

A mathematical framework for inverse wave problems in heterogeneous media

Kirk D Blazek¹, Christiaan Stolk² and William W Symes³

¹ Rochester Community and Technical College, Rochester, MN 55904, USA

² KdV Institute for Mathematics, University of Amsterdam, PO Box 94216, 1090 GE Amsterdam, The Netherlands

³ The Rice Inversion Project, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77251-1892, USA

E-mail: kirk.blazek@gmail.com, c.c.stolk@uva.nl and symes@caam.rice.edu

Received 24 November 2012, in final form 18 March 2013

Published 19 April 2013

Online at stacks.iop.org/IP/29/065001

Abstract

This paper provides a theoretical foundation for some common formulations of inverse problems in wave propagation, based on hyperbolic systems of linear integro-differential equations with bounded and measurable coefficients. The coefficients of these time-dependent partial differential equations represent parametrically the spatially varying mechanical properties of materials. Rocks, manufactured materials, and other wave propagation environments often exhibit spatial heterogeneity in mechanical properties at a wide variety of scales, and coefficient functions representing these properties must mimic this heterogeneity. We show how to choose domains (classes of nonsmooth coefficient functions) and data definitions (traces of weak solutions) so that optimization formulations of inverse wave problems satisfy some of the prerequisites for application of Newton's method and its relatives. These results follow from the properties of a class of abstract first-order evolution systems, of which various physical wave systems appear as concrete instances. Finite speed of propagation for linear waves with bounded, measurable mechanical parameter fields is one of the by-products of this theory.

1. Introduction

Inverse problems for waves in heterogeneous materials occur in seismology, ultrasonic nondestructive evaluation and biomedical imaging, some electromagnetic imaging technologies, and elsewhere. The data of such problems are often idealized as traces of



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](http://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

wave fields on space-time hypersurfaces (or surfaces of even lower dimension), and nonlinear least-squares (or other data-fitting) methods developed for their numerical solution, that is, for finding the coefficients in the systems of partial differential equations chosen to model the wave physics. Newton's method (or one of its relatives) is a natural choice of solution method for these optimization formulations, as its rapid convergence compensates to some extent for the very large computational size of reasonable discretizations—see for example Epanomeritakis *et al* (2009).

Natural and man-made materials may be so heterogeneous that few limits can be placed on the spatial regularity of coefficients representing material parameters (for the case of sedimentary rocks, see for instance (Walden and Hosken 1986, Bourbie *et al* 1987, White *et al* 1990)). However, traces are well defined only for fields possessing some space-time regularity, and optimal convergence of Newton-like methods requires that objective functions and constraints be regular functions of the model parameters, in some sense. Since nonsmoothness of the material parameter (coefficient) fields implies nonsmoothness of dynamical (solution) fields, it is *a priori* unclear that the required traces exist, or that the fields depend sufficiently smoothly on the material parameter (coefficient) functions to permit Newton-like optimization methods to be applied. While discretized models may *a fortiori* produce nominally well-defined data simulations and smooth objective (data misfit) functions, accurate approximation of the continuum fields implies that the continuum limit will determine the behavior of solution algorithms for the corresponding discretized inverse problems. Also, numerical solutions to discretized inverse problems respect the underlying continuum physics only insofar as they converge under refinement of the discretization (at least in principle), and such convergence generally requires that the continuum problems have well-behaved solutions.

The theory presented here provides a mathematical foundation for some common idealizations of inverse problems in wave propagation and approaches to their solution, for material models of bounded and measurable dependence on space variables. Measurability means roughly the existence of averages over arbitrary volumes, which seems a reasonable requirement for physical parameter fields. Most such fields are bounded as a matter of principle (positivity of density) or observation (p-wave velocity range in metals, . . .).

The dynamical laws considered in this paper may be cast as *symmetric hyperbolic systems*, either differential or integro-differential, for vector-valued fields u defined on a domain $\Omega \subset \mathbf{R}^d$ (typically $d = 1, 2$ or 3). Denoting the dimension of the dynamical state vector u by $k \in \mathbf{N}$, the systems treated in this paper take the form

$$a \frac{\partial u}{\partial t} + p(\nabla)u + bu + q * u = f \text{ in } \Omega \times \mathbf{R}; u = 0 \text{ for } t < 0, \quad (1)$$

in which a, b are $k \times k$ matrix-valued functions on Ω , q is a $k \times k$ matrix-valued function on $\Omega \times \mathbf{R}$ with $q(\mathbf{x}, t) = 0$ for $t < 0$, and $p(\nabla)$ is a $k \times k$ matrix of constant-coefficient first-order differential operators in the space variables:

$$p(\nabla)u = \sum_{j=1}^d p_j \frac{\partial u}{\partial x_j}, \quad p_j \in \mathbf{R}^{k \times k}, \quad j = 1, \dots, d. \quad (2)$$

Such a system is *symmetric hyperbolic* if $a(\mathbf{x}), q(\mathbf{x}, t)$ and p_j are symmetric for all $\mathbf{x} \in \Omega, t \in \mathbf{R}$, respectively $j = 1, \dots, d$, and $a(\mathbf{x})$ is uniformly positive-definite: $C_*, C^* \in \mathbf{R}$ exist so that $0 < C_* \leq C^*$, and

$$C_* I \leq a(\mathbf{x}) \leq C^* I, \quad \text{a.e. } \mathbf{x} \in \Omega \quad (3)$$

in which I is the $d \times d$ identity matrix, and the inequality sign signifies that the matrix on the right differs from the matrix on the left by a positive definite symmetric matrix.

This paper has three central objectives. First, we describe constraints on data (coefficients a , b , p , q and right-hand side f) and solution under which (1) is well posed, even when the coefficient functions a , b , q are permitted to be quite irregular in their dependence on $\mathbf{x} \in \Omega$ —in fact, merely bounded and measurable. Second, we determine a sense in which the solution of (1) is regular (C^k , $k \geq 0$) as a function of its coefficients. Third, we characterize conditions under which traces of solutions on time-like hypersurfaces are well defined, and regular as functions of the coefficients.

The complete description of the system (1) requires the specification of domains and ranges for all of the operators appearing in it. We will require that these be chosen so that the expression $p(\nabla)$ defines a skew-adjoint operator.

We will describe two important examples (acoustics and viscoelasticity) in detail below, as symmetric hyperbolic problems of the form (1). The coefficient matrices a , b and q encapsulate spatially varying relations between space and time rates of change of the dynamical state u . A particularly important example, a common component of all linear continuum mechanical systems, is the linearized Newton's law, in which the product of the material density and the time rate of change of velocity balances the sum of internal and external forces acting on the material. The internal forces are expressed as the divergence of a stress tensor. The material density may vary with position and appears somewhere in a , the velocity and stress fields form part of u , and the (spatially invariant) divergence is part of p . These relations may be expressed in algebraically equivalent form via other constitutive laws governing particular examples, but are always present in some form. The time convolution kernel q expresses dissipative and dispersive mechanisms built into the mechanical laws. These mechanisms may originate in upscaling of microscopic physics (for example, poroelasticity), or express generic dissipative mechanisms (viscoelasticity), or be absent (elasticity, acoustics), in particular examples of (1). The right-hand side vector f is a k -vector valued function on $\Omega \times \mathbf{R}$, vanishing for $t < 0$ and representing energy input to the system, in the form of external forces, defects in the constitutive laws, and so on.

As in the similar study of Stolk (2000) for second-order hyperbolic systems, we represent (1) as an instance of a class of abstract autonomous integro-differential systems. We study these abstract systems within a framework modeled after the duality approach introduced by Lions and his collaborators in the 1960s (Lions 1971, Lions and Magenes 1972). A natural notion of *weak solution* is central to this approach, and complements that of strong (pointwise) solution. The class of first-order systems studied here possesses a remarkable property: smoothing solutions in time converts weak solutions to strong (pointwise in t) solutions. Specialized to systems such as acoustics or elastodynamics, this property implies that appropriate traces of solutions are well defined, provided that the right-hand-side or source term is minimally smooth in time (only!). Existence, uniqueness, and regularity of the solution as function of problem data—both right-hand side and (operator) coefficients—also follows from this fact.

The abstract theory accomplishes even more than that, as it applies to a much wider range of dynamics than those commonly occurring in continuum mechanics, accommodating first-order systems with operator coefficients. This added generality creates no additional difficulty for the basic theory. It is in fact very useful: as will be explained in the *Discussion* section, it justifies certain infeasible-model methods for inverse problems in wave propagation.

Another useful by-product of the theory is the finite speed of propagation property for hyperbolic systems with bounded, measurable coefficients, a result which so far as we can tell is new. This follows from the similar property for systems with smooth coefficients via the continuous dependence of solutions on the coefficients in the sense of convergence in measure.

The next section of this paper states and discusses the main results to be established for symmetric hyperbolic systems (1). In the third section, we show how acoustics may be

construed as an example of our framework: we formulate the physics of linear acoustics in the form of a symmetric hyperbolic system, and show how to define the operators involved so that our main results apply to this example, in particular so that the differential operator $p(\nabla)$ is skew-adjoint. In the following section, we do the same for viscoelastodynamics. Following these extended examples, the fifth section of this paper introduces a class of abstract evolution systems encompassing all of the concrete systems mentioned earlier, and lists the main results to be established concerning it. Having explained our main results and some important examples to which they apply, we devote the sixth section to a discussion of the related literature and various implications for inverse problems in wave propagation, including the importance of the general (operator coefficient) case of the abstract theory in formulating certain inversion algorithms, representation of energy sources, and several other matters not addressed in the remainder of this paper. The succeeding sections develop proofs of the main results for the abstract evolution class described in the fifth section. The final section of this paper specializes these abstract results to provide proofs of the theorems stated in the next section. Two appendices discuss skew-adjointness of the acoustic grad-div operator, and the case of systems with no memory term ($q = 0$ in (1)), for which initial value problems make sense.

2. Main results for symmetric hyperbolic systems

Our results hinge on an assumption concerning the skew-symmetric differential operator $p(\nabla)$, which to begin with may be viewed as a densely-defined operator on $L^2(\Omega)^k$, whose domain is a subset of $C^\infty(\Omega^{\text{int}})^k$ that contains $C_0^\infty(\Omega^{\text{int}})^k$. We assume that $p(\nabla)$ extends to a skew-adjoint operator P with dense domain $V \subset L^2(\Omega)^k$, and that V is metrized with the graph norm of $p(\nabla)$ or its equivalent, which we denote by $\|\cdot\|_V$.

The first of our two major sets of results on symmetric hyperbolic systems addresses well posedness.

Theorem 1. *Suppose that, in addition to the hypotheses explained above, the right-hand side f has a square-integrable t -derivative: $f \in H_{\text{loc}}^1(\mathbf{R}, L^2(\Omega)^k)$, and is causal, that is, $f(t) = 0$ for $t < 0$. Then there exists a unique causal $u \in C^1(\mathbf{R}, L^2(\Omega)^k) \cap C^0(\mathbf{R}, V)$ satisfying (1) a.e. in Ω for each $t \in \mathbf{R}$. Moreover, there exists an increasing function $C : \mathbf{R}_+ \rightarrow \mathbf{R}_+$, depending on C_* , C^* , and bounds for b, q , so that for every $t \in \mathbf{R}$,*

$$\left\| \frac{\partial u}{\partial t}(\cdot, t) \right\|_{L^2(\Omega)^k}^2 + \|u(\cdot, t)\|_V^2 \leq C(t) \int_0^t d\tau \left[\|f(\cdot, \tau)\|_{L^2(\Omega)^k}^2 + \left\| \frac{\partial u}{\partial t}(\cdot, \tau) \right\|_{L^2(\Omega)^k}^2 \right] \quad (4)$$

The second main result addresses the existence of traces and their regular dependence on the coefficients. Suppose that $\Gamma \subset \Omega$ is a smoothly embedded m -dimensional submanifold, $m < d$. Denote by Σ_Γ the trace map $C_0^\infty(\Omega^{\text{int}})^k \rightarrow L^2(\Gamma)^k$. Suppose that $l \in \mathbf{N}$ and $m \in C_0^\infty(\Gamma, \mathbf{R}^{l \times k})$. Define $S[m] : C_0^\infty(\Omega^{\text{int}})^k \rightarrow L^2(\Gamma)^l$ by

$$S[m]u = m\Sigma_\Gamma u.$$

Note that $S[m]$ induces a map $L^2([0, T], C_0^\infty(\Omega^{\text{int}})^k) \rightarrow L^2([0, T], L^2(\Gamma)^l)$, for any $T > 0$. Abusing notation, call this induced map $S[m]$ as well.

Having parametrized the various possible trace maps, we define the data prediction, or forward, operator, $\mathcal{F}_{f,m}$, by

$$\mathcal{F}_{f,m}(a, b, q) = S[m]u$$

in which u is the causal solution of (1) provided by theorem 1. A suitable domain for $\mathcal{F}_{f,m}$ is $M \subset L^\infty(\Omega, \mathbf{R}_{\text{symm}}^{k \times k}) \times L^\infty(\Omega, \mathbf{R}^{k \times k}) \times L^1(\mathbf{R}_+, L^\infty(\Omega, \mathbf{R}_{\text{symm}}^{k \times k}))$, defined by

$$M(C_*, C^*, C_B, C_Q) = \{(a, b, q) : C_* I < a(\mathbf{x}) < C^* I \text{ for all } \mathbf{x} \in \Omega, \|b\|_{L^\infty(\Omega, \mathbf{R}^{k \times k})} < C_B,$$

$$\|q\|_{L^1(\mathbf{R}_+, L^\infty(\Omega, \mathbf{R}_{\text{symm}}^{k \times k}))} < C_Q, q(t) = 0 \text{ for } t < 0\},$$

$$M = \bigcup \{M(C_*, C^*, C_B, C_Q) : 0 < C_* \leq C^*, C_B, C_Q \in \mathbf{R}_+\}.$$

Theorem 2. In addition to the hypotheses of theorem 1, suppose that $S[m]$ extends continuously to V . Then for any $T > 0$,

(i) if $f \in H_{\text{loc}}^s(\mathbf{R}, L^2(\Omega)^k)$, $s \geq 1$, then $\mathcal{F}_{f,m}$ defines a continuous map: $M \rightarrow C^{s-1}([0, T], L^2(\Gamma)^l)$;

(ii) if also $s \geq 2$ and, then $\mathcal{F}_{f,m} : M \rightarrow C^{s-2}([0, T], L^2(\Gamma)^l)$ is of class C^{s-1} .

A stronger result concerning continuity *per se*, involving a weaker topology on the coefficient set M , is also important. Recall that a sequence of real-valued measurable functions $\{g_m : m \in \mathbf{N}\} \subset L^\infty(\Omega)$ converges to zero in measure iff for any $\epsilon > 0$,

$$e_m = |\{\mathbf{x} : |g_m(\mathbf{x})| > \epsilon\}|$$

is a null sequence.

Theorem 3. Suppose that $0 < C_* \leq C^*, 0 \leq C_B, 0 \leq C_Q, \{(a_m, b_m, q_m) : m \in \mathbf{N}\} \subset M(C_*, C^*, C_B, C_Q)$, and $(a, b, q) \in M(C_*, C^*, C_B, C_Q)$. Suppose further that

(i) $a_m \rightarrow a$ in measure,

(ii) $b_m \rightarrow b$ in measure,

(iii) $\int_0^t |q_m - q| \rightarrow 0$ in measure

Denote by u_m the (strong) solution of (1) provided by theorem 1, with coefficients a_m, b_m , and q_m , likewise by u the solution with coefficients a, b , and q , with common causal right-hand side $f \in H_{\text{loc}}^1(\mathbf{R}, L^2(\Omega)^k)$. Then for any $T > 0$,

$$\lim_{m \rightarrow \infty} \|u_m - u\|_{L^\infty([0, T], L^2(\Omega))} = 0.$$

One way to generate sequences converging in measure is by mollification. We exploit this observation to use theorem 3 in conjunction with a well-known fact about symmetric hyperbolic systems with smooth coefficients, to prove

Theorem 4. Suppose that $(a, b, q) \in M$, $\epsilon > 0$, and that $\tau \in \mathbf{R}_+$ satisfies

$$\tau a(\mathbf{x}) + \sum_{i=1}^d p_i \xi_i \geq 0 \quad (5)$$

for almost every $\mathbf{x} \in \Omega$ and every $\xi \in \mathbf{R}^d$ for which $|\xi| = 1$. Suppose $\omega \subset \Omega$, $T > 0$, and that $f \in H_{\text{loc}}^1(\mathbf{R}, L^2(\Omega)^k)$ is causal and satisfies $f(\mathbf{x}, t) = 0$ if $|\mathbf{x} - \mathbf{x}_0| \leq \tau(T - t)$ for all $\mathbf{x}_0 \in \omega$. Then the causal solution $u \in C^1(\mathbf{R}, L^2(\Omega)^k) \cap C^0(\mathbf{R}, V)$ of (1) with coefficients a, b, q and right-hand side f vanishes in $\omega \times [0, T]$.

Remark. Note that theorem 2 expresses ‘differentiability with a loss of one derivative’, in two senses. First, taking $k = 2$, even though $f \in H_{\text{loc}}^2(\mathbf{R}, L^2(\Omega)^k)$, the map $\mathcal{F}_{f,m}$ is only once differentiable. Second, even though $\mathcal{F}_{f,m}$ a priori takes values in $C^1([0, T], L^2(\Gamma)^l)$, it is of class C^1 only as a map with range $C^0([0, T], L^2(\Gamma)^l)$.

This ‘loss of derivative’ phenomenon is a fundamental feature of the coefficient-to-solution map for symmetric hyperbolic PDEs, and accounts for much of the ornery nature of the related inverse problems. A simple explicit example illustrates the second sense of derivative loss, and shows that the hypotheses of theorems 1 and 2 cannot be significantly weakened.

