ON NONLINEAR PROBLEMS OF PARABOLIC TYPE WITH IMPLICIT CONSTITUTIVE EQUATIONS INVOLVING FLUX

MIROSLAV BULÍČEK, JOSEF MÁLEK, AND ERIKA MARINGOVÁ

ABSTRACT. We study systems of nonlinear partial differential equations of parabolic type, in which the elliptic operator is replaced by the first order divergence operator acting on a flux function, which is related to the spatial gradient of the unknown through an additional implicit equation. This setting, broad enough in terms of applications, significantly expands the paradigm of nonlinear parabolic problems. Formulating four conditions concerning the form of the implicit equation, we first show that these conditions describe a maximal monotone p-coercive graph. We then establish the global-in-time and large-data existence of (weak) solution and its uniqueness. Towards this goal, we adopt and significantly generalize the Minty method of monotone mappings. A unified theory, containing several novel tools, is developed in a way to be tractable numerically.

1. INTRODUCTION

An initial and boundary value problem for a scalar linear parabolic equation is usually formulated in the following way:

$$
\begin{gathered}
\text{For any given } \Omega \subset \mathbb{R}^d,\ T>0,\ u_0:\Omega\to\mathbb{R},\ u_D:\Sigma_D\to\mathbb{R},\\
f:Q\to\mathbb{R},\ g:\Sigma_N\to\mathbb{R}, \text{ find a function } u:Q\to\mathbb{R} \text{ satisfying}^1
\end{gathered}
$$

$$
\begin{aligned}
\partial_t u - \Delta u &= f && \text{in } Q,\\
u &= u_D && \text{on } \Sigma_D,\\
\nabla u\cdot \boldsymbol{n} &= g && \text{on } \Sigma_N,\\
u(0,\cdot) &= u_0 && \text{in } \Omega.
\end{aligned} \tag{1.1}
$$

Here, $\Omega \subset \mathbb{R}^d$, $d \geq 2$, is supposed to be an open, bounded, connected set with a Lipschitz boundary $\partial\Omega$ consisting of two mutually disjoint parts Γ_D and Γ_N so that $\overline{\Gamma_D \cup \Gamma_N} = \partial\Omega$. Furthermore, $\boldsymbol{n} : \partial\Omega \to \mathbb{R}^d$ is the outer unit normal, $Q := (0,T)\times\Omega$, $\Sigma_D := (0,T)\times\Gamma_D$, and $\Sigma_N := (0,T)\times\Gamma_N$. The set Ω having the above properties will be called Lipschitz domain in what follows.

2010 *Mathematics Subject Classification.* 35K55, 35J66 (primary), and 35Q74, 35Q35 (secondary).

Key words and phrases. nonlinear parabolic systems, implicit constitutive theory, weak solutions, existence, uniqueness.

M. Bulíček and J. Málek acknowledge the support of the project No. 18-12719S financed by the Czech Science foundation (GAČR). E. Maringová acknowledges support from Charles University Research program UNCE/SCI/023, the grant SVV-2020-260583 by the Ministry of Education, Youth and Sports, Czech Republic and from the Austrian Science Fund (FWF), grants P30000, W1245, and F65. M. Bulíček and J. Málek are members of the Nečas Center for Mathematical Modelling.

[1] Instead of Δu, we could consider a general linear elliptic operator of the form $\operatorname{div}(\mathbb{A}(t,\boldsymbol{x})\nabla u)$, where $\mathbb{A}$ fulfills the ellipticity condition: there exists $\alpha > 0$ such that for all $\boldsymbol{z} \in \mathbb{R}^d$, $\boldsymbol{z}\cdot\mathbb{A}(t,\boldsymbol{x})\boldsymbol{z} \geq \alpha|\boldsymbol{z}|^2$. Although this extension has some positive aspects, we avoid it for keeping the introductory section simpler from the point of view of notation.

The problem (1.1) can be equivalently rewritten into the form:

To find, for any data given in (1.1), a couple of functions $(u, \boldsymbol{j}) : Q \to \mathbb{R} \times \mathbb{R}^d$ satisfying

$$\begin{aligned} \partial_t u - \operatorname{div} \boldsymbol{j} &= f && \text{in } Q, \\ \boldsymbol{j} &= \nabla u && \text{in } Q, \\ u &= u_D && \text{on } \Sigma_D, \\ \boldsymbol{j} \cdot \boldsymbol{n} &= g && \text{on } \Sigma_N, \\ u(0, \cdot) &= u_0 && \text{in } \Omega. \end{aligned} \tag{1.2}$$

The mixed formulation (1.2) has several advantages: it frequently reflects how the problem is generated (as the first equation in (1.2) is in the form of balance equation and the second equation is the simplest example of the *constitutive* equations describing how the flux $\boldsymbol{j}$ and ∇u are related); it is focused simultaneously on the quantities of interest, i.e., on u and the flux $\boldsymbol{j}$; and it also serves as the starting point of numerical methods that are different from those designed for (1.1).

The a priori information associated with (1.2), the so-called energy (in)equality, provides a natural functional setting in which a robust mathematical theory should be developed. Taking for simplicity $u_D = 0$, the energy equality associated with (1.2) takes the form

$$\frac{1}{2}\|u(t)\|_2^2 + \int_0^t \int_\Omega \boldsymbol{j} \cdot \nabla u \, \mathrm{d}x \, \mathrm{d}t = \int_0^t \int_\Omega f u \, \mathrm{d}x \, \mathrm{d}t + \int_0^t \int_{\Gamma_N} g u \, \mathrm{d}S \, \mathrm{d}t + \frac{1}{2}\|u_0\|_2^2.$$

Using $\boldsymbol{j} = \nabla u$ twice, one observes that $\boldsymbol{j} \cdot \nabla u = \frac{|\boldsymbol{j}|^2}{2} + \frac{|\nabla u|^2}{2}$, hence the energy equality leads to

$$\|u(t)\|_2^2 + \int_0^t \|\boldsymbol{j}\|_2^2 \, \mathrm{d}t + \int_0^t \|\nabla u\|_2^2 \, \mathrm{d}t = 2\int_0^t \int_\Omega f u \, \mathrm{d}x \, \mathrm{d}t + 2\int_0^t \int_{\Gamma_N} g u \, \mathrm{d}S \, \mathrm{d}t + \|u_0\|_2^2. \tag{1.3}$$

It is well known that in the setting dictated by (1.3), the problem (1.1), as well as (1.2), are well-posed. For example, the following theorem can be found in [18, Theorem 10.1, page 616], see also [22, Chapter 2] for the elliptic version:

Theorem 1.1. *Let $\Omega \subset \mathbb{R}^d$ be a Lipschitz domain and $T > 0$. Furthermore, assume that $u_0 \in L^2(\Omega)$, $u_D \in W^{1,2}(0, T; (W^{1,2}_{\Gamma_D}(\Omega))^*) \cap L^2(0, T; W^{1,2}(\Omega))$, $f \in L^2(0, T; (W^{1,2}(\Omega))^*)$ and $g \in L^2(0, T; (W^{\frac{1}{2},2}(\partial\Omega))^*)$. Then there exists a unique $u : Q \to \mathbb{R}$ fulfilling*[2]

$$\begin{aligned} u &\in L^2(0, T; W^{1,2}(\Omega)) \cap \mathcal{C}([0, T]; L^2(\Omega)), \\ \partial_t u &\in L^2(0, T; (W^{1,2}_{\Gamma_D}(\Omega)^*)) \end{aligned}$$

such that $u = u_D$ on Σ_D and for a.a. $t \in (0, T)$ there holds:

$$\langle \partial_t u, \varphi \rangle_{W^{1,2}_{\Gamma_D}(\Omega)} + \int_\Omega \nabla u \cdot \nabla \varphi \, dx = \langle f, \varphi \rangle_{W^{1,2}_{\Gamma_D}(\Omega)} \quad \text{for all } \varphi \in W^{1,2}_{\Gamma_D}(\Omega). \tag{1.4a}$$

The initial condition is attained in the strong sense, i.e.,

$$\lim_{t \to 0_+} \|u(t) - u_0\|_{L^2(\Omega)} = 0. \tag{1.4b}$$

The aim of this study is to present a robust and possibly elegant mathematical theory for a class of problems similar to (1.2), with one remarkable difference, namely, the linear relation between the flux $\boldsymbol{j}$ and ∇u is replaced by an implicit constitutive equation

$$\boldsymbol{g}(\boldsymbol{j}, \nabla u) = \boldsymbol{0} \quad \text{in } Q, \tag{1.5}$$

where $\boldsymbol{g} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is a given continuous (nonlinear) function. The key examples we have in mind and will be covered by the theory presented below are listed in Table 1.

[2] Here, $W^{1,2}_{\Gamma_D}(\Omega)$ is the standard Sobolev space consisting of functions vanishing on Γ_D.

$\boldsymbol{j} = \boldsymbol{k}(\nabla u)$	$\nabla u = \boldsymbol{k}(\boldsymbol{j})$
$\boldsymbol{j} = \lvert\nabla u\rvert^{p-2}\nabla u$	$\nabla u = \lvert\boldsymbol{j}\rvert^{p'-2}\boldsymbol{j}$
$\boldsymbol{j} = (1+\lvert\nabla u\rvert)^{p-2}\nabla u$	$\nabla u = (1+\lvert\boldsymbol{j}\rvert)^{p'-2}\boldsymbol{j}$
$\boldsymbol{j} = (1+\lvert\nabla u\rvert^2)^{\frac{p-2}{2}}\nabla u$	$\nabla u = (1+\lvert\boldsymbol{j}\rvert^2)^{\frac{p'-2}{2}}\boldsymbol{j}$
$\boldsymbol{j} = (\lvert\nabla u\rvert-\delta_*)^+\frac{\nabla u}{\lvert\nabla u\rvert}$	$\nabla u = (\lvert\boldsymbol{j}\rvert-\sigma_*)^+\frac{\boldsymbol{j}}{\lvert\boldsymbol{j}\rvert}$

TABLE 1. The implicit relation (1.5) contains two classes of explicit relations (the left and the right column) that include various power-law relations as well as the problems with activations (jumps and degeneracies). Regarding the parameters, $p \in (1,+\infty)$, $p' = p/(p-1)$, and $\delta_*, \sigma_* > 0$. The structure of these relations is motivated by the classification of incompressible fluid models presented recently in [3].

Examples that belong to the class (1.5) and are covered by the theory presented below, but cannot be included into any column in Table 1, are sketched in Figure 1. These are simple examples underlying the full strength of the implicit constitutive theory.

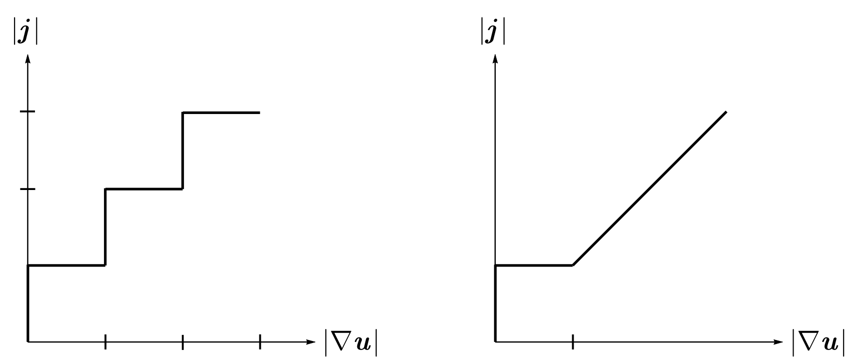

FIGURE 1. The drawing at left represents a relation that can remind the step function both from $\boldsymbol{j}$ and ∇u viewpoint. Note that one can obtain this drawing by considering the $\sqrt{2}$-periodic zig-zag function with the magnitude $\sqrt{2}/2$ rotated by 45 degrees in $(\boldsymbol{j}, \nabla u)$-plane. The drawing at right represents a relation characterized by one simple step followed by the linear relation $\boldsymbol{j} = \nabla \boldsymbol{u}$. Both curves are continuous, none of then can be written in the form $\boldsymbol{j} = \boldsymbol{k}(\nabla u)$ or $\boldsymbol{k} = \boldsymbol{k}(\nabla u)$.

Note that Table 1 includes relations that lead to standard p-Laplace operator $\operatorname{div}(|\nabla u|^{p-2}\nabla u)$ and their variants $\operatorname{div}((1+|\nabla u|^2)^{\frac{p-2}{2}}\nabla u)$ or $\operatorname{div}((1+|\nabla u|)^{p-2}\nabla u)$, however, it also contains less investigated forms, namely

$$\operatorname{div}\boldsymbol{j} \quad \text{with} \quad \nabla u = (1+|\boldsymbol{j}|^2)^{\frac{p'-2}{2}}\boldsymbol{j}.$$

It also covers degenerate operators of type $\operatorname{div}((|\nabla u| - \delta_*)^+ \frac{\nabla u}{|\nabla u|})$ as well as the "multivalued" relations of the type

$$\operatorname{div} \boldsymbol{j} \quad \text{with} \quad \nabla u = (|\boldsymbol{j}| - \sigma_*)^+ \frac{\boldsymbol{j}}{|\boldsymbol{j}|},$$

which are more frequently written in the form

$$\begin{aligned} |\boldsymbol{j}| \le \sigma_* &\iff \nabla u = \mathbf{0}, \\ |\boldsymbol{j}| > \sigma_* &\iff \boldsymbol{j} = \sigma_* \frac{\nabla u}{|\nabla u|} + \nabla u. \end{aligned}$$

In the theory developed in this work we show that the implicit relation (1.5) is fulfilled almost everywhere (i.e., point-wise) in Q provided that $\boldsymbol{g}$ fulfills the following conditions:

(g1) $\boldsymbol{g}$ is Lipschitz continuous[3], i.e., $\boldsymbol{g} \in \mathcal{C}^{0,1}(\mathbb{R}^d \times \mathbb{R}^d)^d$ and $\boldsymbol{g}(\mathbf{0}, \mathbf{0}) = \mathbf{0}$,

(g2) for almost all $(\boldsymbol{j}, \boldsymbol{d}) \in \mathbb{R}^d \times \mathbb{R}^d$:

$$\boldsymbol{g}_{\boldsymbol{j}}(\boldsymbol{j}, \boldsymbol{d}) \ge 0, \quad \boldsymbol{g}_{\boldsymbol{d}}(\boldsymbol{j}, \boldsymbol{d}) \le 0, \quad \boldsymbol{g}_{\boldsymbol{j}}(\boldsymbol{j}, \boldsymbol{d}) - \boldsymbol{g}_{\boldsymbol{d}}(\boldsymbol{j}, \boldsymbol{d}) > 0 \quad \text{and} \quad \boldsymbol{g}_{\boldsymbol{d}}(\boldsymbol{j}, \boldsymbol{d})(\boldsymbol{g}_{\boldsymbol{j}}(\boldsymbol{j}, \boldsymbol{d}))^T \le 0,$$

(g3) one of the following conditions holds:

$$\text{either} \quad \forall \boldsymbol{d} \in \mathbb{R}^d \quad \liminf_{|\boldsymbol{j}| \to +\infty} \boldsymbol{g}(\boldsymbol{j}, \boldsymbol{d}) \cdot \boldsymbol{j} > 0 \quad \text{or} \quad \forall \boldsymbol{j} \in \mathbb{R}^d \quad \limsup_{|\boldsymbol{d}| \to +\infty} \boldsymbol{g}(\boldsymbol{j}, \boldsymbol{d}) \cdot \boldsymbol{d} < 0,$$

(g4) for an arbitrary but fix $p \in (1, \infty)$ there exist $c_1, c_2 > 0$ such that for all $(\boldsymbol{j}, \boldsymbol{d}) \in \mathbb{R}^d \times \mathbb{R}^d$ fulfilling $\boldsymbol{g}(\boldsymbol{j}, \boldsymbol{d}) = \mathbf{0}$ the following condition holds:

$$\boldsymbol{j} \cdot \boldsymbol{d} \ge c_1(|\boldsymbol{j}|^{p'} + |\boldsymbol{d}|^p) - c_2. \qquad (p' := p/(p-1))$$

In (g1)–(g4), we used the following notation. The mappings $\boldsymbol{g}_{\boldsymbol{j}}, \boldsymbol{g}_{\boldsymbol{d}} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{d \times d}$ are defined via

$$\boldsymbol{g}_{\boldsymbol{j}}(\boldsymbol{j}, \boldsymbol{d}) := \frac{\partial \boldsymbol{g}(\boldsymbol{j}, \boldsymbol{d})}{\partial \boldsymbol{j}} \quad \text{and} \quad \boldsymbol{g}_{\boldsymbol{d}}(\boldsymbol{j}, \boldsymbol{d}) := \frac{\partial \boldsymbol{g}(\boldsymbol{j}, \boldsymbol{d})}{\partial \boldsymbol{d}},$$

which written component-wise means (here, $\boldsymbol{g} = (g_1, \ldots, g_d)$, $\boldsymbol{j} = (j_1, \ldots, j_d)$, and $\boldsymbol{d} = (d_1, \ldots, d_d)$)

$$(\boldsymbol{g}_{\boldsymbol{j}})_{ab} := \frac{\partial g_a(\boldsymbol{j}, \boldsymbol{d})}{\partial j_b} \quad \text{and} \quad (\boldsymbol{g}_{\boldsymbol{d}})_{ab} := \frac{\partial g_a(\boldsymbol{j}, \boldsymbol{d})}{\partial d_b}.$$

Further, $(\boldsymbol{g}_{\boldsymbol{d}})^T$ denotes the transposed matrix to $\boldsymbol{g}_{\boldsymbol{d}}$ and $\boldsymbol{g}_{\boldsymbol{j}}(\boldsymbol{j}, \boldsymbol{d})(\boldsymbol{g}_{\boldsymbol{d}}(\boldsymbol{j}, \boldsymbol{d}))^T$ is the standard matrix multiplication. Also, for any matrix $A \in \mathbb{R}^{d \times d}$, the expression $A \ge 0$ means that for any $\boldsymbol{x} \in \mathbb{R}^d$ there holds

$$A\boldsymbol{x} \cdot \boldsymbol{x} \ge 0 \qquad \text{(which written in components is } \sum_{i,j=1}^d A_{ij} x_i x_j \ge 0).$$

In addition, if we write $A > 0$ then it means that the above inequality is strict for all $\boldsymbol{x} \neq \mathbf{0}$. Also, since we can replace $\boldsymbol{g}$ by $-\boldsymbol{g}$, it is clear that all inequalities in (g2) and (g3) can be equivalently formulated with the opposite sign except the last inequality in (g2).

The assumptions (g1)–(g4) are fulfilled by all constitutive equations listed in Table 1. As the assumptions (g1)–(g4) might not seem intuitive at the first sight, for reader's convenience, we show the validity of (g1)–(g4) for few selected constitutive equations listed above in Appendix A. Defining $\boldsymbol{\alpha} := \{(\boldsymbol{j}, \boldsymbol{d}) \in \mathbb{R}^d \times \mathbb{R}^d : \boldsymbol{g}(\boldsymbol{j}, \boldsymbol{d}) = \mathbf{0}\}$, we will also show (see Section 3, Lemma 3.4 and

---

[3] Lipschitz continuity is required in order to guarantee the existence of derivatives, which are used in $(g2)$. It would be possible to require merely continuity of $\boldsymbol{g}$ and replace (g2) by the condition:

(g2)* for any $(\boldsymbol{j}_1, \boldsymbol{d}_1), (\boldsymbol{j}_2, \boldsymbol{d}_2) \in \mathbb{R}^d \times \mathbb{R}^d$ satisfying $\boldsymbol{g}(\boldsymbol{j}_i, \boldsymbol{d}_i) = \mathbf{0}$, $i = 1, 2$, the following condition holds:

$$(\boldsymbol{j}_1 - \boldsymbol{j}_2) \cdot (\boldsymbol{d}_1 - \boldsymbol{d}_2) \ge 0.$$

However, in the setting of implicit equations of the form (1.5), it seems easier to check (g2) than to prove that (g2)* holds, unless the considered constitutive equation belongs to one of explicit classes given in Table 1.

Lemma 3.3) that $\boldsymbol{\alpha}$ is a maximal monotone p-coercive graph (see Section 3 for definitions). More precisely, the assumption (g2) implies that $\boldsymbol{\alpha}$ is monotone and then (g3) in combination with continuity of $\boldsymbol{g}$ guarantees that $\boldsymbol{\alpha}$ is maximal monotone. We prefer to start with the assumptions (g1)–(g4) as they can be verified directly from a given form of $\boldsymbol{g}$ and this verification is easier than showing that corresponding $\boldsymbol{\alpha}$ is a maximal monotone p-coercive graph.

As said above, we aim to develop a robust (i.e., large data) theory for parabolic problems with implicit relations between $\boldsymbol{j}$ and ∇u of the form (1.5) assuming that $\boldsymbol{g}$ fulfills (g1)–(g4). Since the tools we are using are not restricted to scalar problems, we develop the theory for general systems. Thus, instead of considering a vector valued $\boldsymbol{g}$, we impose our assumption on its tensorial analogue $\boldsymbol{G}$. Hence, we assume that for some $p \in (1,\infty)$ and $N \in \mathbb{N}$, the function $\boldsymbol{G}$ and its derivatives $\boldsymbol{G_J}$ and $\boldsymbol{G_D}$, defined in analogous way as $\boldsymbol{g_j}$ and $\boldsymbol{g_d}$, fulfill

(G1) $\boldsymbol{G} \in \mathcal{C}^{0,1}(\mathbb{R}^{d\times N} \times \mathbb{R}^{d\times N})^{d\times N}$ and $\boldsymbol{G}(\boldsymbol{0},\boldsymbol{0}) = \boldsymbol{0}$,

(G2) for almost all $(\boldsymbol{J},\boldsymbol{D}) \in \mathbb{R}^{d\times N} \times \mathbb{R}^{d\times N}$:
$$\boldsymbol{G_J}(\boldsymbol{J},\boldsymbol{D}) \geq 0, \quad \boldsymbol{G_D}(\boldsymbol{J},\boldsymbol{D}) \leq 0, \quad \boldsymbol{G_J}(\boldsymbol{J},\boldsymbol{D}) - \boldsymbol{G_D}(\boldsymbol{J},\boldsymbol{D}) > 0,$$
$$\text{and} \quad \boldsymbol{G_D}(\boldsymbol{J},\boldsymbol{D})(\boldsymbol{G_J}(\boldsymbol{J},\boldsymbol{D}))^T \leq 0$$

(G3) one of the following holds:
$$\begin{aligned} &\text{either} \quad \forall \boldsymbol{D} \in \mathbb{R}^{d\times N} \quad \liminf_{|\boldsymbol{J}|\to+\infty} \boldsymbol{G}(\boldsymbol{J},\boldsymbol{D}) : \boldsymbol{J} > 0 \\ &\text{or} \quad \forall \boldsymbol{J} \in \mathbb{R}^{d\times N} \quad \limsup_{|\boldsymbol{D}|\to+\infty} \boldsymbol{G}(\boldsymbol{J},\boldsymbol{D}) : \boldsymbol{D} < 0, \end{aligned}$$

(G4) there exist $c_1, c_2 > 0$ such that for all $(\boldsymbol{J},\boldsymbol{D}) \in \mathbb{R}^{d\times N} \times \mathbb{R}^{d\times N}$ fulfilling $\boldsymbol{G}(\boldsymbol{J},\boldsymbol{D}) = \boldsymbol{0}$ we have
$$\boldsymbol{J} : \boldsymbol{D} \geq c_1(|\boldsymbol{J}|^{p'} + |\boldsymbol{D}|^p) - c_2.$$

Recall that the constitutive equation $\boldsymbol{G}(\boldsymbol{J},\boldsymbol{D}) = \boldsymbol{0}$ can be replaced by $-\boldsymbol{G}(\boldsymbol{J},\boldsymbol{D}) = \boldsymbol{0}$. Then all inequalities in (G2) and (G3) have the opposite signs except the last inequality in (G2). This ambiguity could be fixed for example by requiring that $\boldsymbol{G}$ is such that the first condition in (G2) holds.

For simplicity, in what follows, we restrict ourselves to homogeneous boundary data, and then the vector-valued analogue of (1.1) reads as follows:

$$\begin{gathered} \text{For any given } \Omega \subset \mathbb{R}^d, T > 0, \boldsymbol{u}_0 : \Omega \to \mathbb{R}^N, \boldsymbol{f} : Q \to \mathbb{R}^N \text{ and} \\ \boldsymbol{G} : \mathbb{R}^{d\times N} \times \mathbb{R}^{d\times N} \to \mathbb{R}^{d\times N} \text{ satisfying } \quad \text{(G1)–(G4)}, \\ \text{find a couple } \boldsymbol{u} : Q \to \mathbb{R}^N \text{ and } \boldsymbol{J} : Q \to \mathbb{R}^{d\times N} \text{ solving the problem} \end{gathered}$$
$$\begin{aligned} \partial_t \boldsymbol{u} - \operatorname{div} \boldsymbol{J} &= \boldsymbol{f} && \text{in } Q, \\ \boldsymbol{G}(\boldsymbol{J}, \nabla \boldsymbol{u}) &= \boldsymbol{0} && \text{in } Q, \\ \boldsymbol{u} &= \boldsymbol{0} && \text{on } \Sigma_D, \\ \boldsymbol{J}\boldsymbol{n} &= \boldsymbol{0} && \text{on } \Sigma_N, \\ \boldsymbol{u}(0,\cdot) &= \boldsymbol{u}_0 && \text{in } \Omega. \end{aligned} \tag{1.6}$$

Considering a system of equations is of real importance. Indeed, all models depicted in Table 1 are of the so-called diagonal form. However, in many real applications, one has to deal with systems of equations that contain the non-diagonal terms in order to describe observed physical effects. The Maxwell–Stefan systems may serve as prototypic examples. The Maxwell–Stefan system describes the diffusive transport of multicomponent mixtures (see for example [14], [4], [16]; regarding the notation we follow [16]), where the governing equations for the concentrations

$u_\nu : (0,T) \times \Omega \to \mathbb{R}$, $0 \le u_\nu \le 1$, take, for $\nu = 1, \dots, N$, $N \ge 2$, the form

$$\partial_t u_\nu - \operatorname{div} \boldsymbol{j}_\nu = r_\nu(\boldsymbol{u}), \tag{1.7}$$

$$\nabla u_\nu = \sum_{\mu=1,\mu\neq\nu}^{N} \alpha_{\mu\nu} \left(c_\mu \boldsymbol{j}_\nu - c_\nu \boldsymbol{j}_\mu \right). \tag{1.8}$$

Here, the constants $\alpha_{\nu\mu}$ are all positive for $\nu \neq \mu$ and $\boldsymbol{u} := (u_1, \dots, u_N)$. Denoting $\boldsymbol{d}_\nu := \nabla u_\nu$, $\boldsymbol{D} := (\boldsymbol{d}_1, \dots, \boldsymbol{d}_N)^T$ and $\boldsymbol{J} := (\boldsymbol{j}_1, \dots, \boldsymbol{j}_N)^T$, the equations (1.8) can be written in the form

$$\boldsymbol{D} = \mathbb{B}(\boldsymbol{u})\boldsymbol{J}, \tag{1.9}$$

where $\mathbb{B}$ is $N \times N$-matrix. It is shown in Appendix B, that $\boldsymbol{G} : \mathbb{R}^{N\times d} \times \mathbb{R}^{N\times d}$ defined through

$$\boldsymbol{G}(\boldsymbol{J}, \boldsymbol{D}) = \mathbb{B}(\boldsymbol{u})\boldsymbol{J} - \boldsymbol{D}$$

fulfill the conditions (G1)–(G3). Of course, since the matrix $\mathbb{B}$ and the right-hand side of the first set of equations in (1.7) depend on the unknown $\boldsymbol{u}$, the theory developed below cannot be applied to (1.7) and the existence analysis of relevant initial- and boundary-value problems must be done in a more delicate way, for which we refer to the above studies [4] and [16]. A thermodynamical basis for diffusive transport of multicomponent mixtures that go beyond Maxwell-Stefan systems (1.9) and are more appropriate for realistic description of mixtures and that belong into the class of fully implicit relations is developed in a recent study [5].

We conclude this introductory section by formulating freely the main result of this study:

For any $\Omega \subset \mathbb{R}^d, T > 0, \boldsymbol{u}_0, \boldsymbol{f}, p \in (1, \infty)$ and $\boldsymbol{G}$ satisfying (G1)–(G4), there is $(\boldsymbol{u}, \boldsymbol{J})$ solving (1.6).

The structure of the remaining parts of the paper is the following. In Section 2 we provide the precise formulation of our main result and summarize its novelties. We also formulate analogous result for the boundary-value problem in the elliptic (i.e., time-independent) case. Then, in Section 3, we recall the concept of the maximal monotone p-coercive graph and establish its connection to implicit constitutive equation $\boldsymbol{G}(\boldsymbol{J}, \boldsymbol{D}) = \boldsymbol{0}$. In particular, we show that the assumptions (G1)–(G4) imply that the null points of $\boldsymbol{G}$ form a maximal monotone p-coercive graph. In Section 4, we construct the appropriate approximating 2-coercive graphs, parameterized by ε, that are shown to be Lipschitz continuous and uniformly monotone mappings. This construction is made very explicitly by using an algebraic structure of monotone graphs. We then investigate the convergence properties between the maximal monotone p-coercive graphs and their approximations and we also add a few additional results useful on its own to this section. Then, in Section 5, we prove the main theorem using the approximation of the null points of $\boldsymbol{G}$ by the Lipschitz continuous and uniformly monotone single-valued mappings constructed and analyzed in the previous section and letting the approximation parameter ε tend to zero. The standard theory regarding the well-posedness of the approximate problems, based on the classical Minty method [23], is, for the sake of completeness, proved in Appendix C. As indicated above, in the Appendices A and B, we focus on several constitutive equations that belong to the class $\boldsymbol{G}(\boldsymbol{J}, \boldsymbol{D}) = \boldsymbol{0}$ and verify that they fulfill the structural assumptions (G1)–(G4).

