Journal of Computational Physics 503 (2024) 112828

Contents lists available at ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

Deep level-set method for Stefan problems

Mykhaylo Shkolnikov, H. Mete Soner, Valentin Tissot-Daguette *

Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ, 08540, USA

ARTICLE INFO

Keywords:
Level-set method
Mushy region
Neural networks
Probabilistic solutions
Stefan problem
Surface tension

ABSTRACT

We propose a level-set approach to characterize the region occupied by the solid in Stefan problems with and without surface tension, based on their recent probabilistic reformulation. The level-set function is parameterized by a feed-forward neural network, whose parameters are trained using the probabilistic formulation of the Stefan growth condition. The algorithm can handle Stefan problems where the liquid is supercooled and can capture surface tension effects through the simulation of particles along the moving boundary together with an efficient approximation of the mean curvature. We demonstrate the effectiveness of the method on a variety of examples with and without radial symmetry.

1. Introduction

The Stefan problem [22,41–44] is central to partial differential equations involving free boundaries. It aims to capture the moving interface separating a solid from a liquid region, as well as the evolution of the temperature in both regions. Despite its simple description and many deep results obtained since its introduction (see, e.g., [13] and the references therein), many intriguing questions remain open. In particular, weak solutions to the Stefan problem are non-unique in general, while strong solutions may fail to exist. Thus, further restrictions are needed to obtain a unique characterization. A recent approach developed by Delarue, Guo, Nadtochiy and the first author [12,27,16] provides stochastic representations and proposes the notion of *physical probabilistic* solutions as a selection principle. This new notion is expected to lie between weak and strong solutions, as shown in [27] for the one-phase supercooled Stefan problem.

Probabilistic solutions satisfy a growth condition relating the change in volume of the solid region, denoted by Γ_t, to the proportion of absorbed "heat" particles in the two phases:

$$\underbrace{|\Gamma_0| - |\Gamma_t|}_{\text{Volume change of the solid}} = \underbrace{\eta\,\mathbb{P}(\tau^1 \le t) - \mathbb{P}(\tau^2 \le t)}_{\text{Absorbed liquid \& solid heat particles}}, \tag{1}$$

where τ^1 (resp. τ^2) stands for the hitting time of the moving interface $\partial\Gamma = (\partial\Gamma_t)_{t\in[0,T]}$ for particles in the liquid (resp. solid). The binary parameter η captures the effect of particles in the liquid when the latter has a nonnegative temperature ($\eta = 1$) or is *supercooled* ($\eta = -1$). An exact statement is given in Definitions 1 and 2, below. While probabilistic solutions of the Stefan problem are quite well understood in one space dimension [12,10] and for radially symmetric solids [16,26], less is known in the general case. Additional challenges arise when incorporating surface tension effects through the classical *Gibbs-Thomson law*, [2,26,32], which postulates that the temperature at the interface is below (resp. above) the equilibrium melting point where the solid is locally convex (resp. concave).

* Corresponding author.
E-mail addresses: mykhaylo@princeton.edu (M. Shkolnikov), soner@princeton.edu (H. Mete Soner), v.tissot-daguette@princeton.edu (V. Tissot-Daguette).

https://doi.org/10.1016/j.jcp.2024.112828
Received 8 July 2023; Received in revised form 1 February 2024; Accepted 1 February 2024

Available online 7 February 2024

We leverage the probabilistic solutions of Stefan problems, the celebrated level-set method of Osher and Sethian [28], and the recent advances in training neural networks in order to produce efficient numerical algorithms. More specifically, we represent $\Gamma = (\Gamma_t)_{t\in[0,T]}$, the evolving region occupied by the solid, by a level-set function which is parameterized by a feed-forward neural network. That is, the evolving solid is given by the zero sublevel set of an appropriate function $\Phi : [0,T]\times\mathbb{R}^d \to \mathbb{R}$, focusing on $d \in \{2,3\}$ in our numerical experiments. The level-set method is widely used to describe the evolution of moving interfaces in arbitrary dimensions as it imposes no assumptions on the geometry of the unknown region and can easily handle changes in topological properties. For example, level sets are able to capture the separation of a connected set into several components and vice versa. As demonstrated in the numerical examples below (Section 4.2.4), this flexibility is of crucial importance for general Stefan problems. We refer the reader to the excellent books by Osher and Fedkiw [29] and Sethian [36] for a comprehensive description of level-set methods.

