# UNIFIED ANALYSIS OF ALGORITHMS FOR EQUILIBRIUM, NON-EQUILIBRIUM, AND HYSTERESIS MODELS OF PHASE TRANSITION IN PERMAFROST

MALGORZATA PESZYNSKA AND NICHOLAS SLUGG

Abstract. In this paper we consider a nonlinear partial differential equation describing heat flow with ice-water phase transition in permafrost soils. Such models and their numerical approximations have been well explored in the applications literature. In this paper we describe a new direction in which the allow relaxation and hysteresis of the phase transition which introduce additional nonlinear terms and complications for the analysis. We present numerical algorithms as well as analysis of the well-posedness and convergence of the fully implicit iterative schemes. The analysis we propose handles the equilibrium, non-equilibrium, and hysteresis cases in a unified way. We also illustrate with numerical examples for a model ODE and PDE.

## 1. Introduction

In this paper we study algorithms for approximation of solutions to a nonlinear heat equation modeling the thermal processes of thawing and freezing in a soil. Such processes are important for modeling in permafrost regions (e.g., in the Arctic) which respond to the daily and seasonal variation of temperature at the ground surface. The specific challenge we consider in this paper is that the phase change processes (from liquid to ice and vice-versa) need not be instantaneous. Rather, they are assumed to follow non-equilibrium (also known as kinetic or relaxation laws) or hysteretic relationships.

The models we consider are nonlinear, and the appropriate numerical schemes for the approximation of solutions already for the equilibrium case present interesting challenges well documented in our previous work [3, 25]. In the paper we propose numerical discretization algorithms extending those for the equilibrium relationships to the cases of non-equilibrium and hysteretic relationships. We also analyze them within a common framework. In particular, we provide results on the well-posedness of the discrete schemes as well as conditions on the convergence of iterative algorithms used to solve the nonlinear problem.

Specifically, we consider a family of nonlinear evolution problems in an open bounded domain $\Omega \subset \mathbb{R}^d$

$$\partial_t(c(u)+\chi) - \nabla\cdot(k(u)\nabla u) = f(x,t),\ x\in\Omega, t>0. \tag{1}$$

The unknown $u$ in (2) represents the temperature of thawing or freezing soil, i.e, undergoing phase transition. The term $c(u)$ is associated with variable heat capacity, and the coefficient $k(u)$ represents heat conductivity, and is a symmetric uniformly positive definite diffusivity coefficient. One also needs initial conditions at $t=0$ imposed on $c(u)+\chi$ and boundary conditions imposed on the boundary $\partial\Omega$ of $\Omega$ which is assumed to be Lipschitz [24].

To complete (1) we need to define the relationship between $\chi$ and $u$. In the model (1) $\chi$ is associated with the phase transition and describes the heat amount necessary for thawing,

so that $c(u)+\chi(u)$ represent the total energy density. In this paper we consider exactly one of the relationships coupling $\chi$ to $u$ discussed below.

$$\text{(2a)} \qquad \text{(EQ)} \quad \chi = F(u),$$

$$\text{(2b)} \qquad \text{(KIN)} \quad \partial_t \chi + B(\chi - F(u)) = 0,$$

$$\text{(2c)} \qquad \text{(HYST)} \quad \partial_t \chi + \mathcal{C}(\chi - F(u)) \ni 0.$$

In these models $F$ is a bounded nonnegative function which is monotone nondecreasing and smooth except at $u = 0$ where it quickly grows to 1.

The choices in (2) describe a phase transition that can occur instantaneously (in equilibrium) as in (2a), or slowly. The non-equilibrium (kinetic) evolution model (KIN) (2b) allows $\chi$ to evolve towards (2a) with some rate $B > 0$. Such non-equilibrium models are common in the applications involving phase transition and chemical reactions; see our work in [9] and [15]. With the hysteretic model (2c) when $u$ increases, $\chi$ follows a different path than at freezing when $u$ decreases; this well-known phenomenon is due to the physics of nucleation [24, 7]. The model represents evolution under constraints given by $\mathcal{C}$, a multivalued monotone constraint graph.

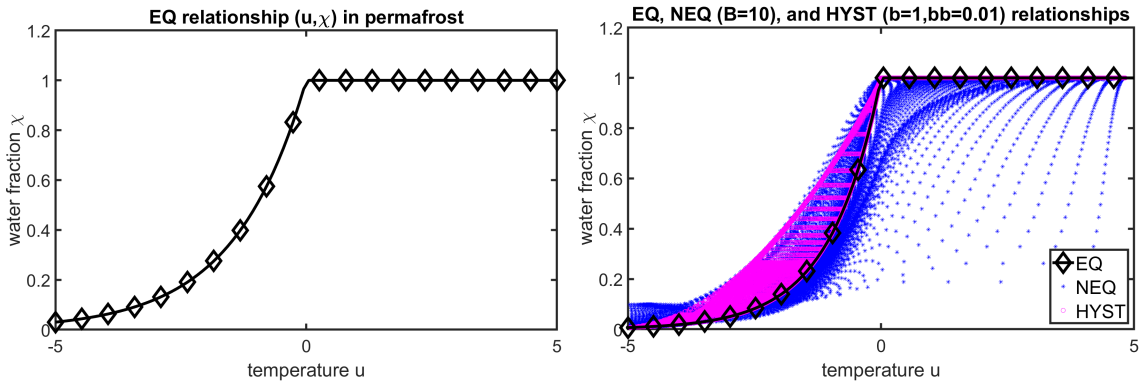


FIGURE 1. Illustration of of relationships (2). On left, we plot of $\chi = F(u)$ in (2a). On right, we plot the same relationship $\chi = F(u)$ again as well as a scatter plot of the results $(u, \chi)$ of simulations+ for (2b) and (2c). Details of this example are given in Section 5.3.

We aim to approximate the solutions $(u, \chi)$ satisfying (1) and one of (2a)–(2c), supplemented with appropriate boundary and initial conditions. For simplicity, we use finite differences (FD) in space, and in our examples we use specifically cell-centered FD called CCFD. In time, we use fully implicit first order scheme. The fully discrete model reads: at every time step $t_n$, solve

$$\text{(3)} \qquad \tfrac{1}{\tau}(C(U^n) - C(U^{n-1})) + \tfrac{1}{\tau}(\Upsilon^n - \Upsilon^{n-1}) + A(U^n)U^n = f^n,$$

where $U^n = (U_j^n)_j, \Upsilon^n = (\Upsilon_j^n)_j$ contain the approximations to $u(x_j, t_n), \chi(x_j, t_n)$ at spatial points $x_j \in \Omega$ of a uniform rectangular grid over $\Omega$. Here $C(U) = (c(U_j^n))_j$ is a vector, and $A(U)$ is a matrix which is symmetric positive definite for any $U$, and $A(U)U$ is an approximation to $-\nabla \cdot k(u)\nabla u$ under suitable boundary conditions, with $A(U)$ deriving its properties from $k = k(u)$. The system (3) is complemented with appropriately discretized equations from (2). We refer to [3, 25] for details on $C, A$.

The resulting system is nonlinear and must be solved by some iteration, while its well-posedness is only guaranteed under some assumptions. The presence of multiple nonlinearities including $F(\cdot), c(\cdot), k(\cdot)$ as well as the complications due to non-equilibrium models makes this task challenging to implement and to analyze. The analysis in this paper addresses these difficulties in a unified way. We also present simulations for a few selected examples of (2) and discuss robustness of the solver.

We believe the results to be presented are new, even if they are related to some work in the literature on other applications. In particular, we are not aware of any computational models or rigorous work for permafrost applications with hysteresis. However, there is ample work on analysis and abstract framework on numerical schemes for the Stefan problem in equilibrium for Galerkin FEM and with relaxation, e.g., see [24, 5]. In our own prior work [3, 25] we used CCFD but worked only with the equilibrium problem and considered only the time-lagging of $A(U)$.

**Outline.** The outline of the paper is as follows. In Section 2 we develop some notation and recall technical preliminaries. In Section 3 we provide details of the model (2) and its numerical discretization. In Section 4 we provide analysis of the well-posedness of the discrete schemes and of the iterative schemes. In Section 5 we illustrate the results with examples for the simplified ODE counterpart of the model and for the PDE. In Section 6 we summarize and outline future and current work.

## 2. Preliminaries and technical results

In this Section we recall some material from the literature needed for the model developments and for the proofs to follow. In particular, we provide information on the evolution problems under constraints.

### 2.1. Operator equation in a Hilbert space.

In the paper we will consider a Hilbert space $V$, with inner product $\langle \cdot, \cdot \rangle$, and norm $||\cdot||$. When $V = \mathbb{R}$, we will use $\langle x, y \rangle = xy, ||x|| = |x|$. On $V = \mathbb{R}^n$ we will use the Euclidean inner product and norm.

Let $I$ be an identity operator on $V$, $I(v) = v$.

For vector spaces $V, W$, and any function $T : V \to W$ we will denote by $L_T$ its Lipschitz constant:

$$||T(v_1) - T(v_2)||_W \leq L_T ||v_1 - v_2||_V. \tag{4}$$

We say that $T$ is a contraction if $L_T < 1$.

Let $v_1, v_2 \in V$ be arbitrary. We say that $T : V \to V$ is monotone if

$$\langle T(v_1) - T(v_2), v_1 - v_2 \rangle \geq 0. \tag{5}$$

We also say that $T$ is strongly monotone if there is some $c_0 > 0$:

$$\langle T(v_1) - T(v_2), v_1 - v_2 \rangle \geq c_0 ||v_1 - v_2||^2. \tag{6}$$

The following equation is of interest in this paper

$$T(u) = b. \tag{7}$$

Of interest is existence and uniqueness of the solutions to (7). Let $T : V \to V$.

**Theorem 1** ([2], Thm. 5.1.4, p211)**.** *Let $T : V \to V$ be (i) strongly monotone and (ii) Lipschitz continuous. Then for any $b \in V$ there is a unique solution $u \in V$ to* (7)*. Moreover, if $T(u_1) = b_1, T(u_2) = b_2$ then $||u_1 - u_2|| \leq \frac{1}{c_0}||b_1 - b_2||$..*

The following result is also well known, and we will apply it in $K = V$.

**Theorem 2** ([2], Thm. .5.1.3, p209)**.** *Let $T : K \to K$ be a contraction where $K \subset V$ is a nonempty closed set. Then for any $v^0$ the sequence*

$$v^m = T(v^{m-1}), m = 1, 2, \dots \tag{8}$$

*converges to the unique fixed point $v$ of $T$ defined by $Tv = v$.*

We also paraphrase the "local convergence" result for Newton iteration.

**Theorem 3** ([2], Thm. .5.4.1, p236)**.** *Let $T : V \to V$ be Fréchet differentiable and let $T(U_*) = 0$. Let also (*) $T'(U_*)^{-1}$ exist be a continuous linear map, and let $T'(U)$ be locally Lipschitz continuous in some neighborhood of $U_*$. Then the Newton's iteration: given $U^{(o)}$, iterate*

$$U^{(m)} = U^{(m-1)} - T'(U^{(m-1)})^{-1}T(U^{(m-1)}) \tag{9}$$

*converges provided (**) $U^{(0)}$ is close enough to $U_*$.*

In many instances the functions we work with do not have continuous or Lipschitz derivatives. Thus we replace $T'$ in (9) by some selection from its subgradient $\partial T$. Weaker hypotheses in [22] allow to work in the *semismooth* framework. We paraphrase some results below as they are used for $T(U_*) = 0$.