2. Main result

Before we state the main result of the paper, we fix some notation. We recall that throughout the whole paper $\Omega \subset \mathbb{R}^d$ is a Lipschitz domain (with two mutually non-intersecting essential parts Γ_D and Γ_N of the boundary $\partial\Omega$), as defined in Section 1 after (1.1). For $t \in (0, T]$, we denote $Q_t := [0, t) \times \Omega$ and we also set $Q := Q_T$. The shortcut *a.a.* t stands for *almost all* t.

We employ small boldfaced letters to denote vectors and bold capitals for tensors. We do not relabel the original sequence when selecting a subsequence. The symbols $\boldsymbol{j} \cdot \boldsymbol{d}$ and $\boldsymbol{J} : \boldsymbol{D}$ stand for the scalar product of vectors $\boldsymbol{j}$ and $\boldsymbol{d}$ or tensors $\boldsymbol{J}$ and $\boldsymbol{D}$, respectively. In a time-space domain, the standard differential operators, like gradient (∇) and divergence (div), are always related to the spatial variables only. Also, we use standard notation for partial ($\partial.$) and total ($\frac{d}{d\cdot}$) derivatives. Generic constants, that depend just on data, are denoted by C and may vary on every line.

For a Banach space X, its dual is denoted by X^*. For $x \in X$ and $x^* \in X^*$, the duality is denoted by $\langle x^*, x\rangle_X$. For $p \in [1,\infty]$, we denote $(L^p(\Omega), \|\cdot\|_{L^p(\Omega)})$ and $(W^{1,p}(\Omega), \|\cdot\|_{W^{1,p}(\Omega)})$ the corresponding Lebesgue and Sobolev spaces with the norms defined in standard way,

$$\|f\|_{L^p(\Omega)} := \begin{cases} \left(\int_\Omega |f|^p \,\mathrm{d}x\right)^{\frac{1}{p}} & \text{if } p \in [1,\infty), \\ \operatorname{ess\,sup}_{x\in\Omega} |f(x)| & \text{if } p = \infty, \end{cases}$$
$$\|f\|_{W^{1,p}(\Omega)} := \|f\|_{L^p(\Omega)} + \|\nabla f\|_{L^p(\Omega)}.$$

Bochner space is designated by $L^p(0,T;X)$ and we set

$$\mathcal{C}([0,T];X) := \{f \in L^\infty(0,T;X); [0,T] \ni t^n \to t \implies f(t^n) \to f(t) \text{ strongly in } X\}.$$

We use the notation $L^p(\Omega;\mathbb{R}^N)$ and $L^p(\Omega;\mathbb{R}^{d\times N})$ for Lebesgue spaces of vector- or matrix-valued functions, respectively.

Next, we define the function spaces related to our setting. We set

$$\begin{aligned} V_p &:= \{\boldsymbol{u}; \boldsymbol{u} \in W^{1,p}(\Omega;\mathbb{R}^N) \cap L^2(\Omega;\mathbb{R}^N),\ \boldsymbol{u} = \boldsymbol{0} \text{ on } \Gamma_D\}, \\ H &:= L^2(\Omega;\mathbb{R}^N), \\ V_p^* &:= (V_p)^*, \end{aligned} \tag{2.1}$$

and equip the space V_p by the norm[4] $\|\boldsymbol{u}\|_{V_p} := \|\nabla\boldsymbol{u}\|_{L^p(\Omega)} + \|\boldsymbol{u}\|_{L^2(\Omega)}$. Then for any $p \in (1,\infty)$

$$V_p \hookrightarrow H \equiv H^* \hookrightarrow V_p^*, \tag{2.2}$$

and both embeddings are continuous and dense. Therefore, these spaces form a Gelfand triplet. For simplicity, we also set $V := V_2$ and $V^* := V_2^*$. Note that V and H are Hilbert spaces. Also, the duality in V_p is defined via

$$\langle \boldsymbol{f}, \boldsymbol{\varphi}\rangle_{V_p} := \lim_{k\to+\infty} \int_\Omega \boldsymbol{f}^k \cdot \boldsymbol{\varphi} \,\mathrm{d}x \tag{2.3}$$

for any $\boldsymbol{\varphi} \in V_p$, where $\{\boldsymbol{f}^k\}_{k\in\mathbb{N}}$ is a sequence in H converging to $\boldsymbol{f}$ in V_p^*. Note that in the case when $\boldsymbol{f} \in L^2(\Omega;\mathbb{R}^N)$, this definition just means

$$\langle \boldsymbol{f}, \boldsymbol{\varphi}\rangle_{V_p} = \int_\Omega \boldsymbol{f} \cdot \boldsymbol{\varphi} \,\mathrm{d}x. \tag{2.4}$$

Having introduced the notation, we can now formulate the main result of the paper.

Theorem 2.1. *Let $\Omega \subset \mathbb{R}^d$ be a Lipschitz domain, $T > 0$ and $p \in (1,\infty)$. Let $\boldsymbol{f} \in L^{p'}(0,T;V_p^*)$ and $\boldsymbol{u}_0 \in H$. Assume that $\boldsymbol{G}$ satisfies (G1)–(G4). Then there exists a weak solution to the*

[4]Note that in case $p \geq 2d/(d+2)$, we have, due to the Sobolev embedding and the Poincaré inequality, that $V_p = W^{1,p}_{\Gamma_D}(\Omega;\mathbb{R}^N)$, where

$$W^{1,p}_{\Gamma_D}(\Omega;\mathbb{R}^N) := \{\boldsymbol{u}; \boldsymbol{u} \in W^{1,p}(\Omega;\mathbb{R}^N),\ \boldsymbol{u} = \boldsymbol{0} \text{ on } \Gamma_D\}$$

and the norm in V_p is equivalent to the standard Sobolev norm.

problem (1.6)*, i.e., there exists a couple* $(\boldsymbol{u}, \boldsymbol{J})$ *fulfilling*

$$\begin{aligned} \boldsymbol{u} &\in L^p(0,T;V_p) \cap \mathcal{C}([0,T];H),\\ \partial_t \boldsymbol{u} &\in L^{p'}(0,T;V_p^*),\\ \boldsymbol{J} &\in L^{p'}(Q;\mathbb{R}^{d\times N}),\end{aligned}$$

so that

$$\langle \partial_t \boldsymbol{u}, \boldsymbol{\varphi}\rangle_{V_p} + \int_\Omega \boldsymbol{J} : \nabla \boldsymbol{\varphi}\, dx = \langle \boldsymbol{f}, \boldsymbol{\varphi}\rangle_{V_p} \quad \text{for a.a. } t \in (0,T) \text{ and for all } \boldsymbol{\varphi} \in V_p, \tag{2.5a}$$

$$\boldsymbol{G}(\boldsymbol{J}, \nabla \boldsymbol{u}) = \boldsymbol{0} \quad \text{almost everywhere in } Q, \tag{2.5b}$$

and the initial condition is attained in the strong sense, i.e.,

$$\lim_{t\to 0_+} \|\boldsymbol{u}(t) - \boldsymbol{u}_0\|_H = 0. \tag{2.5c}$$

In addition, $\boldsymbol{u}$ *is uniquely determined.*

Several comments regarding this result and its novelties are in order:

(i) As explicit constitutive equations (as those listed in Table 1) represent important subparts of implicit constitutive equations, there are plenty of (even classical) examples in various areas of science (solid and fluid mechanics, heat transfer, chemistry, electro-magnetism, etc.) named Hooke's, Fourier's, Fick's laws and their various non-linear generalizations that are covered by the equation $\boldsymbol{G}(\boldsymbol{J}, \boldsymbol{D}) = \boldsymbol{0}$. The systematic study of implicit constitutive equations is however more recent and goes back to works by Rajagopal (see [24, 25] for original papers in elasticity and fluid mechanics[5], and also survey papers [10, 26] and Section 4.5 in [21]). This new viewpoint on constitutive theory in terms of implicit equations has been reflected into the analysis of general problems in fluid mechanics in [8, 9, 7], where both the stationary and evolutionary situations have been treated. However, there are, in our opinion, two shortcomings in the theory developed in [8, 9, 7] for incompressible fluid flow problems and in [11] for flows in porous media, and used in some subsequent studies. First, the theory assumes the existence of a *Borel measurable* selection and the whole proof stems from this a priori existence of such a selection. Second, the proof is highly nonconstructive as it applies standard mollification (to the selection) by convolution, which is very hard to implement numerically, see recent studies [12, 17, 27, 13] devoted to the analysis of finite element discretizations of implicitly constituted fluid flow problems. In our proof below, we do not use the convolution at all. Instead we introduce a very simple algebraic modification/approximation of $\boldsymbol{G}$, which is easy to implement In fact, we offer two possible approximations of $\boldsymbol{G}(\boldsymbol{J}, \boldsymbol{D}) = \boldsymbol{0}$. In addition, these approximative schemes always lead to the setting in the Hilbert spaces on which the approximation graph is Lipschitz continuous and uniformly monotone, which seems to be the most friendly situation for numerical purposes.

(ii) Although our proof uses the concepts as monotone and maximal monotone mappings/graphs, we formulate the result without using these terms. This is due to the fact that we have found easy-to-verify conditions on the function $\boldsymbol{G}$, see the conditions (G1)–(G4) (almost) characterizing that the corresponding $\mathcal{A}$ is a maximal monotone p-coercive graph.

(iii) If we identify the null points of $\boldsymbol{G}$ with a set $\mathcal{A}$, subset of the Cartesian product $\mathbb{R}^{d\times N} \times \mathbb{R}^{d\times N}$, then one can reformulate the problem in terms of $\mathcal{A}$, which leads to the theory of monotone mappings. This theory goes back to the seminal work [23]. This theory when further extended to the analysis of partial differential equations or to the problems of calculus of variations has been however stick to the assumption that the flux is a (possibly multivalued)

[5]Note that when reducing the governing equations for incompressible implicitly constituted fluids to simple shear flows one obtains a scalar version of the problem (1.6) studied here.

function of the gradient of the unknown function. It means it reflects mostly the first row in Table 1. Several concepts such as multivalued sets, subdifferential calculus, variational inequalities, differential inclusions, etc. have been used to set-up rigorous mathematical background for relevant problems. One of the aims of this study is to avoid using such objects and provide, in our opinion, more simple mathematical description, see also the point (iv) below.

(iv) There are many results where the null points of $\boldsymbol{G}$ are assumed to be described by a convex potential. To be more precise, if one assumes that there exist a convex $\Phi : \mathbb{R}^{d\times N} \to \mathbb{R}$ and its convex conjugate $\Phi^* : \mathbb{R}^{d\times N} \to \mathbb{R}$ such that

$$\boldsymbol{G}(\boldsymbol{J},\boldsymbol{D}) = \boldsymbol{0} \qquad \Longleftrightarrow \qquad \boldsymbol{J} : \boldsymbol{D} = \Phi(\boldsymbol{J}) + \Phi^*(\boldsymbol{D}) \tag{2.6}$$

then the condition "$\boldsymbol{G}(\boldsymbol{J},\nabla\boldsymbol{u}) = \boldsymbol{0}$ almost everywhere in Q" stated in Theorem 2.1, can be equivalently[6] replaced by the following inequality

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{u}\|_H^2 + \int_\Omega \Phi(\boldsymbol{J}) + \Phi^*(\nabla\boldsymbol{u})\,\mathrm{d}x \le \langle \boldsymbol{f}, \nabla\boldsymbol{u}\rangle_{V_p}. \tag{2.7}$$

It is then obvious that due to the convexity of Φ and Φ^*, one can usually pass to the inequality (2.7) easily without any major difficulties. Unfortunately, such a procedure works only in the case (2.6), which decreases the applicability of such theory significantly. The second (and more important) limitation of this approach is that for consistency, it requires the possibility of using $\boldsymbol{u}$ as a test function in (2.5a), which is typically not the case in problems arising in fluid dynamics, see e.g. [1], where the inequality of the type (2.7) is used to define a notion of solution.

(v) We wish to emphasize that there are interesting constitutive equations of the form $\boldsymbol{G}(\boldsymbol{J},\boldsymbol{D}) = \boldsymbol{0}$ that generate the *non-monotone* graph and consequently the analysis of corresponding problems is not covered by the theory developed in this paper. In fact, a sounding analysis for problems with non-monotone graphs is a challenging open problem. We refer to [19, 15] for more details.

We complete this section by stating the result for the time-independent case. We however do not give the explicit proof of this result here since it is easier than in the time-dependent situation and in fact the proof can be deduced from the detailed proof of Theorem 2.1 directly by eliminating the steps that are due to the dependency of the quantities on time.

Theorem 2.2. *Let $p \in (1,\infty)$, $\Omega \subset \mathbb{R}^d$ be a Lipschitz domain, $\Gamma_D \neq \emptyset$ and $\boldsymbol{f} \in (W^{1,p}_{\Gamma_D}(\Omega;\mathbb{R}^N))^*$. Assume that $\boldsymbol{G}$ satisfies (G1)–(G4). Then there exists a couple $(\boldsymbol{u},\boldsymbol{J})$ fulfilling*

$$\boldsymbol{u} \in W^{1,p}_{\Gamma_D}(\Omega;\mathbb{R}^N),$$
$$\boldsymbol{J} \in L^{p'}(\Omega;\mathbb{R}^{d\times N}),$$
$$\boldsymbol{G}(\boldsymbol{J},\nabla\boldsymbol{u}) = \boldsymbol{0} \text{ almost everywhere in } \Omega,$$

which satisfies for all $\boldsymbol{\varphi} \in W^{1,p}_{\Gamma_D}(\Omega;\mathbb{R}^N)$

$$\int_\Omega \boldsymbol{J} : \nabla\boldsymbol{\varphi}\,dx = \langle \boldsymbol{f},\boldsymbol{\varphi}\rangle_{W^{1,p}_{\Gamma_D}(\Omega;\mathbb{R}^N)}. \tag{2.8a}$$

In addition, if $\boldsymbol{G}$ satisfies

[6]Indeed, if we set $\boldsymbol{\varphi} := \boldsymbol{u}$ in (2.5a) and compare it with (2.7), we see that

$$\int_\Omega \Phi(\boldsymbol{J}) + \Phi^*(\nabla\boldsymbol{u})\,\mathrm{d}x \le \int_\Omega \boldsymbol{J} : \nabla\boldsymbol{u}\,\mathrm{d}x.$$

Then it follows due to the Young inequality that $\Phi(\boldsymbol{J}) + \Phi^*(\nabla\boldsymbol{u}) = \boldsymbol{J} : \nabla\boldsymbol{u}$ almost everywhere and consequently, the assumption (2.6) gives $\boldsymbol{G}(\boldsymbol{J},\nabla\boldsymbol{u}) = \boldsymbol{0}$ almost everywhere.

*(G2)** *for any* $(\boldsymbol{J}_1, \boldsymbol{D}_1), (\boldsymbol{J}_2, \boldsymbol{D}_2) \in \mathbb{R}^{d\times N} \times \mathbb{R}^{d\times N}$ *fulfilling* $\boldsymbol{G}(\boldsymbol{J}_i, \boldsymbol{D}_i) = \boldsymbol{0}$ *and* $\boldsymbol{D}_1 \neq \boldsymbol{D}_2$*:*

$$(\boldsymbol{J}_1 - \boldsymbol{J}_2) : (\boldsymbol{D}_1 - \boldsymbol{D}_2) > 0,$$

then the solution $\boldsymbol{u}$ *is unique.*

The above result does not include the purely Neumann problem, i.e., $\Gamma_N = \partial\Omega$. Nevertheless, the existence statement of Theorem 2.2 remains true provided that the right hand side fulfills the necessary compatibility condition

$$\langle \boldsymbol{f}, \boldsymbol{\varphi} \rangle_{W^{1,p}(\Omega;\mathbb{R}^N)} = 0 \qquad \text{for all constant } \boldsymbol{\varphi} \in \mathbb{R}^N.$$

Moreover, the uniqueness result holds true if restricted to functions with prescribed mean value.

3. Null points of $\boldsymbol{G}$ and maximal monotone graphs

In this part, we identify the null set of $\boldsymbol{G}$ with a subset $\mathcal{A}$ of $\mathbb{R}^{d\times N} \times \mathbb{R}^{d\times N}$ and show that the assumptions (G1)–(G4) delimiting the structure of $\boldsymbol{G}$ imply that $\mathcal{A}$ is a maximal monotone p-coercive graph (see the definition below). Before doing so, we develop a generalized monotone operator theory following the original work [23] as well as [2] and [6] where a similar but less general approach is used.

Let us start with recalling the notion of maximal monotone graphs.

Definition 3.1 (Maximal monotone p-coercive graph)**.** *Let* $p \in (1, \infty)$ *and* $p' := \frac{p}{p-1}$*. We say that a subset* $\mathcal{A}$ *of* $\mathbb{R}^{d\times N} \times \mathbb{R}^{d\times N}$ *is a maximal monotone p-coercive graph if*

(A1) $(\boldsymbol{0}, \boldsymbol{0}) \in \mathcal{A}$.
(A2) For any $(\boldsymbol{J}_1, \boldsymbol{D}_1), (\boldsymbol{J}_2, \boldsymbol{D}_2) \in \mathcal{A}$

$$(\boldsymbol{J}_1 - \boldsymbol{J}_2) : (\boldsymbol{D}_1 - \boldsymbol{D}_2) \geq 0.$$

(A3) If for some $(\boldsymbol{J}, \boldsymbol{D}) \in \mathbb{R}^{d\times N} \times \mathbb{R}^{d\times N}$ *and for all* $(\overline{\boldsymbol{J}}, \overline{\boldsymbol{D}}) \in \mathcal{A}$

$$(\boldsymbol{J} - \overline{\boldsymbol{J}}) : (\boldsymbol{D} - \overline{\boldsymbol{D}}) \geq 0,$$

then $(\boldsymbol{J}, \boldsymbol{D}) \in \mathcal{A}$.
(A4) There exist $C_1, C_2 > 0$ *such that for all* $(\boldsymbol{J}, \boldsymbol{D}) \in \mathcal{A}$

$$\boldsymbol{J} : \boldsymbol{D} \geq C_1(|\boldsymbol{J}|^{p'} + |\boldsymbol{D}|^p) - C_2.$$

The condition (A1) means that $\mathcal{A}$ passes through the origin, (A2) states that the graph $\mathcal{A}$ is <u>monotone</u>, while (A3) states that $\mathcal{A}$ is <u>maximal monotone</u>, i.e., $\mathcal{A}$ *cannot be extended to a properly larger domain while preserving its monotoneity*[7]. Finally, (A4) states that the graph $\mathcal{A}$ is <u>p-coercive</u>.

Remark 3.2. *For further generality, one could replace (A4) with the following condition:*

(A4) There exist* $c^*, c_* > 0$ *and a Young function* ψ *such that for all* $(\boldsymbol{J}, \boldsymbol{D}) \in \mathcal{A}$

$$\boldsymbol{J} : \boldsymbol{D} \geq c^*(\psi(|\boldsymbol{D}|) + \psi^*(|\boldsymbol{J}|)) - c_*.$$

Here, $\psi : \mathbb{R} \to \mathbb{R}^+$ *is a Young function, i.e.,* ψ *is an even continuous convex function such that*

$$\lim_{s\to 0_+} \frac{\psi(s)}{s} = 0 \quad \textit{and} \quad \lim_{s\to+\infty} \frac{\psi(s)}{s} = +\infty,$$

and the convex conjugate function ψ^* *is defined as the Legendre transform of* ψ*, i.e.,*

$$\psi^*(s) := \sup_{l\in\mathbb{R}} (s \cdot l - \psi(l)).$$

[7]The text in italics is exact citation from [23].

The study of the models related via maximal monotone ψ-graphs*, i.e., the graphs satisfying (A1)–(A3) and (A4*), are of interest; however, such an extension is nowdays rather routine and it is not included in this work.*

Next, we prove an auxiliary result adopted from Minty [23] and [2, Proposition 1.1 (applied to dimension $d \times N$)] and adjusted to our situation. More precisely, having a monotone graph $\mathcal{A}$ we can identify it with two possibly multivalued (monotone) mappings $\boldsymbol{J}^*$ and $\boldsymbol{D}^*$, each defined on a subset of $\mathbb{R}^{d\times N}$ through

$$(\overline{\boldsymbol{J}}, \overline{\boldsymbol{D}}) \in \mathcal{A} \iff \overline{\boldsymbol{J}} \in \boldsymbol{J}^*(\overline{\boldsymbol{D}}) \iff \overline{\boldsymbol{D}} \in \boldsymbol{D}^*(\overline{\boldsymbol{J}}). \tag{3.1}$$

Then, $\boldsymbol{J}^*$ is maximal monotone if and only if $\boldsymbol{J}^* + \varepsilon \boldsymbol{I}$ is onto for any $\varepsilon \in (0,1]$ and $\boldsymbol{D}^*$ is maximal monotone if and only if $\boldsymbol{D}^* + \varepsilon \boldsymbol{I}$ is onto for any $\varepsilon \in (0,1]$. In the next lemma, we will prove the first statement noting that the proof of the second equivalence can be done in the same way just by interchanging the role of $\boldsymbol{D}$ and $\boldsymbol{J}$.

Lemma 3.3. *Let $\varepsilon \in (0,1]$ and $\mathcal{A}$ be a monotone graph identified with $\boldsymbol{J}^*$ via* (3.1)*, i.e., $\mathcal{A}$ satisfies (A1) and (A2). Then, $\mathcal{A}$ is maximal monotone, i.e., $\mathcal{A}$ satisfies (A3), if and only if the mapping $\boldsymbol{J}^* + \varepsilon \boldsymbol{I}$ is onto.*

Proof. We split the proof into two steps. In the first one, we show that if $\mathcal{A}$ is maximal monotone graph then $\boldsymbol{J}^* + \boldsymbol{I}$ is onto (i.e., the domain of $(\boldsymbol{J}^* + \boldsymbol{I})^{-1}$ is $\mathbb{R}^{d\times N}$). In the second step we prove the converse implication. In the proof, we restrict ourselves to the case $\varepsilon = 1$ as the proof can be easily extended to the case $\varepsilon \in (0,1)$.

Step 1. We start with defining a set

$$\mathrm{Im} := \{\boldsymbol{Z} \in \mathbb{R}^{d\times N}; \exists (\boldsymbol{J}, \boldsymbol{D}) \in \mathcal{A}, \boldsymbol{Z} = \boldsymbol{J} + \boldsymbol{D}\} \tag{3.2}$$

and our goal is to define $(\boldsymbol{J}^* + \boldsymbol{I})^{-1}$ on Im and show that $\mathrm{Im} = \mathbb{R}^{d\times N}$.

Im is nonempty and closed. As $\boldsymbol{0} \in \mathrm{Im}$, the set Im is nonempty. In addition, Im is closed, i.e., the following condition holds:

$$\boldsymbol{Z}_j \in \mathrm{Im} \text{ for } j \in \mathbb{N},\ \boldsymbol{Z}_j \to \boldsymbol{Z} \text{ in } \mathbb{R}^{d\times N} \text{ as } j \to \infty \implies \boldsymbol{Z} \in \mathrm{Im}. \tag{3.3}$$

Indeed, since $\boldsymbol{Z}_j \in \mathrm{Im}$ there exist $(\boldsymbol{J}_j, \boldsymbol{D}_j) \in \mathcal{A}$ such that $\boldsymbol{Z}_j = \boldsymbol{J}_j + \boldsymbol{D}_j$ for every j and since $\{\boldsymbol{Z}_j\}$ is bounded, it follows from (A2) and (A1) that

$$|\boldsymbol{J}_j|^2 + |\boldsymbol{D}_j|^2 \le |\boldsymbol{J}_j + \boldsymbol{D}_j|^2 = |\boldsymbol{Z}_j|^2 \le C.$$

Thus, the sequences $\{\boldsymbol{J}_j\}$ and $\{\boldsymbol{D}_j\}$ are bounded and there exists a couple $(\boldsymbol{J}, \boldsymbol{D})$ such that for a subsequence (that we do not relabel) $\boldsymbol{J}_j \to \boldsymbol{J}$ and $\boldsymbol{D}_j \to \boldsymbol{D}$ as $j \to \infty$. As $\boldsymbol{Z}_j \to \boldsymbol{Z}$, we conclude that $\boldsymbol{Z} = \boldsymbol{J} + \boldsymbol{D}$. Due to (A2) and the fact that $(\boldsymbol{J}_j, \boldsymbol{D}_j) \in \mathcal{A}$ we also get

$$(\boldsymbol{J} - \boldsymbol{A}) : (\boldsymbol{D} - \boldsymbol{B}) = \lim_{j\to\infty} (\boldsymbol{J}_j - \boldsymbol{A}) : (\boldsymbol{D}_j - \boldsymbol{B}) \ge 0 \qquad \text{for all } (\boldsymbol{A}, \boldsymbol{B}) \in \mathcal{A}.$$

Then, in virtue of the maximality (A3), $(\boldsymbol{J}, \boldsymbol{D}) \in \mathcal{A}$. Hence, $\boldsymbol{Z} \in \mathrm{Im}$.

Definition of the mapping $(\boldsymbol{J}^* + \boldsymbol{I})^{-1}$. On Im, we define

$$(\boldsymbol{J}^* + \boldsymbol{I})^{-1}(\boldsymbol{Z}) := \{\boldsymbol{D} \in \mathbb{R}^{d\times N};\ \exists \boldsymbol{J},\ (\boldsymbol{J}, \boldsymbol{D}) \in \mathcal{A},\ \boldsymbol{J} + \boldsymbol{D} = \boldsymbol{Z}\} \tag{3.4}$$

and show in the following lines that $(\boldsymbol{J}^* + \boldsymbol{I})^{-1}$ is well-defined single-valued mapping and $\mathrm{Im} = \mathbb{R}^{d\times N}$.

We first check that $(\boldsymbol{J}^* + \boldsymbol{I})^{-1}$ is 1-Lipschitz on Im. Towards this goal, let us take $\boldsymbol{Z}_1, \boldsymbol{Z}_2 \in \mathrm{Im}$, $\boldsymbol{Z}_1 \ne \boldsymbol{Z}_2$, and $\boldsymbol{D}_1 \in (\boldsymbol{J}^* + \boldsymbol{I})^{-1}(\boldsymbol{Z}_1)$, $\boldsymbol{D}_2 \in (\boldsymbol{J}^* + \boldsymbol{I})^{-1}(\boldsymbol{Z}_2)$. Then, with help of (A2), we have

$$(\boldsymbol{Z}_1 - \boldsymbol{Z}_2) : (\boldsymbol{D}_1 - \boldsymbol{D}_2) = (\boldsymbol{J}_1 - \boldsymbol{J}_2 + \boldsymbol{D}_1 - \boldsymbol{D}_2) : (\boldsymbol{D}_1 - \boldsymbol{D}_2) \ge |\boldsymbol{D}_1 - \boldsymbol{D}_2|^2.$$

Then, 1-Lipschitz continuity directly follows from the Cauchy-Schwarz inequality.