To the best of our knowledge, the use of the level-set method for the Stefan problem was first proposed in [9]. Their algorithm alternately approximates the moving interface through level-set functions and the temperature in the two phases via a finite difference scheme for the heat equation. The method is capable of accurately reproducing known solutions to the Stefan problem, as well as of generating realistic dendritic growth for a variety of solids (see also [15], focusing on dendritic crystallization, and [29, Section 23], presenting applications to general heat flows). Herein, the level-set function is parameterized by a time-space feedforward neural network $\Phi : [0,T]\times\mathbb{R}^d\times\Theta \to \mathbb{R}$ where Θ is a finite-dimensional parameter set. The parameters are trained using stochastic gradient descent by converting the growth condition of probabilistic solutions into a loss function. The computation of the latter involves the simulation of reflected Brownian particles in the two phases. The proposed method has the advantage that the normal vector to the interface – which is essential in level-set methods – can be effortlessly computed through automatic differentiation of the deep level-set function.

The training of neural network parameters is achieved by minimizing a loss function that involves stopped particles. Since the hitting times of sharp interfaces lead to vanishing gradient issues and prevent the training of the deep level-set function, we use a relaxation procedure as in [34,35,40], developed for optimal stopping. This relaxation consists of introducing a *mushy region* separating the solid and the liquid as in phase-field models [5,8,39], and then using stopping probabilities. The loss function defined in (13), below, tries to enforce the local version (4) of (1) for a large number of randomly chosen test functions. Such a construction based on an identity holding for a class of test functions is novel and may find other applications.

The trained network encapsulates the two phases, and it is important to emphasize that the temperature function does not need to be approximated during training. Due to the probabilistic nature of the algorithm, the temperature function can be estimated later through the empirical measure of the surviving particles. Moreover, surface tension effects can be seamlessly integrated into the algorithm, see Section 3.5. We refer the interested reader to [32] surveying computational methods for differential equations with surface tension in fluid mechanics.

Combining the level-set method with deep learning to solve free boundary problems appears to be new, although our work shares some similarities with [1], which provides a numerical resolution of controlled front propagation flows with the level-set method, and where the velocity is approximated by neural networks and not the level-set function. The present approach is motivated by the recent advances of deep learning to solve complex and/or high-dimensional problems in partial differential equations [19,37,46,31], optimal stopping [6,7,34,35], as well as general stochastic control problems [4,18,34]. In the case of physical phenomena, *physics-informed* neural networks [33,46] successfully combine observed data with known physical laws to learn the solution for a variety of problems in physics. The neural parameterization of level-set functions has already shown promising results in computer vision. Among others, a neural network is trained to approximate the signed distance function associated with three-dimensional objects in [30] and their *occupancy probability* in [24].

We believe that the proposed deep level-set method can be applied to problems in a variety of contexts. This includes free boundary problems in physics, such as Hele-Shaw and Stokes flows [11]. In mathematical finance, the method can be used to generalize the *neural optimal stopping boundary method* of [35] for the exercise boundary of American options with known geometric structure. Also, optimal portfolio problems with transaction costs [25] could benefit from our method once the so-called no-trade zone is represented by a deep level-set function.

Structure of the paper. We introduce the Stefan problem and its probabilistic solution in Section 2. The deep level-set method is described in Section 3 and extended in Section 3.5 for the Stefan problem with surface tension. Section 4 is devoted to the numerical results in the radially symmetric case (Section 4.1) and for general shapes of the solid (Section 4.2). Section 5 concludes, and Appendix A contains the proofs of the main results.

Notations. Given a Lebesgue measurable set $A \subset \mathbb{R}^d$, $d \geq 1$, and a measurable function $\psi : \mathbb{R}^d \to \mathbb{R}$, we employ the shorthand notations $\int_A \psi = \int_A \psi(x)\,\mathrm{d}x$ and $\int_{\partial A} \psi = \int_{\partial A} \psi(x)\,\mathrm{d}\mathcal{H}^{d-1}(x)$, where $\mathcal{H}^{d-1}$ is the $(d-1)$-dimensional Hausdorff measure. We also write $|A|$ for the Lebesgue measure of A.

2. Stefan problems and probabilistic solutions

Let Ω be a bounded domain in $\mathbb{R}^d$, $d \geq 1$, and $T \in (0,\infty)$. Given a closed subset $\Gamma_{0-} \subseteq \overline{\Omega}$ and $u_0^1 : \overline{\Omega\setminus\Gamma_{0-}} \to \mathbb{R}$, $u_0^2 : \Gamma_{0-} \to \mathbb{R}_-$, the strong formulation of the *two-phase Stefan problem* amounts to finding a triplet $(u^1,u^2,\Gamma) = (u^1(t,\cdot),u^2(t,\cdot),\Gamma_t)_{t\in[0,T]}$ such that

