wigner[70] in quantum scattering. It is based on the concept of time delay of a spherical multipole wave packet due to the scattering process.[71]

Consider an incoming spherical wave packet, formed by the superposition of a narrow range of frequencies. The effect of the scattering on the corresponding outgoing wave packet is to shift the phase of each frequency component, through multiplication by the S-matrix element

$$S(\omega) = \exp\left[2i\eta(\omega)\right] , \qquad (8.1)$$

where η is the scattering phase shift.

Under suitable conditions, the "center" of the outgoing wave packet can be determined by applying the principle of stationary phase. This involves differentiating the total phase, including, among others, the contribution from the time factor $\exp(-i\omega t)$, with respect to ω. It then follows from (8.1) that the effect of the scattering is to shift the center of the outgoing wave packet, relative to the situation when the scatterer is not present, by introducing a <u>time delay</u>

$$\Delta t = 2\, d\eta/d\omega \;, \qquad (8.2)$$

which corresponds to a spatial displacement

$$\Delta r = 2\, d\eta/dk \;. \qquad (8.3)$$

For classical scattering by a spherically symmetric scatterer of radius a, it follows[2] from the analytic properties of S that

$$d\eta/dk \geq -a \;. \qquad (8.4)$$

This result has an immediate interpretation in terms of causality. According to (8.3), it implies that the maximum possible time advance of the outgoing wave packet corresponds to a spatial displacement of $-2a$, i.e., to its appearance as soon as (but no sooner than) the incoming wave packet reaches the surface of the scatterer.

On the other hand, causality allows an arbitrarily large positive time delay. Thus, if S, within the frequency spectrum of the wave packet, is dominated by an isolated narrow resonance, it is readily found from (8.2) that the associated time delay is just the lifetime of the resonance.

The extension of (8.4) to quantum scattering, known as Wigner's causal inequality,[70] contains an extra term related with the uncertainty principle.

The concept of time delay can be extended to the scattering of plane wave packets.[72] In terms of the scattering amplitude $A(\vec{k},\vec{k}')$, the time delay of the scattered wave packet in the direction of $\vec{k}'$ is given by

$$\Delta t = \frac{\partial}{\partial\omega} \arg A(\vec{k},\vec{k}') \qquad (\vec{k}' \neq \vec{k}). \qquad (8.5)$$

This result is not valid in the forward direction, where there is an extra term[73] due to interference with the incident wave. For light propagation in a medium, the forward time delay is related with the real refractive index.

The corresponding spatial displacement of a bounded light beam in total reflection is the well-known Goos-Hänchen effect.[74] Light penetrates into the rarer medium at a beam boundary and travels along the interface as a surface wave before exiting again at the boundary of the reflected beam.

For the conjugate variable pair angular momentum and angle, there is a corresponding concept of angular deflection due to scattering.[75] This is given by

$$\Theta = 2\, d\eta/d\lambda \;, \qquad (8.6)$$

where λ is the continuous (interpolating) angular momentum variable and Θ is the deflection angle for a wave packet formed by superposing a range of angular momenta (impact parameters), i.e., a cylindrical pencil of rays. The total reflection of such a beam on a homogeneous (optically rarer) sphere gives rise to an angular beam displacement[75] that can be derived from (8.6) and represents the spherical analogue of the Goos-Hänchen effect.

9. ANALYTICITY IN DIFFRACTION THEORY

Sommerfeld's celebrated exact solution[76] of the half-plane diffraction problem was an early application of analytic function theory to diffraction. Several other problems, besides this one, were exactly solved by extensions of the Wiener-Hopf technique.[39]

Other exact solutions of diffraction problems have been known for a long time in the form of eigenfunction expansions (partial wave series). A classic example is the Mie solution[77,68] for the scattering of a monochromatic plane wave by a homogeneous sphere. The trouble with such expansions is that they converge extremely slowly when the wavelength is much smaller than the dimensions of the diffracting object, as usually happens for visible light. The remedy is again found in analytic continuation: the complex angular momentum method was invented for this purpose.

The original Watson transformation had several shortcomings; it could only be applied in a few disjoint spatial regions. During the past few years, however, a modified version of the Watson transformation has been developed, allowing one to derive the asymptotic short-wavelength behavior of the partial wave amplitude in any region of space. We give only a very brief outline of the method, referring to previous reviews[78,79] for a more detailed survey.

