IOP PUBLISHING EUROPEAN JOURNAL OF PHYSICS

Eur. J. Phys. **33** (2012) 1635–1650 doi:10.1088/0143-0807/33/6/1635

Linear response laws and causality in electrodynamics

Alex J Yuffa and John A Scales

Department of Physics, Colorado School of Mines, Golden, CO 80401, USA

E-mail: ayuffa@gmail.com and jscales@mines.edu

Received 27 February 2012, in final form 9 July 2012
Published 12 September 2012
Online at stacks.iop.org/EJP/33/1635

Abstract

Linear response laws and causality (the effect cannot precede the cause) are of fundamental importance in physics. In the context of classical electrodynamics, students often have a difficult time grasping these concepts because the physics is obscured by the intermingling of the time and frequency domains. In this paper, we analyse the linear response laws and causality in the time and frequency domains with the aim of pedagogical clarity. We will show that it is easy to violate causality in the frequency domain by making a vanishing absorption approximation. Furthermore, we will show that there can be subtle differences between Fourier transforming Maxwell equations and using a monochromatic source function. We discuss how these concepts can be obscured and offer some suggestions to improve the situation.

1. Introduction

When encountering Maxwell's equations in matter for the first time, students are faced with many conceptual, as well as mathematical, difficulties. In the time domain, the four macroscopic Maxwell's equations, namely

$$\begin{aligned} \nabla \cdot \mathbf{D} &= 4\pi\rho, & \nabla \cdot \mathbf{B} &= 0, \\ \nabla \times \mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} &= 0, & \nabla \times \mathbf{H} - \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t} &= \frac{4\pi}{c}\mathbf{J}, \end{aligned} \tag{1}$$

are straightforward enough. The problem arises in the constitutive relations connecting **E** and **D** or **B** and **H**. While the time-domain Maxwell equations are real and involve only real quantities, the associated response functions are temporally non-local and their very definition involves integration with respect to time. To professionals, this issue is well understood and is almost always side-stepped notationally by writing the response equations as

$$\mathbf{D} = \epsilon\mathbf{E} \qquad \text{and} \qquad \mathbf{H} = \mu^{-1}\mathbf{B}, \tag{2}$$

and mentally juggling the time and frequency domains. Here, for simplicity, we have assumed that the medium is linear, isotropic and homogeneous (LIH). Experts understand that (2) means either

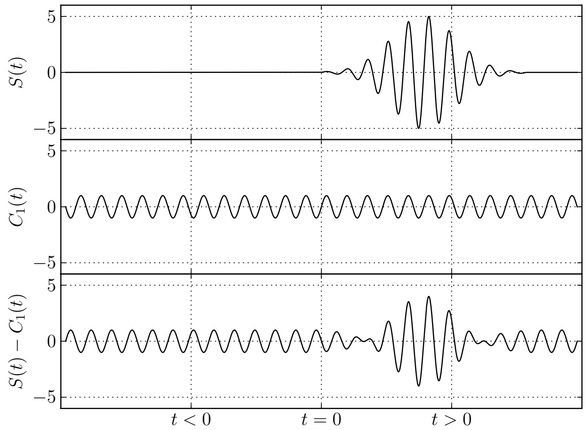

Figure 1. The input signal $S(t)$ is shown along with the output signal $S(t) - C_1(t)$ of a system that only absorbs the $C_1(t)$ component of $S(t)$, without affecting other components.

(a) Fourier-domain relations of general validity (within LIH assumptions) or

(b) time-domain relations valid only for monochromatic fields.

But simultaneously presenting (1) and (2) to students (as is done in popular textbooks [1–3]) can mislead and confuse them; it obscures the important temporal non-locality of the response functions, mixes time and frequency domain concepts and inserts complex quantities into manifestly real equations. To avoid a possible source of confusion, we use the phrase 'Fourier domain' instead of 'frequency domain' because the latter may also refer to a time-domain relation with a monochromatic source.

Moreover, students fail to understand the relationship between absorption, dispersion and causality (the effect cannot precede the cause [4]). To see the interplay between absorption, dispersion and causality pictorially, we follow Toll's ingenious presentation [5] by considering an input signal $S(t)$ that is zero for $t < 0$. The input $S(t)$ is a weighted sum of many sinusoidal components, such as $C_1(t) = \sin(\omega_1 t + \psi_1)$, each of which extends from $t = -\infty$ to $t = \infty$. Note that the weighted sum of all sinusoidal components produces a zero input signal for $t < 0$. If a system only absorbs one component, e.g., $C_1(t)$, without affecting other components, then the output of such a system would simply be $S(t) - C_1(t)$, which is non-zero for $t < 0$; see figure 1. Such a system is impossible because it violates causality; the output is non-zero before the onset of the input signal. Therefore, for causal systems, an absorption of one frequency must be accompanied by phase shifts of other frequencies in order to produce a zero output for $t < 0$, and the necessary phase shifts are prescribed by the dispersion relation. Moreover, the converse is true as well; namely, a phase shift of one frequency is necessarily accompanied by an absorption at other frequencies. From Toll's argument, we conclude that it is impossible to design a physical system that is causal and dispersionless. Therefore, when one speaks of dispersionless media, one violates a sacred physical principle (causality).

This paper is intended not only for instructors who teach advanced undergraduate and beginning graduate students, but also for the graduate students. The mathematical sophistication needed to understand this paper is essentially that of a beginning graduate student. However, throughout the paper, we extensively use the theory of the tempered distributions, with which a typical beginning graduate student may not be familiar. To remedy this deficiency, we have included an example-driven tutorial on the formal theory of tempered distributions in the appendix.

2. Background

The time-domain relationship between $\mathbf{D}$ and $\mathbf{E}$ for an LIH and time-translationally invariant medium is given by[1]

$$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P} \qquad \text{and} \qquad \mathbf{P}(t) = \int_{-\infty}^{\infty} \chi(t-t')\mathbf{E}(t')\,\mathrm{d}t', \tag{3a}$$

where

$$\chi(t-t') = \alpha(t-t')\Theta(t-t') \quad \text{and} \quad \Theta(t-t') = \begin{cases} 1, & t-t' > 0 \\ 0, & t-t' < 0 \end{cases}. \tag{3b}$$

The appearance of the Heaviside step function Θ in the definition of the electric susceptibility χ reminds us that the polarization vector $\mathbf{P}$ can only depend on the past values of the applied electric field $\mathbf{E}$. Therefore, (3) gives a causal relationship between the displacement field $\mathbf{D}$ and the applied electric field. Taking the Fourier transform of (3) yields

$$\widetilde{\mathbf{D}}(\omega) = \widetilde{\mathbf{E}}(\omega) + 4\pi\widetilde{\mathbf{P}}(\omega) \qquad \text{and} \qquad \widetilde{\mathbf{P}}(\omega) = \widetilde{\chi}(\omega)\widetilde{\mathbf{E}}(\omega), \tag{4}$$

with the Fourier transform pair given by

$$\widetilde{f}(\omega) = \mathcal{F}[f(t)] = \int_{-\infty}^{\infty} f(t)\,\mathrm{e}^{+\mathrm{i}\omega t}\,\mathrm{d}t, \tag{5a}$$

$$f(t) = \mathcal{F}^{-1}[\widetilde{f}(\omega)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widetilde{f}(\omega)\,\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}\omega. \tag{5b}$$

Under suitable mathematical conditions, the principle of causality in the time domain translates into the Kramers–Kronig (KK) relations (the Hilbert transform pair) in the Fourier domain, namely

$$\mathrm{Im}\,\widetilde{\chi}(\omega) = \mathcal{H}[\mathrm{Re}\,\widetilde{\chi}(\eta)] = -\frac{1}{\pi}\,⨍_{-\infty}^{\infty} \frac{\mathrm{Re}\,\widetilde{\chi}(\eta)}{\eta-\omega}\,\mathrm{d}\eta, \tag{6a}$$

$$\mathrm{Re}\,\widetilde{\chi}(\omega) = \mathcal{H}^{-1}[\mathrm{Im}\,\widetilde{\chi}(\eta)] = \frac{1}{\pi}\,⨍_{-\infty}^{\infty} \frac{\mathrm{Im}\,\widetilde{\chi}(\eta)}{\eta-\omega}\,\mathrm{d}\eta, \tag{6b}$$

where $\mathrm{Re}\,\widetilde{\chi}(\omega)$ and $\mathrm{Im}\,\widetilde{\chi}(\omega)$ are the real and the imaginary parts of $\widetilde{\chi}(\omega)$, respectively, and $⨍$ denotes the Cauchy principal value integral. In free space, by definition, $\chi(t-t') = 0$ and consequently, the KK relations are trivially satisfied. The KK relations can be derived in 'two lines' by using the Fourier representation of the Heaviside step function and utilizing the freedom to define $\alpha(t-t')$ for $t-t' < 0$ [6]. A lengthier but more traditional derivation of the KK relations is available in [7]. This derivation is self-contained and does not assume *a priori* knowledge of the theory of functions of a complex variable. For a historically accurate account on how the KK relations were first derived, see [8].