$\underline{\mathrm{Im} = \mathbb{R}^{d\times N}}$. Next, for contradiction, assume that $\mathrm{Im} \subsetneq \mathbb{R}^{d\times N}$. We then define an auxiliary function

$$\boldsymbol{F}(\boldsymbol{Z}) := \boldsymbol{Z} - \sqrt{2}(\boldsymbol{J}^* + \boldsymbol{I})^{-1}(\sqrt{2}\boldsymbol{Z}), \qquad \boldsymbol{Z} \in \frac{1}{\sqrt{2}}\mathrm{Im} \tag{3.5}$$

and observe that $\boldsymbol{F}$ is also 1-Lipschitz on $\frac{1}{\sqrt{2}}\mathrm{Im}$. Indeed, let $\boldsymbol{Z}_1, \boldsymbol{Z}_2 \in \frac{1}{\sqrt{2}}\mathrm{Im}$, $\boldsymbol{Z}_1 \neq \boldsymbol{Z}_2$ and consider $\boldsymbol{J}_i, \boldsymbol{D}_i$, $i = 1, 2$, such that $\boldsymbol{J}_1 + \boldsymbol{D}_1 = \sqrt{2}\boldsymbol{Z}_1$, $\boldsymbol{J}_2 + \boldsymbol{D}_2 = \sqrt{2}\boldsymbol{Z}_2$. Then with the help of (3.4) and (A2) we observe that

$$\begin{aligned}|\boldsymbol{F}(\boldsymbol{Z}_1) - \boldsymbol{F}(\boldsymbol{Z}_2)|^2 &= |\boldsymbol{Z}_1 - \boldsymbol{Z}_2|^2 + 2|\boldsymbol{D}_1 - \boldsymbol{D}_2|^2 - 2\sqrt{2}(\boldsymbol{Z}_1 - \boldsymbol{Z}_2) : (\boldsymbol{D}_1 - \boldsymbol{D}_2)\\ &= |\boldsymbol{Z}_1 - \boldsymbol{Z}_2|^2 - 2(\boldsymbol{J}_1 - \boldsymbol{J}_2) : (\boldsymbol{D}_1 - \boldsymbol{D}_2) \le |\boldsymbol{Z}_1 - \boldsymbol{Z}_2|^2.\end{aligned}$$

As we assume that $\frac{1}{\sqrt{2}}\mathrm{Im}$ is proper subset of $\mathbb{R}^{d\times N}$, i.e., $\frac{1}{\sqrt{2}}\mathrm{Im} \subsetneq \mathbb{R}^{d\times N}$, we can extend $\boldsymbol{F}$ defined on $\frac{1}{\sqrt{2}}\mathrm{Im}$ to $\tilde{\boldsymbol{F}}$ defined on $\mathbb{R}^{d\times N}$ in such a way that

$$\tilde{\boldsymbol{F}}(\boldsymbol{Z}) = \begin{cases}\boldsymbol{F}(\boldsymbol{Z}) & \boldsymbol{Z} \in \dfrac{1}{\sqrt{2}}\mathrm{Im},\\ \text{is 1-Lipschitz} & \text{on } \mathbb{R}^{d\times N}.\end{cases}$$

Let us now define, for an arbitrary $\sqrt{2}\tilde{\boldsymbol{Z}} \in \mathbb{R}^{d\times N} \setminus \mathrm{Im}$,

$$\tilde{\boldsymbol{J}} := \frac{1}{\sqrt{2}}(\tilde{\boldsymbol{Z}} + \tilde{\boldsymbol{F}}(\tilde{\boldsymbol{Z}})) \quad \text{and} \quad \tilde{\boldsymbol{D}} := \frac{1}{\sqrt{2}}(\tilde{\boldsymbol{Z}} - \tilde{\boldsymbol{F}}(\tilde{\boldsymbol{Z}})). \tag{3.6}$$

If we prove that

$$(\tilde{\boldsymbol{J}}, \tilde{\boldsymbol{D}}) \in \mathcal{A}, \tag{3.7}$$

then $\tilde{\boldsymbol{J}} + \tilde{\boldsymbol{D}} = \sqrt{2}\tilde{\boldsymbol{Z}}$ and, due to the definition of Im, $\sqrt{2}\tilde{\boldsymbol{Z}} \in \mathrm{Im}$, which is a sought contradiction. Thus, the definition domain of $(\boldsymbol{J}^* + \boldsymbol{I})^{-1}$ is $\mathbb{R}^{d\times N}$. It means that $\boldsymbol{J}^* + \boldsymbol{I}$ is onto.

It remains to verify (3.7). For this purpose, we use the maximality of $\mathcal{A}$, i.e., the assumption (A3), and show that for all $(\boldsymbol{J}, \boldsymbol{D}) \in \mathcal{A}$ the following holds:

$$(\tilde{\boldsymbol{J}} - \boldsymbol{J}) : (\tilde{\boldsymbol{D}} - \boldsymbol{D}) \ge 0. \tag{3.8}$$

Taking an arbitrary $(\boldsymbol{J}, \boldsymbol{D}) \in \mathcal{A}$ and setting $\sqrt{2}\boldsymbol{Z} := \boldsymbol{J} + \boldsymbol{D}$, we observe that $\sqrt{2}\boldsymbol{Z} \in \mathrm{Im}$. Then, in virtue of the definition $\boldsymbol{F}$ (see (3.5)), we have

$$\boldsymbol{J} = \frac{1}{\sqrt{2}}(\boldsymbol{Z} + \boldsymbol{F}(\boldsymbol{Z})) \quad \text{and} \quad \boldsymbol{D} = \frac{1}{\sqrt{2}}(\boldsymbol{Z} - \boldsymbol{F}(\boldsymbol{Z})). \tag{3.9}$$

Using then (3.6), (3.9) and the fact that $\tilde{\boldsymbol{F}}$ is 1-Lipschitz continuous on $\mathbb{R}^{d\times N}$, we obtain

$$\begin{aligned}2(\tilde{\boldsymbol{J}} - \boldsymbol{J}) : (\tilde{\boldsymbol{D}} - \boldsymbol{D}) &= (\tilde{\boldsymbol{Z}} - \boldsymbol{Z} + \tilde{\boldsymbol{F}}(\tilde{\boldsymbol{Z}}) - \boldsymbol{F}(\boldsymbol{Z})) : (\tilde{\boldsymbol{Z}} - \boldsymbol{Z} - (\tilde{\boldsymbol{F}}(\tilde{\boldsymbol{Z}}) - \boldsymbol{F}(\boldsymbol{Z})))\\ &= |\tilde{\boldsymbol{Z}} - \boldsymbol{Z}|^2 - |\tilde{\boldsymbol{F}}(\tilde{\boldsymbol{Z}}) - \boldsymbol{F}(\boldsymbol{Z})|^2 \ge 0,\end{aligned}$$

which proves (3.8). Thus, the proof of (3.7) is complete.

Step 2. It remains to prove the second implication, i.e., if $\boldsymbol{J}^* + \boldsymbol{I}$ is onto, then the monotone graph $\mathcal{A}$ is maximal, i.e., (A3) holds. Let $(\tilde{\boldsymbol{J}}, \tilde{\boldsymbol{D}}) \in \mathbb{R}^{d\times N} \times \mathbb{R}^{d\times N}$ be such that for all $(\boldsymbol{J}, \boldsymbol{D}) \in \mathcal{A}$ (3.8) holds. Our goal is to show that $(\tilde{\boldsymbol{J}}, \tilde{\boldsymbol{D}}) \in \mathcal{A}$. Since $\boldsymbol{J}^* + \boldsymbol{I}$ is onto, we know that for $(\tilde{\boldsymbol{J}}, \tilde{\boldsymbol{D}})$ there exists a couple $(\boldsymbol{J}, \boldsymbol{D}) \in \mathcal{A}$ such that

$$(\boldsymbol{J}^* + \boldsymbol{I})(\boldsymbol{D}) = \boldsymbol{J} + \boldsymbol{D} = \tilde{\boldsymbol{J}} + \tilde{\boldsymbol{D}}. \tag{3.10}$$

Now, since $(\boldsymbol{J}, \boldsymbol{D}) \in \mathcal{A}$, we use the decomposition (3.10) in (3.8) and deduce that

$$\begin{aligned}0 &\le 2(\tilde{\boldsymbol{J}} - \boldsymbol{J}) : (\tilde{\boldsymbol{D}} - \boldsymbol{D}) = (\tilde{\boldsymbol{J}} - (\tilde{\boldsymbol{J}} + \tilde{\boldsymbol{D}} - \boldsymbol{D})) : (\tilde{\boldsymbol{D}} - \boldsymbol{D}) + (\tilde{\boldsymbol{J}} - \boldsymbol{J}) : (\tilde{\boldsymbol{D}} - (\tilde{\boldsymbol{J}} + \tilde{\boldsymbol{D}} - \boldsymbol{J}))\\ &= -|\tilde{\boldsymbol{D}} - \boldsymbol{D}|^2 - |\tilde{\boldsymbol{J}} - \boldsymbol{J}|^2.\end{aligned}$$

Consequently, $\tilde{\boldsymbol{D}} = \boldsymbol{D}$ and $\tilde{\boldsymbol{J}} = \boldsymbol{J}$ and $(\tilde{\boldsymbol{J}}, \tilde{\boldsymbol{D}}) \in \mathcal{A}$, which finishes the proof. □

Lemma 3.4. *Let* $\boldsymbol{G}$ *satisfy assumptions (G1)–(G4). Let*

$$\mathcal{A} := \{(\boldsymbol{J}, \boldsymbol{D}) \in \mathbb{R}^{d\times N} \times \mathbb{R}^{d\times N} : \ \boldsymbol{G}(\boldsymbol{J}, \boldsymbol{D}) = \boldsymbol{0}\}. \tag{3.11}$$

Then $\mathcal{A}$ *is a maximal monotone* p*-coercive graph.*

Here, we would like to emphasize that the assumptions (G1)–(G4) are associated with the implicit constitutive equation $\boldsymbol{G}(\boldsymbol{J}, \boldsymbol{D}) = \boldsymbol{0}$ and it is not evident a priori that the null set of $\boldsymbol{G}$ is a maximal monotone p-coercive graph. However, we will show below that the monotone condition (A2) is in fact a consequence of (G2) while the maximality (A3) follows from the continuity of $\boldsymbol{G}$ and (G3).

Proof of Lemma 3.4. We start the proof with several simple observations. Recalling the definition of $\mathcal{A}$ given in (3.11), we see directly that (G1) $\Longrightarrow$ (A1) and (G4) $\Longleftrightarrow$ (A4). Thus, it remains to verify the conditions (A2) and (A3). We show that they follow from (G2), (G3) and the continuity of $\boldsymbol{G}$. Without loss of generality, we assume here that the second condition in (G3) is fulfilled. In case that the first condition of (G3) was true, we would have to change the role of $\boldsymbol{J}$ and $\boldsymbol{D}$ in the proof below. We split the proof into several steps.
Step 1. We first show that for every $\varepsilon \in (0, 1]$ and every $\boldsymbol{Z} \in \mathbb{R}^{d\times N}$ there exists $\boldsymbol{D} \in \mathbb{R}^{d\times N}$ such that

$$\boldsymbol{G}(\boldsymbol{Z} - \varepsilon\boldsymbol{D}, \boldsymbol{D}) = \boldsymbol{0}. \tag{3.12}$$

Once (3.12) is proved and once we show that the graph $\mathcal{A}$ is monotone (which we shall prove in Step 4 below), then (3.12) implies that $\mathcal{A}$ is maximal by means of Lemma 3.3. Indeed, we need to check that $\boldsymbol{J}^* + \varepsilon\boldsymbol{I}$ is onto, i.e., that for every $\boldsymbol{Z} \in \mathbb{R}^{d\times N}$ there exists a couple $(\boldsymbol{J}, \boldsymbol{D}) \in \mathbb{R}^{d\times N} \times \mathbb{R}^{d\times N}$ such that

$$\boldsymbol{G}(\boldsymbol{J}, \boldsymbol{D}) = \boldsymbol{0} \qquad \text{and} \qquad \boldsymbol{J} + \varepsilon\boldsymbol{D} = \boldsymbol{Z}.$$

However, substituting the second relation into the first one, we observe that it is exactly (3.12).
To summarize, once we verify (3.12) and show that $\mathcal{A}$ is monotone, the proof of Lemma 3.4 is complete.
Proof of (3.12). For arbitrary $\boldsymbol{Z}_1, \ \boldsymbol{Z}_2 \in \mathbb{R}^{d\times N}$, set $\boldsymbol{Z}_t := t\boldsymbol{Z}_1 + (1 - t)\boldsymbol{Z}_2$. Then, in virtue of (G2), $\boldsymbol{G}_{\boldsymbol{J}}(\boldsymbol{Z}_t, \boldsymbol{D}) \geq 0$ for all $\boldsymbol{D} \in \mathbb{R}^{d\times N}$. Consequently

$$\begin{aligned}(\boldsymbol{G}(\boldsymbol{Z}_1, \boldsymbol{D}) - \boldsymbol{G}(\boldsymbol{Z}_2, \boldsymbol{D})) : (\boldsymbol{Z}_1 - \boldsymbol{Z}_2) &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{G}(\boldsymbol{Z}_t, \boldsymbol{D}) : (\boldsymbol{Z}_1 - \boldsymbol{Z}_2)\,\mathrm{d}t \\ &= \int_0^1 \boldsymbol{G}_{\boldsymbol{J}}(\boldsymbol{Z}_t, \boldsymbol{D})(\boldsymbol{Z}_1 - \boldsymbol{Z}_2) : (\boldsymbol{Z}_1 - \boldsymbol{Z}_2)\,\mathrm{d}t \geq 0.\end{aligned}$$

Taking, in particular, $\boldsymbol{Z}_1 = \boldsymbol{Z} - \varepsilon\boldsymbol{D}$ and $\boldsymbol{Z}_2 = \boldsymbol{Z}$, where $\boldsymbol{Z}$ and $\boldsymbol{D} \in \mathbb{R}^{d\times N}$ are arbitrary, it implies that

$$-\varepsilon(\boldsymbol{G}(\boldsymbol{Z} - \varepsilon\boldsymbol{D}, \boldsymbol{D}) - \boldsymbol{G}(\boldsymbol{Z}, \boldsymbol{D})) : \boldsymbol{D} \geq 0. \tag{3.13}$$

Using then (G3), we observe that for arbitrary $\varepsilon \in (0, 1]$ and any $\boldsymbol{Z} \in \mathbb{R}^{d\times N}$

$$\limsup_{|\boldsymbol{D}|\to\infty} \boldsymbol{G}(\boldsymbol{Z} - \varepsilon\boldsymbol{D}, \boldsymbol{D}) : \boldsymbol{D} \overset{(3.13)}{\leq} \limsup_{|\boldsymbol{D}|\to\infty} \boldsymbol{G}(\boldsymbol{Z}, \boldsymbol{D}) : \boldsymbol{D} < 0. \tag{3.14}$$

Thus, there exists $R > 0$ such that for all $\boldsymbol{D} \in \mathbb{R}^{d\times N}$ fulfilling $|\boldsymbol{D}| \geq R$, we have

$$\boldsymbol{G}(\boldsymbol{Z} - \varepsilon\boldsymbol{D}, \boldsymbol{D}) : \boldsymbol{D} \leq 0. \tag{3.15}$$

Having this piece of information, we prove (3.12) by contradiction. We thus assume that for all $\boldsymbol{D} \in \mathbb{R}^{d\times N}$ such that $|\boldsymbol{D}| \le R$ it holds:

$$\boldsymbol{G}(\boldsymbol{Z} - \varepsilon\boldsymbol{D}, \boldsymbol{D}) \neq \boldsymbol{0}. \tag{3.16}$$

Then, due to (3.16) and the continuity of $\boldsymbol{G}$, the mapping

$$\boldsymbol{D} \mapsto R\frac{\boldsymbol{G}(\boldsymbol{Z} - \varepsilon\boldsymbol{D}, \boldsymbol{D})}{|\boldsymbol{G}(\boldsymbol{Z} - \varepsilon\boldsymbol{D}, \boldsymbol{D})|}$$

is defined in a closed ball of radius R, is continuous and maps a closed ball of radius R into itself (in fact it maps the ball of radius R onto its sphere). Consequently, by Browder fixed point theorem, there is $\boldsymbol{D} \in \mathbb{R}^{d\times N}$, $|\boldsymbol{D}| = R$, such that

$$\boldsymbol{D} = R\frac{\boldsymbol{G}(\boldsymbol{Z} - \varepsilon\boldsymbol{D}, \boldsymbol{D})}{|\boldsymbol{G}(\boldsymbol{Z} - \varepsilon\boldsymbol{D}, \boldsymbol{D})|}.$$

Taking the scalar product of both sides of this equality with $\boldsymbol{D}$ and using (3.15), we see that

$$R^2 = \boldsymbol{D} : \boldsymbol{D} = R\frac{\boldsymbol{G}(\boldsymbol{Z} - \varepsilon\boldsymbol{D}, \boldsymbol{D}) : \boldsymbol{D}}{|\boldsymbol{G}(\boldsymbol{Z} - \varepsilon\boldsymbol{D}, \boldsymbol{D})|} \le 0,$$

a contradiction. Hence, for an arbitrarily given $\boldsymbol{Z} \in \mathbb{R}^{d\times N}$ there is $\boldsymbol{D}$ satisfying (3.12).

Step 2. Next we show that for any couple $(\boldsymbol{Z}_1, \boldsymbol{D}_1), (\boldsymbol{Z}_2, \boldsymbol{D}_2) \in \mathbb{R}^{d\times N} \times \mathbb{R}^{d\times N}$ fulfilling, for $i = 1, 2$, $\boldsymbol{G}(\boldsymbol{Z}_i - \varepsilon\boldsymbol{D}_i, \boldsymbol{D}_i) = \boldsymbol{0}$, the following condition holds:

$$|\boldsymbol{D}_1 - \boldsymbol{D}_2| \le C(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{D}_1, \boldsymbol{D}_2)|\boldsymbol{Z}_1 - \boldsymbol{Z}_2|. \tag{3.17}$$

It means that $\boldsymbol{D}$ can be understood as a locally Lipschitz function of $\boldsymbol{Z}$.

Proof of (3.17). Let us denote $\boldsymbol{Z}_t := t\boldsymbol{Z}_1 + (1-t)\boldsymbol{Z}_2$ and $\boldsymbol{D}_t := t\boldsymbol{D}_1 + (1-t)\boldsymbol{D}_2$. Then it follows from the assumption (G2) that

$$\begin{aligned}
\boldsymbol{0} &= \boldsymbol{G}(\boldsymbol{Z}_1 - \varepsilon\boldsymbol{D}_1, \boldsymbol{D}_1) - \boldsymbol{G}(\boldsymbol{Z}_2 - \varepsilon\boldsymbol{D}_2, \boldsymbol{D}_2) \\
&= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{G}(\boldsymbol{Z}_t - \varepsilon\boldsymbol{D}_t, \boldsymbol{D}_t)\,\mathrm{d}t \\
&= \int_0^1 \boldsymbol{G}_{\boldsymbol{J}}(\boldsymbol{Z}_t - \varepsilon\boldsymbol{D}_t, \boldsymbol{D}_t)(\boldsymbol{Z}_1 - \boldsymbol{Z}_2 - \varepsilon(\boldsymbol{D}_1 - \boldsymbol{D}_2)) + \boldsymbol{G}_{\boldsymbol{D}}(\boldsymbol{Z}_t - \varepsilon\boldsymbol{D}_t, \boldsymbol{D}_t)(\boldsymbol{D}_1 - \boldsymbol{D}_2)\,\mathrm{d}t.
\end{aligned}$$

Consequently,

$$\int_0^1 \varepsilon\boldsymbol{G}_{\boldsymbol{J}}(\dots) - \boldsymbol{G}_{\boldsymbol{D}}(\dots)\,\mathrm{d}t\,(\boldsymbol{D}_1 - \boldsymbol{D}_2) = \int_0^1 \boldsymbol{G}_{\boldsymbol{J}}(\dots)\,\mathrm{d}t\,(\boldsymbol{Z}_1 - \boldsymbol{Z}_2), \tag{3.18}$$

where $(\dots)$ stands for $(\boldsymbol{Z}_t - \varepsilon\boldsymbol{D}_t, \boldsymbol{D}_t)$. Thanks to (G2), we know that $\boldsymbol{G}_{\boldsymbol{J}} \ge 0$, $-\boldsymbol{G}_{\boldsymbol{D}} \ge 0$ and $\boldsymbol{G}_{\boldsymbol{J}} - \boldsymbol{G}_{\boldsymbol{D}} > 0$. Consequently, for arbitrary $\varepsilon > 0$ we also have $\varepsilon\boldsymbol{G}_{\boldsymbol{J}} - \boldsymbol{G}_{\boldsymbol{D}} > 0$, and also

$$I := \int_0^1 \varepsilon\boldsymbol{G}_{\boldsymbol{J}}(\dots) - \boldsymbol{G}_{\boldsymbol{D}}(\dots)\,\mathrm{d}t > 0.$$

It means that I is positive definite, and consequently, I is an invertible matrix. It thus follows from (3.18) that

$$\boldsymbol{D}_1 - \boldsymbol{D}_2 = \left(\int_0^1 \varepsilon\boldsymbol{G}_{\boldsymbol{J}}(\dots) - \boldsymbol{G}_{\boldsymbol{D}}(\dots)\,\mathrm{d}t\right)^{-1} \int_0^1 \boldsymbol{G}_{\boldsymbol{J}}(\dots)\,\mathrm{d}t\,(\boldsymbol{Z}_1 - \boldsymbol{Z}_2)$$

and (3.17) follows.

Step 3. In this step, we show that for arbitrary $(\boldsymbol{Z}_1, \boldsymbol{D}_1), (\boldsymbol{Z}_2, \boldsymbol{D}_2) \in \mathbb{R}^{d\times N} \times \mathbb{R}^{d\times N}$ fulfilling, for $i = 1, 2$, $\boldsymbol{G}(\boldsymbol{Z}_i - \varepsilon \boldsymbol{D}_i, \boldsymbol{D}_i) = \boldsymbol{0}$, there holds

$$(\boldsymbol{D}_1 - \boldsymbol{D}_2) : (\boldsymbol{Z}_1 - \boldsymbol{Z}_2) \geq 0. \tag{3.19}$$

Proof of (3.19). It follows from Step 2, that for the null points of $\boldsymbol{G}(\boldsymbol{Z} - \varepsilon\boldsymbol{D}, \boldsymbol{D})$, we can understand $\boldsymbol{D}$ as a locally Lipschitz mapping of $\boldsymbol{Z}$ and we can write $\boldsymbol{D}(\boldsymbol{Z})$. Since $\boldsymbol{D}$ is Lipschitz, its derivative $\boldsymbol{D}_{\boldsymbol{Z}}(\boldsymbol{Z})$ exists for almost all $\boldsymbol{Z}$. By applying this derivation to $\boldsymbol{G}(\boldsymbol{Z}-\varepsilon\boldsymbol{D}(\boldsymbol{Z}), \boldsymbol{D}(\boldsymbol{Z})) = \boldsymbol{0}$, we obtain

$$\begin{aligned}\boldsymbol{0} &= \frac{\partial}{\partial \boldsymbol{Z}} \boldsymbol{G}(\boldsymbol{Z} - \varepsilon\boldsymbol{D}(\boldsymbol{Z}), \boldsymbol{D}(\boldsymbol{Z})) \\ &= \boldsymbol{G}_{\boldsymbol{J}}(\boldsymbol{Z} - \varepsilon\boldsymbol{D}(\boldsymbol{Z}), \boldsymbol{D}(\boldsymbol{Z}))(\boldsymbol{I} - \varepsilon\boldsymbol{D}_{\boldsymbol{Z}}(\boldsymbol{Z})) + \boldsymbol{G}_{\boldsymbol{D}}((\boldsymbol{Z} - \varepsilon\boldsymbol{D}(\boldsymbol{Z}), \boldsymbol{D}(\boldsymbol{Z}))\boldsymbol{D}_{\boldsymbol{Z}}(\boldsymbol{Z})\end{aligned}$$

It implies that

$$(\varepsilon\boldsymbol{G}_{\boldsymbol{J}}(\boldsymbol{Z} - \varepsilon\boldsymbol{D}(\boldsymbol{Z}), \boldsymbol{D}(\boldsymbol{Z})) - \boldsymbol{G}_{\boldsymbol{D}}(\boldsymbol{Z} - \varepsilon\boldsymbol{D}(\boldsymbol{Z}), \boldsymbol{D}(\boldsymbol{Z})))\boldsymbol{D}_{\boldsymbol{Z}}(\boldsymbol{Z}) = \boldsymbol{G}_{\boldsymbol{J}}(\boldsymbol{Z} - \varepsilon\boldsymbol{D}(\boldsymbol{Z}), \boldsymbol{D}(\boldsymbol{Z})).$$

Since the matrix on the left-hand side is regular thanks to the assumption (G2), we observe (we omit writing the dependence on $\boldsymbol{Z}$ for simplicity) that

$$\boldsymbol{D}_{\boldsymbol{Z}} = (\varepsilon\boldsymbol{G}_{\boldsymbol{J}} - \boldsymbol{G}_{\boldsymbol{D}})^{-1}\boldsymbol{G}_{\boldsymbol{J}}. \tag{3.20}$$

Our next goal is to show that

$$\boldsymbol{D}_{\boldsymbol{Z}} \geq 0. \tag{3.21}$$

To do so, consider arbitrary nonzero $\boldsymbol{X} \in \mathbb{R}^{d\times N}$. Since $(\varepsilon\boldsymbol{G}_{\boldsymbol{J}} - \boldsymbol{G}_{\boldsymbol{D}})$ is invertible, we can also define $\boldsymbol{Y} := (\varepsilon\boldsymbol{G}_{\boldsymbol{J}} - \boldsymbol{G}_{\boldsymbol{D}})^{-T}\boldsymbol{X}$ and with the help of (3.20) obtain

$$\begin{aligned}\boldsymbol{D}_{\boldsymbol{Z}}\boldsymbol{X} : \boldsymbol{X} &= (\varepsilon\boldsymbol{G}_{\boldsymbol{J}} - \boldsymbol{G}_{\boldsymbol{D}})^{-1}\boldsymbol{G}_{\boldsymbol{J}}\boldsymbol{X} : \boldsymbol{X} = (\boldsymbol{G}_{\boldsymbol{J}}\boldsymbol{X}) : ((\varepsilon\boldsymbol{G}_{\boldsymbol{J}} - \boldsymbol{G}_{\boldsymbol{D}})^{-T}\boldsymbol{X}) \\ &= \boldsymbol{G}_{\boldsymbol{J}}(\varepsilon\boldsymbol{G}_{\boldsymbol{J}} - \boldsymbol{G}_{\boldsymbol{D}})^{T}\boldsymbol{Y} : \boldsymbol{Y} = ((\varepsilon\boldsymbol{G}_{\boldsymbol{J}} - \boldsymbol{G}_{\boldsymbol{D}})^{T}\boldsymbol{Y}) : ((\boldsymbol{G}_{\boldsymbol{J}})^{T}\boldsymbol{Y}) \\ &= \varepsilon|(\boldsymbol{G}_{\boldsymbol{J}})^{T}\boldsymbol{Y}|^2 - \boldsymbol{G}_{\boldsymbol{D}}(\boldsymbol{G}_{\boldsymbol{J}})^{T}\boldsymbol{Y} : \boldsymbol{Y} \geq 0,\end{aligned}$$

where the last inequality follows from the third assumption in (G2). Finally, since

$$\boldsymbol{D}(\boldsymbol{Z}_1) - \boldsymbol{D}(\boldsymbol{Z}_2) = \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{D}(\boldsymbol{Z}_t)\,\mathrm{d}t = \int_0^1 \boldsymbol{D}_{\boldsymbol{Z}}(\boldsymbol{Z}_t)\,\mathrm{d}t\,(\boldsymbol{Z}_1 - \boldsymbol{Z}_2)$$

with $\boldsymbol{Z}_t := t\boldsymbol{Z}_1 + (1-t)\boldsymbol{Z}_2$, we can use (3.21) to deduce that

$$(\boldsymbol{D}(\boldsymbol{Z}_1) - \boldsymbol{D}(\boldsymbol{Z}_2)) : (\boldsymbol{Z}_1 - \boldsymbol{Z}_2) = \int_0^1 \boldsymbol{D}_{\boldsymbol{Z}}(\boldsymbol{Z}_t)\,\mathrm{d}t(\boldsymbol{Z}_1 - \boldsymbol{Z}_2) : (\boldsymbol{Z}_1 - \boldsymbol{Z}_2) \geq 0$$

and (3.19) follows.

Step 4. Finally, it remains to verify that $\mathcal{A}$ is monotone, i.e., (A2) holds. Let $\boldsymbol{J}_1$, $\boldsymbol{J}_2$, $\boldsymbol{D}_1$ and $\boldsymbol{D}_2 \in \mathbb{R}^{d\times N}$ fulfilling, for $i = 1, 2$, $\boldsymbol{G}(\boldsymbol{J}_i, \boldsymbol{D}_i) = \boldsymbol{0}$ be arbitrary. The aim is to show that

$$(\boldsymbol{J}_1 - \boldsymbol{J}_2) : (\boldsymbol{D}_1 - \boldsymbol{D}_2) \geq 0. \tag{3.22}$$

To prove it, we define $\boldsymbol{Z}_i := \boldsymbol{J}_i + \varepsilon\boldsymbol{D}_i$, $i = 1, 2$. Then it follows from the assumptions on $(\boldsymbol{J}_i, \boldsymbol{D}_i)$ that $\boldsymbol{G}(\boldsymbol{Z}_i - \varepsilon\boldsymbol{D}_i, \boldsymbol{D}_i) = \boldsymbol{0}$. Hence the inequality (3.19) from Step 3 is valid for $(\boldsymbol{Z}_1, \boldsymbol{D}_1)$ and $(\boldsymbol{Z}_2, \boldsymbol{D}_2)$. Thus, we can continue as follows

$$\begin{aligned}(\boldsymbol{J}_1 - \boldsymbol{J}_2) : (\boldsymbol{D}_1 - \boldsymbol{D}_2) &= (\boldsymbol{Z}_1 - \boldsymbol{Z}_2 - \varepsilon(\boldsymbol{D}_1 - \boldsymbol{D}_2)) : (\boldsymbol{D}_1 - \boldsymbol{D}_2) \\ &= (\boldsymbol{Z}_1 - \boldsymbol{Z}_2) : (\boldsymbol{D}_1 - \boldsymbol{D}_2) - \varepsilon|\boldsymbol{D}_1 - \boldsymbol{D}_2|^2 \geq -\varepsilon|\boldsymbol{D}_1 - \boldsymbol{D}_2|^2,\end{aligned}$$

where the last inequality follows from (3.19). Since the left-hand side is independent of ε, letting $\varepsilon \to 0_+$, we deduce (3.22). The proof of Lemma 3.4 is complete. □

4. Algebraic ε-approximations of the graph $\mathcal{A}$

In this section, we construct two different suitable ε-approximations of the maximal monotone p-coercive graph and show that these approximate graphs are Lipschitz continuous and uniformly monotone 2-coercive graphs. Another advantage of these approximate graphs comes from their algebraic construction that is easy to incorporate into numerical schemes and their implementation. Finally, we study the convergence properties. In fact, the first approximation in Definition 4.1 starts with the notion of maximal monotone p-coercive graph, while the second approximation in Definition 4.3 is directly linked to the implicit constitutive equation $\boldsymbol{G}(\boldsymbol{J},\boldsymbol{D})=\boldsymbol{0}$ with $\boldsymbol{G}$ fulfilling (G1)-(G4).