**Theorem 4** ([22], Prop. 2.12, p29, Prop. 2.26, p35)**.** *Let $V = \mathbb{R}^M$, and let $T : V \to V$ be continuous and piecewise smooth. Let also (*) all subgradients from $\partial T(U_*)$ be nonsingular in some neighborhood of $U_*$ and let $U^{(0)}$ be sufficiently close to $U_*$. Then the Newton's iteration* (9) *converges.*

2.2. **Evolution equation on $V$ and finite difference approximation.** We consider a nonlinear evolution equation on $V$ with some $G : V \to V$

$$\frac{d}{dt}u(t) + G(u) = 0, u(0) = u_{init} \in V. \tag{10}$$

Let $\tau > 0$ be the time step, and let $t_n = n\tau$. The fully implicit finite difference approximation of (10) is

$$\frac{1}{\tau}(U^n - U^{n-1}) + G(U^n) = 0 \tag{11}$$

where $U^n \approx u(t_n)$. The scheme (11) is of first order in $\tau$ if $G(\cdot)$ is smooth enough.

It follows from Theorem 1 that (11) rewritten as

$$U^n + \tau G(U^n) = U^{n-1} \tag{12}$$

has a unique solution at each time step $t_n$ under some assumptions.

**Lemma 1.** *Let (i) $G$ be Lipschitz and monotone, or (ii) $G$ be Lipschitz and $\tau$ small enough. Then* (12) *has a unique solution.*

*Proof.* For the case (i) the proof follows immediately by Theorem 1 since then $I + \tau G$ is strongly monotone. Consider (ii) when $G$ is not monotone. For sufficiently small $\tau$, $(I+\tau G)$ is strongly monotone: expanding $\langle (I + \tau G)(u_1) - (I + \tau G)(u_2), u_1 - u_2\rangle$

$$\langle u_1 - u_2, u_1 - u_2\rangle + \tau\langle G(u_1) - G(u_2), u_1 - u_2\rangle \\ = ||u_1 - u_2||^2 - \tau\langle G(u_1) - G(u_2), u_2 - u_1\rangle. \tag{13}$$

The second term can be bounded from below by $-\tau L_G||u_1 - u_2||^2$ since $\langle G(u_1) - G(u_2), u_1 - u_2\rangle \leq L_G||u_1 - u_2||^2$. With small $\tau$ so $c_T = 1-\tau L_G > 0$ and $(I+\tau G)$ is strongly monotone. □

Formally we write the solution to (11) as $U^n = (I + \tau G)^{-1}U^{n-1}$, even if we do not form $(I + \tau G)^{-1}$ explicitly. In addition, if $G$ is nonlinear, solving (11) requires an iteration. For example, we can set up fixed point iteration at every time step using previous time step

$$U^{n,(m)} + \tau G(U^{n,(m-1)}) = U^{n-1}, m = 1, 2, ...; \;\; U^{n,(0)} = U^{n-1}. \tag{14}$$

By Theorem 2, the iteration converges if the map $T$ in $U^{n,(m)} = T(U^{n,(m-1)}) = U^{n-1} - \tau G(U^{n,(m-1)})$ is a contraction which follows if $\tau L_G$ is small enough.

We can also set-up Newton's iteration (9) for $T(U) = U + \tau G(U) - U^{n-1}$. Here, we might need to require that $T'(U) = I + \tau G'(U)$ exists and satisfies at least (*) from Theorem 3. Usually if $\tau$ is small enough, the initial guess from the previous time step solution $U^{n,(0)} = U^{n-1}$ suffices for (**).

The proofs in this paper will be similar to those we outline for (12).

2.3. **Evolution equation under constraints on $\mathbb{R}$ and generalized play model of hysteresis.** The results below are needed to develop the hysteresis models. We introduce the notion of constraint graphs on $\mathbb{R}$, and of evolution under constraints, well developed in [4]. We refer also to [21], and to our recent work on adsorption with hysteresis in [12, 17, 16, 18] using linear play, nonlinear play, and generalized play models of hysteresis for the adsorption applications. In this paper we use generalized play models.

2.3.1. *Constraint graph on* $\mathbb{R}$. Let $a \leq b \in \mathbb{R}$, and consider the non-empty closed interval $[a, b] \subset \mathbb{R}$. If $a < b$ we define the set-valued relationship

$$\mathcal{C}(a, b; s) = \begin{cases} (-\infty, 0] \text{ if } s = a, \\ \{0\} \text{ if } a < s < b, \\ [0, \infty) \text{ if } s = b. \end{cases} \tag{15}$$

This definition indicates that $\mathcal{C}(a, b; \cdot)$ is set-valued with domain $\mathrm{Dom}(\mathcal{C}(a, b; \cdot)) = [a, b]$; its graph is denoted by $\mathcal{C}(a, b) = \{a\} \times (-\infty, 0] \cup (a, b) \times \{0\} \cup \{b\} \times [0, \infty]$. The relation $\mathcal{C}(a, b; )$ is *monotone:* $\langle v_1 - v_2, w_1 - w_2\rangle_V \geq 0$ for any $v_j, w_j \in \mathbb{R}$: $w_j \in \mathcal{C}(v_j)$. It is also *maximal monotone* because the range $\mathrm{Rg}(\mathcal{C}+I) = \mathbb{R}$, and $\mathrm{Rg}(\lambda\mathcal{C}+I) = \mathbb{R}$ for all $\lambda > 0$. The resolvent $R(a, b; s) = (I + \tau\mathcal{C}(a, b; \cdot))^{-1}(s)$ is a function defined on $\mathbb{R}$ which can be written explicitly

$$R(a, b; s) = (I + \tau\mathcal{C}(a, b; \cdot))^{-1} = \begin{cases} a \text{ if } s \leq a, \\ s \text{ if } a < s < b, \\ b \text{ if } s \geq b, \end{cases} \qquad s \in \mathbb{R}, \tag{16}$$

and is independent of $\tau > 0$. The resolvent $R(a, b;)$ is used to determine the solution to a stationary problem

$$v + \tau \mathcal{C}(a, b; v) \ni f \in \mathbb{R}, \tag{17}$$

in which the symbol $\ni$ indicates that the left hand side of (17) is a set. But once $v = R(a, b; f)$ is found, the selection $\mathcal{C}(a, b; v) \ni c^* = \frac{1}{\tau}(f - v)$ is unique.

The following result is straightforward.

**Lemma 2.** *For any $\alpha < \beta$ the resolvent $R(a, b; s)$ is monotone in $s$ and Lipschitz continuous in $s$ with $L_R = 1$, i.e, we have $\forall \xi_1, \xi_2 \in \mathbb{R}$*

$$|\, R(a, b; \xi_1) - R(a, b; \xi_2)\,| \leq |\, \xi_1 - \xi_2)\,|, \tag{18a}$$

$$0 \leq \langle \xi_1 - \xi_2, R(a, b; \xi_1) - R(a, b; \xi_2) \rangle \leq |\, \xi_1 - \xi_2\,|^2 . \tag{18b}$$

*$R$ is also differentiable except at $a, b$ with $0 \leq R' \leq 1$ at the points of differentiability. We also have from* [22]*(Ex.2.1, p33) that $R$ is semismooth.*

2.3.2. *ODE with a constraint graph.* Now we consider the Cauchy problem

$$\tfrac{d}{dt} v(t) + \mathcal{C}(a, b; v(t)) \ni f(t),\ t \in (0, T],\ v(0) = v_{init} \in [a, b]. \tag{19}$$

Its solutions are defined as limits of step functions $v_\tau$ built from implicit finite difference approximations $V^n \approx v(t_n)$ which solve

$$\tfrac{V^n - V^{n-1}}{\tau} + \mathcal{C}(a, b; V^n) \ni f^n, \quad 1 \leq n \leq N,\ V^0 = v_{init}, \tag{20}$$

for some suitable $f^n \approx f(t_n)$. Note that $V^n$ is well defined since (20) has the form (17) and thus we can use the resolvent

$$V^n = R(a, b; V^{n-1} + \tau f^n). \tag{21}$$

In turn, the selection $c^n \in \mathcal{C}(a, b; V^n)$ given from (20) as $\tau c^n = V^{n-1} + \tau f^n - V^n$ is the unique element of minimal norm [4] (pp. 66, 28) for which (20) holds.

Based on the assumption that $v_{init} \in [a, b]$ and $f \in L^1(0, T)$, one can prove convergence of the finite difference solutions to $v(t)$; the rate is usually $O(\tau)$ in a Hilbert space. More details including the construction of $v_\tau$ and regularity of the solutions can be found in [4, 21].

The construction (21) for the approximation to (19) is fundamental in the models in this paper. We illustrate this convergence in Section 5.2.

2.3.3. *ODE with a time-dependent constraint in generalized play hysteresis.* Now we let the constraints in $\mathcal{C}(a, b; \cdot)$ vary in time. To distinguish from fixed $a \leq b$, we consider the graph $C(\alpha, \beta; \cdot)$ with $\alpha(t) = \gamma_r(t) \leq \beta(t) = \gamma_l(t))$, with the resolvent $R(\alpha, \beta; \cdot)$. For more modeling flexibility, $\alpha, \beta$ can depend on some input $u(t)$, namely, $\alpha(t) = \gamma_r(u(t))$, $\beta(t) = \gamma_l(u(t))$. Here $\gamma_r, \gamma_l$ are continuous monotone functions with $\gamma_r \leq \gamma_l$.

Given input $u(t)$ and $\gamma_r, \gamma_l$, we generalize (19) to the Cauchy problem

$$\tfrac{d}{dt} v + \mathcal{C}(\gamma_r(u(t)), \gamma_l(u(t)); v) \ni 0,\ t \in (0, T],\ v(0) = v_{init}. \tag{22}$$

The approximation of (22) follows similar to that for (20) but requires that we know $U^1, U^2, \ldots$. Also, we require $v_{init} \in [\gamma_r(U^0), \gamma_l(U^0)]$. We define the approximation $V^n$ as the solution to

$$\frac{1}{\tau}(V^n - V^{n-1}) + \mathcal{C}(\gamma_r(U^n), \gamma_l(U^n); V^n) \ni 0\,, \quad V^0 = v_{init}\,. \tag{23}$$

Given $U^n$, the solution $V^n$ follows using the resolvent (21).

$$V^n = R(\gamma_r(U^n), \gamma_l(U^n);\ \overline{V}) = \begin{cases} \gamma_r(U^n) \text{ if } \overline{V} \le \gamma_r(U^n), \\ \overline{V} \text{ if } \gamma_r(U^n) < \overline{V} < \gamma_l(U^n), \\ \gamma_l(U^n) \text{ if } \overline{V} \ge \gamma_l(U^n). \end{cases} \quad \overline{V} = V^{n-1}. \tag{24}$$

## 3. PDE model with equilibrium, non-equilibrium, and hysteresis variants and its fully discrete approximation

In this section we present details of the model (2). We closely follow our presentation in [3] and recent in [25, 13] which we augment with the non-equilibrium or hysteretic close for $\chi(u)$. Our equilibrium model is similar to those presented in the applications literature; e.g., in [20, 10, 11], but we make simplifying assumptions and do not consider any particular physical scenarios.

We start with the presentation of an equilibrium model in physical quantities in Section 3.1. Next we reformulate that model in simplified notation in Section 3.1.1. We follow up with the non-equilibrium models, and their discretization.

3.1. **Physical model.** We consider the energy conservation in a partially frozen soil at the Darcy scale, where the soil is treated as a continuum mixture of solid rock grains and water in the liquid or ice phase. The void space available to the water has volume fraction $\eta$ also called the porosity, and the total energy density is given by the enthalpy variable $W$; its evolution is balanced by the heat flux $Q^{\mathcal{T}}$ and the sources. The model is closed by Fourier's heat conduction with a nonlinear heat conductivity $\Lambda^{\mathcal{T}}$ and reads

$$\partial_t(W(\theta)) + \nabla \cdot Q^{\mathcal{T}} = 0, Q^{\mathcal{T}} = -\lambda^{\mathcal{T}} \nabla\theta. \tag{25a}$$

The definitions of $W(\theta), \Lambda^{\mathcal{T}}(\theta)$ we adopt in this paper rely on the following modeling assumptions.