The linear advection equation

$$\left(\frac{1}{c} \frac{\partial u}{\partial t} - \frac{\partial u}{\partial x}\right)(t, x) = f(t, x), \quad (6)$$

for waves moving to the left with speed $c > 0$, conforms to the requirements of theorem 1, with $\Omega = \mathbf{R}$, with the choices

$$p(\nabla) = \frac{\partial}{\partial x}, V = H^1(\mathbf{R}),$$

and $A =$ multiplication by $1/c$. A proof that $p(\nabla)$ is self-adjoint with domain V follows from use of lemma 6 from appendix A, together with the Fourier transform characterization of $H^1(\mathbf{R})$.

The trace manifold in this example is $\Gamma = \{x = 0\}$, and $k = l = 1$; the map m is simply the identity, and is suppressed from the notation. Thus traces take values in $L^2(\Gamma) = \mathbf{R}$.

The causal solution of (6) is formally

$$u(t, x) = \int_x^\infty f\left(t + \frac{x-y}{c}, y\right) dy.$$

That is, this expression gives the solution guaranteed by theorem 1 when $f \in C_0^\infty(\mathbf{R}^2)$.

It is easy to check that for causal $f \in L_{\text{loc}}^2(\mathbf{R}^2)$, the integral defines a locally square-integrable distribution with well-defined trace on $x = 0$ —in fact, a weak solution of (6) in the sense explained in the section 6. Weak solutions do not generally possess well-defined traces; the reasons for this exception are mentioned briefly below, and at more length in the *Discussion* section. In this particular example, the trace on $x = 0$ defines a continuous map $\mathbf{R} \mapsto L_{\text{loc}}^2(\mathbf{R})$. For the particular choice

$$f(t, x) = H(t + x - 1)H(t)$$

(H being the Heaviside function), one finds that for $c = 1$,

$$\mathcal{F}_f[1](c) = tH(t - 1),$$

(m is suppressed, as it is simply the identity in this case). Thus with $f \in L_{\text{loc}}^2(\mathbf{R}, L^2(\mathbf{R}))$, we cannot conclude that \mathcal{F}_f takes values in $C^0(\mathbf{R}, \mathbf{R})$.

For $c > 1$,

$$u(t, 0) = \begin{cases} 0, & t < \frac{c}{2c-1} \\ \frac{c}{c-1}((2c-1)t - c), & \frac{c}{2c-1} \leq t \leq 1 \\ ctH(t-1), & t > 1 \end{cases}$$

It follows that $\mathcal{F}_f[c] \rightarrow \mathcal{F}_f[1]$ as $c \rightarrow 1$ in the mean-square sense, but not in the sense of $C^0(\mathbf{R}, \mathbf{R})$. Thus the condition $f \in H_{\text{loc}}^1(\mathbf{R}, L^2(\mathbf{R}))$ cannot be replaced by $f \in L_{\text{loc}}^2(\mathbf{R}, L^2(\mathbf{R}))$ in theorem 2, part (i).

To illuminate the connection between right-hand side regularity and regularity of \mathcal{F} (theorem 2, part (ii)), select $w \in L_{\text{loc}}^2(\mathbf{R})$ and multiply by the characteristic function of an interval $[a, b]$, $0 < a < b$, to generate a right-hand side:

$$g(t, x) = w(t)\mathbf{1}_{[a,b]}.$$

A brief computation shows that

$$\mathcal{F}_g[c](t) = w * \mathbf{1}_{\left[\frac{a}{c}, \frac{b}{c}\right]}(t).$$

The Lebesgue differentiation theorem implies that for any $T > 0$, $\mathcal{F}_g : \mathbf{R}_+ \rightarrow C^0([0, T], \mathbf{R})$ is continuous, but not differentiable: indeed, if w is discontinuous, then the Newton quotient converges point-wise a.e. to a discontinuous limit.

If $w \in H^1_{loc}(\mathbf{R})$ is causal, then \mathcal{F}_g is well defined as a map $\mathbf{R}_+ \rightarrow C^1(\mathbf{R}, \mathbf{R})$. The formal computation of the derivative,

$$(D\mathcal{F}_g[c]\delta c)(t) = \frac{\delta c}{c^2} \left(bw \left(t - \frac{b}{c} \right) - aw \left(t - \frac{a}{c} \right) \right),$$

actually expresses the derivative of $\mathcal{F}_g : \mathbf{R}_+ \rightarrow C^0(\mathbf{R}, \mathbf{R})$, but $\mathcal{F}_g : \mathbf{R}_+ \rightarrow C^1(\mathbf{R}, \mathbf{R})$ is not differentiable—if it were, the value of the derivative would necessarily be the same distribution as with respect to the weaker metric, but under the assumption on w one can only conclude that $D\mathcal{F}_g[c]\delta c \in C^0(\mathbf{R}, \mathbf{R})$.

In general, if $w \in H^s_{loc}(\mathbf{R})$ is causal, then \mathcal{F}_g is well defined and continuous $\mathbf{R}_+ \rightarrow C^s(\mathbf{R}, \mathbf{R})$ but differentiable only as a map $\mathbf{R}_+ \rightarrow C^{s-1}(\mathbf{R}, \mathbf{R})$. This example exhibits the property ‘differentiable with loss of derivative’ in the second sense mentioned after the statement of theorem 2.

This example does not illustrate first type of derivative loss described in the theorem, that is, the further drop of regularity in t required in general to define the trace. The reason for this is related to the existence of traces even for merely square-integrable right-hand sides, flowing in turn from the propagation of singularity principle for the system (6) (and indeed solution via the method of characteristics, an even more special property). Propagation of singularities along bicharacteristics holds for symmetric or strictly hyperbolic systems with smooth coefficients (see for example Taylor (1981)), and to some limited extent for systems with less regular coefficients (Beals and Reed 1982, 1984, Symes 1986, Lewis and Symes 1991, Bao and Symes 1996). Anomalous trace regularity follows for such systems: the system (6) is a particularly simple example. Some other consequences of this phenomenon will be mentioned in the Discussion section.

3. Example: acoustics

Linear acoustics provides an important example of the theoretical framework described in the last section. Acoustic wave propagation does not include the memory effect modeled by the convolutional term in (1), but illustrates several other features of the class of problems studied in this paper.

For this example, $d = 3$, and we suppose that the domain $\Omega \subset \mathbf{R}^3$ is smoothly bounded (the smoothness requirement can be relaxed to some extent, as discussed briefly in appendix A). The momentum balance and constitutive laws of linear acoustics relate the excess pressure $p(t, \mathbf{x})$ and velocity fluctuations $\mathbf{v}(t, \mathbf{x}) = (v_1(\mathbf{x}, t), v_2(\mathbf{x}, t), v_3(\mathbf{x}, t))^T$, $\mathbf{x} \in \mathbf{R}^3$, to mass density $\rho(\mathbf{x})$, bulk modulus $\kappa(\mathbf{x})$, constitutive law defect $g(t, \mathbf{x})$, and body force density $\mathbf{f}(t, \mathbf{x})$ by

$$\begin{aligned} \rho \frac{\partial \mathbf{v}}{\partial t} &= -\nabla p + \mathbf{f}, \\ \frac{1}{\kappa} \frac{\partial p}{\partial t} &= -\nabla \cdot \mathbf{v} + g. \end{aligned} \tag{7}$$

In this example, $k = 4$, and

$$au = \text{diag} \left(\frac{1}{\kappa}, \rho, \rho, \rho \right) u, u = \begin{pmatrix} p \\ v_1 \\ v_2 \\ v_3 \end{pmatrix} \in L^2(\Omega)^4$$

Densities and bulk moduli of sedimentary rocks in the upper crust, for example, lie between definite bounds. To express these natural bound conditions in a sufficiently flexible manner for present purposes, choose $0 < C_* \leq C^*$, and define

$$M_{ac}(C_*, C^*) = \{(\kappa, \rho) \in L^\infty(\Omega)^2 : C_* \leq 1/\kappa, \rho \leq C^* \text{ a.e. in } \Omega\}.$$

The notation is chosen so that if $(\kappa, \rho) \in M_{ac}(C_*, C^*)$, the matrix a satisfies the bounds (3).

Remark. It is assumed that the fields appearing in equations (7), particularly κ and ρ , have been rendered dimensionless by appropriate scaling. Similar assumptions will be made for other physical systems mentioned in this paper, without comment.

The right-hand side of (7) involves the matrix partial differential operator

$$p(\nabla) = - \begin{pmatrix} 0 & \frac{\partial}{\partial x_1} & \frac{\partial}{\partial x_2} & \frac{\partial}{\partial x_3} \\ \frac{\partial}{\partial x_1} & 0 & 0 & 0 \\ \frac{\partial}{\partial x_2} & 0 & 0 & 0 \\ \frac{\partial}{\partial x_3} & 0 & 0 & 0 \end{pmatrix}. \quad (8)$$

This matrix operator defines a skew-adjoint operator with domain $V = H_0^1(\Omega) \times H_{\text{div}}^1(\Omega) \subset L^2(\Omega)^4$. ($H_{\text{div}}^1(\Omega)$ consists of $\mathbf{v} \in L^2(\Omega)^3$ for which the distribution $\nabla \cdot \mathbf{v} \in L^2(\Omega)^3$. For the reader's convenience, we present a proof that $p(\nabla) : V \rightarrow L^2(\Omega)^4$ is skew-adjoint, in appendix A.)

Finally, the source vector f is defined by

$$f(t) = (g(t, \cdot), \mathbf{f}(t, \cdot))^T \in L^2(\Omega)^4,$$

With these conventions, the acoustics system (7) is symmetric hyperbolic and satisfies the conditions of theorem 1.

Note that if κ and/or ρ are discontinuous, then the form of the equations (7) immediately implies that no solutions of class C^1 may exist, even if the right-hand side f (that is, the body force density \mathbf{f}) is smooth. Physically reasonable fluid configurations thus exist for which solutions in the classical sense cannot be defined, for example piecewise homogeneous mixtures with jump discontinuities of density and/or bulk modulus across smooth interfaces. However, theorem 1 provides a solution with physical sense, as the strain energy

$$E(t) = \frac{1}{2} \int_{\Omega} \left[\rho \mathbf{v} \cdot \mathbf{v} + \frac{p^2}{\kappa} \right] (\cdot, t) = \frac{1}{2} \langle u(t), au(t) \rangle_{L^2(\Omega)^4}$$

is well defined for any such solution, and even of class C^1 in t . The equivalence of the quadratic form defining E and the (square of the) norm in $L^2(\Omega)^4$ follows from the assumption that $(\kappa, \rho) \in M_{ac}(C_*, C^*)$, as already noted.

Theorem 1 gives $u \in C^0(\mathbf{R}, V)$, provided that $f \in H_{\text{loc}}^1(\mathbf{R}, L^2(\Omega)^4)$. To obtain differentiable dependence on the coefficients, it is necessary to sacrifice one more time derivative (theorem 2). This sacrifice seems minor, as modeled energy sources (represented by the right-hand side in (1)) are in many cases quite smooth in time, reflecting the finite bandwidth of energy generation and recording equipment and the loss of high frequencies to dissipative mechanisms during propagation. Membership in $V = H_0^1(\Omega) \times H_{\text{div}}^1(\Omega)$ implies that some (but not all!) traces of u are well defined. In the notation of theorem 2, given a smoothly embedded hypersurface Γ , choose $l = 1$ and $m = m_D$ or $m = m_N$, where

$$m_D(\mathbf{x}) = [1, 0, 0, 0], \quad m_N(\mathbf{x}) = [0, n_1(\mathbf{x}), n_2(\mathbf{x}), n_3(\mathbf{x})], \quad \mathbf{x} \in \Gamma,$$

and $\mathbf{n} = (n_1, n_2, n_3)^T$ is a smooth normal field on Γ . Both $S[m_D]$ and $S[m_N]$ extend continuously to V . Note that traces of tangential components of \mathbf{v} do not so extend. The end result is that

both \mathcal{F}_{f,m_D} (pressure data) or \mathcal{F}_{f,m_N} (normal velocity data) are well defined and differentiable as functions of the coefficients, of class C^{s-1} if $f \in H_{loc}^s(\mathbf{R}, L^2(\Omega)^4)$ (theorem 2). The domain of either version of \mathcal{F} in this case is the open set $M_{ac} \subset L^\infty(\Omega)^2$, defined by

$$M_{ac} = \bigcup \{M_{ac}(C_*, C^*) : 0 < C_* \leq C^*\}.$$

The relation (5) in theorem 4 states that τ is larger than the largest generalized eigenvalue of the matrix $\sum p_i \xi_i$ with ‘mass’ matrix a , at a.e. $\mathbf{x} \in \Omega$. There are two nonzero generalized eigenvalues for each $\mathbf{x} \in \Omega$, namely $\pm \sqrt{\kappa(\mathbf{x})/\rho(\mathbf{x})}$, as is easily verified. Thus under the conditions of theorem 4, the solution u vanishes at in ω up to time T if at every intermediate time t , the support of the right-hand side lies at least distance $\tau(T - t)$ away from ω . Since $\tau \geq \sqrt{\kappa(\mathbf{x})/\rho(\mathbf{x})}$ for a.e. $\mathbf{x} \in \Omega$, this conclusion is just what one would expect: waves move with speed at most

$$\|\sqrt{\kappa/\rho}\|_{L^\infty(\Omega)}.$$

4. Example: viscoelasticity

Viscoelasticity provides an example of the symmetric hyperbolic dynamics with dissipation and frequency dispersion of wave energy, unlike acoustics.

$$\begin{aligned} \rho \frac{\partial \mathbf{v}}{\partial t} &= \nabla \cdot \sigma + \mathbf{f}, \\ S * \frac{\partial \sigma}{\partial t} &= \frac{1}{2} (\nabla \mathbf{v} + \nabla \mathbf{v}^T). \end{aligned} \tag{9}$$

in which \mathbf{v} is the particle velocity field, σ the stress tensor, \mathbf{f} a body force density, ρ the mass density, and S the inverse Hooke operator, or compliance tensor (Christensen 1983, Pipkin 1986). Viscoelasticity generalizes elasticity in that the inverse Hooke operator is a causal convolution operator (in time), rather than a temporally local multiplication operator. Of course, elasticity is a special case of viscoelasticity, in which S is proportional to $\delta(t)$.

We will assume instantaneous elastic response: a nonzero strain rate arises immediately from a stress impulse. Under this assumption, the kernel S can be decomposed as

$$S(t) = S^e \delta(t) + s(t),$$

in which S^e is the elastic inverse Hooke tensor (inverse of the unrelaxed modulus), and s is a causal kernel. Both the elastic kernel S^e and the memory kernel $s(t)$ act as spatially-variable, symmetry-preserving linear operators on symmetric tensor fields. The conventional representation of such things by 4-index tensors,

$$S^e = (S_{ijkl}^e)_{i,j,k,l=1}^3, \quad s = (s_{ijkl})_{i,j,k,l=1}^3,$$

thus entail the symmetries

$$S_{ijkl}^e = S_{jikl}^e = S_{ijlk}^e = S_{klij}^e, \quad i, j, k, l = 1, 2, 3, \tag{10}$$

and similarly for s .

To avoid technical complications, assume that the viscoelastic material occupies all of \mathbf{R}^3 . We require that, for some $0 < g_* \leq g^*$,

- (1) S^e is elliptic: for any symmetric $\sigma \in \mathbf{R}^{3 \times 3}$,

$$g_* \|\sigma\| \leq \|S^e(\mathbf{x})\sigma\| \leq g^* \|\sigma\|, \quad \mathbf{x} \in \mathbf{R}^3; \tag{11}$$

- (2) $S^e \in L^\infty(\mathbf{R}^3, \mathcal{B}(\mathbf{R}^{3 \times 3}_{\text{symm}}))$;
- (3) $s \in W^{1,1}(\mathbf{R}, L^\infty(\mathbf{R}^3, \mathcal{B}(\mathbf{R}^{3 \times 3}_{\text{symm}})))$.

For the ‘state space’ H of the viscoelastic system we choose $H = L^2(\mathbf{R}^3, \mathbf{R}^9) \equiv L^2(\mathbf{R}^3, \mathbf{R}_{\text{symm}}^{3 \times 3}) \times L^2(\mathbf{R}^3, \mathbf{R}^3)$. The inner product in H is defined by

$$\langle u_1, u_2 \rangle = \int_{\mathbf{R}^3} \text{tr} \sigma_1^T \sigma_2 + \mathbf{v}_1^T \mathbf{v}_2, u = \begin{pmatrix} \sigma \\ \mathbf{v} \end{pmatrix}.$$

The assumptions 1–3 above and the symmetries (10) imply that

$$Au = \begin{pmatrix} S^e \sigma \\ \rho \mathbf{v} \end{pmatrix}, u = \begin{pmatrix} \sigma \\ \mathbf{v} \end{pmatrix} \in H$$

defines a bounded, self-adjoint positive-definite operator $A \in \mathcal{B}(H)$.

Define the differential operator $p(\nabla) : C_0^\infty(\mathbf{R}^3, \mathbf{R}^9) \rightarrow C_0^\infty(\mathbf{R}^3, \mathbf{R}^9)$ by

$$p(\nabla)u = - \begin{pmatrix} \frac{1}{2}(\nabla \mathbf{v} + \nabla \mathbf{v}^T) \\ \nabla \cdot \sigma \end{pmatrix}, u = \begin{pmatrix} \sigma \\ \mathbf{v} \end{pmatrix} \in C_0^\infty(\mathbf{R}^3, \mathbf{R}^9) \equiv C_0^\infty(\mathbf{R}^3, \mathbf{R}_{\text{symm}}^{3 \times 3}) \times C_0^\infty(\mathbf{R}^3, \mathbf{R}^3).$$

$p(\nabla)$ is antisymmetric and densely defined in H . Define

- $H_{\text{div}}^1(\Omega, \mathbf{R}_{\text{symm}}^{3 \times 3})$ to be the subspace of $L^2(\Omega, \mathbf{R}_{\text{symm}}^{3 \times 3})$ consisting of square-integrable symmetric matrix valued functions, each column of which has a square-integrable divergence;
- $H_{\text{free}}^1(\Omega)$ to be the subspace of vectors $\sigma \in H_{\text{div}}^1(\Omega, \mathbf{R}_{\text{symm}}^{3 \times 3})$ satisfying

$$\sigma(\mathbf{x}) \cdot \mathbf{n} = 0, \text{ a.e. } \mathbf{x} \in \partial\Omega;$$
- $V = H_{\text{free}}^1(\Omega) \times H^1(\Omega)^3$.