In general, it is difficult to establish an equivalence between the KK relations (6) and the causality (3*b*). If $\widetilde{\chi}(\omega)$ is a square-integrable function, then the Titchmarsh theorem [9] guarantees that (6) and (3*b*) are equivalent. In other words, $\chi(t-t') \propto \Theta(t-t')$ if and only if $\widetilde{\chi}(\omega)$ satisfies (6). The square-integrability requirement on $\widetilde{\chi}(\omega)$ may be somewhat

[1] For our purposes, it is more convenient to work with (3) than with $\mathbf{D} = \int_{-\infty}^{\infty} \epsilon(t-t')\mathbf{E}(t')\,\mathrm{d}t'$, where $\epsilon(t-t') = \delta(t-t') + 4\pi\chi(t-t')$.

Table 1. A brief description of the symbols introduced in the appendix.

Symbol	Description
S	The Schwartz class
S'	The class of tempered distributions
$\langle f, \phi\rangle = \int_{-\infty}^{\infty} f(t)\phi(t)\,\mathrm{d}t$	Tempered distribution f; $f \in S'$ and $\phi \in S$
$(f * \psi)(t) = \int_{-\infty}^{\infty} f(t-t')\psi(t')\,\mathrm{d}t'$	Convolution of f with ψ; $f \in S'$ and $\psi \in S$
$\mathrm{supp}\,[f]$	Support of f; $f \in S'$

relaxed. It can be shown that if $\widetilde{\chi}(\omega)$ is bounded but $\mathbf{E}(t)$ and $\mathbf{P}(t)$ are square-integrable functions, then (6) and (3*b*) are equivalent [5]. Unfortunately, these conditions are generally not satisfied, in fact, they are not even satisfied in the idealized examples considered in this paper. Therefore, we must part with our naïve notion that χ is a classical function and treat it as a generalized function (distribution). For our purposes, it will be sufficient to treat $\chi(t)$ as a tempered distribution. A reader not familiar with the formal theory of the tempered distributions should read the appendix which, for our purposes, serves as a self-contained tutorial on tempered distributions. For the reader's convenience, table 1 provides a summary of the notation introduced in the appendix. From this point on, unless explicitly noted otherwise, the symbol $\chi(t)$ should be interpreted as a tempered distribution, i.e. $\chi(t) \in S'$. Consequently, the convolution integral in (3*a*) and the Fourier transform pair given by (5) should also be interpreted in a distributional sense; see appendix A.1. Strictly speaking, the Hilbert transform pair given by (6) should also be interpreted in a distribution sense. However, for our purposes, it will be sufficient to think of the Hilbert transform as a 'generalized' convolution of the singular function $1/\omega$ (Hilbert kernel) with a tempered distribution $\widetilde{f}(\omega)$. For a more mathematical treatment of the Hilbert transform of the tempered distributions, see [10–13]. In 1958, Taylor [14] (also, see discussion in [15]) rigorously established the equivalence between the causality and the KK relations when $\widetilde{\chi}(\omega)$ is a tempered distribution. This is a very important result that we will use repeatedly. We will force $\widetilde{\chi}(\omega) \in S'$ to satisfy the KK relations, thereby guaranteeing that $\chi(t-t')$ vanishes for $t - t' < 0$. In other words, Taylor's result gives causality meaning in the Fourier domain.

To explore these fundamental issues in a pedagogical context, we will consider two textbook models that are used to derive or explain the dispersion of electromagnetic waves. The first is the Drude model of conduction in metals. The second is the damped harmonic oscillator, which arises in semi-classical models of atomic absorption. The former is just a damped harmonic oscillator model with a spring constant of zero, and where we interpret the damping in terms of electron collisions.

3. The Drude model

In a metal, the motion of a conduction electron of charge $-e$ and mass m under the influence of an electric field $\mathbf{E}$ is given by

$$\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} + \frac{1}{\tau}\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = -\frac{e}{m}\mathbf{E}, \tag{7a}$$

where τ is the collision mean free time [16]. Substituting $\delta(t-t')\mathbf{n}$ for $\mathbf{E}$ and taking the Fourier transform, see appendix A.1, yields

$$\omega\left(\omega + \frac{\mathrm{i}}{\tau}\right)\widetilde{\mathbf{g}}(\omega) = \frac{e}{m}\widetilde{\mathbf{E}}(\omega) \qquad \text{and} \qquad \widetilde{\mathbf{E}}(\omega) = \mathrm{e}^{+\mathrm{i}\omega t'}\mathbf{n}, \tag{7b}$$

where $\widetilde{\mathbf{g}}(\omega)$ denotes the Green function, $\mathbf{n} = \frac{1}{\sqrt{3}}(\mathbf{x}+\mathbf{y}+\mathbf{z})$, and $\{\mathbf{x},\mathbf{y},\mathbf{z}\}$ is the standard basis for the three-dimensional space. Using $\widetilde{\mathbf{P}}(\omega) = -n_c e\widetilde{\mathbf{g}}(\omega)$, we see that the electric susceptibility is given by

$$\omega\left(\omega+\frac{\mathrm{i}}{\tau}\right)\widetilde{\chi}(\omega) = -\frac{e^2 n_c}{m}, \tag{8a}$$

where n_c denotes the conduction electron density (assumed to be constant). When solving (8*a*) for $\widetilde{\chi}(\omega)$ we must remember that $\widetilde{\chi}(\omega)$ is a tempered distribution; in other words, we seek a *distributional* solution to (8*a*). The distributional solution consists of two parts: the *particular* part $\widetilde{\chi}_p(\omega)$ given by the 'naïve division' of (8*a*), namely

$$\widetilde{\chi}_p(\omega) = -\frac{\sigma_0}{\omega(\tau\omega+\mathrm{i})}, \tag{8b}$$

where $\sigma_0 = e^2 n_c\tau/m$ is the static conductivity, and the *homogeneous* part $\widetilde{\chi}_h(\omega)$, which satisfies

$$\left\langle\omega\left(\omega+\frac{\mathrm{i}}{\tau}\right)\widetilde{\chi}_h(\omega),\phi(\omega)\right\rangle = \langle 0,\phi(\omega)\rangle \tag{8c}$$

for all $\phi \in S$. The solution to (8*c*) is of the form of (A.28) and *not* of (A.36) because $\omega + \mathrm{i}/\tau$ never equals zero for real ω and finite τ. In other words, $\operatorname{supp}[\widetilde{\chi}_h(\omega)] = \{0\}$ and not $\{0, -\mathrm{i}/\tau\}$. Therefore, as shown in the first example in appendix A.2, the form of the full solution to (8*a*) is given by

$$\widetilde{\chi}(\omega) = -\sigma_0\left(\frac{1}{\omega(\tau\omega+\mathrm{i})}+\widetilde{\chi}_h(\omega)\right), \qquad \text{where} \qquad \widetilde{\chi}_h(\omega) = c_0\delta(\omega). \tag{8d}$$

To find the unknown constant c_0, we require that the response of the system be causal. According to Taylor [14], the causality requirement is the same thing as requiring that $\mathrm{Im}\widetilde{\chi} = \mathcal{H}[\mathrm{Re}\widetilde{\chi}]$, which yields $c_0 = -\pi$. Therefore, the *causal* electric susceptibility is given by

$$\widetilde{\chi}(\omega) = -\sigma_0\left(\frac{1}{\omega(\tau\omega+\mathrm{i})} - \pi\delta(\omega)\right), \tag{9a}$$

$$\chi(t-t') = \mathcal{F}^{-1}[\widetilde{\chi}(\omega)] = \sigma_0\left(1-\mathrm{e}^{-\frac{t-t'}{\tau}}\right)\Theta(t-t'). \tag{9b}$$

Note that $\chi(t-t')$ is zero for $t < t'$ and increases monotonically to its maximum value of σ_0 for $t > t'$. The rate of the increase is controlled by the collision mean free time. In the next two subsections, we will study the effects of $\chi(t-t')$ on the response functions when a constant or monochromatic electric field is applied.