*Physical assumptions:* The porous medium is rigid and homogeneous, i.e. the porosity $\eta$ is constant, and the only fluid present is water in the ice and liquid phases. The following physical material properties of the liquid, ice, and rock grains are constant including the densities $\rho_l, \rho_i, \rho_r$, the heat capacities $c_l, c_i, c_r$, and the heat conductivities $k_l, k_i, k_r$, respectively. We also assume the latent heat $L$ is constant.

Denoting by $S(\theta)$ the volume fraction of liquid, $W$ includes the volume fraction weighted thermal energy densities plus the latent heat for the water material. Given $\eta$, and denoting $c_u = \eta c_l + (1-\eta)c_r$ and $c_f = \eta c_i + (1-\eta)c_r$ we have as in [3]

$$W \ = \ \int_0^\theta (c_u S(v) + c_f(1 - S(v))dv + L\eta S. \tag{25b}$$

We also require $\Lambda^{\mathcal{T}}$ which can be found by upscaling as in [19] or with simple parametric volume fraction weighting. Wlog, in this paper we assume arithmetic weighting. We denote $k_u = \eta k_l + (1-\eta)k_r$ and $k_f = \eta k_i + (1-\eta)k_r$, and define

$$\Lambda^{\mathcal{T}} = k_u S + k_f(1 - S) = (k_u - k_f)S + k_f. \tag{25c}$$

To complete the model we require $S(\theta)$. In equilibrium it is described by an algebraic expression which we denote by $F(\cdot)$. In this paper we adopt a simplification of the parametric

model [10]

$$S = F(\theta) = \begin{cases} 1; & \theta \geq 0, \\ e^{b\theta}; & \theta < 0. \end{cases} \tag{25d}$$

Here $b > 0$ can be found empirically or from upscaling as in [13]. The physical model (25) is now complete and only requires initial and boundary conditions.

3.1.1. **Simplified PDE model.** We derive now a simplified model which retains the qualitative character of (25a) but uses rescaled primary unknowns and data, under the assumption that $\eta$ is constant. To distinguish between the physical set-up and the simplified model, we replace $\theta$ by $u$. We set $\tilde{c}_u = \frac{c_u}{\eta L}, \tilde{c}_f = \frac{c_f}{\eta L}, \tilde{k}_u = \frac{k_u}{\eta L}; \tilde{k}_f = \frac{k_f}{\eta L}$, and divide (25a) by $L\eta$. We have

$$\frac{W(u)}{\eta L} = c(u) + S = (\tilde{c}_u - \tilde{c}_f) \int_o^u S(v)dv + \tilde{c}_f u + S. \tag{26a}$$

$$\frac{\Lambda^{\mathcal{T}}}{L\eta} = k(u) = \tilde{k}_f + (\tilde{k}_u - \tilde{k}_f)S. \tag{26b}$$

With these, if we replace $S$ by $\chi$, the model reads the same as (2)

$$\partial_t(c(u) + \chi) - \nabla \cdot (k(u)\nabla u) = \tilde{f}. \tag{27a}$$

$$u(x, 0) = u_{init}(x), \tag{27b}$$

$$u(x, t)|_{\partial\Omega} = u_D(t), \tag{27c}$$

where $u_D(t), u_{init}(\cdot)$ as well as the rescaled source term $\tilde{f}$ are given. This model will be next supplemented by (EQ) or (NEQ) or (HYST) relationships which define the relationship $(u, \chi)$; in general, $\chi$ is not the same as $F(u)$ and has its own dynamics.

One must make a choice how $k(u), c(u)$ defined in (26) (depending on $S$) is evaluated. One option is to use the current value of $S = \chi$. Another simpler option is to substitute the equilibrium relationship $S = F(u)$ in (26).

**Remark 1.** *In this paper we make the following choice. To calculate $c(u), k(u)$, we plug in the equilibrium formula $S = F(u)$* (25d) *in* (25b) *and* (25c) *rather than recalculate $c(u)$ and $k(u)$ based on the current $\chi$, an unknown in the (NONEQ) and (HYST) models. This approach allows to focus on the abundant remaining challenges.*

**Example 1.** *We take $S = F(u)$ given in* (25d) *and and evaluate $c(u), k(u)$*

$$c(u) = (\tilde{c}_u - \tilde{c}_f) \int_o^u F(v)dv + \tilde{c}_f u = \begin{cases} \frac{\tilde{c}_u - \tilde{c}_f}{b}(e^{bu} - 1) + \tilde{c}_f u, & u \leq 0, \\ \tilde{c}_u u, & u > 0, \end{cases} \tag{28a}$$

$$k(u) = \tilde{k}_f + (\tilde{k}_u - \tilde{k}_f)F(u) = \begin{cases} (\tilde{k}_u - \tilde{k}_f)e^{bu} + \tilde{k}_f, & u \leq 0, \\ k_u, & u > 0. \end{cases} \tag{28b}$$

**Remark 2.** *We can comment now on the behavior of $c(u), k(u)$ in* (28)*. Both are continuous but neither is differentiable at $u = 0$. Therefore it is not recommended in numerical models to replace $\partial_t c(u)$ by $\frac{dc}{du}\partial_t u$ since $c'(u)$ features a large jump and causes numerical oscillations. Implicit treatment of $c(u)$ is recommended.*

**Remark 3.** *We can absorb the constants $L\eta$ in* (26) *making the time variable nondimensional to adhere to the characteristic time of thawing/freezing of about* $[\text{day}] = 8.64 \times 10^4 [\text{s}]$. *See our simulations in Section 5.3.*

**Assumption 1.** *We will make the following assumptions on the data $c, k$ and $F$.*
**(AC):** *The energy density $c(u)$ is continuous monotone increasing in $u$ and smooth except at $u = 0$ where $c(0) = 0$. It is also differentiable except at $u = 0$, with $L_c \geq c'(u) \geq c_{min} > 0$.*
**(AK):** *Heat conductivity $k(u)$ is bounded above and below by positive constants, is smooth except at $u = 0$, and Lipschitz continuous with constant $L_k$.*
**(AF):** *The equilibrium relationship $S = F(u)$ is nonnegative and bounded above, monotone nondecreasing in $u$, smooth except at $u = 0$, and globally Lipschitz, i.e., there are nonnegative constants $L_F, F_\infty, F_0$ as follows: for every $u_1, u_2 \in \mathbb{R}, u \in \mathbb{R}$*

$$(F(u_1) - F(u_2))(u_1 - u_2) \geq 0, \tag{29a}$$

$$|F(u_1) - F(u_2)| \leq L_F |u_1 - u_2|, \tag{29b}$$

$$0 \leq F(u) \leq F_\infty, \tag{29c}$$

$$F(0) = F_0 \leq F_\infty. \tag{29d}$$

The data $c(u), k(u)$ from Example 1 and $F(u)$ from (25d) satisfy these assumptions, and we have $L_F = b, F_0 = F_\infty = 1$.

3.2. **Complete simplified model and fully discrete counterpart.** Now we write the spatially discrete approximations to the IBVP system (27). For simplicity, in this paper we consider only the cell-centered finite differences, with some rectangular grid $\mathcal{T}_h$ over $\Omega$, with cell centers $x_j, j = 1, 2 \dots M$.

The discrete system is solved for the approximations $U_j^n \approx u(x_j, t_n), \Upsilon_j^n \approx \chi(x_j, t_n)$ with the source terms denoted by $f_j^n$. The corresponding time continuous variables are, respectively, $U_j(t) \approx u(x_j, t)$ and $\Upsilon_j(t) \approx \chi(x_j, t)$. These are collected in the vectors $U^n, \Upsilon^n$, and $U(t), \Upsilon(t)$, respectively. The choice of these approximations on a uniform grid does not require the use of non-diagonal mass matrices, thus it makes the notation easier and allows to focus on the features of (non)equilibria. Also, with the CCFD approximation, the boundary data $u_D(t)$ is absorbed in the right hand side $f_j^n$, and there are no degrees of freedom on the boundary $\partial\Omega$ which is assumed to align with the edges of elements in $\mathcal{T}_h$. We refer to [14, 3] for extensive details on the discretization, and to [1] for recent analysis involving nonlinear diffusion terms similar to $-\nabla \cdot (k(u)\nabla u)$.

**Remark 4.** *The node-centered FD approximation can be handled with the same analysis we employ below, as long as it is done for the interior unknowns only.*

The continuous in time discrete in space approximation to (27) is an ODE in $V = \mathbb{R}^M$ and reads

$$\frac{d}{dt}(C(U) + \Upsilon) + A(U)U = f(t);\ (C(U) + \Upsilon)|_{t=0} = W_{init}. \tag{30}$$

Here the diffusion matrix $A = A(U)$ because $k = k(u)$ is nonlinear, and the entries of $A$ are linear in $k$. Thus the relationship $U \to A(U)$ is Lipschitz because of Assumption (AK). However, $L_A$ depends on $L_k$ and the dimension $M$.

**Remark 5.** *Based on Assumption 1 we see that* $\langle C(U) - C(V), U - V\rangle \geq c_{min}||U - V||^2$. *Also, since* $c(0) = 0$, *we have* $\langle C(U), U\rangle \geq c_{min}||U||^2$, *i.e.* $C(U)$ *is strongly monotone. It is also Lipschitz. Thus it plays a similar role to* $I$ *in analysis. We shall thus consider* $C(U) = U$ *in what follows which will considerably simplify the notation.*

After this simplification we consider

$$\frac{d}{dt}(U + \Upsilon) + A(U)U = f(t);\ (U + \Upsilon)|_{t=0} = W_{init}. \tag{31}$$

The fully discrete version of (31) is

$$\tfrac{1}{\tau}(U^n + \Upsilon^n) + A(U^n)U^n = f^n; (U^0 + \Upsilon^0) = W_{init}. \tag{32}$$

Next we must add to (32) some relationship binding $U$ and $\Upsilon$ pointwise; specifically we add the discrete counterparts of (2a) or (2b) or (2c). This is done below.

3.2.1. *Equilibrium system based on* (2a). In equilibrium, we have (32) and discrete form of (2a)

$$\tfrac{1}{\tau}(U^n + \Upsilon^n) + A(U^n)U^n = f^n; \tag{33a}$$

$$\Upsilon^n = (\Upsilon^n_j)_j = (F(U^n_j))_j = \mathcal{F}(U^n). \tag{33b}$$

With the pointwise relationship (33b) at every degree of freedom, we see that $U \to \mathcal{F}(U)$ is Lipschitz with $L_{\mathcal{F}} = L_F$ since $F$ is Lipschitz component-wise (29b).

The system (33) is nonlinear with two nonlinearities $\mathcal{F}$ and $A$, and must be solved by iteration. Its analysis will be given in Section 4.2.

3.2.2. *Non-equilibrium model based on* (2b). Now we discuss the non-equilibrium relationship (2b). Such relationships are fairly common in systems with large spatial scales. In principle, the model (2) is postulated at any spatial scale $x$ and time scale $t$. Then the temperature $u(x, t)$ as an intensive variable is well understood as the pointwise quantity and the formation and disappearance of ice phase is in complete equilibrium with $u$. This understanding requires that we can resolve the fine details of the phase behavior, i.e, we work at a microscopic scale at which the ice/liquid interfaces are visible.

However, it is in general difficult to approximate the solutions to (2) at that scale. Therefore $u(x, t)$ should be understood as an average at time $t$ of the temperature over some local region $\omega(x)$. Then it is natural that in any such region one can observe simultaneously both the ice crystals and liquid regions.

**Remark 6.** *A simple example of non-equilibrium is provided by a cup of water to which we add a few cubes of ice. If* $u(t)$ *is the average temperature in the entire cup at time* $t$, *then one can talk about an equilibrium between* $u$ *and the average water fraction* $S = \chi(u)$ *in the cup only after sufficiently large time. The process of getting to the equilibrium is modeled by* (2b) *with rate* $B > 0$.