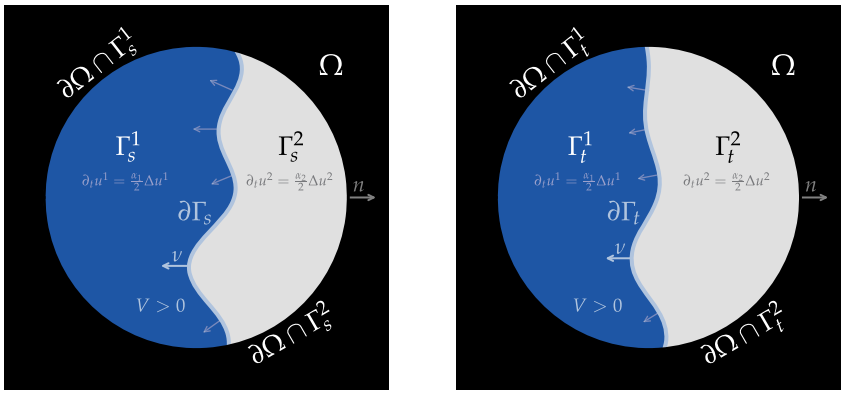


Fig. 1. Illustration of the two-phase Stefan problem (2a)-(2e). Evolution of the liquid region (blue) and solid region (white) in a freezing regime for some time points $0 \le s < t \le T$. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

$$\begin{cases} \partial_t u^i = \frac{\alpha_i}{2}\Delta u^i, & \text{on } int\Gamma^i := \{(t,x) \in (0,T] \times \Omega : \ x \in \text{int}\, \Gamma_t^i\},\ i = 1,2, & \text{(a)} \\ u^i(0-,\cdot) = u_0^i, & \text{on } \Gamma_{0-}^i,\ i = 1,2, & \text{(b)} \\ \partial_{\mathfrak{n}} u^i = 0, & \text{on } \partial\Omega \cap \Gamma^i := \{(t,x) \in [0,T] \times \partial\Omega \ : \ x \in \Gamma_t^i\},\ i = 1,2, & \text{(c)} \\ V = \frac{\alpha_2}{2L}\partial_\nu u^2 - \frac{\alpha_1}{2L}\partial_\nu u^1, & \text{on } \partial\Gamma := \{(t,x) \in [0,T] \times \Omega \ : \ x \in \partial\Gamma_t\}, & \text{(d)} \\ u^1 = u^2 = 0, & \text{on } \partial\Gamma, & \text{(e)} \end{cases} \tag{2}$$

where $\Gamma_t^1 = \Omega \setminus \Gamma_t$, $\Gamma_t^2 = \Gamma_t$ represent respectively the liquid and the solid region, and $\mathfrak{n}$ is the outward normal vector field on $\partial\Omega$. In (2d), V is the outward normal velocity of $\partial\Gamma_t$, ν the outward normal vector field on $\partial\Gamma_t$, L the latent heat of fusion, and α_1, α_2 the thermal diffusivities. An illustration of the two-phase Stefan problem is given in Fig. 1. We suppose that the boundary $\partial\Gamma_t$ is in the solid region so that Γ_t^1 (resp. $\Gamma_t = \Gamma_t^2$) is relatively open (resp. closed) in Ω.

The initial temperature in the solid u_0^2 is always assumed to be non-positive. At the same time, the temperature in the liquid u_0^1 is assumed to be either non-negative or non-positive everywhere. For convenience, we introduce the parameter $\eta := \text{sign}(u_0^1)$ which indicates whether the liquid is initially regular ($\eta = 1$) or *supercooled* ($\eta = -1$).

2.1. Probabilistic solutions

We assume throughout that $\|u_0^i\|_{L^1(\Gamma_{0-}^i)} = 1$, $i = 1,2$, for ease of presentation and interpretation. Indeed, the initial temperature function (or its absolute value) can be regarded as the initial probability density function. Nevertheless, this restriction can be easily removed; see Remark 2. Further, we let X^i be a Brownian motion with diffusivity α_i that is normally reflected along $\partial\Omega$ and absorbed when hitting the moving interface $\partial\Gamma$ (in light of (2a), (2c), (2e)). More specifically,

$$X_t^i = X_0^i + \sqrt{\alpha_i}\, W_{t\wedge\tau^i}^i + l_{t\wedge\tau^i}^i,\ 0 \le t \le T, \quad \tau^i = \inf\{t \in [0,T] : \ X_0^i + \sqrt{\alpha_i}\, W_t^i + l_t^i \notin \Gamma_t^i\}, \tag{3}$$

where W^i is a standard Brownian motion in $\mathbb{R}^d$ and l^i the *local time process* at $\partial\Omega$ [14, Chapter 1]. In addition, we take $X_0^i \sim |u_0^i|(x)\,\mathrm{d}x$. We can now state the definition of a probabilistic solution.

