

Introduction to discrete functional analysis techniques for the numerical study of diffusion equations with irregular data

J. Droniou¹

(Received 28 February 2015; revised 3 November 2015)

Abstract

We give an introduction to discrete functional analysis techniques for stationary and transient diffusion equations. We show how these techniques are used to establish the convergence of various numerical schemes without assuming non-physical regularity of the data. For simplicity of exposure, we mostly consider linear elliptic equations, and we briefly explain how these techniques are adapted and extended to non-linear time-dependent meaningful models (Navier–Stokes equations, flows in porous media, etc.). These convergence techniques rely on Sobolev norms and discrete forms of functional analysis results. The discrete functional analysis tools presented here are versatile and applicable to a number of numerical methods and models.

<http://journal.austms.org.au/ojs/index.php/ANZIAMJ/article/view/9365> gives this article, © Austral. Mathematical Soc. 2015. Published December 9, 2015, as part of the Proceedings of the 17th Biennial Computational Techniques and Applications Conference. ISSN 1446-8735. (Print two pages per sheet of paper.) Copies of this article must not be made otherwise available on the internet; instead link directly to this URL for this article.

Contents

1	Introduction	C102
2	Convergence by compactness techniques	C105
2.1	Model and preliminary considerations	C105
2.2	General path for the convergence analysis	C106
2.3	Examples	C109
2.3.1	Two-point flux approximation finite volume scheme	C109
2.3.2	Non-conforming \mathbb{P}^1 finite element	C112
3	Extension to non-linear models	C115
3.1	Stationary equations	C116
3.1.1	Academic example	C116
3.1.2	Physical models	C117
3.2	Time-dependent and Navier–Stokes equations	C119
4	Conclusions and perspectives	C121
	References	C122

1 Introduction

A number of real-world problems are modelled by partial differential equations (PDEs) which involve some form of singularity. For example, oil engineers deal with underground reservoirs made of stacked geological layers with different rock properties, which are described by discontinuous data (permeability tensor, porosity, etc.) in the corresponding mathematical model. Another example from reservoir engineering is the modelling of wellbores; the relative scales of the wellbores ($\sim 10\text{--}20$ cm in diameter) and the reservoir ($\sim 1\text{--}2$ km wide) justify representing injection and production terms at the wells by Radon measures [34]. The mathematical analysis of PDEs involving

singular data is challenging. The meanings of the terms in the equations have to be re-thought; classical derivatives can no longer be used, and weak/distribution derivatives and Sobolev spaces must be introduced [4]. Beyond these now well-known tools, other techniques had to be developed for the most complex models to define appropriate notions of solutions, and to prove their existence (and uniqueness, if possible): renormalised solutions [10], entropy solutions [3], monotone operators and semi-groups [4], elliptic and parabolic capacity [10, 25], etc. The main purpose of this analysis is to ensure that the models are well-posed; that is, that they make sense from a mathematical perspective. It is rarely possible to give explicit forms for, or even detailed qualitative behaviour of, the solutions to the extremely complex models involved in field applications. Precise quantitative information that can be used for decision-making is obtained only through numerical approximation.

The role of mathematics in obtaining accurate approximate solutions to PDEs is twofold. First, algorithms have to be designed to compute these solutions. But, even based on sound reasoning, in some circumstances algorithms can fail to approximate the expected model [35, Chap. III, Sec. 3]. Benchmarking (testing the algorithms in well-documented cases) is useful to ensure the quality of numerical methods, but it cannot cover all situations that may occur in field applications. The second role of mathematics in the numerical approximation of real-world models is to provide rigorous analysis of the properties and convergence of the schemes; this analysis is not restricted to particular cases and is essential to ensure the reliability of numerical methods for PDEs.

The usual way to prove the convergence of a scheme is to establish error estimates; if $\bar{\mathbf{u}}$ is the solution to the PDE and \mathbf{u}_h is the solution provided by the scheme (where h is, for example, the mesh size), then one will try to establish a bound of the kind

$$\|\mathbf{u}_h - \bar{\mathbf{u}}\|_X \leq \mathcal{C}h^\alpha, \quad (1)$$

where $\|\cdot\|_X$ is an adequate norm, $\alpha > 0$, and \mathcal{C} is a positive constant. Such an inequality provides an estimate for the h that must be selected in order

to achieve a pre-determined accuracy of the approximation. However, major limitations exist:

- Estimates of the kind (1) can be established only if the uniqueness of the solution $\bar{\mathbf{u}}$ to the PDE is known (if (1) holds, then $\bar{\mathbf{u}}$ is unique; often the proof of (1) mimics a proof of uniqueness of $\bar{\mathbf{u}}$).
- The constant \mathcal{C} usually depends on higher derivatives of $\bar{\mathbf{u}}$ or the PDE data, and (1) therefore requires some regularity assumptions on the solution or data.

For many non-linear real-world models, including those from reservoir engineering [36] and the famous Navier–Stokes equations, uniqueness of the solution is not known unless strong regularity properties on the solution are assumed. These properties cannot be established in field applications. Hence convergence analysis based on error estimates is doomed to be somewhat disconnected from applications. This article presents an introduction to techniques that were recently developed to deal with this issue. These techniques enable the convergence analysis of numerical schemes under assumptions that are compatible with real-world data and constraints.

Section 2 details the convergence technique on a simple linear stationary diffusion equation. After recalling some basic energy estimates on the model, we present the general path (in Section 2.2) to establish the convergence of schemes without any regularity assumptions on the data; this path relies on compactness techniques and *discrete functional analysis* tools, translations to the discrete setting of functional analysis results for functions of continuous variables. Section 2.3 shows, using the two-point finite volume scheme and the non-conforming \mathbb{P}^1 finite element scheme, how these tools are constructed in practice. In Section 3 we discuss the extension of this convergence technique to non-linear and non-stationary models, which are more realistic representations of physical phenomena. We briefly show that virtually no adaptation is required from the technique used in the linear setting to deal with the simplest non-linear models. We then give a brief overview of physical models whose numerical analysis was successfully tackled using discrete functional

analysis tools. These include the Navier–Stokes equations, PDEs involved in glaciology, models of oil recovery, and models of melting materials.

2 Convergence by compactness techniques

2.1 Model and preliminary considerations

For our initial presentation consider the linear diffusion equation

$$\begin{cases} -\operatorname{div}(\mathbf{A}\nabla\bar{\mathbf{u}}) = \mathbf{f} & \text{in } \Omega, \\ \bar{\mathbf{u}} = 0 & \text{on } \partial\Omega. \end{cases} \quad (2)$$

In the context of reservoir engineering, (2) corresponds to a steady single-phase single-component Darcy problem with no gravitational effects [11]; $\bar{\mathbf{u}}$ is the pressure and \mathbf{A} is the matrix-valued permeability field. This field is usually considered piecewise constant (constant in each geological layer), and it is therefore discontinuous. Equation (2) cannot be considered under the classical sense, that is, with div and ∇ denoting standard derivatives, and must be re-written in a weak form. This weak form is obtained by multiplying by a test function \mathbf{v} which vanishes on $\partial\Omega$ and by using the Stokes formula [4]:

$$\begin{cases} \text{Find } \bar{\mathbf{u}} \in \mathbf{H}_0^1(\Omega) \text{ such that for all } \mathbf{v} \in \mathbf{H}_0^1(\Omega), \\ \int_{\Omega} \mathbf{A}(\mathbf{x})\nabla\bar{\mathbf{u}}(\mathbf{x}) \cdot \nabla\mathbf{v}(\mathbf{x}) \, \mathrm{d}\mathbf{x} = \int_{\Omega} \mathbf{f}(\mathbf{x})\mathbf{v}(\mathbf{x}) \, \mathrm{d}\mathbf{x}. \end{cases} \quad (3)$$

Here, $\mathbf{H}_0^1(\Omega)$ is the Sobolev space of functions $\mathbf{v} \in \mathbf{L}^2(\Omega)$ (square-integrable functions equipped with the norm $\|\mathbf{v}\|_{\mathbf{L}^2(\Omega)}^2 = \int_{\Omega} |\mathbf{v}(\mathbf{x})|^2 \, \mathrm{d}\mathbf{x}$) that have a weak (distribution) gradient $\nabla\mathbf{v}$ in $\mathbf{L}^2(\Omega)^d$ and a zero value (trace) on $\partial\Omega$. Under the following assumptions, all terms in (3) are well-defined:

1. Ω is a bounded open set of \mathbb{R}^d ($d \geq 1$) and $\mathbf{f} \in \mathbf{L}^2(\Omega)$;
2. $\mathbf{A} : \Omega \mapsto \mathcal{M}_d(\mathbb{R})$ is a measurable matrix-valued mapping;

3. There exists $0 < \underline{\alpha} \leq \bar{\alpha} < \infty$ such that $|\mathbf{A}(\mathbf{x})\xi| \leq \bar{\alpha}|\xi|$ and $\mathbf{A}(\mathbf{x})\xi \cdot \xi \geq \underline{\alpha}|\xi|^2$ for almost every $\mathbf{x} \in \Omega$ and all $\xi \in \mathbb{R}^d$.