We rewrite now (32) and finite difference approximation to (2a). The fully implicit finite difference scheme reads, after some rearranging and setting $\bar{B} = \frac{1}{1+\tau B}$.

$$\frac{1}{\tau}(U^n - U^{n-1}) + \frac{1}{\tau}(\Upsilon^n - \Upsilon^{n-1}) + A(U^n)U^n = f^n, \tag{34a}$$

$$\Upsilon^n = \frac{B\tau}{1 + \tau B}\mathcal{F}(U^n) + \frac{1}{1 + \tau B}\Upsilon^{n-1} = \mathcal{F}(U^n)(1 - \bar{B}) + \bar{B}\Upsilon^{n-1}. \tag{34b}$$

Comparing (34b) to (33b) we see that if $\tau B >> 1$, then $\bar{B}$ is small. Also, (34b) and (33b) have similar right hand sides, with a small role played by $V^{n-1}$.

From this observation we see that the solvability of the system (34) follows similarly to that of (33) once we plug (34b) to (34a). Details will be given in Section 4.3.

3.2.3. *Hysteresis model with* (2c). Now we generalize the model in Section 3.2.2 to acknowledge the fact that the freezing processes have characteristics distinct from thawing. In particular, continuing Remark 6, supplying the heat to the cup will eventually result in all water in the liquid state with rate $B_{thaw}$. However, if the cup is placed in a freezer, the water will gradually turn to ice, but this process due to the slow nucleation of ice crystals will likely proceed with a different rate $B_{freeze} \neq B_{thaw}$, and it might even follow a different $F_{freeze}(u)$ as $u$ is decreasing than $F_{thaw}$ as $u$ is increasing. This is the simplest conceptual model of hysteresis phenomena from physical standpoint.

Hysteresis models have been well studied [23, 6, 8]. In this paper we use the concept of generalized play models of hysteresis recently discussed in the practical setting in [18] and given in Section 2.3.3. We now apply it to the specific characteristics of hysteresis in the permafrost.

We wish to use a particular generalized play model (22) relating $\chi$ to $u$. We must now make a particular selection of $\gamma_r(u), \gamma_l(u)$ to fit the model to a permafrost application. We shall consider a given equilibrium function $F$ satisfying (AF), and a function $G$ satisfying (AF) which together satisfy $G \geq F$, and additional conditions that will be relevant later. In particular, we wish to model

$$F(u) \leq \chi \leq G(u), \tag{35}$$

which we rewrite $0 \leq \chi - F(u) \leq G(u) - F(u)$. In other words, we set up $\alpha = \gamma_r = 0$ and $\beta = \gamma_l(t) = G(u(t)) - F(u(t))$, and apply (22)

$$\frac{d}{dt}\chi + C(0, \beta; \chi - F(u)) \ni 0; \chi(x, 0) = \chi_{init}(x). \tag{36}$$

This equation now supplements (27).

With this specific choice we apply the scheme (23) with some discrete value $\beta_j^n$ known at every degree of freedom $j$, to obtain

$$\frac{1}{\tau}(\Upsilon_j^n - \Upsilon_j^{n-1}) + \mathcal{C}(0, \beta_j^n; \Upsilon_j^n - F(U_j^n)) \ni 0\,;\ \Upsilon_j^0 = \Upsilon_{init,j}. \tag{37}$$

Now we rearrange, subtracting $F(U_j^n)$ from both sides

$$\Upsilon_j^n - F(U_j^n) + \tau\mathcal{C}(0, \beta_j^n; \Upsilon_j^n - F(U_j^n)) \ni \Upsilon_j^{n-1} - F(U_j^n)\,. \tag{38}$$

Applying the resolvent (21) we get $\Upsilon_j^n - F(U_j^n) = (R(0, \beta_j^n; \overline{V}_j))_j$ with $\overline{V}_j = \Upsilon_j^{n-1} - F(U_j^n)$

$$\Upsilon^n = \mathcal{F}(U^n) + (R(0, \beta_j^n;\ \overline{V}_j))_j; \Upsilon_j^n = \begin{cases} F(U_j^n) \text{ if } \overline{V}_j \leq 0, \\ F(U_j^n) + \overline{V}_j \text{ if } 0 < \overline{V}_j < \beta_j^n, \\ \beta_j^n \text{ if } \overline{V}_j \geq \beta_j^n. \end{cases} \tag{39}$$

It remains to specify $\beta^n$.

The fully implicit choice where pointwise $\beta_j^n = G(U_j^n) - F(U_j^n)$ seems natural, but requires iteration and is challenging in analysis.

In this paper we consider therefore the time lagging (sequential approach) with $\beta_j^n = G(U_j^{n-1}) - F(U_j^{n-1})$. The fully implicit finite difference scheme for the system (27), (36) reads now

$$\frac{1}{\tau}(U^n - U^{n-1}) + \frac{1}{\tau}(\Upsilon^n - \Upsilon^{n-1}) + A(U^n)U^n \ni (f_j^n)_j, \tag{40a}$$

$$\Upsilon^n = \mathcal{F}(U^n) + (R(0, \beta_j^n;\ \Upsilon_j^{n-1} - F(U_j^n)))_j, \tag{40b}$$

$$\beta_j^n = G(U_j^{n-1}) - F(U_j^{n-1}).$$

The analysis of this system is given in Section 4.4. We see it will have some similarity with that of (34b) since the right hand side in (40b) is somewhat similar to that in (34b) and (33b), with a small role played by $V^{n-1}$.

## 4. Well-posedness of discrete systems and convergence of iterations

Now we study systems (33), (34), (40) as variants of

$$U^n + \Upsilon^n + \tau A(U^n)U^n \ = \ g^n = \tau f^n + U^{n-1} + \Upsilon^{n-1}, \tag{41}$$

coupled with a relationship for $\Upsilon^n$ in terms of $\Upsilon^{n-1}$ and $\mathcal{F}(U^n)$ specific to the particular model (EQ), (NEQ), or (HYST).

Because of the double nonlinearity in $\mathcal{F}$ and $A$, we start with introductory material on handling $A(U)U$ to be followed by the material on $\mathcal{F}(U)$.

Next we study well-posedness as well as a fixed point iteration; we give proofs for the (EQ) model (33), (NEQ) model (34), and (HYST) model (40) in Sections 4.2, 4.3, and 4.4, respectively.

### 4.1. Case of trivial $\Upsilon^n = 0$ in (41).

We first set $\Upsilon^n = 0$ in (41) and solve

$$U^n + \tau A(U^n)U^n \ = \ g^n. \tag{42}$$

**Theorem 5.** *Let $V = \mathbb{R}^M$. For any $u \in V$, let $A(u)$ be a symmetric uniformly positive definite and uniformly bounded operator which is Lipschitz in $u$*

$$\langle A(u)\xi, \xi\rangle \geq \kappa_0 ||\xi||^2, \tag{43a}$$

$$\langle A(u)\xi, \zeta\rangle \leq A^{max} ||\xi|| ||\zeta||. \tag{43b}$$

$$||(A(u) - A(v))\xi|| \leq L_A ||u - v|| ||\xi|| \tag{43c}$$

*Then if $\tau$ is small enough, there exists a unique solution to*

$$U + \tau A(U)U \ = \ g, \tag{44}$$

*which satisfies the bound*

$$||U|| \leq \tfrac{1}{(1+\tau\kappa_0)} ||g||. \tag{45}$$

*Proof.* We set up fixed point iteration. Given some initial guess $U^{(0)} \in V$ we iterate

$$U^{(m)} + \tau A(U^{(m-1)})U^{(m)} \ = \ g, m = 1, 2, \ldots \tag{46}$$

Fix $m$ now. Taking inner product with $U^{(m)}$ of (46) and applying (43a) we get

$$(1 + \tau\kappa_0)||U^{(m)}||^2 \leq ||U^{(m)}||^2 + \tau\langle A(U^{(m-1)})U^{(m)}, U^{(m)}\rangle = \langle g, U^{(m)}\rangle, \tag{47}$$

with the right hand side bounded by $||g||||U^{(m)}||$. Dividing both sides by $||U^{(m)}||$ we obtain the bound

$$||U^{(m)}|| \leq \tfrac{1}{(1+\tau\kappa_0)}||g||, \ \forall m = 1, 2, \ldots. \tag{48}$$

Next we study the map $U^{(m-1)} \to U^{(m)} = T(U^{(m-1)})$ given by (46), and show $T$ is a contraction. To this aim we subtract (46) for $m$ and that for $m-1$

$$U^{(m-1)} + \tau A(U^{(m-2)})U^{(m-1)} \ = \ g,$$

and take the inner product of the result with $U^{(m)} - U^{(m-1)}$. We get, after adding and subtracting $A(U^{(m-1)})U^{(m-1)}$ inside the inner product

$$\begin{aligned} &\langle U^{(m)} - U^{(m-1)}, U^{(m)} - U^{(m-1)}\rangle \\ &\qquad + \tau\langle A(U^{(m-1)})U^{(m)} - A(U^{(m-2)})U^{(m-1)}, U^{(m)} - U^{(m-1)}\rangle] \\ &= \langle U^{(m)} - U^{(m-1)}, U^{(m)} - U^{(m-1)}\rangle + \tau\langle A(U^{(m-1)})(U^{(m)} - U^{(m-1)}), U^{(m)} - U^{(m-1)}\rangle \\ &\qquad + \tau\langle A(U^{(m-1)})U^{(m-1)} - A(U^{(m-2)})U^{(m-1)}, U^{(m)} - U^{(m-1)}\rangle = 0. \end{aligned} \tag{49}$$

Now shifting the last term on the left hand side to the right and estimating from (43c) using Cauchy-Schwarz inequality we obtain

$$\begin{aligned} &\tau\langle (A(U^{(m-1)}) - A(U^{(m-2)}))U^{(m-1)}, U^{(m)} - U^{(m-1)}\rangle \\ &\qquad \leq \tau L_A||U^{(m-1)} - U^{(m-2)}||||U^{(m-1)}||||U^{(m)} - U^{(m-1)}||.s \end{aligned}$$

For the second term we apply (43a) to get

$$\begin{aligned} &(1+\tau\kappa_0)||U^{(m)} - U^{(m-1)}||^2 \\ &\qquad \leq \tau L_A||U^{(m-1)} - U^{(m-2)}||||U^{(m-1)}||||U^{(m)} - U^{(m-1)}||, \end{aligned}$$

and dividing by $||U^{(m)} - U^{(m-1)}||$ we see

$$||U^{(m)} - U^{(m-1)}|| \leq \frac{\tau L_A}{(1+\tau\kappa_0)}||U^{(m-1)} - U^{(m-2)}||||U^{(m-1)}||.$$

Combining with (48) for $||U^{(m-1)}||$ we finally obtain that the map $T$ has Lipschitz constant

$$L_T = \frac{\tau L_A}{(1+\tau\kappa_0)}||g||. \tag{50}$$

From this we conclude that if $\tau$ is small enough, then $L_T < 1$, and the iteration (46) converges to the unique solution of (44) which satisfies the desired bound (45). □

**Corollary 1.** *Applying Theorem 5 we see that at every time step the solution $U^n$ to (42) exists and is unique. Moreover, the fixed point iteration (46) converges if $\tau$ is small enough. A natural choice of the initial guess $U^{n,(0)} = U^{(n-1)}$.*

**Remark 7.** *The technique we applied does not work in the infinite dimensional setting since there, in particular, one cannot assume that $A(U)$ is uniformly bounded or Lipschitz.*

4.2. **Well-posedness and convergence for the equilibrium problem** (33)**.** Now we consider (41) in which we set $\Upsilon^n = \mathcal{F}(U^n)$. We have

$$U^n + \mathcal{F}(U^n) + \tau A(U^n)U^n \;\; = \;\; g^n; \tag{51}$$

Now we fix and suppress $n$ and study

$$U + \mathcal{F}(U) + \tau A(U)U \;\; = \;\; g; \tag{52}$$

This problem features double nonlinearity $\mathcal{F}(U)$ and $A(U)U$. The challenge is that we do not have lower bounds on $\langle \mathcal{F}(U), U\rangle$, only (29a), thus we have to shift the terms involving $\mathcal{F}(\cdot)$ to the right hand side which results in some restrictive estimates involving $L_F$. We give analysis, and consider alternatives.