3.1. A constant electric field

Suppose now that a constant electric field $\mathbf{E}_0$ was turned on at $t' = 0$. Substituting $\mathbf{E}(t') = \mathbf{E}_0\Theta(t')$ and (9*b*) into (3*a*) yields

$$\mathbf{P}(t) = \sigma_0\left(t-\tau+\tau\,\mathrm{e}^{-\frac{t}{\tau}}\right)\mathbf{E}_0\Theta(t), \tag{10a}$$

and the current density is given by

$$\mathbf{J}(t) = \frac{\partial\mathbf{P}}{\partial t} = \sigma_0\left(1-\mathrm{e}^{-\frac{t}{\tau}}\right)\mathbf{E}_0\Theta(t). \tag{10b}$$

The time dependence in (10*b*) reminds us that when we connect a wire to a battery, the current in the wire does not *instantaneously* reach its Ohm's law steady-state value. The rate at which the current reaches the steady-state value is controlled by the collision mean free time τ, as evident from the exponential term in (10*b*). Of course, (10*b*) reduces to Ohm's law, $\mathbf{J} = \sigma_0\mathbf{E}_0$, when $t \gg \tau$.

3.2. A monochromatic field

Another simple situation is a monochromatic driving field of angular frequency ω_d that has existed since the beginning of time ($t' = -\infty$). Substituting $\mathbf{E}(t') = \mathbf{E}_0 \cos(\omega_\mathrm{d} t')$ and (9*b*) into (3*a*) yields

$$\mathbf{D}(t) = A_1 \mathbf{E}_0 \cos(\omega_\mathrm{d} t) + A_2 \mathbf{E}_0 \sin(\omega_\mathrm{d} t), \tag{11a}$$

where

$$A_1 = 1 - \frac{4\pi\sigma_0\tau}{1+\tau^2\omega_\mathrm{d}^2} \qquad \text{and} \qquad A_2 = \frac{4\pi\sigma_0}{\omega_\mathrm{d}\left(1+\tau^2\omega_\mathrm{d}^2\right)}. \tag{11b}$$

From (11*a*), we see that $\mathbf{D}(t)$ has one component that oscillates in phase and one component that oscillates out of phase with the applied field. The out-of-phase oscillations are caused by collisions of the electrons with the ions (absorption). We can put (11) into a form of (2) if we let

$$\check{\epsilon} = A_1 + \mathrm{i}A_2 \qquad \text{and} \qquad \check{\mathbf{E}} = \mathbf{E}_0\,\mathrm{e}^{-\mathrm{i}\omega_\mathrm{d} t}; \tag{12a}$$

and then, (11) becomes

$$\mathbf{D}(t) = \mathrm{Re}(\check{\epsilon}\,\check{\mathbf{E}}). \tag{12b}$$

Note that, in general, $\mathbf{D}(t) \neq \mathrm{Re}(\check{\epsilon})\mathrm{Re}(\check{\mathbf{E}})$. Thus, what professionals mean by (2) is really (12) if they are talking about a monochromatic applied field. But what is $\check{\epsilon}$? To answer this question, let us rewrite (12*a*) in an illuminating form, namely

$$\frac{\check{\epsilon}-1}{4\pi} = \check{\chi}(\omega_\mathrm{d}) = -\frac{\sigma_0}{\omega_\mathrm{d}(\tau\omega_\mathrm{d}+\mathrm{i})}. \tag{13}$$

Note that (13) has a simple pole at $\omega_\mathrm{d} = 0$. The root cause of this pole is our assumption that the medium is homogeneous, and the definition of homogeneity clearly depends on the spatial wavelength of the interrogating field [17, 18]. Thus, (13) may be devoid of physical reality near $\omega_\mathrm{d} = 0$. Be that as it may, the resemblance between $\check{\chi}(\omega_\mathrm{d})$ and $\widetilde{\chi}(\omega)$ is uncanny if we replace ω_d with ω in (13). But is such a comparison of $\check{\chi}$ and $\widetilde{\chi}$ meaningful? Strictly speaking, it is *not*, because $\check{\chi}$ is a classical (ordinary) function and $\widetilde{\chi}$ is a tempered distribution (generalized function). Nevertheless, we want to compare these two objects by 'promoting' $\check{\chi}$ to be a generalized function. Let $X(\omega) = \widetilde{\chi}(\omega) - \check{\chi}(\omega)$, and then

$$\langle X(\omega), \phi(\omega)\rangle = \int_{-\infty}^{\infty} \pi\sigma_0\delta(\omega)\phi(\omega)\,\mathrm{d}\omega = \pi\sigma_0\phi(0) \tag{14}$$

for all $\phi \in S$. From (14), we see that $\mathrm{supp}\,[X(\omega)] = \{0\}$, i.e. the 'promoted' $\check{\chi}(\omega)$ and $\widetilde{\chi}(\omega)$ differ in a *neighbourhood* of $\omega = 0$. In the laboratory, we actually make a Fourier-domain 'measurement' in the *time domain* by driving the system for a very long time with a monochromatic field [19]. Thus, we experimentally measure $\check{\chi}(\omega_\mathrm{d})$ and not $\widetilde{\chi}(\omega)$. Moreover, it is $\widetilde{\chi}(\omega)$ that satisfies the KK relations and not $\check{\chi}(\omega_\mathrm{d})$, even if we 'promote' $\check{\chi}(\omega_\mathrm{d})$ to a generalized function. This crucial difference between $\widetilde{\chi}(\omega)$ and $\check{\chi}(\omega_\mathrm{d})$ is often missed by students.

A mathematically inclined reader might object to the usage of $\mathbf{E}(t') = \mathbf{E}_0\cos(\omega_\mathrm{d} t')$ for a driving field. Clearly, $\cos(\omega_\mathrm{d} t') \notin S$ but we have only defined the convolution (A.18) between a distribution in S' and a function in S. Of course, physically we know that a starving graduate student had to turn the field on and off. If we assume that the field was turned on/off in a smooth and slow fashion, then it could be approximated by $\cos(\omega_\mathrm{d} t')\exp[-(at')^2]$, which does belong to the space S. A similar physical argument can be used to justify the usage of the Heaviside step function in section 3.1. Moreover, under certain conditions, it is possible to define convolution of two tempered distributions [20]. This definition is beyond the scope of this paper but, if used, it would eliminate the need for the above physical argument.