Definition 1. We say that (μ^1, μ^2, Γ) is a probabilistic solution of the Stefan problem (2a)–(2e) if for all $t \in [0,T]$, one has

$$\int_{\Gamma_{0-}} \psi - \int_{\Gamma_t} \psi = \frac{1}{L}\left(\eta\, \mathbb{E}^{\mu^1}[\psi(X_{\tau^1}^1)\, \mathbb{1}_{\{\tau^1 \le t\}}] - \mathbb{E}^{\mu^2}[\psi(X_{\tau^2}^2)\, \mathbb{1}_{\{\tau^2 \le t\}}]\right), \quad \psi \in C_c^\infty(\Omega), \tag{4}$$

where μ^i is the law of X^i, $i = 1,2$.

We next prove that the definition of a probabilistic solution is consistent with that of a classical one, as already shown for the one-phase problem in [27, Proposition 5.5].

Proposition 1. *Suppose that (u^1, u^2, Γ) is a classical solution of the Stefan problem (2a)–(2e). Let X^i be as in (3) and set $\mu^i = \mathbb{P} \circ (X^i)^{-1}$. Then, (μ^1, μ^2, Γ) is a probabilistic solution of (2a)–(2e).*

Proof. See Appendix A.1. □

In the course of the proof of Proposition 1, we find that

$$|u^i|(t,x)\,\mathrm{d}x = \mathbb{P}(X^i_t \in \mathrm{d}x,\, \tau^i > t), \quad i = 1,2. \tag{5}$$

The temperature, albeit not our primary focus, can thus be retrieved from the (sub)density of "survived" particles (i.e., those not yet absorbed by the moving boundary $\partial\Gamma$) once $\Gamma = (\Gamma_t)_{t\in[0,T]}$ has been estimated. Further details are given in Section 4.1.2.

Remark 1. Suppose that the boundary of the ice region lies entirely in Ω for all $t \in [0,T]$. Then, a direct approximation argument shows that (4) holds also for smooth test functions which are not compactly supported in Ω, and we may choose $\psi \equiv 1$ to obtain

$$|\Gamma_{0-}| - |\Gamma_t| = \eta\,\mathbb{P}(\tau^1 \le t) - \mathbb{P}(\tau^2 \le t).$$

The above identity relates the change in volume of the solid region on the left-hand side to the exit probabilities of liquid and solid particles from their respective regions on the right-hand side, and can be interpreted as energy conservation.

Remark 2. Suppose that (u^1,u^2,Γ) is a classical solution of (2a)–(2e) with initial temperature functions u^i_0 such that $\|u^i_0\|_{L^1(\Gamma^i_{0-})} = c_i \in (0,\infty)$, $i = 1,2$. After a simple adaptation of the proof of Proposition 1, the growth condition (4) becomes

$$\int_{\Gamma_{0-}} \psi - \int_{\Gamma_t} \psi = \frac{1}{L}\Big(\eta\, c_1\, \mathbb{E}^{\mu^1}[\psi(X^1_{\tau^1})\,\mathbb{1}_{\{\tau^1\le t\}}] - c_2\, \mathbb{E}^{\mu^2}[\psi(X^2_{\tau^2})\,\mathbb{1}_{\{\tau^2\le t\}}]\Big), \quad \psi \in C^\infty_c(\Omega). \tag{6}$$

The additional flexibility in the initial temperature when $c_i \neq 1$ proves useful in the numerical experiments (see Sections 4.1.2 and 4.1.3).

Example 1. The *one-phase Stefan problem* consists of setting $u^2(t,x) \equiv 0$ in the solid region. If the liquid temperature is initially positive, then the solid is necessarily melting, i.e.: $\Gamma_t \subseteq \Gamma_s \subseteq \Gamma_{0-}$ for all $0 \le s \le t \le T$. Writing $X = X^1$, $\tau = \tau^1$, and $\mu = \mu^1$, the growth condition (4) simply reads

$$\int_{\Gamma_{0-}\setminus\Gamma_t} \psi = \frac{\eta}{L}\,\mathbb{E}^{\mu}[\psi(X_\tau)\,\mathbb{1}_{\{\tau\le t\}}], \quad \psi \in C^\infty_c(\Omega). \tag{7}$$

2.2. Adding surface tension

Consider the Stefan problem (2a)–(2e) with the Dirichlet boundary condition (2e) replaced by

$$u^1 = u^2 = -\gamma\,\kappa_{\partial\Gamma}, \quad \text{on } \partial\Gamma. \tag{2e'}$$

The term $\kappa_{\partial\Gamma_t}$ is the mean curvature of $\partial\Gamma_t$ and $\gamma > 0$ the surface tension coefficient. We use the convention that $\kappa_{\partial\Gamma_t}(x)$ is nonnegative when Γ_t is locally convex at $x \in \partial\Gamma_t$. Equation (2e') captures the so-called *Gibbs-Thomson* effect which postulates that the temperature at the interface is negative for strictly convex boundaries [45,23]. In other words, the freezing point of the liquid decreases at points of convexity and increases at points of concavity. Clearly, surface tension effects only appear if $d \ge 2$. We now revisit Definition 1 in the presence of surface tension.