An argument similar to that explained in appendix A shows that $p(\nabla)$ has a self-adjoint extension $P : V \rightarrow H$, and that the natural norm on V is equivalent to the graph norm of P .

Let

$$b = \lim_{t \rightarrow 0^+} s(t, \cdot) \in L^\infty(\mathbf{R}^3, \mathcal{B}(\mathbf{R}_{\text{symm}}^{3 \times 3})),$$

and

$$q = \lim_{t \rightarrow 0^+} \mathbf{1}_{[t, \infty)} \frac{\partial s}{\partial t} \in L^1(\mathbf{R}, L^\infty(\mathbf{R}^3, \mathcal{B}(\mathbf{R}_{\text{symm}}^{3 \times 3}))).$$

Then

$$s * \frac{\partial \sigma}{\partial t} = b\sigma + q * \sigma.$$

Define $B \in \mathcal{B}(H)$ and $Q \in L^1(\mathbf{R}, \mathcal{B}(H))$ by

$$Bu = \begin{pmatrix} b\sigma \\ 0 \end{pmatrix}, Q(t)u = \begin{pmatrix} q(t)\sigma \\ 0 \end{pmatrix}, u = \begin{pmatrix} \sigma \\ \mathbf{v} \end{pmatrix}$$

Finally, define $f \in L_{\text{loc}}^2(\mathbf{R}, H)$ by $f = (0, \mathbf{f})^T$.

With these definitions, the system (9) is formally equivalent to the evolution problem (13). The theory developed here thus assures the existence of weak solutions of (9), in material models including discontinuities of densities and/or elastic moduli and/or relaxation moduli.

An immediate consequence of theorem 4 is

Corollary 1. Denote by c_p the maximum quasi- p -wave velocity of the viscoelastic system (9), defined as

$$c_p = \text{ess sup} \{ \lambda_{\max}(S^e(\mathbf{x})[\xi \xi^T]) / \rho(\mathbf{x}) : \mathbf{x}, \xi \in \mathbf{R}^3, \xi^T \xi = 1 \}.$$

Suppose that (\mathbf{x}_0, t_0) satisfies

$$|\mathbf{x} - \mathbf{x}_0| > c_p(t_0 - t)$$

for every $(\mathbf{x}, t) \in \text{supp } \mathbf{f}$. Then the causal weak solution (σ, \mathbf{v}) of (9) vanishes in a neighborhood of (\mathbf{x}_0, t_0) .

5. A class of abstract first order evolution equations and their properties

Much of the theory developed in this paper applies equally well to a type of abstract evolution equation, of which (1) and its specializations are concrete realizations. Both to clarify the logical structure of the arguments, and because these systems have intrinsic interest, we formulate the basic facts about symmetric hyperbolic systems in terms of this abstraction.

Let H be a separable real Hilbert space, with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. We will denote by $V \subset H$ a dense subspace, itself a separable Hilbert space with inner product $\langle \cdot, \cdot \rangle_V$ and norm $\| \cdot \|_V$, defining a stronger topology on V than that induced by H . $\mathcal{B}(H)$ is the Banach space of bounded linear operators on H , with the operator (uniform) norm. $\mathcal{B}_{\text{symm}}(H)$ is the subspace of bounded self-adjoint operators, and $\mathcal{B}_{\text{symm}}^+(H)$ is the cone of bounded self-adjoint coercive operators. We suppose that

- $A \in \mathcal{B}_{\text{symm}}^+(H)$, $0 < C_* \leq C^*$ so that $C_*I \leq A \leq C^*I$.
- $B \in \mathcal{B}(H)$;
- $Q \in L^1(\mathbf{R}, \mathcal{B}_{\text{symm}}(H)) \cap C^0(\mathbf{R}_+, \mathcal{B}(H))$, and is *causal*, that is, $Q(t) = 0$ for $t < 0$;
- P is a skew-adjoint operator with domain V , for which the graph norm is equivalent to $\| \cdot \|_V$.

The kernel Q defines a continuous linear operator $R : L_+^2(\mathbf{R}, H) \rightarrow L_+^2(\mathbf{R}, H)$ on the space

$$L_+^2(\mathbf{R}, H) = \{v \in L_{\text{loc}}^2(\mathbf{R}, H) : \text{for each } T \in \mathbf{R}, v \in L^2((-\infty, T], H)\}$$

with the natural countably normed topology, by

$$R[u](t) = \int Q(t-s)u(s) ds. \quad (12)$$

R is well defined since Q is causal. Note that if $u \in L_{\text{loc}}^2(\mathbf{R}, H)$ is *causal*, that is, $\text{supp } u \subset [T, \infty)$ for some $T \in \mathbf{R}$, then $u \in L_+^2(\mathbf{R}, H)$, $R[u]$ is causal also and $\text{supp } R[u] \subset [T, \infty)$. The formal (distribution) adjoint

$$R^*[u](t) = \int Q(s-t)u(s) ds$$

satisfies a similar condition: if $\text{supp } u \subset (-\infty, T]$ for some $T \in \mathbf{R}$, then $\text{supp } R^*[u] \subset (-\infty, T]$.

The components described above combine to yield the *formal evolution problem*: find an H -valued function of t , say u , which solves, in a suitable sense,

$$Au' + Pu + Bu + R[u] = f, \quad (13)$$

in which the right-hand side f is also an H -valued function of t .

We follow (Lions 1971, Lions and Magenes 1972, Stolk 2000) in defining *weak solutions* $u \in L_{\text{loc}}^2(\mathbf{R}, H)$ of the formal evolution problem (13) by integration against smooth test functions. Inspection of (13) suggests that a similar constraint must also be placed on the right-hand side: henceforth, we assume $f \in L_{\text{loc}}^2(\mathbf{R}, H)$. Because the operator kernel Q may have unbounded support, we must constrain the growth of candidate members of $L_{\text{loc}}^2(\mathbf{R}, H)$ on the negative half-axis. Accordingly, a *weak solution* of the formal evolution problem (13) is a member of $u \in L_{\text{loc}}^2(\mathbf{R}, H)$ satisfying

- (1) $u \in L_+^2(\mathbf{R}, H)$;
- (2)

$$\int \langle u(t), (A\phi' + P\phi - B^*\phi - R^*[\phi])(t) \rangle dt = - \int \langle f(t), \phi(t) \rangle dt; \quad (14)$$

for all $\phi \in C_0^\infty(\mathbf{R}, V)$.

Note that since $R^*[\phi]$ is supported in the half-axis $(-\infty, \sup \text{supp } \phi]$, and is square-integrable, assumption 1 implies that the last term on the left-hand side of (14) is well defined.

We will reserve the term *strong solution* to designate a function $u \in C^1(\mathbf{R}, H) \cap C^0(\mathbf{R}, V) \cap L^2_+(\mathbf{R}, H)$ which solves (13) pointwise. Clearly a strong solution is a weak solution. Existence of a strong solution in this sense also implies that $f \in C^0(\mathbf{R}, H)$.

Because of the causal assumption on the convolution kernel Q , existence of a causal weak solution vanishing for $t < T_0$ implies that the right-hand side f must also be causal, in fact vanish for $t < T_0$.

Concerning the fundamental questions of uniqueness and existence, we prove

Theorem 5. *Assume that H, V, A, B, P and Q have the properties listed above, and that $f \in L^2_{\text{loc}}(\mathbf{R}, H)$ is causal. Choose T_0 so that $\text{supp } f \subset [T_0, \infty)$. Then there exists a unique causal weak solution $u \in C^0(\mathbf{R}, H)$ of (13) with $\text{supp } u \subset [T_0, \infty)$. For every $T \geq T_0$, there exists $C_{T_0, T} \geq 0$ depending on $T_0, T, C_*, C^*, \|B\|_{\mathcal{B}(H)}$, and $\|Q\|_{L^1(\mathbf{R}, \mathcal{B}(H))}$, so that for any $t \leq T$,*

$$\|u(t)\|_H^2 \leq C_{T_0, T} \int_{-\infty}^t ds \|f(s)\|^2. \tag{15}$$

Regularity in time of the right-hand side implies that the weak solution is strong:

Theorem 6. *In addition to the assumptions of theorem 5, suppose that $f \in L^2_+(\mathbf{R}, H) \cap H^k_{\text{loc}}(\mathbf{R}, H)$, $k \geq 1$. Then the causal weak solution u of (14) satisfies $u \in C^k(\mathbf{R}, H) \cap C^{k-1}(\mathbf{R}, V)$. For every $T \geq T_0$, there exists $C_{T_0, T} \geq 0$ depending on $T_0, T, C_*, C^*, \|B\|_{\mathcal{B}(H)}$, and $\|Q\|_{L^1(\mathbf{R}, \mathcal{B}(H))}$, so that for any $t \leq T$,*

$$\|u^{(j)}(t)\|_V^2 \leq C_{T_0, T} \sum_{l=0}^{j+1} \int_{-\infty}^t ds \|f^{(l)}(s)\|^2, \quad j = 0, \dots, k-1. \tag{16}$$

in which $u^{(j)}$ denotes the j th time derivative of u . In particular, $u \in C^1(\mathbf{R}, H) \cap C^0(\mathbf{R}, V)$ is a strong solution of (13). For each $j = 0, \dots, k$, $u^{(j)}$ is the weak solution of (13) with right-hand side $f^{(j)}$.

Regularity of weak solutions u as functions of the coefficients A, B, Q is expressed through estimates which are uniform over certain sets of problems (that is, certain sets of coefficients), which are important enough to merit a definition: given $0 < C_* \leq C^*, 0 \leq C_B, 0 \leq C_Q$, define

$$\mathcal{P}(C_*, C^*, C_B, C_Q) = \{(A, B, Q) \in \mathcal{B}_{\text{symm}}^+(H) \times \mathcal{B} \times [L^1(\mathbf{R}, \mathcal{B}_{\text{symm}}(H)) \cap C^0(\mathbf{R}_+, \mathcal{B}(H))]\};$$

$$C_* I \leq A \leq C^* I, \|B\|_{\mathcal{B}(H)} \leq C_B, \|Q\|_{L^1(\mathbf{R}, H)} + \|Q\|_{L^\infty(\mathbf{R}, H)} \leq C_Q;$$

$$\mathcal{P} = \cup \{\mathcal{P}(C_*, C^*, C_B, C_Q) : 0 < C_* \leq C^*, 0 \leq C_B, 0 \leq C_Q\}. \tag{17}$$

Strong convergence of the coefficient operators, in an appropriate sense, leads to uniform convergence of the corresponding weak solutions on compact sets in \mathbf{R} . For simplicity, and because all of the examples we have in mind satisfy this restriction, we assume that all of the systems appearing the formulation of this continuity result share the same ‘spatial’ operator P . It is possible to weaken this assumption, that is, to approximate P as well; we leave the formulation and proof of this (slightly) stronger result to the reader.

Theorem 7. *Suppose that H, V are as described, $P : V \rightarrow H$ is skew-adjoint, $0 < C_* \leq C^*, 0 \leq C_B, 0 \leq C_Q$, $\{(A_m, B_m, Q_m) : m \in \mathbf{N}\} \subset \mathcal{P}(C_*, C^*, C_B, C_Q)$, $(A, B, Q) \in \mathcal{P}(C_*, C^*, C_B, C_Q)$, and*

- (1) $\lim_{m \rightarrow \infty} \|(A_m - A)w\| \rightarrow 0$ for all $w \in H$;
- (2) $\lim_{m \rightarrow \infty} \|(B_m - B)w\| \rightarrow 0$ for all $w \in H$;
- (3) the convolution operators R_m, R with kernels Q_m, Q satisfy $\lim_{m \rightarrow \infty} \|R_m[w] - R[w]\|_{L^2((-\infty, T])} \rightarrow 0$ for any $T \in \mathbf{R}$, all $w \in L^2_+(\mathbf{R}, H)$.

Let u_m , respectively u , be causal weak solutions of the differential equation (13) with coefficients (A_m, P, B_m, Q_m) , respectively (A, P, B, Q) . and (common) causal right-hand side $f \in L^2_{loc}(\mathbf{R}, H)$. Then for any choice of $T_0 \leq T \in \mathbf{R}$,

$$\lim_{m \rightarrow \infty} \|u_m - u\|_{L^\infty([T_0, T], H)} = 0.$$

With additional regularity of the right-hand side, solutions of (13) have directional (Gâteaux) derivatives as functions of the coefficient operators:

Theorem 8. Suppose that H, V are as described, $P : V \rightarrow H$ is skew-adjoint, $(A, B, Q) \in \mathcal{P}$, and $f \in H^1(\mathbf{R}, H)$ is causal. Denote by $u \in C^1(\mathbf{R}, H) \cap C^0(\mathbf{R}, V)$ the causal strong solution of (13) with these choices of coefficients and right-hand side. Assume that $\delta A, \delta B \in \mathcal{B}(H)$, δA is self-adjoint, and $\delta Q \in L^1(\mathbf{R}, \mathcal{B}_{\text{symm}}(H)) \cap C^0(\mathbf{R}_+, \mathcal{B}(H))$, $\delta Q = 0$ for $t < 0$. Define for $h \in \mathbf{R}$

$$A_h = A + h\delta A, B_h = B + h\delta B, Q_h = Q + h\delta Q.$$

For sufficiently small h , $(A_h, B_h, Q_h) \in \mathcal{P}$, so that the problem (13) with coefficients A_h, P, B_h, Q_h and right-hand side f has a unique (strong) solution $u_h \in C^1(\mathbf{R}, H) \cap C^0(\mathbf{R}, V)$. Denote by $\delta u \in C^0(\mathbf{R}, H)$ the weak solution of the formal evolution problem

$$A\delta u' + P\delta u + B\delta u + R[\delta u] = -\delta A u' - \delta B u - \delta R[u], \tag{18}$$

in which R (δR) is the convolution operator with kernel Q (δQ), as usual. Then for any choice of $T_0 \leq T \in \mathbf{R}$,

$$\lim_{h \rightarrow 0} \left\| \frac{u_h - u}{h} - \delta u \right\|_{L^\infty([T_0, T], H)} = 0. \tag{19}$$

With the correct choice of topology, the solution is (Fréchet) differentiable as a function of the coefficients, assuming as in theorem 8 that the right-hand side is somewhat regular in t . To express this fact in a form most useful for application to inverse problems, introduce another Hilbert space W (of ‘measurements’), and a ‘sampling’ map $S : V \rightarrow W$, assumed continuous. Of course S induces a continuous map $C^k(\mathbf{R}, V) \rightarrow C^k(\mathbf{R}, W)$ for any $k \in \mathbf{N}$; we abuse notation by writing S for this induced map also.

For causal $f \in H^1_{loc}(\mathbf{R}, H)$, define $\mathcal{F}_f : \mathcal{P} \mapsto C^1(\mathbf{R}, W)$ by

$$\mathcal{F}_f[A, B, Q] = Su, \tag{20}$$

in which $u \in C^1(\mathbf{R}, H) \cap C^0(\mathbf{R}, V)$ is the causal (strong) solution of (13) with coefficients A, P, B, Q and right-hand side f . It follows from theorem 6 that \mathcal{F}_f is well defined.

Theorem 9. Suppose that $f \in H^k_{loc}(\mathbf{R}, H)$ is causal, $k \geq 2$, and view \mathcal{P} as an open subset of the Banach space $\mathcal{M} = \mathcal{B}_{\text{symm}}(H) \times \mathcal{B}(H) \times [L^1(\mathbf{R}, \mathcal{B}_{\text{symm}}(H)) \cap C^0(\mathbf{R}_+, \mathcal{B}(H))]$, with norm

$$(A, B, Q) \mapsto \|(A, B, Q)\|_{\mathcal{M}} = \|A\|_{\mathcal{B}(H)} + \|B\|_{\mathcal{B}(H)} + \max(\|Q\|_{L^1(\mathbf{R}_+, \mathcal{B}(H))}, \|Q\|_{L^\infty(\mathbf{R}_+, \mathcal{B}(H))}).$$

Then for any $T_0 \leq T_1 \in \mathbf{R}$, $\mathcal{F}_f \in C^{k-1}(\mathcal{P}, C^0([T_0, T_1], W))$.

6. Discussion

The theory developed in this paper established a number of properties of the relation between coefficients and solutions of symmetric hyperbolic integro-differential systems, such as continuity, differentiability, finite domain of influence, and so on. These properties are assumed (usually without comment) in the applied literature on inverse problems for wave propagation in highly heterogeneous materials. The minimal hypotheses required by our theory allow for a degree of irregularity in the spatial dependence of coefficients which seems adequate to model the heterogeneity observed in real materials. The methods of proof date to the middle of the last century, yet so far as we are aware these results have not previously been explicitly formulated or collected in one place.

It remains to place our work in the context of related literature, and describe consequences for the solutions of inverse problems. This section begins with a review of related work on hyperbolic systems with non-smooth coefficients. Following this review, we discuss several by-products of the theory with implications for inverse problems in wave propagation: systems with operator coefficients, recently arisen as a key element of convexifying relaxations for seismic waveform inversion; computing descent directions for functions on non-reflexive Banach spaces; trace regularity under additional constraints and implications for the definition of ‘Dirichlet-to-Neumann’ maps; and various other possible extensions of the theory and open questions.