Here, $|\cdot|$ is the Euclidean norm on \mathbb{R}^d . Take $\mathbf{v} = \bar{\mathbf{u}}$ in (3) and apply the Cauchy–Schwarz inequality on the right-hand side to find

$$\begin{aligned} \underline{\alpha} \|\nabla \bar{\mathbf{u}}\|_{L^2(\Omega)}^2 &\leq \int_{\Omega} \mathbf{A}(\mathbf{x}) \nabla \bar{\mathbf{u}}(\mathbf{x}) \cdot \nabla \bar{\mathbf{u}}(\mathbf{x}) \, d\mathbf{x} = \int_{\Omega} f(\mathbf{x}) \bar{\mathbf{u}}(\mathbf{x}) \, d\mathbf{x} \\ &\leq \|f\|_{L^2(\Omega)} \|\bar{\mathbf{u}}\|_{L^2(\Omega)}, \end{aligned} \quad (4)$$

where $\|\cdot\|$ is the $L^2(\Omega)$ norm of the Euclidean norm. Essential to the analysis of elliptic equations is the Poincaré inequality:

$$\text{for all } \mathbf{v} \in H_0^1(\Omega), \quad \|\mathbf{v}\|_{L^2(\Omega)} \leq \text{diam}(\Omega) \|\nabla \mathbf{v}\|_{L^2(\Omega)}. \quad (5)$$

Substituted into (4), this inequality leads to the following energy estimate, in which the left-hand side defines the norm in $H_0^1(\Omega)$:

$$\|\bar{\mathbf{u}}\|_{H_0^1(\Omega)} := \|\nabla \bar{\mathbf{u}}\|_{L^2(\Omega)} \leq \text{diam}(\Omega) \underline{\alpha}^{-1} \|f\|_{L^2(\Omega)}. \quad (6)$$

2.2 General path for the convergence analysis

Estimate (6) shows that $H_0^1(\Omega)$ is the natural energy space of problem (2). This estimate is at the core of the theoretical study of (2) and its non-linear variants, partly due to the Rellich compactness theorem [4].

Theorem 1 (Rellich compact embedding). *If Ω is a bounded subset of \mathbb{R}^d , $d \geq 1$, and if $(\mathbf{v}_n)_{n \in \mathbb{N}}$ is bounded in $H_0^1(\Omega)$, then $(\mathbf{v}_n)_{n \in \mathbb{N}}$ has a subsequence that converges in $L^2(\Omega)$. Furthermore, any limit in $L^2(\Omega)$ of a subsequence of $(\mathbf{v}_n)_{n \in \mathbb{N}}$ belongs to $H_0^1(\Omega)$.*

This theorem justifies the general path for a convergence analysis that is applicable without smoothness assumption on the data or the solution, and that can be adapted to non-linear equations. As described by Droniou [14], this path to prove convergence comprises three steps.

1. Establish a priori energy estimates similar to (6) on the solutions to the scheme in a mesh- and scheme-dependent discrete norm that mimics the H_0^1 norm.
2. Prove a compactness result which is the discrete equivalent of Theorem 1: if $(\mathbf{u}_h)_h$ is a sequence of discrete functions that are bounded in the norms introduced in step 1, then as the mesh size h goes to zero there is a subsequence of $(\mathbf{u}_h)_h$ that converges (at least in $L^2(\Omega)$) to a function $\bar{\mathbf{u}} \in H_0^1(\Omega)$.
3. Prove that if $\bar{\mathbf{u}} \in H_0^1(\Omega)$ is the limit in $L^2(\Omega)$ as $h \rightarrow 0$ of solutions to the scheme, then $\bar{\mathbf{u}}$ satisfies (3).

Remark 2. The existence of a solution to the PDE does not need to be known. It is obtained as a consequence of the convergence proof.

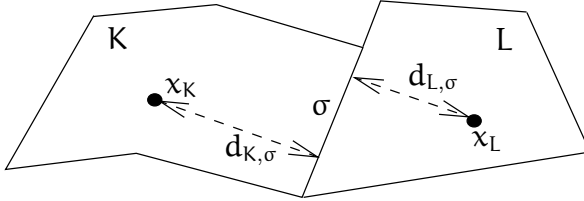
The discrete $H_0^1(\Omega)$ norm is dictated by the scheme. It must be a norm for which (i) a priori estimates on the numerical solutions can be obtained, and (ii) the compactness result in step 2 holds. However, there is a norm applicable to a number of numerical methods. Let us assume that Ω is polytopal (polygonal in 2D, polyhedral in 3D, etc.), and that \mathcal{M} is a mesh of Ω made of polytopal cells. We denote the size of \mathcal{M} by $h_{\mathcal{M}} = \max_{K \in \mathcal{M}} \text{diam}(K)$ and the space of piecewise constant functions in the cells by $X_{\mathcal{M}}$. We identify $\mathbf{v} \in X_{\mathcal{M}}$ with the family of its values $(v_K)_{K \in \mathcal{M}}$ in the cells. The set of all faces of the mesh (edges in 2D) is $\mathcal{E}_{\mathcal{M}}$ and $|\sigma|$ denotes the $(d-1)$ -dimensional measure of a face σ (i.e., length in 2D, area in 3D). We take one point x_K in each cell K , and we let $d_{K,\sigma} = \text{dist}(x_K, \sigma)$ (see Figure 1). If σ is an interface between two cells K and L , then we define $d_{\sigma} = d_{K,\sigma} + d_{L,\sigma}$; otherwise, $d_{\sigma} = d_{K,\sigma}$ with K the unique cell whose σ is an face.

A discrete H_0^1 norm on $X_{\mathcal{M}}$ is defined by

$$\|\mathbf{v}\|_{H_0^1, \mathcal{M}}^2 := \sum_{\sigma \in \mathcal{E}_{\mathcal{M}}} |\sigma| d_{\sigma} \left(\frac{v_K - v_L}{d_{\sigma}} \right)^2. \quad (7)$$

Here, and in subsequent similar sums, we use the convention that K and L

Figure 1: Notations associated with a polytopal mesh.



are the cells on each side of σ , and that $\mathbf{v}_L = 0$ if $\sigma \subset \partial\Omega$ is a face of K . This choice accounts for the homogeneous boundary conditions on $\partial\Omega$.

The usefulness of the discrete H_0^1 norm, in view of the convergence path 1–3, is apparent in the two following theorems, proved by Eymard et al. [30]. Theorem 3 is the key to reproducing, at the discrete level, the sequence of inequalities (4)–(6), leading to the energy estimates in convergence step 1; this requires the scheme to have suitable *coercivity* properties. Theorem 4 covers step 2. Convergence step 3 is more scheme-dependent and relies on *consistency* and *limit-conformity* properties of the scheme. Theorems 3 and 4 are examples of discrete functional analysis results.

Theorem 3 (Discrete Poincaré inequality). *Let \mathcal{M} be a mesh of Ω and set*

$$\theta_{\mathcal{M}} = \max \left[\frac{d_{K,\sigma}}{d_{L,\sigma}} : \sigma \in \mathcal{E}_{\mathcal{M}} \text{ and } K, L \text{ are cells on each side of } \sigma \right]. \quad (8)$$

If $\bar{\theta} \geq \theta_{\mathcal{M}}$, then there exists C_1 only depending on $\bar{\theta}$ such that for any $\mathbf{v} \in X_{\mathcal{M}}$ we have $\|\mathbf{v}\|_{L^2(\Omega)} \leq C_1 \|\mathbf{v}\|_{H_0^1, \mathcal{M}}$.

Theorem 4 (Discrete Rellich theorem). *Let $(\mathcal{M}_n)_{n \in \mathbb{N}}$ be a sequence of discretisations of Ω such that $(\theta_{\mathcal{M}_n})_{n \in \mathbb{N}}$ is bounded and $h_{\mathcal{M}_n} \rightarrow 0$ as $n \rightarrow \infty$. If $\mathbf{v}_n \in X_{\mathcal{M}_n}$ is such that $(\|\mathbf{v}_n\|_{H_0^1, \mathcal{M}_n})_{n \in \mathbb{N}}$ is bounded, then $(\mathbf{v}_n)_{n \in \mathbb{N}}$ is relatively compact in $L^2(\Omega)$. Furthermore, any limit in $L^2(\Omega)$ of a subsequence of $(\mathbf{v}_n)_{n \in \mathbb{N}}$ belongs to $H_0^1(\Omega)$.*

2.3 Examples

Besides Theorems 3 and 4, an important feature of the discrete norm (7) is its versatility; it is suitable for numerous schemes, even with degrees of freedom that are not cell-centred. Here we give two practical illustrations of the usage of the convergence path 1–3 and of the discrete norm (7).

