ESAIM: M2AN 51 (2017) 727–757
DOI: 10.1051/m2an/2016038

ESAIM: Mathematical Modelling and Numerical Analysis
www.esaim-m2an.org

UNIFORM DISCRETE SOBOLEV ESTIMATES OF SOLUTIONS TO FINITE DIFFERENCE SCHEMES FOR SINGULAR LIMITS OF NONLINEAR PDES

Liat Even-Dar Mandel[1,2] and Steven Schochet[2]

Abstract. Uniform discrete Sobolev space estimates are proven for a class of finite-difference schemes for singularly-perturbed hyperbolic-parabolic systems. The estimates obtained improve previous results even when the PDEs do not involve singular perturbations. These estimates are used in a companion paper to prove the convergence of solutions as the discretization parameter and/or the singular perturbation parameter tends to zero.

Mathematics Subject Classification. 65M10.

Received July 9, 2015. Revised February 7, 2016, Accepted May 16, 2016.

1. Introduction

This paper initiates a theory for finite difference schemes analogous to the theory of singular limits of systems of PDEs such as

$$A(\varepsilon u)u_t = \sum_{j=1}^{d} A^j(t,x,u)u_{x_j} + \sum_{j,k=1}^{d} \partial_{x_j}\left(B^{j,k}\partial_{x_k}u\right) + \frac{1}{\varepsilon}\left(\sum_{j=1}^{d} C^j u_{x_j} + Du\right) + F(t,x,u) \tag{1.1}$$

[14, 17, 20, 26, 33]. When the matrices A, A^j, and C^j are symmetric, D is anti-symmetric, the matrices $B^{j,k}$, C^j and D are constant, and assuming for simplicity that the second-order operator in which the $B^{j,k}$ appear either vanishes identically or is strongly elliptic, solutions of (1.1) with fixed smooth initial data exist for a time independent of ε, and the difference between those solutions and the solutions of certain limit or profile equations tends to zero with ε. More general assumptions on the second-order operator are presented in Section 2.2. In particular those assumptions are satisfied by the slightly-compressible barotropic Navier–Stokes equations (*e.g.* [34])

$$\begin{aligned}\frac{P'(\rho_0+\varepsilon r)}{\rho_0+\varepsilon r}[r_t + u\cdot\nabla r] + \frac{P'(\rho_0+\varepsilon r)}{\varepsilon}\nabla\cdot u = 0\\ (\rho_0+\varepsilon r)[u_t + (u\cdot\nabla)u] + \frac{P'(\rho_0+\varepsilon r)}{\varepsilon}\nabla r = \mu\nabla\cdot\nabla u + (\mu+\lambda)\nabla(\nabla\cdot u),\end{aligned}$$

Keywords and phrases. Uniform estimates, finite-difference methods, discrete Sobolev spaces, fully-discrete sharp Gårding inequality, singular limits.

[1] Department of Mathematics and Computer Science, Open University, 43107 Raanana, Israel.
[2] School of Mathematical Sciences, Tel Aviv University, 69978 Tel Aviv, Israel. schochet@post.tau.ac.il

Article published by EDP Sciences

which may be kept in mind as a typical system under consideration. Singular limits of equations of the form (1.1) and variants thereof occur not only in the original motivating example of slightly-compressible fluid dynamics [20, 26, 34] and its variants [5, 32, 46] but also in a variety of other fields (*e.g.*, [1, 7, 36]). It is therefore of much interest to obtain, and prove estimates for, numerical methods for equations of the form (1.1) whose accuracy is uniform in the parameter ε.

There is a vast literature on numerical methods for PDEs like (1.1) without the large terms, *i.e.*, with the C^j and D all vanishing, and for various special cases of such equations. The part of this literature that proves uniform bounds and convergence may be classified according to the corresponding bounds known for PDEs: BV or L^∞ bounds for scalar equations (*e.g.*, [6, 23, 30, 41]), small BV bounds for systems in one spatial variable (*e.g.*, [2, 4]), L^2 or Sobolev H^s bounds for constant-coefficient (*e.g.*, [10]) and variable-coefficient linear systems, and H^s bounds with a sufficiently high index s for smooth solutions of nonlinear systems [40, 44].

However, neither the well-known convergence result of Strang [40] for finite difference approximations to nonlinear evolution equations nor the somewhat improved result of Tomoeda [44] yield uniform bounds for discretizations of (1.1) when large terms are present. This difficulty is inherent in their analysis, which makes use of higher-order time derivatives of the PDE solution that are generally not uniformly bounded. Nevertheless, even in the absence of any theory guaranteeing uniform bounds and convergence, much progress has been made in designing and analyzing schemes whose leading formal asymptotics when the parameter ε tends to zero is consistent with the asymptotics of the PDE, including some first steps towards the stability analysis of such schemes (see [8, 21, 22, 28, 29, 37] and their references). One of the first conclusions of this research was that the large terms must be treated implicitly, since for explicit schemes the CFL condition mandates taking time steps Δt no larger than $O(\varepsilon\Delta x)$, where for simplicity the spatial grid is assumed to be of size $O(\Delta x)$ in all directions.

In this paper we prove discrete Sobolev norm bounds that are independent of both ε and the discretization parameters for solutions to two classes of finite difference approximations to equations of the form (1.1) satisfying the ab ove-mentioned assumptions, one very specific and the other fairly general. As far as we know these are the first uniform bounds obtained for such approximations.

The only finite difference approximation for which uniform discrete bounds can be obtained by following the discrete analogue of every step of the energy estimates used to obtain uniform bounds for the PDE is the Crank–Nicolson scheme, in which the time derivative term u_t in (1.1) is approximated by a forward time difference $\frac{u(t+\Delta t,x)-u(t,x)}{\Delta t}$, all other appearances of the dependent variable u are approximated by the time average $\frac{u(t+\Delta t,x)+u(t,x)}{2}$, and the spatial derivative operators are approximated by arbitrary central differences. We prove here uniform bounds not only for the Crank–Nicolson scheme but also for the θ-scheme in which the equally-weighted time average is replaced by $\theta u(t+\Delta t,x)+(1-\theta)u(t,x)$ with $\theta\in[\frac{1}{2},1]$, without requiring any relationship between the time discretization parameter Δt and the spatial discretization parameter Δx. Some estimates in the discrete ℓ^2 norms defined in Section 2.1 have been obtained previously for the θ-scheme for linear systems without large terms: for the constant-coefficient case without assuming the symmetry of the A^j but with the ratio $\frac{\Delta t}{(\Delta x)^2}$ assumed to be constant [16], for the cases $\theta=\frac{1}{2}$ and $\theta=1$ ([18], Thms. 5.3.2–3), and for other values of θ ([31], Sects. 5.1.8 and 6.2) using scheme-dependent norms.

In order to obtain uniform bounds for a wider class of numerical schemes we treat the large terms, the remaining first-order terms, and the higher-order terms separately from each other. The large terms are treated purely implicitly, which together with their anti-symmetry ensures that they drop out of the discrete energy estimates obtained by taking the inner product of the scheme with $u(t+\Delta t,x)$. This key step towards obtaining bounds uniform in ε mirrors the dropping out of the large terms from energy estimates for the PDE (1.1), first observed by Klainerman and Majda [20]. The second-order terms of the PDE (1.1) will only be assumed to be weakly dissipative in the sense that they do not increase the L^2 norm of solutions, and the PDE will be also be allowed to contain even higher-order terms having the same property. All such terms will be approximated by differences that do not increase an appropriate discrete norm. For dissipative terms of the PDE this leads to numerical approximations of the second-order and higher-order terms similar to but more general than those in the θ-scheme. For purely dispersive terms of the PDE the numerical approximation must be purely implicit as in the case $\theta = 1$ of the θ-scheme in order to be dispersive at the discrete level. The main increase in generality comes from the treatment of the first-order terms that are independent of ε, which are treated purely explicitly: an estimate for that part of the scheme is obtained *via* a fully discrete and nonlinear version of Lax and Nirenberg's sharp Gårding inequality [24], which only requires a condition on the symbol of the operator that describes the scheme. Precise statements of the results including basic examples of discretizations are presented in Section 2, and the results are shown in Section 3 to apply to a representative sample of specific schemes.

Previous results for general schemes for systems like (1.1) without large terms include discrete ℓ^2 estimates for linear systems under conditions on the eigenvalues of the spatial difference operator of the finite difference scheme, without assuming the symmetry of the A^j but with the ratio $\frac{\Delta t}{(\Delta x)^2}$ assumed to be constant [27,38,39,47], estimates in continuous L^2 norms for linear systems without second-order or higher-order terms whose spatial operator satisfies a sharp Gårding inequality [24], and the above-mentioned results for nonlinear systems [40,44] discussed further below. The results here could also be extended to the case when the A^j are not symmetric if $\Delta t = O((\Delta x)^2)$, but a more refined analysis to be presented elsewhere will relax that restriction.

Because of the separate treatment of the various terms, the schemes other than Crank–Nicolson presented here for equations involving large terms or higher-order terms are limited to being first order in time. For equations with large terms this seems to be an inherent limitation. More refined estimates needed to allow first and higher-order terms to be treated together in equations without large terms will be presented elsewhere. Nevertheless, the bounds obtained here improve on earlier results [40,44] even for systems without large terms. First, [40,44] only analyze explicit schemes for first-order hyperbolic systems, and although [40] mentions that the results extend to systems like (1.1) containing second-order terms, the approach suggested there, following [19], requires the usual assumption that $\frac{\Delta t}{(\Delta x)^2}$ be constant. That restriction is avoided here by treating the second-order and higher-order terms implicitly or at least semi-implicitly. Second, by directly estimating discrete Sobolev norms of solutions to finite difference schemes, we show that solutions have as many discrete Sobolev derivatives as the initial data. For equations of the form (1.1) the bounds implicit in the above-mentioned results lose $2\lfloor d/2\rfloor + 2$ derivatives, where d is the spatial dimension. The improved bounds here are used in [9] to show that for schemes without large terms the rate of convergence equals the order of accuracy of the scheme under less restrictive conditions on the smoothness of the PDE solution than in [40]. In addition, the stability condition developed here is easier to apply than the one in [40].

As for the PDE itself, the uniform bounds proven here imply the convergence of solutions as ε or $\Delta x + \Delta t$ or both tend to zero, albeit without a rate. Moreover, the norm in which solutions converge is strong enough to ensure that the limit of the solutions satisfies the relevant limit equation. The details of the convergence argument are given in the companion paper [9], where the solutions constructed here are also shown under additional assumptions to converge uniformly in ε at the rate $O((\Delta t)^{1/2})$.

The framework required for our method includes discrete versions of a variety of calculus inequalities that are well-known in the continuous case, including the Sobolev embedding estimate, certain Gagliardo–Nirenberg inequalities and Moser estimates, and the sharp Gårding inequality. Certain special cases of these estimates have been shown previously: discrete versions of Gagliardo–Nirenberg-Sobolev inequalities involving only derivatives through first order were proven for general meshes used in finite-volume schemes in [3, 15] and references therein, but we require such estimates involving arbitrary orders. A wide variety of discrete Gagliardo–Nirenberg inequalities and certain cases of the discrete Sobolev embedding estimate have been proven by Zhou [48–50] even for nonuniform meshes, but in the case of multiple space dimensions the results are restricted to particular values of the index p in the discrete ℓ^p norms, whereas we require all values of p in $[2, \infty]$. Discrete sharp Gårding inequalities were proven for pseudo-difference operators on half-spaces in ([27], Sect. 4) and references therein, but their translation to the periodic case is not equivalent to the estimates used here. Since the estimates available in the literature do not include all those needed here, the required estimates will be derived briefly in Section 4 for the rectangular meshes used in this paper, following as much as possible the methods used for the original continuous versions, which yield simpler and more uniform proofs. To simplify the notation we will only consider periodic meshes; the case of infinite meshes is similar.

The main results will be proven in Section 5. Various applications and extensions will be considered elsewhere. In particular, the results here and in [9] will be used in [35] to prove the stability and convergence to smooth solutions of upwind finite-volume schemes on uniform rectangular grids that use actual or approximate Riemann solvers.

2. Notation and statement of results

The estimates for the θ-scheme follow closely the estimates for the PDEs themselves, so the conditions for that scheme will be stated in terms of the coefficients of the PDE. The conditions under which estimates can be obtained *via* a sharp Gårding inequality will be stated in terms of the scheme itself. Before describing those schemes we recall some notations for shift and difference operators and some discrete Sobolev spaces that will be needed for both cases.

2.1. Notation and spaces

Let e_j denote the vector whose component j equals one and other components equal zero. The forward and backward shift operators in the jth coordinate are defined by $[S_{j,\Delta x}u](x) := u(x+\Delta x e_j)$ and $[(S_{j,\Delta x})^{-1}u](x) := u(x-\Delta x e_j)$, respectively, and the forward difference operator in that coordinate is $D_{j,\Delta x} := \frac{S_{j,\Delta x}-1}{\Delta x}$. Note that in operator formulas a number denotes the operator of multiplication by that number; these are all scalar operators, but will be extended to operate on vectors componentwise. Higher-order shift operators are defined by $S^{\alpha}_{\Delta x} := (S_{1,\Delta x})^{\alpha_1}\dots(S_{d,\Delta x})^{\alpha_d}$, where α is a multi-index vector with integer components, and higher-order difference operators are defined by $D^{\alpha}_{\Delta x} := (D_{1,\Delta x})^{\alpha_1}\dots(D_{d,\Delta x})^{\alpha_d}$, where α is a multi-index vector with nonnegative integer components. In examples presented in one spatial dimension the index j will be omitted from both shifts and difference operators.

Similarly, the forward time-shift operator and forward time-difference operator are defined by $[S_{\Delta t}u](t,x) = u(t+\Delta t, x)$ and $D_{\Delta t} := \frac{S_{\Delta t}-1}{\Delta t}$, respectively. For the θ-scheme we will also need the θ-averaging operator $u^{\theta} := [\theta S_{\Delta t} + (1-\theta)]u$.

A general first-order difference operator $\partial_{j,\Delta x} := \frac{1}{\Delta x}\sum_{|m|\le M} c_m (S_{j,\Delta x})^m$ in the direction j is assumed to have real coefficients and to be an approximation to the differential operator ∂_{x_j} satisfying $\partial_{j,\Delta x}u = \partial_{x_j}u + O(\Delta x)$ for any $u \in C^2$, which is equivalent to the conditions $\sum_{|m|\le M} c_m = 0$ and $\sum_{|m|\le M} m c_m = 1$. A central first-order difference operator $\partial_{j,\Delta x,c}$ in the direction j is a first-order difference operator in that direction satisfying $c_{-m} = -c_m$. Any first-order difference operator in the direction j can be written as a linear combination $\partial_{j,\Delta x} = \left[\sum_{|m|\le M} d_m (S_{j,\Delta x})^m\right] D_{j,\Delta x}$ of shift operators applied to the forward difference operator in that direction. However, it will be more convenient to write a central difference operator in the direction j as a linear combination $\partial_{j,\Delta x,c} = \sum_{0<m\le M} d_m D_{j,m\Delta x,c}$ of the basic central difference operators $D_{j,m\Delta x,c} := \frac{(S_{j,\Delta x})^m - (S_{j,\Delta x})^{-m}}{2m\Delta x}$ in that direction. The length $|\alpha|$ of a multi-index is the sum $\sum_j |\alpha_j|$ of the absolute values of its components. We will let $P_{\Delta x} := \sum_{|\alpha|\le M} P_\alpha S^{\alpha}_{\Delta x}$, and similarly $Q_{\Delta x}$, $G_{\Delta x}$, *etc.*, denote general spatial shift operators.

Example 2.1. The difference scheme $S_{\Delta t}u = [G_{\Delta x}(u)]u$ with $G_{\Delta x}(u) = 1 + \frac{\Delta t}{\Delta x}a(u)(S_{\Delta x} - 1)$ is an approximation of the PDE $u_t = a(u)u_x$. A more complicated scheme that is second order in both space and time can be obtained by combining the central-difference approximation $\frac{S_{\Delta x} - (S_{\Delta x})^{-1}}{2\Delta x}$ to the derivative operator ∂_x with midpoint time-stepping. Defining $\lambda := \frac{\Delta t}{\Delta x}$ and $u^{\text{mid}} := u + \frac{\lambda}{4}a(u)\left(S_{\Delta x} - (S_{\Delta x})^{-1}\right)u$, the resulting scheme can be written as $S_{\Delta t}u = G_{\Delta x}(\Delta t, \Delta x, S_{\Delta x}u, u, (S_{\Delta x})^{-1}u)u$ with

$$\begin{aligned} &G_{\Delta x}(\Delta t, \Delta x, S_{\Delta x}u, u, (S_{\Delta x})^{-1}u) \\ &\quad = \left[1 + \tfrac{\lambda}{2}a(u^{\text{mid}})\left(S_{\Delta x} - (S_{\Delta x})^{-1}\right) + \tfrac{\lambda^2}{8}a(u^{\text{mid}})\left\{a(S_{\Delta x}u)\left(S^2_{\Delta x} - 1\right) - a((S_{\Delta x})^{-1}u)\left(1 - S^{-2}_{\Delta x}\right)\right\}\right]. \end{aligned} \tag{2.1}$$

The adjoint $S^*_{j,\Delta x}$ of the shift operator $S_{j,\Delta x}$ equals $(S_{j,\Delta x})^{-1}$, no matter whether that adjoint is taken with respect to the continuum L^2 inner product or the discrete ℓ^2 inner product defined below. Together with the standard results that $(P+Q)^* = P^* + Q^*$ and $(PQ)^* = Q^*P^*$, this allows us to calculate the adjoints $\partial^*_{j,\Delta x}$ and $P^*_{\Delta x}$ of any difference or shift operator. In particular, a central difference operator is anti-symmetric.

The symbol of a spatial shift or difference operator is defined by

$$\left[\mathrm{Symb}\left(\sum_\alpha c_\alpha(x) S^\alpha_{\Delta x}\right)\right](\xi) := \sum_\alpha c_\alpha(x) \mathrm{e}^{i\Delta x \alpha\cdot\xi}. \tag{2.2}$$

Besides occurring in difference schemes, difference operators also appear in formulas for discrete Sobolev norms in place of the differential operators in ordinary continuum Sobolev norms. For concreteness we will define discrete norms in the periodic case, in which the domain of the spatial variables is $X_L := [-L, L)^d$ with the endpoints in each direction identified. The discrete domain $X_{\Delta x} = [-L_{\Delta x}, L_{\Delta x})^d \cap \Delta x \mathbb{Z}^d$, where $L_{\Delta x} := \Delta x \left\lfloor \frac{L}{\Delta x} \right\rfloor$, is also taken periodic, *via* the rule that if x belongs to $X_{\Delta x}$ but $x \pm \Delta x e_j$ does not, then $x \pm \Delta x e_j$ is taken to equal $x \mp m\Delta x e_j$ where m is the largest integer for which the latter point belongs to $X_{\Delta x}$. The cases when the periods of different components differ, some or all of the spatial components lie in $\mathbb{R}$, or the distances between grid points of different components differ requires for the most part only notational adjustments. The discrete ℓ^2 norm is $\|v\|_{\ell^2} := \sqrt{\langle v, v\rangle_{\ell^2}}$, where $\langle v, w\rangle_{\ell^2} := \sum_{x\in X_{\Delta x}} v(x)\cdot w(x)\ (\Delta x)^d$ is the discrete ℓ^2 inner product. The argument $X_{\Delta x}$, as in $\ell^2(X_{\Delta x})$, will sometimes be added to distinguish these from the analogous norm and inner product on Fourier space. Similarly to the PDE case, when the matrix coefficient A of the time-difference operator is not simply the identity matrix then the ℓ^2 estimate will usually not be obtained directly in the ℓ^2 norm but in the time-varying equivalent norm $\|v\|_{\ell^2_A} := \sqrt{\langle v, Av\rangle_{\ell^2}}$. The discrete Sobolev norms are then defined for nonnegative integers s by $\|u\|_{h^s_A} := \sqrt{\langle u, u\rangle_{h^s_A}}$, where $\langle u, v\rangle_{h^s_A} := \sum_{|\alpha|\le s} \langle D^\alpha_{\Delta x} u, A D^\alpha_{\Delta x} v\rangle_{\ell^2}$ is the discrete Sobolev inner product, any points in the formula for $D^\alpha_{\Delta x}$ that lie outside the discrete domain $X_{\Delta x}$ are understood in the periodic sense defined above, and the positive-definite matrix A is omitted from the inner product if it is omitted from the notation for the norm. Since the sum in the inner product defining the h^s_A norm consists of nonnegative terms, includes the ℓ^2 norm as the case $\alpha = 0$ of the sum, and involves differences of order at most s, there exists a constant $C(s, d, \Delta_0)$ such that

$$\|u\|_{\ell^2} \le \|u\|_{h^s} \le \tfrac{C(s,d,\Delta_0)}{(\Delta x)^s} \|u\|_{\ell^2} \tag{2.3}$$

for $0 < \Delta x \le \Delta_0$ where d denotes the spatial dimension. The discrete ℓ^∞ norm is $\|u\|_{\ell^\infty} = \sup_{x\in X_{\Delta x}} |u(x)|$. The discrete version of Sobolev's theorem, which is a particular case of Lemma 4.1 below, says that there exists a constant $c(d)$ independent of Δx such that

$$\|u\|_{\ell^\infty} \le c(d) \|u\|_{h^\sigma}, \tag{2.4}$$

where $\sigma := \lfloor d/2 \rfloor + 1$ is called the Sobolev embedding exponent.