Definition 2. We say that (μ^1,μ^2,Γ) is a probabilistic solution of the Stefan problem (2a)–(2e') with surface tension if for all $t \in [0,T]$, one has

$$\int_{\Gamma_{0-}} \psi - \int_{\Gamma_t} \psi = \frac{1}{L}\Big(\eta\,\mathbb{E}^{\mu^1}[\psi(X^1_{\tau^1})\,\mathbb{1}_{\{\tau^1\le t\}}] - \mathbb{E}^{\mu^2}[\psi(X^2_{\tau^2})\,\mathbb{1}_{\{\tau^2\le t\}}] + \mathcal{K}_t\Big), \quad \text{where} \tag{8}$$

$$\mathcal{K}_t = \gamma \int_0^t \int_{\partial\Gamma_s} \Big(\frac{\alpha_2}{2}\partial_\nu K^2 - \frac{\alpha_1}{2}\partial_\nu K^1\Big)\,\psi\,, \tag{9}$$

$$K^i(t,x) = \mathbb{E}\left[\kappa_{\partial\Gamma_{\overleftarrow{\tau}^i_t}}(X^i_{\overleftarrow{\tau}^i_t})\,\mathbb{1}_{\{\overleftarrow{\tau}^i_t\le t\}} \;\middle|\; X^i_0 = x\right], \tag{10}$$

with the backward exit times $\overleftarrow{\tau}^i_t = \inf\{s \in [0,t] :\ X^i_0 + \sqrt{\alpha_i}\,W^i_s + l^i_s \notin \Gamma^i_{t-s}\}$.

Proposition 2. *Suppose that (u^1,u^2,Γ) is a classical solution of the Stefan problem with surface tension (2a)–(2e'). If $\mu^i := \mathbb{P}\circ(X^i)^{-1}$ with X^i as in (3), then (μ^1,μ^2,Γ) is a probabilistic solution of (2a)–(2e').*

Proof. See Appendix A.2. □

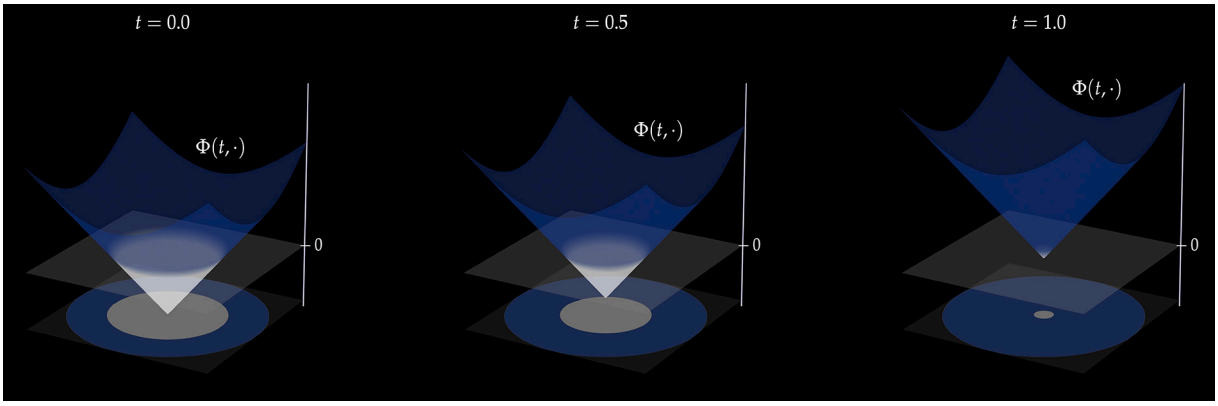


Fig. 2. Time evolution of the level-set function in the radial case and projected solid region (bottom horizontal plane).