2.3.1 Two-point flux approximation finite volume scheme

The two-point flux approximation (TPFA) scheme for (2) is given by flux balances (obtained by integrating (2) over the cells), and a finite difference approximation of the flux $-\int_{\sigma} \mathbf{A}(\mathbf{x}) \nabla \bar{\mathbf{u}}(\mathbf{x}) \cdot \mathbf{n}_{\mathbf{K}}(\mathbf{x}) \, d\mathbf{x}$ using the two unknowns on each side of σ :

$$\text{for all } \mathbf{K} \in \mathcal{M}, \quad \sum_{\sigma \in \mathcal{E}_{\mathbf{K}}} F_{\mathbf{K},\sigma} = \int_{\mathbf{K}} f(\mathbf{x}) \, d\mathbf{x}; \quad (9)$$

$$\text{for all } \mathbf{K} \in \mathcal{M} \text{ and for all } \sigma \in \mathcal{E}_{\mathbf{K}}, \quad F_{\mathbf{K},\sigma} = \tau_{\sigma}(\mathbf{u}_{\mathbf{K}} - \mathbf{u}_{\mathbf{L}}). \quad (10)$$

Here, $\mathcal{E}_{\mathbf{K}}$ is the set of faces of a cell $\mathbf{K} \in \mathcal{M}$, and the transmissivity $\tau_{\sigma} \in (0, \infty)$ depends on \mathbf{A} and the local mesh geometry [28]. Under usual non-degeneracy assumptions on the mesh, there exists $C_2 > 0$ only depending on $\bar{\mathbf{a}}$ and $\underline{\mathbf{a}}$ such that

$$\tau_{\sigma} \geq C_2 \frac{|\sigma|}{d_{\sigma}}. \quad (11)$$

Convergence step 1 The inequalities (4)–(6) that lead to the a priori estimates on $\bar{\mathbf{u}}$ are obtained by the following sequence of manipulations: (i) multiply (2) by $\mathbf{v} = \bar{\mathbf{u}}$ and integrate the resulting equation; (ii) apply the Stokes formula; and (iii) use the Poincaré inequality. Since the flux balance (9) is the discrete expression of (2), we reproduce these manipulations at the discrete level.

- (i). *Multiply and integrate:* we multiply (9) by $\mathbf{v}_K = \mathbf{u}_K$ and we sum on $K \in \mathcal{M}$. Accounting for (10), this gives

$$\sum_{K \in \mathcal{M}} \sum_{\sigma \in \mathcal{E}_K} \tau_\sigma(\mathbf{u}_K - \mathbf{u}_L) \mathbf{u}_K = \sum_K \left[\int_K f(x) \, dx \right] \mathbf{u}_K = \int_\Omega f(x) \mathbf{u}(x) \, dx. \quad (12)$$

- (ii). *Apply the Stokes formula:* this consists of gathering by faces the sum in the left-hand side of (13). The contributions of a face are $\tau_\sigma(\mathbf{u}_K - \mathbf{u}_L) \mathbf{u}_K$ and $\tau_\sigma(\mathbf{u}_L - \mathbf{u}_K) \mathbf{u}_L = -\tau_\sigma(\mathbf{u}_K - \mathbf{u}_L) \mathbf{u}_L$. Hence, using (11) and the Cauchy–Schwarz inequality on the right-hand side, we find

$$C_2 \sum_{\sigma \in \mathcal{E}_\mathcal{M}} \frac{|\sigma|}{d_\sigma} (\mathbf{u}_K - \mathbf{u}_L)^2 \leq \sum_{\sigma \in \mathcal{E}_\mathcal{M}} \tau_\sigma (\mathbf{u}_K - \mathbf{u}_L)^2 \leq \|f\|_{L^2(\Omega)} \|\mathbf{u}\|_{L^2(\Omega)}. \quad (13)$$

- (iii). *Use the Poincaré inequality:* the left-hand side of (13) is $C_2 \|\mathbf{u}\|_{H_0^1, \mathcal{M}}^2$. Invoking the discrete Poincaré inequality (Theorem 3), we find C_3 only depending on an upper bound of $\theta_\mathcal{M}$ such that

$$\|\mathbf{u}\|_{H_0^1, \mathcal{M}} \leq C_3 \|f\|_{L^2(\Omega)}. \quad (14)$$

Estimate (14) is the discrete equivalent of (6) for the solution of the TPFA scheme.

Convergence step 2 This step is straightforward from (14) after applying the discrete Rellich Theorem 4. This estimate shows that if $(\mathcal{M}_n)_{n \in \mathbb{N}}$ is a sequence of meshes and if \mathbf{u}_n is the solution of the TPFA scheme on \mathcal{M}_n , then $(\|\mathbf{u}_n\|_{H_0^1, \mathcal{M}_n})_{n \in \mathbb{N}}$ remains bounded. Hence, up to a subsequence, \mathbf{u}_n converges in $L^2(\Omega)$ towards some function $\bar{\mathbf{u}} \in H_0^1(\Omega)$.

Convergence step 3 As mentioned above, proving that $\bar{\mathbf{u}}$ is the solution to (3) hinges on adequate consistency properties enjoyed by the scheme. Here,

it all comes down to the proper choice of transmissivities τ_σ , and to the geometry of the mesh. By taking $\varphi \in C_c^\infty(\Omega)$, multiplying (9) by $\varphi(\mathbf{x}_K)$ when $\mathcal{M} = \mathcal{M}_n$, and summing over all K we find

$$\sum_{K \in \mathcal{M}_n} \sum_{\sigma \in \mathcal{E}_K} \tau_\sigma [(\mathbf{u}_n)_K - (\mathbf{u}_n)_L] \varphi(\mathbf{x}_K) = \sum_{K \in \mathcal{M}_n} \int_K f(\mathbf{x}) \varphi(\mathbf{x}_K) \, d\mathbf{x}.$$

We then gather the sums in the left-hand side by terms involving $(\mathbf{u}_n)_K$:

$$\sum_{K \in \mathcal{M}_n} (\mathbf{u}_n)_K \sum_{\sigma \in \mathcal{E}_K} \tau_\sigma [\varphi(\mathbf{x}_K) - \varphi(\mathbf{x}_L)] = \sum_{K \in \mathcal{M}_n} \int_K f(\mathbf{x}) \varphi(\mathbf{x}_K) \, d\mathbf{x}, \quad (15)$$

where $\varphi(\mathbf{x}_L) = 0$ if $\sigma \in \mathcal{E}_K$ lies on $\partial\Omega$. The choice of τ_σ , the geometrical assumptions constraining the meshes of the TPFA method (that is, an orthogonality requirement of $(\mathbf{x}_K \mathbf{x}_L)$ and σ for a scalar product induced by A^{-1}), and the smoothness of φ , ensure that $\sum_{\sigma \in \mathcal{E}_K} \tau_\sigma [\varphi(\mathbf{x}_K) - \varphi(\mathbf{x}_L)] = -\int_K \operatorname{div}(A \nabla \varphi) + |K| \mathcal{O}(\mathbf{h}_{\mathcal{M}_n})$, where $|K|$ is the d -dimensional measure of K . Relation (15) thus gives

$$-\int_\Omega \mathbf{u}_n(\mathbf{x}) \operatorname{div}(A \nabla \varphi)(\mathbf{x}) \, d\mathbf{x} + \mathcal{O}(\|\mathbf{u}_n\|_{L^1(\Omega)} \mathbf{h}_{\mathcal{M}_n}) = \int_\Omega f(\mathbf{x}) \varphi(\mathbf{x}) \, d\mathbf{x} + \mathcal{O}(\mathbf{h}_{\mathcal{M}_n}),$$

where we use the smoothness of φ on the right-hand side. By the convergence of \mathbf{u}_n to $\bar{\mathbf{u}}$ in $L^2(\Omega)$, in the limit $n \rightarrow \infty$ we find that, for all $\varphi \in C_c^\infty(\Omega)$, $\bar{\mathbf{u}} \in H_0^1(\Omega)$ satisfies

$$-\int_\Omega \bar{\mathbf{u}}(\mathbf{x}) \operatorname{div}(A \nabla \varphi)(\mathbf{x}) \, d\mathbf{x} = \int_\Omega f(\mathbf{x}) \varphi(\mathbf{x}) \, d\mathbf{x},$$

which is classically equivalent to (3).