2.2. Results for the θ-scheme

The basic form of θ-schemes for the PDE (1.1) is

$$\begin{aligned} A(\varepsilon u) D_{\Delta t} u = \sum_{j=1}^d A^j(t, x, u^\theta) \partial_{j,\Delta x,c} u^\theta - \sum_{j,k=1}^d (\partial_{j,\Delta x})^* \left(B^{j,k} \partial_{k,\Delta x} u^\theta\right) \\ + \tfrac{1}{\varepsilon}\left(\sum_{j=1}^d C^j \partial_{j,\Delta x,c} u^\theta + D u^\theta\right) + F(t, x, u^\theta), \end{aligned} \tag{2.5}$$

where $\theta \in [\frac{1}{2}, 1]$ is a parameter. Since our assumptions will allow the matrices $B^{j,k}$ to be identically zero, (2.5) includes schemes for first-order systems as a special case; systems with higher-order spatial derivative terms can be handled similarly under appropriate assumptions.

The assumptions needed on the coefficients of system (2.5) are mostly the same as those used to obtain the simplest analogous results for the PDE system (1.1). First, all the coefficients are assumed to be smooth in all variables and uniformly bounded in the independent variables, *i.e.*, for sufficiently large s,

$$\sum_{|\alpha|\le s} |D^{\alpha}_{(t,x,u)}\left(A(u), \{A^j(t,x,u)\}_{j=1}^d, F(t,x,u)\right)| \le M_s(|u|), \tag{2.6}$$

where for any vector w the derivative operator D^α_w means $\prod_j \partial_{w_j}^{\alpha_j}$, $|\,|$ denotes the norm of a vector or matrix, not summed or integrated over the spatial variables, and both here and later all bounds depending on variables are assumed for simplicity to be continuous and nondecreasing.

Second, the first-order terms are assumed to form a symmetric-hyperbolic system, which means that

$$\text{the matrices } A\text{, the } A^j\text{, and the } C^j \text{ are symmetric} \tag{2.7}$$

and A is positive definite, *i.e.*, satisfies

$$A(w) \ge \tfrac{1}{m_0(|w|)} I. \tag{2.8}$$

Third, the terms of order $\frac{1}{\varepsilon}$ and the second-order terms in (1.1) will be assumed to have constant coefficients, *i.e.*,

$$\text{the matrices } B^{j,k},\ C^j\text{, and } D \text{ are constant.} \tag{2.9}$$

Fourth, the large operator $\sum C^j\partial_{x_j} + D$ in (1.1) will be assumed to be antisymmetric, *i.e.*

$$\left(C^j\right)^T = C^j, \quad D^T = -D. \tag{2.10}$$

The reason for repeating the condition $\left(C^j\right)^T = C^j$, which already appeared in (2.7), will be explained in Remark 2.3.

Fifth, the second-order operator in (1.1) will assumed to satisfy a somewhat weaker condition than ellipticity. Specifically, either

$$\sum_{j,k} w^j \cdot B^{j,k} w^k \ge b_0 \sum_j |w^j|^2 \qquad \text{with } b_0 \ge 0, \tag{2.11}$$

or else both

$$\sum_{j,k} \xi_j\xi_k \frac{B^{j,k} + (B^{j,k})^T}{2} \ge b_0|\xi|^2 I \qquad \text{with } b_0 \ge 0 \tag{2.12}$$

and

$$\text{the difference operators } \partial_{k,\Delta x} \text{ in the terms in (2.5) involving } B^{j,k} \text{ are central.} \tag{2.13}$$

Condition (2.11) implies (2.12), but the matrices $B^{1,1} = \left(\begin{smallmatrix}5&0\\0&1\end{smallmatrix}\right)$, $B^{1,2} = B^{2,1} = \left(\begin{smallmatrix}0&2\\2&0\end{smallmatrix}\right)$, and $B^{2,2} = \left(\begin{smallmatrix}1&0\\0&5\end{smallmatrix}\right)$ provide an example in which (2.12) is satisfied but (2.11) is not. By definition, the second-order terms of (1.1) are strongly elliptic when (2.12) holds with $b_0 > 0$. The case $b_0 = 0$ has been allowed not only to cover the case of hyperbolic systems having no second-order terms, but also to include systems having purely dispersive second-order terms. For example, the nonlinear Schrödinger equation and the Zakharov equations can be written as real systems of the form (1.1) with $b_0 = 0$ (see [36] and Sect. 3.3 below). When (2.12) holds but (2.11) does not then the additional condition (2.13) is needed to enable the terms involving the $B^{j,k}$ in (2.5) to be estimated in Fourier space.

Finally, because we have not assumed that b_0 in (2.11)–(2.12) is strictly positive, just as for PDEs we need to assume that

$$\text{either } A \text{ is constant or the } B^{j,k} \text{ all vanish.} \tag{2.14}$$

The above assumptions will suffice to obtain a uniform h^s estimate for solutions of (2.5) having uniformly bounded initial data $u_{0,\varepsilon,\Delta x}$. However, in order to obtain a uniform estimate for the time difference $D_{\Delta t}u$ as well it is necessary to assume in addition that the initial time difference $D_{\Delta t}u\big|_{t=0}$ is bounded uniformly in ε. That condition is not automatic on account of the $O(\frac{1}{\varepsilon})$ term in the difference scheme, but will hold provided that

$$\left\|\left(\sum C^j\partial_{j,\Delta x,c}+D\right)u_{0,\varepsilon,\Delta x}\right\|_{h^r}\le c\varepsilon \tag{2.15}$$

for an appropriate value of r. As in the PDE case, such initial data will be called well prepared.

Unlike the results in [16,38,39,47], the following theorem does not require that $\frac{\Delta t}{(\Delta x)^2}$ be bounded.

Theorem 2.2. *Let s be an integer greater or equal to $\sigma+2$, where $\sigma := \lfloor d/2\rfloor+1$ is the Sobolev embedding exponent. Assume that the above-mentioned conditions hold,* i.e., *(2.6) with this value of s, (2.7), (2.8), (2.9), (2.10), (2.14), and either (2.11) or both (2.12) and (2.13).*

Then for every K_0 and ε_0 there are Δ_0, K, and T such that for all $\varepsilon\in(0,\varepsilon_0]$, Δx in $(0,1]$, $\Delta t\in(0,\Delta_0]$, $\theta\in[\frac12,1]$, and initial value $u_{0,\varepsilon,\Delta x}$ satisfying $\sup_{\varepsilon\in(0,\varepsilon_0],\Delta x\in(0,1]}\|u_{0,\varepsilon,\Delta x}\|_{h^s}\le K_0$ there exists a unique solution of (2.5) plus initial condition

$$u(0)=u_{0,\varepsilon,\Delta x} \tag{2.16}$$

in $\ell^\infty([0,T]\cap\Delta t\mathbb{Z};h^s)$ satisfying $\sup_{t\in[0,T]\cap\Delta t\mathbb{Z}}\|u\|_{h^s}\le K$. The maximum allowed time step Δ_0, solution bound K, and guaranteed time of existence T depend only on K_0, ε_0, the bounds M_s and m_0 in (2.6)–(2.7), the smoothness parameter s and the dimension d.

If in addition (2.15) holds for some $r\le s-2$, then $D_{\Delta t}u$ is uniformly bounded in $\ell^\infty([0,T-\Delta t]\cap\Delta t\mathbb{Z};h^r)$.

Remark 2.3. The result of Theorem 2.2 can be generalized in various ways:

1. Essentially the same proof yields the following additional results:
   (a) The central difference operators multiplied by A^j in (2.5) may be different from those multiplied by C^j there.
   (b) Rather than depending solely on u^θ, the coefficients A^j and F in (2.5) may depend on an arbitrary average $u^\rho := [\rho S_{\Delta t}+(1-\rho)]u$, or even on an arbitrary finite set $\{S^\alpha_{\Delta x}u, S^\alpha_{\Delta x}S_{\Delta t}u\}_{|\alpha|\le M}$ of spatial shifts of u and its time shift $S_{\Delta t}u$, as long as A^j and F depend smoothly on all those variables. Dependence of coefficients on shifts of u is discussed further in the next subsection.
   (c) When the large terms are absent then the coefficient A of the time difference can also be allowed to depend on t and x as well as u, as long as the bound m_0 in (2.7) is independent of those variables.
   (d) When the $B^{j,k}$ all vanish then the minimum value of s is $\sigma+1$ rather than $\sigma+2$ and the maximum value of r in the last part of the theorem is $s-1$ instead of $s-2$.
   (e) The discrete spatial domain may be the discretization $\Delta x\mathbb{Z}^d$ of the whole space $\mathbb{R}^d$ provided that the inhomogeneous part $F(t,x,0)$ of F is bounded in h^s uniformly in Δx and t. For the periodic case considered in Theorem 2.2 that condition follows from the smoothness assumption on F.
2. When the second-order terms are strongly elliptic, *i.e.* b_0 in (2.11) or (2.12) is positive, then those terms contribute a favorable term to the energy estimates, which can be used to cancel various unfavorable terms. First, condition (2.14) is no longer needed, since the favorable term so obtained dominates the problematic term arising from (5.24) below. As for PDEs ([20], Thm. 4), condition (2.14) may also be omitted in the hyperbolic-parabolic case familiar from the compressible Navier–Stokes equations in which second-order terms only appear in certain equations and their restriction to those equations satisfies (2.11) or (2.12) with $b_0>0$.
   Second, when $b_0>0$ then the the first-order terms A^j do not need to be symmetric, but the coefficients C^j appearing in the large terms must still be symmetric to ensure the anti-symmetry of the large operator.

2.3. Results *via* sharp Gårding inequality

In order to obtain estimates for finite difference schemes other than the θ-scheme, we will write those schemes using shift operators rather than difference operators. Before defining the exact form of the schemes to be considered, let us consider various terms of the PDE (1.1) in order to elucidate what form the terms in the numerical scheme used to approximate them should have. However, these considerations are used only to motivate the form of the difference scheme; in particular, the assumptions actually made on the difference scheme allow terms that approximate partial differential operators of order higher than two and large terms involving differential operators of order higher than one. Such higher-order terms are, however, required to have the same properties as the one appearing in (1.1), *i.e.* the large terms are required to be antisymmetric and the higher-order terms that are independent of ε are required to not increase the energy.

The first-order terms in (1.1) are quasilinear, *i.e.*, their coefficients depend on the dependent variables, and hence the coefficients of the finite difference schemes will also depend on u in general. As the scheme in (2.1) illustrates, the form of the difference scheme approximating the first-order terms need not be simply the first-order terms from the PDE with derivatives replaced by differences as in the θ-scheme, but will still be related to the terms in the PDE. In accordance with the example shift operator in (2.1) we will allow the coefficients of shift operators that depend on u to also depend on any finite set of its spatial shifts. To keep the notation compact let us denote such a collection by

$$\widetilde{u} := \{S^{\alpha}_{\Delta x} u\}_{|\alpha|\le M}, \tag{2.17}$$

and use the abbreviation

$$\varDelta := \{\varDelta t, \varDelta x\}. \tag{2.18}$$

Another example of a scheme that requires shifts in coefficients is the fourth-order Runge–Kutta time-stepping scheme presented in Section 3.2. Although the PDE and forward time-shift in Example 2.1 are not multiplied by matrices, we will allow schemes in which the time shift is multiplied by a matrix $A(\varepsilon\widetilde{u})$, in accordance with the PDE (1.1). In similar fashion to the conditions for (2.5), the matrix A will be required to satisfy

$$(A(\varepsilon\widetilde{u}))^T = A(\varepsilon\widetilde{u}), \quad A(w) \ge \tfrac{1}{m_0(|w|)} I, \quad \sum_{|\beta|\le s} |\partial^{\beta}_w A(w)| \le M_s(|w|). \tag{2.19}$$

While we will not assume that the coefficients of the shift operator $G_{\Delta x}$ are symmetric matrices, in some applications the symmetry (Sect. 3.1) or at least symmetry up to $O((\varDelta t)^2)$ (Sect. 3.2) of the coefficients of $G_{\Delta x}$ can be used to show that the stability condition (2.26) below holds.

Another difference between (2.1) and the general scheme to be considered is that while the right side of (2.1) consists of u plus terms of order $\varDelta t$ arising from difference operators, in general we will allow an A-weighted average of shifts of u plus terms of order $\varDelta t$ arising from difference operators. Here and later, the factor $\varDelta t$ occurs on account of the shift-operator form in which we write the numerical scheme. Specifically, we assume that

$$G_{\Delta x} = \sum_{|\alpha|\le M} G_\alpha S^{\alpha}_{\Delta x}, \quad G_\alpha = A(\varepsilon\widetilde{u})C_\alpha + \tfrac{\Delta t}{\Delta x} D_\alpha(\varDelta, t, x, \widetilde{u}), \tag{2.20}$$

where

$$\sum_{\alpha} C_\alpha = I, \;\; |C_\alpha| \le c, \quad \sum_{\alpha} D_\alpha = 0, \sum_{|\beta|\le s} |\partial^{\beta}_{t,x,w} D_\alpha(\varDelta, t, x, w)| \le C_s(|w|). \tag{2.21}$$

For most difference schemes the only nonzero C_α is the coefficient $C_{\mathbf{0}}$ that multiplies the unshifted term, with the Lax–Friedrichs scheme discussed in Section 3.1 being the main exception.

The second-order and higher-order terms independent of ε may be treated either implicitly or by a combination of implicit and explicit terms. The implicit terms will be written as $-\varDelta t B_{\Delta x} S_{\Delta t} u$, and the explicit terms as

$\Delta t H_{\Delta x} u$. Some specific examples of shift operators $B_{\Delta x}$ and $H_{\Delta x}$ will be presented in Section 3.3. Both B_Δ and H_Δ will be assumed to have constant coefficients, in accordance with the fact that the second-order terms in (1.1) have constant coefficients. Because the implicit terms will appear in the expression for which an estimate is obtained, the symmetric part of the operator $B_{\Delta x}$ needs to be semi-positive definite; more precisely, we assume that

$$\langle u, B_{\Delta x} u\rangle_{\ell^2} \geq 0 \tag{2.22}$$

and

$$|\,\langle v, B_{\Delta x} w\rangle_{\ell^2}\,| \leq c_B \|v\|_{h^\mu} \|w\|_{h^\mu} \tag{2.23}$$

for some nonnegative integer μ. In addition, assume that

$$\|H_{\Delta x} u\|_{\ell^2} \leq c\|u\|_{h^\nu} \tag{2.24}$$

for some nonnegative integer ν, which says that $H_{\Delta x}$ is a difference operator of finite order.

The energy estimates used in this paper require that the first-order terms satisfy an estimate by themselves, and the second-order and higher-order terms are merely required to not increase the energy of solutions. When the second-order terms are multiplied by a "viscosity" parameter then this approach yields estimates that are uniform in that parameter. More specifically, our stability assumptions are

$$|\,\langle v, H_{\Delta x} u\rangle_{\ell^2}\,| \leq \sqrt{\langle v, B_{\Delta x} v\rangle}\sqrt{\langle u, B_{\Delta x} u\rangle} \tag{2.25}$$

on the second-order terms, and

$$\operatorname{Symb}(G_{\Delta x})^* A^{-1} \operatorname{Symb}(G_{\Delta x}) \leq (1 + c(|u|)\Delta t)A \tag{2.26}$$

on the lower-order terms, where we have used the definition (2.2). Due to a technicality arising in the derivation of the sharp Gårding inequality, we will also require that

$$\text{Either } A \text{ is a constant matrix or } \Delta t \geq \delta\Delta x \text{ for some } \delta > 0. \tag{2.27}$$

Assuming that at least one coefficient D_α from (2.20) is nonzero, $G_{\Delta x}$ is an explicit finite-difference approximation to a first-order partial differential operator, and hence the well-known CFL condition says that in order for (2.26) to hold we must assume some restriction

$$\Delta t \leq c\Delta x \tag{2.28}$$

on the size of the time steps. Since the constant δ in the second alternative of (2.27) can be arbitrarily small that condition is compatible with (2.28) no matter how small the constant c there is.

In order to make the $O(\frac{1}{\varepsilon})$ terms in the numerical scheme drop out of the energy estimates for the numerical scheme of the type considered here, as they do from the energy estimates for the PDE (1.1), we will need to treat those terms purely implicitly, *i.e.*, by an approximation of the form $\frac{1}{\varepsilon} Q_{\Delta x} S_{\Delta t} u$. Furthermore, just as large operator in (1.1) is antisymmetric, the operator $Q_{\Delta x} = \sum Q_\alpha S^\alpha_{\Delta x}$ will need to be antisymmetric. More specifically, we will require that

$$Q_{-\alpha} = -Q_\alpha^T, \qquad \|Q_{\Delta x} u\|_{\ell^2} \leq c\|u\|_{h^{\widetilde{\nu}}} \tag{2.29}$$

for some nonnegative integer $\widetilde{\nu}$. The second condition in (2.29) implies that $Q_{\Delta x}$ corresponds to a difference operator of finite order.

It will be convenient to use the notations

$$\mathcal{A}_{\Delta x}(\Delta, \varepsilon\widetilde{u}) := A(\varepsilon\widetilde{u}) + \Delta t B_{\Delta x} \tag{2.30}$$

and

$$\mathcal{G}_{\Delta x}(\varDelta, t, x, \varepsilon\widetilde{u}, \widetilde{u}) := G_{\Delta x}(\varDelta, t, x, \varepsilon\widetilde{u}, \widetilde{u}) + \varDelta t H_{\Delta x}. \tag{2.31}$$

Then, after adding to the elements already considered a term of order zero yields a numerical scheme of the form

$$\mathcal{A}_{\Delta x}(\varDelta, \varepsilon\widetilde{u})S_{\Delta t}u + \tfrac{\Delta t}{\varepsilon}Q_{\Delta x}(\varDelta)S_{\Delta t}u = \mathcal{G}_{\Delta x}(\varDelta, t, x, \varepsilon\widetilde{u}, \widetilde{u})\, u + \varDelta t F(t, x, \widetilde{u}). \tag{2.32}$$

Before stating the theorem we need to define the norm in which the size of the initial data and the solution will be measured. Analogously to the norm $\|\,\|_{h^s_A}$ we define $\|u\|_{h^s_{\mathcal{A}(\varepsilon V)}} = \sqrt{\langle u, u\rangle_{h^s_{\mathcal{A}(\varepsilon V)}}}$, where $\langle u, v\rangle_{h^s_{\mathcal{A}(\varepsilon V)}} = \sum_{|\alpha|\le s}\left\langle D^\alpha_{\Delta x}u, \frac{\mathcal{A}_{\Delta x}(\varDelta,\varepsilon V)+(\mathcal{A}_{\Delta x}(\varDelta,\varepsilon V))^*}{2}D^\alpha_{\Delta x}v\right\rangle_{\ell^2}$. When it is clear what the argument εV of $\mathcal{A}$ is then that argument will sometimes be omitted from the notation for the norm. By (2.19) and (2.22) plus Lemma 4.2 below,

$$\frac{1}{m_0(\varepsilon|V|)}\|u\|^2_{h^s} \le \|u\|^2_{h^s_{\mathcal{A}(\varepsilon V)}} \le M_s(\varepsilon|V|)\|u\|^2_{h^s} + c_B\varDelta t\|u\|^2_{h^{s+2\mu}}. \tag{2.33}$$

The assumption on the initial data will be framed in terms of the $h^s_{\mathcal{A}(0)}$ norm, in order to have a fixed norm to work with while still including the term involving $\varDelta t B_{\Delta x}$. Although the $h^s_{\mathcal{A}(0)}$ norm is stronger than the h^s norm, the factor of $\varDelta t$ multiplying the $h^{s+2\mu}$ norm in (2.33) ensures that when following theorem is applied to approximate solutions of a PDE then only a uniform h^s bound will be required for the initial data to the PDE.