4. Plasma

A limiting case of the Drude model is dilute neutral plasma with very large collision mean free time. If we try to expand (9*a*) around $1/\tau = 0$, we will obtain a non-causal $\widetilde{\chi}(\omega)$ because the homogeneous solution $\widetilde{\chi}_h(\omega)$ changes form in this limiting case. To see this, expand (8*c*) around $1/\tau = 0$ to obtain

$$\langle \omega^2 \widetilde{\chi}_h(\omega), \phi(\omega) \rangle = \langle 0, \phi(\omega) \rangle, \tag{15a}$$

for all $\phi \in S$. The form of the solution to (15*a*) is given by, see (A.31),

$$\widetilde{\chi}_h(\omega) = c_0 \delta(\omega) + c_1 \delta'(\omega), \tag{15b}$$

which differs from the homogeneous part of (9*a*) by a $\delta'(\omega)$ term. Therefore, the full solution to $\omega^2 \widetilde{\chi}(\omega) = -e^2 n_c / m$ is given by

$$\widetilde{\chi}(\omega) = -\frac{\omega_p^2}{4\pi} \left(\frac{1}{\omega^2} + c_0 \delta(\omega) + c_1 \delta'(\omega) \right), \tag{15c}$$

where $\omega_p^2 = 4\pi e^2 n_c / m$ and ω_p is called the angular plasma frequency. To find the unknown constants c_0 and c_1, we use Taylor's causality requirement $\mathrm{Im}\widetilde{\chi} = \mathcal{H}[\mathrm{Re}\widetilde{\chi}]$, which yields $c_0 = 0$ and $c_1 = \mathrm{i}\pi$. Therefore, the *causal* electric susceptibility is given by

$$\widetilde{\chi}(\omega) = -\frac{\omega_p^2}{4\pi} \left(\frac{1}{\omega^2} + \mathrm{i}\pi \frac{\mathrm{d}}{\mathrm{d}\omega} \delta(\omega) \right), \tag{16a}$$

$$\chi(t - t') = \mathcal{F}^{-1}[\widetilde{\chi}(\omega)] = \frac{e^2 n_c}{m} (t - t') \Theta(t - t'). \tag{16b}$$

We compare (16*a*) to a common textbook expression for the electric susceptibility of dilute neutral plasma, given by [16]

$$\widetilde{\chi}(\omega) = -\frac{\omega_p^2}{4\pi} \frac{1}{\omega^2}. \tag{17}$$

This differs from (16*a*) in a neighbourhood of $\omega = 0$. The difference is caused by the fact that (16*a*) is causal and valid for all ω, unlike the textbook version (17), which is only valid for high enough frequencies, namely $\omega \gg 1/\tau$.

The above example illustrates that the approximation in the Fourier domain must be made with care if we insist on a causal electric susceptibility. It is interesting to note that no such care is necessary if the approximation is made in the time domain. For example, expanding (9*b*) around $(t - t')/\tau = 0$ immediately yields (16*b*), which is clearly causal.

4.1. Plasma in a constant electric field

Next, let us consider a plasma in a constant electric field $\mathbf{E}_0$ that was turned on at $t' = 0$. Substituting $\mathbf{E}(t') = \mathbf{E}_0 \Theta(t')$ and (16*b*) into (3*a*) yields

$$\mathbf{P}(t) = \frac{e^2 n_c}{2m} t^2 \mathbf{E}_0 \Theta(t). \tag{18}$$

From (16*b*) and (18), we see that the electric susceptibility is linear in time and that the polarization vector is quadratic in time. The quadratic dependence of $\mathbf{P}(t)$ on time signifies that the dilute neutral plasma is accelerating uniformly under the influence of the applied static electric field. Moreover, from (3*a*), we see that $\mathbf{D}$ grows quadratically in time and that at $t = 0$ the response field $\mathbf{D}$ is equal to the applied field $\mathbf{E}_0$.

4.2. Plasma in a monochromatic electric field

Finally, if we drive the plasma with a monochromatic field of angular frequency ω_d that existed since $t' = -\infty$, we will see another source of confusion for students. Substituting $\mathbf{E}(t') = \mathbf{E}_0 \cos(\omega_d t')$ and (16*b*) into (3*a*) yields

$$\mathbf{D}(t) = \check{\epsilon}\, \mathbf{E}(t) \qquad \text{and} \qquad \check{\epsilon} - 1 = 4\pi \check{\chi}\,(\omega_d) = -\frac{\omega_p^2}{\omega_d^2}. \tag{19}$$

Again, if we replace ω_d with ω in (19) and 'promote' $\check{\chi}(\omega)$ to be a tempered distribution, then the 'promoted' $\check{\chi}(\omega)$ differs from $\widetilde{\chi}(\omega)$ only in a neighbourhood of $\omega = 0$. Also, note that all quantities in (19) are purely real. Of course, $\mathbf{D}(t)$ oscillates in phase with $\mathbf{E}(t)$ because we have effectively ignored collisions (absorption) in our approximation. If we let $\check{\mathbf{E}} = \mathbf{E}_0 \exp(-\mathrm{i}\omega_d t)$, then (19) may be written as $\mathbf{D}(t) = \mathrm{Re}(\check{\epsilon})\mathrm{Re}(\check{\mathbf{E}})$, which to a student may confirm an improper interpretation of (2).

5. Damped harmonic oscillator

A damped harmonic oscillator is a simple model for the motion of a bound electron in a dielectric. The equation of motion in the time domain is given by

$$\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} + \gamma \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} + \omega_0^2 \mathbf{r} = -\frac{e}{m}\mathbf{E}, \tag{20a}$$

where ω_0 is the natural angular frequency and γ is the radiation damping. The Green function for (20*a*) (obtained in the same manner and notation as in section 3) is

$$(\omega - \omega_+)\,(\omega - \omega_-)\,\widetilde{\mathbf{g}}(\omega) = \frac{e}{m}\widetilde{\mathbf{E}}(\omega), \tag{20b}$$

where

$$\omega_\pm = \pm s - \mathrm{i}\frac{\gamma}{2} \qquad \text{and} \qquad s = \sqrt{\omega_0^2 - \left(\frac{\gamma}{2}\right)^2}. \tag{20c}$$

If we assume that each electron in the dielectric oscillates with the same natural angular frequency, then using $\widetilde{\mathbf{P}}(\omega) = -en_b\widetilde{\mathbf{g}}(\omega)$ yields

$$(\omega - \omega_+)(\omega - \omega_-)\widetilde{\chi}(\omega) = -\frac{e^2 n_b}{m}, \tag{21a}$$

where n_b denotes the density of the bounded electrons (assumed to be constant). From (21*a*), we see that $\widetilde{\chi}(\omega)$ has two simple poles in the complex ω-plane. As we will soon see, the location of these poles in the complex ω-plane will dictate the form of the homogeneous solution $\widetilde{\chi}_h(\omega)$. First, consider the simplest under-damped case, namely when $s > 0$ and $\gamma \neq 0$. In this case, $\omega - \omega_+$ and $\omega - \omega_-$ never equal zero for real ω. Therefore, the homogeneous solution is simply zero, $\widetilde{\chi}_h(\omega) = 0$, and the *full* solution is given by

$$\widetilde{\chi}(\omega) = -\frac{e^2 n_b}{m} \frac{1}{(\omega - \omega_+)(\omega - \omega_-)}, \tag{21b}$$

$$\chi(t - t') = \mathcal{F}^{-1}[\widetilde{\chi}(\omega)] = \frac{e^2 n_b}{m} \frac{\mathrm{e}^{-\frac{\gamma}{2}(t-t')} \sin[s(t - t')]}{s} \Theta(t - t'). \tag{21c}$$

Note that we did not have to impose Taylor's causality requirement, $\mathrm{Im}\widetilde{\chi} = \mathcal{H}[\mathrm{Re}\widetilde{\chi}]$, as it was 'automatically' satisfied by (21*b*).

To get a better understanding of (21*c*), let us put it in a constant electric field that turns on at $t' = 0$. Substituting $\mathbf{E}(t') = \mathbf{E}_0\Theta(t')$ and (21*c*) into (3*a*) yields

$$\mathbf{P}(t) = \mathbf{P}_0\left[1 - \mathrm{e}^{-\frac{\gamma}{2}t}\left(\cos(st) + \frac{\gamma \sin(st)}{2s}\right)\right]\Theta(t), \tag{22a}$$

where

$$\mathbf{P}_0 = \frac{e^2 n_b}{m\omega_0^2}\mathbf{E}_0. \qquad (22b)$$

From (22), we see that $\mathbf{P}(t)$ monotonically increases from zero at $t = 0$ to some maximum value, and then oscillates around $\mathbf{P}_0$ before finally settling at $\mathbf{P}_0$. The oscillations around $\mathbf{P}_0$ remind us that the bounded electrons oscillate around the new equilibrium position.