3. Deep level-set method

Let $\Gamma = (\Gamma_t)_{t\in[0,T]}$ be the time-varying region occupied by the solid, and let $\Phi : [0,T]\times\Omega \to \mathbb{R}$ be a *level-set function* such that

$$\Gamma_t = \{x \in \Omega : \Phi(t,x) \le 0\}$$

and the free boundary $\partial\Gamma_t$ is given by the zero isocontour $\{\Phi(t,\cdot) = 0\}$. We also assume that the initial region Γ_{0-} is encoded as the zero sublevel set of some function $\Phi_0 : \Omega \to \mathbb{R}$. Whenever $\Phi \in C^{1,2}([0,T]\times\Omega)$ and $\nabla_x\Phi \neq 0$, the normal velocity, outward normal vector, and mean curvature of $\partial\Gamma$ read[1]

$$V = -\frac{\partial_t\Phi}{|\nabla_x\Phi|}, \quad \nu = \frac{\nabla_x\Phi}{|\nabla_x\Phi|}, \quad \kappa_{\partial\Gamma} = \frac{\operatorname{div}(\nu)}{d-1}.$$

In particular, the Stefan condition (2d) can be rewritten as

$$\partial_t\Phi = \frac{\alpha_1}{2L}\nabla_x\Phi\cdot\nabla_x u^1 - \frac{\alpha_2}{2L}\nabla_x\Phi\cdot\nabla_x u^2.$$

The *signed distance function* defined by

$$\rho(t,x) = (-1)^{\mathbb{1}_{\Gamma_t}}\, dist(x,\partial\Gamma_t)$$

can always be used as a level-set function and encodes all geometric properties of the interface; see [3,38] and [29, Section I.2]. In particular, suppose that $\partial\Gamma_t$ is a smooth manifold. Then, ρ is differentiable in a neighborhood of $\partial\Gamma_t$ and satisfies

$$|\nabla_x\rho| = 1, \quad V = -\partial_t\rho, \quad \nu = \nabla_x\rho, \quad \kappa_{\partial\Gamma} = \frac{\Delta_x\rho}{d-1}, \qquad \text{on } \partial\Gamma_t.$$

To approximate a general level-set function Φ, we parameterize the difference $\Phi(t,\cdot) - \Phi_0$ by a neural network $G : [0,T]\times\Omega\times\Theta \to \mathbb{R}$, for some parameter set $\Theta \subset \mathbb{R}^p$, $p \in \mathbb{N}$. This leads to the *deep level-set function*

$$\Phi(t,x;\theta) := \Phi_0(x) + G(t,x;\theta), \quad t \in [0,T], \tag{11}$$

with the associated regions $\Gamma^{1,\theta} = \{\Phi(\cdot\,;\theta) > 0\}$ and $\Gamma^\theta = \Gamma^{2,\theta} = \{\Phi(\cdot\,;\theta) \le 0\}$. By parameterizing the difference we make sure that the initial condition holds automatically. It also allows us to capture the occasional jumps in the solid region at $t = 0$ when the initial data (u_0^1, u_0^2) exhibits discontinuities; see Section 4.1.4.

Example 2. (Radial case) Let $\Omega = B_R := \{|x| \le R\}$ and $\Gamma_{0-} = B_{r_0}$ for some $r_0 \in (0,R)$. We can then initialize the level-set function with the signed distance $\Phi_0(x) = |x| - r_0$. If the initial temperature inside the solid and the liquid is radially symmetric, then $\Gamma_t = B_{r(t)}$ for some càdlàg function $r : [0,T] \to [0,R]$. We can therefore set $\Phi(t,x) = |x| - r(t)$, as illustrated in Fig. 2. This implies that $\Phi(t,x) - \Phi_0(x) = r_0 - r(t)$ for all $x \in \Omega$. Hence, the neural network, although unaware of the radial symmetry, "simply" needs to learn a function of time. Indeed, in our numerical experiments, we do not impose radial symmetry on the solution, but rather let the neural network learn this invariance through training.

[1] The factor $\frac{1}{d-1}$ in the mean curvature ensures that $\kappa_{\partial B}(x) = \frac{1}{|x|}$ for any ball B around $0 \in \mathbb{R}^d$.