Remark 5. The above reasoning apparently only shows the convergence of a *subsequence* of $(\mathbf{u}_n)_{n \in \mathbb{N}}$. However, since there is only one possible limit (namely, the unique solution $\bar{\mathbf{u}}$ to (3)), this proves that the whole sequence $(\mathbf{u}_n)_{n \in \mathbb{N}}$ converges to $\bar{\mathbf{u}}$.

2.3.2 Non-conforming \mathbb{P}^1 finite element

Usage of the discrete norm (7) is not limited to numerical methods with only/primarily cell unknowns. Let us consider a triangulation \mathcal{T} of 2D polygonal domain Ω (what follows also generalises to tetrahedral meshes of a 3D polyhedral domain). The non-conforming Crouzeix–Raviart \mathbb{P}^1 finite element [9] for (2) has degrees of freedom at the midpoints $(\bar{x}_\sigma)_{\sigma \in \mathcal{E}_\mathcal{T}}$ of the triangulation's edges. The discrete space $Y_\mathcal{T}$ of unknowns is made of families of reals $\mathbf{u} = (\mathbf{u}_\sigma)_{\sigma \in \mathcal{E}_\mathcal{T}}$, where $\mathbf{u}_\sigma = 0$ if $\sigma \subset \partial\Omega$. These families are identified with functions $\mathbf{u} : \Omega \rightarrow \mathbb{R}$ that are piecewise linear on the mesh, with values $(\mathbf{u}_\sigma)_{\sigma \in \mathcal{E}_\mathcal{T}}$ at $(\bar{x}_\sigma)_{\sigma \in \mathcal{E}_\mathcal{T}}$. The non-conforming \mathbb{P}^1 approximation of (3) is

$$\begin{cases} \text{Find } \mathbf{u} \in Y_\mathcal{T} \text{ such that for all } \mathbf{v} \in Y_\mathcal{T}, \\ \int_\Omega \mathbf{A}(\mathbf{x}) \nabla_b \mathbf{u}(\mathbf{x}) \cdot \nabla_b \mathbf{v}(\mathbf{x}) \, d\mathbf{x} = \int_\Omega f(\mathbf{x}) \mathbf{v}(\mathbf{x}) \, d\mathbf{x}, \end{cases} \quad (16)$$

where ∇_b is the broken gradient and $(\nabla_b \mathbf{u})|_K$ is the constant gradient of the linear function \mathbf{u} in the triangle $K \in \mathcal{T}$.

Convergence step 1 To benefit from Theorems 3 and 4, we need to introduce the norm (7), which requires some choice of cell unknowns. Here, the most natural choice is to set \mathbf{u}_K equal to \mathbf{u} at the centre of gravity \bar{x}_K of K . Since \mathbf{u} is linear in K this gives

$$\text{for all } K \in \mathcal{T}, \quad \mathbf{u}_K = \mathbf{u}(\bar{x}_K) = \frac{1}{3} \sum_{\sigma \in \mathcal{E}_K} \mathbf{u}_\sigma.$$

This choice associates (in a non-injective way) to each $\mathbf{u} \in Y_\mathcal{T}$ a $\tilde{\mathbf{u}} = (\mathbf{u}_K)_{K \in \mathcal{T}} \in X_\mathcal{T}$. Two simple inequalities, both based on the linearity of \mathbf{u} inside each triangle, are useful to conclude step 1.

Lemma 6. *Let $\eta_\mathcal{T}$ be the maximum over $K \in \mathcal{T}$ of the ratio of the exterior diameter of K over the interior diameter of K . Assume that $\bar{\eta} \geq \eta_\mathcal{T}$. Then*


there exists C_4 only depending on $\bar{\eta}$ such that, for all $\mathbf{u} \in Y_{\mathcal{T}}$,

$$\|\tilde{\mathbf{u}}\|_{H^1_{0,\mathcal{T}}} \leq C_4 \|\nabla_{\mathbf{b}} \mathbf{u}\|_{L^2(\Omega)}, \quad (17)$$

$$\|\tilde{\mathbf{u}} - \mathbf{u}\|_{L^2(\Omega)} \leq h_{\mathcal{T}} \|\nabla_{\mathbf{b}} \mathbf{u}\|_{L^2(\Omega)}. \quad (18)$$

Proof: Start with (17). There exists C_5 only depending on $\bar{\eta}$ such that for all $\sigma \in \mathcal{K}$ we have $\text{dist}(\bar{x}_{\mathbf{K}}, \bar{x}_{\sigma}) \leq C_5 \mathbf{d}_{\sigma}$. Hence, since \mathbf{u} is linear inside each triangle,

$$\begin{aligned} \frac{|\tilde{\mathbf{u}}_{\mathbf{K}} - \tilde{\mathbf{u}}_{\mathbf{L}}|}{\mathbf{d}_{\sigma}} &\leq C_5 \frac{|\mathbf{u}(\bar{x}_{\mathbf{K}}) - \mathbf{u}(\bar{x}_{\sigma})|}{\text{dist}(\bar{x}_{\mathbf{K}}, \bar{x}_{\sigma})} + C_5 \frac{|\mathbf{u}(\bar{x}_{\mathbf{L}}) - \mathbf{u}(\bar{x}_{\sigma})|}{\text{dist}(\bar{x}_{\mathbf{L}}, \bar{x}_{\sigma})} \\ &\leq C_5 |(\nabla_{\mathbf{b}} \mathbf{u})_{\mathbf{K}}| + C_5 |(\nabla_{\mathbf{b}} \mathbf{u})_{\mathbf{L}}|. \end{aligned} \quad (19)$$

By squaring (19), multiplying by $|\sigma| \mathbf{d}_{\sigma}$, summing over the edges and using $\sum_{\sigma \in \mathcal{E}_{\mathbf{K}}} |\sigma| \mathbf{d}_{\sigma} \leq C_6 |\mathbf{K}|$ with C_6 only depending on $\bar{\eta}$, we obtain (17). The proof of (18) is simpler and follows directly from $\tilde{\mathbf{u}}(\mathbf{x}) - \mathbf{u}(\mathbf{x}) = \mathbf{u}(\bar{x}_{\mathbf{K}}) - \mathbf{u}(\mathbf{x}) = (\nabla_{\mathbf{b}} \mathbf{u})_{\mathbf{K}} \cdot (\bar{x}_{\mathbf{K}} - \mathbf{x})$ for all $\mathbf{x} \in \mathbf{K}$. 

Equipped with (17) and (18), we now delve into convergence step 1. Substituting $\mathbf{v} = \mathbf{u}$ in the formulation (16) of the scheme, the coercivity of \mathbf{A} entails

$$\underline{\mathbf{a}} \|\nabla_{\mathbf{b}} \mathbf{u}\|_{L^2(\Omega)}^2 \leq \|\mathbf{f}\|_{L^2(\Omega)} \|\mathbf{u}\|_{L^2(\Omega)}.$$

Using (18) and $h_{\mathcal{T}} \leq \text{diam}(\Omega)$, this gives

$$\underline{\mathbf{a}} \|\nabla_{\mathbf{b}} \mathbf{u}\|_{L^2(\Omega)}^2 \leq \|\mathbf{f}\|_{L^2(\Omega)} \left[\|\tilde{\mathbf{u}}\|_{L^2(\Omega)} + \text{diam}(\Omega) \|\nabla_{\mathbf{b}} \mathbf{u}\|_{L^2(\Omega)} \right].$$

A bound on $\eta_{\mathcal{T}}$ implies a bound on $\theta_{\mathcal{T}}$ (defined by (8)). Hence, the discrete Poincaré inequality (Theorem 3) and (17) lead to

$$\|\nabla_{\mathbf{b}} \mathbf{u}\|_{L^2(\Omega)} \leq [C_1 C_4 + \text{diam}(\Omega)] \underline{\mathbf{a}}^{-1} \|\mathbf{f}\|_{L^2(\Omega)}. \quad (20)$$

Estimate (20) is the discrete equivalent of the energy estimate (6). In conjunction with (17) it gives

$$\|\tilde{\mathbf{u}}\|_{H^1_{0,\mathcal{T}}} \leq C_4 [C_1 C_4 + \text{diam}(\Omega)] \underline{\mathbf{a}}^{-1} \|\mathbf{f}\|_{L^2(\Omega)}. \quad (21)$$

Convergence step 2 This is similar to the same step in the TPGA method. If $(\mathcal{T}_n)_{n \in \mathbb{N}}$ is a sequence of uniformly regular triangulations whose size tends to zero, then combining (21) (with $\mathcal{T} = \mathcal{T}_n$) and the discrete Rellich theorem (Theorem 4) shows that $\tilde{\mathbf{u}}_n \rightarrow \bar{\mathbf{u}}$ in $L^2(\Omega)$ up to a subsequence, for some $\bar{\mathbf{u}} \in H_0^1(\Omega)$. Moreover, by (18) and (20), we also have $\mathbf{u}_n \rightarrow \bar{\mathbf{u}}$ in $L^2(\Omega)$.