In the case of vanishing radiation damping, we may set $\gamma = 0$ in (21c) to obtain

$$\chi(t-t') = \frac{e^2 n_b}{m\omega_0}\sin[\omega_0(t-t')]\Theta(t-t'), \qquad (23)$$

which is clearly causal. But if we set $\gamma = 0$ in (21b), we would violate causality! To obtain a causal $\widetilde{\chi}(\omega)$, we set $\gamma = 0$ in (21a) to obtain

$$(\omega-\omega_0)(\omega+\omega_0)\,\widetilde{\chi}(\omega) = -\frac{e^2 n_b}{m}. \qquad (24a)$$

From (A.35) and (A.36), we see that the form of the full solution to (24a) is given by

$$\widetilde{\chi}(\omega) = \frac{e^2 n_b}{2m\omega_0}\left[\frac{1}{\omega+\omega_0} - \frac{1}{\omega-\omega_0} + \widetilde{\chi}_h(\omega)\right], \qquad (24b)$$

where

$$\widetilde{\chi}_h(\omega) = b_0\delta(\omega+\omega_0) + c_0\delta(\omega-\omega_0). \qquad (24c)$$

As in previous examples, we find the unknown constants, b_0 and c_0, via Taylor's causality requirement, $\mathrm{Im}\widetilde{\chi} = \mathcal{H}[\mathrm{Re}\widetilde{\chi}]$, which yields $b_0 = -\mathrm{i}\pi$ and $c_0 = \mathrm{i}\pi$. Therefore, the full casual solution to (24a) is given by

$$\widetilde{\chi}(\omega) = \frac{n_b e^2}{2m\omega_0}\left[\frac{1}{\omega+\omega_0} - \mathrm{i}\pi\delta(\omega+\omega_0) - \frac{1}{\omega-\omega_0} + \mathrm{i}\pi\delta(\omega-\omega_0)\right], \qquad (24d)$$

and the inverse Fourier transform of (24d), of course, yields (23). From the above example, we again conclude that when considering limiting cases of the electric susceptibility in the Fourier domain, we must be careful not to inadvertently violate causality. However, in the time domain, we do not have to worry about the solution not reducing to a proper form in these limiting cases.

6. Concluding remarks

In this paper, we have taken a somewhat contrarian approach to the linear response laws of classical electrodynamics by looking at the response functions in the time domain. The advantage of the time domain is that all quantities are purely real and causality is naturally expressed in terms of time. The disadvantage is that the response functions are temporally non-local, so most of us get tired of writing convolutions on the blackboard and quickly slip into a short-hand mix of time and frequency/Fourier domain notations that can confuse students profoundly.

While it is perfectly reasonable to avoid complications, such as dispersion, in introductory physics courses, by the time students are in their third or fourth year of physics study, it is important to expose them to the fundamental principles associated with a classical, macroscopic picture of matter. In particular, we believe that the following should be emphasized.

- Causality is easy to enforce in the time domain, but the constitutive relations are non-local in time and involve convolution integrals.

- The constitutive relations are mathematically simple in the Fourier domain, but causality is given by the KK relations (the Hilbert transform pair).
- We should make it clear whether we are really transforming into the Fourier domain, or whether we are assuming a monochromatic source in a time-domain experiment. Very often the results look the same, but as we have shown, confusing the two can lead to serious misunderstandings.
- Finally, it should be emphasized that Maxwell's equations (in the time domain) are purely real and involve only purely real quantities. The Fourier transformation promotes variables to the complex plane. As teachers, we should be careful when speaking of the real and imaginary parts of the response function and preface our remarks with a note that we are working in the non-physical, but highly useful, Fourier domain.

Acknowledgments

This material is based upon work supported in part by the US Office of Naval Research as a multi-disciplinary University Research Initiative on sound and electromagnetic interacting waves under grant number N00014-10-1-0958.

Appendix. Distribution theory

The ordinary (classical) functions may be thought of as a mapping between two sets of numbers. We can extend the idea of an ordinary function by considering a mapping (functional, if you will) between a set of *functions* and a set of numbers. The physical reason for extending the idea of an ordinary function lies in our inability to *experimentally* measure a function at a point, e.g., see [21, pp 1–2], [22]. Let $f(x)$ represent temperature at some point x. To measure the temperature at that point, we place a thermometer there to obtain a value for $f(x)$, but do we actually obtain a value of the temperature at point x? The bulb of the thermometer has a finite size; thus, what we measure is an average temperature *around* the point x. Mathematically, an average is a weighted sum, and our measurement only reveals the value $T_1 = \int f(x)\phi_1(x)\,\mathrm{d}x$, where $\phi_1(x)$ is essentially zero away from the bulb of the thermometer. If we make another measurement of the temperature at point x using a different thermometer, then we would measure $T_2 = \int f(x)\phi_2(x)\,\mathrm{d}x$ and, hopefully, the 'true' temperature $T \approx (T_1 + T_2)/2$. The above discussion is meant to serve as a physical motivation for defining what we will call *generalized functions* as certain linear functionals. Generalized functions are also called *distributions*, and we will use both terms interchangeably. Our presentation of the generalized functions closely follows that of Strichartz [21]. A reader interested in a more detailed study may also find [23–26] helpful.

For our purposes, it will be sufficient to consider only a special class of distributions, namely the *tempered* distributions. Before we can formally define tempered distributions, we must first define a set of 'good' functions. This set of 'good' functions is also called the *space of test functions*. For the tempered distributions, the space of test functions, denoted by S, contains all real or complex-valued functions $\phi(t)$ that are classically infinitely differentiable and, along with all its derivatives, vanish at infinity faster than the reciprocal of any polynomial. For example, any function of the form $\sum_{n=0}^{N} c_n t^n \exp(-t^2)$ belongs to S. We are now ready to define the class of tempered distributions, denoted by S', as all continuous[2] *linear functionals* on S. A linear functional f on S is a rule by which we assign to *every* test function $\phi(t)$ a (real or complex) number denoted by $\langle f, \phi\rangle$, such that the identity

[2] The definition of continuity of linear functionals is rather technical and not necessary for our purposes.

$\langle f, c_1\phi_1 + c_2\phi_2\rangle = c_1\langle f, \phi_1\rangle + c_2\langle f, \phi_2\rangle$ is satisfied for arbitrary test functions ϕ_1 and ϕ_2 and (real or complex) numbers c_1 and c_2. The terminology and notation used for distributions can be confusing at times because the phrase 'function f or even generalized function (distribution) f' may refer to f itself or to the value of $\langle f, \phi\rangle$. In other words, no distinction is made between a distribution and a 'function' from which the distribution was obtained[3]. To make the notion of tempered distributions more concrete, let us consider a few simple examples. Let us find a distribution defined by $f = \Theta'(t)$:

$$\mathcal{T}_f = \left\langle \frac{\mathrm{d}\Theta(t)}{\mathrm{d}t}, \phi(t)\right\rangle$$
$$= \int_{-\infty}^{\infty} \frac{\mathrm{d}\Theta(t)}{\mathrm{d}t}\phi(t)\,\mathrm{d}t = -\int_{-\infty}^{\infty} \Theta(t)\frac{\mathrm{d}\phi(t)}{\mathrm{d}t}\,\mathrm{d}t \tag{A.1}$$

$$= -\int_{0}^{\infty} \frac{\mathrm{d}\phi(t)}{\mathrm{d}t}\,\mathrm{d}t = -\left(\phi(\infty) - \phi(0)\right) = \phi(0), \tag{A.2}$$

where we integrated by parts in (A.1) and the integrated terms vanished because ϕ is a 'good' function, i.e. $\phi \in S$. By comparing (A.2) to the sifting property of the Dirac delta function

$$\langle \delta(t), \phi(t)\rangle = \int_{-\infty}^{\infty} \delta(t)\phi(t)\,\mathrm{d}t = \phi(0), \tag{A.3}$$

we conclude that the (generalized) derivative of the Heaviside step function equals the Dirac delta function. We can even differentiate (in a distributional sense, of course) more complicated functions. Let $f = g(t)\delta'(t)$, where $g(t)$ is a polynomial of any degree; then,

$$\mathcal{T}_f = \left\langle g(t)\frac{\mathrm{d}\delta(t)}{\mathrm{d}t}, \phi(t)\right\rangle$$
$$= \int_{-\infty}^{\infty} \left(g(t)\frac{\mathrm{d}\delta(t)}{\mathrm{d}t}\right)\phi(t)\,\mathrm{d}t = \int_{-\infty}^{\infty} \frac{\mathrm{d}\delta(t)}{\mathrm{d}t}(g(t)\phi(t))\,\mathrm{d}t.$$