Theorem 2.4. *Consider the finite difference scheme (2.32), where the definitions (2.30), (2.31), (2.17), and (2.18) have been used. Assume that that (2.23) holds for some nonnegative μ, that (2.24) holds for some nonnegative ν, that (2.29) holds for some nonnegative $\widetilde{\nu}$, and that (2.21) holds for some $s \ge \sigma+\max\{2, 2\mu, \nu, \widetilde{\nu}\}$, where $\sigma := \lfloor d/2\rfloor+1$ is the Sobolev embedding exponent, and that (2.19), (2.20), and (2.22) hold. Finally, assume in addition that (2.25), (2.26), (2.27), and (2.28) hold.*

Choose any positive K_0 and any $K > K_0$. Then there is a positive T such that for all initial values $u_{0,\varepsilon,\Delta x}$ satisfying $\sup_{\varepsilon\in(0,1],\Delta x\in(0,1]}\|u_{0,\varepsilon,\Delta x}\|_{h^s_{\mathcal{A}(0)}} \le K_0$ there exists a unique solution of (2.32), (2.16) satisfying $\|u\|_{h^s_{\mathcal{A}(0)}} \le K$ for $0 \le t \le T$. The time of existence T depends only on K_0, K, the bounds in all the assumptions listed above, the smoothness parameter s and the dimension d, and so in particular is independent of $\varepsilon \in (0, 1]$ and of $\varDelta t$ and $\varDelta x$ in $(0, 1]$ satisfying any restrictions placed above on $\varDelta t$.

If in addition $\|Q_{\Delta x}u_{0,\varepsilon,\Delta x}\|_{h^r_{\mathcal{A}(0)}} \le c\varepsilon$ for some $r \le s - \max\{1, 2\mu, \nu, \widetilde{\nu}\}$, then $D_{\Delta t}u$ is uniformly bounded in $\ell^\infty([0, T - \varDelta t] \cap \varDelta t\mathbb{Z}; h^r_{\mathcal{A}(0)})$.

Remark 2.5. Essentially the same proof yields extensions of the results of Theorem 2.4 analogous to the extensions (c) and (e) of Theorem 2.2 in Remark 2.3.

However, when only first-order terms are present then the minimum value of s required in Theorem 2.4 is still $\sigma + 2$ instead of $\sigma + 1$ as in Theorem 2.2. The reason is that Theorem 2.2, like the corresponding result for PDEs, requires that the coefficients belong to C^1, whereas the sharp Gårding inequality used for Theorem 2.4 requires that the coefficients belong to C^2.

3. Specific schemes

In order to make clear the meaning and applicability of Theorem 2.4, some schemes will now be presented and shown to satisfy the conditions of that theorem, although the proof for the Runge–Kutta scheme in Section 3.2 requires two facts from Section 4.

3.1. Lax–Friedrichs and local Lax–Friedrichs schemes

The local Lax–Friedrichs scheme (*e.g.*, [25], Sect. 12.5) for the PDE (1.1), also known as the Rusanov scheme, with both the large terms and the second-order terms treated implicitly, is:

$$\Big[A(\varepsilon v) - \tfrac{\Delta t}{\varepsilon}\Big(\sum_{j=1}^{d} C^j \tfrac{S_{j,\Delta x}-(S_{j,\Delta x})^{-1}}{2\Delta x} + D\Big) - \tfrac{\Delta t}{(\Delta x)^2}\sum_{j,k=1}^{d}(1-(S_{j,\Delta x})^{-1})B^{j,k}(S_{k,\Delta x}-1)\Big]S_{\Delta t}v$$
$$= \Big[A(\varepsilon v) + \tfrac{\Delta t}{2\Delta x}A(\varepsilon v)\sum_{j=1}^{d}\gamma_j(v,\varepsilon)(S_{j,\Delta x}-2+S^{-1}_{j,\Delta x}) + \tfrac{\Delta t}{2\Delta x}\sum_{j=1}^{d}A^j(v)\left(S_{j,\Delta x}-(S_{j,\Delta x})^{-1}\right)\Big]v + \Delta t\,F(v), \tag{3.1}$$

where for notational simplicity we have omitted the dependence of coefficients on the independent variables. There are several possible variants of this scheme. First, although (3.1) is suitable for (1.1) since that PDE is not in conservation form, many treatments of the local Lax–Friedrichs scheme use a conservation formulation, in which the artificial viscosity term

$$\frac{\Delta t}{2\Delta x}A(\varepsilon v)\sum_{j=1}^{d}\gamma_j(v,\varepsilon)\left(S_{j,\Delta x}-2+S^{-1}_{j,\Delta x}\right) \tag{3.2}$$

from (3.1) is replaced by $\frac{\Delta t}{2\Delta x}A(\varepsilon v)\sum_{j=1}^{d}\Big[\gamma_j(S_{\Delta x}\widetilde{v},\varepsilon)(S_{j,\Delta x}-1)-\gamma_j(\widetilde{v},\varepsilon)(1-S^{-1}_{j,\Delta x})\Big]$, where as usual $\widetilde{v}$ denotes a finite set of spatial shifts of v. Since $\widetilde{v}$ differs from v by $O(\Delta x)$, and the coefficients in which it appears are multiplied by $\frac{\Delta t}{\Delta x}$, this variant differs from the version in (3.1) by $O(\Delta t)$, which does not affect the stability of the scheme because condition (2.26) allows arbitrary terms of size $O(\Delta t)$. A more significantly different variant is the original Lax–Friedrichs scheme, in which the artificial viscosity term from (3.1) is replaced by

$$\tfrac{1}{2}A(\varepsilon v)\sum_{j=1}^{d}\gamma_j\left(S_{j,\Delta x}-2+S^{-1}_{j,\Delta x}\right) \tag{3.3}$$

with the now-constant γ_j satisfying

$$\sum_{j=1}^{d}\gamma_j = 1. \tag{3.4}$$

The artificial viscosity of the original Lax–Friedrichs scheme has a simpler form, at the price of being generally much larger than the artificial viscosity of the local version. The dropping of the factor $\frac{\Delta t}{\Delta x}$ from the artificial viscosity term prevents that term from being incorporated into the coefficients $\frac{\Delta t}{\Delta x}D_\alpha$ in (2.20), so the coefficients C_α there are used instead. The following result covers both the local and original Lax–Friedrichs schemes.

Corollary 3.1. *Let $s \geq \sigma+2$ be an integer, where $\sigma := \lfloor d/2\rfloor+1$ is the Sobolev embedding exponent, and assume that A, the A^j and F satisfy assumptions (2.6)–(2.8) and that the C^j, D, and the $B^{j,k}$ satisfy (2.9)–(2.11). Pick any positive K_0 and $K > K_0$, and define $u_{\max} = c_S(d)K$, where $c_S(d)$ is the discrete Sobolev embedding constant from (2.4). Define $\widetilde{A}^j(v,\varepsilon) := (A(\varepsilon v))^{-1/2}A^j(v)(A(\varepsilon v))^{-1/2}$, let $\|M\|$ denote the L^2 operator norm of a matrix M, and assume that*

$$\left\|\widetilde{A}^j(v,\varepsilon)\right\| \leq \gamma_j(v,\varepsilon), \qquad j = 1,\dots,d. \tag{3.5}$$

Assume further that Δt satisfies the stability condition

$$\left[\max_{0\leq\varepsilon\leq1,|v|\leq u_{\max}}\sum_{j=1}^{d}\gamma_j(v,\varepsilon)\right]\frac{\Delta t}{\Delta x} \leq 1. \tag{3.6}$$

If A is not constant then also restrict Δt by the condition $\frac{\Delta t}{\Delta x} \geq \delta$, where δ is an arbitrary positive constant.

Then there exists $T > 0$ independent of $\varepsilon \in (0,1]$, Δx, and Δt satisfying the above restriction(s), such that for all initial data $u_{0,\varepsilon,\Delta x}$ satisfying $\sup_{\varepsilon\in(0,1],\Delta x\in(0,1]} \|u_{0,\varepsilon,\Delta x}\|_{h^s} \le K_0$ the local Lax–Friedrichs scheme (3.1) with initial data $u_{0,\varepsilon,\Delta x}$ has a unique solution in $\ell^\infty([0,T]\cap \Delta t\mathbb{Z}; h^s)$ satisfying $\|u\|_{h^s} \le K$ for $0 \le t \le T$.

Moreover, if in addition $\left\| \left(\sum_{j=1}^d C^j \frac{S_{j,\Delta x}-(S_{j,\Delta x})^{-1}}{2\Delta x} + D \right) u_{0,\varepsilon,\Delta x} \right\|_{h^{s-2}} \le c\varepsilon$ then $\|D_{\Delta t} u\|_{h^{s-2}}$ is also uniformly bounded up to time T.

Furthermore, after replacing conditions (3.5)–(3.6) by (3.4) plus

$$\left[\max_{0\le\varepsilon\le 1, |v|\le u_{\max}} \left\| \widetilde{A}^j(v,\varepsilon) \right\| \right] \frac{\Delta t}{\Delta x} \le \gamma_j, \qquad j = 1,\dots,d, \tag{3.7}$$

the same result holds for the original Lax–Friedrichs scheme in which (3.3) is substituted for the term (3.2) in (3.1).

Proof. Define

$$\begin{aligned}
G_{\Delta x} &:= A(\varepsilon v) + \frac{\Delta t}{2\Delta x} \left[A(\varepsilon v) \sum_{j=1}^d \gamma_j(v,\varepsilon)(S_{j,\Delta x} - 2 + S_{j,\Delta x}^{-1}) + \frac{\Delta t}{2\Delta x} \sum_{j=1}^d A^j(v) \left(S_{j,\Delta x} - (S_{j,\Delta x})^{-1} \right) \right], \\
Q_{\Delta x} &:= \sum C^j \frac{S_{j,\Delta x} - (S_{j,\Delta x})^{-1}}{2\Delta x} + D, \\
B_{\Delta x} &:= -\frac{1}{(\Delta x)^2} \sum (1 - (S_{j,\Delta x})^{-1}) B^{j,k} (S_{k,\Delta x} - 1),
\end{aligned}$$

and $H_{\Delta x} := 0$. Then (3.1) has the form (2.32), where definitions (2.30)–(2.31) have been used. The symbol of $G_{\Delta x}$ is

$$\operatorname{Symb}(G_{\Delta x}) = \left(1 - \frac{\Delta t}{\Delta x} \sum_{j=1}^d \gamma_j(v,\varepsilon)(1 - \cos(\Delta x\,\xi_j)) \right) A(\varepsilon v) + i \frac{\Delta t}{\Delta x} \sum_{j=1}^d \sin(\Delta x\,\xi_j) A^j(v). \tag{3.8}$$

The assumption that A and the A^j are all symmetric implies that the adjoint $\operatorname{Symb}(G_{\Delta x})^*$ is obtained simply by replacing i by $-i$ in (3.8). Substituting these into the left side of (2.26), factoring out factors of $(A(\varepsilon v))^{1/2}$ at the beginning and end of that expression, and using the definition of the $\widetilde{A}^j$ yields

$$\begin{aligned}
&\operatorname{Symb}(G_{\Delta x})^* \left(A^0\right)^{-1} \operatorname{Symb}(G_{\Delta x}) - A \\
&= A^{\frac12} \Bigg\{ -2 \frac{\Delta t}{\Delta x} \sum_{j,k=1}^d \gamma_j(v,\varepsilon)(1 - \cos(\Delta x\,\xi_j)) \\
&\quad + \left(\frac{\Delta t}{\Delta x} \right)^2 \left(\left[\sum_{j,k=1}^d \gamma_j(v,\varepsilon)(1 - \cos(\Delta x\,\xi_j)) \right]^2 + \left[\sum_j \sin(\Delta x\,\xi_j) \widetilde{A}^j \right]^2 \right) \Bigg\} A^{\frac12}.
\end{aligned} \tag{3.9}$$

By assumption (3.5) plus the fact that if a symmetric matrix M satisfies $\|M\| \le cI$ for some positive constant c then $M^2 \le c^2 I$,

$$\left[\sum_j^d \sin(\Delta x\,\xi_j) \widetilde{A}^j \right]^2 \le \left[\sum_j^d \gamma_j(v,\varepsilon) |\sin(\Delta x\,\xi_j)| \right]^2. \tag{3.10}$$

Writing the squared sums involving γ_j from (3.9) and (3.10) as double sums, estimating the products $(1-\cos(\Delta x\,\xi_j))(1-\cos(\Delta x\,\xi_k))$ and $|\sin(\Delta x\,\xi_j)|\,|\sin(\Delta x\,\xi_k)|$ by the elementary bound $ab \le \frac12(a^2+b^2)$, and simplifying yields

$$\begin{aligned}
&\left[\sum_{j=1}^{d} \gamma_j(v,\varepsilon)(1-\cos(\Delta x\,\xi_j))\right]^2 + \left[\sum_{j=1}^{d} \gamma_j(v,\varepsilon)|\sin(\Delta x\,\xi_j)|\right]^2 \\
&= \sum_{j,k=1}^{d} \gamma_j(v,\varepsilon)\gamma_k(v,\varepsilon)\left[(1-\cos(\Delta x\,\xi_j))(1-\cos(\Delta x\,\xi_k)) + |\sin(\Delta x\,\xi_j)|\,|\sin(\Delta x\,\xi_k)|\right] \\
&\le \tfrac{1}{2} \sum_{j,k=1}^{d} \gamma_j(v,\varepsilon)\gamma_k(v,\varepsilon)\left[(1-\cos(\Delta x\,\xi_j))^2 + (1-\cos(\Delta x\,\xi_k))^2 + \sin(\Delta x\,\xi_j)^2 + \sin(\Delta x\,\xi_k)^2\right] \\
&= \left[\sum_{k=1}^{d} \gamma_k(v,\varepsilon)\right]\left[\sum_{j=1}^{d} \gamma_j(v,\varepsilon)\left[(1-\cos(\Delta x\,\xi_j))^2 + \sin(\Delta x\,\xi_j)^2\right]\right] \\
&= 2\left[\sum_{j=1}^{d} \gamma_j(v,\varepsilon)\right]\left[\sum_{j=1}^{d} \gamma_j(v,\varepsilon)(1-\cos(\Delta x\,\xi_j))\right]. \qquad (3.11)
\end{aligned}$$

Substituting (3.10) and then (3.11) into (3.9) and using assumption (3.6) yields

$$\operatorname{Symb}(G_{\Delta x})^* \left(A^0\right)^{-1} \operatorname{Symb}(G_{\Delta x}) - A \le -2\frac{\Delta t}{\Delta x}\sum_{j=1}^{d} \gamma_j(v,\varepsilon)(1-\cos(\Delta x\,\xi_j)) \left(1 - \frac{\Delta t}{\Delta x}\sum_{j=1}^{d} \gamma_j(v,\varepsilon)\right) A \le 0,$$

so the stability criterion (2.26) of Theorem 2.4 holds. Since assumption (2.11) ensures that $\langle u, B_{\Delta x} u\rangle_{\ell^2} \ge 0$, condition (2.25) also holds. The operator $Q_{\Delta x}$ is anti-symmetric, and the remaining hypotheses of Theorem 2.4 may also be checked straightforwardly, so the conclusion of the lemma for the local Lax–Friedrichs scheme (3.1) follows from that theorem.

The original Lax–Friedrichs scheme can be viewed as the special case of the local version in which $\gamma_j(v,\varepsilon)$ is chosen to be the constant $\frac{\Delta x}{\Delta t}\gamma_j$. As noted above, even though $\gamma_j(v,\varepsilon)$ now contains a factor of $\frac{\Delta x}{\Delta t}$ the resulting scheme still has the form defined in (2.32) and (2.31) with $G_{\Delta x}$ given by (2.20). Assumption (3.4) ensures that (3.6) now holds with equality, while assumption (3.7) ensures that (3.5) still holds. Hence the result for the original Lax–Friedrichs scheme follows from the result for the local version. □

3.2. A fourth order classical explicit Runge–Kutta scheme

When the PDE (1.1) contains large terms there is little reason to attempt to obtain a higher-order-in-time scheme, because the error of a scheme of order $p \ge 2$ in time is $(\Delta t)^p O(\|\partial_t^p u\|)$, and even for well-prepared initial data $\partial_t^p u$ will in general be of order $\varepsilon^{-(p-1)}$, which is large when ε is small. Hence in this subsection we consider the PDE (1.1) with ε set equal to one. In addition, we will assume for simplicity that all the $B^{j,k}$, C^j, D, and F vanish.

The centered difference operators $\frac{1}{\Delta x} L_{j,\Delta x}$, where

$$L_{j,\Delta x} := \tfrac{2}{3}\left(S_{j,\Delta x} - (S_{j,\Delta x})^{-1}\right) - \tfrac{1}{12}\left(S_{j,2\Delta x} - (S_{j,2\Delta x})^{-1}\right),$$

are fourth-order approximations to ∂_{x_j} because

$$\operatorname{Symb}(L_{j,\Delta x}) = i\left(\tfrac{4}{3}\sin(\Delta x\,\xi_j) - \tfrac{1}{6}\sin(2\Delta x\,\xi_j)\right) = i\Delta x\,\xi_j + O((\Delta x)^5).$$

Using $\frac{1}{\Delta x} L_{j,\Delta x}$ to approximate the spatial derivatives, the first-order-in-time approximation to the solution after a time-step Δt is $(I + \widetilde{G}_{\Delta x})u$, where

$$\widetilde{G}_{\Delta x}(t,x,v) := \lambda\,(A(v))^{-1} \sum_{j=1}^{d} A^j(t,x,v) L_{j,\Delta x} \qquad (3.12)$$

and as before $\lambda := \frac{\Delta t}{\Delta x}$. The classical fourth-order Runge–Kutta scheme for the spatial discretization can be written in the form (2.32) as $A(v)S_{\Delta t}v = G_{\Delta x}(t,x,v)v$, where $G_{\Delta x} := A(v)\left[1 + \frac{1}{6}\left(\mathcal{K}_1 + 2\mathcal{K}_2 + 2\mathcal{K}_3 + \mathcal{K}_4\right)\right]$ with

$$\mathcal{K}_1 = \widetilde{G}_{\Delta x}(t,x,v),$$
$$\mathcal{K}_2 = \widetilde{G}_{\Delta x}(t + \tfrac{\Delta t}{2}, x, v + \tfrac{1}{2}\mathcal{K}_1 v)(1 + \tfrac{1}{2}\mathcal{K}_1),$$
$$\mathcal{K}_3 = \widetilde{G}_{\Delta x}(t + \tfrac{\Delta t}{2}, x, v + \tfrac{1}{2}\mathcal{K}_2 v)(1 + \tfrac{1}{2}\mathcal{K}_2),$$

and

$$\mathcal{K}_4 = \widetilde{G}_{\Delta x}(t + \Delta t, x, v + \mathcal{K}_3 v)(1 + \mathcal{K}_3).$$

Since the $L_{j,\Delta x}$ are fourth order in Δx, and the Runge–Kutta scheme applied to the discretized problem is fourth order in Δt, the above scheme is a fourth order approximation to the PDE.