Convergence step 3 Assume now that

$$\nabla_b \mathbf{u}_n \rightarrow \nabla \bar{\mathbf{u}} \text{ weakly in } L^2(\Omega)^d \text{ as } n \rightarrow \infty. \quad (22)$$

For $\varphi \in C_c^\infty(\Omega)$ we define the interpolant $\mathbf{v}_n \in Y_{\mathcal{T}_n}$ by $(\mathbf{v}_n)_\sigma = \varphi(\bar{\mathbf{x}}_\sigma)$. The smoothness of φ ensures that $\mathbf{v}_n \rightarrow \varphi$ in $L^\infty(\Omega)$ and $\nabla_b \mathbf{v}_n \rightarrow \nabla \varphi$ in $L^\infty(\Omega)^d$. The convergence (22) therefore allows us to pass to the limit in (16) written for \mathbf{u}_n and \mathbf{v}_n . We deduce that $\bar{\mathbf{u}} \in H_0^1(\Omega)$ satisfies $\int_\Omega \mathbf{A} \nabla \bar{\mathbf{u}} \cdot \nabla \varphi \, dx = \int_\Omega f \varphi \, dx$ for all smooth φ , which is equivalent to (3).

The proof of (22) relies on well-established techniques. By (20) the sequence $(\nabla_b \mathbf{u}_n)_{n \in \mathbb{N}}$ is bounded, and therefore converges weakly in $L^2(\Omega)^d$ to some $\boldsymbol{\chi}$, up to a subsequence. We just need to prove that $\boldsymbol{\chi} = \nabla \bar{\mathbf{u}}$. Take $\boldsymbol{\psi} \in C_c^\infty(\Omega)^d$ and, by the Stokes formula in each triangle,

$$\begin{aligned} \int_\Omega \nabla_b \mathbf{u}_n(\mathbf{x}) \cdot \boldsymbol{\psi}(\mathbf{x}) \, dx &= \sum_{K \in \mathcal{T}_n} \int_K \nabla_b \mathbf{u}_n(\mathbf{x}) \cdot \boldsymbol{\psi}(\mathbf{x}) \, dx \\ &= \sum_{K \in \mathcal{T}_n} \int_{\partial K} (\mathbf{u}_n)|_K(\mathbf{x}) \mathbf{n}_K \cdot \boldsymbol{\psi}(\mathbf{x}) \, dS(\mathbf{x}) - \sum_{K \in \mathcal{T}_n} \int_K \mathbf{u}_n(\mathbf{x}) \operatorname{div} \boldsymbol{\psi}(\mathbf{x}) \, dx \\ &= Z_n - \int_\Omega \mathbf{u}_n(\mathbf{x}) \operatorname{div} \boldsymbol{\psi}(\mathbf{x}) \, dx, \end{aligned} \quad (23)$$

where \mathbf{n}_K is the outer normal to K and $(\mathbf{u}_n)|_K$ denotes values on σ from K . Since $\boldsymbol{\psi} = 0$ on $\partial\Omega$ and $\boldsymbol{\psi} \cdot \mathbf{n}_K + \boldsymbol{\psi} \cdot \mathbf{n}_L = 0$ on the interface σ between

K and L, we have

$$\begin{aligned} & \sum_{K \in \mathcal{T}_n} \sum_{\sigma \in \mathcal{E}_K} \int_{\sigma} (\mathbf{u}_n)_{\sigma} \mathbf{n}_K \cdot \boldsymbol{\psi}(\mathbf{x}) \, dS(\mathbf{x}) \\ &= \sum_{\sigma \in \mathcal{E}, \sigma \subset \Omega} \int_{\sigma} (\mathbf{u}_n)_{\sigma} [\mathbf{n}_K \cdot \boldsymbol{\psi}(\mathbf{x}) + \mathbf{n}_L \cdot \boldsymbol{\psi}(\mathbf{x})] \, dS(\mathbf{x}) = 0, \end{aligned}$$

and thus

$$Z_n = \sum_{K \in \mathcal{T}_n} \sum_{\sigma \in \mathcal{E}_K} \int_{\sigma} [(\mathbf{u}_n)_{|K}(\mathbf{x}) - (\mathbf{u}_n)_{\sigma}] \mathbf{n}_K \cdot \boldsymbol{\psi}(\mathbf{x}) \, dS(\mathbf{x}).$$

By the definition of $(\mathbf{u}_n)_{\sigma}$ we have $\int_{\sigma} [(\mathbf{u}_n)_{|K}(\mathbf{x}) - (\mathbf{u}_n)_{\sigma}] \, dS(\mathbf{x}) = 0$. Using $|(\mathbf{u}_n)_{|K} - (\mathbf{u}_n)_{\sigma}| \leq \text{diam}(K) |(\nabla_b \mathbf{u}_n)_{|K}|$ and the smoothness of $\boldsymbol{\psi}$, we infer

$$\begin{aligned} |Z_n| &= \left| \sum_{K \in \mathcal{T}_n} \sum_{\sigma \in \mathcal{E}_K} \int_{\sigma} [(\mathbf{u}_n)_{|K}(\mathbf{x}) - (\mathbf{u}_n)_{\sigma}] \mathbf{n}_K \cdot [\boldsymbol{\psi}(\mathbf{x}) - \boldsymbol{\psi}(\bar{\mathbf{x}}_{\sigma})] \, dS(\mathbf{x}) \right| \\ &\leq C_{\boldsymbol{\psi}} h_{\mathcal{M}_n} \sum_{K \in \mathcal{T}_n} \sum_{\sigma \in \mathcal{E}_K} |\sigma| h_K |(\nabla_b \mathbf{u}_n)_{|K}| \leq 3C_{\boldsymbol{\psi}} C_7 h_{\mathcal{M}_n} \| \nabla_b \mathbf{u}_n \|_{L^1(\Omega)}, \end{aligned}$$

with C_7 not depending on n (we used the regularity assumption on \mathcal{T}_n to write $|\sigma| h_K \leq C_7 |K|$). Invoking the discrete energy estimate (20), we deduce that $Z_n \rightarrow 0$ and we therefore evaluate the limit of (23) since $\mathbf{u}_n \rightarrow \mathbf{u}$ in $L^2(\Omega)$ and $\nabla_b \mathbf{u}_n \rightarrow \boldsymbol{\chi}$ weakly in $L^2(\Omega)^d$. This gives $\int_{\Omega} \boldsymbol{\chi}(\mathbf{x}) \cdot \boldsymbol{\psi}(\mathbf{x}) \, d\mathbf{x} = - \int_{\Omega} \bar{\mathbf{u}}(\mathbf{x}) \operatorname{div} \boldsymbol{\psi}(\mathbf{x}) \, d\mathbf{x}$, which proves that $\boldsymbol{\chi} = \nabla \bar{\mathbf{u}}$ as required.

3 Extension to non-linear models

The previous technique, based on the convergence path 1–3 and on the discrete Rellich theorem and the discrete Poincaré inequality, would not be very useful if it only applied to the linear diffusion equation (2). Convergence of numerical

methods for this equation is well-known, and best obtained through error estimates. The power of the compactness techniques presented above is that they seamlessly apply to non-linear models, including models of physical relevance such as oil recovery and the Navier–Stokes equations. Presenting a complete review of these techniques on such models is beyond the scope of this article, but we give an overview of some of the latest developments.

3.1 Stationary equations

3.1.1 Academic example

We first show with an academic example how to apply the previous techniques to a non-linear model. We consider

$$\begin{cases} -\operatorname{div}[A(\cdot, \bar{\mathbf{u}})\nabla\bar{\mathbf{u}}] = F(\bar{\mathbf{u}}) & \text{in } \Omega, \\ \bar{\mathbf{u}} = 0 & \text{on } \partial\Omega, \end{cases} \quad (24)$$

where $F : \mathbb{R} \mapsto \mathbb{R}$ is continuous and bounded, and $A : \Omega \times \mathbb{R} \mapsto \mathcal{M}_d(\mathbb{R})$ is a Caratheodory function (measurable with respect to $\mathbf{x} \in \Omega$, continuous with respect to $\mathbf{s} \in \mathbb{R}$) such that for all $\mathbf{s} \in \mathbb{R}$ the function $A(\cdot, \mathbf{s})$ satisfies assumption 3 with $\underline{\mathbf{a}}$ and $\bar{\mathbf{a}}$ not depending on \mathbf{s} . The weak form of (24) consists of (3) with $f(\mathbf{x})$ and $A(\mathbf{x})$ replaced with $F(\bar{\mathbf{u}}(\mathbf{x}))$ and $A(\mathbf{x}, \bar{\mathbf{u}}(\mathbf{x}))$, respectively.