3.1. Loss function and training

Let us explain how the parameters in Γ^θ can be trained so as to find a probabilistic solution, which is equivalent to requiring the growth condition (4) to hold for every test function. This is similar to Leray type weak solutions of nonlinear partial differential equations that demand certain equations to be satisfied for a class of test functions, and our approach can possibly be applied to compute this type of solutions as well. We transform the growth condition into a loss function by forcing (4) to hold for a large but finite number of test functions. Specifically, letting $\Psi \subset C_c^\infty(\Omega)$ be a finite collection of test functions, we aim to

$$\text{minimize} \quad \mathcal{L}(\theta) = \sum_{\psi\in\Psi} \mathcal{L}(\theta,\psi) \ \text{ over } \ \theta\in\Theta, \quad \text{where} \tag{12}$$

$$\mathcal{L}(\theta,\psi) = \left\| \int_{\Gamma_{0-}} \psi - \int_{\Gamma^\theta_\cdot} \psi - \frac{1}{L}\left(\eta\, \mathbb{E}[\psi(X^1_{\tau^{1,\theta}})\mathbb{1}_{\{\tau^{1,\theta}\le\cdot\}}] - \mathbb{E}[\psi(X^2_{\tau^{2,\theta}})\mathbb{1}_{\{\tau^{2,\theta}\le\cdot\}}]\right)\right\|^2_{L^2([0,T])}. \tag{13}$$

We expect that Γ^θ induces a probabilistic solution if $\mathcal{L}(\theta)=0$ and Ψ grows to a dense subset of $C_c^\infty(\Omega)$. In order to compute the loss function in (12), we simulate reflected Brownian particles $X^j = (X^j_{t_n})_{t_n\in\Pi^N}$, $j=1,\dots,J$, on a regular time grid $\Pi_N := \{t_n = n\Delta t : n=0,\dots,N\}$, $\Delta t = \frac{T}{N}$. This gives the empirical loss

$$\mathcal{L}(\theta) = \sum_{\psi\in\Psi}\sum_{n=1}^N \ell_n(\theta,\psi)^2, \tag{14}$$

$$\ell_n(\theta,\psi) = \int_{\Gamma_{0-}} \psi - \int_{\Gamma^\theta_{t_n}} \psi - \frac{1}{JL}\sum_{j=1}^J \left(\eta\, \psi(X^{1,j}_{\tau^{1,\theta,j}})\mathbb{1}_{\{\tau^{1,\theta,j}\le t_n\}} - \psi(X^{2,j}_{\tau^{2,\theta,j}})\mathbb{1}_{\{\tau^{2,\theta,j}\le t_n\}}\right). \tag{15}$$

Regarding the test functions, we choose the Gaussian kernels[2]

$$\Psi = \left\{ \psi_k(x) = e^{-\beta_k|x-z_k|^2} \,:\, \beta_k>0,\ z_k\in\Omega,\ k=1,\dots,K \right\}, \quad K\in\mathbb{N}.$$

The centers z_k and widths β_k are randomized in each training iteration to cover the whole domain. The randomization of the centers z_k is particularly helpful when $d=3$, where considering a mesh would become computationally challenging. This choice of test functions is motivated both by theoretical intuition and numerical tests. First, the family Ψ must be dense enough to approximate any functions in $C_c^\infty(\Omega)$. One possibility is to use monomials, although they did not lead to satisfactory results (and are not compactly supported). As we are primarily interested in the free boundary $\partial\Gamma$ (or equivalently, the regions Γ^1, Γ^2), it is expected that functions with small support lead to better precision than monomials capturing global behaviors. Gaussian kernels indeed allow to verify the growth condition locally by choosing large width parameters β_k.

We then train the deep level-set function by gradient descent. That is, the parameters are updated according to

$$\theta_{m+1} = \theta_m - \zeta_m \nabla_\theta \mathcal{L}(\theta_m), \quad m=1,\dots,M, \tag{16}$$

for some $M\in\mathbb{N}$. However, the map $\theta\mapsto\tau^\theta$ is piecewise constant, where τ^θ denotes a stopping time on the right-hand side of (15). For concreteness, let us suppose that τ^θ is the exit time of a liquid particle from the liquid region. In terms of the parameterized level-set function, this reads

$$\tau^\theta = \inf\{t_n, n\le N \,:\, \Phi(t_n, X^1_{t_n};\theta) < 0\}.$$

Unless a trajectory $(X^1_{t_n}(\omega))_{n=0}^N$ is exactly on the solid-liquid interface at the stopping time $\tau^\theta(\omega)$, i.e., $\Phi(\tau^\theta(\omega), X^1_{\tau^\theta}(\omega);\theta)=0$, the value of $\tau^\theta(\omega)$ will remain the same after an infinitesimal change in the parameter vector θ. Hence, the gradient descent would not converge due to the vanishing gradient $\nabla_\theta\tau^\theta$. To circumvent this issue, the stopping times are relaxed according to a procedure explained in the next section.