Corollary 3.2. *Let $s \geq \sigma + 2$ be an integer, where $\sigma := \lfloor d/2 \rfloor + 1$ is the Sobolev embedding exponent, and assume that A and the A^j satisfy (2.6)–(2.8). Pick any positive K_0 and $K > K_0$, and define $u_{\max} = c_S(d)K$, where $c_S(d)$ is the discrete Sobolev embedding constant from (2.4). Let γ be a number such that $\left|A^j(t,x,u)\right| \leq \gamma A(\varepsilon u)$ for all j, t, x, $\varepsilon \in (0,1]$, and $|u| \leq u_{\max}$, and restrict Δt by the condition*

$$\delta \leq \frac{\Delta t}{\Delta x} \leq \frac{4\sqrt{2}}{3}\frac{1}{\gamma d}, \tag{3.13}$$

where δ is an arbitrary positive constant.

Then there exists $T > 0$ independent of $\varepsilon \in (0,1]$, Δx, and Δt satisfying (3.13) such that for all initial data $u_{0,\varepsilon,\Delta x}$ satisfying $\sup_{\varepsilon \in (0,1], \Delta x \in (0,1]} \|u_{0,\varepsilon,\Delta x}\|_{h^s} \leq K_0$ the scheme (3.1) with initial data $u_{0,\varepsilon,\Delta x}$ has a unique solution in $\ell^\infty([0,T] \cap \Delta t\mathbb{Z}; h^s)$ satisfying $\|u\|_{h^s} \leq K$ for $0 \leq t \leq T$.

Moreover $\|D_{\Delta t}u\|_{h^{s-1}}$ is also uniformly bounded up to time T.

Proof. Since the scheme is explicit, its solution is well defined. Since the other hypotheses of Theorem 2.4 are easily verified, in order to show the boundedness of the solution it suffices to show that the symbol of $G_{\Delta x}$ satisfies (2.26). Because the equation has no large terms the uniform boundedness of the time difference of the solution then follows from the scheme itself.

Since λ is bounded and $L_{j,\Delta x}$ is a bounded combination of shift operators, we obtain by induction and the definition (3.12) of the operator $\widetilde{G}_{\Delta x}$ that the $\mathcal{K}_j$ are also bounded operators on h^r spaces. On the other hand, writing λ as $\frac{\Delta t}{\Delta x}$ and using the fact that $\frac{1}{\Delta x}L_{j,\Delta x}$ is a first-order difference operator, the discrete Sobolev inequality (2.4), the boundedness of the $\mathcal{K}_j$, and the definition of the $\mathcal{K}_j$ yields $\|\mathcal{K}_j v\|_{\ell^\infty} \leq c_1\|\mathcal{K}_j v\|_{h^\sigma} \leq c_2\Delta t\|v\|_{h^{\sigma+1}}$, and hence also that

$$\left\|\left(A(v + c\mathcal{K}_j v)\right)^{-1} A^j(t + c\Delta t, x, v + c\mathcal{K}_j v) - \left(A(v)\right)^{-1} A^j(t,x,v)\right\|_{\ell^\infty} \leq c_3\|\mathcal{K}_j v\|_{\ell^\infty} \leq c_4\Delta t\|v\|_{h^{\sigma+1}}. \tag{3.14}$$

Since arbitrary terms of size $O(\Delta t)$ are permitted in estimate (2.26), let us modify $G_{\Delta x}$ by replacing the arguments $t + c\Delta t$ and $v + c\mathcal{K}_\ell v$ appearing in $\widetilde{G}_{\Delta x}$ in the definition of the $\mathcal{K}_j$ by t and v, respectively. This yields the modified operator

$$G_{\Delta x,\mathrm{mod}}(t,x,v) := A(v)\sum_{k=0}^{4}\frac{1}{k!}\left(\widetilde{G}_{\Delta x}(t,x,v)\right)^k,$$

which has the familiar form of the fourth-order Taylor series approximation to the exponential function that the classical Runge–Kutta method takes when applied to linear equations. In view of (3.14), $G_{\Delta x} = G_{\Delta x,\mathrm{mod}} + O(\Delta t)$. Using this plus the definition of $G_{\Delta x,\mathrm{mod}}$, the estimate

$$\|\mathrm{Symb}(P_{\Delta x}Q_{\Delta x}) - \mathrm{Symb}(P_{\Delta x})\,\mathrm{Symb}(Q_{\Delta x})\|_{\ell^\infty} \leq c\Delta x$$

from Lemma 4.7 below, and the boundedness of λ from below yields $\mathrm{Symb}\,(G_{\Delta x}) = A\sum_{k=0}^{4}\frac{1}{k!}\left(\mathrm{Symb}(\widetilde{G}_{\Delta x})\right)^k + O(\Delta t)$. The symbol of $\widetilde{G}_{\Delta x}$ can be written in terms of $\widetilde{A}^j := A^{-1/2}A^jA^{-1/2}$ and

$$\zeta_j := \tfrac{4}{3}\sin(\Delta x\,\xi_j) - \tfrac{1}{6}\sin(2\Delta x\,\xi_j)$$

as $\mathrm{Symb}(\widetilde{G}_{\Delta x}) = A^{-1/2}\left[i\lambda\sum_{j=1}^{d}\zeta_j\widetilde{A}^j\right]A^{1/2}$. Together with the algebraic identity

$$\left[\sum_{k=0}^{4}\frac{(-iy)^k}{k!}\right]\left[\sum_{k=0}^{4}\frac{(iy)^k}{k!}\right] = 1 - \frac{y^6}{72}\left(1-\tfrac{y^2}{8}\right),$$

these formulas imply that $\mathrm{Symb}(G_{\Delta x})^*A^{-1}\,\mathrm{Symb}(G_{\Delta x})$ equals

$$A - \tfrac{A^{1/2}}{72}\left(\lambda\sum_{j=1}^{d}\zeta_j\widetilde{A}^j\right)^3\left[1-\left(\tfrac{\lambda}{2\sqrt{2}}\sum_{j=1}^{d}\zeta_j\widetilde{A}^j\right)^2\right]\left(\lambda\sum_{j=1}^{d}\zeta_j\widetilde{A}^j\right)^3A^{1/2} + O(\Delta t).$$

Hence (2.26) will hold provided that $\left(\frac{\lambda}{2\sqrt{2}}\sum_{j=1}^{d}\zeta_j\widetilde{A}^j\right)^2 \le I$, and the trig estimate $\left|\frac{4}{3}\sin(x)-\frac{1}{6}\sin(2x)\right| \le \frac{3}{2}$ plus the definition of γ and assumption (3.13) ensure that that condition indeed holds. □

3.3. Examples of approximations of second-order terms

A purely implicit treatment of second-order terms was given in (3.1). That method works even when the second-order terms are purely dispersive, as in the subsystem ([36], Eqs. (2.9)–(2.10)) of the Zakharov equations $F_t = -\Delta G$, $G_t = \Delta F$, where Δ denotes the Laplacian operator and lower-order terms have been omitted for simplicity. For the PDE system, adding F times the first equation to G times the second, integrating over the spatial variables, and noting that the terms arising from the right side form exact derivatives yields $\frac{\mathrm{d}}{\mathrm{d}t}\int(F^2+G^2)\,\mathrm{d}x = 2\int(G\Delta F - F\Delta G)\,\mathrm{d}x = \int\nabla\cdot(G\nabla F - F\nabla G)\,\mathrm{d}x = 0$. For the discretization $S_{\Delta t}F = F - \frac{\Delta t}{(\Delta x)^2}\sum_{j=1}^{d}(S_{j,\Delta x}-2+(S_{j,\Delta x})^{-1})S_{\Delta t}G$, $S_{\Delta t}G = G + \frac{\Delta t}{(\Delta x)^2}\sum_{j=1}^{d}(S_{j,\Delta x}-2+(S_{j,\Delta x})^{-1})S_{\Delta t}F$ of these terms, adding the ℓ^2 inner product of $S_{\Delta t}F$ with the first equation to the ℓ^2 inner product of $S_{\Delta t}G$ with the second equation, and using the fact that $(S_{j,\Delta x})^{-1} = S^*_{j,\Delta x}$ yields

$$\begin{aligned}
&\|S_{\Delta t}F\|^2_{\ell^2} + \|S_{\Delta t}G\|^2_{\ell^2} = \langle S_{\Delta t}F, F\rangle_{\ell^2} + \langle S_{\Delta t}F, F\rangle_{\ell^2}\\
&\quad + \tfrac{\Delta t}{(\Delta x)^2}\sum_{j=1}^{d}\left[\left\langle S_{\Delta t}G, (1-(S_{j,\Delta x})^{-1})(S_{j,\Delta x}-1)S_{\Delta t}F\right\rangle_{\ell^2} - \left\langle S_{\Delta t}F, (1-(S_{j,\Delta x})^{-1})(S_{j,\Delta x}-1)S_{\Delta t}G\right\rangle_{\ell^2}\right]\\
&= \langle S_{\Delta t}F, F\rangle_{\ell^2} + \langle S_{\Delta t}F, F\rangle_{\ell^2}\\
&\quad - \tfrac{\Delta t}{(\Delta x)^2}\sum_{j=1}^{d}\left[\langle(S_{j,\Delta x}-1)S_{\Delta t}G, (S_{j,\Delta x}-1)S_{\Delta t}F\rangle_{\ell^2} - \langle(S_{j,\Delta x}-1)S_{\Delta t}F, (S_{j,\Delta x}-1)S_{\Delta t}G\rangle_{\ell^2}\right]\\
&= \langle S_{\Delta t}F, F\rangle_{\ell^2} + \langle S_{\Delta t}F, F\rangle_{\ell^2} \le \tfrac{1}{2}\left[\|S_{\Delta t}F\|^2_{\ell^2} + \|S_{\Delta t}G\|^2_{\ell^2} + \|F\|^2_{\ell^2} + \|G\|^2_{\ell^2}\right],
\end{aligned}$$

which shows that the second-order terms drop out of the discrete energy estimate, just like they drop out of the continuuous energy estimate for the PDE.

We next consider an example in which second-order terms are treated only semi-implicitly. The scalar PDE $u_t = u_{xx}+u_{xy}+u_{yy}$ may be discretized as $S_{\Delta t}u + \Delta t B_{\Delta x}S_{\Delta t}u = u + \Delta t H_{\Delta x}u$ with $B_{\Delta x}u := \frac{1}{(\Delta x)^2}[(S_{1,\Delta x})^{-1} - 1)(S_{1,\Delta x}-1) + (S_{2,\Delta x})^{-1} - 1)(S_{2,\Delta x}-1)]$ and $H_{\Delta x}u := \frac{1}{(\Delta x)^2}(S_{1,\Delta x}-1)(S_{2,\Delta x}-1)$. Since

$$\begin{aligned}
|\langle v, H_{\Delta x}u\rangle_{\ell^2}| &= \tfrac{1}{(\Delta x)^2}|\langle v, (S_{1,\Delta x}-1)(S_{2,\Delta x}-1)u\rangle_{\ell^2}| = \tfrac{1}{(\Delta x)^2}\left|\left\langle((S_{1,\Delta x})^{-1}-1)v, (S_{2,\Delta x}-1)u\right\rangle_{\ell^2}\right|\\
&\le \tfrac{1}{(\Delta x)^2}\|((S_{1,\Delta x})^{-1}-1)v\|_{\ell^2}\|(S_{2,\Delta x}-1)u\|_{\ell^2} = \tfrac{1}{(\Delta x)^2}\|(S_{1,\Delta x}-1)v\|_{\ell^2}\|(S_{2,\Delta x}-1)u\|_{\ell^2}
\end{aligned}$$

and

$$\begin{aligned}\langle u, B_{\Delta x}u\rangle_{\ell^2} &= \tfrac{1}{(\Delta x)^2}\left\langle u, \left[((S_{1,\Delta x})^{-1}-1)(S_{1,\Delta x}-1) + ((S_{2,\Delta x})^{-1}-1)(S_{2,\Delta x}-1)\right]u\right\rangle_{\ell^2}\\ &= \tfrac{1}{(\Delta x)^2}\left[\langle (S_{1,\Delta x}-1)u, (S_{1,\Delta x}-1)u\rangle_{\ell^2} + \langle (S_{2,\Delta x}-1)u, (S_{2,\Delta x}-1)u\rangle_{\ell^2}\right]\\ &= \tfrac{1}{(\Delta x)^2}\left[\|(S_{1,\Delta x}-1)u\|_{\ell^2}^2 + \|(S_{2,\Delta x}-1)u\|_{\ell^2}^2\right]\end{aligned}$$

the stability condition (2.25) is satisfied.

4. Discrete versions of continuous estimates

In this section we formulate and prove discrete versions of various well-known continuous estimates that are important in the analysis of linear and nonlinear PDEs. The discrete versions will play an analogous role in our theory of difference schemes.

4.1. Discrete Fourier transform

In order to prove the discrete Sobolev embedding estimate (2.4) and other results that will be needed for the proofs of Theorems 2.2 and 2.4 we will need certain results about the discrete Fourier transform. Recall that the physical-space variable, denoted x, lies in the set $X_{\Delta x} = [-L_{\Delta x}, L_{\Delta x})^d \cap \Delta x\mathbb{Z}^d$, which contains $N := 2\frac{L_{\Delta x}}{\Delta x}$ points in each direction. The domain $\Pi_{\Delta x} := \left[-\frac{\pi}{\Delta x}, \frac{\pi}{\Delta x}\right)^d \cap \frac{\pi}{L_{\Delta x}}\mathbb{Z}^d$ of the corresponding Fourier variable ξ also contains N points in each direction. The discrete Fourier and inverse Fourier transforms may be written in the form

$$\widehat{f}(\xi) := \sum_{x\in X_{\Delta x}} f(x)\mathrm{e}^{-i\xi\cdot x}(\Delta x)^d, \tag{4.1}$$

$$f(x) = \frac{1}{(2\pi)^d}\sum_{\xi\in \Pi_{\Delta x}} \widehat{f}(\xi)\mathrm{e}^{i\xi\cdot x}(\Delta\xi)^d, \tag{4.2}$$

where $\Delta\xi := \frac{\pi}{L_{\Delta x}}$. These formulas can be derived from the standard discrete Fourier and inverse Fourier transform formulas (*e.g.* [12], pp. 250–252) $\widehat{a}_m := \sum_{n=0}^{N-1} a_n\mathrm{e}^{-2\pi imn/N}$ and $a_n = \frac{1}{N}\sum_{m=0}^{N-1}\widehat{a}_m\mathrm{e}^{2\pi imn/N}$ by making the changes of independent and dependent variables

$$x = n\Delta x - L_{\Delta x}, \quad \xi = \tfrac{\pi m}{L_{\Delta x}}, \quad f(x) = a_n, \quad \widehat{f}(\xi) = \mathrm{e}^{\pi i m}\widehat{a}_m\Delta x \tag{4.3}$$

and generalizing to multiple dimensions by applying one-dimensional transforms in each variable. The discrete Fourier and inverse Fourier transforms satisfy the Plancherel identity

$$\|f\|_{\ell^2(X_{\Delta x})}^2 = \tfrac{1}{(2\pi)^d}\|\widehat{f}\|_{\ell^2(\Pi_{\Delta x})}^2, \tag{4.4}$$

where the ℓ^2 norm $\|g\|_{\ell^2(\Pi_{\Delta x})} := \sqrt{\langle g, g\rangle_{\ell^2(\Pi_{\Delta x})}}$ on $\Pi_{\Delta x}$ is defined in terms of the inner product $\langle g, h\rangle_{\ell^2(\Pi_{\Delta x})} := \sum_{\xi\in\Pi_{\Delta x}} g(\xi)\overline{h(\xi)}(\Delta\xi)^d$. The corresponding Parseval identity is

$$\langle f, g\rangle_{\ell^2(X_{\Delta x})} = \tfrac{1}{(2\pi)^d}\left\langle \widehat{f}, \widehat{g}\right\rangle_{\ell^2(\Pi_{\Delta x})}. \tag{4.5}$$

Formula (4.4) can be derived from the Plancherel identity $\sum_{n=0}^{N-1}|a_n|^2 = \frac{1}{N}\sum_{m=0}^{N-1}|\widehat{a}_m|^2$ ([12], Lem. 7.1, p. 251) for the standard form of the discrete Fourier transform via the changes of variables (4.3).

Formula (4.1) and the definition (2.2) of the symbol of a difference operator imply that

$$\widehat{P_{\Delta x}f}(\xi) = [\mathrm{Symb}(P_{\Delta x})](\xi)\widehat{f}(\xi) \tag{4.6}$$

for any constant-coefficient shift operator $P_{\Delta x}$. The discrete ℓ^p and $w^{k,p}$ norms are defined by

$$\|u\|_{\ell^p} := \left[\sum_{x\in X_{\Delta x}} |u(x)|^p (\Delta x)^d\right]^{1/p}, \quad \|u\|_{w^{k,p}} := \sum_{|\alpha|\le k} \|D^\alpha_{\Delta x} u\|_{\ell^p}.$$

The following lemma is the discrete version of the Sobolev embedding lemma, which says that the H^s norm dominates the C^0 norm when $s > \frac{d}{2}$, and, more generally, dominates the C^k norm when $s > k + \frac{d}{2}$.

Lemma 4.1 (Discrete sobolev embedding estimate)**.** *For any nonnegative integer k there is a constant $c_S(d,k)$ depending only on k and the spatial dimension d such that*

$$\|u\|_{w^{k,\infty}} \le c_S(d,k)\|u\|_{h^{k+\sigma}}, \tag{4.7}$$

where $\sigma := \lfloor d/2 \rfloor + 1$ is the Sobolev embedding exponent.

Proof. The standard proof (*e.g.* [11], pp. 243–244) for the differential case translates directly to the difference case once we notice that the discrete Fourier transform of a difference is $\widehat{D^\alpha_{\Delta x} f}(\xi) = \prod_{j=1}^d \left(\frac{e^{i\xi_j \Delta x}-1}{\Delta x}\right)^{\alpha_j} \widehat{f}(\xi)$ by (4.6), and

$$\left|\frac{e^{i\xi_j \Delta x}-1}{\Delta x}\right| \ge c|\xi_j| \tag{4.8}$$

for $|\xi| \le \frac{\pi}{\Delta x}$, because the fact that $\sigma > \frac{d}{2}$ ensures that the Fourier-space sum $\sum_{\xi\in\Pi_{\Delta x}} (1+|\xi|^2)^{-\sigma} (\Delta\xi)^d$ is bounded by a constant independent of Δx just like it ensures that the integral $\int_{\xi\in\mathbb{R}^d} (1+|\xi|^2)^{-\sigma}\, d\xi$ is finite. □

The following elementary result says that, just as for differential operators, the order of a difference operator is bounded by twice the order of the bilinear form that it induces.