Definition 4.1 (Construction of the approximate graphs)**.** *Let $\mathcal{A}$ be a maximal monotone p-coercive graph, see Definition 3.1, and let $\varepsilon>0$. We define*

$$\mathcal{A}_\varepsilon := \{(\tilde{\boldsymbol{J}},\tilde{\boldsymbol{D}}) \in \mathbb{R}^{d\times N}\times\mathbb{R}^{d\times N}; \exists(\overline{\boldsymbol{J}},\overline{\boldsymbol{D}})\in\mathcal{A}, \tilde{\boldsymbol{J}}=\overline{\boldsymbol{J}}, \tilde{\boldsymbol{D}}=\overline{\boldsymbol{D}}+\varepsilon\overline{\boldsymbol{J}}\}, \tag{4.1a}$$

$$\mathcal{A}^\varepsilon_\varepsilon := \{(\boldsymbol{J},\boldsymbol{D}) \in \mathbb{R}^{d\times N}\times\mathbb{R}^{d\times N}; \exists(\tilde{\boldsymbol{J}},\tilde{\boldsymbol{D}})\in\mathcal{A}_\varepsilon, \boldsymbol{J}=\tilde{\boldsymbol{J}}+\varepsilon\tilde{\boldsymbol{D}}, \boldsymbol{D}=\tilde{\boldsymbol{D}}\}. \tag{4.1b}$$

Remark 4.2. *There is no apparent reason for the lower and the upper index in the definition of the graph $\mathcal{A}^\varepsilon_\varepsilon$ to be the same. However, making them different (e.g., $\mathcal{A}^e_\varepsilon$ for $\varepsilon, e>0$) would not bring any analytical advantage, generality, or simplicity.*

Definition 4.3 (Construction of the approximative constitutive equations)**.** *Let $\boldsymbol{G}$ satisfy (G1)–(G4) and let $\varepsilon>0$. We set $\boldsymbol{G}_\varepsilon(\boldsymbol{J},\boldsymbol{D}):=\boldsymbol{G}(\boldsymbol{J}-\varepsilon\boldsymbol{D},\boldsymbol{D}-\varepsilon\boldsymbol{J})$.*

Remark 4.4. *Instead of null points of $\boldsymbol{G}_\varepsilon$, one could also use an alternative approximation $\tilde{\boldsymbol{G}}_\varepsilon(\boldsymbol{J},\boldsymbol{D}):=\boldsymbol{G}(\boldsymbol{J},\boldsymbol{D})\pm\varepsilon(\boldsymbol{J}-\boldsymbol{D})$ where positive sign is used if $\boldsymbol{G}_{\boldsymbol{J}}(\boldsymbol{J},\boldsymbol{D})\geq 0$ in (G2), while negative sign would be associated if all inequality signs (besides the last one in (G2)) are opposite. Such an approximation then also leads to the strictly monotone and locally Lipschitz graphs but we cannot guarantee* 2*-coercivity and therefore the Hilbert structure of the approximative problem is lost. On the other hand, $\tilde{\boldsymbol{G}}_\varepsilon$ seems to be the easiest way how to approximate the constitutive equation $\boldsymbol{G}(\boldsymbol{J},\boldsymbol{D})=\boldsymbol{0}$.*

For the approximations introduced in Definitions 4.1 and 4.3 above, we establish the following results playing a key role in the subsequent analysis developed in this paper.

Lemma 4.5. *Let $\mathcal{A}$ be a maximal monotone p-coercive graph. Then for every $\varepsilon\in(0,1)$, $\mathcal{A}^\varepsilon_\varepsilon$ is a maximal monotone* 2*-coercive graph. Moreover, there exists a unique single-valued mapping $\boldsymbol{J}^*_\varepsilon:\mathbb{R}^{d\times N}\to\mathbb{R}^{d\times N}$ satisfying*

$$(\boldsymbol{J},\boldsymbol{D})\in\mathcal{A}^\varepsilon_\varepsilon \iff \boldsymbol{J}=\boldsymbol{J}^*_\varepsilon(\boldsymbol{D}). \tag{4.2}$$

Moreover, $\boldsymbol{J}^*_\varepsilon(\boldsymbol{0})=\boldsymbol{0}$ and $\boldsymbol{J}^*_\varepsilon$ is Lipschitz continuous and uniformly monotone, i.e., there exist C_1, $C_2>0$ such that for all $\boldsymbol{D}_1$, $\boldsymbol{D}_2\in\mathbb{R}^{d\times N}$

$$\begin{aligned}|\boldsymbol{J}^*(\boldsymbol{D}_1)-\boldsymbol{J}^*(\boldsymbol{D}_2)| &\leq C_2|\boldsymbol{D}_1-\boldsymbol{D}_2|,\\ (\boldsymbol{J}^*(\boldsymbol{D}_1)-\boldsymbol{J}^*(\boldsymbol{D}_2)):(\boldsymbol{D}_1-\boldsymbol{D}_2) &\geq C_1|\boldsymbol{D}_1-\boldsymbol{D}_2|^2.\end{aligned} \tag{4.3}$$

Also, for any arbitrary measurable and bounded $U\subset Q$, let $\boldsymbol{J}^\varepsilon,\boldsymbol{D}^\varepsilon:U\to\mathbb{R}^{d\times N}$ be such that $(\boldsymbol{J}^\varepsilon,\boldsymbol{D}^\varepsilon)\in\mathcal{A}^\varepsilon_\varepsilon$ almost everywhere in U and there is a $C>0$ such that

$$\int_U \boldsymbol{J}^\varepsilon:\boldsymbol{D}^\varepsilon\,dx\,dt\leq C \quad \textit{uniformly with respect to } \varepsilon. \tag{4.4}$$

Then there exist $\boldsymbol{J}\in L^{p'}(U;\mathbb{R}^{d\times N})$, $\boldsymbol{D}\in L^p(U;\mathbb{R}^{d\times N})$ so that (modulo subsequences)

$$\begin{aligned}\boldsymbol{J}^\varepsilon &\rightharpoonup \boldsymbol{J} \quad \textit{weakly in } L^{\min\{2,p'\}}(U;\mathbb{R}^{d\times N}),\\ \boldsymbol{D}^\varepsilon &\rightharpoonup \boldsymbol{D} \quad \textit{weakly in } L^{\min\{2,p\}}(U;\mathbb{R}^{d\times N}).\end{aligned} \tag{4.5}$$

Moreover, if

$$\limsup_{\varepsilon\to 0_+}\int_U \boldsymbol{J}^\varepsilon : \boldsymbol{D}^\varepsilon \,dx\,dt \le \int_U \boldsymbol{J} : \boldsymbol{D}\,dx\,dt, \tag{4.6}$$

then $(\boldsymbol{J},\boldsymbol{D})\in\mathcal{A}$ *almost everywhere in* U *and,*

$$\boldsymbol{J}^\varepsilon : \boldsymbol{D}^\varepsilon \rightharpoonup \boldsymbol{J}:\boldsymbol{D} \quad \text{weakly in } L^1(U). \tag{4.7}$$

The next assertion concerns the properties of $\boldsymbol{G}_\varepsilon$, see Definition 4.3.

Lemma 4.6. *Let* $\boldsymbol{G}$ *satisfy (G1)–(G4) with* $p\in(1,\infty)$*. Then for every* $\varepsilon\in(0,1)$*, the null points of* $\boldsymbol{G}_\varepsilon$ *generate a maximal monotone* 2*-coercive graph* $\tilde{A}^\varepsilon_\varepsilon$ *and* (4.2)–(4.5) *hold. Moreover, if* (4.6) *holds true then* $\boldsymbol{G}(\boldsymbol{J},\boldsymbol{D})=\boldsymbol{0}$ *almost everywhere in* U *and* (4.7) *holds as well.*

The importance and the novelty of these results are twofold. First, we approximate, in a constructive way, a general maximal monotone p-coercive graph $\mathcal{A}$ by Lipschitz continuous and uniformly monotone 2-coercive graphs $\mathcal{A}^\varepsilon_\varepsilon$ that can be identified with a single-valued (Lipschitz continuous and uniformly monotone) mapping. For such mappings there are many tools to obtain the existence of solution to corresponding systems of PDEs. (We provide one such proof in Appendix C.) Then, referring to the convergence part of the above lemma, we observe that to identify the limiting graph, it is just enough to check the validity of (4.6). Note that several subtle tools have been developed to achieve (4.6) for various nonlinear problems of elliptic or parabolic type, mostly in fluid and solid mechanics, that can be used even if the energy equality is not available (or saying differently, if the solution itself is not regular enough to be admissible test function for the limiting problem). We refer to [3] for details. The second point worth noticing is that we do not a priori assume the existence of a Borel measurable selection $\boldsymbol{J}^*$ linked to the graph $\mathcal{A}$. Indeed, in all previous papers we are aware of, this assumption seems to be crucial for solving the corresponding system of partial differential equations. However, here, we are able to avoid such an assumption by a proper definition of the approximate graphs $\mathcal{A}^\varepsilon_\varepsilon$ that are identified with a continuous single-valued mappings $\boldsymbol{J}^*_\varepsilon$. Last, we would like to emphasize, that the approximation by using $\boldsymbol{G}_\varepsilon$ might be more efficient when solving the problem (1.6) numerically, while the approximation by using $\mathcal{A}^\varepsilon_\varepsilon$ is easier to handle from theoretical point of view and follows the classical approaches in the theory of maximal monotone graphs.

Proof of Lemma 4.5. Throughout the proof of Lemma 4.5, we follow the notation indicated in the Definition 4.1, namely $(\boldsymbol{J},\boldsymbol{D})\in\mathcal{A}^\varepsilon_\varepsilon$, $(\tilde{\boldsymbol{J}},\tilde{\boldsymbol{D}})\in\mathcal{A}_\varepsilon$, and $(\overline{\boldsymbol{J}},\overline{\boldsymbol{D}})\in\mathcal{A}$, possibly with indices. The only exception is the limiting object defined in (4.5) as it is not a priori defined to be in any of the graphs. We hope this notation may help to clarify the construction as well as the limiting procedure.

Step 1. <u>The existence of $\boldsymbol{J}^*_\varepsilon$</u>. In (3.1), we identified the maximal monotone graph $\mathcal{A}$ with a possibly multivalued maximal monotone mapping $\boldsymbol{D}^*$ defined on a subset of $\mathbb{R}^{d\times N}$. Thanks to Lemma 3.3 we know that $\boldsymbol{D}^*+\varepsilon\boldsymbol{I}$ is onto $\mathbb{R}^{d\times N}$ for any $\varepsilon\in(0,1]$. This surjectivity and the definition of $\boldsymbol{D}^*$ then imply that for any $\tilde{\boldsymbol{D}}\in\mathbb{R}^{d\times N}$ there is $(\overline{\boldsymbol{J}},\overline{\boldsymbol{D}})\in\mathcal{A}$ such that $\tilde{\boldsymbol{D}}=\overline{\boldsymbol{D}}+\varepsilon\overline{\boldsymbol{J}}$. Setting simply $\tilde{\boldsymbol{J}}:=\overline{\boldsymbol{J}}$, then

$$\begin{gathered}\text{for any } \tilde{\boldsymbol{D}}\in\mathbb{R}^{d\times N} \text{ there exist } (\overline{\boldsymbol{J}},\overline{\boldsymbol{D}})\in\mathcal{A} \text{ and } \tilde{\boldsymbol{J}}\in\mathbb{R}^{d\times N} \text{ such that}\\ \tilde{\boldsymbol{J}}=\overline{\boldsymbol{J}}, \qquad \tilde{\boldsymbol{D}}=\overline{\boldsymbol{D}}+\varepsilon\overline{\boldsymbol{J}}, \quad \text{and} \quad (\tilde{\boldsymbol{J}},\tilde{\boldsymbol{D}})\in\mathcal{A}_\varepsilon.\end{gathered} \tag{4.8}$$

By definition of $\mathcal{A}^\varepsilon_\varepsilon$, for any $(\boldsymbol{J},\boldsymbol{D})\in\mathcal{A}^\varepsilon_\varepsilon$ there exists $(\tilde{\boldsymbol{J}},\tilde{\boldsymbol{D}})\in\mathcal{A}_\varepsilon$ such that $\boldsymbol{J}=\tilde{\boldsymbol{J}}+\varepsilon\tilde{\boldsymbol{D}}$, $\boldsymbol{D}=\tilde{\boldsymbol{D}}$. However, thanks to (4.8), we obtain that

$$\text{for any } \boldsymbol{D}\in\mathbb{R}^{d\times N}, \text{ there exists } \boldsymbol{J}\in\mathbb{R}^{d\times N} \text{ such that } (\boldsymbol{J},\boldsymbol{D})\in\mathcal{A}^\varepsilon_\varepsilon,$$

which guarantees the existence a single-valued mapping $\boldsymbol{J}^*_\varepsilon$ as defined in (4.2).

Step 2. Properties of $\boldsymbol{J}_\varepsilon^*$. To prove its properties, for $i=1,2$, let $(\boldsymbol{J}_i,\boldsymbol{D}_i)\in\mathcal{A}_\varepsilon^\varepsilon$, $(\tilde{\boldsymbol{J}}_i,\tilde{\boldsymbol{D}}_i)\in\mathcal{A}_\varepsilon$, and $(\overline{\boldsymbol{J}}_i,\overline{\boldsymbol{D}}_i)\in\mathcal{A}$, which relate to each other according to the definitions in (4.1). By means of the monotonicity (A2), we obtain

$$
\begin{aligned}
(\tilde{\boldsymbol{J}}_1-\tilde{\boldsymbol{J}}_2):(\tilde{\boldsymbol{D}}_1-\tilde{\boldsymbol{D}}_2)&=(\overline{\boldsymbol{J}}_1-\overline{\boldsymbol{J}}_2):(\overline{\boldsymbol{D}}_1-\overline{\boldsymbol{D}}_2+\varepsilon(\overline{\boldsymbol{J}}_1-\overline{\boldsymbol{J}}_2))\\
&\geq\varepsilon|\overline{\boldsymbol{J}}_1-\overline{\boldsymbol{J}}_2|^2=\varepsilon|\tilde{\boldsymbol{J}}_1-\tilde{\boldsymbol{J}}_2|^2,
\end{aligned}
\tag{4.9}
$$

which implies that $|\tilde{\boldsymbol{J}}_1-\tilde{\boldsymbol{J}}_2|^2\leq c(\varepsilon)|\tilde{\boldsymbol{D}}_1-\tilde{\boldsymbol{D}}_2|^2$. Moreover, using (4.9), we obtain

$$
\begin{aligned}
(\boldsymbol{J}_1-\boldsymbol{J}_2):(\boldsymbol{D}_1-\boldsymbol{D}_2)&=(\tilde{\boldsymbol{J}}_1-\tilde{\boldsymbol{J}}_2+\varepsilon(\tilde{\boldsymbol{D}}_1-\tilde{\boldsymbol{D}}_2)):(\tilde{\boldsymbol{D}}_1-\tilde{\boldsymbol{D}}_2)\\
&\geq\varepsilon|\tilde{\boldsymbol{J}}_1-\tilde{\boldsymbol{J}}_2|^2+\varepsilon|\tilde{\boldsymbol{D}}_1-\tilde{\boldsymbol{D}}_2|^2\\
&=\varepsilon|\boldsymbol{J}_1-\boldsymbol{J}_2-\varepsilon(\boldsymbol{D}_1-\boldsymbol{D}_2)|^2+\varepsilon|\boldsymbol{D}_1-\boldsymbol{D}_2|^2\\
&=\varepsilon\left(|\boldsymbol{J}_1-\boldsymbol{J}_2|^2+(1+\varepsilon^2)|\boldsymbol{D}_1-\boldsymbol{D}_2|^2-2\varepsilon(\boldsymbol{J}_1-\boldsymbol{J}_2):(\boldsymbol{D}_1-\boldsymbol{D}_2)\right),
\end{aligned}
$$

and consequently

$$
(\boldsymbol{J}_1-\boldsymbol{J}_2):(\boldsymbol{D}_1-\boldsymbol{D}_2)\geq\frac{\varepsilon}{1+2\varepsilon^2}\left(|\boldsymbol{J}_1-\boldsymbol{J}_2|^2+(1+\varepsilon^2)|\boldsymbol{D}_1-\boldsymbol{D}_2|^2\right),
$$

which proves the Lipschitz continuity and the uniform monotonicity of $\boldsymbol{J}_\varepsilon^*$. It also gives the 2-coercivity (A4) of $\mathcal{A}_\varepsilon^\varepsilon$ (and $\boldsymbol{J}_\varepsilon^*$ as well) by taking $(\boldsymbol{J}_2,\boldsymbol{D}_2)=(\boldsymbol{0},\boldsymbol{0})$.
Step 3. Proof of (4.5). Let $(\boldsymbol{J}^\varepsilon,\boldsymbol{D}^\varepsilon)\in\mathcal{A}_\varepsilon^\varepsilon$ almost everywhere in U, and let $\int_U\boldsymbol{J}^\varepsilon:\boldsymbol{D}^\varepsilon\,\mathrm{d}x\leq C$. From the 2-coercivity of the graph $\mathcal{A}_\varepsilon^\varepsilon$, we know that $\boldsymbol{J}^\varepsilon,\boldsymbol{D}^\varepsilon\in L^2(Q;\mathbb{R}^{d\times N})$ and for

$$
\overline{\boldsymbol{J}^\varepsilon}:=\boldsymbol{J}^\varepsilon-\varepsilon\boldsymbol{D}^\varepsilon,\qquad\overline{\boldsymbol{D}^\varepsilon}:=\boldsymbol{D}^\varepsilon-\varepsilon\overline{\boldsymbol{J}^\varepsilon}
\tag{4.10}
$$

it holds that $(\overline{\boldsymbol{J}^\varepsilon},\boldsymbol{D}^\varepsilon)\in\mathcal{A}_\varepsilon$ almost everywhere in Q and $(\overline{\boldsymbol{J}^\varepsilon},\overline{\boldsymbol{D}^\varepsilon})\in\mathcal{A}$ almost everywhere in Q. Thanks to the monotonicity (A2),

$$
\boldsymbol{J}^\varepsilon:\boldsymbol{D}^\varepsilon=(\overline{\boldsymbol{J}^\varepsilon}+\varepsilon\boldsymbol{D}^\varepsilon):\boldsymbol{D}^\varepsilon=\varepsilon|\boldsymbol{D}^\varepsilon|^2+\overline{\boldsymbol{J}^\varepsilon}:\boldsymbol{D}^\varepsilon=\varepsilon|\boldsymbol{D}^\varepsilon|^2+\varepsilon|\overline{\boldsymbol{J}^\varepsilon}|^2+\overline{\boldsymbol{J}^\varepsilon}:\overline{\boldsymbol{D}^\varepsilon}\geq 0,
\tag{4.11}
$$

but also

$$
\boldsymbol{J}^\varepsilon:\boldsymbol{D}^\varepsilon\geq\varepsilon|\boldsymbol{D}^\varepsilon|^2+\varepsilon|\overline{\boldsymbol{J}^\varepsilon}|^2+C_1|\overline{\boldsymbol{J}^\varepsilon}|^{p'}+C_1|\overline{\boldsymbol{D}^\varepsilon}|^p-C_2,
\tag{4.12}
$$

due to the p-coercivity (A4) of $\mathcal{A}$. Therefore, using the assumption (4.4),

$$
\int_U\varepsilon|\boldsymbol{D}^\varepsilon|^2+\varepsilon|\overline{\boldsymbol{J}^\varepsilon}|^2+|\overline{\boldsymbol{J}^\varepsilon}|^{p'}+|\overline{\boldsymbol{D}^\varepsilon}|^p\,\mathrm{d}x\,\mathrm{d}t\leq C\ \text{ uniformly with respect to }\varepsilon.
\tag{4.13}
$$

Using the definitions in (4.10),

$$
\begin{aligned}
\int_U|\boldsymbol{J}^\varepsilon|^{\min\{2,p'\}}\,\mathrm{d}x\,\mathrm{d}t&=\int_U|\overline{\boldsymbol{J}^\varepsilon}+\varepsilon\boldsymbol{D}^\varepsilon|^{\min\{2,p'\}}\,\mathrm{d}x\,\mathrm{d}t\leq C,\\
\int_U|\boldsymbol{D}^\varepsilon|^{\min\{2,p\}}\,\mathrm{d}x\,\mathrm{d}t&=\int_U|\overline{\boldsymbol{D}^\varepsilon}+\varepsilon\overline{\boldsymbol{J}^\varepsilon}|^{\min\{2,p\}}\,\mathrm{d}x\,\mathrm{d}t\leq C,
\end{aligned}
$$

and due to reflexivity of $L^p(Q;\mathbb{R}^{d\times N})$ for any $p>1$, there exist $\boldsymbol{J}$, $\boldsymbol{D}$, $\overline{\boldsymbol{J}}$, and $\overline{\boldsymbol{D}}$ such that

$$
\begin{aligned}
\boldsymbol{J}^\varepsilon&\rightharpoonup\boldsymbol{J}&&\text{weakly in }L^{\min\{2,p'\}}(U;\mathbb{R}^{d\times N}),\\
\boldsymbol{D}^\varepsilon&\rightharpoonup\boldsymbol{D}&&\text{weakly in }L^{\min\{2,p\}}(U;\mathbb{R}^{d\times N}),\\
\overline{\boldsymbol{J}^\varepsilon}&\rightharpoonup\overline{\boldsymbol{J}}&&\text{weakly in }L^{p'}(U;\mathbb{R}^{d\times N}),\\
\overline{\boldsymbol{D}^\varepsilon}&\rightharpoonup\overline{\boldsymbol{D}}&&\text{weakly in }L^{p}(U;\mathbb{R}^{d\times N}).
\end{aligned}
\tag{4.14}
$$

Next, we show that $\boldsymbol{J} = \overline{\boldsymbol{J}}$ almost everywhere in U and $\boldsymbol{D} = \overline{\boldsymbol{D}}$ almost everywhere in U. From (4.13) and (4.14) we have

$$\begin{aligned} \varepsilon \boldsymbol{D}^\varepsilon &\rightharpoonup \mathbf{0} \quad \text{weakly in } L^2(U;\mathbb{R}^{d\times N}),\\ \varepsilon \overline{\boldsymbol{J}^\varepsilon} &\rightharpoonup \mathbf{0} \quad \text{weakly in } L^2(U;\mathbb{R}^{d\times N}), \end{aligned} \tag{4.15}$$

and also

$$\begin{aligned} \boldsymbol{J} \leftharpoonup \boldsymbol{J}^\varepsilon = \overline{\boldsymbol{J}^\varepsilon} + \varepsilon \boldsymbol{D}^\varepsilon \rightharpoonup \overline{\boldsymbol{J}} \quad &\Longrightarrow \quad \boldsymbol{J} = \overline{\boldsymbol{J}} \ \text{ in } L^{\min\{2,p'\}}(U;\mathbb{R}^{d\times N}),\\ \boldsymbol{D} \leftharpoonup \boldsymbol{D}^\varepsilon = \overline{\boldsymbol{D}^\varepsilon} + \varepsilon \overline{\boldsymbol{J}^\varepsilon} \rightharpoonup \overline{\boldsymbol{D}} \quad &\Longrightarrow \quad \boldsymbol{D} = \overline{\boldsymbol{D}} \ \text{ in } L^{\min\{2,p\}}(U;\mathbb{R}^{d\times N}). \end{aligned} \tag{4.16}$$

Together, (4.14) and (4.16) prove the statement (4.5).

Step 4. Proof of (4.7). First, we converge in $\mathcal{A}$ itself using very special sequences, and later we show that these special sequences really can be approximated by the points from $\mathcal{A}^\varepsilon_\varepsilon$ (in fact, they are constructed in such way). For $\eta \in (0,1)$, let $(\boldsymbol{J}^\varepsilon, \boldsymbol{D}^\varepsilon) \in \mathcal{A}^\varepsilon_\varepsilon$, and $(\boldsymbol{J}^\eta, \boldsymbol{D}^\eta) \in \mathcal{A}^\eta_\eta$. Then, using (4.10) as an inverse definition to (4.1), $(\overline{\boldsymbol{J}^\varepsilon}, \overline{\boldsymbol{D}^\varepsilon})$ and $(\overline{\boldsymbol{J}^\eta}, \overline{\boldsymbol{D}^\eta}) \in \mathcal{A}$. From the monotonicity of $\mathcal{A}$,

$$|(\overline{\boldsymbol{J}^\varepsilon} - \overline{\boldsymbol{J}^\eta}) : (\overline{\boldsymbol{D}^\varepsilon} - \overline{\boldsymbol{D}^\eta})| = (\overline{\boldsymbol{J}^\varepsilon} - \overline{\boldsymbol{J}^\eta}) : (\overline{\boldsymbol{D}^\varepsilon} - \overline{\boldsymbol{D}^\eta}). \tag{4.17}$$

Also, for any fixed $\varepsilon, \eta \in (0,1)$ it holds that $\overline{\boldsymbol{J}^\varepsilon}, \overline{\boldsymbol{J}^\eta}, \overline{\boldsymbol{D}^\varepsilon}, \overline{\boldsymbol{D}^\eta} \in L^2(Q;\mathbb{R}^{d\times N})$. Then, for any $U \subset Q$, using the weak convergence results (4.14) and (4.16),

$$\begin{aligned} &\limsup_{\eta\to 0_+} \int_U (\overline{\boldsymbol{J}^\varepsilon} - \overline{\boldsymbol{J}^\eta}) : (\overline{\boldsymbol{D}^\varepsilon} - \overline{\boldsymbol{D}^\eta}) \,\mathrm{d}x\,\mathrm{d}t\\ &\qquad = \limsup_{\eta\to 0_+} \int_U \overline{\boldsymbol{J}^\varepsilon} : (\overline{\boldsymbol{D}^\varepsilon} - \overline{\boldsymbol{D}^\eta}) + \overline{\boldsymbol{J}^\eta} : (\overline{\boldsymbol{D}^\eta} - \overline{\boldsymbol{D}^\varepsilon}) \,\mathrm{d}x\,\mathrm{d}t\\ &\qquad = \int_U \overline{\boldsymbol{J}^\varepsilon} : (\overline{\boldsymbol{D}^\varepsilon} - \boldsymbol{D}) \,\mathrm{d}x\,\mathrm{d}t - \int_U \boldsymbol{J} : \overline{\boldsymbol{D}^\varepsilon} \,\mathrm{d}x\,\mathrm{d}t + \limsup_{\eta\to 0_+} \int_U \overline{\boldsymbol{J}^\eta} : \overline{\boldsymbol{D}^\eta} \,\mathrm{d}x\,\mathrm{d}t \end{aligned} \tag{4.18}$$

and

$$\begin{aligned} &\limsup_{\varepsilon\to 0_+} \limsup_{\eta\to 0_+} \int_U (\overline{\boldsymbol{J}^\varepsilon} - \overline{\boldsymbol{J}^\eta}) : (\overline{\boldsymbol{D}^\varepsilon} - \overline{\boldsymbol{D}^\eta}) \,\mathrm{d}x\,\mathrm{d}t\\ &\qquad = \limsup_{\varepsilon\to 0_+} \int_U \overline{\boldsymbol{J}^\varepsilon} : \overline{\boldsymbol{D}^\varepsilon} \,\mathrm{d}x\,\mathrm{d}t - \liminf_{\varepsilon\to 0_+} \int_U \overline{\boldsymbol{J}^\varepsilon} : \boldsymbol{D} \,\mathrm{d}x\,\mathrm{d}t\\ &\qquad\quad - \liminf_{\varepsilon\to 0_+} \int_U \boldsymbol{J} : \overline{\boldsymbol{D}^\varepsilon} \,\mathrm{d}x\,\mathrm{d}t + \limsup_{\eta\to 0_+} \int_U \overline{\boldsymbol{J}^\eta} : \overline{\boldsymbol{D}^\eta} \,\mathrm{d}x\,\mathrm{d}t\\ &\qquad = 2\left(\limsup_{\varepsilon\to 0_+} \int_U \overline{\boldsymbol{J}^\varepsilon} : \overline{\boldsymbol{D}^\varepsilon} \,\mathrm{d}x\,\mathrm{d}t - \int_U \boldsymbol{J} : \boldsymbol{D} \,\mathrm{d}x\,\mathrm{d}t \right). \end{aligned} \tag{4.19}$$