6.1. *Prior art*

We begin with a brief overview of prior work on the solution of (1). The well posedness of symmetric hyperbolic systems with regular (smooth) coefficients has been well-understood for decades (Courant and Hilbert 1962, Lax 2006). Some effort has also been devoted to describing solutions for first- or second-order hyperbolic systems with less-than-smooth (but still continuous) coefficients, mostly focused on the propagation of regularity along bicharacteristics, or the existence of traces (for example, Beals and Reed (1982), Symes (1983), Beals and Reed (1984), Lasiecka (1986), Lasiecka and Trigianni (1989), Bao and Symes (1991), ,1993), Smith (1998). Those works treating the continuity or differentiability of the solution as a function of the coefficients have mostly required more smoothness of the coefficients than is allowed in this work, or dealt only with one-dimensional problems (Bamberger *et al* 1979, Symes 1986, Lewis and Symes 1991, Bao and Symes 1996, Salo 2006, Stefanof and Uhlmann 2009). An exception is Fernandez-Berdaguer *et al* (1996), in which (results which imply) a special case of theorem 2 is established.

A large number of articles have appeared on abstract first- or second-order hyperbolic equations and related inverse problems, for example Lavrentiev *et al* (1986), Choulli (1991), Lorenzi and Ramm (2001), Ramm and Koshkin (2001), Awawdeh (2010), Orlovsky *et al* (2010), mostly using the theory of strongly continuous semigroups (or cosine operators, in the second order case) to obtain a hold on well posedness. Of these, the closest in spirit to this work is that of Orlovsky *et al* (2010), which treats second order systems via the cosine operator approach but observes that smoothness *in time* of data implies that weak solutions are strong solutions, analogous to a crucial intermediate result in our work. Choulli (1991) established that a first-order abstract integro-differential initial value problem, resembling those of our abstract framework to be detailed in the next section, is well posed, and also proved uniqueness for the solution of a related inverse problem. The form of the integral term is actually a generalization of our convolutional memory operator, and the requirement on the ‘spatial’ operator is also a generalization (generates a strongly continuous semigroup—in

the problem considered here, this operator is skew-adjoint, and generates a unitary group). However Choulli formulates an inverse problem which involves only determining lower-order terms, rather than the principal part, the inverse problem data does not involve time-like traces, and Choulli does not investigate the regularity of any analogue of \mathcal{F} .

The immediate predecessor of our work is chapter 2 of the second-named author's PhD thesis (Stolk 2000), which treated second-order hyperbolic systems with nonsmooth coefficients, including the displacement formulation of elastodynamics, beginning with the techniques of Lions and his collaborators (Lions 1971, Lions and Magenes 1972) but going beyond well posedness to study the dependence of solutions on coefficients. The results on strong convergence and Gâteaux differentiability (theorems 7 and 8) are direct translations of arguments from this source, as is the observation that convergence in measure of L^∞ multipliers implies strong convergence of the corresponding multiplication operators on L^2 (lemma 4).

6.2. Descent directions

We have shown how to formulate the 'forward map' ($\mathcal{F}_{f,m}$ in the notation of theorem 2) as a continuously differentiable map on an open subset of a suitable Banach space, with values in a (subset of a) suitable Hilbert space. Thus least squares objective functions for such problems are continuously differentiable in a well-defined sense. However, the domain metric specified by the theory is some version of L^∞ , in the case of concrete problems of form (1), or the operator norm in the case of abstract problems (theorem 9). Thus the ambient Banach space is in all cases nonreflexive, and does not have the properties required for a sensible definition of gradient, as explained for instance by Kaltenbacher *et al* (2008). Thus differentiability does not necessarily bring with it natural access to descent directions.

Additional information about the derivative is available, however, via the well-known adjoint state method. For brevity, we explain this construction in the case of the abstract evolution problem (13) in the differential case and without lower-order term, that is, $B, R = 0$, and we proceed formally. With notation as in the statement of theorem 9, we define

$$\mathcal{F}_{f,m} : \mathcal{B}_{\text{symm}}^+(H) \rightarrow L^2([0, T], W)$$

by

$$\mathcal{F}_{f,m}[A] = S[m]u,$$

where $u \in C^1(\mathbf{R}, H) \cap C^0(\mathbf{R}, V)$ is the solution of

$$Au' + Pu = f \in H_{\text{loc}}^k(\mathbf{R}, H), f(t), u(t) = 0, t < 0$$

guaranteed by the theory. If $k \geq 2$, then $\mathcal{F}_{f,m}$ is of class C^1 , and its derivative is given by

$$D\mathcal{F}_{f,m}[A]\delta A = S[m]\delta u,$$

where

$$A\delta u' + P\delta u = -\delta Au', \delta u(t) = 0, t < 0$$

Suppose that we are provided data $d \in L^2([0, T], W)$. The adjoint state method (Plessix 2006) represents the derivative of the least-squares function

$$J_{f,m}[A; d] = \frac{1}{2} \|\mathcal{F}_{f,m}[A] - d\|^2$$

as follows. Suppose that w is a strong solution of

$$A'w + Pw = S[m]^*(d - \mathcal{F}_{f,m}[A]); w(t) = 0, t > T. \quad (21)$$

Then for any $\delta A \in \mathcal{B}_{\text{symm}}(H)$,

$$DJ_{f,m}[A; d]\delta A = \int_0^T dt \langle \delta A u'(t), w(t) \rangle. \quad (22)$$

It is possible to extract a *gradient*, that is, a direction of fastest ascent for $J_{f,m}$, from this representation of the derivative. Write (for $u, v \in H$) $u \otimes v$ for the rank 1 member of $\mathcal{B}(H)$ defined by $u \otimes v(w) = \langle v, w \rangle u$. Such rank-1 operators are very special instances of operators of *trace class*. We refer to Conway (1990), pp 267–8 and 275 for the properties of this subspace of $\mathcal{B}(H)$ cited here, in particular these: the product of a bounded operator and a trace-class operator is of trace class; a real-valued linear *trace function* tr is defined on the trace class, generalizing the trace of matrices. The adjoint-state expression (22) is equivalent to

$$DJ_{f,m}[A; d]\delta A = \text{tr} \left(\delta A \int_0^T dt u'(t) \otimes w(t) \right), \quad (23)$$

well defined since the integral inside the parenthesis defines a trace-class operator. This trace of the product defines a duality pairing, expressing the trace class as the ‘pre-dual’ of $\mathcal{B}(H)$.

Assuming that f, d are sufficiently smooth in t , the integrand in (23) is well-approximated by quadrature:

$$DJ_{f,m}[A; d]\delta A = \text{tr} \left(\delta A \Delta t \sum_{n=0}^{[T/\Delta t]} u'(n\Delta t) \otimes w(n\Delta t) \right) \quad (24)$$

Linear functionals on $\mathcal{B}(H)$ of the form given by the right-hand side of (24) form precisely the subset of the dual $\mathcal{B}(H)^*$ consisting of functionals continuous in the weak operator topology. Since the unit ball of $\mathcal{B}(H)$ is compact with respect to weak convergence, the functional defined by the right-hand side of (24) has a maximizer over the ball, which is a candidate for an approximate gradient of $J_{f,m}$ at A .

This observation remains correct if A and δA are restricted to subspaces of $\mathcal{B}(H)$, for example L^∞ matrix-valued functions as in the definition of symmetric hyperbolic systems (1).

The program outlined in the preceding paragraphs leaves a number of details to be filled in. For example, actually *finding* the maximizer of the linear functional in (24) is in some sense a finite dimensional problem, but not a particularly simple one. Also, the precise relation between right-hand sides of (23) and (24) remains to be established. Finally, the adjoint state evolution problem (21) involves a right-hand side $(S[m]^*(d - \mathcal{F}_{m,f}[A])) \in L_{\text{loc}}^2(\mathbf{R}, V^*)$ outside of the class for which solutions, strong or weak, have been shown to exist. We will address this latter problem in the next subsection.

We also note that the gradient, defined as a direction of fastest ascent, is not smooth, or even necessarily a function, even under hypotheses which guarantee that $J_{f,m}$ has several derivatives. This observation leads to various developments in convex optimization, beyond the scope of this paper. We note merely that Bamberger *et al* (1979), a landmark early paper on inverse problems in wave propagation, addressed this point explicitly.

6.3. Sharpness of trace regularity and source definition

Theorem 2 gives sufficient conditions for the trace of a solution of the symmetric hyperbolic system to be well defined and depend smoothly on the coefficients and right-hand side in (1). However these conditions are only sufficient, not necessary. Under some conditions, the trace of even a weak solution may be well defined and depend continuously on the coefficients and right-hand side. We saw an example of such anomalously well-behaved traces already in the discussion of the advection equation (6) in section 2, however the phenomenon is much more general.

For example, suppose that the coefficients of (1) are smooth in a neighborhood ω of the boundary $\partial\Omega$ and containing the measurement surface Γ , that the right-hand side is supported outside of this neighborhood, and that no rays of geometric optics originating outside this neighborhood are tangent to the boundary. Then microlocal propagation of regularity shows that the map $f \in H_0^1([0, T], L^2(\Omega \setminus \omega)^k) \rightarrow L^2([0, T], L^2(\Gamma)^l)$ extends continuously to a map $L^2([0, T], L^2(\Omega \setminus \omega)^k) \rightarrow L^2([0, T], L^2(\Gamma)^l)$. See for example Symes (1983), Lasić (1986).

By duality, sources supported on Γ also give rise to well-defined solutions with estimates similar to those proven in the body of this paper. Combining these two observations leads to definition of variants of the *Dirichlet-to-Neumann map* (Sylvester and Uhlmann 1990) which forms the natural abstraction of ‘data’ for many inverse problems. These mappings, from suitable (source) data supported on Γ to other data supported on Γ (or on another similar surface), thus inherit a definition and regularity properties for minimally regular coefficient classes as described above.

6.4. Nonphysical extension via operator coefficients

Perhaps surprisingly, the results on systems with operator coefficients (theorems 5–9) seem likely to be useful in themselves. This utility originates in the resistance of inverse problems for hyperbolic systems to the least-squares (or least-error) approaches that have successfully treated many other types of inverse problems in science and engineering. Natural objective functions change rapidly in some directions, not in others, and appear to possess many stationary points far from any useful model estimate, a nearly fatal feature for variants of Newton’s method, often the only feasible approach to model estimation (Gauthier *et al* 1986, Santosa and Symes 1989).

The last-named author has suggested that use of operator coefficients as an *extension* of normal continuum physics could conceivably convexify the normal least-squares objective (Symes 2008). Far from an arbitrary introduction of additional degrees of freedom, this reformulation of the inverse problem is closely related to well-tested ideas in seismic data processing, and has already a partial theoretical justification (Stolk *et al* 2009, Shen and Symes 2008). Very recently, implementations of extended least squares inversion using operator coefficients have produced very promising early results (Biondi and Almomin 2012) which tend to support the convexification hypothesis.

6.5. And so on . . .

The fundamental model problem (1) does not include static differential constraints, so the results established above do not apply directly to Maxwell’s equations, for example. We suspect that the theory can be extended to accommodate such problems.

Imposition of additional regularity requirements on coefficient matrices, beyond boundedness and measurability, leads to existence of well-behaved solutions with more regularity, and (by duality) less regularity as well. Stolk (2000) offers some results concerning second-order problems with additional regularity of both coefficients and data.

Both physical heuristics and numerical evidence for inverse problems in seismology (Virieux and Operto 2009) suggest strongly that the mapping \mathcal{F} is more linear when the right-hand side is smoother in time. Estimates similar to those developed in the proof of theorem 9 show that for right-hand sides of the form $f(\epsilon t)$, the Newton quotient remainder is $O(h\epsilon)$ relative to the directional derivative, for any $(\delta A, \delta B, \delta Q)$ not in the null space of $D\mathcal{F}$. This observation can be developed into a theoretical justification of the frequency continuation

strategy employed in every successful contemporary algorithm for least squares inversion of seismic data.

7. The energy inequality

Define the *energy* $E(t)$ of a weak solution u of (13) by

$$E(t) = \frac{1}{2} \langle u(t), Au(t) \rangle \quad (25)$$

It follows from the definition of weak solution that E is well defined almost everywhere, and locally integrable. Because A is positive-definite,

$$C_* \|u(t)\|^2 \leq E(t) \leq C^* \|u(t)\|^2 \quad (26)$$

hold for almost all $t \in \mathbf{R}$, for suitable $C^* \geq C_* > 0$.

Remark. In the linear acoustics and viscoelasticity examples described earlier in this paper, $E(t)$ is precisely the mechanical (kinetic plus potential) energy of the dynamical field u at time t .

In this and the following sections, we will use C to denote a generic nonnegative constant depending on C_* , C^* , $\|B\|_{\mathcal{B}(H)}$, and $\|Q\|_{L^1(\mathbf{R}, \mathcal{B}(H))}$, and possibly on other quantities as noted.

We shall repeatedly use the following property of weak solutions:

Proposition 1. *Suppose that $u \in L^2_{\text{loc}}(\mathbf{R}, H)$ is a weak solution of (13). Then for any $\eta \in C_0^\infty(\mathbf{R})$, the mollification $\eta * u$ is a strong solution of (13) with right-hand side $\eta * f$, and in fact*

$$\eta * u \in C^\infty(\mathbf{R}, V). \quad (27)$$

Remark. In applications such as the acoustic and viscoelastic systems described earlier, H and V are function spaces, and membership in V entails additional regularity beyond that required for membership in H . In such contexts, the content of this theorem is that smoothing in time also ‘smooths in space’, in the sense that the values of the smoothed weak solution are confined to the subspace V .

Proof. Choose the test function ϕ in (14) to have the special form $\phi(s) = \eta(t - s)w$, where $t \in \mathbf{R}$, $w \in V$, and $\eta \in C_0^\infty(\mathbf{R})$. Then

$$\begin{aligned} \langle \eta * u(t), Pw \rangle &= \left\langle \int \eta(t - s)u(s) \, ds, Pw \right\rangle \\ &= \int \langle u(s), P(\eta(t - s)w) \rangle \, ds \\ &= \int [\langle u(s), A\eta'(t - s)w + B^*\eta(t - s)w \\ &\quad + R^*[\eta(t - \cdot)w](s)] \, ds - \langle \eta * f(t), w \rangle \\ &= \langle A(\eta' * u)(t) + B(\eta * u)(t) + R[\eta * u](t) - (\eta * f)(t), w \rangle, \end{aligned} \quad (28)$$

where the third equality is simply a rearrangement of (14) with the special choice of test function $\phi(s) = \eta(t - s)w$ mentioned above, and the last holds amongst other reasons because R is also convolutional. The right-hand side of (28) is bounded by a w -independent multiple of $\|w\|_H$, therefore so is the left. Therefore $\eta * u$ takes values in the domain $\mathcal{D}(P^*)$ of

the adjoint P^* for any $\eta \in C_0^\infty(\mathbf{R})$. But P is skew-adjoint, so $\mathcal{D}(P^*) = \mathcal{D}(P) = V$. Therefore we can shift P to the left-hand side of the inner product on left-hand side of (28), to obtain

$$\langle -P(\eta * u)(t), w \rangle = \langle A(\eta' * u)(t) + B(\eta * u)(t) + R[\eta * u](t) - (\eta * f)(t), w \rangle. \quad (29)$$

As (29) holds for every $w \in V$ and $V \subset H$ is dense, it follows that

$$-P(\eta * u)(t) = A(\eta' * u)(t) + B(\eta * u)(t) + R[\eta * u](t) - (\eta * f)(t), \quad (30)$$

that is, (13) is satisfied pointwise in t .

Since the right-hand side of (30) is of class $C^\infty(\mathbf{R}, H)$, so is the left-hand side. Denote by $T_{\delta t} \in \mathcal{B}(L_{\text{loc}}^2(\mathbf{R}, H))$ the translation operator by δt , defined by $T_{\delta t}u(t) = u(t + \delta t)$. For each $k \in \mathbf{N}$, let

$$\Delta_{k, \delta t} = \delta t^{-k} \sum_{j=-m_k}^{m_k} a_{k,j} T_{\delta t}^j$$

be a finite difference operator for which $u \in C^0(\mathbf{R}, H)$ has k derivatives if and only if $\Delta_{i, \delta t}u$ converges pointwise in t for each $i = 1, \dots, k$ to $u^{(i)}$ as $\delta t \rightarrow 0$. From (30) it follows that for each $k \in \mathbf{N}$,

$$\begin{aligned} -\Delta_{k, \delta t}(P(\eta * u))(t) &= -P(\Delta_{k, \delta t}(\eta * u)(t)) \\ &= \Delta_{k, \delta t}(A(\eta' * u) + B(\eta * u) + R[\eta * u] - \eta * f)(t). \end{aligned} \quad (31)$$

Since the right-hand side of (31) converges for each t , so does the left-hand side, both versions, as $\delta t \rightarrow 0$. Thus $\Delta_{k, \delta t}(\eta * u)(t)$ converges in the graph norm of P , hence in the sense of $\|\cdot\|_V$, as $\delta t \rightarrow 0$ for each $t \in \mathbf{R}$, $k \in \mathbf{N}$. Thus $\eta * u \in C^\infty(\mathbf{R}, V)$. \square

Proposition 2. Let $u \in L_{\text{loc}}^2(\mathbf{R}, H)$ be a weak solution of (13), $E \in L_{\text{loc}}^1(\mathbf{R})$ its energy as defined in (25). Then

- after modification on a set of measure zero, E is continuous;
- if in addition f is causal, $\text{supp } f \subset [T_0, \infty)$, then for any $T \in \mathbf{R}$ there exists $C_{T_0, T} \geq 0$ so that for $t \in (-\infty, T]$,

$$E(t) \leq C_{T_0, T} \int_{-\infty}^t \|f\|^2. \quad (32)$$

Proof. Let $\{\eta_n \in C_c^\infty(\mathbf{R})\}$ be a sequence of mollifiers approximating the Dirac delta distribution: explicitly, $\eta_n(t) = n\eta(nt)$, where

$$\eta \in C_0^\infty(\mathbf{R}), \quad \eta \geq 0, \quad \int \eta(t) dt = 1, \quad \text{supp } \eta \subset [-1, 1]. \quad (33)$$

Define

$$E_n(t) = \frac{1}{2} \langle (\eta_n * u)(t), A(\eta_n * u)(t) \rangle. \quad (34)$$

Since $\eta_n * u \rightarrow u$ in $L_{\text{loc}}^2(\mathbf{R}, H)$, $E_n \rightarrow E$ in $L_{\text{loc}}^1(\mathbf{R})$. Thanks to proposition 1, $\eta_n * u \in C^\infty(\mathbf{R}, V)$ is a strong solution of (13) for each $n \in \mathbf{N}$, with f replaced by $\eta_n * f$.