The passage to the λ plane proceeds by first applying to the partial wave series the Poisson sum formula,

$$\sum_{\ell=0}^{\infty} \phi\left(\ell + \frac{1}{2}, \vec{r}\right) = \sum_{m=-\infty}^{\infty} (-)^m \int_0^{\infty} \phi(\lambda, \vec{r}) \exp(2im\pi\lambda) d\lambda \ . \qquad (9.1)$$

Analiticity in λ is then employed to deform the path of integration in the λ plane. The freedom thus gained is the main advantage of the method.

Slow convergence of the left-hand side of (9.1) at short wavelengths means that significant contributions to the result are spread over many terms of the partial-wave series. The object of the path deformation on the right-hand side is to concentrate the dominant contributions in a relatively small number of "critical points" in the λ plane. This requires different path deformations in different spatial regions. When Regge poles are swept through, we also get their residue contributions besides the deformed path integrals.

Thus, one finds two types of critical points: saddle points (real and/or complex), which dominate the contributions to the deformed path integrals, and poles.

Real saddle points correspond to the usual geometric-optic rays; their contributions lead to the well-known WKB series. Complex saddle points are associated with complex rays, a kind of analytic continuation of real ones. Such rays occur in total reflection, where they describe the evanescent waves penetrating into the rarer medium; the quantum analogue is the barrier tunnelling effect. They also occur on the shadow side of caustics.[80]

Regge pole contributions (cf. Sect. 7) represent both the effects of resonances and of surface waves, i.e., "creeping modes", which can also be described in terms of diffracted rays.[80]

Near the boundary between spatial regions where different path deformations are required, several interesting diffraction effects occur. One of them, involving transitions between saddle points and poles, is a penumbra region, described by a generalization of Fock's theory of diffraction.[81] A collision between real saddle points, followed by their transformation into complex ones, corresponds to a rainbow.[79] Both complex saddle points and complex poles give the dominant contributions to the glory,[79] an effect that was first explained by complex angular momentum theory.

Diffraction, by definition, is the penetration of light into regions that are inaccessible to the real rays of geometrical optics. In terms of the Hamiltonian analogy between geometrical optics and classical mechanics, these are "classically forbidden" regions. Thus, in wave-mechanical terms, diffraction effects may be regarded as a kind of tunnelling through potential barriers. These barriers represent "inertial forces" related with the geometry of the diffracting objects;[82] e.g., in the angular momentum representation, we get the centrifugal barrier.

The strongest effects are found near the top of such a potential barrier, where it is most easily penetrable. In the Mie problem, this is the "edge domain", corresponding to rays near tangential incidence. The dominant contributions to the glory arise from this domain, where sur face waves are launched. It also yields significant contributions to the Mie efficiency factors;[83] such contributions represent an extension of the geometrical-optic ones to complex angles of incidence and refraction.

Diffraction effects may therefore be explained by a kind of analytic continuation of geometrical optics to complex values of ray parameters (complex trajectories). This is a vast new domain, which we are just beginning to explore.

REFERENCES

1. H. A. Kramers, Atti Congr. Intern. Fis. Como 2, 545 (1927); Phys. Z. 30, 522 (1929); R. de L. Kronig, J. Opt. Soc. Am. 12, 547 (1926); Ned. T. Natuurk. 9, 402 (1942).
2. H. M. Nussenzveig, Causality and Dispersion Relations (Academic, N.Y., 1972).
3. V. Vladimirov, Distributions en Physique Mathématique (Mir, Moscow, 1979), and references therein.
4. R. L. Weaver and Y.-H. Pao, to appear in J. Math. Phys.
5. L. D. Landau and E. M. Lifshitz, Statistical Physics (Pergamon, London, 1958), p. 394; Electrodynamics of Continuous Media (Pergamon, London, 1960), p. 256.
6. V. M. Agranovich and V. L. Ginzburg, Spatial Dispersion in Crystal Optics and the Theory of Excitons (Interscience, London, 1966), pp. 9,49.
7. M. Altarelli, D. L. Dexter, H. M. Nussenzveig and D. Y. Smith, Phys. Rev. B6, 4502 (1972).
8. P. C. Martin, Phys. Rev. 161, 143 (1967).
9. E. U. Condon, Rev. Mod. Phys. 9, 432 (1937); J. P. Mathieu, Handb. d. Physik 28 (Springer, Berlin, 1957), p. 333.
10. D. Y. Smith, J. Opt. Soc. Am. 66, 454 (1976).
11. M. T. C. S. Thomaz, Master's Thesis (Univ. of São Paulo, 1979); M. T. C. S. Thomaz and H. M. Nussenzveig, to be published; W. P. Healy, J. Phys. B9, 2499 (1976).
12. M. Altarelli, D. L. Dexter, H. M. Nussenzveig and D. Y. Smith, Phys. Rev. B6, 4502 (1972).
13. M. Altarelli and D. Y. Smith, Phys. Rev. B9, 1290 (1974).
14. D. Y. Smith, Phys. Rev. B13, 5303 (1976).
15. H. J. Hagemann, W. Gudat and C. Kunz, DESY SR-74/7 (Hamburg, 1974), J. Opt. Soc. Am. 65, 742 (1975).
16. B. W. Veal and A. P. Paulikas, Phys. Rev. B10, 1280 (1974).
17. T. Inagaki, E. T. Arakawa, R. N. Hamm and M. W. Williams, Phys. Rev. B15, 3243 (1977).