As in the linear model, establishing the convergence of a numerical method for (24) by using discrete functional analysis techniques consists of mimicking estimates on the continuous equation. Here, these estimates are obtained as for the linear model; substituting $\mathbf{v} = \bar{\mathbf{u}}$ in the weak form of (24) and using the coercivity of A , the bound on F and the Poincaré inequality, it is seen that $\bar{\mathbf{u}}$ satisfies

$$\|\bar{\mathbf{u}}\|_{H_0^1(\Omega)} \leq \operatorname{diam}(\Omega)\underline{\mathbf{a}}^{-1}|\Omega|^{1/2}\|F\|_{L^\infty(\mathbb{R})}.$$

Writing a numerical method for (24) using a method for the linear equation (2) is usually quite straightforward: all $f(x)$ and $A(x)$ appearing in the definition of the method (e.g., through τ_σ for the TPFA method) have to be replaced with $F(\mathbf{u}(x))$ and $A(x, \mathbf{u}(x))$, where \mathbf{u} is the approximation sought through the scheme. A quick inspection of convergence steps 1 in Sections 2.3.1 and 2.3.2 shows that the discrete energy estimates (14), (20) and (21) hold with $\|f\|_{L^2(\Omega)}$ replaced with $|\Omega|^{1/2}\|F\|_{L^\infty(\mathbb{R})}$.

Convergence step 2 then follows from Theorem 4 exactly as in the linear case, and we find $\bar{\mathbf{u}} \in H_0^1(\Omega)$ such that up to a subsequence $\mathbf{u}_n \rightarrow \bar{\mathbf{u}}$ in $L^2(\Omega)$. This ensures that $F(\mathbf{u}_n) \rightarrow F(\bar{\mathbf{u}})$ in $L^2(\Omega)$, and that up to a subsequence $A(\cdot, \mathbf{u}_n) \rightarrow A(\cdot, \bar{\mathbf{u}})$ almost everywhere while remaining uniformly bounded. These convergences enable us to evaluate the limit of the scheme by following the exact same technique as in convergence step 3 for the linear model. This establishes that $\bar{\mathbf{u}}$ is a weak solution of (24).

Remark 7. Although the strong convergence of \mathbf{u}_n to $\bar{\mathbf{u}}$ is not necessary in the linear case (weak convergence would suffice), it is essential for non-linear models such as (24). Indeed, if $(\mathbf{u}_n)_{n \in \mathbb{N}}$ only converges weakly, then $F(\mathbf{u}_n)$ and $A(\cdot, \mathbf{u}_n)$ may not converge to the correct limits $F(\bar{\mathbf{u}})$ and $A(\cdot, \bar{\mathbf{u}})$, respectively.

3.1.2 Physical models

As mentioned in the introduction, the strength of a convergence analysis via compactness techniques is that it applies to fully non-linear models that are relevant in a number of applications.

Elliptic equations with measure data Equations of the form (2) appear in models of oil recovery, in which f models wells. The relative scales of the reservoir and the wellbores justify taking a Radon measure for this source term [34, 26]. The ensuing analysis is more complex. To start with, the weak formulation (3) is no longer suitable [3, 10]. Moreover, due to the singularity

of the source term, the solution has very weak regularity properties and may not be unique. This prevents any proof of error estimates for numerical approximations of these models.

Discrete functional analysis tools were developed to establish the convergence of the TPFA finite volume scheme for diffusion and (possibly non-coercive) convection-diffusion equations with measures as source terms [37, 24]. Key elements to obtaining a priori estimates on the solutions to these equations are the Sobolev spaces $W_0^{1,p}(\Omega)$ (which is $H_0^1(\Omega)$ if $p = 2$), and the Sobolev embeddings. The corresponding numerical analysis requires the discrete $W_0^{1,p}$ norm on $X_{\mathcal{M}}$,

$$\|\mathbf{v}\|_{W_0^{1,p},\mathcal{M}}^p := \sum_{\sigma \in \mathcal{E}_{\mathcal{M}}} |\sigma| d_{\sigma} \left(\frac{v_K - v_L}{d_{\sigma}} \right)^p,$$

in order to generalise the discrete Poincaré and Rellich theorems to a discrete non-Hilbertian setting, and to establish discrete Sobolev embeddings, that is, if $p \in (1, d)$ and $q \leq \frac{dp}{d-p}$, then

$$\|\mathbf{v}\|_{L^q(\Omega)} \leq C \|\mathbf{v}\|_{W_0^{1,p},\mathcal{M}}. \tag{25}$$

Remark 8. The most efficient proofs of the discrete Poincaré and Rellich theorems use the discrete Sobolev embeddings [30, 19].

Remark 9. The numerical study of (2) with f measure is currently (mostly) limited to the TPFA scheme, since, in general, no other method has the structure that enables the mimicking of the continuous estimates [14].

Leray–Lions and p -Laplace equations These models are non-linear generalisations of (2) that appear in models of gaciology [39]. They have a more severe non-linearity than (24) since they involve both $\bar{\mathbf{u}}$ and $\nabla \bar{\mathbf{u}}$. The general form of these equations is obtained by replacing $\operatorname{div}(\mathbf{A}\nabla \bar{\mathbf{u}})$ in (2) with $\operatorname{div}[\mathbf{a}(\cdot, \bar{\mathbf{u}}, \nabla \bar{\mathbf{u}})]$, where $\mathbf{a} : \Omega \times \mathbb{R} \times \mathbb{R}^d \mapsto \mathbb{R}^d$ satisfies growth, monotonicity and coercivity assumptions. The simplest form is possibly the p -Laplace equation $-\operatorname{div}(|\nabla \bar{\mathbf{u}}|^{p-2} \nabla \bar{\mathbf{u}}) = f$ for $p \in (1, \infty)$.

Uniqueness may fail for these equations [21, Remark 3.4], which completely prevents classical error estimates for their numerical approximations. Compactness techniques were used to study the convergence of at least three different schemes for Leary–Lions equations: the mixed finite volume method [13]; the discrete duality finite volume method [1]; and a cell-centred finite volume scheme [29]. These studies make use of discrete scheme-dependent $W_0^{1,p}$ norms and related discrete Rellich and Poincaré theorems. They also require an (easy) adaptation to the discrete setting of the Minty monotonicity method to deal with the non-linearity involving $\nabla \bar{\mathbf{u}}$.

3.2 Time-dependent and Navier–Stokes equations

Studying non-linear time-dependent models requires space–time compactness results. In the context of Sobolev spaces, these results are usually variants of the Aubin–Simon theorem [2, 40] which, roughly speaking, ensures the compactness in $L^p(\Omega \times (0, T))$ of a sequence $(\mathbf{u}_n)_{n \in \mathbb{N}}$, provided that $(\nabla \mathbf{u}_n)_{n \in \mathbb{N}}$ is bounded in $L^p(\Omega \times (0, T))^d$ and that $(\partial_t \mathbf{u}_n)_{n \in \mathbb{N}}$ is bounded in $L^q(0, T; W^{-1,r}(\Omega))$, where $W^{-1,r}(\Omega) = (W_0^{1,r'}(\Omega))'$. These are natural spaces in which solutions to parabolic PDEs can be estimated.

Carrying out the numerical analysis of these equations with irregular data necessitates the development of discrete versions of the Aubin–Simon theorem; this often includes designing a discrete dual norm mimicking the norm in $W^{-1,r'}(\Omega)$. This analysis was done for various schemes and models: transient Leray–Lions equations [21], including non-local dependencies of $\mathbf{a}(\mathbf{x}, \bar{\mathbf{u}}, \nabla \bar{\mathbf{u}})$ with respect to $\bar{\mathbf{u}}$ (as in image segmentation [33]); a model of miscible fluid flows in porous media from oil recovery [6, 7]; the Stefan model of melting material [27]; and Richards’ model and multi-phase flows in porous media [32]. Discrete Aubin–Simon theorems also sometimes need to be completed with other compactness results, such as compactness results involving sequences of discrete spaces [38] or discrete compensated compactness theorems [17], to deal with degenerate parabolic PDEs.