For each n , E_n is smooth; differentiating E_n , obtain for any $s, t \in \mathbf{R}$

$$\begin{aligned} E_n(t) - E_n(s) &= \int_s^t \frac{dE_n}{ds}(\tau) d\tau \\ &= \int_s^t \langle (\eta_n * u)(\tau), A(\eta_n * u)'(\tau) \rangle d\tau \end{aligned}$$

$$\begin{aligned}
 &= \int_s^t [-\langle (\eta_n * u)(\tau), (P(\eta_n * u)(\tau) + B(\eta_n * u)(\tau) + R[\eta_n * u])(\tau) \rangle \\
 &\quad + \langle (\eta_n * f)(\tau), (\eta_n * u)(\tau) \rangle] d\tau \\
 &= \int_s^t [-\langle (\eta_n * u)(\tau), (B(\eta_n * u)(\tau) + R[\eta_n * u])(\tau) \rangle \\
 &\quad + \langle (\eta_n * f)(\tau), (\eta_n * u)(\tau) \rangle] d\tau.
 \end{aligned}
 \tag{35}$$

thanks to the skew-symmetry of P .

Since convolution with η commutes with the convolution operator R , and with the actions of the other operators appearing in (13), the identity (35) implies that

$$\begin{aligned}
 |E_n(t) - E_n(s)| &= \left| \int_s^t \frac{dE_n}{d\tau}(\tau) d\tau \right| \\
 &\leq \int_s^t [|\langle \eta_n * (Bu(\cdot))(\tau), \eta_n * u(\tau) \rangle| \\
 &\quad + |\langle (\eta_n * R[u])(\tau), \eta_n * u(\tau) \rangle| \\
 &\quad + |\langle \eta_n * f(\tau), \eta_n * u(\tau) \rangle|] d\tau \\
 &\leq (\|B\| + 1) \left(\int_s^t \|(\eta_n * u)(\tau)\|^2 d\tau \right) \\
 &\quad + \int_s^t (\|(\eta_n * R[u])(\tau)\|^2 + \|(\eta_n * f)(\tau)\|^2) d\tau
 \end{aligned}
 \tag{36}$$

Since u , $R[u]$, and f are locally square-integrable, for each $t \in \mathbf{R}$ and $\epsilon > 0$, there exist $\Delta t(t, \epsilon) > 0$ and an $N(t, \epsilon) \in \mathbf{N}$ so that for $|s - t| < \Delta t(t, \epsilon)$ and $n > N(t, \epsilon)$,

$$\int_{s-1/n}^{t+1/n} \|u\|^2 < \epsilon, \int_{s-1/n}^{t+1/n} \|R[u]\|^2 < \epsilon, \text{ and } \int_{s-1/n}^{t+1/n} \|f\|^2 < \epsilon,$$

whence (36) implies that for $n > N(t, \epsilon)$, $|s - t| < \Delta t(t, \epsilon)$,

$$|E_n(t) - E_n(s)| < C\epsilon.
 \tag{37}$$

Continuity of E_n implies existence of $\overline{\Delta t}(t, \epsilon) > 0$ so that for $|s - t| < \overline{\Delta t}(t, \epsilon)$ and $n \leq N$, $|E_n(t) - E_n(s)| < C\epsilon$. Thus the inequality (37) holds for all $n \in \mathbf{N}$ if s satisfies $|s - t| < \min(\Delta t(t, \epsilon), \overline{\Delta t}(t, \epsilon))$. Since $t \in \mathbf{R}$, $\epsilon > 0$ are arbitrary, we have shown that the sequence $\{E_n\} \subset C^0(\mathbf{R})$ is equicontinuous.

Choose $T_0 \leq T \in \mathbf{R}$: it follows from the definition (34) and Young's inequality that

$$\int_{T_0}^T E_n \leq C \int_{T_0-1}^{T+1} \|u\|^2
 \tag{38}$$

is bounded independently of n . For $t \in [T_0, T]$,

$$(T - T_0)E_n(t) = \int_{T_0}^T (E_n(t) - E_n(s)) ds + \int_{T_0}^T E_n(s) ds
 \tag{39}$$

Apply Young's inequality to the convolutions with η_n appearing in the inequality (36) to conclude that for $T_0 \leq s, t \leq T$,

$$|E_n(t) - E_n(s)| \leq C \int_{T_0-1}^{T+1} (\|u\|^2 + \|R[u]\|^2 + \|f\|^2).
 \tag{40}$$

Taken together, (38)–(40) imply that $\{E_n\}$ is a bounded subset of $C^0([T_0, T])$. According to Ascoli's theorem, $\{E_n\}$ is precompact in $C^0([T_0, T])$, hence a subsequence converges uniformly to a continuous limit. Since the subsequence is necessarily also L^1 -convergent, and $T \in \mathbf{R}$ is arbitrary, the first assertion of the theorem is established.

In view of the continuity of E , we may take the limit $n \rightarrow \infty$ on both sides of the inequality (36) along the uniformly convergent subsequence whose existence we have just established. Since $\eta_n * u \rightarrow u$ in $L^2([T_0, T], H)$, the right hand side converges, and we obtain

$$|E(t) - E(s)| \leq C \int_s^t (\|u\|^2 + \|R[u]\|^2 + \|f\|^2). \quad (41)$$

We have assumed u to be causal, but this assumption has not appeared in the reasoning up to now. It allows us to take $s \rightarrow -\infty$ in (41). In view of the equivalence of \sqrt{E} and the norm $\|\cdot\|$ (inequalities (26)), the inequality (41) implies that

$$E(t) \leq C \int_{-\infty}^t (E + \|f\|^2).$$

Gronwall's inequality then yields the second conclusion. \square

Corollary 2. *The energy E of a weak solution u of (13), as defined above, satisfies for any $s, t \in \mathbf{R}$*

$$E(t) - E(s) = \int_s^t \langle -Bu(\tau) - R[u](\tau) + f(\tau), u(\tau) \rangle d\tau. \quad (42)$$

Proof. Continuity of E and convergence of $\eta_n * u$ to u in $L^2_{\text{loc}}(\mathbf{R}, H)$ allows us to take limits on both sides of (35). \square

Corollary 3. *Suppose that $u_1, u_2 \in L^2_{\text{loc}}(\mathbf{R}, H)$ are causal weak solutions of (13). Then $u_1 = u_2$.*

Proof. The conclusion follows immediately from the energy inequality (32), applied to the difference $u = u_1 - u_2$, which is a weak solution with $f \equiv 0$. \square

Corollary 4. *Suppose that $u \in L^2_{\text{loc}}(\mathbf{R}, H)$ is a causal weak solution of (13). Then $u \in C^0(\mathbf{R}, H)$.*

Proof. For $\delta t \in \mathbf{R}$, denote by $u_{\delta t}$ the member of $L^2_{\text{loc}}(\mathbf{R}, H)$ defined by $u_{\delta t}(t) = u(t + \delta t)$. Then $u_{\delta t}$ is a causal weak solution (the only one, thanks to corollary 3) of (13) with f replaced by $f_{\delta t} \in L^2(\mathbf{R}, H)$, defined by $f_{\delta t}(t) = f(t + \delta t)$. The translation group acts strongly continuously on L^2 , i.e. $\|f_{\delta t} - f\|_{L^2(\mathbf{R}, H)} \rightarrow 0$ as $\delta t \rightarrow 0$. Since the difference $u_{\delta t} - u$ is a causal solution of (13) with right-hand side $f_{\delta t} - f$, it follows immediately from (32) that $\|u_{\delta t}(t) - u(t)\|_H \rightarrow 0$ as $\delta t \rightarrow 0$ for any $t \in \mathbf{R}$, that is, $u \in C^0(\mathbf{R})$. \square

Corollary 5. *Suppose that*

- (1) $\mathcal{K} \subset \mathcal{B}(H)$ is a bounded set;
- (2) $\mathcal{L} \subset \mathcal{B}(V, H)$ is a bounded set of skew-adjoint operators on H with (common) domain V , whose graph norms are all equivalent (to each other and to the norm in V);
- (3) $\mathcal{M} \subset \mathcal{B}(H)$ is a bounded set of self-adjoint, uniformly positive definite operators: there exist constants $0 < C_* \leq C^*$ so that for all $A \in \mathcal{M}$,

$$C_* I \leq A \leq C^* I;$$

- (4) $\mathcal{Q} \subset L^1(\mathbf{R}, \mathcal{B}_{\text{symm}}(H)) \cap C^0(\mathbf{R}_+, \mathcal{B}(H))$ is a bounded set of causal operator-valued functions: if $Q \in \mathcal{Q}$, then $Q(t) = 0$ for $t < 0$.

Let the set $\mathcal{P} \subset \mathcal{M} \times \mathcal{L} \times \mathcal{K} \times \mathcal{Q}$ parametrize a family of formal evolution problems of for (13), with coefficients $A \in \mathcal{M}$, $P \in \mathcal{L}$, $B \in \mathcal{K}$, and $Q \in \mathcal{Q}$, with common right-hand side $f \in L^2(\mathbf{R}, H)$, and let $\mathcal{U} \subset L^2_{\text{loc}}(\mathbf{R}, H)$ be a corresponding family of causal weak solutions. Then $\mathcal{U} \subset C^0(\mathbf{R}, H)$ is equicontinuous.

Proof. That $\mathcal{U} \subset C^0(\mathbf{R}, H)$ is the content of the last corollary. It follows from the proof of the basic energy estimate (32) that the constant C appearing in its right-hand side may be chosen uniform over \mathcal{P} - indeed, the bounds defining the sets listed in conditions 1–4 above are precisely those on which our constants, canonically notated C , depend. Therefore (32) implies that for $u \in \mathcal{U}$,

$$\|u(t + \delta t) - u(t)\|^2 \leq \frac{1}{C_*} E_{u_{\delta t} - u}(t) \leq C \int_{-\infty}^t \|f_{\delta t} - f\|^2 = C \int_t^{t+\delta t} \|f\|^2$$

from which a uniform modulus of continuity follows. \square

Additional regularity in time of the right-hand side f translates into additional regularity in time of the solution.

Corollary 6. Suppose that $f \in L^2_{T_0}(\mathbf{R}, H) \cap H^k_{\text{loc}}(\mathbf{R}, H)$, $k \in \mathbf{N}$, and $u \in L^2_{\text{loc}}(\mathbf{R}, H)$ is a weak solution of (13). Then $u \in C^k(\mathbf{R}, H)$; moreover, for $j = 1, \dots, k$, $u^{(j)}$ is the weak solution of (14) with right-hand side $f^{(j)}$.

Proof. Corollary 4 is the case $k = 0$.

The case $k = 1$ provides the induction step, so we discuss it first. Since $f \in H^1_{\text{loc}}(\mathbf{R}, H)$, using the notation of the proof of corollary 4,

$$\frac{1}{\delta t} (f_{\delta t} - f) \rightarrow f' \text{ as } \delta t \rightarrow 0$$

in mean square. The energy estimate (32) implies that for each t ,

$$\frac{1}{\delta t} (u_{\delta t}(t) - u(t))$$

has a limit as $\delta t \rightarrow 0$, whence u is differentiable and u' satisfies (14) with right-hand-side f' . The previous corollary shows that u' is continuous.

This argument also serves as the induction step to establish the assertion of the corollary for $k > 1$. \square

Smoothing the solution in time leads to a different sort of ‘regularity’:

Corollary 7. Suppose that $T_0 \in \mathbf{R}$, $f \in L^2_{T_0}(\mathbf{R}, H)$, $\Delta T > 0$, $\eta \in C^\infty_0((-\Delta T, \Delta T))$, and let u denote a causal weak solution of (14). Then for every $T \geq T_0$, there exists $C_{T_0, T, \Delta T} \geq 0$ depending on $T_0, T, \Delta T$, $\|\eta\|_{L^1(\mathbf{R})}$, $\|\eta'\|_{L^1(\mathbf{R})}$, C_* , C^* , $\|B\|_{B(H)}$, and $\|Q\|_{L^1(\mathbf{R}, B(H))}$, so that for any $t \leq T$,

$$\|(\eta * u)(t)\|_V^2 \leq C_{T_0, T, \Delta T} \int_{-\infty}^{T+\Delta T} ds \|f(s)\|^2. \quad (43)$$

Remark. Note that this inequality is a pointwise bound on $\eta * u$ in the norm of V . In applications, V is compactly embedded in H , so this is potentially a much stronger statement than the obvious H -norm bound which follows directly from (32).

Proof. The left hand side makes sense, thanks to proposition 1. The identity (28) may be re-written as

$$-\langle P(\eta * u)(t), w \rangle = \int [\langle u(s), -A\eta'(t-s)w + B^*\eta(t-s)w + R^*[\eta(t-\cdot)w](s) \rangle] ds - \langle \eta * f(t), w \rangle$$

for any $w \in H$ (not just V !). Since $\text{supp}(\eta) \subset [-\Delta T, \Delta T]$, for any $t \leq T$,

$$|\langle P(\eta * u)(t), w \rangle| \leq C \|w\| \int_{-\infty}^{T+\Delta T} ds (\|u(s)\| + \|f(s)\|) \quad (44)$$

$$\leq C_{T_0, T, \Delta T} \|w\| \left(\int_{-\infty}^{T+\Delta T} ds \|f(s)\|^2 \right)^{\frac{1}{2}} \quad (45)$$

thanks to the energy estimate (32) and by-now familiar use of Young's inequality and the assumptions on the various operators and quantities in the problem formulation. Now choose $w = P(\eta * u)(t)$ to obtain a bound on its H -norm. In combination with (32) and Young's inequality, this estimate implies a bound of the required form on $(\eta * u)(t)$ in the graph norm of P , that is, in the norm of V . \square

8. Existence of weak solutions: proofs of theorems 5 and 6

The proof of existence follows the pattern laid out by Lions (1971), which in turn echoes Cauchy's proof of the fundamental theorem of ordinary differential equations. We define a Galerkin method, show that it converges, and finally that the limit is a weak solution. Note that no rate of convergence follows from this argument; in fact it is easy to see that none can be expected. Of course, proposition 2 has already assured that the solution so constructed is the only solution.

Proof of theorem 5. In view of the energy estimate (proposition 2), at most one such solution exists, and any sequence of weak solutions, corresponding to an $L^2(\mathbf{R}, H)$ -convergent sequence of right hand sides f , must itself be $L^2(\mathbf{R}, H)$ -convergent. Therefore it suffices to establish existence of solutions for a $L^2(\mathbf{R}, H)$ -dense set of right hand sides. In particular, we may assume that $f \in C^0(\mathbf{R}, H)$, without loss of generality.

Since V is separable (with respect to the graph norm of P) and $V \subset H$ is dense, countable linearly independent subsets $\{w_k\}_{k=1}^\infty \subset V$ exist for which finite linear combinations are dense in V , hence in H . Without loss of generality, assume that $\{w_k\}_{k=1}^\infty$ is $(H-)$ orthonormal: $\langle w_k, w_l \rangle = \delta_{kl}$, $k, l \in \mathbf{N}$.

Define $m \times m$ matrices A^m (symmetric positive definite), P^m and B^m by

$$A_{kl}^m = \langle Aw_k, w_l \rangle, \quad (46)$$

$$P_{kl}^m = \langle Pw_k, w_l \rangle, \quad (47)$$

$$B_{kl}^m = \langle Bw_k, w_l \rangle, \quad (48)$$

for $1 \leq k, l \leq m$, and the operator R^m on $L^2_{\text{loc}}(\mathbf{R})^m$ defined analogously to (12) by

$$R^m U^m(t) = \int_{-\infty}^t \langle Q^m(t-s)U^m(s) \rangle ds, \quad Q_{kl}^m(t) = \langle Q(t)w_k, w_l \rangle, \quad 1 \leq k, l \leq m.$$

Note that $Q^m \in L^1(\mathbf{R}, \mathcal{B}_{\text{symm}}(\mathbf{R}^m)) \cap C^0(\mathbf{R}_+, \mathcal{B}(\mathbf{R}^m))$ is causal ($Q^m(t) = 0, t < 0$).

Define $F^m \in C^0(\mathbf{R})^m$ by

$$F_k^m(t) = \langle f(t), w_k \rangle, 1 \leq k \leq m.$$

A minor modification of a standard contraction mapping argument (see for example Coddington and Levinson (1955)) shows that for each $m \in \mathbf{N}$, the initial value problem

$$\begin{aligned} A^m \frac{dU^m}{dt} + P^m U^m + B^m U^m + R^m U^m &= F^m, \\ U^m(t) &= 0, \quad t < T_0. \end{aligned} \quad (49)$$

has a unique solution $U^m \in C^1(\mathbf{R}, \mathbf{R}^m)$.