4.2.1. *Handling only nonlinearity in F while lagging nonlinearity in A.* We can apply Theorem 1 to the problem

$$U + \mathcal{F}(U) + \tau \bar{A}U \;\; = \;\; g; \tag{53}$$

in which $\bar{A}$ is fixed, and satisfies all the desired properties (43) with a trivial $L_A = 0$. Since $F(\cdot)$ is monotone and Lipschitz, and since $\bar{A}$ linear spd, we obtain that the operator $T(U) = U + \mathcal{F}(U) + \tau \bar{A}U$ is strongly monotone with constant $c_0 = 1 + \tau\kappa_0$ and Lipschitz with the (large) constant $L_T = 1 + L_{\mathcal{F}} + \tau||\bar{A}||$. We obtain the following.

**Lemma 3.** *Let $\bar{A}$ be constant matrix which satisfies* (43a), (43b)*, and let $F(\cdot)$ satisfy* (29)*. Then the problem* (53) *is well-posed and*

$$||U|| \leq \frac{1}{1 + \tau\kappa_0}||g||. \tag{54}$$

Next we check if the iteration $U^{(m-1)} \to U^{(m)} = T(U^{(m-1)})$ defined by

$$U^{(m)} + \mathcal{F}(U^{(m-1)}) + \tau \bar{A}U^{(m)} \;\; = \;\; g, \tag{55}$$

converges. To do so, we use estimates similar to those in the Proof of Theorem 5.

$$\begin{aligned}(1 + \tau\kappa_0)||U^{(m)} - U^{(m-1)}||^2 &\leq ||\mathcal{F}(U^{(m-1)}) - \mathcal{F}(U^{(m-2)})||||U^{(m)} - U^{(m-1)}|| \\ &\leq L_F||U^{(m-1)} - U^{(m-2)}||||U^{(m)} - U^{(m-1)}||.\end{aligned} \tag{56}$$

The Lipschitz estimate employed above works since

$$||\mathcal{F}(U) - \mathcal{F}(V)||^2 = \sum_j |\, F(U_j) - F(V_j)\,|^2 \leq L_F^2 \sum_j |\, U_j - V_j\,|^2 \,.$$

Still, we require $\frac{L_F}{1+\tau\kappa_0} < 1$ for convergence, and even if this holds, convergence may be slow. Thus Newton solver may be a better option for this particular iteration.

To ensure (9) can be applied, we must check the conditions in Theorem 3 or Theorem 4. Consider the Jacobian $\mathcal{J} = T'(U) = I + \mathcal{F}'(U) + \tau\bar{A}$ for $T(U) = U + \mathcal{F}(U) + \tau\bar{A}U - g$. Here $\mathcal{F}'(U)$ is a diagonal matrix with the $j'th$ entry equal $F'(U_j)$. Now $\mathcal{J}$ is nonsingular for any $U$, since as we said above $U + \mathcal{F}(U) + \bar{A}U$ is strongly monotone. However, since $F'$ is not continuous, $T'$ is only given piecewise, and we are not able to verify that $T'$ is Lipschitz or that $(T')^{-1}$ is continuous (see already the scalar case $M = 1$). However, $\mathcal{F}$ is semismooth, thus $\mathcal{J}$ is as well. Thus Newton iteration converges.

**Lemma 4.** *(Semismooth) Newton iteration converges for* (53)*.*

An alternative proof of well-posedness exploiting the fact that the monotone operator $I + \mathcal{F}' + \tau\bar{A}$ is a subgradient is given in [3].

4.2.2. *Handling the double nonlinearity F and A directly.* Now we wish to handle (52) rather than (53). For this we derive additional estimates but they require rather unrealistic restrictive assumptions. We illustrate the difficulties. We set up the iteration combining (46) and (55) to get $U^{(m-1)} \to U^{(m)} = T(U^{(m-1)})$ defined by

$$U^{(m)} + \mathcal{F}(U^{(m-1)}) + \tau A(U^{(m-1)})U^{(m)} \;=\; g. \tag{57}$$

We follow the proof of Theorem 5 adding the elements of the proof of Lemma 3.

First we obtain an upper bound for $||U^{(m)}||$ by following calculations that led to (47) modifying its right hand side from $g$ to $g - \mathcal{F}(U^{(m-1)})$ and estimating its norm by $||g|| + F_\infty$, from $\langle \mathcal{F}(U^{(m-1)}), U^{(m)}\rangle \le ||\mathcal{F}||_\infty ||U^{(m)}|| \le F_\infty ||U^{(m)}||$ to obtain

$$||U^{(m)}|| \le \frac{1}{1+\tau\kappa_0}(||g|| + F_\infty). \tag{58}$$

Handling next the question of convergence of the iteration and proceeding as in (49) we have an additional term on its right hand side $\langle F(U^{(m-1)}) - F(U^{(m-2)}), U^{(m)} - U^{(m-1)}\rangle$, similar as in (55). Thus

$$\begin{aligned}(1+\tau\kappa_0)||U^{(m)} - U^{(m-1)}||^2 \\ \le \tau L_A ||U^{(m-1)} - U^{(m-2)}||||U^{(m-1)}||||U^{(m)} - U^{(m-1)}|| \\ + L_F ||U^{(m-1)} - U^{(m-2)}||||U^{(m)} - U^{(m-1)}||.\end{aligned} \tag{59}$$

Dividing by $||U^{(m)} - U^{(m-1)}||$ and incorporating (58)

$$||U^{(m)} - U^{(m-1)}|| \le \frac{\tau L_A + L_F}{(1+\tau\kappa_0)}(||g|| + F_\infty)||U^{(m-1)} - U^{(m-2)}||,$$

we see that $T$ is a contraction under a rather restrictive condition that

$$L_T = \frac{\tau L_A + L_F}{(1+\tau\kappa_0)}(||g|| + F_\infty) < 1, \tag{60}$$

and we conclude with the result as follows.

**Proposition 1.** *Let $F, A$ be as in Theorem 5 and Lemma 3, respectively, and let* (60) *hold. Then a unique solution to* (52) *exists and the iteration* (57) *converges.*

In general the assumption (60) cannot be verified. Therefore we proceed differently in the next section.

4.2.3. *Handling double nonlinearity by double iteration.* As mentioned above, solving (52) with one fixed iteration is not a practical option since (60) can be restrictive. On the other hand, since $U + \mathcal{F}(U)$ is strongly monotone, there are better algorithms than the single monolithic iteration (57).

We solve (52) by double iteration exploiting (53). Given $U^{(m-1)}$, we find $U^{(m)}$ as follows:

$$\text{Calculate } \bar{A} = A(U^{(m-1)}). \text{ Solve } U^{(m)} + \mathcal{F}(U^{(m)}) + \tau\bar{A}U^{(m)} = g; \tag{61}$$

using Newton's method. Here we iterate $U^{(m,0)}, U^{(m,1)}, \dots U^{(m,r)}, \dots$. In each iteration $r$ we have the Jacobian

$$\mathcal{J} = I + \mathcal{F}'(U^{(m,r)}) + \tau \bar{A}$$

which is spd for any $U^{(m,r)}$. From Lemma 4 we have therefore convergence of Newton's step within the double iteration. With a good initial guess from previous time step $U^{n,(0)} = U^{n-1}$, the iteration is likely very robust for the equilibrium problems, as we have shown, e.g., in [25].

Next we wish to show that the external iteration converges. To this aim we work with the solutions to (61) in iteration $m$ and $m-1$, which we subtract.

$$U^{(m)} + \mathcal{F}(U^{(m)}) + \tau A(U^{(m-1)})U^{(m)} = U^{(m-1)} + \mathcal{F}(U^{(m-1)}) + \tau A(U^{(m-2)})U^{(m-1)}.$$

Introducing $-\tau A(U^{(m-1)})U^{(m-1)}$ on both sides, moving the first two right hand side terms to the left, rearranging and taking inner product with $U^{(m)} - U^{(m-1)}$ as before, and estimating the monotone terms involving $I + \mathcal{F} + \tau A(U^{(m-1)})$ on the left hand side from below we obtain

$$\begin{aligned}(62)\quad (1+\tau\kappa_0)||U^{(m)} - U^{(m-1)}||^2 &\\ \leq \tau\langle A(U^{(m-2)})U^{(m-1)} - \mathbb{A}(U^{(m-1)})U^{(m-1)}, U^{(m)} - U^{(m-1)}\rangle &\\ \leq \tau L_A ||U^{(m-1)}|| ||U^{(m-1)} - U^{(m-2)}|| ||U^{(m)} - U^{(m-1)}||.&\end{aligned}$$

Now we can apply the a-priori bound (54) from Theorem 5 which holds for $U^{(m-1)}$ and we see that the double iteration converges provided

$$\frac{\tau L_A}{(1+\tau\kappa_0)^2}||g|| < 1. \tag{63}$$

Even though this is a restrictive bound on $\tau$, at least it does not involve $L_F$. We summarize these calculations below.

**Proposition 2.** *Under the assumptions of Lemma 3 the double iteration* (61) *is well-posed and converges.*

The efficiency of (61) depends on the ability find a good guess for the Newton step so that the semismooth Newton step converges indeed q-superlinearly.

4.3. **Well-posedness and convergence for non-equilibrium problem** (34)**.** Now we wish to study (34) which we rewrite suppressing $n$, and plugging the expression for $\Upsilon^n$ into that for $U^n$. In every time step $t_n$ we solve

$$U^n + (1-\bar{B})\mathcal{F}(U^n) + \tau A(U^n)U^n = g^n = \tau f^n + U^{n-1} + (1-\bar{B})\Upsilon^{n-1}, \tag{64}$$

Suppressing $n$ we see we solve

$$U + (1-\bar{B})\mathcal{F}(U) + \tau A(U)U = g. \tag{65}$$

Since $1 > \bar{B} > 0$, we see immediately that one can apply the analysis of Section 4.2. In particular. we apply the double iteration (61) and Proposition 2.

4.4. **Well-posedness and convergence for hysteresis problem** (40)**.** Now we focus on solving (40a) to which we plug (40b). At time step $t_n$, we solve

$$U^n + \mathcal{F}(U^n) + (R(0, \beta_j^n;\ \Upsilon_j^{n-1} - F(U_j^n)))_j + \tau A(U^n)U^n = g^n \tag{66}$$
$$g^n = \tau f^n + U^{n-1} + \Upsilon^{n-1}.$$

With the sequential approach, $(\beta_j)_j$ is known. In what follows we suppress the dependence of $R(0, \beta; v)$ on its first two arguments and write $R(v)$ instead.

Thus, given a fixed $\Upsilon, \beta, g$ we must solve for $u$

$$U + \mathcal{F}(U) + (R(\Upsilon_j - F(U_j)))_j + \tau \bar{A}U = g.$$

The term $\mathcal{F}_R(U) = \mathcal{F}(U) + (R(\Upsilon_j - F(U_j)))_j$ requires additional analysis.

We prove that $\mathcal{F}_R$ is Lipschitz componentwise with constant $L_F$ and monotone. Then we apply the analysis of Section 4.2.