Lemma 4.2. *If a constant coefficient difference operator $B_{\Delta x}$ satisfies (2.23) for some nonnegative integer μ then $\|B_{\Delta x} u\|_{\ell^2} \le \widetilde{c}\|u\|_{h^{2\mu}}$.*

Proof. Define an operator Λ by $\widehat{\Lambda f}(\xi) = (1+|\xi|)\hat{f}(\xi)$. Then by the definition of the h^μ norm, (4.4), (4.6), and (4.8) plus the trivial reverse inequality $\left|\frac{e^{i\xi_j\Delta x}-1}{\Delta x}\right| \le c_2|\xi_j|$,

$$c_-(\mu,s)\|f\|_{h^{\mu+s}} \le \|\Lambda^\mu f\|_{h^s} \le c_+(\mu,s)\|f\|_{h^{\mu+s}}. \tag{4.9}$$

By the fact that $B_{\Delta x}$ has constant coefficients, (2.23), and (4.9),

$$|\langle v, B_{\Delta x} u\rangle_{\ell^2}| = |\langle \Lambda^{-\mu} v, B_{\Delta x}\Lambda^\mu u\rangle| \le c\,\|\Lambda^{-\mu} v\|_{h^\mu}\|\Lambda^\mu u\|_{h^\mu} \le \widetilde{c}\,\|v\|_{\ell^2}\|u\|_{h^{2\mu}}. \tag{4.10}$$

Setting $v := B_{\Delta x} u$ in (4.10) yields the claimed result. □

It will sometimes be useful to extend functions defined on the discrete lattice $X_{\Delta x}$ so as to be defined on a continuous domain $X_{L_{\Delta x}} := [-L_{\Delta x}, L_{\Delta x})^d$. Such an interpolation can be obtained by taking the inverse discrete Fourier transform of the discrete Fourier transform of a function, but letting the new spatial variable vary over $X_{L_{\Delta x}}$ rather than just $X_{\Delta x}$. The fact that the inverse and direct Fourier transforms are inverses ensures that the restriction of the interpolation back to the lattice yields the original function. By formulas (4.1) and (4.2), this interpolation operator Int_x is defined for $x \in X_{L_{\Delta x}}$ by

$$[\mathrm{Int}_x f](x) := \frac{1}{(2\pi)^d} \sum_{\xi\in\Pi_{\Delta x}} \left[\sum_{y\in X_{\Delta x}} f(y) e^{-i\xi\cdot y} (\Delta x)^d\right] e^{i\xi\cdot x} (\Delta\xi)^d.$$

Lemma 4.3. *The interpolation operator* Int_x *is a bounded map from* $h^s(X_{\Delta x})$ *to* $H^s(X_{L_{\Delta x}})$ *for every integer* $s \geq 0$*, uniformly in* Δx*.*

Proof. Using Plancherel's identities for $X_{L_{\Delta x}}$ and $X_{\Delta x}$, the fact that $0 < c_1 \leq \frac{\sum_{|\alpha|\leq s}\prod_j |\xi_j|^{2\alpha_j}}{(1+|\xi|^2)^s} \leq c_2 < \infty$, and (4.8), we obtain that

$$\|\,\mathrm{Int}_x f\|^2_{H^s(X_{L_{\Delta x}})} \leq c_1 \sum_{\xi\in\Pi_{\Delta x}} (1+|\xi|^2)^s |\widehat{f}(\xi)|^2 \left(\tfrac{\pi}{L_{\Delta x}}\right)^d$$
$$\leq c_2 \sum_{\xi\in\Pi_{\Delta x}} \left(1+\sum_j \left|\tfrac{\mathrm{e}^{i\xi_j\Delta x}-1}{\Delta x}\right|^2\right)^s |\widehat{f}(\xi)|^2 \left(\tfrac{\pi}{L_{\Delta x}}\right)^d \leq c_3\|f\|^2_{h^s(X_{\Delta x})}. \qquad \square$$

4.2. Discrete calculus inequalities

Our development of the theory of nonlinear difference equations will require not only the discrete Sobolev embedding estimate but also a discrete version of certain Gagliardo–Nirenberg inequalities, which interpolate between various discrete Sobolev norms, and of certain Moser inequalities, which estimate Sobolev norms of smooth functions of $u(x)$ in terms of the corresponding norms of $u(x)$.

The discrete version of the chain rule is the identity

$$D_{j,\Delta x}F(t,x,u(t,x)) = \frac{1}{\Delta x}\int_0^1 \frac{\mathrm{d}}{\mathrm{d}s}F(t,x+s\Delta x e_j, sS_{j,\Delta x}u+(1-s)u)\,\mathrm{d}s$$
$$= \int_0^1 F_{x_j}(t,x+s\Delta x e_j, sS_{j,\Delta x}u+(1-s)u)\,\mathrm{d}s + \int_0^1 F_u(t,x+s\Delta x e_j, sS_{j,\Delta x}u+(1-s)u)\,\mathrm{d}s\cdot D_{j,\Delta x}u,$$

which can be written in terms of $F_{x_j,\Delta x}(t,x,u,S_{j,\Delta x}u) := \int_0^1 F_{x_j}(t,x+s\Delta x e_j, sS_{j,\Delta x}u+(1-s)u)\,\mathrm{d}s$ and $F_{u,j,\Delta x}(t,x,u,S_{j,\Delta x}u) := \int_0^1 F_u(t,x+s\Delta x e_j, sS_{j,\Delta x}u+(1-s)u)\,\mathrm{d}s$ as

$$D_{j,\Delta x}F(t,x,u(t,x)) = F_{x_j,\Delta x}(t,x,u,S_{j,\Delta x}u) + F_{u,j,\Delta x}(t,x,u,S_{j,\Delta x}u)\cdot D_{j,\Delta x}u. \tag{4.11}$$

A similar identity holds for the time difference. Note that the discrete effective derivatives $F_{x_j,\Delta x}$ and $F_{u,j,\Delta x}$ depend on both u and its shift $S_{j,\Delta x}u$. In particular, for the function $F(v,w)=vw$ calculating the last integral in (4.11) explicitly yields $D_{j,\Delta x}(vw) = (\mathrm{Av}_{j,\Delta x}\,v)\,(D_{j,\Delta x}w) + (D_{j,\Delta x}v)\,(\mathrm{Av}_{j,\Delta x}\,w)$, where $Av_{j,\Delta x} := \frac{(S_{j,\Delta x})+1}{2}$. We will also sometimes use the unsymmetric versions

$$D_\Delta(vw) = vD_\Delta w + (D_\Delta v)\,S_\Delta w \tag{4.12}$$

where Δ denotes either Δt or Δx, and

$$D_{j,m\Delta x,c}(vw) = vD_{j,m\Delta x,c}w + (D_{j,m\Delta x}v)\,S_{j,m\Delta x}w + \left(D_{j,m\Delta x}(S_{j,m\Delta x})^{-1}v\right)(S_{j,m\Delta x})^{-1}w. \tag{4.13}$$

Identities (4.12) and (4.13) can be verified by substituting in the definitions of the shift and difference operators and simplifying. By induction, (4.11) and its time-difference variant imply analogous but more complicated chain-rule formulas for higher-order differences $(D_{\Delta t})^m D^\alpha_{\Delta x}F(u)$.

The Gagliardo–Nirenberg inequalities (*e.g.*, [13], Thm. 9.3) are estimates for the L^p norms of derivatives of order k in terms of L^q norms of derivatives of some higher order m and L^r norms of the original function. Specifically, $\sum_{|\alpha|=k}\|D^\alpha u\|_{L^p} \leq c\left[\sum_{|\beta|=m}\|D^\beta u\|^\alpha_{L^q}\right]\|u\|^{1-\alpha}_{L^r}$ when $1\leq q\leq\infty$, $1\leq r\leq\infty$, $\frac{k}{m}\leq\alpha<1$ and $\frac{1}{p} = \frac{k}{d}+\alpha\left(\frac{1}{q}-\frac{m}{d}\right)+(1-\alpha)\frac{1}{r}$. We prove the discrete version of these inequalities in the special case when $\alpha=\frac{k}{m}$, with q and r restricted to be at least two.

Lemma 4.4. *For any $2 \le q, r \le \infty$ with at least one of q or r finite, and integers $1 \le k < m$, the p defined by*

$$\frac{1}{p} = \frac{k}{m}\frac{1}{q} + \left(1 - \frac{k}{m}\right)\frac{1}{r} \tag{4.14}$$

also satisfies $2 \le p < \infty$, and there is a constant c depending only on q, r, k, m and the spatial dimension such that

$$\sum_{|\alpha|=k} \|D^\alpha_{\Delta x} u\|_{\ell^p} \le c \left[\sum_{|\beta|=m} \|D^\beta_{\Delta x} u\|_{\ell^q}^{k/m} \right] \|u\|_{\ell^r}^{1-k/m}. \tag{4.15}$$

Proof. The proof for the continuous case outlined in ([13], pp. 24–27) relies mostly on induction and Hölder's inequality, and so only requires proving directly the case when $k = 1$, $m = 2$, both q and r are finite, and the spatial dimension equals one. To prove that case note that identity (4.12) with $v = |D_{\Delta x} u|^{p-2} D_{\Delta x} u$ and $w = u$ can be written as

$$|D_{\Delta x} u|^p = D_{\Delta x}\left(u |D_{\Delta x} u|^{p-2} D_{\Delta x} u\right) - D_{\Delta x}\left(D_{\Delta x} u|^{p-2} D_{\Delta x} u\right) S_{\Delta x} u.$$

Summing over the spatial grid, noting that the sum of the exact difference vanishes by periodicity, and using the discrete chain rule (4.11), the fact that shifts preserve ℓ^p norms, and Hölder's inequality with the three factors $\frac{p}{p-2}$, q and r, we obtain

$$\begin{aligned}
\sum_{X_{\Delta x}} |D_{\Delta x} u|^p \Delta x &= -\sum_{X_{\Delta x}} D_{\Delta x}\left(|D_{\Delta x} u|^{p-2} D_{\Delta x} u\right) S_{\Delta x} u \Delta x \\
&= -(p-2) \sum_{X_{\Delta x}} \left[\int_0^1 |s S_{\Delta x} D_{\Delta x} u + (1-s) D_{\Delta x} u|^{p-2} \, \mathrm{d}s \right] (D_{\Delta x})^2 u \, S_{\Delta x} u \Delta x \\
&\le c(p) \sum_{X_{\Delta x}} \left(|S_{\Delta x} D_{\Delta x} u|^{p-2} + |D_{\Delta x} u|^{p-2} \right) |(D_{\Delta x})^2 u| \, |S_{\Delta x} u| \Delta x \\
&\le 2c(p) \left(\sum_{X_{\Delta x}} |D_{\Delta x} u|^p \Delta x \right)^{\frac{p-2}{p}} \left(\sum_{X_{\Delta x}} |(D_{\Delta x})^2 u|^q \Delta x \right)^{1/q} \left(\sum_{X_{\Delta x}} |u|^r \Delta x \right)^{1/r}
\end{aligned} \tag{4.16}$$

provided that $\frac{p-2}{p} + \frac{1}{q} + \frac{1}{r} = 1$, which is equivalent to (4.14) for the case $k = 1$, $m = 2$. Solving the inequality (4.16) for $\sum_{X_{\Delta x}} |D_{\Delta x} u|^p \Delta x$ yields the above-mentioned case of (4.15). Since the constant c in (4.16) depends only on p, which remains finite as either q or r but not both tends to infinity, the estimate is also valid under those circumstances. As indicated above, the rest of the proof follows the proof for the differential case. $\square$

Moser estimates bound norms of derivatives of compositions of functions and of commutators of multiplication and derivative operators in terms of the norms of their component parts. Using Lemmas 4.1 and 4.4, discrete versions of the Moser estimates can be obtained. Although we will only use here the following weak versions of those estimates, the discrete versions of the standard strong versions (*e.g.* [26], Prop. 2.1, p. 43 are also valid, and their proofs are identical to those of the standard differential versions, which use only the above Gagliardo–Nirenberg estimates, the Sobolev embedding estimate, Hölder's inequality, and standard interpolation estimates between L^p norms that are also valid for the discrete ℓ^p norms.

Lemma 4.5. *Let s be an integer, and define $\widetilde{s} := \max\{s, \sigma\}$, where $\sigma := \lfloor d/2 \rfloor + 1$ is the Sobolev embedding exponent. Assume that $\mathcal{F}$ is a $C^{\widetilde{s}}$ function satisfying $\mathcal{F}(t, x, 0) \equiv 0$, A is a matrix-valued $C^{\widetilde{s}}$ function satisfying $A(t, x, 0) \equiv 0$, and B is a matrix-valued $C^{\widetilde{s}}$ function of the independent variables t and x. Then there is a* *function c_F and constant c, depending on F, A, B, s and the spatial dimension such that*

$$\|\mathcal{F}(t,x,U)\|_{h^s} \le c_{\mathcal{F}}\left(\|U\|_{h^{\tilde{s}}}\right)\|U\|_{h^s}, \tag{4.17}$$

$$\sum_{|\alpha|\le s}\|[D^\alpha_{\Delta x}, A(t,x,U)]u\|_{\ell^2} \le c\|A(t,x,U)\|_{h^{\tilde{s}}}\|u\|_{h^{s-1}}, \tag{4.18}$$

$$\sum_{|\alpha|\le s}\|[D^\alpha_{\Delta x}, B(t,x)]u\|_{\ell^2} \le c\|u\|_{h^{s-1}}, \tag{4.19}$$

where $[D^\alpha_{\Delta x}, A]$ denotes the commutator of $D^\alpha_{\Delta x}$ with the operator of multiplication by the matrix A.

The differential Gronwall lemma says that if a function u satisfies $u' \le c(t)u + k(t)$ then $u(t) \le \mathrm{e}^{\int_0^t c(r)\,dr}u(0) + \int_0^t \mathrm{e}^{\int_s^t c(r)\,dr}k(s)\,\mathrm{d}s$. We will make use of the following discrete constant-coefficient version:

Lemma 4.6. *Suppose that a nonnegative function w defined on $[0,T]\cap\Delta t\mathbb{Z}$ satisfies $w(t+\Delta t) \le (1+c\Delta t)w(t) + k\Delta t$. Then $w(t) \le \mathrm{e}^{ct}w(0) + \frac{\mathrm{e}^{ct}-1}{c}k$ for $t\in[0,T]\cap\Delta t\mathbb{Z}$.*

Proof. By induction, the formula for the sum of a geometric series, and the estimate $(1+x)^n \le \mathrm{e}^{nx}$, for $t = n\Delta t$, $w(t) = w(n\Delta t) \le (1+c\Delta t)^n w(0) + k\Delta t\sum_{j=0}^{n-1}(1+c\Delta t)^j = (1+c\Delta t)^n w(0) + k\frac{(1+c\Delta t)^n-1}{c} \le \mathrm{e}^{cn\Delta t}w(0) + k\frac{\mathrm{e}^{cn\Delta t}-1}{c} = \mathrm{e}^{ct}w(0) + \frac{\mathrm{e}^{ct}-1}{c}k$. □

4.3. Discrete sharp Gårding inequality

The sharp Gårding inequality states if the symbol of an operator is self-adjoint and nonnegative then the symmetric part of the operator is almost nonnegative, in some sense. Both the original and sharp forms of Gårding's inequality play an important role in the theory of differential and pseudodifferential operators ([43], Sect. 7.6, [42], Chap. 7). Although a version has been proven for shift operators ([24], Thm. 1.1) and applied to difference schemes, that version estimates the action of shift operators on the space L^2 of functions of continuous variables. Specifically, if the coefficients $P_\alpha(x)$ of a shift operator $P_{\Delta x}$ are defined for all $x\in X_L$ and satisfy $\sum_{\alpha\in\mathbb{Z}^d}\left[\sum_{0\le\beta\le2}\|D^\beta P_\alpha\|_{C^0} + (1+|\alpha|^2)\|P_\alpha\|_{C^0}\right] < \infty$, and the symbol $\mathrm{Symb}(P_{\Delta x}) = \sum_{\alpha\in\mathbb{Z}^d}P_\alpha(x)\mathrm{e}^{i\alpha\cdot\xi}$ of the shift operator is self-adjoint and nonnegative, then $P_{\Delta x} + P^*_{\Delta x} \ge -c\Delta x$ when considered as an operator acting on L^2. Here we formulate and prove a fully discrete version, in which the shift operator acts on the space ℓ^2 and its coefficients need only be defined at grid points.

By the definition of symbols, the operator corresponding to a symbol is $\mathrm{Op}\left(\sum_\alpha P_\alpha(x)\mathrm{e}^{i\Delta x\alpha\cdot\xi}\right) = \sum_\alpha P_\alpha(x)S^\alpha_{\Delta x}$. Then $\mathrm{Op}\circ\mathrm{Symb}$ is the identity operator on shift operators, and $\mathrm{Symb}\circ\mathrm{Op}$ is the identity operator on symbols. By direct calculation we obtain the following lemma, in which $\|P\|_{\ell^2\to\ell^2}$ denotes the operator norm $\sup_{\|v\|_{\ell^2}\le1}\|Pv\|_{\ell^2}$, since moving a shift past a coefficient shifts that coefficient by an amount $O(\Delta x)$, but only if the coefficient depends on x.

Lemma 4.7. *Suppose that the coefficients $P_\alpha(x)$ and $Q_\alpha(x)$ of bounded shift operators $P_{\Delta x}$ and $Q_{\Delta x}$ belong to $w^{1,\infty}$. Then*

$$\begin{aligned}\left\|P^*_{\Delta x} - \mathrm{Op}\left((\mathrm{Symb}\,P_{\Delta x})^*\right)\right\|_{\ell^2\to\ell^2} &\le c\Delta x,\\ \|P_{\Delta x}Q_{\Delta x} - \mathrm{Op}\left(\mathrm{Symb}\,P_{\Delta x}\,\mathrm{Symb}\,Q_{\Delta x}\right)\|_{\ell^2\to\ell^2} &\le c\Delta x,\\ \|\mathrm{Symb}(P_{\Delta x}Q_{\Delta x}) - \mathrm{Symb}(P_{\Delta x})\,\mathrm{Symb}(Q_{\Delta x})\|_{\ell^\infty} &\le c\Delta x.\end{aligned}$$

Moreover, if $P_\alpha(x) = P^{(0)}_\alpha + \delta P^{(1)}_\alpha(x)$ then $\left\|P^*_{\Delta x} - \mathrm{Op}\left((\mathrm{Symb}\,P_{\Delta x})^*\right)\right\|_{\ell^2\to\ell^2} \le c\delta\Delta x$, and if in addition $Q_\alpha(x) = Q^{(0)}_\alpha + \delta Q^{(1)}_\alpha(x)$ then

$$\|P_{\Delta x}Q_{\Delta x} - \mathrm{Op}\left(\mathrm{Symb}\,P_{\Delta x}\,\mathrm{Symb}\,Q_{\Delta x}\right)\|_{\ell^2\to\ell^2} \le c\delta\Delta x. \tag{4.20}$$

We now prove the discrete sharp Gårding inequality:

Theorem 4.8. *Let $P_{\Delta x} := \sum_{\alpha\in\mathbb{Z}} P_\alpha(x) S^\alpha_{\Delta x}$ be a shift operator, possibly involving infinitely many shifts, and assume that*

$$\sum_{\alpha\in\mathbb{Z}^d}\left[\sum_{0\le|\beta|\le 2}\|D^\beta_{\Delta x}P_\alpha\|_{h^\sigma}+(1+|\alpha|^2)\|P_\alpha\|_{h^\sigma}\right]\le c_1. \tag{4.21}$$

Suppose in addition that the symbol of $P_{\Delta x}$ satisfies

$$\mathrm{Symb}(P_{\Delta x})^* = \mathrm{Symb}(P_{\Delta x}), \qquad \mathrm{Symb}(P_{\Delta x}) \ge 0. \tag{4.22}$$

Then the operator $P_{\Delta x}$ is almost positive in the sense that

$$P_{\Delta x} + P^*_{\Delta x} \ge -c\Delta x, \tag{4.23}$$

where c depends only on c_1 in (4.21). Moreover, if

$$P_\alpha(x) = P^{(0)}_\alpha + \delta P^{(1)}_\alpha(x) \tag{4.24}$$

and (4.21) holds with P_α replaced by each $P^{(j)}_\alpha$ then

$$P_{\Delta x} + P^*_{\Delta x} \ge -c_2\delta\Delta x. \tag{4.25}$$

Proof. Begin by extending the definition of the coefficients $P_\alpha(x)$ from the discrete lattice $X_{\Delta x}$ to the torus $X_{L_{\Delta x}}$, by replacing $P_\alpha(x)$ by its interpolation $[\mathrm{Int}_x P_\alpha](x)$, hereafter abbreviated to $\mathrm{Int}\, P_\alpha$. Although interpolation maintains self-adjointness of the symbol, it does not necessarily maintain positivity, as the scalar example $f(x) = 1-(\Delta x)^2 - \cos(\pi(x-\frac{\Delta x}{2}))$ with $\Delta x = 1/m$ for some integer $m > 1$ illustrates: its restriction to the lattice $\Delta x\mathbb{Z}$ is positive but it is negative at $x = \frac{\Delta x}{2}$. However, as that example also illustrates, the most negative value of $\mathrm{Symb}(\mathrm{Int}\, P_{\Delta x})$ is of the order of

$$\sum_\alpha \|P_\alpha\|_{w^{2,\infty}}(\Delta x)^2, \tag{4.26}$$

as a Taylor series expansion around the most negative point y shows in view of the fact that the value at the closest grid point must be nonnegative. By (4.21) plus the Sobolev embedding estimate (Lem. 4.1), $\sum_\alpha \|P_\alpha\|_{w^{2,\infty}}$ is uniformly bounded. Hence nonnegativity of the interpolated symbol can be restored by adding a constant times $(\Delta x)^2 I$ to $P_{\Delta x}$, which does not affect the conclusion to be proven.