For each $m \in \mathbf{N}$, define $u_m \in C^1(\mathbf{R}, V)$, $f_m \in C^0(\mathbf{R}, H)$ by

$$u_m(t) = \sum_{k=1}^m U_k^m(t) w_k, \quad f_m(t) = \sum_{k=1}^m F_k^m(t) w_k.$$

Then the system (49) satisfied by U^m , together with the H -orthonormality of $\{w_k\}$, implies that u_m is the weak solution of the evolution equation (13) with right-hand side f_m . The energy estimate (32) shows that the sequence $\{u_m\}$ is bounded in $L_{\text{loc}}^2(\mathbf{R}, H)$, hence weakly precompact in $L_{\text{loc}}^2(\mathbf{R}, H)$ thanks to the Tychonoff–Alaoglu theorem and a diagonal process argument. Denote by $\{u_{m(l)}\}$ a weakly convergent subsequence, and by u its weak limit. Since $u_m(t) = 0$ for $t < T_0$ and all $m \in \mathbf{N}$, the same is true for u .

To see that the limit u is a weak solution of (13), introduce for each $m_0 \in \mathbf{N}$ test functions ψ of the form

$$\psi = \sum_{k=1}^{m_0} \phi_k \otimes w_k, \quad \phi_k \in C_0^\infty(\mathbf{R}). \quad (50)$$

For l sufficiently large that $m(l) > m_0$, $\langle f^m(t), \psi(t) \rangle = \langle f(t), \psi(t) \rangle$, $\langle P u_{m(l)}, \psi \rangle = -\langle u_{m(l)}, P \psi \rangle$, etc. So

$$\int \langle u_{m(l)}, A \psi' + P \psi - B^* \psi - R^*[\psi] \rangle dt = - \int \langle f, \psi \rangle dt.$$

Letting $l \rightarrow \infty$, it follows that u satisfies (14) for all test functions ψ of the form given in equation (50). Since linear combinations of w_m 's are dense in V , the set of functions of the form (50) is dense in $C_0^1(\mathbf{R}, V)$, whence u is a weak solution of (13). \square

Proof of theorem 6. We give the proof for $k = 1$; the general case follows by a straightforward induction argument.

Choose a Dirac sequence $\{\eta_n : n \in \mathbf{N}\} \subset C_0^\infty(\mathbf{R})$ as in the proof of proposition 2. According to corollary 7, $\eta_n * u \in C^k(\mathbf{R}, V)$ for each $n \in \mathbf{N}$. According to corollary 6, $u \in C^1(\mathbf{R}, H)$, so the first term on the right-hand side of (28) may be integrated by parts to yield

$$\langle (\eta_n * u), Pw \rangle = \eta_n * \langle u, Pw \rangle \quad (51)$$

$$= \eta_n * \langle Au' + Bu + R[u] - f, w \rangle \quad (52)$$

for any $w \in V$. Both sides are continuous in t , so the limit as $n \rightarrow \infty$ is valid pointwise. Thus

$$\langle u(t), Pw \rangle = \langle Au'(t) + Bu(t) + R[u](t) - f(t), w \rangle \quad (53)$$

for any $w \in V$, $t \in \mathbf{R}$. Pointwise bounds on $\|u(t)\|$ (proposition 2), $\|u'(t)\|$ (corollary 6) and the standing assumptions on the various operators imply that

$$\langle u(t), Pw \rangle \leq C_{T_0, T} \|w\| \left(\int_{-\infty}^T [\|f\|^2 + \|f'\|^2] \right)^{\frac{1}{2}}$$

for all $w \in V$, $t \in (-\infty, T]$ (and any T , but the constant $C_{T_0, T}$ depends on T_0 and T), which shows that $u(t) \in \mathcal{D}(P^*) = \mathcal{D}(P) = V$ for each $t \in \mathbf{R}$. Thus

$$\langle Pu(t), w \rangle \leq C_{T_0, T} \|w\| \left(\int_{-\infty}^T [\|f\|^2 + \|f'\|^2] \right)^{\frac{1}{2}}$$

for all $w \in H$, $t \in (-\infty, T]$; taking $w = Pu(t)$, obtain the pointwise bound (16) for $k = 1$. That $u \in C^0(\mathbf{R}, V)$ follows exactly as in the proof of corollary 4, via the strong continuity of the translation group on $H^1(\mathbf{R}, H)$. \square

9. Continuous dependence on parameters: proofs of theorems 7–9

The proof of theorem 7 divides into three steps. In the first step (lemma 1), we show that the assumptions of the theorem imply $L^2_{\text{loc}}(\mathbf{R}, H)$ -weak convergence of the sequence of solutions. The second step leverages this result to show that the sequence converges H -weakly, pointwise (lemma 2). The last step combines these two observations with strong convergence of the coefficients to obtain convergence in norm of the solution sequence.

Lemma 1. *Under the conditions of theorem 7, u_m converges weakly to u in $L^2_{\text{loc}}(\mathbf{R}, H)$.*

Remark. In order that R_m converge to R pointwise, as assumed in the statement of the preceding theorem, it is sufficient that $Q_m \rightarrow Q$ uniformly in \mathbf{R}_+ .

Proof. The bounds implied by the energy estimate (proposition 2) are uniform over bounded sets of coefficients as described in the statement of the theorem. Therefore $\{u_m\}$ is bounded in $L^2_{\text{loc}}(\mathbf{R}, H)$, hence has an $L^2_{\text{loc}}(\mathbf{R}, H)$ -weakly convergent subsequence $u_{m(l)}$, with limit $\bar{u} \in L^2_{\text{loc}}(\mathbf{R}, H)$. Note that $L^2_{\text{loc}}(\mathbf{R}, H)$ -weak convergence implies convergence in the sense of H -valued distributions on \mathbf{R} . Choose a test function $\phi \in C_0^\infty(\mathbf{R}, V)$: then

$$-\int \langle f(s), \phi(s) \rangle ds = \int \langle u_{m(l)}(s), (A_{m(l)}\phi' + P_{m(l)}\phi - B_{m(l)}^*\phi - R_{m(l)}^*[\phi])(s) \rangle ds \quad (54)$$

$$= \int \langle \bar{u}(s), (A\phi' + P\phi - B^*\phi - R^*[\phi])(s) \rangle ds \quad (55)$$

$$+ \int \langle (u_{m(l)}(s) - \bar{u}(s)), (A\phi' + P\phi - B^*\phi - R^*[\phi])(s) \rangle ds \quad (56)$$

$$+ \int \langle u_{m(l)}(s), ((A_{m(l)} - A)\phi' + (P_{m(l)} - P)\phi \quad (57)$$

$$- (B_{m(l)} - B)^*\phi - (R_{m(l)} - R)^*[\phi])(s) \rangle ds \quad (58)$$

The second term vanishes in the limit $l \rightarrow \infty$ because of the weak convergence of $u_{m(l)}$ to \bar{u} . The coefficients A_m, \dots range over bounded sets of operators, so we may replace ϕ and ϕ' in the third term with simple V -valued functions, taking finitely many values, at the price of an arbitrarily small perturbation in this term, uniformly in l . However the strong convergence of the coefficient operators assumed in the statement of the theorem then implies that the resulting integrals become arbitrarily small as $l \rightarrow \infty$. Thus \bar{u} is a weak solution of the problem (13), and must therefore be the same as the (unique) weak solution u constructed in the preceding section. Thus no other weak accumulation point of the bounded sequence $\{u_m\}$ may exist, hence $u_m \rightharpoonup u$ in $L^2_{\text{loc}}(\mathbf{R}, H)$ as claimed. \square

Lemma 2. Under the conditions of theorem 7, u_m converges to u weakly, pointwise in $t \in \mathbf{R}$ and uniformly on compact sets. That is, $u_m(t) \rightharpoonup u(t)$ for all $t \in \mathbf{R}$, and for any $w \in H$, $T_0 \leq T \in \mathbf{R}$,

$$\lim_{m \rightarrow \infty} |\langle u_m - u, w \rangle|_{L^\infty([T_0, T])} = 0.$$

Proof. According to corollary 5, the conditions described in the statement of theorem 7 imply that $\{u_m : m \in \mathbf{N}\}$ is equicontinuous, hence uniformly equicontinuous on compact sets. Choose $T_0 \leq T \in \mathbf{R}$. Given $\epsilon > 0$, choose $\Delta t > 0$ so that if $|\delta t| < \Delta t$, $t \in [T_0, T]$, then

$$\|u_m(t + \delta t) - u_m(t)\| < \epsilon, m \in \mathbf{N}; \|u(t + \delta t) - u(t)\| < \epsilon,$$

which implies that

$$\left\| u_m(t) - \frac{1}{2\Delta t} \int_{t-\Delta t}^{t+\Delta t} u_m \right\| < \epsilon, m \in \mathbf{N}; \left\| u(t) - \frac{1}{2\Delta t} \int_{t-\Delta t}^{t+\Delta t} u \right\| < \epsilon.$$

However, according to lemma 1, for any $w \in H$,

$$\frac{1}{2\Delta t} \int_{t-\Delta t}^{t+\Delta t} \langle u_m - u, w \rangle = \int \left\langle u_m - u, w \frac{1}{2\Delta t} \mathbf{1}_{[t-\Delta t, t+\Delta t]} \right\rangle \rightarrow 0, m \rightarrow \infty.$$

Therefore, assuming without loss of generality that $\|w\| = 1$,

$$|\langle u_m(t) - u(t), w \rangle| \leq 3\epsilon$$

for m sufficiently large, $t \in [T_0, T]$. Since T_0, T , and $\epsilon > 0$ are arbitrary, the proof is complete. \square

We require one more fact about the memory operators R :

Lemma 3. Suppose that $\{Q_m : m \in \mathbf{N}\} \subset L^1(\mathbf{R}, \mathcal{B}_{\text{symm}}(H)) \cap C^0(\mathbf{R}_+, \mathcal{B}(H))$ defines a sequence of operators $R_m : L^2_+(\mathbf{R}, H) \rightarrow L^2_+(\mathbf{R}, H)$ converging strongly to zero. Let $\mathcal{K} \subset L^2_+(\mathbf{R}, H)$ be bounded, and assume that all $u \in \mathcal{K}$ are causal, with common support $\text{supp } u \subset [T_0, \infty)$. Then for any sequence $\{u_m : m \in \mathbf{N}\} \subset \mathcal{K}$, and any $v \in L^2_{\text{loc}}(\mathbf{R}, H)$, any $T \in \mathbf{R}$,

$$\int_{-\infty}^T dt \langle v(t), R_m[u_m](t) \rangle \rightarrow 0, m \rightarrow \infty$$

Proof.

$$\begin{aligned} \int_{-\infty}^T dt \int_{-\infty}^t ds \langle v(t), R_m[u_m](t) \rangle &= \int_{-\infty}^T \int_{T_0}^t ds \langle v(t), Q_m(t-s)u_m(s) \rangle \\ &= \int_{-T}^{-T_0} d\sigma \int_{-T}^{\sigma} d\tau \langle u_m(-\sigma), Q_m(\sigma-\tau)v(-\tau) \rangle \\ &= \int dt \langle \mathbf{1}_{[-T, -T_0]}(t) \check{u}_m(t), R_m[\mathbf{1}_{[-T, -T_0]}\check{v}](t) \rangle. \end{aligned}$$

Note that we have used the symmetry of Q_m , and it is for this reason that the assumption was introduced. According to the hypothesis, $\|R_m[w]\|_{L^2((-\infty, t])} \rightarrow 0$ as $m \rightarrow \infty$ for any $w \in L^2_+(\mathbf{R}, H)$, $t \in \mathbf{R}$. Setting $t = -T_0$, $w = \mathbf{1}_{[-T, -T_0]}\check{v}$. Since $\mathcal{K} \in L^2_{\text{caus}}(\mathbf{R}, H)$ is bounded, the sequence

$$\{\|\mathbf{1}_{[-T, -T_0]}(t) \check{u}_m\|_{L^2(\mathbf{R}, H)}\} \subset \mathbf{R}$$

is bounded, whence the conclusion follows via the Cauchy-Schwarz inequality. \square

Proof of theorem 7. We will use repeatedly the algebraic identity: for $K, L \in \mathcal{B}(H)$, $v, w \in H$,

$$\langle v, Kv \rangle - \langle w, Lw \rangle = \langle v - w, K(v - w) \rangle + \langle v - w, Lw \rangle \\ + \langle w, L(v - w) \rangle + \langle v, (K - L)w \rangle + \langle w, (K - L)(v - w) \rangle.$$

If K and L are symmetric, this identity simplifies to

$$\langle v, Kv \rangle - \langle w, Lw \rangle = \langle v - w, K(v - w) \rangle + 2\langle v - w, Lw \rangle + \langle 2v - w, (K - L)w \rangle$$

Application of the second version of this identity with $v = u_m$, $w = u$, $K = A_m$, $L = A$ yields

$$\langle u_m, A_m u_m \rangle - \langle u, Au \rangle = \langle u_m - u, A_m(u_m - u) \rangle \\ + \langle 2u_m - u, (A_m - A)u \rangle + 2\langle u_m - u, Au \rangle. \quad (59)$$

The right-hand side f in the formal evolution equation (13) for both u_m and u vanishes for sufficiently large negative t , else u could not be causal, but then u_m and u must vanish on a common (negative) half-axis, thanks to corollary 3. The energy identity (42) implies that

$$\langle u_m, A_m u_m \rangle - \langle u, Au \rangle(t) \\ = - \int_{-\infty}^t [\langle B_m u_m, u_m \rangle - \langle Bu, u \rangle + \langle R_m[u_m], u_m \rangle - \langle R[u], u \rangle - \langle f, u_m - u \rangle].$$

Application of the algebraic identities stated above shows that the right-hand side is

$$= - \int_{-\infty}^t [\langle u_m - u, B_m(u_m - u) \rangle + \langle 2u_m - u, (B_m - B)u \rangle + 2\langle u_m - u, Bu \rangle \\ \times \langle u_m - u, R_m[u_m - u] \rangle + \langle u_m, R_m[u] - R[u] \rangle + \langle u, R_m[u_m - u] - R[u_m - u] \rangle \\ + \langle u_m - u, R[u] \rangle + \langle u, R[u_m - u] \rangle + \langle f, u_m - u \rangle]. \quad (60)$$

Identities (59) and (60) combine to yield

$$\langle u_m - u, A_m(u_m - u) \rangle(t) = - \int_{-\infty}^t \langle B_m(u_m - u) + R_m[u_m - u], u_m - u \rangle + g_m(t), \quad (61)$$

in which $g_m \in C^0(\mathbf{R})$ is defined by

$$g_m(t) = - \langle 2u_m(t) - u(t), (A_m - A)u(t) \rangle - 2\langle (u_m - u)(t), Au(t) \rangle \\ - \int_{-\infty}^t [\langle 2u_m - u, (B_m - B)u \rangle + 2\langle u_m - u, Bu \rangle \\ + \langle u_m, R_m[u] - R[u] \rangle + \langle u, R_m[u_m - u] - R[u_m - u] \rangle \\ + \langle u_m - u, R[u] \rangle + \langle u, R[u_m - u] \rangle + \langle f, u_m - u \rangle]. \quad (62)$$

Since the B_m 's are uniformly bounded operators on H and the R_m 's are uniformly bounded operators on $L^2((-\infty, t], H)$ for every $t \in \mathbf{R}$ (with norm independent of t), (61) implies that

$$\|u_m - u\|^2(t) \leq C \langle u_m - u, A_m(u_m - u) \rangle(t) \leq C \int_{-\infty}^t \|u_m - u\|^2 + |g_m(t)|. \quad (63)$$

in which C means something different each time it occurs, as before, but depends on the quantities indicated the second section.

Select T_0 for which $u(t) = u_m(t) = 0$ for all $t < T_0$, $m \in \mathbf{N}$. Choose $T \in \mathbf{R}$. Application of Gronwall's inequality to (63) yields, for $T_0 \leq t \leq T$ and C depending on T along with everything else,

$$\|u_m - u\|^2(t) \leq C_{T-T_0} |g_m|(t). \quad (64)$$

in which C_{T-T_0} depends on $T - T_0$ in addition to the standard dependences.

It remains to see that $g_m(t) \rightarrow 0$ uniformly in $t \in [T_0, T]$, as $m \rightarrow \infty$. The first term in (62) tends to zero pointwise (in t) thanks to the assumption that $A_m \rightarrow A$ strongly, and to the energy estimate (theorem 5) which assures that $\|u_m - u\|$ is bounded uniformly in $m \in \mathbf{N}$ and $t \in [T_0, T]$. On the other hand, corollary 5 and the uniform bounds on $\{A_m : m \in \mathbf{N}\}$ implied by the assumption that $\{A_m, B_m, Q_m\} : m \in \mathbf{N} \subset \mathcal{P}(C_*, C^*, C_B, C_Q)$ in turn imply that the first term in (62) defines an equicontinuous sequence of continuous functions on $[T_0, T]$. Since any convergent subsequence converges pointwise to the zero function, so does the entire sequence, and uniformly. The second term in (62) tends to zero, uniformly on $[T_0, T]$, thanks to lemma 2, the third because $B_m \rightarrow B$ strongly, the fourth, seventh, eighth and ninth because of lemma 1 and theorem 5, the fifth because $R_m \rightarrow R$ strongly in $L^2_+(\mathbf{R}, H)$, the sixth because of lemma 3. \square

Remark. This result is sharp, in the sense that nothing stronger than continuity can be expected without additional constraints on the various components of the formal evolution problem (13). In particular, the modulus of continuity cannot be uniform in the right-hand side ($f \in L^2(\mathbf{R}, H)$), even locally.