**Lemma 5.** *Let $F$ be monotone nondecreasing with Lipschitz constant $L_F$, and let $R = R(0, \beta; v)$ be as defined in* (15)*. Then given a fixed $x \in \mathbb{R}$, the function*

$$F_R(x; v) = F(v) + R(x - F(v)). \tag{67}$$

*is differentiable a.e. in $v$, Lipschitz with constant $L_F$ and monotone in $v$:*

$$\langle v_1 - v_2, F_R(x, v_1) - F_R(x, v_2)\rangle \geq 0. \tag{68}$$

*Proof.* The proof is elementary but tedious since $R$ is only differentiable a.e., and $R(x-F(v))$ is anti-monotone in $V$.

Lipschitz continuity of $F_R$ is immediate; the calculation of $L_{FR}$ follows after we rewrite setting $w_k = F(v_k), y_k = x - w_k$

$$\begin{aligned} F_R(x; v_1) - F_R(x; v_2) &= w_1 + R(x - w_1) - (w_2 + R(x - w_2)) \\ &= y_2 - y_1 + R(y_1) - R(y_2) = (I - R)(y_2) - (I - R)(y_1). \end{aligned} \tag{69}$$

Since $I - R$ has Lipschitz constant 1, thus

$$||F_R(x; v_1) - F_R(x; v_2)|| = ||(I - R)(y_1 - y_2)|| \leq ||w_1 - w_2|| \leq L_F||v_1 - v_2||. \tag{70}$$

To prove monotonicity, recall $0 \leq R' \leq 1$ a.e. Substitute $w_k = F(v_k), \xi_k = x - w_k, k = 1, 2$, and rewrite $F_R(x; v_1) - F_R(x; v_2) = w_1 - w_2 + R(x - w_1) - R(x - w_2)$. For the left hand side of (68) we have

$$\begin{aligned} (v_1 - v_2)(F_R(x; v_1) - F_R(x; v_2)) \\ = (v_1 - v_2)(w_1 - w_2) - (v_1 - v_2)(R(\xi_2) - R(\xi_1)). \end{aligned} \tag{71}$$

From monotonicity of $F, R$ and with $\xi_k = x - w_k$ we have

$$v_1 - v_2 \geq 0 \text{ iff } w_1 - w_2 \geq 0 \text{ iff } R(\xi_2) - R(\xi_1) \geq 0. \tag{72}$$

We study the components of $\langle v_1 - v_2, R(\xi_2) - R(\xi_1)\rangle$ and see that by (72) their product is nonnegative, i.e. equal $|\,(v_1 - v_2)(R(\xi_2) - R(\xi_1))\,|$. We derive thus by repeated application of (72), (18a), and the fact that $\xi_2 - \xi_1 = w_1 - w_2$

$$\begin{aligned} (v_1 - v_2)(R(\xi_2) - R(\xi_1)) &= |\,(v_1 - v_2)(R(\xi_2) - R(\xi_1))\,| \\ &= |\,v_1 - v_2\,||\,R(\xi_2) - R(\xi_1)\,| \leq |\,v_1 - v_2\,||\,\xi_2 - \xi_1\,| \\ &= |\,v_1 - v_2\,||\,w_2 - w_1\,| = (v_1 - v_2)(w_1 - w_2). \end{aligned} \tag{73}$$

Applying this identity, (72) and (71), upon a sign change we obtain (68) from

$$(v_1 - v_2)(F_R(x; v_1) - F_R(x; v_2)) \geq (v_1 - v_2)(w_1 - w_2) - (v_1 - v_2)(w_1 - w_2) = 0. \tag{74}$$

□

With these facts in place, we see that we can easily adapt the proof for the equilibrium case regarding well-posedness of (40) and its solvability by the double iteration (61). However, while our problem is still semismooth, compared to (EQ) and (NEQ) we have now an additional source of lack of smoothness due to only piecewise smoothness of $R(0, \beta; \cdot)$. This fact makes Newton solver work harder.

4.5. **Summary of theoretical results.** We have seen that the Proposition 2 gives convergence of "double" iteration (61) and handling the double nonlinearity with $\mathcal{F}(U)$ and $A(U)U$ for the equilibrium problem (51). With small modification, it also applies to the nonequilibrium (NEQ) and hysteresis (HYST) problems (64) and (66).

With our unified treatment of the algorithms, the main theoretical challenge is not in the individual features of (NEQ) or (HYST) models but rather still in the double nonlinearity.

We implement the double iteration (61) in practice for all the variants (EQ), (NEQ), (HYST). The results and examples are presented below.

## 5. Examples

In this section we show numerical examples for (2) with (EQ), (NEQ), and (HYST) closure. We start with a detailed illustration of calibration of generalized play (HYST) model in Section 2.3.3.

For this model we illustrate how to work with an ODE system, a reduction of (40) from $(U^n, \Upsilon^n) \in R^M \times \mathbb{R}^M$ to $(U^n, \Upsilon^n) \in \mathbb{R} \times \mathbb{R}$. We are able to confirm first order convergence of the numerical scheme; see Section 5.2.

Section 5.3 is devoted to the PDE examples and simulations of (1) coupled to $\chi$ given by the (EQ), (NEQ), (HYST) variants.

5.1. **Calibration of generalized play (HYST) model.** Here we work with the notation of the physical model denoting temperature by $\theta$. We aim to develop a hysteresis model for $\chi(t)$ with which $\gamma_r \leq \chi(t) \leq \gamma_l(t)$, where the lower bound is the same as the curve $\gamma_r(t)$ "on the right" with subscript $r$. Analogously, $\gamma_l$ represents the bound $\gamma_l(t)$ on the left. Both $\gamma_l, \gamma_r$ should be at least Lipschitz continuous.

First we discuss in practice how to get $\gamma_l$ and $\gamma_r$. Typically, assume we are given some equilibrium curve $\chi = F(\theta)$. We aim to define some $G(\theta)$ so that

$$F(\theta) \leq \chi \leq G(\theta), \tag{75}$$

where $G$ is similar to $F$ but has an off-set which models the different rate and delay of nucleation. By design, we want $G$ and $F$ to agree for $\theta > 0$ as well as below certain low temperature $\theta < \theta_0$

$$G(\theta) = F(\theta), \theta < \theta_0;\ G(\theta) = F(\theta), \theta > 0. \tag{76a}$$

We also want these to be smooth, and we parametrize $G(\theta)$ requiring

$$F(\theta_0) = G(\theta_0); F'(\theta_0) = G'(\theta_0); F(0) = G(0). \tag{76b}$$

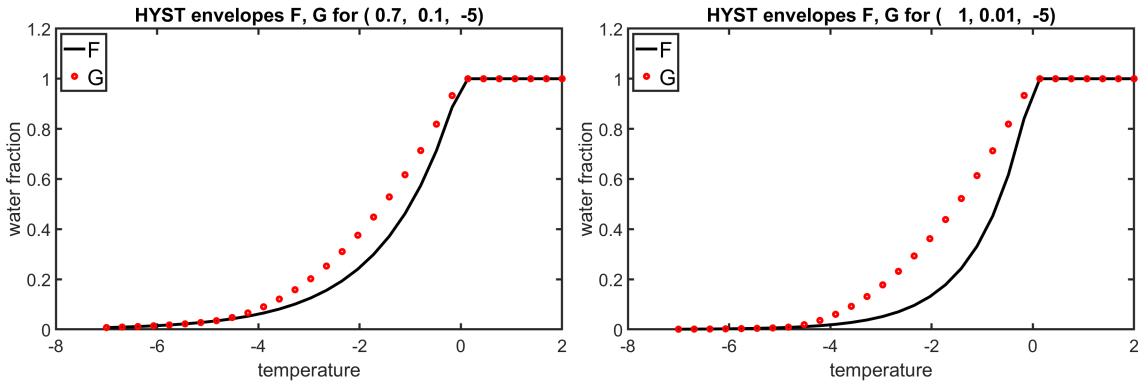


FIGURE 2. The envelopes of the generalized play HYST model calibrated in Example 2 (i) and (ii). The titles indicate the parameters $(b, \bar{b}, \theta_0)$.

**Example 2** (Calibration of $F(\theta), G(\theta)$ for permafrost application)**.** *Let $b > 0$ be given and $F(\theta) = e^{b\theta}$ be as in* (25d)*. For some given $\bar{b} > 0$ amd $\theta_0 < 0$ we propose*

$$G(\theta) = ae^{\bar{b}\theta} + D\theta + C, \tag{77}$$

*where $a, D, C$ are found with the three conditions in* (76b)*. We find*

$$a = \frac{e^{b\theta_0} - b\theta_0 e^{b\theta_0} - 1}{e^{\bar{b}\theta_0} - \bar{b}\theta_0 e^{\bar{b}\theta_0} - 1}; C = 1 - a; D = be^{b\theta_0} - a\bar{b}e^{\bar{b}\theta_0}. \tag{78}$$

*In particular, we obtain data in Table 2 with the graphs $F, G$ plotted in Figure 2.*

TABLE 1. Calibration data in Example 2.

| *CASE* | $b$ | $\bar{b}$ | $\theta_0$ | $a$ | $D$ | $C$ |
|---|---|---|---|---|---|---|
| *(i)* | *0.7* | *0.1* | *-5* | *9.5795* | *-0.5598* | *-8.5795.* |
| *(ii)* | *1* | *0.01* | *-5* | *793.62* | *-7.5424* | *-792.6225* |
| *(iii)* | *0.5* | *0.75* | *-5* | *2.3269* | | *0.0274* |

*We also consider a different case (iii) when the graph $F$ is shifted to the left, and we impose matching with $G(\theta) = ae^{\bar{b}\theta} + C$ where we find $a = \frac{be^{b\theta_0}}{\bar{b}e^{\bar{b}\theta_0}}; C = e^{b\theta_0} - \frac{b}{\bar{b}}e^{b\theta_0}$.*

5.2. **Simulation and convergence of (HYST) for an ODE model.** The envelopes for HYST model developed in Example 2 can now be applied in the evolution model (36). We set-up an example to illustrate that the generalized play model works well and delivers solutions which stay within these envelopes and satisfy (75).

**Example 3.** *We calibrate the hysteresis data $(0, \beta)$ given $F(u)$ and $G(u)$ as in Example 2(ii) and (iii). We solve the ODE* (36) *with scheme* (37) *applying to the resolvent of $\mathcal{C}(0, \beta; \cdot)$ the sequential strategy for $\beta(t) = G(u(t)) - F(u(t))$ based on an assumed input $u(t)$*

$$u(t) = h(t)\cos\left(\frac{\pi}{4}t\right) + g(t),\ h(t) = \begin{cases} 8 & t < 4 \\ 4 & t \geq 4 \end{cases},\ g(t) = \begin{cases} -2 & t < 4 \\ \frac{t}{2} - 8 & t \geq 4. \end{cases} \tag{79}$$

*We consider the interval $(0, T], T = 30$ with a timestep of $\tau = 3.75 \times 10^{-2}$.*

Figure 3 shows computed solution which adheres well to the prescribed envelopes and travels horizontally between these envelopes $F(u)$ and $G(u)$.

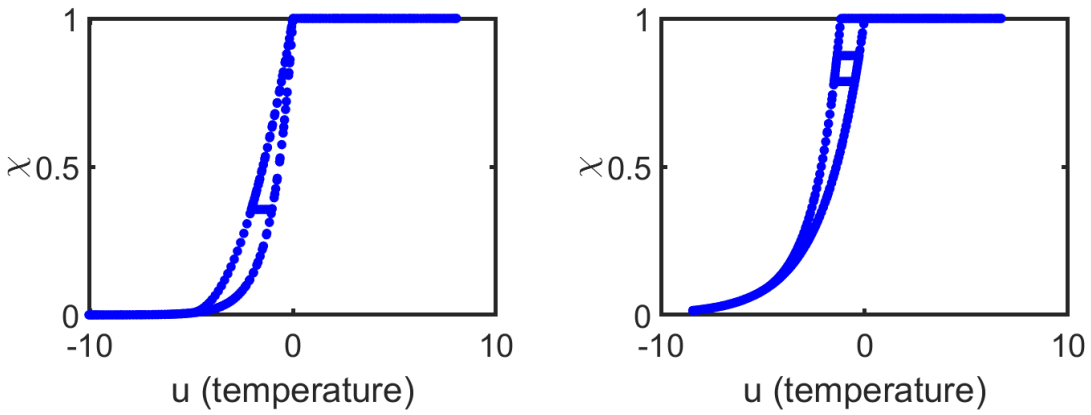


FIGURE 3. Phase plot of solutions $(U^n, \Upsilon^n)_n$ to Example 3 case (ii) and (iii).