All these compactness results only provide strong convergence in a space–time averaged norm (e.g., $L^p(\Omega \times (0, T))$ for some $p < \infty$). However, Droniou et al. [17, 22] recently developed a technique to establish a uniform-in-time convergence result (i.e., in $L^\infty(0, T; L^2(\Omega))$) by combining the initial averaged convergences, energy estimates from the PDE, and a discontinuous weak Ascoli–Arzela theorem. This strong uniform convergence corresponds to the needs of end-users, who are usually more interested in the behaviour of the solution at the final time rather than averaged over time.

Navier–Stokes equations The regularity and uniqueness of the solution of the Navier–Stokes equations is a famous open problem. Therefore, as explained in the introduction, the convergence analysis of numerical schemes for these equations cannot be based on error estimates. If it is to be rigorously carried out under reasonable physical assumptions, then this convergence analysis can only be done through compactness techniques.

Let us first consider the continuous case. Because of the term $(\bar{\mathbf{u}} \cdot \nabla)\bar{\mathbf{u}}$ in

$$\partial_t \bar{\mathbf{u}} - \Delta \bar{\mathbf{u}} + (\bar{\mathbf{u}} \cdot \nabla)\bar{\mathbf{u}} + \nabla \bar{p} = \mathbf{f}, \quad (26)$$

evaluating the limit from a sequence of approximate solutions $(\mathbf{u}_n)_{n \in \mathbb{N}}$ requires a strong space–time L^2 compactness on $(\mathbf{u}_n)_{n \in \mathbb{N}}$ (since $(\nabla \mathbf{u}_n)_{n \in \mathbb{N}}$ converges only in $L^2(\Omega \times (0, T))^d$ -weak). The Kolmogorov theorem ensures this strong compactness on $(\mathbf{u}_n)_{n \in \mathbb{N}}$ provided we can control the space-translates and time-translates of the functions. The space translates are naturally estimated thanks to the bound on $(\nabla \mathbf{u}_n)_{n \in \mathbb{N}}$, and the the time translates $\|\mathbf{u}_n(\cdot + \tau, \cdot) - \mathbf{u}_n\|_{L^1(0, T; L^2(\Omega))}$ are estimated by

$$\int_{\Omega} |\mathbf{u}_n(t+\tau, \mathbf{x}) - \mathbf{u}_n(t, \mathbf{x})|^2 \, d\mathbf{x} = \int_{\Omega} \int_t^{t+\tau} \partial_t \mathbf{u}_n(s, \mathbf{x}) [\mathbf{u}_n(t+\tau, \mathbf{x}) - \mathbf{u}_n(t, \mathbf{x})] \, d\mathbf{x} \, ds.$$

Equation (26) is then used to substitute $\partial_t \mathbf{u}_n$ in terms of \mathbf{u}_n and its space derivatives (since $\operatorname{div} \mathbf{u}_n = 0$, the term involving ∇p_n disappears). Bounding the term $(\mathbf{u}_n \cdot \nabla)\mathbf{u}_n \times \mathbf{u}_n$ that appears after this substitution requires Sobolev

estimates on $(\mathbf{u}_n)_{n \in \mathbb{N}}$; these ensure that, only considering the space integral, $\mathbf{u}_n \in L^6(\Omega)$ and thus $|\mathbf{u}_n|^2 |\nabla \mathbf{u}_n| \in L^{6/5}(\Omega)$ (without Sobolev estimates, $\mathbf{u}_n \in L^2(\Omega)$ and $|\mathbf{u}_n|^2 |\nabla \mathbf{u}_n|$ is not even integrable).

The same issue arises in the convergence analysis of numerical methods for Navier–Stokes equations. Discrete Sobolev estimates of the kind (25) are required to estimate the time-translates of the approximate solutions and ensure the convergence towards the correct model. Droniou and Eymard [16] did this for the mixed finite volume method, and Chenier et al. [8] considered an extension of the marker-and-cell scheme. Both references establish more scheme-specific Sobolev embeddings than (25), but this general inequality is sufficient for the analyses carried out in these works.

4 Conclusions and perspectives

We presented techniques that enable the convergence analysis of numerical schemes for PDEs under assumptions that are compatible with field applications. In particular, discontinuous coefficients or fully non-linear physically relevant models can be handled. These techniques do not require the uniqueness or regularity of the solutions and are based on discrete functional analysis tools—that is, the translation to the discrete setting of the functional analysis used in the study of the PDEs.

These discrete tools were adapted to a number of schemes, including the hybrid mixed mimetic family [20] (which contains the hybrid finite volumes [30], the mimetic finite differences [5] and the mixed finite volumes [15]), the discrete duality finite volumes [1], and the discontinuous Galerkin methods [12].

It might appear from our brief introduction that the discrete Sobolev norms and all related results (Poincaré, Rellich, etc.) require specific adaptations for each scheme or model. This is usually not the case. A framework was recently designed, the gradient scheme framework [31, 21, 19], that enables the unified convergence analysis of many different schemes for many diffusion

PDEs. The idea is to identify a set of five properties that are not related to any model but are intrinsic to the discrete space and operators (gradient, etc.) of the numerical methods. Convergence proofs of numerical approximations of many different models can be carried out based on only these five properties (sometimes even fewer). Generic discrete functional analysis tools exist to ensure that several well-known schemes (including meshless methods) satisfy these properties [23], and therefore that the aforementioned convergence results apply to these schemes. The gradient scheme framework covers several boundary conditions, and also guided the design of new schemes [31, 18].

References

- [1] B. Andreianov, F. Boyer, and F. Hubert. Discrete duality finite volume schemes for Leray–Lions-type elliptic problems on general 2D meshes. *Numer. Meth. Part. D. E.*, 23(1):145–195, 2007. doi:[10.1002/num.20170](https://doi.org/10.1002/num.20170) C119, C121
- [2] J.-P. Aubin. Un théorème de compacité. *C. R. Math. Acad. Sci. Paris*, 256:5042–5044, 1963. <http://gallica.bnf.fr/ark:/12148/bpt6k4006n/f1164.item.r=.zoom> C119
- [3] P. Bénilan, L. Boccardo, T. Gallouët, R. Gariepy, M. Pierre, and J. L. Vázquez. An L^1 -theory of existence and uniqueness of solutions of nonlinear elliptic equations. *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* 22(2):241–273, 1995. http://www.numdam.org/item?id=ASNSP_1995_4_22_2_241_0 C103, C117
- [4] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, 2011. doi:[10.1007/978-0-387-70914-7](https://doi.org/10.1007/978-0-387-70914-7) C103, C105, C106

- [5] F. Brezzi, K. Lipnikov, and V. Simoncini. A family of mimetic finite difference methods on polygonal and polyhedral meshes. *Math. Models Meth. Appl. Sci.* 15(10):1533–1551, 2005. doi:[10.1142/S0218202505000832](https://doi.org/10.1142/S0218202505000832) C121
- [6] C. Chainais-Hillairet and J. Droniou. Convergence analysis of a mixed finite volume scheme for an elliptic-parabolic system modeling miscible fluid flows in porous media. *SIAM J. Numer. Anal.* 45(5):2228–2258, 2007. doi:[10.1137/060657236](https://doi.org/10.1137/060657236) C119
- [7] C. Chainais-Hillairet, S. Krell, and A. Mouton. Convergence analysis of a DDFV scheme for a system describing miscible fluid flows in porous media. *Numer. Meth. Part. D. E.*, 31(3):723–760, 2015. doi:[10.1002/num.21913](https://doi.org/10.1002/num.21913) C119
- [8] E. Chénier, R. Eymard, T. Gallouët, and R. Herbin. An extension of the MAC scheme to locally refined meshes: convergence analysis for the full tensor time-dependent Navier–Stokes equations. *Calcolo* 52(1):69–107, 2015. doi:[10.1007/s10092-014-0108-x](https://doi.org/10.1007/s10092-014-0108-x) C121
- [9] M. Crouzeix and P.-A. Raviart. Conforming and nonconforming finite element methods for solving the stationary Stokes equations I. *Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge*, 7(R-3):33–75, 1973. <https://eudml.org/doc/193250> C112
- [10] G. Dal Maso, F. Murat, L. Orsina, and A. Prignet. Renormalized solutions of elliptic equations with general measure data. *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* 28(4):741–808, 1999. http://www.numdam.org/item?id=ASNSP_1999_4_28_4_741_0 C103, C117
- [11] D. A. Di Pietro and M. Vohralik. A review of recent advances in discretization methods, a posteriori error analysis, and adaptive algorithms for numerical modeling in geosciences. *Oil Gas Sci. Technol.* 69(4):701–730, 2014. doi:[10.2516/ogst/2013158](https://doi.org/10.2516/ogst/2013158) C105