For example, the 1D linear advection problem

$$\left(\frac{1}{c} \frac{\partial u}{\partial t} - \frac{\partial u}{\partial x}\right)(t, x) = f(t, x)$$

conforms to the setting described above, with $H = L^2(\mathbf{R})$. The operator coefficients are: $A =$ multiplication by the positive constant $1/c$, $P = \partial/\partial x$, skew-adjoint with domain $V = H^1(\mathbf{R})$, and $B \equiv 0, Q \equiv 0$. For any $f \in L^2(\mathbf{R}^2) (\equiv L^2(\mathbf{R}, H))$ by Fubini's theorem, the causal weak solution is

$$u[c, f](t, x) = c \int_{-\infty}^t f(\tau, x + c(t - \tau)) \, d\tau, \tag{65}$$

in which we have explicitly indicated the dependence of the weak solution on the coefficient $1/c$ and the right-hand side f .

Suppose $\chi \in C_0^\infty(\mathbf{R})$, $\text{supp } \chi \subset [-1, 1]$, and

$$\int \chi = 1.$$

For $\epsilon > 0$, set $f_\epsilon(t, x) = \cos((x + t)/\epsilon)\chi(x + t)\chi(x)$. Then $u[1, f_\epsilon](t, x) = \cos((x + t)/\epsilon)\chi(x + t)$ for $t > 1$, whereas integration by parts shows that

$$u[c, f_\epsilon](t, x) = O\left(\frac{\epsilon}{|c - 1|}\right)$$

for $c \neq 1$. Thus the modulus of continuity of $(c, f_\epsilon) \mapsto u[c, f_\epsilon](t, \cdot) \in H$ (for $t > 1$) cannot be uniform over the bounded set $[a, b] \times \{f_\epsilon : \epsilon > 0\} \subset \mathbf{R} \times L^2(\mathbf{R}, H)$, if $1 \in (a, b)$, and in particular this map is not locally uniformly continuous. The heuristic reason is that c is the wave speed, so changing c changes the position of arriving waves. This position shift has unboundedly large impact if the frequency of oscillation in the solution is allowed to become arbitrarily large. Oscillations in the data in time translate into oscillations in the solution in space, of similar frequency, because the problem is hyperbolic.

On the other hand, additional regularity *in time* of the right-hand side entails more regular behavior of the solution, as one might expect since such regularity damps temporal frequencies.

Corollary 8. *In addition to the hypotheses of theorem 7, assume that $f \in H_{\text{loc}}^k(\mathbf{R}, H)$, $k \geq 1$. Then for any choice of $T_0 \leq T \in \mathbf{R}$,*

$$\lim_{m \rightarrow \infty} \|u_m - u\|_{C^{k-1}([T_0, T], H)} = 0.$$

Proof. Follows directly from theorems 6 and 7. \square

Proof of theorem 8. The meaning of (18) is that for any $\phi \in C_0^\infty(\mathbf{R}, V)$,

$$\begin{aligned} \int \langle \delta u, A\phi' + P\phi - B^*\phi - R^*[\phi] \rangle &= \int \langle \delta Au' + \delta Bu + \delta R[u], \phi \rangle \\ &= - \int \langle u, \delta A\phi' - \delta B^*\phi - \delta R^*[\phi] \rangle. \end{aligned} \quad (66)$$

On the other hand, both u and u_h , $h > 0$, satisfy (14) with the same right-hand side, so

$$\begin{aligned} 0 &= \frac{1}{h} \left(\int \langle u_h, A_h\phi' + P\phi - B_h^*\phi - R_h^*[\phi] \rangle \right. \\ &\quad \left. - \int \langle u, A\phi' + P\phi - B^*\phi - R^*[\phi] \rangle \right) \\ &= \int \langle u_h, \delta A\phi' - \delta B^*\phi - \delta R^*[\phi] \rangle \\ &\quad + \int \left\langle \frac{u_h - u}{h}, A\phi' + P\phi - B^*\phi - R^*[\phi] \right\rangle. \end{aligned} \quad (67)$$

Subtracting (66) from (67) and rearranging, obtain

$$\begin{aligned} \int \left\langle \left(\frac{u_h - u}{h} - \delta u \right), A\phi' + P\phi - B^*\phi - R^*[\phi] \right\rangle \\ = \int \langle u_h - u, \delta A\phi' - \delta B^*\phi - \delta R^*[\phi] \rangle \\ = - \int \langle \delta A(u_h - u)' + \delta B(u_h - u) + \delta R[u_h - u], \phi \rangle. \end{aligned} \quad (68)$$

In view of equation (68), the Newton quotient remainder

$$\frac{u_h - u}{h} - \delta u$$

is the weak solution of (13) with right-hand side

$$\delta A(u_h - u)' + \delta B(u_h - u) + \delta R[u_h - u] \in L_{\text{loc}}^2(\mathbf{R}, H).$$

In view of corollary 8 and the energy estimate (theorem 5) imply that

$$\|\delta A(u_h - u)' + \delta B(u_h - u) + \delta R[u_h - u]\|_{L^2((-\infty, T], H)} \rightarrow 0$$

as $h \rightarrow 0$ for any $T \in \mathbf{R}$. The conclusion then follows from another use of theorem 5. \square

Proof of theorem 9. We present the case $k = 2$. The general case follows by an induction argument, which we omit.

As usual, denote by $T_0 \in \mathbf{R}$ a lower bound for $\text{supp } f$. In the notation of the proof of theorem 8, $u \in C^2(\mathbf{R}, H) \cap C^1(\mathbf{R}, V)$, thanks to theorem 6. Therefore the right-hand side of (18) has a locally square-integrable derivative, whence the causal weak solution δu actually belongs to the class $C^1(\mathbf{R}, H) \cap C^0(\mathbf{R}, V)$. Applying theorems 5 and 6 repeatedly, one sees that δu satisfies for any $T_1 \in \mathbf{R}$

$$\|\delta u\|_{L^\infty([T_0, T_1], V)} \leq C_{T_1 - T_0} \|(\delta A, \delta B, \delta Q)\|_{\mathcal{P}} \|f\|_{H^2([T_0, T_1])}.$$

It follows that the linear map $D\mathcal{F}[(A, B, Q)] : \mathcal{M} \rightarrow C^0([T_0, T_1], W)$ defined by

$$D\mathcal{F}[(A, B, Q)](\delta A, \delta B, \delta Q) = S\delta u$$

is continuous.

Suppose that $(A_m, B_m, Q_m) \rightarrow (A, B, Q)$ in norm (of \mathcal{M}). Denote by δu_m the solution of (18) with (A, B, Q) replaced by (A_m, B_m, Q_m) . It follows from the definition (equation (18)) that $\delta u_m - \delta u$ is the (strong) solution of

$$\begin{aligned} & A(\delta u_m - \delta u)' + P(\delta u_m - \delta u) + B(\delta u_m - \delta u) + R[\delta u_m - \delta u] \\ &= - [(A_m - A)\delta u_m + (B_m - B)\delta u_m + R_m[\delta u_m] - R_m[\delta u] \\ &\quad + \delta A(u_m - u)' + \delta B(u_m - u) + \delta R[u_m - u]]. \end{aligned} \tag{69}$$

Theorem 6 implies that $\{\delta u_m : m \in \mathbf{N}\}$ is a bounded set in $C^1((-\infty, T_1], H) \cap C^0((-\infty, T_1], V)$, whence in $H^1_{\text{loc}}((-\infty, T_1], H)$ also; moreover $\text{supp } \delta u_m \subset [T_0, \infty)$ for all $m \in \mathbf{N}$. It follows that the first three terms on the right-hand side of (69) tend to zero in $H^1([T_0, T_1], H)$. Applying theorem 7 to $u'_m - u'$, the difference of solutions of equations of the form (13) with the same causal right-hand side $f' \in H^1_{\text{loc}}(\mathbf{R}, H)$, one sees that $u_m - u \rightarrow 0$ in $H^1([T_0, T_1], H)$ also, whence the second three terms also tend to zero in this sense. Thus the entire right-hand side of (69) tends to zero as $m \rightarrow \infty$ in the sense of $H^1([T_0, T_1], H)$. Now it follows from theorem 6 that

$$(D\mathcal{F}[(A_m, B_m, Q_m)] - D\mathcal{F}[(A, B, Q)])(\delta A, \delta B, \delta Q) = S(\delta u_m - \delta u) \rightarrow 0$$

(in the sense of $C^0([T_0, T_1], W)$). That is, $D\mathcal{F} : \mathcal{P} \times \mathcal{M} \rightarrow C^0([T_0, T_1], W)$ is continuous, and the theorem is proved in the case $k = 2$. □

10. Symmetric hyperbolic systems: continuous dependence, proofs of theorems 1–4

For convenience, we repeat the key definitions from the second section.

Recall that $k \times k$ symmetric hyperbolic systems take the form

$$a \frac{\partial u}{\partial t} + p(\nabla)u + bu + q * u = f \text{ in } \Omega \times \mathbf{R}; u = 0 \text{ for } t < 0, \tag{70}$$

in which the coefficient matrices a, b , and q are $k \times k$, and the $k \times k$ matrix differential operator in the ‘space’ variables $\mathbf{x} \in \Omega$ has symmetric and constant coefficient matrices.

The set of admissible coefficients $M \subset L^\infty(\Omega, \mathbf{R}^{k \times k}_{\text{symm}}) \times L^\infty(\Omega, \mathbf{R}^{k \times k}) \times L^1(\mathbf{R}_+, L^\infty(\Omega, \mathbf{R}^{k \times k}_{\text{symm}}))$ is defined by

$$M(C_*, C^*, C_B, C_Q) = \{(a, b, q) : C_*I < a(\mathbf{x}) < C^*I \text{ for all } \mathbf{x} \in \Omega, \|b\|_{L^\infty(\Omega, \mathbf{R}^{k \times k})} < C_B,$$

$$\|q\|_{L^1(\mathbf{R}_+, L^\infty(\Omega, \mathbf{R}^{k \times k}_{\text{symm}}))} < C_Q, q(t) = 0 \text{ for } t < 0\},$$

$$M = \bigcup \{M(C_*, C^*, C_B, C_Q) : 0 < C_* \leq C^*, C_B, C_Q \in \mathbf{R}_+\}. \tag{71}$$

M corresponds to a collection of problems of the form (70), with common $p(\nabla)$.

The Hilbert space of states is $H = L^2(\Omega)^k$. Provide $p(\nabla)$ with a dense domain $V \subset H$, and assume the $p(\nabla) : V \rightarrow H$ is skew-adjoint, and that the norm in V is equivalent to the graph norm of $p(\nabla)$.

The hypotheses of theorem 5 are all satisfied, so

Corollary 9. *Suppose that $(a, b, q) \in M$, and $f \in L^2_{\text{loc}}(\mathbf{R}, H)$ is causal. Then there exists a unique weak solution of (70), satisfying for every $\phi \in H^1(\mathbf{R}, L^2(\Omega)^k)$*

$$\int_{-\infty}^{\infty} dt \int_{\Omega} d\mathbf{x} [u(-A\phi' - p(\nabla)\phi + b\phi + q * \phi) - f\phi] = 0.$$

There exists an increasing function $C : \mathbf{R}_+ \rightarrow \mathbf{R}_+$, depending on C_*, C^* , and bounds for b, q , so that for every $t \in \mathbf{R}$,

$$\|u(\cdot, t)\|_{L^2(\Omega)^k}^2 \leq C(t) \int_0^t d\tau \|f(\cdot, \tau)\|_{L^2(\Omega)^k}^2. \tag{72}$$

Theorems 1 and 2 now follow directly from theorems 8 and 9 respectively.

Multiplication by a member of $L^\infty(\Omega)$ defines a continuous map from $L^\infty(\Omega)$ to $\mathcal{B}(L^2(\Omega))$, so convergence of the coefficients a, b, q in $L^\infty(\Omega)$ is sufficient to induce uniform, hence strong, convergence of the corresponding operators, hence convergence of the solutions per theorem 7. However, convergence in a weaker sense is also sufficient to induce strong operator convergence. The key observation is the following result, identical to lemma 2.8.5 in Stolk (2000):

Lemma 4. Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, $\{r_m\} \subset L^\infty(\Omega, \mu)$ with $\|r_m\|_{L^\infty(\Omega, \mu)} \leq R \in \mathbf{R}_+$, $\{f_m\} \subset L^2(\Omega, \mu)$ with $\|f_m\|_{L^2(\Omega, \mu)} \leq F \in \mathbf{R}_+$ for all $m \in \mathbf{N}$. Suppose that $r_m \rightarrow 0$ in μ -measure. Then for any $g \in L^2(\Omega, \mu)$,

$$\lim_{m \rightarrow \infty} \int_{\Omega} r_m f_m g \, d\mu = 0. \tag{73}$$

Proof. Suppose on the contrary that such sequences $\{r_m\}, \{f_m\}$ and square-integrable g exist, also an $\eta > 0$, for which the left-hand side of (73) remains $\geq \eta$ along a common subsequence. Without loss of generality, renumber the subsequence so that

$$\left| \int_{\Omega} r_m f_m g \, d\mu \right| \geq \eta, \quad m \in \mathbf{N}. \tag{74}$$

Convergence in measure of $\{r_m\}$ means that for any $\epsilon > 0$,

$$\mu[\Omega_\epsilon(r_m)] \rightarrow 0 \text{ as } m \rightarrow \infty, \text{ where } \Omega_\epsilon(r_m) = \{\mathbf{x} \in \Omega : |r_m(\mathbf{x})| \geq \epsilon\}.$$

Choose ϵ so that $\epsilon F \|g\|_{L^2(\Omega, \mu)} < \eta/2$.

From this definition and the Cauchy–Schwarz inequality, one sees that

$$\begin{aligned} \left| \int_{\Omega} r_m f_m g \, d\mu \right| &\leq \epsilon \int_{\Omega \setminus \Omega_\epsilon(r_m)} |f_m g| \, d\mu + R \int_{\Omega_\epsilon(r_m)} |f_m g| \, d\mu \\ &\leq \epsilon F \|g\|_{L^2(\Omega, \mu)} + RF \left(\int_{\Omega_\epsilon(r_m)} g^2 \, d\mu \right)^{\frac{1}{2}} < \frac{\eta}{2} + RF \left(\int_{\Omega_\epsilon(r_m)} g^2 \, d\mu \right)^{\frac{1}{2}} \end{aligned}$$

By passing if necessary to a further subsequence, we may assume that

$$\mu[\Omega_\epsilon(r_m)] \leq 2^{-m} \Rightarrow \sum_m \mu[\Omega_\epsilon(r_m)] < \infty.$$

Thus the characteristic functions of the sets $\Omega_\epsilon(r_m)$ are almost everywhere convergent to zero as $m \rightarrow \infty$. Since $|g|^2 \in L^1(\Omega, \mu)$, it follows from the Lebesgue dominated convergence theorem that for large enough m ,

$$\left(\int_{\Omega_\epsilon(r_m)} g^2 \, d\mu \right)^{\frac{1}{2}} < \frac{\eta}{2RF}$$

Thus the left-hand side of (74) can be made smaller than η , a contradiction. □

Lemma 5. Suppose that $\{a_m\}_{m=1}^\infty \subset L^\infty(\mathbf{R}^n)$ converges in measure to $a \in L^\infty(\mathbf{R}^n)$, and that $A_m, m \in \mathbf{N}$ and $A \in \mathcal{B}(L^2(\mathbf{R}^n))$ are defined by

$$(A_m u)(\mathbf{x}) = a_m(\mathbf{x})u(\mathbf{x}), \quad (A u)(\mathbf{x}) = a(\mathbf{x})u(\mathbf{x}), \quad u \in L^2(\mathbf{R}^n).$$

Then $A_m \rightarrow A$ strongly. The same is true for similar sequences of operators on $L^2(\mathbf{R}^n)^p$ defined by sequences of $p \times p$ matrix-valued functions whose components converge in measure.

Proof. In fact, the operators so defined are self-adjoint, and

$$\|A_m u - Au\|^2 = \int (a_m - a)[(a_m - a)u]u \rightarrow 0,$$

as follows from lemma 4, taking $a_m - a$ for r_m , $(a_m - a)u$ for f_m , and u for g in the notation of that lemma. \square

Proof of theorem 3. Follows immediately from lemma 5 and theorem 7. \square

One way to obtain convergence in measure is via mollification. Let $\{\eta_m : m \in \mathbf{N}\}$ be a Dirac sequence of mollifiers, as in the proof of proposition 2. Since $a \in L^p_{\text{loc}}(\mathbf{R}^d)$ for any p , $\eta_m * a \rightarrow a$ pointwise almost everywhere, hence in measure. An application of this observation is the *finite speed of propagation* property: for solutions of (70), support expands at finite speed. This property is well-understood for hyperbolic systems with smooth coefficients: for example, Lax (2006), chapter 4, presents a proof of:

Proposition 3. *In addition to the hypotheses of theorem 1, suppose that the coefficients in (70) are smooth (of class $C^\infty(\bar{\Omega})$). Then*

- (1) *the causal weak solution $u \in L^2_{\text{loc}}(\mathbf{R}, H)$ is smooth: $u \in C^\infty(\mathbf{R}^{n+1})^p$, and*
 (2) *if $\phi \in C^\infty(\mathbf{R}^n)$ satisfies*

$$a + \sum p_i \frac{\partial \phi}{\partial x_i} > 0$$

and $\text{supp}(f) \cap \{(\mathbf{x}, t) : \phi(\mathbf{x}) > t\} = \emptyset$, then $u(\mathbf{x}, t) = 0$ if $\phi(\mathbf{x}) \geq t$.

Corollary 10. *In the setting of proposition 3, suppose that $\tau \in \mathbf{R}$ satisfies*

$$\tau a(\mathbf{x}) + \sum_{i=1}^n p_i \xi_i \geq 0, \quad \mathbf{x} \in \mathbf{R}^n, \quad |\xi| = 1, \quad \text{a.e. } \mathbf{x} \in \Omega. \quad (75)$$

If $\mathbf{x}_0 \in \mathbf{R}^n$, $t_0 \in \mathbf{R}$ satisfy

$$f(\mathbf{x}, t) = 0 \text{ if } |\mathbf{x} - \mathbf{x}_0| \leq \tau(t_0 - t),$$

then $u(\mathbf{x}_0, t_0) = 0$.