However, using the definitions (4.10), we obtain the estimate

$$\overline{\boldsymbol{J}^\varepsilon} : \overline{\boldsymbol{D}^\varepsilon} = \overline{\boldsymbol{J}^\varepsilon} : \left(\boldsymbol{D}^\varepsilon - \varepsilon \overline{\boldsymbol{J}^\varepsilon}\right) \le \overline{\boldsymbol{J}^\varepsilon} : \boldsymbol{D}^\varepsilon = (\boldsymbol{J}^\varepsilon - \varepsilon \boldsymbol{D}^\varepsilon) : \boldsymbol{D}^\varepsilon \le \boldsymbol{J}^\varepsilon : \boldsymbol{D}^\varepsilon,$$

and if we combine it with the assumption (4.6),

$$\limsup_{\varepsilon\to 0_+} \int_U \overline{\boldsymbol{J}^\varepsilon} : \overline{\boldsymbol{D}^\varepsilon} \,\mathrm{d}x\,\mathrm{d}t \le \limsup_{\varepsilon\to 0_+} \int_U \boldsymbol{J}^\varepsilon : \boldsymbol{D}^\varepsilon \,\mathrm{d}x\,\mathrm{d}t \le \int_U \boldsymbol{J} : \boldsymbol{D} \,\mathrm{d}x\,\mathrm{d}t. \tag{4.20}$$

Now, the results (4.17), (4.19) and (4.20) together imply that

$$\lim_{\varepsilon\to 0_+} \lim_{\eta\to 0_+} \int_U |(\overline{\boldsymbol{J}^\varepsilon} - \overline{\boldsymbol{J}^\eta}) : (\overline{\boldsymbol{D}^\varepsilon} - \overline{\boldsymbol{D}^\eta})| \,\mathrm{d}x\,\mathrm{d}t = 0, \tag{4.21}$$

which proves that for any $\varphi \in L^\infty(U)$,

$$\lim_{\varepsilon\to 0_+}\lim_{\eta\to 0_+}\int_U (\overline{\boldsymbol{J}^\varepsilon}-\overline{\boldsymbol{J}^\eta}):(\overline{\boldsymbol{D}^\varepsilon}-\overline{\boldsymbol{D}^\eta})\varphi\,\mathrm{d}x\,\mathrm{d}t=0.$$

Using the boundedness of φ and a procedure very similar to that in (4.18) and (4.19) we observe that

$$\begin{aligned}0&=\lim_{\varepsilon\to 0_+}\left(\int_U \overline{\boldsymbol{J}^\varepsilon}:(\overline{\boldsymbol{D}^\varepsilon}-\boldsymbol{D})\varphi\,\mathrm{d}x\,\mathrm{d}t-\int_U \boldsymbol{J}:\overline{\boldsymbol{D}^\varepsilon}\varphi\,\mathrm{d}x\,\mathrm{d}t+\lim_{\eta\to 0_+}\int_U \overline{\boldsymbol{J}^\eta}:\overline{\boldsymbol{D}^\eta}\varphi\,\mathrm{d}x\,\mathrm{d}t\right)\\&=2\left(\lim_{\varepsilon\to 0_+}\int_U \overline{\boldsymbol{J}^\varepsilon}:\overline{\boldsymbol{D}^\varepsilon}\varphi\,\mathrm{d}x\,\mathrm{d}t-\int_U \boldsymbol{J}:\boldsymbol{D}\varphi\,\mathrm{d}x\,\mathrm{d}t\right).\end{aligned}$$

As it is true for any $\varphi \in L^\infty(U)$, we obtain

$$\overline{\boldsymbol{J}^\varepsilon}:\overline{\boldsymbol{D}^\varepsilon}\rightharpoonup \boldsymbol{J}:\boldsymbol{D}\quad\text{weakly in}\quad L^1(U). \tag{4.22}$$

Step 5. $\underline{(\boldsymbol{J},\boldsymbol{D})\in\mathcal{A}}$. Let $\boldsymbol{x}$ be a Lebesgue point of $\boldsymbol{J}$, $\boldsymbol{D}$, and $\boldsymbol{J}:\boldsymbol{D}$. Let $(\overline{\boldsymbol{J}},\overline{\boldsymbol{D}})\in\mathcal{A}$ be arbitrary (independent of ε and $\boldsymbol{x}$). Then for any $\varphi\in L^\infty(U)$, $\varphi\ge 0$, using monotonicity of $\mathcal{A}$ and the weak convergence results (4.14), (4.16), and (4.22), we have that

$$\begin{aligned}0&\le\lim_{\varepsilon\to 0_+}\int_U(\overline{\boldsymbol{J}^\varepsilon}-\overline{\boldsymbol{J}}):(\overline{\boldsymbol{D}^\varepsilon}-\overline{\boldsymbol{D}})\varphi\,\mathrm{d}x\,\mathrm{d}t\\&=\lim_{\varepsilon\to 0_+}\int_U \overline{\boldsymbol{J}^\varepsilon}:\overline{\boldsymbol{D}^\varepsilon}\varphi-\overline{\boldsymbol{J}^\varepsilon}:\overline{\boldsymbol{D}}\varphi-\overline{\boldsymbol{J}}:(\overline{\boldsymbol{D}^\varepsilon}-\overline{\boldsymbol{D}})\varphi\,\mathrm{d}x\,\mathrm{d}t\\&=\int_U(\boldsymbol{J}-\overline{\boldsymbol{J}}):(\boldsymbol{D}-\overline{\boldsymbol{D}})\varphi\,\mathrm{d}x\,\mathrm{d}t.\end{aligned}$$

Set $\varphi:=\frac{1}{|B_\rho(\boldsymbol{x})|}\chi_{B_\rho(\boldsymbol{x})}$, and let $\rho\to 0_+$. Since $\boldsymbol{x}$ is a Lebesgue point,

$$0\le\lim_{\rho\to 0_+}\frac{1}{|B_\rho(\boldsymbol{x})|}\int_{B_\rho(\boldsymbol{x})}(\boldsymbol{J}-\overline{\boldsymbol{J}}):(\boldsymbol{D}-\overline{\boldsymbol{D}})\,\mathrm{d}x\,\mathrm{d}t=(\boldsymbol{J}(\boldsymbol{x})-\overline{\boldsymbol{J}}):(\boldsymbol{D}(\boldsymbol{x})-\overline{\boldsymbol{D}}),$$

and it holds for any $(\overline{\boldsymbol{J}},\overline{\boldsymbol{D}})\in\mathcal{A}$, then from maximality of $\mathcal{A}$, see (A3), we obtain that $(\boldsymbol{J}(\boldsymbol{x}),\boldsymbol{D}(\boldsymbol{x}))\in\mathcal{A}$.

Finally, we converge with $(\boldsymbol{J}^\varepsilon,\boldsymbol{D}^\varepsilon)\in\mathcal{A}^\varepsilon_\varepsilon$, using the assumption (4.6) and the results (4.11) for the first, and (4.19) with (4.21) for the second equality,

$$\begin{aligned}\int_U \boldsymbol{J}:\boldsymbol{D}\,\mathrm{d}x\,\mathrm{d}t&\ge\limsup_{\varepsilon\to 0_+}\int_U \boldsymbol{J}^\varepsilon:\boldsymbol{D}^\varepsilon\,\mathrm{d}x\,\mathrm{d}t\\&=\limsup_{\varepsilon\to 0_+}\int_U \overline{\boldsymbol{J}^\varepsilon}:\overline{\boldsymbol{D}^\varepsilon}+\varepsilon|\boldsymbol{D}^\varepsilon|^2+\varepsilon|\overline{\boldsymbol{J}^\varepsilon}|^2\,\mathrm{d}x\,\mathrm{d}t\\&=\int_U \boldsymbol{J}:\boldsymbol{D}\,\mathrm{d}x\,\mathrm{d}t+\limsup_{\varepsilon\to 0_+}\int_U\varepsilon|\boldsymbol{D}^\varepsilon|^2+\varepsilon|\overline{\boldsymbol{J}^\varepsilon}|^2\,\mathrm{d}x\,\mathrm{d}t.\end{aligned}$$

Hence, the last integral vanishes as $\varepsilon\to 0_+$, and therefore $(\sqrt{\varepsilon}\,\boldsymbol{D}^\varepsilon)$ and $(\sqrt{\varepsilon}\,\overline{\boldsymbol{J}^\varepsilon})$ converge strongly to zero in $L^2(U;\mathbb{R}^{d\times N})$, as opposed to the weak result in (4.15).

Finally, since $(\boldsymbol{J}^\varepsilon:\boldsymbol{D}^\varepsilon)=(\overline{\boldsymbol{J}^\varepsilon}:\overline{\boldsymbol{D}^\varepsilon}+\varepsilon|\overline{\boldsymbol{J}^\varepsilon}|^2+\varepsilon|\boldsymbol{D}^\varepsilon|^2)$, we use that the first term converges weakly in $L^1(U)$ to the desired limit thanks to (4.22) and the last two converge strongly to zero in $L^1(U)$ to obtain the final statement (4.7). □

Proof of Lemma 4.6. First, recalling that $\boldsymbol{G}_\varepsilon(\boldsymbol{J},\boldsymbol{D})=\boldsymbol{G}(\boldsymbol{J}-\varepsilon\boldsymbol{D},\boldsymbol{D}-\varepsilon\boldsymbol{J})$, it is straightforward to observe that $(\bar{\boldsymbol{J}},\bar{\boldsymbol{D}})$ is a null point of $\boldsymbol{G}$ if and only if the couple $(\boldsymbol{J},\boldsymbol{D})$ defined through

$$\boldsymbol{D}=\frac{\bar{\boldsymbol{D}}+\varepsilon\bar{\boldsymbol{J}}}{(1-\varepsilon^2)},\qquad \boldsymbol{J}=\frac{\bar{\boldsymbol{J}}+\varepsilon\bar{\boldsymbol{D}}}{(1-\varepsilon^2)} \tag{4.23}$$

is a null point of $\boldsymbol{G}_\varepsilon$. Then, since $\boldsymbol{D}^* + \varepsilon \boldsymbol{I}$ is onto (see (3.1) for the definition of $\boldsymbol{D}^*$ and Step 1 in proof of Lemma 3.3), we see that $\boldsymbol{D}$ can be understood as a (single valued) function of $\boldsymbol{J}$ and analogously (by interchaning the role of $\boldsymbol{J}$ and $\boldsymbol{D}$) $\boldsymbol{J}$ can be understood as a function of $\boldsymbol{D}$. Next, we show that these mappings are uniformly monotone and Lipschitz continuous, which implies that $\boldsymbol{G}_\varepsilon$ generates maximal monotone 2-coercive graph. Indeed, let $(\boldsymbol{J}_1, \boldsymbol{D}_1)$ and $(\boldsymbol{J}_2, \boldsymbol{D}_2)$ be two null points of $\boldsymbol{G}_\varepsilon$. Then, $(\boldsymbol{J}_i - \varepsilon \boldsymbol{D}_i, \boldsymbol{D}_i - \varepsilon \boldsymbol{J}_i)$ are null points of $\boldsymbol{G}$ and, as the graph generated by $\boldsymbol{G}$ is by Lemma 3.4 monotone, we have

$$\begin{aligned} 0 &\le ((\boldsymbol{J}_1 - \varepsilon \boldsymbol{D}_1) - (\boldsymbol{J}_2 - \varepsilon \boldsymbol{D}_2)) : ((\boldsymbol{D}_1 - \varepsilon \boldsymbol{J}_1) - (\boldsymbol{D}_2 - \varepsilon \boldsymbol{J}_2)) \\ &= ((\boldsymbol{J}_1 - \boldsymbol{J}_2) - \varepsilon(\boldsymbol{D}_1 - \boldsymbol{D}_2)) : ((\boldsymbol{D}_1 - \boldsymbol{D}_2) - \varepsilon(\boldsymbol{J}_1 - \boldsymbol{J}_2)) \\ &= (1+\varepsilon^2)(\boldsymbol{J}_1 - \boldsymbol{J}_2) : (\boldsymbol{D}_1 - \boldsymbol{D}_2) - \varepsilon(|\boldsymbol{D}_1 - \boldsymbol{D}_2|^2 + |\boldsymbol{J}_1 - \boldsymbol{J}_2|^2). \end{aligned}$$

Consequently,

$$\frac{\varepsilon}{1+\varepsilon^2}(|\boldsymbol{D}_1 - \boldsymbol{D}_2|^2 + |\boldsymbol{J}_1 - \boldsymbol{J}_2|^2) \le (\boldsymbol{J}_1 - \boldsymbol{J}_2) : (\boldsymbol{D}_1 - \boldsymbol{D}_2), \tag{4.24}$$

which is the desired uniform monotonicity and which implies, after applying the Cauchy–Schwarz inequality to the right-hand side, the Lipschitz continuity. Consequently, the null points of $\boldsymbol{G}_\varepsilon$ generate a maximal monotone 2-coercive graph.

The rest of the proof coincides with the proof of Lemma 4.5 with necessary minor changes due to a slightly different relation between the null points of $\boldsymbol{G}$ and $\boldsymbol{G}_\varepsilon$ given by (4.23) and the relation between the graphs $\mathcal{A}$ and $\mathcal{A}^\varepsilon_\varepsilon$ given by Definition 4.1. □

Further auxiliary results. We finish this section by stating three results. Two of them, Lemma 4.7 and Lemma 4.9, will be needed in the proof of the main theorem. The third result, see Lemma 4.8, is of independent interest within the context of earlier established results requiring a priori the existence of a Borel measurable selection.

The first result establishes the condition that guarantees the stability of the graph $\mathcal{A}$ with respect to weakly converging sequences. It is a simpler variant of Lemma 4.5 above.

Lemma 4.7. *Let $\mathcal{A}$ be a maximal monotone p-coercive graph and $U \subset \mathbb{R}^d$ be a measurable bounded set. Assume that for every $n \in \mathbb{N}$, the mappings $\boldsymbol{J}^n, \boldsymbol{D}^n : U \to \mathbb{R}^{d\times N}$ are such that $(\boldsymbol{J}^n, \boldsymbol{D}^n) \in \mathcal{A}$ almost everywhere in U. In addition, let*

$$\int_U \boldsymbol{J}^n : \boldsymbol{D}^n \, dx \, dt \le C \quad \text{uniformly with respect to } n \in \mathbb{N}.$$

Then there exist $\boldsymbol{J} \in L^{p'}(U; \mathbb{R}^{d\times N}$ and $\boldsymbol{D} \in L^p(U; \mathbb{R}^{d\times N})$ such that

$$\begin{aligned} \boldsymbol{J}^n &\rightharpoonup \boldsymbol{J} && \text{weakly in } L^{p'}(U; \mathbb{R}^{d\times N}), \\ \boldsymbol{D}^n &\rightharpoonup \boldsymbol{D} && \text{weakly in } L^p(U; \mathbb{R}^{d\times N}). \end{aligned}$$

Moreover, if

$$\limsup_{n\to\infty} \int_U \boldsymbol{J}^n : \boldsymbol{D}^n \, dx \, dt \le \int_U \boldsymbol{J} : \boldsymbol{D} \, dx \, dt,$$

then $(\boldsymbol{J}, \boldsymbol{D}) \in \mathcal{A}$ almost everywhere in U and $\boldsymbol{J}^n : \boldsymbol{D}^n \rightharpoonup \boldsymbol{J} : \boldsymbol{D}$ weakly in $L^1(U)$.

Proof. See Lemma 1.2.2 in [7] or Lemma 4.5 above. □

The next lemma is of interest within the context of the mathematical methods for general constitutive equations of the form $\boldsymbol{G}(\boldsymbol{J}, \boldsymbol{D}) = \boldsymbol{0}$ (associated with the graph $\mathcal{A}$) developed earlier for fluid flow problems, see [8, 9, 7, 11]. In these studies, the assumption on the existence of a Borel measurable selection played an important role both for constructing approximating single-valued mappping (by convolution) and for showing that

$$\text{for each } \boldsymbol{D} \in L^p \text{ there is } \boldsymbol{J} \in L^{p'} \text{ such that } (\boldsymbol{J}, \boldsymbol{D}) \in \mathcal{A}. \tag{4.25}$$

In this study, we do not require the existence of a Borel measurable selection due to a different approximation scheme developed above in this section. For the sake of completeness, we also show that the property (4.25) is available.

Lemma 4.8. *Let $p \in (1,\infty)$ and $\mathcal{A}$ be a maximal monotone p-coercive graph. Then for every $\boldsymbol{D} \in L^p(Q;\mathbb{R}^{d\times N})$ there exists $\boldsymbol{J} \in L^{p'}(Q;\mathbb{R}^{d\times N})$ such that $(\boldsymbol{J},\boldsymbol{D}) \in \mathcal{A}$ almost everywhere in Q.*

Proof. For $k \in \mathbb{N}$, define $\boldsymbol{D}_k := \boldsymbol{D}\chi_{\{|\boldsymbol{D}|\le k\}}$. Recall the definition of $\mathcal{A}^\varepsilon_\varepsilon$ (4.1b) and its selection $\boldsymbol{J}^*_\varepsilon$ (4.2). Then, by definition of selection, $(\boldsymbol{J}^*_\varepsilon(\boldsymbol{D}_k),\boldsymbol{D}_k) \in \mathcal{A}^\varepsilon_\varepsilon$ almost everywhere in Q. We can estimate

$$|\boldsymbol{J}^*_\varepsilon(\boldsymbol{D}_k)| \le C(|\boldsymbol{D}_k|^2 + |\boldsymbol{D}_k|^p) \le C(k^2 + k^p),$$

which implies that there exists $\boldsymbol{J}_k$ such that as $\varepsilon \to 0_+$,

$$\boldsymbol{J}^*_\varepsilon(\boldsymbol{D}_k) \rightharpoonup^* \boldsymbol{J}_k \text{ weakly}^* \text{ in } L^\infty(Q;\mathbb{R}^{d\times N}).$$

Then we have the limit

$$\lim_{\varepsilon\to 0_+}\int_Q \boldsymbol{J}^*_\varepsilon(\boldsymbol{D}_k) : \boldsymbol{D}_k \,\mathrm{d}x\,\mathrm{d}t = \int_Q \boldsymbol{J}_k : \boldsymbol{D}_k \,\mathrm{d}x\,\mathrm{d}t,$$

and thanks to Lemma 4.7 we know that $(\boldsymbol{J}_k,\boldsymbol{D}_k) \in \mathcal{A}$ almost everywhere in Q. Therefore,

$$C_1(|\boldsymbol{J}_k|^{p'} + |\boldsymbol{D}_k|^p) - C_2 \le \boldsymbol{J}_k : \boldsymbol{D}_k \le \frac{C_1}{p'}|\boldsymbol{J}_k|^{p'} + C|\boldsymbol{D}_k|^p,$$

and then

$$\int_Q |\boldsymbol{J}_k|^{p'} + |\boldsymbol{D}_k|^p \,\mathrm{d}x\,\mathrm{d}t \le C\int_Q |\boldsymbol{D}_k|^p \,\mathrm{d}x\,\mathrm{d}t \le \int_Q |\boldsymbol{D}|^p \,\mathrm{d}x\,\mathrm{d}t \le C,$$

where the boundedness follows from the assumption. Finally, as $k \to +\infty$, we have for subsequences that

$$\begin{aligned} \boldsymbol{J}_k &\rightharpoonup \boldsymbol{J} && \text{weakly in } L^{p'}(Q;\mathbb{R}^{d\times N}),\\ \boldsymbol{D}_k &\to \boldsymbol{D} && \text{strongly in } L^p(Q;\mathbb{R}^{d\times N}),\end{aligned}$$

so $\lim_{k\to\infty}\int_Q \boldsymbol{J}_k : \boldsymbol{D}_k \,\mathrm{d}x\,\mathrm{d}t = \int_Q \boldsymbol{J} : \boldsymbol{D}\,\mathrm{d}x\,\mathrm{d}t$, which finishes the proof by use of Lemma 4.7. □

We finish this section by proving the uniform (ε-independent) coercivity estimate for $\mathcal{A}^\varepsilon_\varepsilon$.

Lemma 4.9. *There exist $\tilde{C}_1, \tilde{C}_2 \in \mathbb{R}_+$ such that for all $\varepsilon \in (0,1)$ and all $(\boldsymbol{J}^\varepsilon,\boldsymbol{D}^\varepsilon) \in \mathcal{A}^\varepsilon_\varepsilon$ there holds*

$$\boldsymbol{J}^\varepsilon : \boldsymbol{D}^\varepsilon \ge \tilde{C}_1(|\boldsymbol{J}^\varepsilon|^{\min\{p',2\}}) + |\boldsymbol{D}^\varepsilon|^{\min\{p,2\}}) - \tilde{C}_2. \tag{4.26}$$

Proof. Let $(\boldsymbol{J},\boldsymbol{D}) \in \mathcal{A}$ be the couple corresponding to $(\boldsymbol{J}^\varepsilon,\boldsymbol{D}^\varepsilon) \in \mathcal{A}^\varepsilon_\varepsilon$ according to Definition 4.1. Then

$$\boldsymbol{J}^\varepsilon = \boldsymbol{J} + \varepsilon\boldsymbol{D}^\varepsilon, \quad \text{and} \quad \boldsymbol{D}^\varepsilon = \boldsymbol{D} + \varepsilon\boldsymbol{J}.$$

Now, using the p-coercivity of $\mathcal{A}$, we get (4.12), and if we compute

$$\begin{aligned} |\boldsymbol{D}^\varepsilon|^{\min\{p,2\}} &= |\boldsymbol{D} + \varepsilon\boldsymbol{J}|^{\min\{p,2\}} \le C(|\boldsymbol{D}|^p + \varepsilon|\boldsymbol{J}|^2 + 1),\\ |\boldsymbol{J}^\varepsilon|^{\min\{p',2\}} &= |\boldsymbol{J} + \varepsilon\boldsymbol{D}^\varepsilon|^{\min\{p',2\}} \le C(|\boldsymbol{J}|^{p'} + \varepsilon|\boldsymbol{D}^\varepsilon|^2 + 1),\end{aligned}$$

and combine it together, we obtain

$$|\boldsymbol{J}^\varepsilon|^{\min\{p',2\}} + |\boldsymbol{D}^\varepsilon|^{\min\{p,2\}} \le C(|\boldsymbol{D}|^p + \varepsilon|\boldsymbol{J}|^2 + |\boldsymbol{J}|^{p'} + \varepsilon|\boldsymbol{D}^\varepsilon|^2 + 1) \le C(\boldsymbol{J}^\varepsilon : \boldsymbol{D}^\varepsilon + 1).$$

□

5. Proof of Theorem 2.1

The proof is based on the identification of the null set of $\boldsymbol{G}$ with a maximal monotone p-coercive graph $\mathcal{A}$ and on its subsequent approximation by the Lipschitz continuous and uniformly monotone 2-coercive graphs $\mathcal{A}^\varepsilon_\varepsilon$ constructed and analyzed in Section 4. The solution of the problem is then obtained by limiting process as $\varepsilon \to 0_+$. In order to link original p-coercive graph with approximating 2-coercive graphs, we need to consider smoother right-hand side $\boldsymbol{f}$. More precisely, we define

$$
\begin{aligned}
\mu &:= \min\{p, 2\}, \quad \mu' := \max\{p', 2\}, \\
\nu &:= \min\{p', 2\}, \quad \nu' := \max\{p, 2\}.
\end{aligned} \tag{5.1}
$$

and then, in the first seven steps of the proof, we prove Theorem 2.1 for $\boldsymbol{f} \in L^{\mu'}(0,T;V^*_\mu)$. In the final Step 8, once having a solution for such $\boldsymbol{f}$, we consider a sequence of solutions $\{(\boldsymbol{u}^m, \boldsymbol{J}^m)\}_{m\in\mathbb{N}}$ of the problem (1.6) in the sense of Theorem 2.1 with the right-hand side $\{\boldsymbol{f}^m\}_{m\in\mathbb{N}} \subset L^{\mu'}(0,T;V^*_\mu)$ satisfying[8] $\boldsymbol{f}^m \to \boldsymbol{f}$ in $L^{p'}(0,T;V^*_p)$ and we briefly comment why the weak limits $(\boldsymbol{u}, \boldsymbol{J})$ of suitable subsequences $\{(\boldsymbol{u}^m, \boldsymbol{J}^m)\}_{m\in\mathbb{N}}$ solve the problem (1.6) with the righ-hand side $\boldsymbol{f}$.

Step 1. Approximations. First, we introduce a graph $\mathcal{A}$ by

$$
\mathcal{A} := \{(\boldsymbol{J}, \boldsymbol{D}) : \ \boldsymbol{G}(\boldsymbol{J}, \boldsymbol{D}) = \boldsymbol{0}\}.
$$

Then, due to Lemma 3.4, it follows from the assumptions (G1)–(G4) that $\mathcal{A}$ is a maximal monotone p-coercive graph, i.e., $\mathcal{A}$ satisfies (A1)–(A4) in Definition 3.1. Consequently, for an arbitrary $\varepsilon \in (0,1)$, we use (4.1) to construct ε-approximate graphs $\mathcal{A}^\varepsilon_\varepsilon$. Then, due to Lemma 4.5, we observe that $\mathcal{A}^\varepsilon_\varepsilon$ can be identified with a Lipschitz continuous and uniformly monotone single-valued mapping $\boldsymbol{J}^*_\varepsilon$ so that

$$
(\boldsymbol{J}, \boldsymbol{D}) \in \mathcal{A}^\varepsilon_\varepsilon \iff \boldsymbol{J} = \boldsymbol{J}^*_\varepsilon(\boldsymbol{D}).
$$

Consequently, for every $\varepsilon \in (0,1)$, we can apply Lemma C.1 and find

$$
(\boldsymbol{u}^\varepsilon, \boldsymbol{J}^\varepsilon) \in \big(L^2(0,T;V) \cap \mathcal{C}([0,T];H)\big) \times L^2(Q;\mathbb{R}^{d\times N})
$$

satisfying

$$
\langle \partial_t \boldsymbol{u}^\varepsilon, \boldsymbol{\varphi}\rangle_V + \int_\Omega \boldsymbol{J}^\varepsilon : \nabla\boldsymbol{\varphi} \,\mathrm{d}x = \langle \boldsymbol{f}, \boldsymbol{\varphi}\rangle_V \quad \text{for a.a. } t \in (0,T] \text{ and for any } \boldsymbol{\varphi} \in V, \tag{5.2}
$$

$$
\boldsymbol{J}^\varepsilon = \boldsymbol{J}^*_\varepsilon(\nabla \boldsymbol{u}^\varepsilon) \quad \text{almost everywhere in } Q, \tag{5.3}
$$

$$
\lim_{t\to 0_+} \|\boldsymbol{u}^\varepsilon(t) - \boldsymbol{u}_0\|_H = 0. \tag{5.4}
$$

Step 2. Uniform a priori estimates. We set $\boldsymbol{\varphi} := \boldsymbol{u}^\varepsilon$ in (5.2), integrate over $(0,t)$, use that $\partial_t \boldsymbol{u}^\varepsilon \in L^2(0,T;V^*)$ and properties of the Gelfand triplet (2.2), and obtain

$$
\frac{1}{2}\|\boldsymbol{u}^\varepsilon(t)\|_H^2 + \int_{Q_t} \boldsymbol{J}^\varepsilon : \nabla \boldsymbol{u}^\varepsilon \,\mathrm{d}x\,\mathrm{d}\tau = \int_0^t \langle \boldsymbol{f}, \boldsymbol{u}^\varepsilon\rangle_{V_\mu} \mathrm{d}\tau + \frac{1}{2}\|\boldsymbol{u}_0\|_H^2.
$$

Using the estimate (4.26) from Lemma 4.9, we get

$$
\begin{aligned}
\frac{1}{2}\|\boldsymbol{u}^\varepsilon(t)\|_H^2 + \tilde{C}_1 \int_{Q_t} |\boldsymbol{J}^\varepsilon|^\nu + |\nabla \boldsymbol{u}^\varepsilon|^\mu \,\mathrm{d}x\,\mathrm{d}\tau &\le \frac{1}{2}\|\boldsymbol{u}^\varepsilon(t)\|_H^2 + \int_{Q_t} \boldsymbol{J}^\varepsilon : \nabla \boldsymbol{u}^\varepsilon \,\mathrm{d}x\,\mathrm{d}\tau + C \\
&\le \int_0^t \langle \boldsymbol{f}, \boldsymbol{u}^\varepsilon\rangle_{V_\mu} \mathrm{d}\tau + \frac{1}{2}\|\boldsymbol{u}_0\|_H^2 + C.
\end{aligned} \tag{5.5}
$$

[8] Since V^*_μ is dense in V^*_p for $\mu' \ge p'$, such sequence surely exists.