Integrating by parts and noting that the product of $g(t)$ and $\phi(t)$ is still in S (so that the integrated terms vanish) yields

$$\mathcal{T}_f = -\int_{-\infty}^{\infty} \delta(t)\left(g(t)\frac{\mathrm{d}\phi(t)}{\mathrm{d}t} + \frac{\mathrm{d}g(t)}{\mathrm{d}t}\phi(t)\right)\mathrm{d}t$$
$$= -g(0)\phi'(0) - g'(0)\phi(0). \tag{A.5}$$

We can write (A.5) in a more standard form that does not involve the derivatives of $\phi(t)$. Using (A.3) and noting that

$$\langle \delta'(t), \phi(t)\rangle = \int_{-\infty}^{\infty} \frac{\mathrm{d}\delta(t)}{\mathrm{d}t}\phi(t)\,\mathrm{d}t = -\int_{-\infty}^{\infty} \delta(t)\frac{\mathrm{d}\phi(t)}{\mathrm{d}t}\,\mathrm{d}t = -\phi'(0),$$

we obtain

$$\left\langle g(t)\frac{\mathrm{d}\delta(t)}{\mathrm{d}t}, \phi(t)\right\rangle = \langle g(0)\delta'(t), \phi(t)\rangle \quad - \langle g'(0)\delta(t), \phi(t)\rangle. \tag{A.6}$$

It is a very common abuse of notation to 'drop' the $\langle\cdot,\cdot\rangle$ brackets, along with $\phi(t)$, and write (A.6) simply as

$$g(t)\delta'(t) = g(0)\delta'(t) - g'(0)\delta(t). \tag{A.7}$$

In particular, if we let $g(t) = t$ in (A.4), then (A.7) yields $t\delta'(t) = -\delta(t)$, not just zero as one might have naïvely expected. An alert reader may have noted that in the derivation of

[3] Strictly speaking, this is an abuse of terminology but it is so common that one must be aware of it. Moreover, one often speaks of generalized functions (distributions) as if they were proper functions, e.g., the Dirac delta function $\delta(t)$.

(A.7), we never used the assumption that $g(t)$ is a polynomial; all that the derivation required was $g(t)\phi(t) \in S$. While it is definitely true that $g(t)\phi(t) \in S$ when $g(t)$ is a polynomial, requiring $g(t)$ to be a polynomial is an unnecessary restriction. In other words, (A.7) holds for any function $g(t)$ as long as $g(t)\phi(t) \in S$. For example, $g(t)$ could be $\sin(t)$, but it cannot be $\exp\left(t^4\right)$ because then $g(t)\phi(t) \notin S$ and the integrated terms will not vanish. We considered this example in such detail because we will have numerous opportunities to use (A.7) in appendix A.2.

A.1. The Fourier transform of tempered distributions

The key idea in generalizing the notation of a derivative is to move the derivative from a function (generalized function) onto a set of 'good' functions, namely $\phi(t) \in S$. Moreover, we saw that by integrating by parts enough times, every (generalized) function had a derivative, because the space S is composed of classically infinitely differentiable functions. We will use this 'moving idea' (adjoint operator) to *define* the Fourier transform of a tempered distribution $f(t)$. By the Fourier transform of $f(t) \in S'$, denoted by $\widetilde{f}(\omega)$ or by $\mathcal{F}[f(t)]$, we mean

$$\langle \mathcal{F}[f(t)], \phi(\omega)\rangle = \langle f(t), \mathcal{F}[\phi(\omega)]\rangle, \tag{A.8}$$

where

$$\langle f(t), \mathcal{F}[\phi(\omega)]\rangle = \int_{-\infty}^{\infty} f(t) \left(\int_{-\infty}^{\infty} \phi(\omega)\, \mathrm{e}^{+\mathrm{i}\omega t}\, \mathrm{d}\omega\right) \mathrm{d}t. \tag{A.9}$$

By the inverse Fourier transform of $\widetilde{f}(\omega)$, denoted by $\mathcal{F}^{-1}[\widetilde{f}(\omega)]$, we mean

$$\langle \mathcal{F}^{-1}[\widetilde{f}(\omega)], \widetilde{\phi}(t)\rangle = \langle \widetilde{f}(\omega), \mathcal{F}^{-1}[\widetilde{\phi}(t)]\rangle, \tag{A.10}$$

where

$$\langle \widetilde{f}(\omega), \mathcal{F}^{-1}[\widetilde{\phi}(t)]\rangle = \int_{-\infty}^{\infty} \widetilde{f}(\omega) \left(\frac{1}{2\pi}\int_{-\infty}^{\infty} \widetilde{\phi}(t)\, \mathrm{e}^{-\mathrm{i}\omega t}\, \mathrm{d}t\right) \mathrm{d}\omega. \tag{A.11}$$

By changing the order of integration in (A.9) and (A.11), we see that (A.9) and (A.11) are indeed consistent with the classical Fourier transform pair. The underlining reason for defining the Fourier transform pair by (A.8) and (A.10) lies in the fact that if $\phi \in S$, then $\widetilde{\phi}$ is also in S; the converse is also true. Moreover, it can be shown that $f \in S'$ if and only if $\widetilde{f} \in S'$. For a proof of these and related matters, see [21, chapter 3], [24, chapter 6], [25, chapter 7] and [26, chapter 2]. Now, we will make (A.8)–(A.11) more concrete by considering a few simple examples.

For our first example, we will compute the generalized Fourier transform of $\delta(t)$, i.e. $\langle \mathcal{F}[\delta(t)], \phi(\omega)\rangle$. Substituting $\delta(t)$ into (A.9) and changing the order of integration yields

$$\int_{-\infty}^{\infty} \phi(\omega) \left(\int_{-\infty}^{\infty} \delta(t)\, \mathrm{e}^{+\mathrm{i}\omega t}\, \mathrm{d}t\right) \mathrm{d}\omega = \int_{-\infty}^{\infty} \phi(\omega)\, \mathrm{d}\omega = \langle 1, \phi(\omega)\rangle. \tag{A.12}$$

Thus, we see that $\mathcal{F}[\delta(t)] = 1$. Moreover, by taking the inverse Fourier transform, we obtain the famous integral representation of the Dirac delta function, namely

$$\delta(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\omega t}\, \mathrm{d}\omega. \tag{A.13}$$

It is worth stressing that (A.13) should be interpreted in a distribution sense and, strictly speaking, writing (A.13) as we did is an abuse of notation. However, such abuses of notation are very common in physics, e.g., see the famous graduate electrodynamics textbook [27].

As another simple example, consider the generalized inverse Fourier transform of $2\pi\widetilde{\delta}(\omega - \omega_0)$, where ω_0 is a real number, i.e. $\langle \mathcal{F}^{-1}\left[2\pi\widetilde{\delta}(\omega - \omega_0)\right], \widetilde{\phi}(t)\rangle$ when $\omega_0 \in \mathbb{R}$.

Substituting $2\pi\widetilde{\delta}(\omega-\omega_0)$ into (A.11) and then changing the order of integration and integrating over ω yields

$$\int_{-\infty}^{\infty}\widetilde{\phi}(t)\,\mathrm{e}^{-\mathrm{i}\omega_0 t}\,\mathrm{d}t = \langle \mathrm{e}^{-\mathrm{i}\omega_0 t}, \widetilde{\phi}(t)\rangle. \tag{A.14}$$

Thus, we see that $\mathcal{F}^{-1}\left[\widetilde{\delta}(\omega-\omega_0)\right] = \exp(-\mathrm{i}\omega_0 t)$.