Proof in proposition 3. Take $\phi(\mathbf{x}) = t_0 - |\mathbf{x} - \mathbf{x}_0|/\tau$. \square

Note that the argument given in Lax (2006) does not quite encompass proposition 3 and corollary 10: it does not apply to systems like (70) with memory terms. However the extension is straightforward.

Proof of theorem 4. Using a sequence of mollifiers as before, obtain a sequence $\{(a_m, b_m, q_m) : m \in \mathbf{N}\}$ of smooth coefficients converging in measure to (a, b, q) . Then the corresponding operators converge strongly to those induced by (a, b, q) . Since (5) holds almost everywhere, it follows that a_m satisfies the spectral inequality (75) in \mathbf{R}^n for sufficiently large m , whence corollary 10 implies that u_m vanishes in ω . Since, $u_m \rightarrow u$ in $L^2(\Omega)^k$ the conclusion follows. \square

Acknowledgments

KDB acknowledges the support of his work by the Rice University VIGRE program, funded by the National Science Foundation. The work of KDB and WWS was supported in part by the sponsors of The Rice Inversion Project. The research of CS is supported by the Netherlands Organisation for Scientific Research, through VIDI grant 639.032.509. The authors are grateful to Yin Huang for a very careful reading of a preliminary version and for several helpful suggestions. WWS acknowledges the Isaac Newton Institute for Mathematical Sciences at Cambridge University (programme on Inverse Problems, Fall 2011) for the opportunity to take advantage of its superbly productive environment during the final stages of our work on this project.

Appendix A. The skew-adjoint property of the acoustic grad-div operator

We establish skew-adjointness of the operator D , defined formally in (8), by appealing to an auxiliary result, similar to necessary and sufficient conditions for self-adjointness found in many texts on functional analysis (for example, Conway (1990)).

Lemma 6. *Suppose that H is a Hilbert space, $V \subset H$ a dense subspace, and $L : V \rightarrow H$ a skew-symmetric linear operator. Then L is skew-adjoint if and only if for any $c \in \mathbf{R} \setminus \{0\}$, $L + cI$ is surjective.*

Proof. Suppose first that L is skew-symmetric and $L + cI$ is surjective for every $c \in \mathbf{R} \setminus \{0\}$. Since the domain of L is dense, for any such c ,

$$\ker(L^* + cI) = \text{rng}(L + cI)^\perp = \{0\}.$$

Choose $x \in \mathcal{D}(L^*)$, and $c \in \mathbf{R}$, $c \neq 0$. Then there must be $y \in V$ so that $(L - cI)y = (L^* + cI)x$. However, since L is skew-symmetric, $V \subset \mathcal{D}(L^*)$ so $(L - cI)y = -(L^* + cI)y = (L^* + cI)x$. However, we just saw that $L^* + cI$ is injective, so $x = -y \in V$. Thus $V = \mathcal{D}(L^*)$ whence L is skew-adjoint.

Conversely, if L is skew-adjoint, there can be no nonzero solutions to $(L + cI)x = 0$ with nonzero c , as follows from skew-symmetry. Hence there are no nonzero solutions to $(L^* + cI)x = 0$ with nonzero c , since $L = L^*$. That is, the kernel of $L^* + cI$ is trivial for nonzero c , whence the range of $L + cI$ is dense for any nonzero c . However L is closed (see for instance (Conway 1990), proposition X.1.6, p 305), so its range is closed (this is proved just as is a similar fact for closed symmetric operators, see Conway (1990), proposition X.2.5, p 310) thus $\text{rng}(L + cI) = H$. \square

Proposition 4. *The differential operator $p(\nabla)$ with domain $C_0^\infty(\Omega)^4$, defined formally by (8), extends to a skew-adjoint operator P on H with dense domain*

$$V = H_0^1(\Omega) \times H_{\text{div}}^1(\Omega).$$

Remark. The Hilbert space $H_{\text{div}}^1(\Omega)$ is the dense subspace of $L^2(\Omega)^3$ obtained by completing $C^1(\Omega)^3$ in the graph norm of the divergence operator. Equivalently, $\mathbf{v} \in L^2(\Omega)^3$ belongs to $H_{\text{div}}^1(\Omega)$ if and only if there exists $C \geq 0$ so that for every $\phi \in H_0^1(\Omega)$,

$$|(\nabla\phi, \mathbf{v})_{L^2(\Omega)^3}| \leq C\|\phi\|_{L^2(\Omega)}.$$

Proof. From the definitions, $p(\nabla)$ extends to $P : V \rightarrow L^2(\Omega)^4$. According to lemma 6, it suffices to show that for $(q, \mathbf{w})^T \in H$, $c \in \mathbf{R} \setminus \{0\}$, there exists $(p, \mathbf{v})^T \in V$ for which $P(p, \mathbf{v})^T = (q, \mathbf{w})^T$, that is,

$$\begin{aligned} -\nabla p + c\mathbf{v} &= \mathbf{w}, \\ -\nabla \cdot \mathbf{v} + cp &= q. \end{aligned} \tag{A.1}$$

The first equation is equivalent to the requirement that for $\phi \in H_0^1(\Omega)$,

$$-\langle \nabla \phi, \nabla p \rangle + c \langle \nabla \phi, \mathbf{v} \rangle = \langle \nabla \phi, \mathbf{w} \rangle.$$

To satisfy the second equation in (A.1), $\mathbf{v} \in H_{\text{div}}^1(\Omega)$ necessarily, and

$$\langle \nabla \phi, \mathbf{v} \rangle = -\langle \phi, \nabla \cdot \mathbf{v} \rangle = \langle \phi, q - cp \rangle.$$

Thus (A.1) implies that

$$-\langle \nabla \phi, \nabla p \rangle - c^2 \langle \phi, p \rangle = \langle \nabla \phi, \mathbf{w} \rangle - c \langle \phi, q \rangle. \tag{A.2}$$

The left-hand side of (A.2) defines a bounded and coercive (negative-definite) form on $H_0^1(\Omega)$ for any nonzero c , and the right hand side defines a continuous linear form on the same Hilbert space. The Lax–Milgram theorem (Yosida (1996), III.7) implies the existence of a unique $p \in H_0^1(\Omega)$ for which (A.2) holds for every $\phi \in H_0^1(\Omega)$.

With this choice of p , we are required to solve the second of the two conditions (A.1). In fact, a solution $\mathbf{v}_0 \in H^1(\Omega)^3 \subset H_{\text{div}}^1(\Omega)$ exists satisfying

$$\|\mathbf{v}_0\|_{H^1(\Omega)^3} \leq C \|q - cp\|_{L^2(\Omega)}$$

with a constant C depending only on Ω . For a proof, see Brenner and Scott (2007), lemma 11.2.3, who also give references to other results for domains for polygonal, rather than smooth, boundaries.

Set $\mathbf{v}_1 = c^{-1}\nabla p - \mathbf{v}_0 + c^{-1}\mathbf{w} \in L^2(\Omega)^3$. Then for any $\phi \in H_0^1(\Omega)$,

$$\begin{aligned} -c \langle \nabla \phi, \mathbf{v}_1 \rangle &= \langle \nabla \phi, -\nabla p + c\mathbf{v}_0 - \mathbf{w} \rangle \\ &= \langle \nabla \phi, -\nabla p - \mathbf{w} \rangle - c \langle \phi, \nabla \cdot \mathbf{v}_0 \rangle \\ &= \langle \nabla \phi, -\nabla p - \mathbf{w} \rangle + c \langle \phi, q - cp \rangle \\ &= 0, \end{aligned} \tag{A.3}$$

thanks to (A.2). That is, \mathbf{v}_1 is divergence-free, in the sense of distributions, and in particular $\mathbf{v}_1 \in H_{\text{div}}^1(\Omega)$. Now set $\mathbf{v} = \mathbf{v}_0 + \mathbf{v}_1$. Then the first of the two conditions in (A.1) is satisfied by (p, \mathbf{v}) thanks to the definition of \mathbf{v}_1 , whereas the construction of \mathbf{v}_0 and the divergence-free property of \mathbf{v}_1 imply that the second condition is also satisfied.

Thus we have constructed a solution of (A.1) in $V = H_0^1(\Omega) \times H_{\text{div}}^1(\Omega)$ for any $c \in \mathbf{R} \setminus \{0\}$, whence we conclude that P is skew-adjoint. \square

Appendix B. The differential case

If the memory term (convolution operator R) is absent, then initial data determine solutions uniquely. In this section, we sketch the theory, parallel to that for causal solutions, which holds in this differential case. We assume throughout this section that $Q \equiv 0$.

Corollary 11. *Suppose that $u \in L^2_{\text{loc}}(\mathbf{R}, H)$ is a weak solution of (13). Then $u \in C^0(\mathbf{R}, H)$.*

Proof. Choose $\phi \in C^\infty(\mathbf{R})$ so that $\phi(t) = 1$ for $t > 1$ (say), and $\phi(t) = 0$ for $t < -1$. Set $u_+ = \phi u$, $u_- = (1 - \phi)u$. It is straightforward to verify that u_+ is a causal solution of (13) with f replaced by $f + \phi' u \in L^2(\mathbf{R}, H)$, whence $u_+ \in C^0(\mathbf{R}, H)$ according to corollary 4. Likewise $t \mapsto u_-(-t)$ is also a causal solution of (13) with D, B replaced by $-P, -B$, and f replaced by $t \mapsto -f(-t) + \phi' u(-t)$, which also belongs to $L^2(\mathbf{R}, H)$. Thus u_- is also continuous, but $u = u_+ + u_-$. \square

Corollary 12. *Suppose that u is a weak solution of (13). Then for any $s \leq t \in \mathbf{R}$,*

$$E(t) \leq E(s) + C \int_s^t \|f\|^2.$$

Proof. The energy identity (42) applies to weak solutions, causal or not. Take into account $R = 0$, and use Gronwall's inequality, the boundedness of B , and the equivalence (26) of the energy with the norm in H . \square

Corollary 13. *If u_1 and u_2 are weak solutions of (13) (with the same right-hand side $f \in L^2(\mathbf{R}, H)$), and $u_1(s) = u_2(s)$ for some $s \in \mathbf{R}$, then $u_1 \equiv u_2$.*

Proof. Set $u = u_1 - u_2$: u is a weak solution with right-hand side $f = 0$, and $u(s) = 0$. The result follows immediately from corollary 12. \square

Theorem 10. *Suppose that $T_0 \in \mathbf{R}$ and $u_0 \in H$. Then there exists a unique weak solution of (13) for which $u(T_0) = u_0$.*

Proof. A proof of this result is precisely analogous to the proof of theorem 5: the solution is approximated by a Galerkin procedure and the solution of systems of ordinary differential equations, and the energy in the error estimated (in this instance, via corollary 12). \square

Results precisely analogous to those established in the last section hold concerning regular dependence on the coefficient operators for weak solutions with specified initial data and right-hand side. We leave the reader to formulate these results, whose proofs are minor variants of those given above.

References

- Awawdeh F 2010 Perturbation method for abstract second-order inverse problems *Nonlinear Anal.* **72** 1379–86
- Bamberger A, Chavent G and Lailly P 1979 About the stability of the inverse problem in 1-d wave equation—application to the interpretation of seismic profiles *Appl. Math. Optim.* **5** 1–47
- Bao G and Symes W W 1991 A trace theorem for solutions of linear partial differential equations *Math. Methods Appl. Sci.* **14** 553–62
- Bao G and Symes W W 1993 Trace regularity for a second order hyperbolic equation with nonsmooth coefficients *J. Math. Anal. Appl.* **174** 370–89
- Bao G and Symes W W 1996 On the sensitivity of solutions of hyperbolic equations to the coefficients *Commun. Partial Diff. Eqns* **21** 395–422
- Beals M and Reed M 1982 Propagation of singularities for hyperbolic pseudodifferential operators with nonsmooth coefficients *Commun. Pure Appl. Math.* **35** 169–84

- Beals M and Reed M 1984 Microlocal regularity theorems for nonsmooth pseudo-differential operators and applications to nonlinear problems *Trans. Am. Math. Soc.* **285** 159–84
- Biondi B and Almomin A 2012 Tomographic full waveform inversion (TFWI) bin combining full waveform inversion with wave-equation migration velocity analysis *82nd Annu. Int. Meeting Expanded Abstract SI9.5* (Tulsa, OK: Society of Exploration Geophysicists)
- Bourbie T, Coussy O and Zinszner B 1987 *Acoustics of Porous Media* (Paris: Editions TECHNIP)
- Brenner S and Scott L R 2007 *The Mathematical Theory of Finite Element Methods* 3rd edn (Berlin: Springer)
- Choulli M 1991 An abstract inverse problem *J. Appl. Math. Stoch. Anal.* **4** 117–28
- Christensen R M 1983 *Theory of Viscoelasticity: An Introduction* 2nd edn (New York: Academic)
- Coddington E A and Levinson N 1955 *Theory of Ordinary Differential Equations* (Malabar, FL: Krieger)
- Conway J 1990 *A Course in Functional Analysis* (Berlin: Springer)
- Courant R and Hilbert D 1962 *Methods of Mathematical Physics, volume II* (New York: Wiley-Interscience)
- Epanomeritakis I, Akcelik V, Ghattas O and Bielak J 2009 A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion *Inverse Problems* **24** 034015
- Fernandez-Berdaguer E M, Santos J E and Sheen D 1996 An iterative procedure for estimation of variable coefficients in a hyperbolic system *Appl. Math. Comput.* **76** 213–50
- Gauthier O, Tarantola A and Virieux J 1986 Two-dimensional nonlinear inversion of seismic waveforms *Geophysics* **51** 1387–403
- Kaltenbacher B, Neubauer A and Scherzer O 2008 *Iterative Regularization Methods for Nonlinear Problems* (Berlin: de Gruyter)
- Lasiecka I 1986 Sharp regularity results for mixed hyperbolic problems of second order *Differential Equations in Banach Spaces (Lecture Notes in Mathematics vol 1223)* (Berlin: Springer)
- Lasiecka I and Trigianni R 1989 Trace regularity of the solutions of the wave equation with homogeneous boundary conditions and compactly supported data *J. Math. Anal. Appl.* **141** 49–71
- Lavrentiev M, Romanov V and Shishatskii S 1986 *Ill-Posed Problems of Mathematical Physics* (Providence, RI: American Mathematical Society)
- Lax P 2006 *Hyperbolic Partial Differential Equations* (Providence, RI: American Mathematical Society)
- Lewis R M and Symes W W 1991 On the relation between the velocity coefficient and boundary values for solutions of the one-dimensional wave equation *Inverse Problems* **7** 597–632
- Lions J-L 1971 *Optimal Control of Systems Governed by Partial Differential Equations* (Berlin: Springer)
- Lions J-L and Magenes E 1972 *Non-Homogeneous Boundary Value Problems and Applications* (Berlin: Springer)
- Lorenzi A and Ramm A 2001 Some identification problems for integro-differential operator equations *Nonlinear Funct. Anal. Appl.* **6** 107–23
- Orlovsky D, Piskarev S and Spigler R 2010 On approximation of inverse problems for abstract hyperbolic equations *Taiwanese J. Math.* **14** 1145–67
- Pipkin A 1986 *Lectures on Viscoelasticity Theory* 2nd edn (Berlin: Springer)
- Plessix R-E 2006 A review of the adjoint-state method for computing the gradient of a functional with geophysical applications *Geophys. J. Int.* **167** 495–503
- Ramm A and Koshkin S 2001 An inverse problem for an abstract evolution equation *Appl. Anal.* **79** 475–82
- Salo M 2006 Stability for solutions of wave equations with $C^{1,1}$ coefficients *Inverse Problems Imaging* **1** 537–56
- Santosa F and Symes W W 1989 *An Analysis of Least Squares Velocity Inversion* (Tulsa, OK: Society of Exploration Geophysicists)
- Shen P and Symes W W 2008 Automatic velocity analysis via shot profile migration *Geophysics* **73** VE49–60
- Smith H 1998 A parametrix construction for wave equations with $C^{1,1}$ coefficients *Ann. l'institute Fourier (Grenoble)* **48** 797–835
- Stefanof P and Uhlmann G 2009 Linearizing nonlinear inverse problems and its application to inverse backscattering *J. Funct. Anal.* **256** 2842–66
- Stolk C 2000 On the modeling and inversion of seismic data *PhD Thesis* (The Netherlands: Universiteit Utrecht, Utrecht)
- Stolk C C, de Hoop M V and Symes W W 2009 Kinematics of shot-geophone migration *Geophysics* **74** WCA18–34
- Sylvester J and Uhlmann G 1990 The Dirichlet to Neumann map and applications *Inverse Problems in Partial Differential Equations* (Philadelphia: SIAM) pp 101–39
- Symes W W 1983 Trace theorem for solutions of the wave equation and the remote determination of acoustic sources *Math. Methods Appl. Sci.* **5** 131–52
- Symes W W 1986 On the relation between coefficient and boundary values for solutions of Webster's horn equation *SIAM J. Math. Anal.* **17** 1400–20
- Symes W W 2008 Migration velocity analysis and waveform inversion *Geophys. Prospect.* **56** 765–90
- Taylor M 1981 *Pseudodifferential Operators* (Princeton: Princeton University Press)

- Virieux J and Operto S 2009 An overview of full waveform inversion in exploration geophysics *Geophysics* **74** WCC127–52
- Walden A and Hosken J 1986 The nature of the non-Gaussianity of primary reflection coefficients and its significance for deconvolution *Geophys. Prospect.* **34** 1038–66
- White B, Sheng P and Nair B 1990 Localization and backscattering spectrum of seismic waves in stratified lithology *Geophysics* **55** 1158–65
- Yosida K 1996 *Functional Analysis* (Berlin: Springer)