18. C. Kunz,, ed., Synchrotron Radiation, Techniques and Applications (Springer, Berlin, 1979).
19. Cf., e.g., H. M. Nussenzveig, Introduction to Quantum Optics (Gordon and Breach, N.Y., 1973).
20. Photon detectors that would operate by emission. leading to antinormal ordering, are conceivable in principle; cf. L. Mandel, Phys. Rev. 152, 438 (1966).
21. H. M. Nussenzveig, Proc. II Brazilian Symp. Theor. Phys. (Rio de Janeiro, 1969).
22. R. J. Glauber, Phys. Rev. 131, 2766 (1963).
23. V. Bargmann, Commun. Pure Appl. Math. 14, 187 (1961).
24. J. S. Toll, Phys. Rev. 104, 1760 (1956).
25. T. Inagaki, H. Kuwata and A. Ueda, Phys. Rev. B19, 2400 (1979).
26. E. Wolf, Proc. Phys. Soc. (London) 80, 1269 (1962); Y. Kano and E. Wolf, ibid. 80, 1273 (1962).
27. H. M. Nussenzveig, J. Math. Phys. 8, 561 (1967).
28. R. E. Burge, M. A. Fiddy, A. H. Greenaway and G.Ross, Proc. Roy. Soc. (London) A350, 191 (1976).
29. E. C. Titchmarsh, The Theory of Functions, 2nd ed. (Oxford, 1939).
30. E. N. Leith and J. Upatnieks, J. Opt. Soc. Am. 54, 1295 (1964).
31. O. Bryngdahl and A. Lohmann, J. Opt. Soc. Am. 58, 620 (1968).
32. H. Gamo, in Electromagnetic Theory and Antennas, ed. E. C.Jordan (Pergamon, Oxford 1963), p. 801.
33. C. L. Mehta, J. Opt. Soc. Am. 58, 1233 (1968).
34. T. D. Beard, Appl. Phys. Lett. 15, 227 (1969).
35. D. Kohler and L. Mandel, J. Opt. Soc. Am. 63, 126 (1973).
36. K. Chadan and P. S. Sabatier, Inverse Problems in Quantum Scattering Theory (Springer, N.Y., 1977).
37. P. C. Sabatier, ed., Applied Inverse Problems (Springer, N.Y., 1978).
38. H. P. Baltes, ed., Inverse Source Problems in Optics (Springer, N.Y., 1978).
39. C. J. Bouwkamp, in Rep. Progr. Phys. 17, 41 (1954).
40. J. R. Shewell and E. Wolf, J. Opt. Soc. Am. 58, 1596 (1968).
41. A. J. Devaney, J. Math. Phys. 19, 1526 (1978).
42. I. M. Gelfand and B. M. Levitan, Izv. Akad. Nauk SSSR, Ser. Mat. 15, 309 (1951) [Am. Math. Soc. Transl. (2) 1, 253 (1955).]
43. V. A. Marchenko, Doklad. Akad. Nauk SSSR, 72, 457 (1950); 104, 695 (1955).
44. R. Jost and W. Kohn, Kgl. Danske Vid. Selsk. Mat.-fys. Medd. 27, No. 9 (1953).
45. L. D. Faddeev, J. Math. Phys. 4, 72 (1963).
46. G. N. Balanis, J. Math. Phys. 13, 1001 (1972).