- [12] D. A. Di Pietro and A. Ern. *Mathematical aspects of discontinuous Galerkin methods*, volume 69 of *Mathematics & Applications*. Springer, Heidelberg, 2012. doi:[10.1007/978-3-642-22980-0](https://doi.org/10.1007/978-3-642-22980-0) C121
- [13] J. Droniou. Finite volume schemes for fully non-linear elliptic equations in divergence form. *Math. Model. Numer. Anal.*, 40(6):1069–1100 (2007), 2006. doi:[10.1051/m2an:2007001](https://doi.org/10.1051/m2an:2007001) C119
- [14] J. Droniou. Finite volume schemes for diffusion equations: introduction to and review of modern methods. *Math. Models Meth. Appl. Sci.* 24(8):1575–1619, 2014. doi:[10.1142/S0218202514400041](https://doi.org/10.1142/S0218202514400041) C106, C118
- [15] J. Droniou and R. Eymard. A mixed finite volume scheme for anisotropic diffusion problems on any grid. *Numer. Math.* 105(1):35–71, 2006. doi:[10.1007/s00211-006-0034-1](https://doi.org/10.1007/s00211-006-0034-1) C121
- [16] J. Droniou and R. Eymard. Study of the mixed finite volume method for Stokes and Navier–Stokes equations. *Numer. Meth. Part. D. E.* 25(1):137–171, 2009. doi:[10.1002/num.20333](https://doi.org/10.1002/num.20333) C121
- [17] J. Droniou and R. Eymard. Uniform-in-time convergence of numerical methods for non-linear degenerate parabolic equations. *Numer. Math.*, 2015. doi:[10.1007/s00211-015-0733-6](https://doi.org/10.1007/s00211-015-0733-6) C119, C120
- [18] J. Droniou, R. Eymard, and P. Féron. Gradient schemes for Stokes problem. *IMA J. Numer. Anal.*, 2015. doi:[10.1093/imanum/drv061](https://doi.org/10.1093/imanum/drv061) C122
- [19] J. Droniou, R. Eymard, T. Gallouët, C. Guichard, and R. Herbin. Gradient schemes for elliptic and parabolic problems. Unpublished. C118, C121
- [20] J. Droniou, R. Eymard, T. Gallouët, and R. Herbin. A unified approach to mimetic finite difference, hybrid finite volume and mixed finite volume methods. *Math. Models Meth. Appl. Sci.*, 20(2):265–295, 2010. doi:[10.1142/S0218202510004222](https://doi.org/10.1142/S0218202510004222) C121

- [21] J. Droniou, R. Eymard, T. Gallouët, and R. Herbin. Gradient schemes: a generic framework for the discretisation of linear, nonlinear and nonlocal elliptic and parabolic equations. *Math. Models Meth. Appl. Sci.* 23(13):2395–2432, 2012. doi:[10.1142/S0218202513500358](https://doi.org/10.1142/S0218202513500358) C119, C121
- [22] J. Droniou, R. Eymard, and C. Guichard. Uniform-in-time convergence of numerical schemes for Richards’ and Stefan’s models. In M. Ohlberger J. Fuhrmann and C. Rohde Eds., editors, *Finite Volumes for Complex Applications VII—Methods and Theoretical Aspects*, volume 77, pages 247–254. Springer, 2014. doi:[10.1007/978-3-319-05684-5_23](https://doi.org/10.1007/978-3-319-05684-5_23) C120
- [23] J. Droniou, R. Eymard, and R. Herbin. Gradient schemes: generic tools for the numerical analysis of diffusion equations. *Math. Model. Numer. Anal.*, 2015. doi:[10.1051/m2an/2015079](https://doi.org/10.1051/m2an/2015079) C122
- [24] J. Droniou, T. Gallouët, and R. Herbin. A finite volume scheme for a noncoercive elliptic equation with measure data. *SIAM J. Numer. Anal.*, 41(6):1997–2031, 2003. doi:[10.1137/S0036142902405205](https://doi.org/10.1137/S0036142902405205) C118
- [25] J. Droniou, A. Porretta, and A. Prignet. Parabolic capacity and soft measures for nonlinear equations. *Potential Anal.*, 19(2):99–161, 2003. doi:[10.1023/A:1023248531928](https://doi.org/10.1023/A:1023248531928) C103
- [26] J. Droniou and K. S. Talbot. On a miscible displacement model in porous media flow with measure data. *SIAM J. Math. Anal.*, 46(5):3158–3175, 2014. doi:[10.1137/130949294](https://doi.org/10.1137/130949294) C117
- [27] R. Eymard, P. Féron, T. Gallouët, C. Guichard and R. Herbin. Gradient schemes for the Stefan problem. *Int. J. Finite Vol.*, 10:1–37, 2013. <http://www.i2m.univ-amu.fr/IJFV/spip.php?article47> C119
- [28] R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. In P. G. Ciarlet and J.-L. Lions, editors, *Techniques of Scientific Computing, Part III*, Handbook of Numerical Analysis, VII, pages 713–1020. Elsevier, Amsterdam, 2000. <https://www.elsevier.com/books/>

- [handbook-of-numerical-analysis/ciarlet/978-0-444-89928-6](#)
C109
- [29] R. Eymard, T. Gallouët, and R. Herbin. Cell centred discretisation of non linear elliptic problems on general multidimensional polyhedral grids. *J. Numer. Math.*, 17(3):173–193, 2009. doi:[10.1515/JNUM.2009.010](#) C119
- [30] R. Eymard, T. Gallouët, and R. Herbin. Discretization of heterogeneous and anisotropic diffusion problems on general nonconforming meshes SUSHI: a scheme using stabilization and hybrid interfaces. *IMA J. Numer. Anal.*, 30(4):1009–1043, 2010. doi:[10.1093/imanum/drn084](#) C108, C118, C121
- [31] R. Eymard, C. Guichard, and R. Herbin. Small-stencil 3D schemes for diffusive flows in porous media. *Math. Model. Numer. Anal.* 46(2):265–290, 2012. doi:[10.1051/m2an/2011040](#) C121, C122
- [32] R. Eymard, C. Guichard, R. Herbin, and R. Masson. Gradient schemes for two-phase flow in heterogeneous porous media and Richards equation. *J. Appl. Math. Mech.*, 94(7-8):560–585, 2014. doi:[10.1002/zamm.201200206](#) C119
- [33] R. Eymard, A. Handlovičová, R. Herbin, K. Mikula, and O. Stašová. Gradient schemes for image processing. In *Finite volumes for complex applications VI Problems & Perspectives*, volume 4 of *Springer Proc. Math.*, pages 429–437. Springer, Heidelberg, 2011. doi:[10.1007/978-3-642-20671-9_45](#) C119
- [34] P. Fabrie and T. Gallouët. Modelling wells in porous media flow. *Math. Models Meth. Appl. Sci.*, 10(5):673–709, 2000. doi:[10.1142/S0218202500000367](#) C102, C117
- [35] I. Faille. *Modélisation bidimensionnelle de la genèse et de la migration des hydrocarbures dans un bassin sédimentaire*. PhD thesis, Université Joseph Fourier, Grenoble 1, 1992. C103

- [36] X. B. Feng. On existence and uniqueness results for a coupled system modeling miscible displacement in porous media. *J. Math. Anal. Appl.* 194(3):883–910, 1995. doi:[10.1006/jmaa.1995.1334](https://doi.org/10.1006/jmaa.1995.1334) C104
- [37] T. Gallouët and R. Herbin. Finite volume approximation of elliptic problems with irregular data. In *Finite volumes for complex applications II*, pages 155–162. Hermes Sciences, Paris, 1999. C118
- [38] T. Gallouët and J.-C. Latché. Compactness of discrete approximate solutions to parabolic PDEs—application to a turbulence model. *Commun. Pure Appl. Anal.* 11(6):2371–2391, 2012. doi:[10.3934/cpaa.2012.11.2371](https://doi.org/10.3934/cpaa.2012.11.2371) C119
- [39] R. Glowinski and J. Rappaz. Approximation of a non-linear elliptic problem arising in a non-newtonian fluid flow model in glaciology. *Math. Model. Numer. Anal.* 37(1):175–186, 2003. doi:[10.1051/m2an:2003012](https://doi.org/10.1051/m2an:2003012) C118
- [40] J. Simon. Compact sets in the space $L^p(0, T; B)$. *Annali Mat. Pura appl.* 146:65–96, 1987. doi:[10.1007/BF01762360](https://doi.org/10.1007/BF01762360) C119

Author address

1. **J. Droniou**, School of Mathematical Sciences, Monash University, Clayton, Victoria 3800, Australia.
<mailto:jerome.droniou@monash.edu>