The Sobolev embedding estimate also ensures that assumption (4.21) implies the assumption

$$\sum_\alpha \sup_{x\in X_{L_{\Delta x}}}\left[\sum_{|\beta|\le 2}|\partial^\beta_x \,\mathrm{Int}\, P_\alpha| + (1+|\alpha|^2)|\,\mathrm{Int}\, P_\alpha|\right] \le c < \infty$$

of the ordinary continuous version of the sharp Gårding inequality ([45], Thm. 1.1). Apply that result, namely $\langle v, (\mathrm{Int}\, P_{\Delta x} + \mathrm{Int}\, P^*_{\Delta x})v\rangle_{L^2} \ge -c\Delta x\|v\|^2_{L^2}$, to functions v that are constant in each centered grid cell $C_x := \{y \mid \max_k |x_k - y_k| \le \frac{\Delta x}{2}\}$ when x is a lattice point, and use the fact that for such v

$$\begin{aligned}&|\langle v, (\mathrm{Int}\, P_{\Delta x} + \mathrm{Int}\, P^*_{\Delta x})v\rangle_{L^2} - \langle v, (P_{\Delta x} + P^*_{\Delta x})v\rangle_{\ell^2}|\\ &\quad = \left|\left\langle v(x), \int_{C_x}[(\mathrm{Int}\, P_{\Delta x}(y) + \mathrm{Int}\, P_{\Delta x}(y)^*) - (\mathrm{Int}\, P_{\Delta x}(x) + \mathrm{Int}\, P_{\Delta x}(x)^*)]v(x)\right\rangle_{\ell^2}\right| \le c\Delta x\|v\|^2_{\ell^2}.\end{aligned} \tag{4.27}$$

Since the error of converting from continuous to discrete norms is of the same order as the claimed deviation from positivity, the continuous result implies the discrete one.

Finally, in order to obtain the improved estimate (4.25) it suffices to show it in the continuous case and to show that the estimates (4.26) and (4.27) also improve in the same way. To do that, note that in the proof of Gårding's inequality for the continuous case in [45] the constant c in (4.23) depends only on the estimates in Lemmas 2 and 3 there. The estimates in the proof of Lemma 2 in ([45], p. 156) only involve derivatives of the coefficients P_α, and so contain a factor of δ when (4.24) holds, while the proof of Lemma 3 in ([45], p. 156) is based on estimating the expression, in our notation,

$$\sum_\alpha \left[\int P_\alpha(x-(\Delta x)^{1/2}y)\left\{q(y)^2 - q(y+(\Delta x)^{1/2}\alpha)^2\right\}\,\mathrm{d}y\right]\mathrm{e}^{i\alpha\cdot z}. \tag{4.28}$$

When (4.24) holds then the part of (4.28) involving the constant coefficients $P_\alpha^{(0)}$ cancels since

$$\int \left\{q(y)^2 - q(y+(\Delta x)^{1/2}\alpha)^2\right\}\,\mathrm{d}y = 0,$$

again leaving only terms containing a factor of δ. Hence the extension holds for the continuous sharp Gårding inequality, and similar arguments show that the same improvement is obtained in (4.26)–(4.27). $\square$

Corollary 4.9.

1. *Let $A(x)$ be a positive self-adjoint matrix satisfying$A(x) \ge c_0 I$, such that $\|A(x)\|_{h^{\sigma+2}} \le c_1 < \infty$. Suppose that a shift operator $P_{\Delta x}(x)$ containing finitely-many shifts satisfies*

$$\sum_\alpha \|P_\alpha(x)\|_{h^{\sigma+2}} \le c_2 < \infty$$

and

$$\operatorname{Symb}(P_{\Delta x})^* A^{-1}\operatorname{Symb}(P_{\Delta x}) \le (1+\eta)A. \tag{4.29}$$

Then

$$\left\langle P_{\Delta x}v, A^{-1}P_{\Delta x}v\right\rangle_{\ell^2} \le (1+\eta+c\Delta x)\left\langle v, Av\right\rangle_{\ell^2}, \tag{4.30}$$

where c depends only on the c_j.

2. *Let A be a constant-coefficient positive self-adjoint matrix satisfying $A \ge c_0 I$. Suppose that a shift operator $P_{\Delta x}(x) = P_{\Delta x}^{(0)} + \delta P_{\Delta x}^{(1)}(x)$ containing finitely-many shifts satisfies (4.29) and*

$$\sum_\alpha \left[|P_\alpha^{(0)}| + \|P_\alpha^{(1)}(x)\|_{h^{\sigma+2}}\right] \le c_1 < \infty.$$

Then

$$\left\langle P_{\Delta x}v, A^{-1}P_{\Delta x}v\right\rangle_{\ell^2} \le (1+\eta+c\delta\Delta x)\left\langle v, Av\right\rangle_{\ell^2}, \tag{4.31}$$

where c depends only on the c_j.

Proof. When (4.29) holds, then by Lemma 4.7 and Theorem 4.8,

$$\begin{aligned}\left\langle P_{\Delta x}v, A^{-1}P_{\Delta x}v\right\rangle_{\ell^2} &= \left\langle v, P_{\Delta x}^* A^{-1}P_{\Delta x}v\right\rangle_{\ell^2}\\ &\le \left\langle v, \operatorname{Op}\left[\operatorname{Symb}(P_{\Delta x})^*\operatorname{Symb}(A^{-1}P_{\Delta x})\right]v\right\rangle_{\ell^2} + k_1\Delta x\|v\|_{\ell^2_A}^2\\ &= \left\langle v, \operatorname{Op}\left[\operatorname{Symb}(P_{\Delta x})^*A^{-1}\operatorname{Symb}(P_{\Delta x}) - (1+\eta)A\right]v\right\rangle_{\ell^2} + (1+\eta+k_1\Delta x)\left\langle v, Av\right\rangle_{\ell^2}\\ &\le k_2\Delta x\left\langle v, Av\right\rangle_{\ell^2} + (1+\eta+k_1\Delta x)\left\langle v, Av\right\rangle_{\ell^2} = (1+\eta+c\Delta x)\left\langle v, Av\right\rangle_{\ell^2}.\end{aligned} \tag{4.32}$$

When A has constant coefficients, a calculation similar to (4.32) but using the variant results (4.20) and (4.25) shows that (4.31) holds. $\square$

5. Energy estimates

5.1. Estimates for the θ-scheme

Lemma 5.1. *As usual, let $\sigma := \lfloor d/2 \rfloor + 1$ denote the Sobolev embedding exponent. Assume that the hypotheses of Theorem 2.2 hold for some s, and let K_1 and K_2 be finite constants. Then there exist a positive Δ_0 and finite κ_0 and R_0 such that for all $\Delta t \in (0, \Delta_0]$, all $T > 0$, all $u_{0,\varepsilon,\Delta x} \in \ell^2$, all $\widetilde{F} \in \ell^\infty([0,T] \cap \Delta t\mathbb{Z}; \ell^2)$, and all V satisfying*

$$\sup_{t\in[0,T]\cap\Delta t\mathbb{Z}} \|V\|_{h^{\sigma+1}} \le K_1, \quad \varepsilon \sup_{t\in[0,T-\Delta t]\cap\Delta t\mathbb{Z}} \|D_{\Delta t} V\|_{h^\sigma} \le K_2 \tag{5.1}$$

the linearized numerical scheme

$$A(\varepsilon V) D_{\Delta t} v = \sum_{j=1}^d A^j(t,x,V^\theta)\partial_{j,\Delta x,c} v^\theta - \sum_{j,k=1}^d (\partial_{j,\Delta x})^* \left[B^{j,k}\partial_{k,\Delta x} v^\theta\right] + \frac{1}{\varepsilon}\Big[\sum_{j=1}^d C^j \partial_{j,\Delta x,c} + D\Big] v^\theta + \widetilde{F}(t,x), \tag{5.2}$$

$$v(0) = v_{0,\varepsilon,\Delta x} \tag{5.3}$$

has a unique solution in $\ell^\infty([0,T]\cap\Delta t\mathbb{Z};\ell^2)$ satisfying

$$\|S_{\Delta t} v\|^2_{\ell^2_{S_{\Delta t}A}} \le (1 + R_0\Delta t)\|v\|^2_{\ell^2_A} + \kappa_0 \Delta t \|\widetilde{F}\|^2_{\ell^2} \tag{5.4}$$

for all $t \in [0, T-\Delta t] \cap \Delta t\mathbb{Z}$. The constants Δ_0, κ_0, and R_0 depend only on the bounds K_1 and K_2 on V, the bounds M_s and m_0 in (2.6)–(2.7), and the dimension d.

Proof. The *a priori* estimate (5.4) says that at each time the square of the norm of the solution at the next time step is bounded by a constant times the squares of the norms of the inhomogeneous term and the solution at the current time step. Since the difference scheme (5.2) is linear, such an estimate implies the existence and uniqueness of the solution. Hence it suffices to prove (5.4).

The key estimate for dealing with the time-difference term in (5.2) is

$$D_{\Delta t}(\langle v, Av\rangle_{\ell^2}) \le 2\left\langle v^\theta, A D_{\Delta t} v\right\rangle_{\ell^2} + \langle S_{\Delta t} v, (D_{\Delta t} A)\, S_{\Delta t} v\rangle_{\ell^2}. \tag{5.5}$$

The first step towards deriving (5.5) is to use the definitions of the time-average v^θ and time-difference $D_{\Delta t} v$ and the symmetry of A to obtain

$$(v^\theta)^T A D_{\Delta t} v = \frac{\theta (S_{\Delta t} v)^T A S_{\Delta t} v - (2\theta - 1) v(t)^T A S_{\Delta t} v - (1-\theta) v(t)^T A v(t)}{\Delta t}. \tag{5.6}$$

Since $2\theta - 1 \ge 0$, estimating the mixed term $v(t)^T A S_{\Delta t} v$ on the right side of (5.6) *via* the Cauchy–Schwarz inequality and simplifying yields

$$\frac{\theta (S_{\Delta t} v)^T A S_{\Delta t} v - (2\theta - 1) v(t)^T A S_{\Delta t} v - (1-\theta) v(t)^T A v(t)}{\Delta t} \ge \tfrac{1}{2}\frac{(S_{\Delta t} v)^T A S_{\Delta t} v - v(t)^T A v(t)}{\Delta t}. \tag{5.7}$$

If the matrix A is constant then the right side of (5.7) is an exact time difference. When A is not constant then it is evaluated at time t everywhere in (5.7) so the right side of that expression is not an exact time difference, but it can be expressed as a time difference plus a correction term involving the time difference of A, *i.e.*,

$$\tfrac{1}{2}\frac{(S_{\Delta t} v)^T A S_{\Delta t} v - v(t)^T A v(t)}{\Delta t} = \tfrac{1}{2}\left[D_{\Delta t}(v \cdot Av) - (S_{\Delta t} v)\cdot (D_{\Delta t} A)(S_{\Delta t} v)\right]. \tag{5.8}$$

Combining (5.6)–(5.8) yields (5.5). Substituting for $AD_{\Delta t}v$ in (5.5) the value it has according to the numerical scheme (5.2) yields

$$D_{\Delta t}\left\langle v, Av\right\rangle_{\ell^2} \le \left\langle S_{\Delta t}v, (D_{\Delta t}A)\, S_{\Delta t}v\right\rangle_{\ell^2} + 2\sum_{j=1}^{d}\left\langle v^\theta, A^j\partial_{j,\Delta x,c}v^\theta\right\rangle_{\ell^2}$$
$$-2\sum_{j,k=1}^{d}\left\langle v^\theta, (\partial_{j,\Delta x})^*\left[B^{j,k}\partial_{k,\Delta x}v^\theta\right]\right\rangle_{\ell^2} + \frac{2}{\varepsilon}\left\langle v^\theta, \left[\sum_{j=1}^{d}C^j\partial_{j,\Delta x,c} + D\right]v^\theta\right\rangle_{\ell^2} + 2\left\langle v^\theta, \widetilde{F}\right\rangle_{\ell^2}. \tag{5.9}$$

We now turn to estimating each of the terms on the right side of (5.9).

When the assumption (2.12) on the coefficients $B^{j,k}$ holds and the difference operators $\partial_{j,\Delta x}$ are centered then using successively the definition of an adjoint operator, Parseval's identity (4.5) plus formula (4.6) for the Fourier transform of a difference operator and the identity $2w^TBw = w^T(B+B^T)w$, (2.12) and the fact that the symbol of a centered difference operator is purely imaginary and hence the product of two such symbols is real, and the Plancherel identity (4.4), the second-order terms in (5.9) can be estimated by

$$-2\sum_{j,k=1}^{d}\left\langle v^\theta, (\partial_{j,\Delta x})^*\left[B^{j,k}\partial_{k,\Delta x}v^\theta\right]\right\rangle_{\ell^2} = -2\sum_{j,k=1}^{d}\left\langle \partial_{j,\Delta x}v^\theta, B^{j,k}\partial_{k,\Delta x}v^\theta\right\rangle_{\ell^2}$$
$$= -c\sum_{j,k=1}^{d}\sum_{\xi\in\Pi_{\Delta x}}\mathrm{Symb}(\partial_{j,\Delta x})(\Delta x\,\xi)\overline{\mathrm{Symb}(\partial_{k,\Delta x})(\Delta x\,\xi)}\widehat{v^\theta}(\xi)^T\left[B^{j,k}+(B^{j,k})^T\right]\overline{\widehat{v^\theta}(\xi)}\,(\Delta\xi)^d$$
$$\le -cb_0\sum_{\xi\in\Pi_{\Delta x}}\sum_j\left|\mathrm{Symb}(\partial_{j,\Delta x})(\Delta x\,\xi)\right|^2\left|\widehat{v^\theta}(\xi)\right|^2(\Delta\xi)^d = -cb_0\sum_j\left\|\partial_{j,\Delta x}v^\theta\right\|_{\ell^2}^2, \tag{5.10}$$

since (2.12) for real vectors implies that the same inequality holds for the real part when the vectors are complex and one factor is conjugated, and the fact that the overall expression is real means that the imaginary part can be ignored. Note that this argument uses the assumption that the difference operators $\partial_{j,\Delta x}$ are centered since that ensures that their symbols are purely imaginary and hence that the product of two such symbols is real. When the stronger assumption (2.11) on the $B^{j,k}$ holds then the last line of (5.10) is obtained directly as an estimate for the second line there, without assuming that the $\partial_{j,\Delta x}$ are centered differences.

Using the symmetry of the inner product, formulas for adjoint operators, the antisymmetry of central differences, the fact that the central differences in the first-order terms can be written in terms of elementary central differences as $\partial_{j,\Delta x,c} = \sum_{0<m\le M_j} d_{m,j}D_{j,m\Delta x,c}$, and the product rule (4.13) for $D_{j,m\Delta x,c}$ lets us write and then estimate the first-order terms as

$$2\sum_{j=1}^{d}\left\langle v^\theta, A^j\partial_{j,\Delta x,c}v^\theta\right\rangle_{\ell^2} = \sum_{j=1}^{d}\left\langle v^\theta, \left[A^j\partial_{j,\Delta x,c} + (\partial_{j,\Delta x,c})^*(A^j)^T\right]v^\theta\right\rangle_{\ell^2}$$
$$= \sum_{j=1}^{d}\left\langle v^\theta, \left[A^j\partial_{j,\Delta x,c} - \partial_{j,\Delta x,c}A^j\right]v^\theta\right\rangle_{\ell^2}$$
$$= -\sum_{j=1}^{d}\left\langle v^\theta, \sum_{0<m<M_j}d_{m,j}\Big[\left(D_{j,m\Delta x}A^j\right)S_{j,m\Delta x} + \left(D_{j,m\Delta x}(S_{j,m\Delta x})^{-1}A^j\right)(S_{j,m\Delta x})^{-1}\Big]v^\theta\right\rangle_{\ell^2}$$
$$\le c\sum_{j=1}^{d}\|D_{j,\Delta x}A^j\|_{\ell^\infty}\|v^\theta\|_{\ell^2}^2. \tag{5.11}$$

Since all the terms after the last equals sign in (5.11) involve differences of the A^j, a similar calculation shows that the large terms in (5.9), which have constant coefficients, cancel to zero. Substituting (5.10) and (5.11) into (5.9) and taking into account the cancellation of the large terms, using the Cauchy–Schwarz inequality for the term involving $\widetilde{F}$, using the definition of v^θ and the estimate $\|a+b\|^2 \le 2\|a\|^2 + 2\|b\|^2$, and using the positivity of A guaranteed by (2.8) then yields

$$\begin{aligned} D_{\Delta t}\left\langle v, Av\right\rangle_{\ell^2} &\le \|D_{\Delta t}A\|_{\ell^\infty}\|S_{\Delta t}v\|^2_{\ell^2} + c\sum_{j=1}^d \|D_{j,\Delta x}A^j\|_{\ell^\infty}\|v^\theta\|^2_{\ell^2} + \|v^\theta\|_{\ell^2}\|\widetilde{F}\|_{\ell^2} - cb_0\sum_j \left\|\partial_{j,\Delta x}v^\theta\right\|^2_{\ell^2} \\ &\le c_5\left[1 + \|D_{\Delta t}A\|_{\ell^\infty} + \sum_{j=1}^d \|D_{j,\Delta x}A^j\|_{\ell^\infty}\right]\left[\|S_{\Delta t}v\|^2_{\ell^2} + \|v\|^2_{\ell^2}\right] + \frac{1}{2}\|\widetilde{F}\|^2_{\ell^2} \\ &\le c_6(\|V\|_{\ell^\infty}, \|S_{\Delta t}V\|_{\ell^\infty})\left[1 + \|D_{\Delta t}A\|_{\ell^\infty} + \sum_{j=1}^d \|D_{j,\Delta x}A^j\|_{\ell^\infty}\right]\left[\|S_{\Delta t}v\|^2_{\ell^2_{S_{\Delta t}A}} + \|v\|^2_{\ell^2_A}\right] + \frac{\|\widetilde{F}\|^2_{\ell^2}}{2}. \end{aligned} \tag{5.12}$$

By the discrete chain rule (4.11) and its time-difference variant plus the smoothness assumption (2.6),

$$\begin{aligned} \|D_{\Delta t}A(\varepsilon V)\|_{\ell^\infty} &\le c_7(\|V\|_{\ell^\infty}, \|S_{\Delta t}V\|_{\ell^\infty})\varepsilon\|D_{\Delta t}V\|_{\ell^\infty}, \\ \|D_{j,\Delta x}A^j(t,x,V)\|_{\ell^\infty} &\le c_8(\|V\|_{\ell^\infty})(1 + \|D_{j,\Delta x}V\|_{\ell^\infty}). \end{aligned}$$

Inserting these estimates into (5.12), using the discrete Sobolev embedding estimate (4.7) to replace all ℓ^∞ norms by Sobolev space norms, and using the definition of the time-difference operator in the expression $D_{\Delta t}\left\langle v, Av\right\rangle_{\ell^2}$ yields

$$\begin{aligned} &\left[1 - c_9(\|V\|_{h^{\sigma+1}}, \|S_{\Delta t}V\|_{h^\sigma}, \varepsilon\|D_{\Delta t}V\|_{h^\sigma})\Delta t\right]\|S_{\Delta t}v\|^2_{\ell^2_{S_{\Delta t}A}} \\ &\qquad \le \left[1 + c_9(\|V\|_{h^{\sigma+1}}, \|S_{\Delta t}V\|_{h^\sigma}, \varepsilon\|D_{\Delta t}V\|_{h^\sigma})\Delta t\right]\|v\|^2_{\ell^2_A} + \frac{\Delta t}{2}\|\widetilde{F}\|^2_{\ell^2}. \end{aligned} \tag{5.13}$$