Next, recalling the definition of the V_μ-norm and using Young's inequality, we get

$$\begin{aligned}\langle \boldsymbol{f}, \boldsymbol{u}^\varepsilon\rangle_{V_\mu} &\le \|\boldsymbol{f}\|_{V_\mu^*}\left(\|\boldsymbol{u}^\varepsilon\|_H + \|\nabla \boldsymbol{u}^\varepsilon\|_{L^\mu(\Omega)}\right)\\ &\le \frac{\tilde{C}_1}{2}\|\nabla \boldsymbol{u}^\varepsilon\|^\mu_{L^\mu(\Omega)} + C\left(\|\boldsymbol{f}\|^{\mu'}_{V_\mu^*} + (\|\boldsymbol{u}^\varepsilon\|_H^2 + 1)\|\boldsymbol{f}\|_{V_\mu^*}\right).\end{aligned}$$

Inserting it into (5.5), using the assumptions on data $\boldsymbol{f}$ and $\boldsymbol{u}_0$ and applying then the Gronwall lemma, we get

$$\sup_{t\in(0,T)} \|\boldsymbol{u}^\varepsilon(t)\|_H \le C \text{ uniformly with respect to } \varepsilon \in (0,1). \tag{5.6}$$

Referring again to (5.5) we then also conclude that

$$\sup_{t\in(0,T)} \|\boldsymbol{u}^\varepsilon(t)\|_H^2 + \int_Q |\boldsymbol{J}^\varepsilon|^\nu + |\nabla \boldsymbol{u}^\varepsilon|^\mu \,\mathrm{d}x\,\mathrm{d}\tau \le C \text{ uniformly with respect to } \varepsilon \in (0,1). \tag{5.7}$$

Moreover, we also have

$$\int_Q \boldsymbol{J}^\varepsilon : \nabla \boldsymbol{u}^\varepsilon \,\mathrm{d}x\,\mathrm{d}\tau \le C \text{ uniformly with respect to } \varepsilon \in (0,1). \tag{5.8}$$

Finally, note that (5.7) also implies that

$$\|\boldsymbol{u}^\varepsilon\|_{L^\mu(0,T;V_\mu)} \le C \text{ uniformly with respect to } \varepsilon \in (0,1). \tag{5.9}$$

To estimate the time derivative, denote $\mathcal{W} := \{\boldsymbol{w} \in V_p \cap V; \|\boldsymbol{w}\|_{V_{\nu'}} \le 1\}$. Note that $\mathcal{W} \subset V$, then we can set $\boldsymbol{\varphi} := \boldsymbol{w} \in \mathcal{W}$ in the equation (C.5a) to get the following

$$\begin{aligned}\|\partial_t \boldsymbol{u}^\varepsilon\|_{V_{\nu'}^*} = \sup_{\mathcal{W}} \langle \partial_t \boldsymbol{u}^\varepsilon, \boldsymbol{w}\rangle_{V_{\nu'}} &= \sup_{\mathcal{W}}\left(-\int_\Omega \boldsymbol{J}^\varepsilon : \nabla \boldsymbol{w}\,\mathrm{d}x + \langle \boldsymbol{f}, \boldsymbol{w}\rangle_{V_\mu}\right)\\ &\le \sup_{\mathcal{W}}\left(\|\boldsymbol{J}^\varepsilon\|_{L_\nu(\Omega;\mathbb{R}^{d\times N})}\|\nabla \boldsymbol{w}\|_{L_{\nu'}(\Omega;\mathbb{R}^{d\times N})} + \|\boldsymbol{f}\|_{V_{\mu'}}\|\boldsymbol{w}\|_{V_\mu}\right).\end{aligned}$$

Using the fact that $V_\mu \subset V_{\nu'}$, taking the ν-th power and integrating the result over $(0,T)$ we obtain, using also (5.7),

$$\int_0^T \|\partial_t \boldsymbol{u}^\varepsilon\|^\nu_{V_{\nu'}^*}\,\mathrm{d}t \le \int_0^T \|\boldsymbol{J}^\varepsilon\|^\nu_{L_\nu(\Omega;\mathbb{R}^{d\times N})} + \|\boldsymbol{f}\|^\nu_{V_{\mu'}}\,\mathrm{d}t \le C \text{ uniformly with respect to } \varepsilon \in (0,1). \tag{5.10}$$

Step 3. Limit $\varepsilon \to 0_+$. Using (5.9), (5.7), (5.6) and (5.10), we obtain that as $\varepsilon \to 0_+$,

$$\begin{aligned} \boldsymbol{u}^\varepsilon &\rightharpoonup \boldsymbol{u} && \text{weakly in } L^\mu(0,T;V_\mu),\\ \boldsymbol{J}^\varepsilon &\rightharpoonup \boldsymbol{J} && \text{weakly in } L^\nu(Q;\mathbb{R}^{d\times N}),\\ \boldsymbol{u}^\varepsilon &\rightharpoonup^* \boldsymbol{u} && \text{weakly}^* \text{ in } L^\infty(0,T;H),\\ \partial_t \boldsymbol{u}^\varepsilon &\rightharpoonup \partial_t \boldsymbol{u} && \text{weakly in } L^\nu(0,T;V_{\nu'}^*).\end{aligned} \tag{5.11}$$

Moreover, (5.8) in combination with the result of the Lemma 4.5 gives

$$\boldsymbol{J} \in L^{p'}(Q;\mathbb{R}^{d\times N}) \quad\text{and}\quad \nabla \boldsymbol{u} \in L^p(Q;\mathbb{R}^{d\times N}). \tag{5.12}$$

Next, take $\boldsymbol{w} \in V_{\mu'}$ and $\xi \in L^\infty(0,T)$ arbitrary. Setting $\boldsymbol{\varphi} := \xi\boldsymbol{w}$ in (5.2), integrating the result over $(0,T)$, we obtain

$$\int_0^T \langle \partial_t \boldsymbol{u}^\varepsilon, \xi\boldsymbol{w}\rangle_{V_\mu}\,\mathrm{d}t + \int_Q \boldsymbol{J}^\varepsilon : \nabla \boldsymbol{w}\xi\,\mathrm{d}x\,\mathrm{d}t = \int_0^T \langle \boldsymbol{f}, \xi\boldsymbol{w}\rangle_{V_\mu}\,\mathrm{d}t.$$

Noticing that all terms are well-defined, we can take the limit $\varepsilon \to 0_+$ and, by means of (5.11), we end up with

$$\int_0^T \langle \partial_t \boldsymbol{u}, \xi \boldsymbol{w} \rangle_{V_\mu} \,\mathrm{d}t + \int_Q \boldsymbol{J} : \nabla \boldsymbol{w} \xi \,\mathrm{d}x \,\mathrm{d}t = \int_0^T \langle \boldsymbol{f}, \xi \boldsymbol{w} \rangle_{V_\mu} \,\mathrm{d}t.$$

Since this holds for all ξ, it implies

$$\langle \partial_t \boldsymbol{u}, \boldsymbol{w} \rangle_{V_\mu} + \int_\Omega \boldsymbol{J} : \nabla \boldsymbol{w} \,\mathrm{d}x = \langle \boldsymbol{f}, \boldsymbol{w} \rangle_{V_\mu} \quad \text{for a.a. } t \in (0,T) \text{ and for all } w \in V_{\mu'}. \tag{5.13}$$

Since $V_p \cap V \hookrightarrow V_p$ and this embedding is dense, all terms in (5.13) are well-defined. To verify (2.5a), we need to show that (5.13) holds true for all $\boldsymbol{w} \in V_p$. For this purpose, we need to improve the information about the time derivative.

Step 4. Improved information regarding $\partial_t \boldsymbol{u}$. Thanks to the dense embedding $V_p \cap V \hookrightarrow V_p$, we can use (5.13) for $\mathcal{W}_p := \{\boldsymbol{w} \in V_p \cap V; \|\boldsymbol{w}\|_{V_p} \le 1\}$ as follows

$$\begin{aligned}\|\partial_t \boldsymbol{u}\|_{V_p^*} &= \sup_{\mathcal{W}_p} \langle \partial_t \boldsymbol{u}, \boldsymbol{w} \rangle_{V_p} = \sup_{\mathcal{W}_p} \left(-\int_\Omega \boldsymbol{J} : \nabla \boldsymbol{w} \,\mathrm{d}x + \langle \boldsymbol{f}, \boldsymbol{w} \rangle_{V_p} \right) \\ &\le \sup_{\mathcal{W}_p} \left(\|\boldsymbol{J}\|_{L^{p'}(\Omega;\mathbb{R}^{d\times N})} \|\nabla \boldsymbol{w}\|_{L^p(\Omega;\mathbb{R}^{d\times N})} + \|\boldsymbol{f}\|_{V_p^*} \|\boldsymbol{w}\|_{V_p} \right) \le \|\boldsymbol{J}\|_{L^{p'}(\Omega;\mathbb{R}^{d\times N})} + \|\boldsymbol{f}\|_{V_p^*}.\end{aligned}$$

Applying the power p', integrating over time $t \in (0,T)$, and using the results in (5.11), we obtain

$$\int_0^T \|\partial_t \boldsymbol{u}\|_{V_p^*}^{p'} \,\mathrm{d}t \le \int_0^T \|\boldsymbol{J}\|_{L^{p'}(\Omega;\mathbb{R}^{d\times N})}^{p'} + \|\boldsymbol{f}\|_{V_p^*}^{p'} \,\mathrm{d}t \le C,$$

and again using the density of $V_p \cap V \hookrightarrow V_p$, we conclude that (5.13) is valid for any $\boldsymbol{w} \in V_p$ and for almost every $t \in (0,T)$.

Moreover, thanks to $\boldsymbol{u} \in L^p(0,T;V_p)$, $\partial_t \boldsymbol{u} \in L^{p'}(0,T;V_p^*)$, and the Gelfand triplet (2.2), there holds $\boldsymbol{u} \in \mathcal{C}([0,T];H)$.

Step 5. Attainment of the initial datum. For $0 < \epsilon \ll 1$ and $t \in (0, T-\epsilon)$, we first introduce a cut-off function $\eta \in \mathcal{C}^{0,1}([0,T])$ as a piece-wise linear function of three parameters:

$$\eta(\tau) = \begin{cases} 1 & \text{if } \tau \in [0,t), \\ 1 + \frac{t-\tau}{\epsilon} & \text{if } \tau \in [t, t+\epsilon), \\ 0 & \text{if } \tau \in [t+\epsilon, T]. \end{cases} \tag{5.14}$$

Next, for $\boldsymbol{w} \in V_{\mu'}$, we set $\boldsymbol{\varphi} := \eta \boldsymbol{w}$ in (C.5a) and integrate over $(0,T)$,

$$\frac{1}{\epsilon} \int_t^{t+\epsilon} (\boldsymbol{u}^\varepsilon(\tau), \boldsymbol{w})_H \,\mathrm{d}x \,\mathrm{d}\tau + \int_{Q_{t+\epsilon}} \boldsymbol{J}^\varepsilon : \nabla \boldsymbol{w} \eta \,\mathrm{d}x \,\mathrm{d}\tau = \int_0^{t+\epsilon} \langle \boldsymbol{f}, \boldsymbol{w}\eta \rangle_{V_\mu} \,\mathrm{d}\tau + (\boldsymbol{u}_0, \boldsymbol{w})_H.$$

Letting $\varepsilon \to 0_+$ and using the results established in (5.11), we conclude that

$$\frac{1}{\epsilon} \int_t^{t+\epsilon} (\boldsymbol{u}(\tau), \boldsymbol{w})_H \,\mathrm{d}x \,\mathrm{d}\tau + \int_{Q_{t+\epsilon}} \boldsymbol{J} : \nabla \boldsymbol{w} \eta \,\mathrm{d}x \,\mathrm{d}\tau = \int_0^{t+\epsilon} \langle \boldsymbol{f}, \boldsymbol{w}\eta \rangle_{V_\mu} \,\mathrm{d}\tau + (\boldsymbol{u}_0, \boldsymbol{w})_H.$$

Since $\boldsymbol{u} \in \mathcal{C}([0,T];H)$, we can also pass with $\epsilon \to 0_+$,

$$(\boldsymbol{u}(t), \boldsymbol{w})_H + \int_{Q_t} \boldsymbol{J} : \nabla \boldsymbol{w} \,\mathrm{d}x \,\mathrm{d}\tau = \int_0^t \langle \boldsymbol{f}, \boldsymbol{w} \rangle_{V_\mu} \,\mathrm{d}\tau + (\boldsymbol{u}_0, \boldsymbol{w})_H,$$

and finally, we let $t \to 0_+$ to get

$$\lim_{t \to 0_+} (\boldsymbol{u}(t), \boldsymbol{w})_H = (\boldsymbol{u}_0, \boldsymbol{w})_H.$$

As $\boldsymbol{w} \in V_p \cap V$ was arbitrary and $V_p \cap V$ is dense in H, we obtain that $\boldsymbol{u}(t) \rightharpoonup \boldsymbol{u}_0$ weakly in H, but thanks to the continuity of $\boldsymbol{u}$ in H we obtain the strong convergence (2.5c).

Step 6. Attainment of the constitutive equation. The aim is to show that $(\boldsymbol{J}, \nabla \boldsymbol{u}) \in \mathcal{A}$ almost everywhere in Q, which is equivalent to showing $\boldsymbol{G}(\boldsymbol{J}, \nabla \boldsymbol{u}) = \boldsymbol{0}$ almost everywhere in Q. Towards this goal, we need to verify the assumption (4.6) of Lemma 4.5, i.e., we need to prove that for all $t \in (0,T)$,

$$\limsup_{\varepsilon\to 0_+} \int_{Q_t} \boldsymbol{J}^\varepsilon : \nabla \boldsymbol{u}^\varepsilon \,\mathrm{d}x\,\mathrm{d}\tau \le \int_{Q_t} \boldsymbol{J} : \nabla \boldsymbol{u} \,\mathrm{d}x\,\mathrm{d}\tau. \tag{5.15}$$

Indeed, having (5.15), Lemma 4.5 implies that $(\boldsymbol{J}, \nabla \boldsymbol{u}) \in \mathcal{A}$ almost everywhere in Q_t and that $\boldsymbol{J}^\varepsilon : \boldsymbol{D}^\varepsilon \rightharpoonup \boldsymbol{J} : \nabla \boldsymbol{u}$ weakly in $L^1(Q_t)$. Thus, we obtained the desired result on Q_t for every $t \in (0,T)$, therefore also on Q.

The relation (5.15) is achieved by the standard energy and weak lower semicontinuity techniques used in parabolic systems and for the sake of completeness, we provide the proof also here. In (5.2), we set $\boldsymbol{\varphi} = \boldsymbol{u}^\varepsilon$ and integrate the result over $(0,t)$ for $t \in (0,T)$. We obtain

$$\int_{Q_t} \boldsymbol{J}^\varepsilon : \nabla \boldsymbol{u}^\varepsilon \,\mathrm{d}x\,\mathrm{d}\tau = \int_0^t \langle \boldsymbol{f}, \boldsymbol{u}^\varepsilon \rangle_{V_\mu} \,\mathrm{d}\tau + \frac{1}{2}\|\boldsymbol{u}_0\|_H^2 - \frac{1}{2}\|\boldsymbol{u}^\varepsilon(t)\|_H^2.$$

Applying then the limes superior as $\varepsilon to 0_+$ and using the weak convergence of $\boldsymbol{u}^\varepsilon$ in $L^\mu(0,T;V_\mu)$ we conclude that

$$\limsup_{\varepsilon\to 0_+} \int_{Q_t} \boldsymbol{J}^\varepsilon : \nabla \boldsymbol{u}^\varepsilon \,\mathrm{d}x\,\mathrm{d}\tau = \int_0^t \langle \boldsymbol{f}, \boldsymbol{u} \rangle_{V_\mu} \,\mathrm{d}\tau + \frac{1}{2}\|\boldsymbol{u}_0\|_H^2 - \frac{1}{2}\liminf_{\varepsilon\to 0_+}\|\boldsymbol{u}^\varepsilon(t)\|_H^2. \tag{5.16}$$

On the other hand, setting $\boldsymbol{w} = \boldsymbol{u}$ in (5.13) (we already have the right duality pairings to do so) and integrating it over $(0,t)$ we arrive at

$$\int_{Q_t} \boldsymbol{J} : \nabla \boldsymbol{u} \,\mathrm{d}x\,\mathrm{d}\tau = \int_0^t \langle \boldsymbol{f}, \boldsymbol{u} \rangle_{V_\mu} \,\mathrm{d}\tau + \frac{1}{2}\|\boldsymbol{u}_0\|_H^2 - \frac{1}{2}\|\boldsymbol{u}(t)\|_H^2. \tag{5.17}$$

Subtracting (5.17) from (5.16) gives

$$\limsup_{\varepsilon\to 0_+} \int_{Q_t} \boldsymbol{J}^\varepsilon : \nabla \boldsymbol{u}^\varepsilon \,\mathrm{d}x\,\mathrm{d}\tau = \frac{1}{2}\|\boldsymbol{u}(t)\|_H^2 - \frac{1}{2}\liminf_{\varepsilon\to 0_+}\|\boldsymbol{u}^\varepsilon(t)\|_H^2 + \int_{Q_t} \boldsymbol{J} : \nabla \boldsymbol{u} \,\mathrm{d}x\,\mathrm{d}\tau. \tag{5.18}$$

That is, to verify (5.15), it remains to show that

$$\|\boldsymbol{u}(t)\|_H^2 \le \liminf_{\varepsilon\to 0_+}\|\boldsymbol{u}^\varepsilon(t)\|_H^2. \tag{5.19}$$

In case V_p is compactly embedded into H (i.e., if $p > 2d/(d+2)$), the above relation is for a.a. $t \in (0,T)$ consequence of the convergence results (5.11) and the Aubin–Lions compactness lemma. Therefore, if $p > 2d/(d+2)$, (5.15) holds for almost all time, which is sufficient for finishing the proof. Nevertheless, in case we do not have V_p compactly embedded into H, we proceed slightly differently and even more, we obtain (5.19) for all $t \in (0,T)$ (instead of for almost all t).

Let $0 < \delta \ll T$ and take $\boldsymbol{\varphi} = \boldsymbol{u}^\varepsilon$ in (5.2). Integrating the result over $(t, t+\delta)$ and using then the integration by parts applied to the first term, we obtain

$$\frac{1}{2}\|\boldsymbol{u}^\varepsilon(t+\delta)\|_H^2 + \int_t^{t+\delta}\int_\Omega \boldsymbol{J}^\varepsilon : \nabla \boldsymbol{u}^\varepsilon \,\mathrm{d}x\,\mathrm{d}\tau = \int_t^{t+\delta} \langle \boldsymbol{f}, \boldsymbol{u}^\varepsilon \rangle_{V_\mu} \,\mathrm{d}\tau + \frac{1}{2}\|\boldsymbol{u}^\varepsilon(t)\|_H^2.$$

As $\mathcal{A}_\varepsilon^\varepsilon$ is monotone, we observe that $\boldsymbol{J}^\varepsilon : \nabla \boldsymbol{u}^\varepsilon \ge 0$ and we neglect the corresponding term receiving the inequality. Integrating it with respect to δ over $(0,\gamma)$ for $0 < \gamma \ll 1$, we arrive at

$$\frac{1}{2}\int_0^\gamma \|\boldsymbol{u}^\varepsilon(t+\delta)\|_H^2 \,\mathrm{d}\delta - \int_0^\gamma \int_t^{t+\delta} \langle \boldsymbol{f}, \boldsymbol{u}^\varepsilon \rangle_{V_\mu} \,\mathrm{d}\tau\,\mathrm{d}\delta \le \frac{\gamma}{2}\|\boldsymbol{u}^\varepsilon(t)\|_H^2.$$

Taking limes inferior as $\varepsilon \to 0_+$ and using, on the left hand side, the established weak convergence for $\boldsymbol{u}^\varepsilon$ and the weak lower semicontinuity of the norm, followed by the multiplication of the achieved inequality by $\frac{2}{\gamma}$, we get

$$\frac{1}{\gamma}\int_0^\gamma \|\boldsymbol{u}(t+\delta)\|_H^2 \,\mathrm{d}\delta - \frac{2}{\gamma}\int_0^\gamma \int_t^{t+\delta} \langle \boldsymbol{f}, \boldsymbol{u}\rangle_{V_\mu} \,\mathrm{d}\tau \,\mathrm{d}\delta \le \liminf_{\varepsilon\to 0_+} \|\boldsymbol{u}^\varepsilon(t)\|_H^2.$$

Finally, letting $\gamma \to 0_+$, using the continuity of $\boldsymbol{u}$ in H and the fact that the duality between $\boldsymbol{f}$ and $\boldsymbol{u}$ is well-defined, we obtain (5.19).

Step 7. Uniqueness of $\boldsymbol{u}$. Let $(\boldsymbol{u}_1, \boldsymbol{J}_1)$ and $(\boldsymbol{u}_2, \boldsymbol{J}_2)$ be two solutions to the problem (1.6). If we subtract their weak formulations, we obtain

$$\langle \partial_t(\boldsymbol{u}_1 - \boldsymbol{u}_2), \boldsymbol{\varphi}\rangle_{V_p} + \int_\Omega (\boldsymbol{J}_1 - \boldsymbol{J}_2) : \nabla\boldsymbol{\varphi} \,\mathrm{d}x = 0.$$

Next, we set $\boldsymbol{\varphi} := (\boldsymbol{u}_1 - \boldsymbol{u}_2)$ to get

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_{V_p} + \int_\Omega (\boldsymbol{J}_1 - \boldsymbol{J}_2) : (\nabla\boldsymbol{u}_1 - \nabla\boldsymbol{u}_2) \,\mathrm{d}x = 0,$$

however, due to monotonicity of the graph $\mathcal{A}$, we obtain that each term is equal to zero. Finally, after integration over time $(0,t)$ for every $t \in (0,T)$, we use that both solutions satisfy the same initial condition and conclude that $\boldsymbol{u}_1(t) = \boldsymbol{u}_2(t)$ in V_p for every $t \in (0,T)$.

Step 8. Sketch of the proof of Theorem 2.1 for $f \in L^{p'}(0,T;V_p^*)$. Since V_μ^* is dense in V_p^* for $\mu' \ge p'$, for a given $f \in L^{p'}(0,T;V_p^*)$ there exist $\{\boldsymbol{f}^m\}_{m\in\mathbb{N}} \subset L^{\mu'}(0,T;V_\mu^*)$ satisfying

$$\boldsymbol{f}^m \to \boldsymbol{f} \text{ in } L^{p'}(0,T;V_p^*).$$

For each $m \in \mathbb{N}$ we consider a solution $(\boldsymbol{u}^m, \boldsymbol{J}^m)$ of the problem (1.6) in the sense of Theorem 2.1 with the right-hand side $\boldsymbol{f}^m$. Then, we proceed as in Steps 2–7, i.e., we derive the uniform estimates for $\{(\boldsymbol{u}^m, \boldsymbol{J}^m)\}_{m\in\mathbb{N}}$, find appropriate weak limits $(\boldsymbol{u}, \boldsymbol{J})$, and study the limit as $m \to \infty$. This is all done in the same way (or slightly simpler) than above. In particular, we use Lemma 4.7 for verification that the couple $(\boldsymbol{J}, \nabla\boldsymbol{u})$ belongs to $\mathcal{A}$. Note that $\mathcal{A}$ remains unchanged throughout this step.

Appendix A. Prototypic examples

Following the aim to clarify the conditions (g1)–(g4) formulated in the introductory section and to fix the notation involved in their descriptions, we consider five examples of the implicit constitutive equations $\boldsymbol{g}(\boldsymbol{j}, \boldsymbol{d}) = \boldsymbol{0}$ and show that they satisfy (g1)–(g4).

Example A.1. *The linear case* $\boldsymbol{j} = \boldsymbol{d}$*, i.e.,*

$$\boldsymbol{g}(\boldsymbol{j}, \boldsymbol{d}) = \boldsymbol{j} - \boldsymbol{d}.$$

Validity of (g1)–(g4) for Example A.1. To show that Example A.1 satisfies (g1)–(g4), we first notice that $\boldsymbol{g}(\boldsymbol{0},\boldsymbol{0}) = \boldsymbol{0}$, $\boldsymbol{g}_{\boldsymbol{j}}(\boldsymbol{j},\boldsymbol{d}) = \boldsymbol{I}$, $\boldsymbol{g}_{\boldsymbol{d}}(\boldsymbol{j},\boldsymbol{d}) = -\boldsymbol{I}$, $\boldsymbol{g}_{\boldsymbol{j}}(\boldsymbol{j},\boldsymbol{d}) - \boldsymbol{g}_{\boldsymbol{d}}(\boldsymbol{j},\boldsymbol{d}) = 2\boldsymbol{I}$ and, by a simple computation, $\boldsymbol{g}_{\boldsymbol{d}}(\boldsymbol{j},\boldsymbol{d})(\boldsymbol{g}_{\boldsymbol{j}}(\boldsymbol{j},\boldsymbol{d}))^T = -\boldsymbol{I}$, and therefore (g1) and (g2) obviously hold. Furthermore,

$$\boldsymbol{g}(\boldsymbol{j},\boldsymbol{d}) \cdot \boldsymbol{j} = |\boldsymbol{j}|^2 - \boldsymbol{j}\cdot\boldsymbol{d} \quad \text{and} \quad \boldsymbol{g}(\boldsymbol{j},\boldsymbol{d}) \cdot \boldsymbol{d} = \boldsymbol{j}\cdot\boldsymbol{d} - |\boldsymbol{d}|^2.$$

Consequently, for a fixed $\boldsymbol{d} \in \mathbb{R}^d$,

$$\lim_{|\boldsymbol{j}|\to\infty} \boldsymbol{g}(\boldsymbol{j},\boldsymbol{d}) \cdot \boldsymbol{j} = \infty$$

and, for a fixed $\boldsymbol{j} \in \mathbb{R}^d$,

$$\lim_{|\boldsymbol{d}|\to\infty} \boldsymbol{g}(\boldsymbol{j},\boldsymbol{d}) \cdot \boldsymbol{d} = -\infty,$$

which proves (g3). Finally,

$$\boldsymbol{j}\cdot\boldsymbol{d} = \frac{1}{2}(\boldsymbol{j}\cdot\boldsymbol{d}+\boldsymbol{j}\cdot\boldsymbol{d}) = \frac{1}{2}|\boldsymbol{j}|^2+\frac{1}{2}|\boldsymbol{d}|^2,$$

where, in the last equality, we inserted first $\boldsymbol{d}$ for $\boldsymbol{j}$ and then $\boldsymbol{j}$ for $\boldsymbol{d}$ (using $\boldsymbol{0}=\boldsymbol{g}(\boldsymbol{j},\boldsymbol{d})=\boldsymbol{j}-\boldsymbol{d}$). Hence, (g4) holds.

Note that setting $\boldsymbol{g}(\boldsymbol{j},\boldsymbol{d})=\boldsymbol{d}-\boldsymbol{j}$ leads to sign changes in the all identities in (g2) and (g3) except the last identity in (g2) that remains unchanged.