47. L. D. Faddeev, Trudy Mat. Inst. Steklov 73, 314 (1964) [Am. Math. Soc. Transl. 2, 139 (1965)].
48. P. Deift and E. Trubowitz, Commun. Pure Appl. Math. 32, 121 (1979).
49. R. G. Newton, J. Math. Phys. 21, 493 (1980).
50. L. D. Faddeev, Preprint Kiev ITP 71.106 E; J. Sov. Math. 5, 334 (1976).
51. R. G. Newton, in Scattering Theory in Mathematical Physics, eds. J. A. Lavita and J. P. Marchand (D. Reidel, Dordrecht, 1974), p. 193; Phys. Rev. Lett. 43, 541 (1979).
52. C. S. Gardner, J. M. Greene, M. D. Kruskal and R. M. Miura, Phys. Rev. Lett. 19, 1095 (1967).
53. A. C. Scott, F. Y. F. Chu and D. W. McLaughlin, Proc. IEEE 61, 1443 (1973).
54. M. J. Ablowitz, D. J. Kaup, A. C. Newell and H. Segur, Studies Appl. Math. 53, 249 (1974).
55. S. L. Mc Call and E. L. Hahn, Phys. Rev. 183, 457 (1969).
56. G. L. Lamb, Jr., Phys. Rev. Lett. 31, 196 (1973); Phys. Rev. A9, 422 (1974).
57. M. J. Ablowitz, D. J. Kaup and A. C. Newell, J. Math. Phys. 15, 1852 (1974); D. J. Kaup, Phys. Rev. A16, 704 (1977).
58. H. M. A. de Castro, Master's Thesis (Univ. of São Paulo, 1979); H. M. A. de Castro and H. M. Nussenzveig, to be published.
59. H. A. Haus, Rev. Mod. Phys. 51, 331 (1979).
60. V. E. Zakharov and A. L. Shabat, Sov. Phys. JETP 34, 62 (1972).
61. R. C. T. da Costa, J. Math. Phys. 11, 2799 (1970).
62. F. L. Ridener, Jr. and R. H. Good, Jr., Phys. Rev. B10, 4980 (1974), and references therein.
63. J. J. Thomson, Proc. London Math. Soc. 15 (1), 197 (1884).
64. G. Gamow, Z. Phys. 51, 204; 52, 510 (1928).
65. W. J. Romo, J. Math. Phys. 21, 311 (1980), and references therein.
66. B. J. Hoenders, in Coherence and Quantum Optics, eds. L. Mandel and E. Wolf (Plenum, N.Y., 1978), p. 221; J. Phys. A11, 1815 (1978).
67. E. Wolf and D. N. Pattanayak, in Symposia Mathematica, vol. 18 (Academic, N.Y., 1976); Opt. Commun. 6, 217 (1972).
68. Cf. N. A. Logan, Proc. IEEE 53, 773 (1965).
69. W. Franz, Theorie der Beugung Elektromagnetischer Wellen (Springer, Berlin, 1957).
70. E. P. Wigner, Phys. Rev. 98, 145 (1955).
71. L. Eisenbud, Ph.D. Thesis (Princeton Univ., 1948, unpublished).

72. M. Froissart, M. L. Goldberger and K. M. Watson, Phys. Rev. 131, 2820 (1963).
73. H. M. Nussenzveig, Phys. Rev. D6, 1534 (1972).
74. F. Goos and H. Hänchen, Ann. Physik (6) 1, 333 (1947); 5, 251 (1949); K. Artmann, ibid. 2, 87 (1948).
75. N. F. Ferrari Jr. and H. M. Nussenzveig, to be published.
76. A. Sommerfeld, Math. Ann. 47, 317 (1896).
77. G. Mie, Ann. Physik 25, 377 (1908).
78. H. M. Nussenzveig, in Methods and Problems of Theoretical Physics, ed. J. E. Bowcock (North-Holland, Amsterdam, 1970), p. 203.
79. H. M. Nussenzveig, J. Opt. Soc. Am. 69, 1068 (1979) and references therein.
80. J. B. Keller, in Calculus of Variations and its Applications, ed. L. M. Graves (McGraw-Hill, N.Y., 1958), p. 27.
81. V. A. Fock, Electromagnetic Diffraction and Propagation Problems (Pergamon, Oxford, 1965).
82. G. Beck and H. M. Nussenzveig, Nuovo Cimento 16, 416 (1960).
83. H. M. Nussenzveig and W. J. Wiscombe, to be published.