The derivation of the bound c_9 shows that it may be taken to be continuous and nondecreasing, so under the assumptions of the lemma

$$c_9(\|V\|_{h^{\sigma+1}}, \|S_{\Delta t}V\|_{h^\sigma}, \varepsilon\|D_{\Delta t}V\|_{h^\sigma}) \le c_{10} := c_9(K_1, K_1, K_2) \tag{5.14}$$

for all $t \in [0, T-\Delta t] \cap \Delta t\mathbb{Z}$. Pick any constant η less than one and define $\Delta_0 := \frac{\eta}{c_{10}}$; then for $0 < \Delta t \le \Delta_0$ the quantity $1 - c_{10}\Delta t$ is positive, so the estimate obtained by substituting (5.14) into (5.13) can be solved for $\|S_{\Delta t}v\|^2_{\ell^2_{S_{\Delta t}A}}$, yielding

$$\|S_{\Delta t}v\|^2_{\ell^2_{S_{\Delta t}A}} \le \tfrac{1+c_{10}\Delta t}{1-c_{10}\Delta t}\|v\|^2_{\ell^2_A} + \tfrac{\Delta t}{2(1-\eta)}\|\widetilde{F}\|^2_{\ell^2} \le \left[1 + \tfrac{2c_{10}}{1-\eta}\Delta t\right]\|v\|^2_{\ell^2_A} + \tfrac{\Delta t}{2(1-\eta)}\|\widetilde{F}\|^2_{\ell^2} \tag{5.15}$$

for all $t \in [0, T-\Delta t] \cap \Delta t\mathbb{Z}$. Estimate (5.15) has the desired form (5.4) with $R_0 := \frac{2c_{10}}{1-\eta}$ and $\kappa_0 := \frac{1}{2(1-\eta)}$, and our calculation shows that those bounds and Δ_0 depend only on the claimed quantities. $\square$

Lemma 5.2. *Assume that the hypotheses of Theorem 2.2 hold for some s, and let K_1 and K_2 be finite constants. Then there exist a positive Δ_s and finite κ_s and R_s such that for all $\Delta t \in (0, \Delta_s]$, all $T > 0$, all $u_{0,\varepsilon,\Delta x} \in h^s$, all $\widetilde{F} \in \ell^\infty([0,T] \cap \Delta t\mathbb{Z}; h^s)$, and all V satisfying*

$$\sup_{t\in[0,T]\cap\Delta t\mathbb{Z}} \|V\|_{h^s} \le K_1, \quad \varepsilon \sup_{t\in[0,T-\Delta t]\cap\Delta t\mathbb{Z}} \|D_{\Delta t}V\|_{h^\sigma} \le K_2 \tag{5.16}$$

the solution of linearized numerical scheme (5.2) with initial data (5.3) satisfies the additional estimates

$$\|S_{\Delta t}v\|^2_{h^s_{S_{\Delta t}A}} \le (1 + R_s\Delta t)\|v\|^2_{h^s_A} + \kappa_s\Delta t\|\widetilde{F}\|^2_{h^s} \tag{5.17}$$

for all $t \in [0, T-\Delta t] \cap \Delta t\mathbb{Z}$, *and*

$$\begin{aligned}\|S_{\Delta t}(D_{\Delta t}v)\|^2_{h^r_{S_{\Delta t}A}} &\le (1+R_s\Delta t)\|D_{\Delta t}v\|^2_{h^r_A} \\ &\quad + \kappa_s\Delta t\Big(\|D_{\Delta t}\widetilde{F}\|^2_{h^r} + \|D_{\Delta t}V\|^2_{h^r_A}\|S_{\Delta t}v\|^2_{h^{s-1}_{S_{\Delta t}A}} + \|S_{\Delta t}v\|^2_{h^{s-1}_A} + \|(S_{\Delta t})^2 v\|^2_{h^{s-1}_A}\Big)\end{aligned} \tag{5.18}$$

for $0 \le r \le s-2$, *all* $t \in [0, T-2\Delta t] \cap \Delta t\mathbb{Z}$. *The constants* Δ_s, κ_s, *and* R_s *depend only on the bounds* K_1 *and* K_2 *on* V, *the bounds* M_s *and* m_0 *in* (2.6)–(2.7), *the smoothness parameter* s *and the dimension* d.

Proof. Applying a spatial difference operator $D^\alpha_{\Delta x}$ to the scheme (5.2) yields

$$\begin{aligned}A(\varepsilon V)D_{\Delta t}(D^\alpha_{\Delta x}v) &= \sum_{j=1}^d A^j(t,x,V^\theta)\partial_{j,\Delta x,c}(D^\alpha_{\Delta x}v)^\theta - \sum_{j,k=1}^d (\partial_{j,\Delta x})^*\left[B^{j,k}\partial_{k,\Delta x}(D^\alpha_{\Delta x}v)^\theta\right] \\ &\quad + \frac{1}{\varepsilon}\left[\sum_{j=1}^d C^j\partial_{j,\Delta x,c} + D\right](D^\alpha_{\Delta x}v)^\theta + \widetilde{F}_\alpha,\end{aligned} \tag{5.19}$$

where

$$\widetilde{F}_\alpha = D^\alpha_{\Delta x}\widetilde{F} - [D^\alpha_{\Delta x}, A]D_{\Delta t}v + \sum_{j=1}^d [D^\alpha_{\Delta x}, A^j]\partial_{j,\Delta x,c}v^\theta. \tag{5.20}$$

Since (5.19) has the form of (5.2) with v replaced by $D^\alpha_{\Delta x}v$ and $\widetilde{F}$ replaced by $\widetilde{F}_\alpha$, we can apply the estimate (5.4) of Lemma 5.1 with those substitutions and then sum over $0 \le |\alpha| \le s$ and use the definition of the h^s norm to obtain

$$\|S_{\Delta t}v\|^2_{h^s_{S_{\Delta t}A}} \le (1+R_0\Delta t)\|v\|^2_{h^s_A} + \kappa_0\Delta t\sum_{0\le|\alpha|\le s}\|\widetilde{F}_\alpha\|^2_{\ell^2}. \tag{5.21}$$

We now need to estimate the square of the ℓ^2 norm of the $\widetilde{F}_\alpha$. For any norm,

$$\left\|\sum_{j=1}^N a_j\right\|^2 \le N\sum_{j=1}^N \|a_j\|^2, \tag{5.22}$$

so it suffices to estimate the square of the ℓ^2 norms of each of the terms on the right side of (5.20). For the first term we obtain simply

$$\sum_{0\le|\alpha|\le s}\|D^\alpha_{\Delta x}\widetilde{F}\|^2_{\ell^2} = \|\widetilde{F}\|^2_{h^s}. \tag{5.23}$$

The second term on the right side of (5.20) is the most difficult, because of the presence of the time difference. To treat this term, note first that since $A^o(0)$ is independent of x, $[D^\alpha_{\Delta x}, A(0)] = 0$. In particular, if A is identically constant then the whole second term vanishes. By the alternative in assumption (2.14), we can therefore assume while estimating this terms that all the $B^{j,k}$ vanish. By the fact just noted, the Moser estimates (4.18) and (4.17) show that this term may be estimated by

$$\begin{aligned}\sum_{0\le|\alpha|\le s}\|[D^\alpha_{\Delta x}, A(\varepsilon V)]D_{\Delta t}v\|^2_{\ell^2} &= \|[D^\alpha_{\Delta x}, A(\varepsilon V)-A(0)]D_{\Delta t}v\|^2_{\ell^2} \\ &\le c_1\|A(\varepsilon V)-A(0)\|^2_{h^s}\|D_{\Delta t}v\|^2_{h^{s-1}} \le \varepsilon^2 c_2(\|\varepsilon V\|_{h^s})\|V\|^2_{h^s}\|D_{\Delta t}v\|^2_{h^{s-1}}.\end{aligned} \tag{5.24}$$

By the difference equation (5.2) plus the above assumption, $\varepsilon D_{\Delta t}v$ equals a sum of spatial first-differences of v^θ plus the undifferentiated term $\widetilde{F}$, each multiplied by some matrix of size at most $O(1)$. The Moser estimate (4.17) with s replaced by $s-1$ and $\mathcal{F} = \mathcal{F}(t,x,V,V^\theta,\partial_{j,\Delta x,c}v^\theta) := \varepsilon A(\varepsilon V)^{-1}A^j(t,x,V^\theta)\partial_{j,\Delta x,c}v^\theta$ for one particular term, with similar definitions for the other terms, yields the estimate

$$\varepsilon\|D_{\Delta t}v\|_{h^{s-1}} \le c_3(\|V\|_{h^{s-1}}, \|S_{\Delta t}V\|_{h^{s-1}})\,(\|v\|_{h^s} + \|S_{\Delta t}v\|_{h^s})\,. \tag{5.25}$$

Substituting (5.25) into (5.24) shows that

$$\sum_{0\le|\alpha|\le s}\|[D^\alpha_{\Delta x}, A(\varepsilon V)]D_{\Delta t}v\|^2_{\ell^2} \le c_4(\|V\|_{h^s}, \|S_{\Delta t}V\|_{h^{s-1}})\left(\|v\|^2_{h^s} + \|S_{\Delta t}v\|^2_{h^s}\right). \tag{5.26}$$

Write $A^j(t,x,V^\theta) = A^j(t,x,0) + \left(A^j(t,x,V^\theta) - A^j(t,x,0)\right)$, use (4.19) to estimate the commutator of $D^\alpha_{\Delta x}$ with $A^j(t,x,0)$, and use the same method as for A (but without the need to convert any time difference) to estimate the commutator with $A^j(t,x,V^\theta) - A^j(t,x,0)$. This shows that the remaining terms on the right side of (5.20) are also bounded by the right side of (5.26), possibly with a different c_4. Substituting the estimates for all the terms from (5.20) into (5.21), using (2.8) to convert the h^s norms of v and $S_{\Delta t}v$ into the h^s_A and $h^s_{S_{\Delta t}A}$ norms, respectively, and using the assumed bounds (5.16) on V to replace all bounds involving V by constants yields

$$\|S_{\Delta t}v\|^2_{h^s_{S_{\Delta t}A}} \le (1+R_0\Delta t)\|v\|^2_{h^s_A} + \kappa_0\Delta t\left(\|\widetilde{F}\|^2_{h^s} + c_5\left(\|v\|^2_{h^s_A} + \|S_{\Delta t}v\|^2_{h^s_{S_{\Delta t}A}}\right)\right). \tag{5.27}$$

Picking any constant η less than one and calculating as at the end of the proof of Lemma 5.1 yields (5.17) with $\Delta_s = \min(\Delta_0, \frac{\eta}{c_5\kappa_0})$, $R_s = \frac{R_0+2c_5\kappa_0}{1-\eta}$, and $\kappa_s = \frac{\kappa_0}{1-\eta}$.

Applying to (5.2) the time difference operator $D_{\Delta t}$ while using the product rule (4.12), then applying a spatial difference operator $D^\beta_{\Delta x}$ with $|\beta| \le r$, and estimating in similar fashion to the calculation above yields (5.18), possibly after increasing the values of R_s and κ_s. No additional assumptions are required on $\widetilde{F}$, V or the initial data because the lemma does not claim that the norms on the right side of (5.18) are bounded uniformly in Δt or ε. ☐

Proof of Theorem 2.2. Let $v = \mathcal{T}(V)$ denote the mapping sending the argument V of the coefficients of the linearized difference scheme (5.2), with $\widetilde{F}(t,x)$ chosen to be $F(t,x,V(t,x))$, to the solution v of that scheme. We will prove the existence of a unique solution of the nonlinear difference scheme (2.5) by applying the contraction-mapping theorem to the mapping $\mathcal{T}$ on the set S of functions in $\ell^\infty([0,T]\cap\Delta t\mathbb{Z}; h^s)$ satisfying the bounds $\|w(0)\|_{h^\sigma} \le K_0$, $\sup_{t\in[0,T]\cap\Delta t\mathbb{Z}}\|w(t)\|_{h^s} \le K$, and $\sup_{t\in[0,T-\Delta t]\cap\Delta t\mathbb{Z}}\varepsilon\|D_{\Delta t}w(t)\|_{h^\sigma} \le b := 2c_3(K,K)$, for some appropriate constants T and K, where K_0 is the assumed bound on the h^s norm of the initial data in the statement of the theorem and c_3 is the function appearing in (5.25).

We now show that K and T can be chosen so that $\mathcal{T}$ maps the set S into itself. By (2.6), the assumed bound $\|u_{0,\varepsilon,\Delta x}\|_{h^s} \le K_0$ for the standard norm of the initial data yields the bound

$$\|v\|^2_{h^s_{A(\varepsilon V)}}\Big|_{t=0} = \|u_{0,\varepsilon,\Delta x}\|^2_{h^s_{A(\varepsilon V)}} \le M_s(\varepsilon_0 c(d)K_0)K_0^2 \tag{5.28}$$

for the square of the h^s_A norm, where ε_0 is the constant chosen in the statement of the theorem, M_s is the function appearing in (2.6), and $c(d)$ is the discrete Sobolev embedding bound in (2.4). We will apply Lemma 5.2 with $K_1 := K$ and $K_2 := b$, and the bound Δ_0 on the allowed values of Δt for the theorem will be the bound Δ_s obtained in that lemma for those values of K_1 and K_2. Note that writing $F(t,x,V) = F(t,x,0) + [F(t,x,V) - F(t,x,0)]$ and using the bound (2.6) and the Moser estimate (4.17) yields a bound $\|F(t,x,V)\|^2_{h^s} \le \widetilde{c}_F(\|V\|_{h^s}) \le \widetilde{c}_F(K)$. Substituting this bound into (5.17) yields

$$\|S_{\Delta t}v\|^2_{h^s_{S_{\Delta t}A}} \le (1+R_s\Delta t)\|v\|^2_{h^s_A} + \kappa_s\widetilde{c}_F(K)\Delta t. \tag{5.29}$$

By Lemma 4.6 with $w(t) := \|v(t)\|^2_{h^s_{A(\varepsilon V(t))}}$, (5.28)–(5.29) imply that

$$\|v\|^2_{h^s_A} \le \mathrm{e}^{R_s t}M_s(\varepsilon_0 c(d)K_0)K_0^2 + \tfrac{\mathrm{e}^{R_s t}-1}{R_s}\kappa_s\widetilde{c}_F(K) \tag{5.30}$$

for $t \in [0,T] \cap \Delta t\mathbb{Z}$, where R_s and κ_s also depend on K. Moreover, by the definition of the set S and the fact that $\varepsilon V(t) = \varepsilon V(0) + \sum_{j=0}^{\frac{t}{\Delta t}-1} \varepsilon D_{\Delta t} V(j\Delta t)\,\Delta t$,

$$\varepsilon \|V\|_{\ell^\infty} \le c(d)\varepsilon \|V\|_{h^\sigma} \le c(d)(\varepsilon_0 K_0 + bt). \tag{5.31}$$

Using (5.31) to estimate the argument of m_0 in (2.8) together with the bound (5.30) shows that

$$\|v\|_{h^s}^2 \le m_0(c(d)(\varepsilon_0 K_0 + bt)) \left(\mathrm{e}^{R_s t} M_s(\varepsilon_0 c(d) K_0) K_0^2 + \tfrac{\mathrm{e}^{R_s t}-1}{R_s} \kappa_s \widetilde{c}_F(K) \right). \tag{5.32}$$

The right side of (5.32) is an increasing function of t whose value at time zero is $m_0(c(d)\varepsilon_0 K_0) M_s(\varepsilon_0 c(d) K_0) K_0^2$. We therefore pick K to be any number greater than the square root of that quantity, and let T_1 be the value of t at which the right side of (5.32), with that value of K substituted into the parameters that depend on it, equals K^2. By construction, $T_1 > 0$. Moreover, the initial data for the linearized scheme satisfy the bound $\|v(0)\|_{h^\sigma} = \|v_{0,\varepsilon,\Delta x}\|_{h^\sigma} \le K_0$, and estimate (5.25) shows that $\varepsilon \|D_{\Delta t} v\|_{h^\sigma} \le b$. Hence for any $T \le T_1$ the mapping $\mathcal{T}$ indeed maps S into itself.

Now let V and $\widetilde{V}$ be two elements of S and consider $v := \mathcal{T}(V)$ and $\widetilde{v} := \mathcal{T}(\widetilde{V})$. Take the difference of the equations (5.2) for v and $\widetilde{v}$, with $\widetilde{F}$ in (5.2) defined as $F(t,x,V)$ in the equation for v and as $F(t,x,\widetilde{V})$ in the equation for $\widetilde{v}$ in accordance with the definition of $\mathcal{T}$. After rearranging to express the result in terms of $w := v - \widetilde{v}$ and $W := V - \widetilde{V}$ we obtain

$$A(\varepsilon V) D_{\Delta t} w = \sum_{j=1}^d A^j(t,x,V^\theta)\partial_{j,\Delta x,c} w^\theta - \sum_{j,k=1}^d (\partial_{j,\Delta x})^* \left[B^{j,k}\partial_{k,\Delta x} w^\theta\right] + \frac{1}{\varepsilon}\left[\sum_{j=1}^d C^j \partial_{j,\Delta x,c} + D\right] w^\theta + \mathcal{F}, \tag{5.33}$$

$$w(0) = 0, \tag{5.34}$$

where the inhomogeneous term

$$\mathcal{F} := \left\{ F(t,x,\widetilde{V}+W) - F(t,x,\widetilde{V}) \right\} + \left\{ \tfrac{A(\varepsilon\widetilde{V}+\varepsilon W) - A(\varepsilon \widetilde{V})}{\varepsilon} \right\} \varepsilon D_{\Delta t} v - \sum_{j=1}^d \left\{ A^j(\varepsilon\widetilde{V}+\varepsilon W) - A^j(\varepsilon\widetilde{V}) \right\} \partial_{j,\Delta x,c} v^\theta \tag{5.35}$$

can be estimated in similar fashion to estimates in the lemmas above to yield $\|\mathcal{F}\|_{\ell^2}^2 \le c_{\mathcal{F}}(K,b)\|W\|_{\ell^2}^2$. Using this bound in estimate (5.4) and then applying Lemma 4.6 and translating back to the standard ℓ^2 norm *via* (2.8) yields $\|w(t)\|_{\ell^2} \le \rho(t) \max_{0\le s\le t} \|W\|_{\ell^2}$, where $\rho(t)^2 = m_0(c(d)\varepsilon K)\frac{\mathrm{e}^{R_0 t}-1}{R_0}\kappa_0 c_{\mathcal{F}}(K,b)$. Picking any $\widetilde{\rho} < 1$ and defining T_2 to be the value of t at which $\rho(t) = \widetilde{\rho}$ shows that $\mathcal{T}$ is a contraction on S in the ℓ^2 norm provided that we set $T = \min(T_1, T_2)$. As for the PDE case the set S is closed in the ℓ^2 norm; in fact this is trivial for the periodic finite difference case since the spaces ℓ^2 and h^k are finite-dimensional; in particular, as already noted in (2.3), the corresponding norms are equivalent for fixed Δx. Hence the conditions of the contraction-mapping theorem hold and so $\mathcal{T}$ has a unique fixed point in S, which is the claimed solution of the nonlinear difference scheme. The claimed h^s bound on the solution holds by the definition of the set S.