For our last example, let us compute the Fourier transform of the nth generalized derivative of $f(t)\in S'$. From (A.8), we have

$$\left\langle \mathcal{F}\left[\frac{\mathrm{d}^n}{\mathrm{d}t^n}f(t)\right], \phi(\omega)\right\rangle = \int_{-\infty}^{\infty}\left(\frac{\mathrm{d}^n}{\mathrm{d}t^n}f(t)\right)\widetilde{\phi}(t)\,\mathrm{d}t. \tag{A.15}$$

Performing integration by parts n-times on the right-hand side (where the integrated terms vanished because $\widetilde{\phi}\in S$) yields

$$(-1)^n\int_{-\infty}^{\infty} f(t)\left(\frac{\mathrm{d}^n}{\mathrm{d}t^n}\widetilde{\phi}(t)\right)\mathrm{d}t = (-1)^n\int_{-\infty}^{\infty} f(t)\mathcal{F}[(+\mathrm{i}\omega)^n\phi(\omega)]\,\mathrm{d}t$$

$$= \int_{-\infty}^{\infty}\mathcal{F}[f(t)](-\mathrm{i}\omega)^n\phi(\omega)\,\mathrm{d}\omega, \tag{A.16}$$

where we obtained (A.16) by using the definition (A.8). Finally, comparing (A.16) with the left-hand side of (A.15) yields

$$\mathcal{F}\left[\frac{\mathrm{d}^n}{\mathrm{d}t^n}f(t)\right] = (-\mathrm{i}\omega)^n\,\mathcal{F}\left[f(t)\right]. \tag{A.17}$$

Note that in this example, we have used the definition (A.8) *twice*; Strichartz [21, pp 49–50] appropriately refers to this as 'definition chasing'.

No discussion of the Fourier transform of tempered distributions would be complete without the Fourier transform of a convolution integral. For our purposes, it will be sufficient to only consider convolution of the tempered distribution f with a fixed element ψ from the space of 'good' functions. Let $f\in S'$ and $\psi\in S$; then, by the convolution of f with ψ, denoted by $(f*\psi)\,(t)$, we mean

$$(f*\psi)(t) = \int_{-\infty}^{\infty} f(t-t')\psi(t')\,\mathrm{d}t'. \tag{A.18}$$

By a simple change of variables, we see that convolution is commutative, i.e. $(f*\psi)(t) = (\psi*f)(t)$. Convolution defines an infinitely differentiable function and thus can be viewed as a 'smoothing' process. To see this, let $h(t) = (f*\psi)(t)$ and then

$$\frac{\mathrm{d}^n}{\mathrm{d}t^n}h(t) = \frac{\mathrm{d}^n}{\mathrm{d}t^n}[(f*\psi)(t)] = \int_{-\infty}^{\infty}\left(\frac{\mathrm{d}^n}{\mathrm{d}t^n}f(t-t')\right)\psi(t')\,\mathrm{d}t' \tag{A.19}$$

$$= \frac{\mathrm{d}^n}{\mathrm{d}t^n}[(\psi*f)(t)] = \int_{-\infty}^{\infty}\left(\frac{\mathrm{d}^n}{\mathrm{d}t^n}\psi(t-t')\right)f(t')\,\mathrm{d}t'. \tag{A.20}$$

From (A.20), we conclude that all derivatives of $h(t)$ exist in the classical sense because $\psi\in S$, and from (A.19), we see that it does not matter how 'rough' (e.g., $\delta'(t)$ is 'rougher' than $\delta(t)$) is the distribution. In passing, we note another useful property of (A.18), namely that its Fourier transform corresponds to multiplication in the Fourier domain:

$$\langle\mathcal{F}[(\psi*f)(t)], \phi(\omega)\rangle = \langle\widetilde{\psi}(\omega)\widetilde{f}(\omega), \phi(\omega)\rangle. \tag{A.21}$$

A.2. Support and structure of tempered distributions

We often want to speak about the local properties of distributions as if the distributions were ordinary (classical) functions. When speaking about an ordinary function $f(t)$, the statement '$f(t)$ has a value of $f(t_1)$ when $t = t_1$' has meaning, but the statement is nonsense if $f(t)$ is a distribution[4]. However, we *can* identify a set of points where distribution f is non-zero. Loosely speaking, this set of points is known as the *support* of f. The formal definition of support requires us to define where a distribution f *is zero*. We say a distribution $f(t)$ is zero, $f(t) = 0$, on an *open* interval (a, b) if $\langle f(t), \psi(t)\rangle = 0$ for every infinitely differentiable test function $\psi(t)$ that vanishes in a *neighbourhood* of every point *not* in the (a, b) interval. For example, if $f(t) = 0$ on $(-1, 1)$, then $\psi(t)$ vanishes *outside* the $(-1+\epsilon, 1-\epsilon)$ interval for some $\epsilon > 0$. We now formally define the support of a distribution $f(t)$, denoted by $\operatorname{supp}[f(t)]$, as the *complement* of the set of points t such that $f(t) = 0$ in a *neighbourhood* of t. For example, if $f(t) = 0$ on $(-\infty, \infty)$, then $\operatorname{supp}[f(t)]$ is the empty set and, as a less trivial example, $\operatorname{supp}[\delta(t)] = \{0\}$. Moreover, we will show that

$$\operatorname{supp}\left[\frac{\mathrm{d}^n}{\mathrm{d}t^n}\delta(t)\right] = \{0\}\,. \tag{A.22}$$

Consider any open interval I_1 that does *not* contain the point $t = 0$; then,

$$\left\langle \frac{\mathrm{d}^n}{\mathrm{d}t^n}\delta(t), \psi(t)\right\rangle = (-1)^n\left\langle \delta(t), \frac{\mathrm{d}^n}{\mathrm{d}t^n}\psi(t)\right\rangle = (-1)^n\frac{\mathrm{d}^n}{\mathrm{d}t^n}\psi(0) = 0$$

because $\psi(t)$ vanishes in a *neighbourhood* of $t = 0$. Of course, for any open interval I_2 that *does* contain the point $t = 0$, the derivatives of ψ do not vanish at the point $t = 0$ for every test function ψ. Thus, we see that all derivatives of $\delta(t)$ have the same point support.

The above discussion was necessary to understand the following 'structure' theorem. A tempered distribution $f(t)$ with $\operatorname{supp}[f(t)] = \{t_0\}$ must be of the form [21, pp 82–8]

$$f(t) = \sum_{n=0}^{N} c_n \frac{\mathrm{d}^n}{\mathrm{d}t^n}\delta(t - t_0), \tag{A.23}$$

where the coefficients $c_{n=0,\ldots,N}$ are the complex numbers, i.e. $c_n \in \mathbb{C}$ for $n = 0, \ldots, N$. In other words, any tempered distribution with a point support may be expressed as a *finite* linear combination of the Dirac function and its derivatives; this is a powerful statement! In the next paragraph, we will show how we can use (A.23) to solve 'algebraic' equations in a distributional sense. These types of equations are frequently encountered when we solve differential equations by the Fourier transform technique.