□

Example A.2. *We consider* $\boldsymbol{d}=(1+|\boldsymbol{j}|^2)^{\frac{p'-2}{2}}\boldsymbol{j}$ *with* $p'=p/(p-1)$, $p\in(1,\infty)$. *This means that*

$$\boldsymbol{g}(\boldsymbol{j},\boldsymbol{d}) = (1+|\boldsymbol{j}|^2)^{\frac{p'-2}{2}}\boldsymbol{j}-\boldsymbol{d}.$$

Validity of (g1)–(g4) for Example A.2. Clearly, $\boldsymbol{g}(\boldsymbol{0},\boldsymbol{0})=\boldsymbol{0}$, $\boldsymbol{g}_{\boldsymbol{d}}(\boldsymbol{j},\boldsymbol{d})=-\boldsymbol{I}$ and

$$\boldsymbol{g}_{\boldsymbol{j}}(\boldsymbol{j},\boldsymbol{d}) = (1+|\boldsymbol{j}|^2)^{\frac{p'-2}{2}}\boldsymbol{I}+(p'-2)(1+|\boldsymbol{j}|^2)^{\frac{p'-4}{2}}\boldsymbol{j}\otimes\boldsymbol{j},$$

where $(\boldsymbol{j}\otimes\boldsymbol{j})_{k\ell}:=j_kj_\ell$. Hence, for all $\boldsymbol{x}\in\mathbb{R}^d$, one has

$$\begin{aligned}\boldsymbol{g}_{\boldsymbol{j}}(\boldsymbol{j},\boldsymbol{d})\boldsymbol{x}\cdot\boldsymbol{x} &= (1+|\boldsymbol{j}|^2)^{\frac{p'-4}{2}}\left((1+|\boldsymbol{j}|^2)|\boldsymbol{x}|^2+(p'-2)(\boldsymbol{j}\cdot\boldsymbol{x})^2\right)\\ &\geq \begin{cases}(1+|\boldsymbol{j}|^2)^{\frac{p'-2}{2}}|\boldsymbol{x}|^2 & \text{for } p'\geq 2,\\ (p'-1)(1+|\boldsymbol{j}|^2)^{\frac{p'-2}{2}}|\boldsymbol{x}|^2 & \text{for } p'\in(1,2).\end{cases}\end{aligned}$$

Hence $\boldsymbol{g}_{\boldsymbol{j}}(\boldsymbol{j},\boldsymbol{d})>0$. Consequently, $\boldsymbol{g}_{\boldsymbol{j}}(\boldsymbol{j},\boldsymbol{d})-\boldsymbol{g}_{\boldsymbol{d}}(\boldsymbol{j},\boldsymbol{d})>0$ and $\boldsymbol{g}_{\boldsymbol{d}}(\boldsymbol{j},\boldsymbol{d})(\boldsymbol{g}_{\boldsymbol{j}}(\boldsymbol{j},\boldsymbol{d}))^T<0$ and the validity of (g1) and (g2) is verified. Furthermore, for any $\boldsymbol{d}\in\mathbb{R}^d$, recalling that $p'>1$, we obtain that

$$\boldsymbol{g}(\boldsymbol{j},\boldsymbol{d})\cdot\boldsymbol{j} = (1+|\boldsymbol{j}|^2)^{\frac{p'-2}{2}}|\boldsymbol{j}|^2-\boldsymbol{j}\cdot\boldsymbol{d}\to\infty \quad \text{as } |\boldsymbol{j}|\to\infty.$$

Similarly, for any $\boldsymbol{j}\in\mathbb{R}^d$,

$$\boldsymbol{g}(\boldsymbol{j},\boldsymbol{d})\cdot\boldsymbol{d} = (1+|\boldsymbol{j}|^2)^{\frac{p'-2}{2}}\boldsymbol{j}\cdot\boldsymbol{d}-|\boldsymbol{d}|^2\to-\infty \quad \text{as } |\boldsymbol{d}|\to\infty,$$

and (g3) holds. Finally,

$$\boldsymbol{j}\cdot\boldsymbol{d} = (1+|\boldsymbol{j}|^2)^{\frac{p'-2}{2}}|\boldsymbol{j}|^2 \geq \begin{cases}|\boldsymbol{j}|^{p'-2}|\boldsymbol{j}|^2=|\boldsymbol{j}|^{p'} & \text{if } p'\geq 2,\\ 2^{\frac{p'-2}{2}}|\boldsymbol{j}|^{p'-2}|\boldsymbol{j}|^2=2^{\frac{p'-2}{2}}|\boldsymbol{j}|^{p'} & \text{if } p'\in(1,2) \text{ and } |\boldsymbol{j}|\geq 1,\\ 2^{\frac{p'-2}{2}}|\boldsymbol{j}|^2\geq 2^{\frac{p'-2}{2}}|\boldsymbol{j}|^{p'}-c_0 & \text{if } p'\in(1,2) \text{ and } |\boldsymbol{j}|<1,\end{cases}$$

where, in the last step, we used Young's inequality $|\boldsymbol{j}|^{p'}\leq|\boldsymbol{j}|^2+c$. Since

$$|\boldsymbol{d}|^2 = (1+|\boldsymbol{j}|^2)^{\frac{p'-2}{2}}|\boldsymbol{j}|^2\leq(1+|\boldsymbol{j}|^2)^{p'-1} \quad\Longrightarrow\quad 1+|\boldsymbol{j}|^2\geq|\boldsymbol{d}|^{\frac{2}{p'-1}},$$

we observe that

$$\begin{aligned}\boldsymbol{j}\cdot\boldsymbol{d} &= (1+|\boldsymbol{j}|^2)^{\frac{p'-2}{2}}|\boldsymbol{j}|^2 = (1+|\boldsymbol{j}|^2)^{\frac{p'-2}{2}}(1+|\boldsymbol{j}|^2-1)\\ &= (1+|\boldsymbol{j}|^2)^{\frac{p'}{2}}-(1+|\boldsymbol{j}|^2)^{\frac{p'-2}{2}}\\ &\geq \begin{cases}\frac{1}{2}|\boldsymbol{d}|^p-c & \text{if } p'\geq 2,\\ |\boldsymbol{d}|^p-1 & \text{if } p'\in(1,2)\,,\end{cases}\end{aligned}$$

where in the last step we used Young's inequality $(1+|\boldsymbol{j}|^2)^{\frac{p'-2}{2}}\leq\frac{1}{2}(1+|\boldsymbol{j}|^2)^2+c$ for $p'\geq 2$ and the fact that $(1+|\boldsymbol{j}|^2)^{\frac{p'-2}{2}}\leq 1$ for $p'\in(1,2\rangle$. The last two formulae imply (g4). □

Example A.3. *For* $\boldsymbol{d} = (|\boldsymbol{j}| - \sigma_*)^+ \frac{\boldsymbol{j}}{|\boldsymbol{j}|}$, *we set*[9]

$$\boldsymbol{g}(\boldsymbol{j}, \boldsymbol{d}) = (|\boldsymbol{j}| - \sigma_*)^+ \frac{\boldsymbol{j}}{|\boldsymbol{j}|} - \boldsymbol{d}.$$

Validity of (g1)–(g4) for Example A.3. Clearly, $\boldsymbol{g}$ is continuous on $\mathbb{R}^d \times \mathbb{R}^d$, Lipschitz continuous almost everywhere in $\mathbb{R}^d \times \mathbb{R}^d$, $\boldsymbol{g}(\boldsymbol{0}, \boldsymbol{0}) = \boldsymbol{0}$, $\boldsymbol{g}_{\boldsymbol{d}}(\boldsymbol{j}, \boldsymbol{d}) = -\boldsymbol{I}$ and

$$\boldsymbol{g}_{\boldsymbol{j}}(\boldsymbol{j}, \boldsymbol{d}) = \frac{(|\boldsymbol{j}| - \sigma_*)^+}{|\boldsymbol{j}|} \boldsymbol{I} + \chi_{\{|\boldsymbol{j}| > \sigma_*\}} \frac{\boldsymbol{j} \otimes \boldsymbol{j}}{|\boldsymbol{j}|^2} - (|\boldsymbol{j}| - \sigma_*)^+ \frac{\boldsymbol{j} \otimes \boldsymbol{j}}{|\boldsymbol{j}|^3},$$

where χ_U denotes the characteristic function of $U \subset \mathbb{R}^d$. The last identity leads to

$$\boldsymbol{g}_{\boldsymbol{j}}(\boldsymbol{j}, \boldsymbol{d}) \boldsymbol{x} \cdot \boldsymbol{x} \geq \frac{(|\boldsymbol{j}| - \sigma_*)^+}{|\boldsymbol{j}|} \left(|\boldsymbol{x}|^2 - \frac{(\boldsymbol{j} \cdot \boldsymbol{x})^2}{|\boldsymbol{j}|^2} \right) + \chi_{\{|\boldsymbol{j}| > \sigma_*\}} \frac{(\boldsymbol{j} \cdot \boldsymbol{x})^2}{|\boldsymbol{j}|^2} \geq 0.$$

The above observations imply that $\boldsymbol{g}_{\boldsymbol{j}}(\boldsymbol{j}, \boldsymbol{d}) \geq 0$, $\boldsymbol{g}_{\boldsymbol{j}}(\boldsymbol{j}, \boldsymbol{d}) - \boldsymbol{g}_{\boldsymbol{d}}(\boldsymbol{j}, \boldsymbol{d}) > 0$ and $\boldsymbol{g}_{\boldsymbol{d}}(\boldsymbol{j}, \boldsymbol{d})(\boldsymbol{g}_{\boldsymbol{j}}(\boldsymbol{j}, \boldsymbol{d}))^T \geq 0$. Hence, (g1) and (g2) hold.

Next, it is easy to deduce that

$$\begin{aligned} \boldsymbol{g}(\boldsymbol{j}, \boldsymbol{d}) \cdot \boldsymbol{j} &= (|\boldsymbol{j}| - \sigma_*)^+ |\boldsymbol{j}| - \boldsymbol{d} \cdot \boldsymbol{j} \to \infty && \text{as } |\boldsymbol{j}| \to \infty, \\ \boldsymbol{g}(\boldsymbol{j}, \boldsymbol{d}) \cdot \boldsymbol{d} &= (|\boldsymbol{j}| - \sigma_*)^+ \frac{\boldsymbol{j} \cdot \boldsymbol{d}}{|\boldsymbol{j}|} - |\boldsymbol{d}|^2 \to -\infty && \text{as } |\boldsymbol{d}| \to \infty \end{aligned}$$

and consequently, (g3) follows. Finally, we observe that

$$\boldsymbol{j} \cdot \boldsymbol{d} = (|\boldsymbol{j}| - \sigma_*)^+ |\boldsymbol{j}| \geq \begin{cases} 0 \geq |\boldsymbol{j}|^2 - \sigma_*^2 & \text{if } |\boldsymbol{j}| \leq \sigma_*, \\ |\boldsymbol{j}|^2 - \sigma_* |\boldsymbol{j}| \geq \frac{1}{2} |\boldsymbol{j}|^2 - c & \text{if } |\boldsymbol{j}| > \sigma_*. \end{cases}$$

Since $|\boldsymbol{d}| = (|\boldsymbol{j}| - \sigma_*)^+$ and consequently $\boldsymbol{d} = \boldsymbol{0}$ if $|\boldsymbol{j}| \leq \sigma_*$ and $|\boldsymbol{j}| = |\boldsymbol{d}| + \sigma_*$ is $|\boldsymbol{j}| \geq \sigma_*$, we also get

$$\boldsymbol{j} \cdot \boldsymbol{d} = (|\boldsymbol{j}| - \sigma_*)^+ |\boldsymbol{j}| \geq \begin{cases} 0 = |\boldsymbol{d}|^2 & \text{if } |\boldsymbol{j}| \leq \sigma_*, \\ |\boldsymbol{d}| \left(|\boldsymbol{d}| + \sigma_* \right) \geq |\boldsymbol{d}|^2 - c_0 & \text{if } |\boldsymbol{j}| > \sigma_*. \end{cases}$$

This proves (g4). □

We end up this part by studying the models depicted in Figure 1.

Example A.4. *For* $a : [0, \infty] \to [0, 1]$ *defined through*

$$a(x) := \begin{cases} 1 & \text{for } x \in [0, \sqrt{2}/2], \\ \dfrac{\sqrt{2} - x}{x} & \text{for } x \in (\sqrt{2}/2, \sqrt{2}), \\ 0 & \text{for } x \geq \sqrt{2}, \end{cases} \tag{A.1}$$

we set

$$\boldsymbol{g}(\boldsymbol{j}, \boldsymbol{d}) := \boldsymbol{j} - \boldsymbol{d} - a \left(\frac{\sqrt{2} |\boldsymbol{j} + \boldsymbol{d}|}{2} \right) (\boldsymbol{j} + \boldsymbol{d}). \tag{A.2}$$

Then the null points of $\boldsymbol{g}$ *describes the right graph drawn in Figure 1, i.e.,*

$$|\boldsymbol{j}| \leq 1 \quad \text{if } \boldsymbol{d} = \boldsymbol{0} \qquad \text{and} \qquad \boldsymbol{j} = \max\{1, |\boldsymbol{d}|^{-1}\} \boldsymbol{d} \quad \text{if } \boldsymbol{d} \neq \boldsymbol{0}. \tag{A.3}$$

In addition, $\boldsymbol{g}$ *satisfies the assumptions (g1)-(g4) with* $p = 2$.

[9] Here and in what follows we tacitly assume a continuous extension at $\boldsymbol{j} = \boldsymbol{0}$, namely $\boldsymbol{d} = \boldsymbol{0}$ for $\boldsymbol{j} = \boldsymbol{0}$.

Verification of (A.3). Consider first $\boldsymbol{d} = \boldsymbol{0}$. Then it follows from (A.2) and $\boldsymbol{g}(\boldsymbol{j}, \boldsymbol{0}) = \boldsymbol{0}$ that

$$\boldsymbol{j} = a\left(\frac{\sqrt{2}|\boldsymbol{j}|}{2}\right)\boldsymbol{j}$$

Hence, either $\boldsymbol{j} = \boldsymbol{0}$ or

$$a\left(\frac{\sqrt{2}|\boldsymbol{j}|}{2}\right) = 1.$$

It however follows from the definition of a, see (A.1), that the second option is possible if and only if $|\boldsymbol{j}| \le 1$.

Next, let $\boldsymbol{d} \neq \boldsymbol{0}$. Then it follows from (A.2) that

$$\boldsymbol{g}(\boldsymbol{j}, \boldsymbol{d}) = \boldsymbol{0} \iff \left(1 - a\left(\frac{\sqrt{2}|\boldsymbol{j}+\boldsymbol{d}|}{2}\right)\right)\boldsymbol{j} = \left(1 + a\left(\frac{\sqrt{2}|\boldsymbol{j}+\boldsymbol{d}|}{2}\right)\right)\boldsymbol{d}, \tag{A.4}$$

and, as $\boldsymbol{d} \neq \boldsymbol{0}$, $\left(1 - a\left(\frac{\sqrt{2}|\boldsymbol{j}+\boldsymbol{d}|}{2}\right)\right)$ cannot be zero, which means that $|\boldsymbol{j} + \boldsymbol{d}| > 1$. Hence, the null points of $\boldsymbol{g}$ satisfy

$$\boldsymbol{j} = b\boldsymbol{d}.$$

The goal is to determine b. First, we observe from (A.4) and the definition of a that $\boldsymbol{j} = \boldsymbol{d}$ (and thus $b = 1$) if $|\boldsymbol{j} + \boldsymbol{d}| \ge 2$. It remains to show that $b = |\boldsymbol{d}|^{-1}$ if $1 < |\boldsymbol{j} + \boldsymbol{d}| < 2$. Inserting $\boldsymbol{j} = b\boldsymbol{d}$ into (A.2) we obtain

$$(b-1)\boldsymbol{d} = (1+b)a\left(\frac{\sqrt{2}(1+b)|\boldsymbol{d}|}{2}\right)\boldsymbol{d}$$

and consequently

$$(b-1) = (1+b)a\left(\frac{\sqrt{2}(1+b)|\boldsymbol{d}|}{2}\right).$$

In order to use the fact that $a(x)x = \sqrt{2} - x$ for $x \in (\sqrt{2}/2, \sqrt{2})$, we multiply the last equality by $\frac{\sqrt{2}}{2}|\boldsymbol{d}|$ and conclude that

$$\frac{\sqrt{2}}{2}(b-1)|\boldsymbol{d}| = \sqrt{2} - \frac{\sqrt{2}}{2}(1+b)|\boldsymbol{d}|,$$

which gives $b = |\boldsymbol{d}|^{-1}$.

Validity of (g1)–(g4) for Example A.4. Obviously, (g1) holds. Next, taking the scalar product of $\boldsymbol{g}(\boldsymbol{j}, \boldsymbol{d}) = \boldsymbol{0}$ first by $\boldsymbol{d}$ and then by $-\boldsymbol{j}$ and summing of the results, we obtain

$$\begin{aligned} 2\boldsymbol{j}\cdot\boldsymbol{d} &= |\boldsymbol{j}|^2 + |\boldsymbol{d}|^2 - a(\dots)|\boldsymbol{j}|^2 + a(\dots)|\boldsymbol{d}|^2 \\ &\ge |\boldsymbol{j}|^2 + |\boldsymbol{d}|^2 - a(\dots)|\boldsymbol{j}|^2 \\ &\ge \begin{cases} |\boldsymbol{j}|^2 + |\boldsymbol{d}|^2 & \text{if } |\boldsymbol{j}+\boldsymbol{d}| > 2, \\ |\boldsymbol{j}|^2 + |\boldsymbol{d}|^2 - C & \text{if } |\boldsymbol{j}+\boldsymbol{d}| \le 2, \end{cases} \end{aligned}$$

which gives (g4) with $p = 2$, and also (g3). It remains to show the validity of (g2). Note that

$$\boldsymbol{g}_{\boldsymbol{j}}(\boldsymbol{j}, \boldsymbol{d}) = \left(1 - a\left(\frac{\sqrt{2}|\boldsymbol{j}+\boldsymbol{d}|}{2}\right)\right)\boldsymbol{I} - a'\left(\frac{\sqrt{2}|\boldsymbol{j}+\boldsymbol{d}|}{2}\right)\frac{\sqrt{2}}{2}\frac{(\boldsymbol{j}+\boldsymbol{d})\otimes(\boldsymbol{j}+\boldsymbol{d})}{|\boldsymbol{j}+\boldsymbol{d}|}$$

$$\boldsymbol{g}_{\boldsymbol{d}}(\boldsymbol{j}, \boldsymbol{d}) = -\left(1 + a\left(\frac{\sqrt{2}|\boldsymbol{j}+\boldsymbol{d}|}{2}\right)\right)\boldsymbol{I} - a'\left(\frac{\sqrt{2}|\boldsymbol{j}+\boldsymbol{d}|}{2}\right)\frac{\sqrt{2}}{2}\frac{(\boldsymbol{j}+\boldsymbol{d})\otimes(\boldsymbol{j}+\boldsymbol{d})}{|\boldsymbol{j}+\boldsymbol{d}|}$$

Then, by using the definition of a, we observe that for arbitrary $\boldsymbol{x} \in \mathbb{R}^d$

$$\boldsymbol{g_j}(\boldsymbol{j},\boldsymbol{d})\boldsymbol{x}\cdot\boldsymbol{x} = \begin{cases} 0 & \text{for } |\boldsymbol{j}+\boldsymbol{d}| \leq 1, \\ |\boldsymbol{x}|^2 & \text{for } |\boldsymbol{j}+\boldsymbol{d}| \geq 2, \\ \left(2 - \dfrac{2}{|\boldsymbol{j}+\boldsymbol{d}|}\right)|\boldsymbol{x}|^2 + 2\dfrac{((\boldsymbol{j}+\boldsymbol{d})\cdot\boldsymbol{x})^2}{|\boldsymbol{j}+\boldsymbol{d}|^3} & \text{for } |\boldsymbol{j}+\boldsymbol{d}| \in (1,2) \end{cases}$$

$$\boldsymbol{g_d}(\boldsymbol{j},\boldsymbol{d})\boldsymbol{x}\cdot\boldsymbol{x} = \begin{cases} -2|\boldsymbol{x}|^2 & \text{for } |\boldsymbol{j}+\boldsymbol{d}| \leq 1, \\ -|\boldsymbol{x}|^2 & \text{for } |\boldsymbol{j}+\boldsymbol{d}| \geq 2, \\ -\dfrac{2|\boldsymbol{x}|^2}{|\boldsymbol{j}+\boldsymbol{d}|} + 2\dfrac{((\boldsymbol{j}+\boldsymbol{d}):\boldsymbol{x})^2}{|\boldsymbol{j}+\boldsymbol{d}|^3} & \text{for } |\boldsymbol{j}+\boldsymbol{d}| \in (1,2) \end{cases}$$

Thus, $\boldsymbol{g_j}(\boldsymbol{j},\boldsymbol{d}) \geq 0$ and $\boldsymbol{g_d}(\boldsymbol{j},\boldsymbol{d}) \leq 0$. In addition also $\boldsymbol{g_j}(\boldsymbol{j},\boldsymbol{d}) - \boldsymbol{g_d}(\boldsymbol{j},\boldsymbol{d}) > 0$. Finally, as $\boldsymbol{g_j}$ and $\boldsymbol{g_d}$ are symmetric, it follows directly from nonnegativity of $\boldsymbol{g_j}$ and nonpositivity of $\boldsymbol{g_d}$ that

$$\boldsymbol{g_j}(\boldsymbol{g_d})^T = \boldsymbol{g_j}\boldsymbol{g_d} \leq 0.$$

Thus (g2) holds. □

Example A.5. *Let $b : [0,\infty] \to \mathbb{R}$ be $\sqrt{2}$-periodic and satisfy*

$$b(x) := xa(x) \quad \textit{for } x \in [0,\sqrt{2}],$$

where a is defined in (A.1)*. Defining*

$$\tilde{a}(x) := \begin{cases} 0 & \textit{for } x = 0, \\ \dfrac{b(x)}{x} & \textit{for } x \in (0,\infty), \end{cases} \tag{A.5}$$

we set

$$\boldsymbol{g}(\boldsymbol{j},\boldsymbol{d}) := \boldsymbol{j} - \boldsymbol{d} - \tilde{a}\left(\frac{\sqrt{2}|\boldsymbol{j}+\boldsymbol{d}|}{2}\right)(\boldsymbol{j}+\boldsymbol{d}). \tag{A.6}$$

Then the null points of $\boldsymbol{g}$ describes the graph drawn left in Figure 1. In addition, $\boldsymbol{g}$ satisfies the assumptions (g1)-(g4) with $p = 2$.

We do not verify teh validity of (g1)–(g4) for the graph described by the null points of $\boldsymbol{g}$ defined in (A.6) as the proof is almost identical to the proof for Example A.4.

Appendix B. The Maxwell–Stefan system

Here, we consider the Maxwell–Stefan system given by (1.8). We omit the dependence of parameters on the solution itself and we just focus on the proof of the fulfilment of (G1)–(G3).

Example B.1. *For $d, N \in \mathbb{N}$, $N \geq 2$, consider*

$$(\boldsymbol{G}(\boldsymbol{J},\boldsymbol{D}))_{\nu i} = \sum_{\mu=1}^{N} (\mathbb{A}_{\nu\mu}(c_\mu \boldsymbol{J}_{\nu i} - c_\nu \boldsymbol{J}_{\mu i})) - \boldsymbol{D}_{\nu i}, \qquad i = 1,\dots,d;\ \nu = 1,\dots,N,$$

where $\mathbb{A}$ is a given symmetric matrix in $\mathbb{R}^{N\times N}$ fulfilling $\mathbb{A}_{\nu\mu} > 0$ for $\nu,\mu = 1,\dots,N$ and $\{c_\nu\}_{\nu=1}^N$ fulfil

$$c_\nu \in (0,1) \textit{ for all } \nu = 1,\dots,N \qquad \textit{and} \qquad \sum_{\nu=1}^{N} c_\nu = 1.$$

Validity of (G1)–(G3). We can evaluate

$$\begin{aligned}\frac{\partial(\boldsymbol{G}(\boldsymbol{J},\boldsymbol{D}))_{\nu i}}{\partial \boldsymbol{J}_{\mu j}} &= \delta_{ij}\left(\delta_{\nu\mu}\left(\sum_{\alpha=1}^{N}\mathbb{A}_{\nu\alpha}c_\alpha\right) - \mathbb{A}_{\nu\mu}c_\nu\right),\\ \frac{\partial(\boldsymbol{G}(\boldsymbol{J},\boldsymbol{D}))_{\nu i}}{\partial \boldsymbol{D}_{\mu j}} &= -\delta_{ij}\delta_{\nu\mu}.\end{aligned}$$

Then for arbitrary $\boldsymbol{B}\in\mathbb{R}^{N\times d}$, we have

$$\begin{aligned}\sum_{\nu,\mu=1}^{N}\sum_{i,j=1}^{d}\frac{\partial(\boldsymbol{G}(\boldsymbol{J},\boldsymbol{D}))_{\nu i}}{\partial \boldsymbol{D}_{\mu j}}\boldsymbol{B}_{\nu i}\boldsymbol{B}_{\mu j} &= -|\boldsymbol{B}|^2\le 0,\\ \sum_{\nu,\mu=1}^{N}\sum_{i,j=1}^{d}\frac{\partial(\boldsymbol{G}(\boldsymbol{J},\boldsymbol{D}))_{\nu i}}{\partial \boldsymbol{J}_{\mu j}}\boldsymbol{B}_{\nu i}\boldsymbol{B}_{\mu j} &= \sum_{\nu,\mu=1}^{N}\sum_{i=1}^{d}\boldsymbol{B}_{\nu i}\boldsymbol{B}_{\nu i}\mathbb{A}_{\nu\mu}c_\mu - \sum_{\nu,\mu=1}^{N}\sum_{i=1}^{d}\boldsymbol{B}_{\nu i}\boldsymbol{B}_{\mu i}\mathbb{A}_{\nu\mu}c_\nu.\end{aligned}$$

While the first inequality is exactly of the form we want, we focus on the second inequality. First, we can observe that the second identity can be rewritten into the form

$$\sum_{\nu,\mu=1}^{N}\sum_{i,j=1}^{d}\frac{\partial(\boldsymbol{G}(\boldsymbol{J},\boldsymbol{D}))_{\nu i}}{\partial \boldsymbol{J}_{\mu j}}\boldsymbol{B}_{\nu i}\boldsymbol{B}_{\mu j} = \sum_{i=1}^{d}\sum_{\nu,\mu=1}^{N}\boldsymbol{B}_{\nu i}\boldsymbol{B}_{\mu i}\mathbb{B}_{\nu\mu},$$

where $\mathbb{B}$ is a matrix given as

$$\mathbb{B}_{\nu\mu} := \begin{cases}\displaystyle\sum_{\alpha=1;\,\alpha\neq\nu}^{N}\mathbb{A}_{\nu\alpha}c_\alpha & \text{for } \nu=\mu,\\ -\,\mathbb{A}_{\nu\mu}c_\nu & \text{for } \nu\neq\mu.\end{cases}$$

Next, we can use [16, Lemma 2.1], where it is shown that the spectrum of $\mathbb{B}$ is nonnegative (but contains simple eigen-value 0) and consequently, it follows that $\frac{\partial(\boldsymbol{G}(\boldsymbol{J},\boldsymbol{D}))}{\partial\boldsymbol{J}}\ge 0$. Hence, we see that $\boldsymbol{G}$ satisfy (G1) and (G2). Also it is evident that it satisfies $(G3)_2$. However, since the spectrum of $\mathbb{B}$ also contains 0 it cannot satisfy (G4). Nevertheless, since for all null points we have that (note that all null points must satisfy $\sum_\nu \boldsymbol{D}_\nu = 0$)

$$\boldsymbol{J}:\boldsymbol{D} = \mathbb{B}\boldsymbol{J}:\boldsymbol{J},$$

it follows from the positivity of the spectrum of $\mathbb{B}$, except simple eigen-value zero, that for all $\boldsymbol{J}$ satisfying $\sum_\mu \boldsymbol{J}_\mu = 0$ there holds

$$\mathbb{B}\boldsymbol{J}:\boldsymbol{J}\ge c|\mathbb{B}\boldsymbol{J}|^2,$$

and consequently also,

$$\boldsymbol{J}:\boldsymbol{D}\ge c|\boldsymbol{D}|^2.$$

Hence, (G4) with $p=2$ is fulfilled on the range of $\mathbb{B}$, as also used in the analysis, see [16].

□

Appendix C. Solvability of (1.6) for Lipschitz continuous and uniformly monotone graphs

Here, we consider the following problem: for given $\Omega \subset \mathbb{R}^d$, $T > 0$, $\boldsymbol{f} : Q \to \mathbb{R}^N$ and $\boldsymbol{u}_o : \Omega \to \mathbb{R}^N$, find $(\boldsymbol{u}, \boldsymbol{J}) : Q \to \mathbb{R}^N \times \mathbb{R}^{N\times d}$ satisfying

$$\partial_t \boldsymbol{u} - \operatorname{div} \boldsymbol{J} = \boldsymbol{f} \qquad \text{in } Q, \tag{C.1a}$$

$$\boldsymbol{J} = \boldsymbol{J}^*(\nabla \boldsymbol{u}) \qquad \text{in } Q, \tag{C.1b}$$

$$\boldsymbol{u} = \boldsymbol{0} \qquad \text{on } \Sigma_D, \tag{C.1c}$$

$$\boldsymbol{J}\boldsymbol{n} = \boldsymbol{0} \qquad \text{on } \Sigma_N, \tag{C.1d}$$

$$\boldsymbol{u}(0,\cdot) = \boldsymbol{u}_0 \qquad \text{in } \Omega, \tag{C.1e}$$

where $\boldsymbol{J}^* : \mathbb{R}^{d\times N} \to \mathbb{R}^{d\times N}$ is a Lipschitz continuous and uniformly monotone single-valued mapping, which means that there are C_1, $C_2 > 0$ such that for all $\boldsymbol{D}_1$, $\boldsymbol{D}_2 \in \mathbb{R}^{d\times N}$

$$\begin{aligned} |\boldsymbol{J}^*(\boldsymbol{D}_1) - \boldsymbol{J}^*(\boldsymbol{D}_2)| &\le C_2 |\boldsymbol{D}_1 - \boldsymbol{D}_2|, \\ (\boldsymbol{J}^*(\boldsymbol{D}_1) - \boldsymbol{J}^*(\boldsymbol{D}_2)) : (\boldsymbol{D}_1 - \boldsymbol{D}_2) &\ge C_1 |\boldsymbol{D}_1 - \boldsymbol{D}_2|^2, \\ \boldsymbol{J}^*(\boldsymbol{0}) &= \boldsymbol{0}. \end{aligned} \tag{C.2}$$

Note that taking $\boldsymbol{D}_2 = \boldsymbol{0}$ and relabelling $\boldsymbol{D}_1$ by $\boldsymbol{D}$ in (C.2) we obtain

$$\boldsymbol{J}^*(\boldsymbol{D}) : \boldsymbol{D} \ge \frac{C_1}{2}|\boldsymbol{D}|^2 + \frac{C_1}{2C_2^2}|\boldsymbol{J}^*(\boldsymbol{D})|^2 \ge C\left(|\boldsymbol{D}|^2 + |\boldsymbol{J}|^2\right), \tag{C.3}$$

where we set $\boldsymbol{J} = \boldsymbol{J}^*(\boldsymbol{D})$ and $C := \min\{C_1/2, C_1/(2C_2^2)\}$. Consequently, the graph $\mathcal{A}$ defined through the relation

$$(\boldsymbol{J}, \nabla \boldsymbol{u}) \in \mathcal{A} \iff \boldsymbol{J} = \boldsymbol{J}^*(\nabla \boldsymbol{u}). \tag{C.4}$$

is Lipschitz continuous and uniformly monotone 2-coercive graph.

By the Faedo-Galerkin method, we establish the following well-posedness result.