5.2.1. *Coupled ODE with generalized play (HYST) model.* Next we consider a coupled model where $u(t)$ is not explicitly given as in Example 3 but rather is governed by its own dynamics

$$\frac{d}{dt}u + \frac{d}{dt}\chi + Au = f(t);\ t \in (0,T]; u(0) = u_{init}, \chi(0) = \chi_{init}, \tag{80}$$

coupled to (36). Here $A > 0$ is fixed and $f$ is some input function. We use Newton solver (9) to solve the discrete step (37).

**Example 4.** *We let* $A = 0.02$ *and design an oscillating forcing function*

$$f(t) = h(t)\cos(\pi t) + g(t);\ h(t) = \begin{cases} 16 & t < 1 \\ 4 & t \geq 1 \end{cases};\ g(t) = \begin{cases} -15 & t < 1 \\ 4t - 30 & t \geq 1. \end{cases} \tag{81}$$

*The hysteresis graph is calibrated with* $b = 1$, $\bar{b} = 0.1$, $\theta_0 = -5$. *We use* $u_{init} = -0.2$, $\chi_{init} = e^{-0.5}$, *and consider the time interval* $(0,T], T = 10$.

For this example we study the convergence of the solutions as $\tau \to 0$. Since we do not know the true solution, we use as a proxy the fine grid solution with $\tau = 0.0001$. We compute the order of convergence for the solution vector $\zeta := (\chi, u)$ considering for each $\tau$, $\zeta_\tau(t) - \zeta_{fine}(t)$. In Table 2 we report the errors $||\zeta_\tau - \zeta_{fine}||$, as well as the order of convergence. We see first order convergence in all norms.

TABLE 2. Error and convergence order for Example 3.

| $\tau$ | $\|\|\zeta_{\text{fine}} - \zeta_\tau\|\|_1$ | Order | $\|\|\zeta_{\text{fine}} - \zeta_\tau\|\|_2$ | Order | $\|\|\zeta_{\text{fine}} - \zeta_\tau\|\|_\infty$ | Order |
|---|---|---|---|---|---|---|
| 0.1 | 0.1750 | - | 0.01750 | - | 0.1750 | - |
| 0.01 | 0.0149 | 1.0702 | 0.0149 | 1.0702 | 0.0149 | 1.0702 |
| 0.001 | 0.0013 | 1.0495 | 0.0013 | 1.0495 | 0.0013 | 1.0495 |

5.3. **Computational results for PDE model.** Now we show simulation results for the PDE model (1) with each one of (2a), (2b), (2c) simulated with the schemes (33), (34), (40), respectively. We use Newton's method and the $A$-lagging approach (61) for resolving the double nonlinearity.

The examples are designed to show robustness of the scheme, with $(u(x,t),\chi(x,t))$ responding to the varying boundary conditions. Such conditions are realistic in permafrost examples where the surface boundary temperature varies daily.

We first focus on the (EQ) model, and next consider (NEQ) and (HYST) cases.

**Data for** (1)**:** We set $F(u)=e^{bu}$ with $b=1$. We also premultiply (27) by $10^6$, and set the time scale of $10^6$[sec] $\approx 11.57$ [day]. To calculate $c(u),k(u)$ we assume $\eta=0.32$ and use data from [3] in Example 1. Now we use $\tilde{\tilde{c}}=\tilde{c}$, and $\tilde{\tilde{k}}=10^6\tilde{k}$, and $\tilde{\tilde{k}}_f=2.06\times10^{-2}, \tilde{\tilde{k}}_u=1.51\times10^{-2}$, $\tilde{\tilde{c}}_f=2.21\times10^{-2}; \tilde{\tilde{c}}_u=2.94\times10^{-2}$. We have

$$\begin{aligned} k(u) &= \tilde{\tilde{k}}_f+(\tilde{\tilde{k}}_u-\tilde{\tilde{k}}_f)F(u); \\ c(u) &= \begin{cases} (\tilde{\tilde{c}}_u-\tilde{\tilde{c}}_f)\int_0^u F(v)dv+\tilde{\tilde{c}}_f u, & (u<0), \\ \tilde{\tilde{c}}_u u, & (u\geq 0). \end{cases} \end{aligned}$$

**Example 5.** *Let $\Omega=(0,1)$, and $T=3$ (about 33* [days]*). We set the initial and boundary conditions*

$$u_{init}^{EQ}(x)=-5, u(0,t)=\begin{cases} 5, & t\leq 1; \\ -10(t-1)+5, & 1<t\leq 2; \\ 10(t-2)-5, & 2<t, \end{cases} \quad u(1,t)=-5. \tag{82a}$$

*We conduct simulations with discretization parameters $M=100$, $\tau=0.01$, considering also testing with other $M,\tau$.*

We present the plots of evolution of $u(x,t),\chi(x,t)$ in Figure 4 (left), with each row corresponding to selected time steps $t=0.5,1,1.5,2,2.5,3$.

We see how the temperature $u(x,t)$ adapts to the evolving boundary condition $u(0,t)$, with the water fraction following suit, and the free boundary apparent from the plot of $\chi(x,t)$ at the points $x$ where $u(x,t)\approx 0$. In particular, the temperature $u(x,t), t\leq 5$ increases gradually from $-5$ to 5 while maintaining the right boundary condition (82a). When $t>1$, $u(x,t)$ responds to the fluctuating upper boundary condition $u|_{x=0,t}$. We also see good resolution of the free boundary; see, e.g., the plot of $\chi(x,t)$ at $t=0.5$ (in the image in the upper left corner).

**Example 6.** *We use the same data as in Example 5 for the (NEQ) and (HYST) models. These start out of equilibrium with*

$$\chi_{init}^{NEQ}(x)=F(u_{init}(x))+0.1, \tag{83}$$

$$\chi_{init}^{HYST}(x)=F(u_{init}(x))+0.1. \tag{84}$$

*In the (NEQ) model we use $B=5$. We also use $B=10$ and $B=20$ for comparison. For (HYST) model we calibrate the hysteresis model as shown in Example 2(ii).*

We simulate $(u^{NEQ}(x,t),\chi^{NEQ}(x,t))$ and $(u^{HYST}(x,t),\chi^{HYST}(x,t))$. We add the plots of their approximations to Figure 4 to supplement the equilibrium model plotted for reference.

We see that the solutions $\chi^{NEQ},\chi^{HYST}$ are distinct from $\chi^{EQ}$ reaching even close to $||\chi^{EQ}-\chi^{NEQ}||_\infty\approx 0.85$, with the corresponding difference for the temperature $||u^{EQ}-u^{NEQ}||_\infty\approx 1$. Similarly, we have $||\chi^{EQ}-\chi^{HYST}||_\infty\approx 0.85$, $||u^{EQ}-u^{NEQ}||_\infty\approx 1.3$.

The solutions $\chi^{NEQ}(x,t)$ "follow" $\chi^{EQ}(x,t)$ with some delay proportional to $1-\bar{B}$. Since the free boundary is present and both $\chi,u$ are evolving, the variable $\chi^{NEQ}(x,t)$ is not always

TABLE 3. Number of iterations of nonlinear solver in Examples 5 and 6.

| | | EQ | | | NEQ | | | HYST | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $M$ | $\tau$ | $N_{min}$ | $N_{max}$ | $N_{ave}$ | $N_{min}$ | $N_{max}$ | $N_{ave}$ | $N_{min}$ | $N_{max}$ | $N_{ave}$ |
| 100 | 0.01 | 3 | 8 | 4.34333 | 3 | 8 | 4.28333 | 3 | 12 | 6.88333 |
| 100 | 0.1 | 4 | 10 | 6.26667 | 4 | 9 | 6 | 4 | 12 | 7.9 |
| 50 | 0.01 | 3 | 8 | 4.28667 | 3 | 7 | 4.22667 | 3 | 20 | 7.63667 |
| 50 | 0.1 | 4 | 10 | 6.1 | 4 | 9 | 5.96667 | 4 | 12 | 7.9 |
| 20 | 0.01 | 3 | 20 | 9.87333 | 3 | 16 | 4.10333 | 3 | 20 | 12.9567 |
| 20 | 0.1 | 5 | 9 | 6.03333 | 4 | 9 | 5.96667 | 4 | 20 | 8.43333 |

very close to $\chi^{EQ}(x,t)$. The simulations capture the richness of the dynamics due to variable boundary data.

In turn, the profile $\chi^{HYST}(x,t)$ adheres to the constraints in (36) which only become active when the local values of $u(x,t)$ change from increasing to decreasing. In the end $\chi^{HYST}(x,t)$ follows, as expected, either $F(u(x,t))$ or $G(u(x,t))$, or remains between these, even if it is not easy to track down exactly where the constraint is active.

The solutions corresponding to the (NEQ) case are smoother than those for (HYST). However, of course both can be expected to be only as smooth as the (EQ) solutions.

In the end, the overall dynamics of the problem is best seen in the phase plot of $(u(x_j,t_n),\chi(x_j,t_n))_{j,n}$ given in Figure 5 and earlier in Figure 1. There we give the points collected from all the simulations for $(u^{EQ}(x_j,t_n),\chi^{EQ}(x_j,t_n))_{j,n}$, $(u^{NEQ}(x_j,t_n),\chi^{NEQ}(x_j,t_n))_{j,n}$, and $(u^{HYST}(x_j,t_n),\chi^{HYST}(x_j,t_n))_{j,n}$. We see the hysteresis loops for the (HYST) solutions. We also see the large difference between the (EQ) and the (NEQ) case, and a significant more spread for the (NEQ) case when $B=1$ and the small spread when $B=5$.

5.3.1. *Solver performance.* The solver is quite robust even though the initiation of the free boundary around $t=0$ presents a challenge. Also, more iterations are needed around the time of dramatic change in the boundary conditions such as around $t=2$. We record the number of required iterations needed to get the residual to satisfy $||T^{(r)}||_\infty \leq 10^{-8}$, with maximum $N_{max}$ of iterations in the more difficult time steps, down to minimum $N_{min}$ when conditions stabilize, with average number $N_{ave}$. We report these in Table 3. We do not iterate with more than 20 iterations.

As expected, thanks to the analysis in Sections 4, there is no significant difference in performance between EQ, and NEQ, with the NEQ case appearing to be slightly "easier", due to the delay effects of relaxation. However, HYST cases seem to require more effort; we believe this is because they feature lower degree of semi-smoothness.

There is some but small difference in the dependence on the dimension $M$ (primarily in $N_{init}$ for a fixed $\tau$. There is a larger difference for fixed $M$ on the time step $\tau$. Unlike in other nonlinear PDEs where smaller $\tau$ improves Newton convergence and that coarse grids requires fewer iterations, we see here that the cases with coarse discretization but small time step seem to struggle more, especially for HYST case. We believe this is related to the fact that the solver has difficulty getting the free boundary right, and thus requires a larger number of iterations. Clearly more work is needed on the solvers.

## 6. Summary and acknowledgements

In this paper we outlined the basic ingredients of a model for phase transitions out of equilibrium for the applications to thawing and freezing in permafrost soils. Our focus was on a unified presentation and analysis of fully implicit schemes for these phenomena.