When the initial data satisfies (2.15) with $r \le s-2$ then the first step towards obtaining a uniform bound for $\|D_{\Delta t} u\|_{h_A^r}$ is to show that it is uniformly bounded at time zero. To do this we apply $D_{\Delta x}^\beta$ to (2.5), take the ℓ^2 inner product of the result with $D_{\Delta x}^\beta D_{\Delta t} v$, add the result over $|\beta| \le \widetilde{r} \le r$, and set $t = 0$. On the left side we obtain $\|D_{\Delta t} u\,|_{t=0}\|_{h_A^{\widetilde{r}}}^2$ plus the commutator term $\sum_{|\beta|\le\widetilde{r}} \left\langle D_{\Delta x}^\beta D_{\Delta t} u\,|_{t=0}, [D_{\Delta x}^\beta, A(\varepsilon u)] D_{\Delta t} u\,|_{t=0} \right\rangle_{\ell^2}$, which vanishes when $\widetilde{r} = 0$ and is bounded in absolute value by $c\|D_{\Delta t} u\,|_{t=0}\|_{h_A^{\widetilde{r}}} \|D_{\Delta t} u\,|_{t=0}\|_{h_A^{\widetilde{r}-1}}$ when $\widetilde{r} > 0$. The terms on the right side containing A^j, $B^{j,k}$ and F are bounded by a constant times $\|D_{\Delta t} u\,|_{t=0}\|_{h_A^s}$ on account of the h^s bound for u, the Moser estimates, and the fact that $\widetilde{r} \le s-2$. To estimate the large terms, write the time average $u^\theta|_{t=0}$ appearing in them as $u_{0,\varepsilon,\Delta x} + \theta\Delta t D_{\Delta t} u\,|_{t=0}$; the part of the large terms containing $\theta \Delta t D_{\Delta t} u\,|_{t=0}$

vanishes by the antisymmetry of the operator in those terms, while by assumption the part containing $u_{0,\varepsilon,\Delta x}$ is bounded by a constant times $\|D_{\Delta t}u\,|_{t=0}\|_{h_A^{\tilde r}}$. Combining these estimates and dividing by the common factor of $\|D_{\Delta t}u\,|_{t=0}\|_{h_A^s}$ yields $\|D_{\Delta t}u\,|_{t=0}\|_{h_A^0} \le c$ and $\|D_{\Delta t}u\,|_{t=0}\|_{h_A^{\tilde r}} \le c + \|D_{\Delta t}u\,|_{t=0}\|_{h_A^{\tilde r-1}}$ for $1 \le \widetilde r \le r$, which by finite induction yields the claimed uniform bound at time zero. Combining this bound with estimate (5.18), Lemma 4.6, and (2.8) then yield the uniform bound on $D_{\Delta t}u$ at later times claimed in the theorem. ☐

5.2. Energy estimates *via* Gårding's inequality

Lemma 5.3. *Assume that the hypotheses of Theorem 2.4 hold for some s, and let K_1 and K_2 be finite constants. Then there exist finite κ_0 and R_0 such that for all $\Delta t \in (0,\Delta_0]$, all $T>0$, all $u_{0,\varepsilon,\Delta x} \in \ell^2$, all $\widetilde F \in \ell^\infty([0,T]\cap \Delta t\mathbb{Z};\ell^2)$, and all V satisfying*

$$\sup_{t\in[0,T]\cap\Delta t\mathbb{Z}} \|V\|_{h^{\sigma+2}} \le K_1, \quad \varepsilon \sup_{t\in[0,T-\Delta t]\cap\Delta t\mathbb{Z}} \|D_{\Delta t}V\|_{h^\sigma} \le K_2 \tag{5.36}$$

the linearized numerical scheme

$$\mathcal{A}_{\Delta x}(\Delta,\varepsilon\widetilde V)S_{\Delta t}v + \tfrac{\Delta t}{\varepsilon}Q_{\Delta x}(\Delta)S_{\Delta t}v = \mathcal{G}_{\Delta x}\left(\Delta,t,x,\varepsilon\widetilde V,\widetilde V\right)v + \Delta t\widetilde F(t,x), \tag{5.37}$$

$$v(0) = v_{0,\varepsilon,\Delta x}, \tag{5.38}$$

where $\widetilde V$ denotes any finite collection $\{S^\alpha_{\Delta x}V\}_{|\alpha|\le M}$ of spatial shifts of V, has a unique solution in $\ell^\infty([0,T]\cap\Delta t\mathbb{Z};\ell^2)$ satisfying

$$\|S_{\Delta t}v\|^2_{h^0_{\mathcal{A}(\varepsilon S_{\Delta t}V)}} \le (1+R_0\Delta t)\|v\|^2_{h^0_{\mathcal{A}(\varepsilon V)}} + \kappa_0\Delta t\|\widetilde F\|^2_{\ell^2} \tag{5.39}$$

for all $t\in[0,T-\Delta t]\cap\Delta t\mathbb{Z}$. The constants κ_0 and R_0 depend only on the bounds K_1 and K_2 on V, the bounds in the assumptions of Theorem 2.4 and the dimension d.

Proof. Taking the ℓ^2 inner product of (5.37) with $S_{\Delta t}v$ and using the antisymmetry of the operator $Q_{\Delta x}$ yields

$$\|S_{\Delta t}v\|^2_{h^0_{\mathcal{A}(\varepsilon V)}} = \langle S_{\Delta t}v, \mathcal{A}_{\Delta x}S_{\Delta t}v\rangle_{\ell^2} = \langle S_{\Delta t}v,\mathcal{G}_{\Delta x}v\rangle_{\ell^2} + \Delta t\,\langle S_{\Delta t}v,F\rangle_{\ell^2} \tag{5.40}$$

Using (2.31), Cauchy–Schwarz, estimate (2.25), estimate (2.26) plus formula (2.20) and the second part of Corollary 4.9 with $\eta = c\Delta t$ and $\delta = \frac{\Delta t}{\Delta x}$ if A is constant, or the first part of that corollary, still with $\eta = c\Delta t$ and using the fact that $\Delta x \le c\Delta t$ by (2.27) if A is not constant, and the elementary bound $ab \le \frac12 a^2 + \frac12 b^2$ yields

$$\begin{aligned}\langle S_{\Delta t}v,\mathcal{G}_{\Delta x}v\rangle_{\ell^2} &= \langle S_{\Delta t}v, G_{\Delta x}v\rangle_{\ell^2} + \Delta t\,\langle S_{\Delta t}v, H_{\Delta x}v\rangle_{\ell^2}\\ &\le \tfrac12\langle S_{\Delta t}v, AS_{\Delta t}v\rangle_{\ell^2} + \tfrac12\left\langle v, G^*_{\Delta x}(A)^{-1}G_{\Delta x}v\right\rangle_{\ell^2} + \tfrac{\Delta t}{2}\langle S_{\Delta t}v, B_{\Delta x}S_{\Delta t}v\rangle_{\ell^2} + \tfrac{\Delta t}{2}\langle v, B_{\Delta x}v\rangle_{\ell^2}\\ &\le \tfrac12\langle S_{\Delta t}v, AS_{\Delta t}v\rangle_{\ell^2} + \tfrac12(1+c\Delta t)\,\langle v, Av\rangle_{\ell^2} + \tfrac{\Delta t}{2}\langle S_{\Delta t}v, B_{\Delta x}S_{\Delta t}v\rangle_{\ell^2} + \tfrac{\Delta t}{2}\langle v, B_{\Delta x}v\rangle_{\ell^2}\\ &\le \tfrac12\langle S_{\Delta t}v, \mathcal{A}_{\Delta x}S_{\Delta t}v\rangle_{\ell^2} + \tfrac12(1+c\Delta t)\,\langle v,\mathcal{A}_{\Delta x}v\rangle_{\ell^2}\,.\end{aligned} \tag{5.41}$$

By (2.26) and Corollary 4.9, the constant c in (5.41) depends at most on $\|V\|_{h^{\sigma+2}}$. Since the assumptions (2.8) plus (2.22) imply that $\mathcal{A}_{\Delta x} \ge \frac{1}{m(\varepsilon\|V\|_{\ell^\infty})}I$,

$$\Delta t\,\langle S_{\Delta t}v,F\rangle_{\ell^2} \le \tfrac{\Delta t}{2}\left[\|S_{\Delta t}v\|^2_{\ell^2} + \|F\|^2_{\ell^2}\right] \le c\Delta t\,\langle S_{\Delta t}v,\mathcal{A}_{\Delta x}S_{\Delta t}v\rangle_{\ell^2} + \tfrac{\Delta t}{2}\|F\|^2_{\ell^2}. \tag{5.42}$$

In addition, by writing the difference as the integral of a derivative and using the Moser estimate (4.17) and the Sobolev embedding estimate (4.7) we obtain

$$\|A(\varepsilon S_{\Delta t}V) - A(\varepsilon S_{\Delta t}V)\|_{\ell^\infty} = \|\int_0^1 \frac{\mathrm{d}}{\mathrm{d}s}A(\varepsilon(sS_{\Delta t}V + (1-s)V)\,\mathrm{d}s\|_{\ell^\infty} \le c(\|V\|_{h^\sigma},\|S_{\Delta t}V\|_{h^\sigma})\|D_{\Delta t}V\|_{h^\sigma}, \tag{5.43}$$

which implies that

$$\|S_{\Delta t}v\|^2_{h^0_{\mathcal{A}(\varepsilon S_{\Delta t}V)}} \le (1+c\Delta t)\|S_{\Delta t}v\|^2_{h^0_{\mathcal{A}(\varepsilon V)}}. \tag{5.44}$$

Substituting (5.41)–(5.42) into (5.40), solving the result for $\langle S_{\Delta t}v, \mathcal{A}_{\Delta t}S_{\Delta t}v\rangle_{\ell^2}$, and substituting the result into the right side of (5.44) yields (5.39). □

Proof of Theorem 2.4. We first obtain the analogue of Lemma 5.2. Before applying a spatial difference operator $D^\alpha_{\Delta x}$ to equation (5.37), we rearrange that equation to show that the commutator terms that arise will have a similar form to those in Lemma 5.2. Assumption (2.21) implies that, after subtracting Av from both sides of (5.37) and multiplying both sides of the result by $(\Delta t)^{-1}$, the terms multiplied by A on the left side can be written as a time difference and the terms arising from $\mathcal{G}_{\Delta x}v - Av$ can be written as spatial differences. Note that although the term $\frac{A(\varepsilon v)[\sum C_\alpha S^\alpha_{\Delta x} - I]v}{\Delta t}$ is a sum of spatial differences divided by Δt rather than by Δx, it can be written as $\frac{\Delta x}{\Delta t}$ times $\frac{A(\varepsilon v)[\sum C_\alpha S^\alpha_{\Delta x} - I]v}{\Delta x}$, and by assumption (2.27) either $\frac{\Delta x}{\Delta t}$ is bounded or else A is constant and hence this term produces no commutator terms. Hence $D^\alpha_{\Delta x}$ applied to (5.37) can be written as that same equation with v replaced by $D^\alpha_{\Delta x}v$ plus Δt times commutators that can be estimated by a constant times spatial differences of order at most $|\alpha|$, which can be estimated in similar fashion to the estimates for the commutators of the θ-scheme. Estimating as in the proof of Lemma 5.2 but using the basic estimate of Lemma 5.3 instead of Lemma 5.1, and using Lemma 4.2 to bound the term involving $B_{\Delta x}$ in the analogue of (5.25), therefore yields the desired estimate for $\langle v, \mathcal{A}v\rangle_{h^s}$. As in the proof of Lemma 5.2, an estimate for the time difference is obtained similarly. Given those estimates, the rest of the proof is similar to that of Theorem 2.2. In particular, the same trick used in the proof of that theorem can be used to show that when $\|Q_{\Delta x}v_{0,\varepsilon,\Delta x}\|_{h^r_{\mathcal{A}(0)}} \le c\varepsilon$ then $\|D_{\Delta t}u\,|_{t=0}\|_{h^r_{\mathcal{A}(0)}}$ is uniformly bounded, since the purely-implicit form $\frac{1}{\varepsilon}Q_{\Delta x}S_{\Delta t}u$ of the large term in (2.32) corresponds to the case $\theta = 1$ of the θ-scheme. □

References

[1] G. Alì and L. Chen, The zero-electron-mass limit in the Euler-Poisson system for both well- and ill-prepared initial data. *Nonlinearity* **24** (2011) 2745–2761.

[2] F. Ancona and A. Marson, A locally quadratic Glimm functional and sharp convergence rate of the Glimm scheme for nonlinear hyperbolic systems. *Arch. Ration. Mech. Anal.* **196** (2010) 455–487.

[3] M. Bessemoulin-Chatard, C. Chainais-Hillairet and F. Filbet, On discrete functional inequalities for some finite volume schemes. *IMA J. Numer. Anal.* **35** (2015) 1125–1149.

[4] A. Bressan, Hyperbolic systems of conservation laws. The one-dimensional Cauchy problem. Vol. 20 of *Oxford Lect. Ser. Math. Appl.* Oxford University Press, Oxford (2000).

[5] L. Chen, D. Donatelli and P. Marcati, Incompressible type limit analysis of a hydrodynamic model for charge-carrier transport. *SIAM J. Math. Anal.* **45** (2013) 915–933.

[6] F. Coquel and P. LeFloch, Convergence de schémas aux différences finies pour des lois de conservation à plusieurs dimensions d'espace. *C. R. Acad. Sci. Paris Sér. I Math.* **310** (1990) 455–460.

[7] S. Cordier and E. Grenier, Quasineutral limit of an Euler-Poisson system arising from plasma physics. *Commun. Partial Differ. Eq.* **25** (2000) 1099–1113.

[8] S. Dellacherie, Analysis of Godunov type schemes applied to the compressible Euler system at low Mach number. *J. Comput. Phys.* **229** (2010) 978–1016.

[9] L. Even-Dar Mandel and S. Schochet, Convergence of solutions to finite difference schemes for singular limits of nonlinear evolutionary PDEs. To appear in *ESAIM: M2AN* (2016) `DOI: 10.1051/m2an/2016029`.

[10] U.S. Fjordholm, S. Mishra and E. Tadmor, Arbitrarily high-order accurate entropy stable essentially nonoscillatory schemes for systems of conservation laws. *SIAM J. Numer. Anal.* **50** (2012) 544–573.

[11] G.B. Folland, *Introduction to partial differential equations.* Princeton University Press, Princeton, N.J. (1976).

[12] G.B. Folland, *Fourier analysis and its applications.* Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA (1992).

[13] A. Friedman, *Partial differential equations*, original edn. Robert E. Krieger Publishing Co., Huntington, New York (1976).

[14] I. Gallagher, Applications of Schochet's methods to parabolic equations. *J. Math. Pures Appl.* **77** (1998) 989–1054.

[15] A. Glitzky and J.A. Griepentrog, Discrete Sobolev-Poincaré inequalities for Voronoi finite volume approximations. *SIAM J. Numer. Anal.* **48** (2010) 372–391.

[16] M. Goldberg, Stable difference schemes for parabolic systems – a numerical radius approach. *SIAM J. Numer. Anal.* **35** (1998) 478–493.

[17] E. Grenier, Pseudo-differential energy estimates of singular perturbations. *Commun. Pure Appl. Math.* **50** (1997) 821–865.

[18] B. Gustafsson, H.O. Kreiss and J. Oliger, Time dependent problems and difference methods. *Pure and Applied Mathematics.* John Wiley & Sons Inc., New York (1995).

[19] F. John, On integration of parabolic equations by difference methods. I. Linear and quasi-linear equations for the infinite interval. *Commun. Pure Appl. Math.* **5** (1952) 155–211.

[20] S. Klainerman and A. Majda, Singular limits of quasilinear hyperbolic systems with large parameters and the incompressible limit of compressible fluids. *Commun. Pure Appl. Math.* **34** (1981) 481–524.

[21] R. Klein, Semi-implicit extension of a Godunov-type scheme based on low Mach number asymptotics. I. One-dimensional flow. *J. Comput. Phys.* **121** (1995) 213–237.

[22] R. Klein, N. Botta, T. Schneider, C.D. Munz, S. Roller, A. Meister, L. Hoffmann and T. Sonar, Asymptotic adaptive methods for multi-scale problems in fluid mechanics. *J. Engrg. Math.* **39** (2001) 261–343.

[23] D. Kröner and M. Rokyta, Convergence of upwind finite volume schemes for scalar conservation laws in two dimensions. *SIAM J. Numer. Anal.* **31** (1994) 324–343.

[24] P.D. Lax and L. Nirenberg, On stability for difference schemes: A sharp form of Gårding's inequality. *Commun. Pure Appl. Math.* **19** (1966) 473–492.

[25] R.J. LeVeque, Finite volume methods for hyperbolic problems. *Cambridge Texts in Applied Mathematics.* Cambridge University Press, Cambridge (2002).

[26] A. Majda, Compressible fluid flow and systems of conservation laws in several space variables. Vol. 53 of *Appl. Math. Sci.* Springer-Verlag, New York (1984).

[27] D. Michelson, Stability theory of difference approximations for multidimensional initial-boundary value problems. *Math. Comput.* **40** (1983) 1–45.

[28] C.D. Munz, S. Roller, R. Klein and K.J. Geratz, The extension of incompressible flow solvers to the weakly compressible regime. *Comput. Fluids* **32** (2003) 173–196.

[29] S. Noelle, G. Bispen, K.R. Arun, M. Lukáčová-Medviďová and C.D. Munz, A weakly asymptotic preserving low Mach number scheme for the Euler equations of gas dynamics. *SIAM J. Sci. Comput.* **36** (2014) B989–B1024.

[30] S. Osher and E. Tadmor, On the convergence of difference approximations to scalar conservation laws. *Math. Comput.* **50** (1988) 19–51.

[31] A.A. Samarskii, The theory of difference schemes. Vol. 240 of *Monographs and Textbooks in Pure and Applied Mathematics.* Marcel Dekker Inc., New York (2001).

[32] S. Schochet, The incompressible limit in nonlinear elasticity. *Commun. Math. Phys.* **102** (1985) 207–215.

[33] S. Schochet, Fast singular limits of hyperbolic PDEs. *J. Differ. Eq.* **114** (1994) 476–512.

[34] S. Schochet, The mathematical theory of low Mach number flows. *ESAIM: M2AN* **39** (2005) 441–458.

[35] S. Schochet, Convergence of finite-volume schemes to smooth solutions of multidimensional hyperbolic systems. In preparation (2017).

[36] S. Schochet and M.I. Weinstein, The nonlinear Schrödinger limit of the Zakharov equations governing Langmuir turbulence. *Commun. Math. Phys.* **106** (1986) 569–580.

[37] J. Schütz and S. Noelle, Flux splitting for stiff equations: a notion on stability. *J. Sci. Comput.* **64** (2015) 522–540.

[38] E.L. Smirnova, The stability of parabolic difference schemes with variable coefficients. Russian original published (1975). *Vestnik Leningrad. Univ. Math.* (1980) 373–381.

[39] E.L. Smirnova, The invertibility and stability of difference operators with variable coefficients. Russian original published (1977). *Vestnik Leningrad. Univ. Math.* (1982) 209–215.

[40] G. Strang, Accurate partial difference methods. II. Non-linear problems. *Numer. Math.* **6** (1964) 37–46.

[41] E. Tadmor, Local error estimates for discontinuous solutions of nonlinear hyperbolic equations. *SIAM J. Numer. Anal.* **28** (1991) 891–906.

[42] M.E. Taylor, Pseudodifferential operators. Vol. 34 of *Princeton Mathematical Series.* Princeton University Press, Princeton, N.J. (1981).

[43] M.E. Taylor, Partial differential equations, II. Qualitative studies of linear equations. Vol. 116 of *Appl. Math. Sci.* Springer-Verlag, New York (1996).

[44] K. Tomoeda, Convergence of difference approximations for quasilinear hyperbolic systems. *Hiroshima Math. J.* **11** (1981) 465–491.

[45] R. Vaillancourt, A simple proof of Lax–Nirenberg theorems. *Commun. Pure Appl. Math.* **23** (1970) 151–163.

[46] D. Wang and C. Yu, Incompressible limit for the compressible flow of liquid crystals. *J. Math. Fluid Mech.* **16** (2014) 771–786.

[47] O.B. Widlund, On the stability of parabolic difference schemes. *Math. Comput.* **19** (1965) 1–13.

[48] Y.L. Zhou, *Applications of discrete functional analysis to the finite difference method.* International Academic Publishers, Beijing (1991).

[49] Y.L. Zhou, On the general interpolation formulas for discrete functional spaces. I. *J. Comput. Math.* **11** (1993) 188–192.

[50] Y.L. Zhou, General interpolation formulas for spaces of discrete functions with nonuniform meshes. *J. Comput. Math.* **13** (1995) 70–93.