Lemma C.1. *Let $\Omega \subset \mathbb{R}^d$ be a Lipschitz domain, $T > 0$, $\boldsymbol{f} \in L^2(0,T;V^*)$, $\boldsymbol{u}_0 \in H$ and $\boldsymbol{J}^*$ satisfy* (C.2)*. Then there exists a unique couple $(\boldsymbol{u}, \boldsymbol{J})$ such that*

$$\begin{aligned} \boldsymbol{u} &\in L^2(0,T;V) \cap \mathcal{C}([0,T];H), \\ \partial_t \boldsymbol{u} &\in L^2(0,T;V^*), \\ \boldsymbol{J} &\in L^2(Q;\mathbb{R}^{d\times N}), \end{aligned}$$

satisfying

$$\langle \partial_t \boldsymbol{u}, \boldsymbol{\varphi}\rangle_V + \int_\Omega \boldsymbol{J} : \nabla \boldsymbol{\varphi}\, dx = \langle \boldsymbol{f}, \boldsymbol{\varphi}\rangle_V \qquad \text{for a.a. } t \in (0,T) \text{ and for all } \boldsymbol{\varphi} \in V, \tag{C.5a}$$

$$\boldsymbol{J} = \boldsymbol{J}^*(\nabla \boldsymbol{u}) \qquad \text{almost everywhere in } Q, \tag{C.5b}$$

$$\lim_{t\to 0_+} \|\boldsymbol{u}(t) - \boldsymbol{u}_0\|_H = 0. \tag{C.5c}$$

Remark C.2. *Obviously, we could completely avoid using $\boldsymbol{J}$ in the formulation of Lemma C.1 and merely require that $\boldsymbol{u}$ fulfills, instead of* (C.5a)-(C.5b)*,*

$$\langle \partial_t \boldsymbol{u}, \boldsymbol{\varphi}\rangle_V + \int_\Omega \boldsymbol{J}^*(\nabla \boldsymbol{u}) : \nabla \boldsymbol{\varphi}\, dx = \langle \boldsymbol{f}, \boldsymbol{\varphi}\rangle_V \qquad \text{for a.a. } t \in (0,T) \text{ and for all } \boldsymbol{\varphi} \in V.$$

The formulation used in Lemma C.1 is more suitable for proving Theorem 2.1 in this text.

Proof. We follow the original Minty method, see [23], with small modifications adapted to our setting. The whole proof is split into several steps.

Step 1. Galerkin approximations. Let $\{\boldsymbol{w}_i\}_{i\in\mathbb{N}}$ and corresponding λ_i be the solutions of the eigenvalue problem $((\boldsymbol{w}_i,\boldsymbol{Z}))=\lambda_i(\boldsymbol{w}_i,\boldsymbol{Z})$ valid for all $\boldsymbol{Z}\in V$. Here, $((\cdot,\cdot))$ stands for the scalar product in V and $(\cdot,\cdot)$ is the scalar product in H, whereas the spaces V and H are defined in Section 2. Then $\{\boldsymbol{w}_i\}_{i\in\mathbb{N}}$ forms an orthogonal basis in V that is in addition orthonormal in H. Furthermore, the projection P^n of V to the linear hull of $\{\boldsymbol{w}_i\}_{i=1}^n$ defined by

$$P^n\boldsymbol{u}:=\sum_{i=1}^{n}(\boldsymbol{u},\boldsymbol{w}_i)_H\boldsymbol{w}_i \tag{C.6}$$

satisfy $\|P^n\boldsymbol{u}\|_H\le\|\boldsymbol{u}\|_H$ and $\|P^n\boldsymbol{u}\|_V\le\|\boldsymbol{u}\|_V$. See, for example [20, Section 6.4] for details.

For every $n\in\mathbb{N}$, we set

$$\boldsymbol{u}^n(t,\boldsymbol{x}):=\sum_{i=1}^{n}c_i^n(t)\boldsymbol{w}_i(\boldsymbol{x})\quad\text{for}\quad(t,\boldsymbol{x})\in Q, \tag{C.7}$$

where the functions $c_i^n(t)$ solve the following system of ordinary differential equations

$$(\partial_t\boldsymbol{u}^n,\boldsymbol{w}_i)_H+\int_\Omega\boldsymbol{J}^*(\nabla\boldsymbol{u}^n):\nabla\boldsymbol{w}_i\,\mathrm{d}x=\langle\boldsymbol{f},\boldsymbol{w}_i\rangle_V,\qquad i=1,\dots,n, \tag{C.8a}$$

with the initial conditions

$$c_i^n(0)=\int_\Omega\boldsymbol{u}_0\cdot\boldsymbol{w}_i\,\mathrm{d}x=(\boldsymbol{u}_0,\boldsymbol{w}_i)_H,\qquad i=1,\dots,n. \tag{C.8b}$$

Due to the Picard–Lindelöf theory, there exists a unique solution defined on an interval $[0,t^n)$. In virtue of the uniform estimates established in (C.10) below, one observes that $t^n\ge T$ for all n.

Step 2. Uniform estimates. Multiplying the i-th equation in (C.8a) by $c_i^n(t)$ and summing the result over $i=1,\dots,n$, we obtain

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{u}^n\|_H^2+\int_\Omega\boldsymbol{J}^*(\nabla\boldsymbol{u}^n):\nabla\boldsymbol{u}^n\,\mathrm{d}x=\langle\boldsymbol{f},\boldsymbol{u}^n\rangle_V. \tag{C.9}$$

Then, by means of Hölder's and Young's inequalities and (C.3), followed by the integration over $(0,t)$, we conclude, using also the assumption on data $\boldsymbol{u}_0$ and $\boldsymbol{f}$, that

(C.10)

$$\sup_{t\in(0,T)}\|\boldsymbol{u}^n(t)\|_H^2+\int_0^T\|\boldsymbol{u}^n\|_V^2+\|\boldsymbol{J}^*(\nabla\boldsymbol{u}^n)\|_{L^2(\Omega)}^2\,\mathrm{d}t\le C\left(\int_0^T\|\boldsymbol{f}\|_{V^*}^2\,\mathrm{d}t+\|\boldsymbol{u}_0\|_H^2\right)\le C.$$

This implies the following n-independent estimate

$$\|\boldsymbol{u}^n\|_{L^2(0,T;V)\cap L^\infty(0,T;H)}+\|\boldsymbol{J}^*(\nabla\boldsymbol{u}^n)\|_{L^2(Q)}\le C\quad\text{uniformly with respect to } n\in\mathbb{N}. \tag{C.11}$$

Furthermore, for any $\boldsymbol{\varphi}\in V$, we obtain from (C.8a)

$$\langle\partial_t\boldsymbol{u}^n,\boldsymbol{\varphi}\rangle_V=(\partial_t\boldsymbol{u}^n,P^n\boldsymbol{\varphi})_H=\int_\Omega\partial_t\boldsymbol{u}^n\cdot(P^n\boldsymbol{\varphi})\,\mathrm{d}x=-\int_\Omega\boldsymbol{J}^*(\nabla\boldsymbol{u}^n):\nabla(P^n\boldsymbol{\varphi})\,\mathrm{d}x+\langle\boldsymbol{f},P^n\boldsymbol{\varphi}\rangle_V.$$

The standard duality and scalar product estimates together with (C.11) and the continuity of P^n mentioned above imply that

$$\int_0^T\|\partial_t\boldsymbol{u}^n\|_{V^*}^2\,\mathrm{d}t\le C\quad\text{uniformly with respect to } n\in\mathbb{N}. \tag{C.12}$$

Step 3. Limit $n \to \infty$. By virtue of the uniform estimates (C.11) and (C.12), reflexivity of spaces V and V^* and the Aubin–Lions lemma, there exist (not relabelled) subsequences and functions $\boldsymbol{u}$ and $\boldsymbol{J}$ such that, for $n \to \infty$,

$$\boldsymbol{u}^n \rightharpoonup^* \boldsymbol{u} \qquad \text{weakly}^* \text{ in } L^\infty(0,T;H), \tag{C.13a}$$

$$\boldsymbol{u}^n \rightharpoonup \boldsymbol{u} \qquad \text{weakly in } L^2(0,T;V), \tag{C.13b}$$

$$\partial_t \boldsymbol{u}^n \rightharpoonup \partial_t \boldsymbol{u} \qquad \text{weakly in } L^2(0,T;V^*), \tag{C.13c}$$

$$\boldsymbol{u}^n \to \boldsymbol{u} \qquad \text{strongly in } L^2(0,T;H), \tag{C.13d}$$

$$\boldsymbol{J}^*(\nabla \boldsymbol{u}^n) \rightharpoonup \boldsymbol{J} \qquad \text{weakly in } L^2(Q;\mathbb{R}^{d\times N}). \tag{C.13e}$$

For any $\xi \in C^1(0,T)$ and $\boldsymbol{\varphi} \in V$, multiplying the i-th equation by $\xi(\boldsymbol{\varphi}, \boldsymbol{w}_i)_H$, summing the result over $i = 1, \dots, k$ for $k \le n$ and integrating then the outcome over $(0,T)$, we get, for every $k = 1, \dots, n$.

$$\int_0^T \left(\partial_t \boldsymbol{u}^n, \xi P^k \boldsymbol{\varphi}\right)_H \,\mathrm{d}t + \int_Q \boldsymbol{J}^*(\nabla \boldsymbol{u}^n) : \nabla(P^k \boldsymbol{\varphi})\xi \,\mathrm{d}x\,\mathrm{d}t = \int_0^T \langle \boldsymbol{f}, \xi P^k \boldsymbol{\varphi}\rangle_V \,\mathrm{d}t.$$

Using the convergence results (C.13), we can easily take the limit for $n \to \infty$. Since the limit terms hold for any smooth ξ, we obtain

$$\langle \partial_t \boldsymbol{u}, P^k \boldsymbol{\varphi}\rangle_V + \int_\Omega \boldsymbol{J} : \nabla(P^k \boldsymbol{\varphi}) \,\mathrm{d}x = \langle \boldsymbol{f}, P^k \boldsymbol{\varphi}\rangle_V \qquad \text{for a.a. } t \in (0,T) \text{ and for all } k \in \mathbb{N}.$$

As $P^k \boldsymbol{\varphi} \to \boldsymbol{\varphi}$ in V as $k \to \infty$, we arrive at the weak formulation (C.5a).

Step 4. Attainment of the initial datum. We first notice that it follows from $\boldsymbol{u} \in L^2(0,T;V)$ and $\partial_t \boldsymbol{u} \in L^2(0,T;V^*)$ that $\boldsymbol{u} \in \mathcal{C}([0,T];H)$. Hence

$$\boldsymbol{u}(t) \to \boldsymbol{u}(0) \text{ strongly in } H \text{ as } t \to 0_+. \tag{C.14}$$

To prove (C.5c), it is then enough to show that

$$\boldsymbol{u}(t) \rightharpoonup \boldsymbol{u}_0 \quad \text{weakly in } H \text{ as } t \to 0_+. \tag{C.15}$$

Towards this goal, let $0 < \varepsilon \ll 1$ and $t \in (0, T-\varepsilon)$. Recalling the definition of an auxiliary η in (5.14), multiplying (C.8a) by such an η and integrating the result with respect to $\tau \in (0,T)$, we obtain, for every $i = 1, \dots, n$,

$$\int_0^T (\partial_t \boldsymbol{u}^n, \boldsymbol{w}_i)_H \eta \,\mathrm{d}\tau + \int_Q \boldsymbol{J}^*(\nabla \boldsymbol{u}^n) : \nabla \boldsymbol{w}_i \eta \,\mathrm{d}x\,\mathrm{d}\tau = \int_0^T \langle \boldsymbol{f}, \boldsymbol{w}_i\rangle_V \eta \,\mathrm{d}\tau.$$

Integration by parts in the first term (using $\eta(T) = 0$) then leads to

$$-\int_0^T (\boldsymbol{u}^n, \boldsymbol{w}_i)_H \eta' \,\mathrm{d}\tau + \int_Q \boldsymbol{J}^*(\nabla \boldsymbol{u}^n) : \nabla \boldsymbol{w}_i \eta \,\mathrm{d}x\,\mathrm{d}\tau = \int_0^T \langle \boldsymbol{f}, \boldsymbol{w}_i\rangle_V \eta \,\mathrm{d}\tau + (P^n \boldsymbol{u}_0, \boldsymbol{w}_i)_H \eta(0).$$

Applying the weak convergence results established in (C.13) as well as the convergence of the projection P^n as $n \to \infty$ we observe, for any $i \in \mathbb{N}$, that

$$-\int_0^T (\boldsymbol{u}, \boldsymbol{w}_i)_H \eta' \,\mathrm{d}\tau + \int_Q \boldsymbol{J} : \nabla \boldsymbol{w}_i \eta \,\mathrm{d}x\,\mathrm{d}\tau = \int_0^T \langle \boldsymbol{f}, \boldsymbol{w}_i\rangle_V \eta \,\mathrm{d}\tau + (\boldsymbol{u}_0, \boldsymbol{w}_i)_H \eta(0).$$

The incorporation of the properties of η, namely $\eta(\tau) = 1$ for $\tau \in [0,t)$, $\eta(\tau) = 0$ for $\tau \in (t+\varepsilon, T]$, and $\eta'(\tau) = -\frac{1}{\varepsilon}$ for $\tau \in (t, t+\varepsilon)$, then yields

$$\frac{1}{\varepsilon}\int_t^{t+\varepsilon} (\boldsymbol{u}, \boldsymbol{w}_i)_H \,\mathrm{d}\tau + \int_{Q_{t+\varepsilon}} \boldsymbol{J} : \nabla \boldsymbol{w}_i \eta \,\mathrm{d}x\,\mathrm{d}\tau = \int_0^{t+\varepsilon} \langle \boldsymbol{f}, \boldsymbol{w}_i\rangle_V \eta \,\mathrm{d}\tau + (\boldsymbol{u}_0, \boldsymbol{w}_i)_H.$$

Finally, we let $\varepsilon \to 0_+$. In the first term, the integrand is well-defined (in fact, $\boldsymbol{u} \in \mathcal{C}([0,T];H)$), and the term converges to $(\boldsymbol{u}(t), \boldsymbol{w}_i)_H$. In the other terms, due to their integrability, we can take the limit as $\varepsilon \to 0_+$ together with $t \to 0_+$ and arrive at

$$\lim_{t\to 0_+} (\boldsymbol{u}(t), \boldsymbol{w}_i)_H = (\boldsymbol{u}_0, \boldsymbol{w}_i)_H.$$

Since $\{\boldsymbol{w}_i\}_{i\in\mathbb{N}}$ forms a basis in H, (C.15) and then also (C.5c) are proved.

Step 5. Attainment of the constitutive equation. It remains to show (C.3). To do so, we multiply (C.9) by piece-wise linear $\eta(\tau)$ defined in (5.14) and integrate the result over $(0,T)$. This yields

$$\int_{Q_{t+\varepsilon}} \boldsymbol{J}^*(\nabla \boldsymbol{u}^n) : \nabla \boldsymbol{u}^n \eta \,\mathrm{d}x\,\mathrm{d}\tau = \int_0^{t+\varepsilon} \langle \boldsymbol{f}, \boldsymbol{u}^n\rangle_V \eta \,\mathrm{d}\tau + \frac{1}{2}\|P^n \boldsymbol{u}_0\|_H^2 - \frac{1}{2\varepsilon}\int_t^{t+\varepsilon} (\boldsymbol{u}^n, \boldsymbol{u}^n)_H \,\mathrm{d}\tau.$$

Since $\boldsymbol{J}^*(\boldsymbol{0}) = \boldsymbol{0}$ and $\boldsymbol{J}^*(\cdot)$ is monotone, we have, for every $n \in \mathbb{N}$,

$$\boldsymbol{J}^*(\nabla \boldsymbol{u}^n) : \nabla \boldsymbol{u}^n \geq 0.$$

Therefore, as $\eta \equiv 1$ in Q_t,

$$\begin{aligned}
\limsup_{n\to\infty} \int_{Q_t} \boldsymbol{J}^*(\nabla \boldsymbol{u}^n) : \nabla \boldsymbol{u}^n \,\mathrm{d}x\,\mathrm{d}\tau &\leq \limsup_{n\to\infty} \int_{Q_{t+\varepsilon}} \boldsymbol{J}^*(\nabla \boldsymbol{u}^n) : \nabla \boldsymbol{u}^n \eta \,\mathrm{d}x\,\mathrm{d}\tau \\
&= \limsup_{n\to\infty} \int_0^{t+\varepsilon} \langle \boldsymbol{f}, \boldsymbol{u}^n\rangle_V \eta \,\mathrm{d}\tau + \frac{1}{2}\|P^n \boldsymbol{u}_0\|_H^2 - \liminf_{n\to\infty} \frac{1}{2\varepsilon}\int_t^{t+\varepsilon} (\boldsymbol{u}^n, \boldsymbol{u}^n)_H \,\mathrm{d}\tau \\
&\leq \int_0^{t+\varepsilon} \langle \boldsymbol{f}, \boldsymbol{u}\rangle_V \eta \,\mathrm{d}\tau + \frac{1}{2}\|\boldsymbol{u}_0\|_H^2 - \frac{1}{2\varepsilon}\int_t^{t+\varepsilon} (\boldsymbol{u}, \boldsymbol{u})_H \,\mathrm{d}\tau,
\end{aligned}$$

where we used the results established in (C.13) and the weak lower-semicontinuity of the norm. Letting $\varepsilon \to 0_+$, we note that the left hand side is independent of ε and the all quantities on the right-hand side are well-defined for such limit (since $\boldsymbol{u} \in \mathcal{C}([0,T];H)$). We thus obtain, for an arbitrary $t \in (0,T)$,

$$\limsup_{n\to\infty} \int_{Q_t} \boldsymbol{J}^*(\nabla \boldsymbol{u}^n) : \nabla \boldsymbol{u}^n \,\mathrm{d}x\,\mathrm{d}\tau \leq \int_0^t \langle \boldsymbol{f}, \boldsymbol{u}\rangle_V \,\mathrm{d}\tau + \frac{1}{2}\left(\|\boldsymbol{u}_0\|_H^2 - \|\boldsymbol{u}(t)\|_H^2\right). \tag{C.16}$$

Now, we set $\boldsymbol{\varphi} := \boldsymbol{u}$ in (C.5a) and integrate the result over time interval $(0,t)$. Using the integration by parts formulae (thanks to the fact that we have the Gelfand triplet) and (C.5c), we get

$$\begin{aligned}
\int_{Q_t} \boldsymbol{J}^*(\nabla \boldsymbol{u}) : \nabla \boldsymbol{u} \,\mathrm{d}x\,\mathrm{d}\tau &= \int_0^t \langle \boldsymbol{f}, \boldsymbol{u}\rangle_V - \langle \partial_t \boldsymbol{u}, \boldsymbol{u}\rangle_V \,\mathrm{d}\tau \\
&= \int_0^t \langle \boldsymbol{f}, \boldsymbol{u}\rangle_V \,\mathrm{d}\tau + \frac{1}{2}\left(\|\boldsymbol{u}_0\|_H^2 - \|\boldsymbol{u}(t)\|_H^2\right).
\end{aligned} \tag{C.17}$$

Once comparing (C.16) and (C.17), we obtain

$$\limsup_{n\to\infty} \int_{Q_t} \boldsymbol{J}^*(\nabla \boldsymbol{u}^n) : \nabla \boldsymbol{u}^n \,\mathrm{d}x\,\mathrm{d}\tau \leq \int_{Q_t} \boldsymbol{J}^*(\nabla \boldsymbol{u}) : \nabla \boldsymbol{u} \,\mathrm{d}x\,\mathrm{d}\tau. \tag{C.18}$$

Now, let $\boldsymbol{W} \in L^2(0,T;L^2(\Omega))$ be arbitrary, then

$$\begin{aligned}
0 &\leq \int_{Q_t} (\boldsymbol{J}^*(\nabla \boldsymbol{u}^n) - \boldsymbol{J}^*(\boldsymbol{W})) : (\nabla \boldsymbol{u}^n - \boldsymbol{W}) \,\mathrm{d}x\,\mathrm{d}\tau \\
&= \int_{Q_t} \boldsymbol{J}^*(\nabla \boldsymbol{u}^n) : \nabla \boldsymbol{u}^n \,\mathrm{d}x\,\mathrm{d}\tau - \int_{Q_t} \boldsymbol{J}^*(\nabla \boldsymbol{u}^n) : \boldsymbol{W} + \boldsymbol{J}^*(\boldsymbol{W}) : (\nabla \boldsymbol{u}^n - \boldsymbol{W}) \,\mathrm{d}x\,\mathrm{d}\tau.
\end{aligned}$$

Letting $n \to \infty$, using the estimate (C.18) and the weak convergence results given in (C.13), we obtain

$$0 \leq \int_{Q_t} (\boldsymbol{J}^*(\nabla \boldsymbol{u}) - \boldsymbol{J}^*(\boldsymbol{W})) : (\nabla \boldsymbol{u} - \boldsymbol{W}) \, \mathrm{d}x \, \mathrm{d}\tau.$$

Finally, setting particularly $\boldsymbol{W} := \nabla \boldsymbol{u} \pm \varepsilon \boldsymbol{Z}$, dividing the result by ε and let $\varepsilon \to 0_+$ (at this point we use the continuity of the Lipschitz continuous single-valued mapping $\boldsymbol{J}^*$), we obtain, for arbitrary $\boldsymbol{Z}$,

$$0 \leq \int_{Q_t} (\boldsymbol{J} - \boldsymbol{J}^*(\nabla \boldsymbol{u})) : \boldsymbol{Z} \, \mathrm{d}x \, \mathrm{d}\tau, \tag{C.19}$$

which implies that $\boldsymbol{J} = \boldsymbol{J}^*(\nabla \boldsymbol{u})$ in Q_t for any $t \in (0, T)$.

Step 6. Uniqueness. Let $(\boldsymbol{u}_1, \boldsymbol{J}_1)$ and $(\boldsymbol{u}_2, \boldsymbol{J}_2)$ be two different weak solutions to (C.1) corresponding to the same set of data. Subtracting their weak formulations and inserting for $\boldsymbol{J}_1$ and $\boldsymbol{J}_2$, we obtain

$$\langle \partial_t (\boldsymbol{u}_1 - \boldsymbol{u}_2), \boldsymbol{\varphi} \rangle_V + \int_\Omega (\boldsymbol{J}^*(\nabla \boldsymbol{u}_1) - \boldsymbol{J}^*(\nabla \boldsymbol{u}_2)) : \nabla \boldsymbol{\varphi} \, \mathrm{d}x = 0 \qquad \text{for all } \boldsymbol{\varphi} \in V.$$

Taking $\boldsymbol{\varphi} := (\boldsymbol{u}_1(t, \cdot) - \boldsymbol{u}_2(t, \cdot))$, we get

$$\frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} \|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_H^2 + \int_\Omega (\boldsymbol{J}^*(\nabla \boldsymbol{u}_1) - \boldsymbol{J}^*(\nabla \boldsymbol{u}_2)) : (\nabla \boldsymbol{u}_1 - \nabla \boldsymbol{u}_2) \, \mathrm{d}x = 0,$$

which, due to the uniform monotonicity of $\boldsymbol{J}^* = \boldsymbol{J}^*(\nabla \boldsymbol{u})$, after integration over $(0, t)$ for an arbitrary $t \in (0, T)$, leads to

$$\|\boldsymbol{u}_1(t) - \boldsymbol{u}_2(t)\|_H^2 \leq \|\boldsymbol{u}_1(0) - \boldsymbol{u}_2(0)\|_H^2 = 0.$$

Necessarily, $\boldsymbol{u}_1(t) = \boldsymbol{u}_2(t)$ in V for almost every $t \in (0, T)$, and obviously, $\boldsymbol{J}_1 = \boldsymbol{J}^*(\nabla \boldsymbol{u}_1) = \boldsymbol{J}^*(\nabla \boldsymbol{u}_2) = \boldsymbol{J}_2$. □

References

[1] A. Abbatiello and E. Feireisl, *On a class of generalized solutions to equations describing incompressible viscous fluids*, Ann. Mat. Pura Appl. (4) **199** (2020), no. 3, 1183–1195.

[2] G. Alberti and L. Ambrosio, *A geometrical approach to monotone functions in* $\mathbf{R}^n$, Math. Z. **230** (1999), no. 2, 259–316.

[3] J. Blechta, J. Málek, and K. R. Rajagopal, *On the classification of incompressible fluids and a mathematical analysis of the equations that govern their motion*, SIAM J. Math. Anal. **52** (2020), no. 2, 1232–1289.

[4] D. Bothe, *On the Maxwell-Stefan approach to multicomponent diffusion*, Parabolic problems, Progr. Nonlinear Differential Equations Appl., vol. 80, Birkhäuser/Springer Basel AG, Basel, 2011, pp. 81–93.

[5] D. Bothe and P.-E. Druet, *On the structure of continuum thermodynamical diffusion fluxes – A novel closure scheme and its relation to the Maxwell–Stefan and the Fick–Onsager approach*, arXiv:2008.05327.

[6] H. Brezis, M. G. Crandall, and A. Pazy, *Perturbations of nonlinear maximal monotone sets in Banach space*, Comm. Pure Appl. Math. **23** (1970), 123–144.

[7] M. Bulíček, P. Gwiazda, J. Málek, K. R. Rajagopal, and A. Świerczewska Gwiazda, *On flows of fluids described by an implicit constitutive equation characterized by a maximal monotone graph*, Mathematical aspects of fluid mechanics, London Math. Soc. Lecture Note Ser., vol. 402, Cambridge Univ. Press, Cambridge, 2012, pp. 23–51.

[8] M. Bulíček, P. Gwiazda, J. Málek, and A. Świerczewska Gwiazda, *On steady flows of incompressible fluids with implicit power-law-like rheology*, Adv. Calc. Var. **2** (2009), no. 2, 109–136.

[9] ———, *On unsteady flows of implicitly constituted incompressible fluids*, SIAM J. Math. Anal. **44** (2012), no. 4, 2756–2801.

[10] M. Bulíček, J. Málek, K. R. Rajagopal, and E. Süli, *On elastic solids with limiting small strain: modelling and analysis*, EMS Surv. Math. Sci. **1** (2014), no. 2, 283–332.

[11] M. Bulíček, J. Málek, and J. Žabenský, *A generalization of the Darcy-Forchheimer equation involving an implicit, pressure-dependent relation between the drag force and the velocity*, J. Math. Anal. Appl. **424** (2015), no. 1, 785–801.

[12] L. Diening, Ch. Kreuzer, and E. Süli, *Finite element approximation of steady flows of incompressible fluids with implicit power-law-like rheology*, SIAM J. Numer. Anal. **51** (2013), no. 2, 984–1015.
[13] P. E. Farrell, P. A. Gazca-Orozco, and E. Süli, *Numerical analysis of unsteady implicitly constituted incompressible fluids: 3-field formulation*, SIAM J. Numer. Anal. **58** (2020), no. 1, 757–787.
[14] V. Giovangigli, *Multicomponent flow modeling*, Modeling and Simulation in Science, Engineering and Technology, Birkhäuser Boston, Inc., Boston, MA, 1999.
[15] A. Janečka, J. Málek, V. Průša, and G. Tierra, *Numerical scheme for simulation of transient flows of non-Newtonian fluids characterised by a non-monotone relation between the symmetric part of the velocity gradient and the Cauchy stress tensor*, Acta Mech. **230** (2019), no. 3, 729–747.
[16] A. Jüngel and I. V. Stelzer, *Existence analysis of Maxwell-Stefan systems for multicomponent mixtures*, SIAM J. Math. Anal. **45** (2013), no. 4, 2421–2440.
[17] Ch. Kreuzer and E. Süli, *Adaptive finite element approximation of steady flows of incompressible fluids with implicit power-law-like rheology*, ESAIM Math. Model. Numer. Anal. **50** (2016), no. 5, 1333–1369.
[18] O. A. Ladyzhenskaya, V. A. Solonnikov, and N. N. Ural'ceva, *Linear and quasilinear equations of parabolic type*, Translated from the Russian by S. Smith. Translations of Mathematical Monographs, Vol. 23, American Mathematical Society, Providence, R.I., 1968.
[19] Ch. Le Roux and K. R. Rajagopal, *Shear flows of a new class of power-law fluids*, Appl. Math. **58** (2013), no. 2, 153–177.
[20] J. Málek, J. Nečas, M. Rokyta, and M. Růžička, *Weak and measure-valued solutions to evolutionary PDEs*, Applied Mathematics and Mathematical Computation, vol. 13, Chapman & Hall, London, 1996.
[21] J. Málek and V. Průša, *Derivation of equations for continuum mechanics and thermodynamics of fluids*, Handbook of mathematical analysis in mechanics of viscous fluids, Springer, Cham, 2018, pp. 3–72.
[22] J. Málek and Z. Strakoš, *Preconditioning and the conjugate gradient method in the context of solving PDEs*, SIAM Spotlights, vol. 1, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2015.
[23] G. J. Minty, *Monotone (nonlinear) operators in Hilbert space*, Duke Math. J. **29** (1962), 341–346.
[24] K. R. Rajagopal, *On implicit constitutive theories*, Appl. Math. **48** (2003), 279–319.
[25] ———, *On implicit constitutive theories for fluids*, J. Fluid Mech. **550** (2006), 243–249 (English).
[26] K. R. Rajagopal and G. Saccomandi, *A novel approach to the description of constitutive relations*, Frontiers in Materials **3** (2016), 36.
[27] E. Süli and T. Tscherpel, *Fully discrete finite element approximation of unsteady flows of implicitly constituted incompressible fluids*, IMA J. Numer. Anal. **40** (2020), no. 2, 801–849.

Charles University, Faculty of Mathematics and Physics, Mathematical Institute, Sokolovská 83, 186 75, Prague, Czech Republic
Email address: mbul8060@karlin.mff.cuni.cz

Charles University, Faculty of Mathematics and Physics, Mathematical Institute, Sokolovská 83, 186 75, Prague, Czech Republic
Email address: malek@karlin.mff.cuni.cz

Institute for Analysis and Scientific Computing, Vienna University of Technology, Wiedner Hauptstr. 8-10, 1040 Vienna, Austria
Email address: erika.maringova@tuwien.ac.at