As our first example, consider the following equation:

$$\langle (t - t_0) f(t), \phi(t)\rangle = \langle 1, \phi(t)\rangle, \tag{A.24}$$

for all $\phi(t) \in S$. Before we find the unknown tempered distribution $f(t)$, we note that it is customary to abuse the notation and write (A.24) simply as

$$(t - t_0) f(t) = 1. \tag{A.25}$$

Naïvely, we might expect that

$$f_{\mathrm{p}}(t) = \frac{1}{t - t_0} \tag{A.26}$$

would solve (A.25), but this is only the particular part of the solution. We could have a tempered distribution $f_{\mathrm{h}}(t)$ with $\operatorname{supp}[f_{\mathrm{h}}(t)] = \{t_0\}$, such that $(t - t_0) f_{\mathrm{h}}(t) = 0$, and therefore, $f(t) = f_{\mathrm{p}}(t) + f_{\mathrm{h}}(t)$ would also satisfy (A.25). We refer to $f_{\mathrm{h}}(t)$ as the homogeneous solution

[4] Recall that we have defined distributions only by their action on the space of test functions.

and, in light of the structure theorem in the previous paragraph, we know it must be of the form

$$f_{\rm h}(t) = \sum_{n=0}^{N} c_n \frac{{\rm d}^n}{{\rm d}t^n}\delta(t-t_0). \tag{A.27}$$

Substituting (A.27) into $\langle (t-t_0) f_{\rm h}(t), \phi(t)\rangle = \langle 0, \phi(t)\rangle$ and integrating by parts until the derivatives only appear on ϕ yields

$$-c_1\phi(t_0) + 2c_2\phi'(t_0) - 3c_3\phi''(t_0) + \cdots + (-1)^N N c_N \phi^{(N-1)}(t_0) = 0.$$

The above equation must hold for all $\phi \in S$. Thus, the coefficients must vanish independently, i.e. $c_n = 0$ for $n = 1, 2, \ldots, N$. Therefore, the homogeneous solution is given by

$$f_{\rm h}(t) = c_0\delta(t-t_0), \tag{A.28}$$

and the full solution to (A.25) (or more formally, to (A.24)) is given by

$$f(t) = \frac{1}{t-t_0} + c_0\delta(t-t_0). \tag{A.29}$$

Loosely speaking, from (A.29), we see that we can divide by zero, provided we add an appropriate tempered distribution with a point support. In general, using the same procedure as above, we can show that the distributional solution to $(t-t_0)^n f(t) = 1$ is given by $f(t) = f_{\rm p}(t) + f_{\rm h}(t)$, where

$$f_{\rm p}(t) = \frac{1}{(t-t_0)^n}, \tag{A.30}$$

$$f_{\rm h}(t) = c_0\delta(t-t_0) + c_1\delta'(t-t_0) + \cdots + c_{n-1}\delta^{(n-1)}(t-t_0). \tag{A.31}$$

For our second and last example, consider (in a distributional sense, of course) the following equation:

$$(t-t_1)(t-t_2) f(t) = 1, \tag{A.32}$$

where $t_1 \neq t_2$. From our previous example, we expect the solution to be of the form

$$f(t) = \frac{1}{(t-t_1)(t-t_2)} + \sum_{n=0}^{N} b_n \delta^{(n)}(t-t_1) + \sum_{m=0}^{M} c_m \delta^{(m)}(t-t_2). \tag{A.33}$$

Substituting (A.33) into (A.32) (of course, we actually mean $\langle (t-t_1)(t-t_2) f(t), \phi(t)\rangle = \langle 1, \phi(t)\rangle$, for all $\phi \in S$) and integrating by parts until the derivatives only appear on ϕ yields

$$\sum_{n=1}^{N} (-1)^n b_n [n(n-1)\phi^{(n-2)}(t_1) + n(t_1-t_2)\phi^{(n-1)}(t_1)]$$
$$+ \sum_{m=1}^{M} (-1)^m c_m [m(m-1)\phi^{(m-2)}(t_2) - m(t_1-t_2)\phi^{(m-1)}(t_2)] = 0. \tag{A.34}$$

From (A.34), we see that $b_n = 0$ for $n = 1, \ldots, N$ and $c_m = 0$ for $m = 1, \ldots, M$. Therefore, the full solution to (A.32) is given by $f(t) = f_{\rm p}(t) + f_{\rm h}(t)$, where

$$f_{\rm p}(t) = \frac{1}{(t-t_1)(t-t_2)}, \tag{A.35}$$

$$f_{\rm h}(t) = b_0\delta(t-t_1) + c_0\delta(t-t_2). \tag{A.36}$$

Note that if $t_1 = t_2$, then (A.36) does *not* yield the correct solution, which is given by (A.31) (with $n = 2$ and $t_1 = t_2 \to t_0$). The reason for this 'discrepancy' is because our conclusion from (A.34), namely that $b_{n=1,\ldots,N} = 0$ and $c_{m=1,\ldots,M} = 0$, is *not* valid if $t_1 = t_2$. In other words, we must be very careful when dealing with distributional solutions in limiting cases such as $t_2 \to t_1$. In the body of the paper, these limiting cases arise when we consider vanishing absorption.

References

[1] Griffiths D J 1999 *Introduction to Electrodynamics* 3rd edn (Upper Saddle River, NJ: Prentice-Hall) pp 382–3
[2] Heald M A and Marion J B 1995 *Classical Electromagnetic Radiation* 3rd edn (Fort Worth, TX: Saunders) pp 167–8
[3] Grant I S and Phillips W R 1994 *Electromagnetism* 2nd edn (Chichester: Wiley) p 356
[4] Kinsler P 2011 How to be causal: time, spacetime and spectra *Eur. J. Phys.* **32** 1687
[5] Toll J S 1956 Causality and the dispersion relation: logical foundations *Phys. Rev.* **104** 1760–70
[6] Hu B Y-K 1989 Kramers–Kronig in two lines *Am. J. Phys.* **57** 821
[7] Reitz J R, Milford F J and Christy R W 1993 *Foundations of Electromagnetic Theory* 4th edn (Reading, MA: Addison-Wesley) pp 610–5
[8] Bohren C F 2010 What did Kramers and Kronig do and how did they do it? *Eur. J. Phys.* **31** 573
[9] Titchmarsh E C 1948 *Introduction to the Theory of Fourier Integrals* 2nd edn (Oxford: Clarendon) pp 128–9
[10] Beltrami E and Wohlers M 1967 The Cauchy integral of tempered distributions and some theorems on analytic continuation *SIAM J. Appl. Math.* **15** 1077–87
[11] Orton M 1973 Hilbert transforms, Plemelj relations and Fourier transforms of distributions *SIAM J. Math. Anal.* **4** 656–70
[12] Ishikawa S 1987 Generalized Hilbert transforms in tempered distributions *Tokyo J. Math.* **10** 119–32
[13] Pandey J N 1996 *The Hilbert Transform of Schwartz Distributions and Applications* (New York: Wiley) chapter 4
[14] Taylor J G 1958 Dispersion relations and Schwartz's distributions *Ann. Phys.* **5** 391–8
[15] King F W 2009 *Hilbert Transforms* vol 2 (New York: Cambridge University Press) pp 110–7
[16] Ashcroft N W and Mermin N D 1976 *Solid State Physics* (Philadelphia, PA: Saunders) chapter 1
[17] Landau L D, Lifshitz E M and Pitaevskiĭ L P 1993 *Electrodynamics of Continuous Media* 2nd edn vol 8 (Oxford: Pergamon) chapter 12
[18] Dressel M and Grüner G 2002 *Electrodynamic of Solids: Optical Properties of Electrons in Matter* (Cambridge: Cambridge University Press) chapter 5
[19] Zadler B J, Grêt A and Scales J A 2005 Spectroscopy versus interferometry: resolving small changes *Am. J. Phys.* **73** 837–44
[20] Schwartz L 1966 *Théorie des Distributions* (Paris: Hermann) pp 243–8
[21] Strichartz R S 2003 *A Guide to Distribution Theory and Fourier Transforms* (River Edge, NJ: World Scientific)
[22] Papoulis A 1962 *The Fourier Integral and Its Applications* (New York: McGraw-Hill) pp 281–2
[23] Lighthill M J 1959 *Introduction to Fourier Analysis and Generalised Functions* (Cambridge: Cambridge University Press)
[24] Kanwal R P 2004 *Generalized Functions Theory and Applications* 3rd edn (Boston, MA: Birkhäuser)
[25] Zemanian A H 1965 *Distribution Theory and Transform Analysis* (New York: McGraw-Hill)
[26] Gel'fand I M and Shilov G E 1964 *Generalized Functions* vol 1 (New York: Academic)
[27] Jackson J D 1999 *Classical Electrodynamics* 3rd edn (New York: Wiley) pp 69–70