The main challenges are the presence of two nonlinearities $F(U)$ describing the water fraction due to the temperature, as well as $A(U)U$ associated with the nonlinear dependence of the heat conductivity on the temperature $u$. The analysis of the schemes we conducted reveals that handling the non-equilibrium and hysteretic relationships can be done in a similar fashion as of those for the equilibrium model. Computational examples show all models are robust, but the HYST case requires slightly more work and is less smooth.

In this paper we presented only a few examples. Work is underway on identifying the appropriate data for the non-equilibrium and hysteretic models depending on the spatial scale of interest. In addition, in this paper we considered fixed formulas for $c(u), k(u)$ not directly involving the variable liquid fraction subject to the lack of equilibrium. We plan to extend our work in this direction in the future.

**Acknowledgments.** This research was partially supported by the grants NSF DMS-1912938 "Modeling with Constraints and Phase Transitions in Porous Media", and NSF DMS-2309682 "Computational mathematics of Arctic processes". We also would like to thank the reviewers and editors whose remarks helped to improve this manuscript.

## References

[1] A. Alhammali, M. Peszynska, and C. Shin. Numerical analysis of a mixed finite element approximation of a coupled system modeling biofilm growth in porous media with simulations. *IJNAM*, 21:20–64, 2024.

[2] Kendall Atkinson and Weimin Han. *Theoretical numerical analysis*, volume 39 of *Texts in Applied Mathematics*. Springer, New York, second edition, 2005. A functional analysis framework.

[3] Lisa Bigler, Malgorzata Peszynska, and Naren Vohra. Heterogeneous Stefan problem and permafrost models with P0-P0 finite elements and fully implicit monolithic solver. *Electronic Research Archive*, 30(4):1477–1531, 2022.

[4] H. Brézis. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*. North-Holland Publishing Co., Amsterdam, 1973. North-Holland Mathematics Studies, No. 5. Notas de Matemática (50).

[5] Xun Jiang and Ricardo Nochetto. A P1–P1 finite element method for a phase relaxation model I: Quasiuniform mesh. *Siam Journal on Numerical Analysis - SIAM J NUMER ANAL*, 35, 06 1998.

[6] M. A. Krasnoselskii and A. V. Pokrovskiĭ. *Systems with hysteresis*. Springer-Verlag, Berlin, 1989. Translated from the Russian by Marek Niezgódka.

[7] T. D. Little and R. E. Showalter. The super-Stefan problem. *Internat. J. Engrg. Sci.*, 33(1):67–75, 1995.

[8] I. D. Mayergoyz. *Mathematical models of hysteresis*. Springer-Verlag, New York, 1991.

[9] F. Patricia Medina and Malgorzata Peszynska. Stability for implicit–explicit schemes for non-equilibrium kinetic systems in weighted spaces with symmetrization. *Journal of Computational and Applied Mathematics*, 328:216–231, 2018.

[10] Radoslaw L. Michalowski. A constitutive model of saturated soils for frost heave simulations. *Cold Regions Science and Technology*, 22(1):47–63, 1993.

[11] Dmitry Nicolsky, Vladimir Romanovsky, and Gennadiy Tipenko. Using in-situ temperature measurements to estimate saturated soil thermal properties by solving a sequence of optimization problems. *The Cryosphere*, 1, 11 2007.

[12] M. Peszynska. A differential model of adsorption hysteresis with applications to chromatography. In Jorge Guinez Angel Domingo Rueda, editor, *III Coloquio sobre Ecuaciones Diferenciales Y Aplicaciones, May 1997*, volume II. Universidad del Zulia, 1998.

[13] M. Peszynska, Z. Hilliard, and N Vohra. Coupled flow and energy models with phase change in permafrost from pore- to Darcy scale: modeling and approximation. *Journal of Computational and Applied Mathematics*, 450:115964, November 2024.

[14] M. Peszynska, E. Jenkins, and M. F. Wheeler. Boundary conditions for fully implicit two-phase flow model. In Xiaobing Feng and Tim P. Schulze, editors, *Recent Advances in Numerical Methods for Partial Differential Equations and Applications*, volume 306 of *Contemporary Mathematics Series*, pages 85–106. American Mathematical Society, 2002.

[15] M. Peszynska and C. Shin. Stability of a numerical scheme for methane transport in hydrate zone under equilibrium and non-equilibrium conditions. *Computational Geosciences*, 5:1855–1886, 2021.

[16] M. Peszynska and R. Showalter. Approximation of scalar conservation law with hysteresis. *SIAM Journal Numerical Analysis*, 58(2):962–987, 2020.

[17] M. Peszynska and R. E. Showalter. A transport model with adsorption hysteresis. *Differential Integral Equations*, 11(2):327–340, 1998.

[18] Malgorzata Peszynska and Ralph E Showalter. Approximation of hysteresis functional. *Journal of Computational and Applied Mathematics*, 389:113356, 2021.

[19] Malgorzata Peszynska, Naren Vohra, and Lisa Bigler. Upscaling an Extended Heterogeneous Stefan Problem from the Pore-Scale to the Darcy Scale in Permafrost. *Multiscale Modeling & Simulation*, 22(1):436–475, 2024.

[20] Vladimir Romanovsky and Tom Osterkamp. Effects of unfrozen water on heat and mass transport in the active layer and permafrost. *Permafrost and Periglacial Processes*, 11:219–239, 07 2000.

[21] R. E. Showalter. *Monotone operators in Banach space and nonlinear partial differential equations*, volume 49 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1997.

[22] Michael Ulbrich. *Semismooth Newton methods for variational inequalities and constrained optimization problems in function spaces*, volume 11 of *MOS-SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2011.

[23] Augusto Visintin. *Differential models of hysteresis*, volume 111 of *Applied Mathematical Sciences*. Springer-Verlag, Berlin, 1994.

[24] Augusto Visintin. *Models of phase transitions*, volume 28 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser Boston, Inc., Boston, MA, 1996.

[25] Naren Vohra and Malgorzata Peszynska. Robust conservative scheme and nonlinear solver for phase transitions in heterogeneous permafrost. *Journal of Computational and Applied Mathematics*, 442:115719, 2024.

Oregon State University, Department of Mathematics, Corvallis, OR 97331, USA
*Email address*: mpesz@math.oregonstate.edu, sluggn@oregonstate.edu

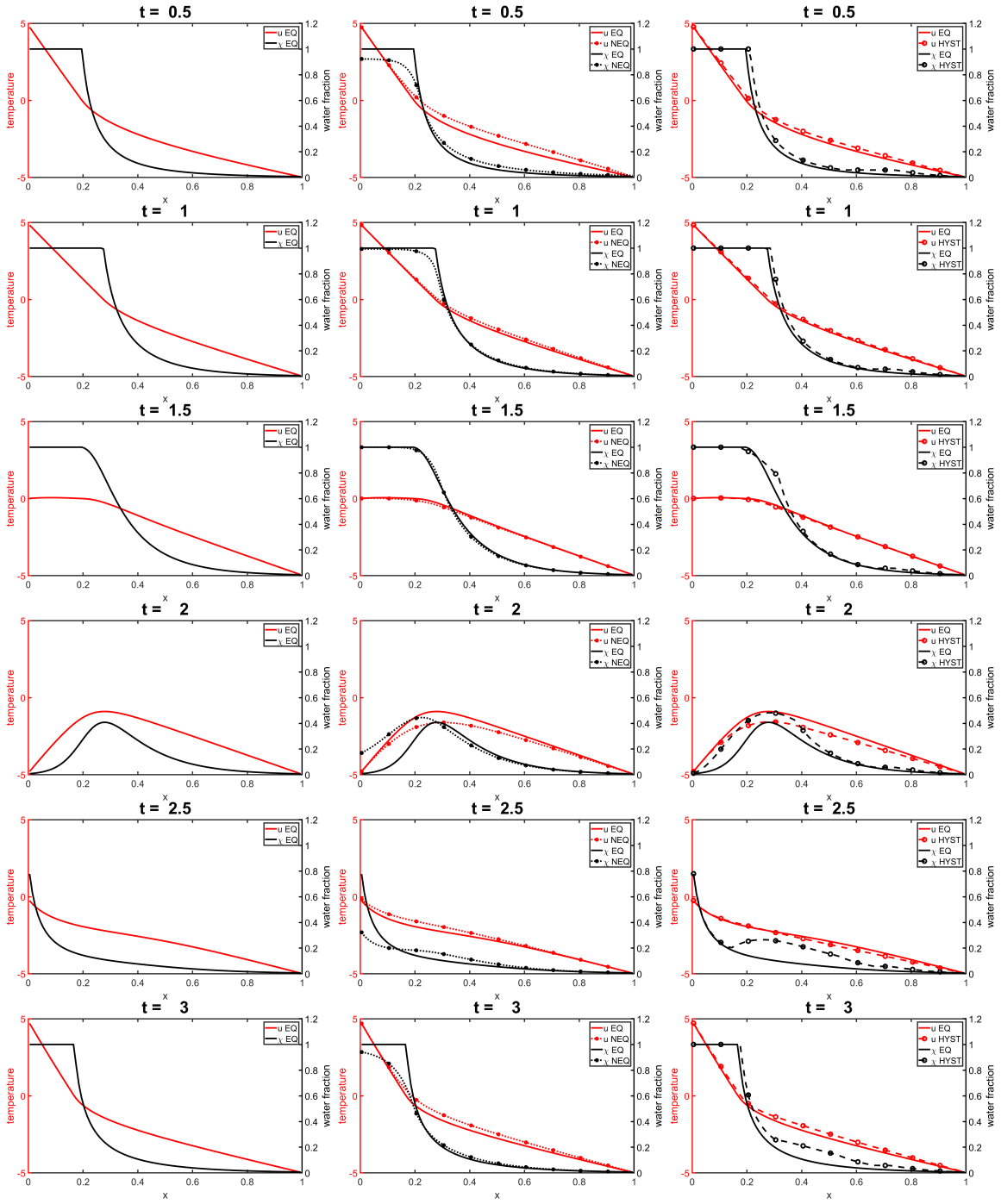


FIGURE 4. Solutions to Example 5 and 6 at time steps $t = 0.5, 1, 1.5, 2, 2.5, 3$. We plot both the temperature $u(x,t)$ (scale on the left axis) and the water fraction $\chi(x,t)$ (scale on the right axis). In the left column we plot only the solutions to the (EQ) model, as indicated in the legend. In the middle and right we plot the solutions to the (NEQ) and (HYST) models, respectively, with the equilibrium solution (EQ) plotted for reference.

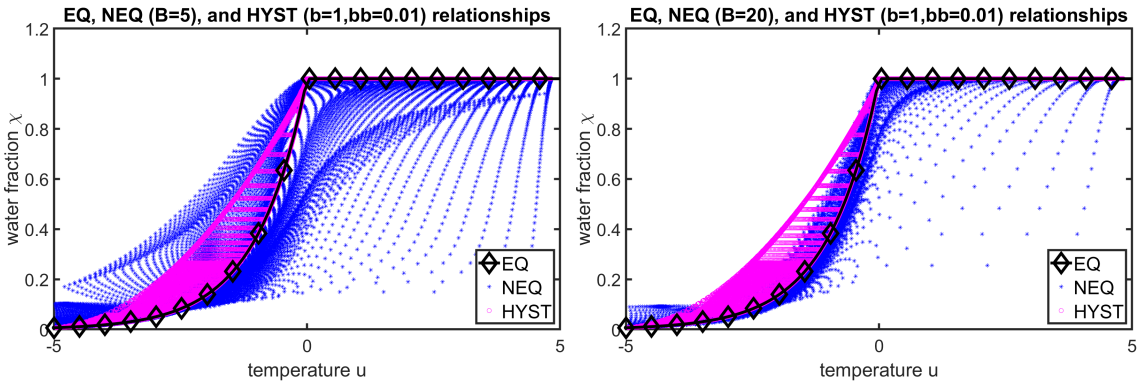


FIGURE 5. Phase plot of the solutions to Example 6 for $B = 5$ (left) and $B = 20$ (right).