3.2. Relaxed stopping times

We adapt the relaxation procedure of [35], proposed in the context of optimal stopping. In short, it consists of replacing the sharp boundary $\partial\Gamma_t$ by a *mushy* (or *fuzzy*) *region* where stopping may or may not take place. First, we transform the neural network $\Phi(t,\cdot\,;\theta)$ to locally approximate the signed distance to the interface $\partial\Gamma_t$. This is achieved by normalization, namely by setting $\rho(t,\cdot\,;\theta) = \frac{\Phi(t,\cdot\,;\theta)}{|\nabla_x\Phi(t,\cdot\,;\theta)|}$.[3] We note that the spatial gradient of $\Phi(\cdot;\theta)$ can be exactly and effortlessly computed using automatic differentiation.

[2] If β_k is large enough, then, at least numerically, ψ_k is compactly supported in Ω.

[3] Indeed, write $\phi = \Phi(t,\cdot\,;\theta)$ and fix $x\notin\partial\Gamma_t$, i.e., $\phi(x)\neq 0$. If ν is the outward normal vector at the point on $\partial\Gamma_t$ closest to x, then the signed distance $\rho\in\mathbb{R}$ of x to $\partial\Gamma_t$ can be characterized as the smallest solution in absolute value to $\phi(x-\rho\nu)=0$. A Taylor approximation gives $0=\phi(x-\rho\nu)\approx\phi(x)-\rho\,\nu\cdot\nabla\phi(x)$, so $\rho\approx\frac{\phi(x)}{|\nabla\phi(x)|}$ as desired. Note that this approximation is only accurate close to the interface, which is precisely where ρ plays a role in the algorithm.

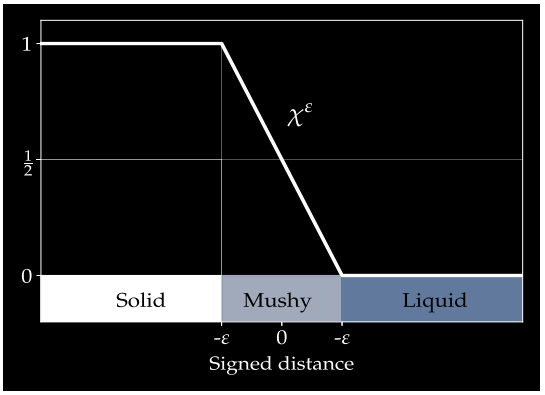


Fig. 3. Illustration of the relaxed phase function $\chi^\epsilon(\rho) = \frac{(1-\rho/\epsilon)^+}{2} \wedge 1$.

Next, the approximate signed distance is converted into a stopping probability. Without loss of generality, we describe the procedure for liquid particles which are stopped when entering the solid region. For solid particles, see Remark 3. Given $\varepsilon > 0$, define the phase indicator variables

$$q_n^{\theta,\varepsilon} = (\chi^\epsilon \circ \rho)(t_n, X^1_{t_n}; \theta), \quad n = 0, \dots, N,$$

with the *relaxed phase function* $\chi^\epsilon(\rho) = \frac{(1-\rho/\epsilon)^+}{2} \wedge 1$; see Fig. 3 for an illustration. The phase indicator variable is equal to 1 in the "ε-interior" of the solid region (where $\rho(t,x;\theta) \le -\varepsilon$) and is 0 when $\rho(t,x;\theta) \ge \varepsilon$. Inside $\partial\Gamma_t^{\theta,\varepsilon}$, the phase indicator variable decreases linearly with the signed distance to the interface. The parameter $\varepsilon > 0$ specifies the width of the *mushy region*

$$\partial\Gamma_t^{\theta,\varepsilon} = \{x \in \Omega : \ |\rho(t,x;\theta)| < \varepsilon\},$$

in which a liquid particle enters the solid region with probability $q_n^{\theta,\varepsilon} \in (0,1)$.

For Brownian particles, a natural choice for ε is $\varepsilon = \sqrt{d\ \Delta t}$, this being the order of the typical Euclidean distance traveled by a d-dimensional standard Brownian motion over a time sub-interval $[t, t+\Delta t]$. For Brownian particles with diffusivity $\alpha > 0$, we set instead $\varepsilon = \sqrt{\alpha d \Delta t}$. When the liquid and the solid particles have different diffusivities, we consider distinct mushy regions. Finally, the stopped values in (15) are replaced by

$$\psi(X^1_{\tau^{1,\theta}})\, \mathbb{1}_{\{\tau^{1,\theta} \le t_n\}} \approx \sum_{l=1}^{n} Q_l^{\theta,\varepsilon}\ \psi(X^1_{t_l}),$$

with the random stopping probabilities recursively given by

$$Q_n^{\theta,\varepsilon} = \underbrace{q_n^{\theta,\varepsilon}\left(1 - \sum_{l<n} Q_l^{\theta,\varepsilon}\right)}_{\text{same phase until } t_n}, \quad n = 0, \dots, N.$$

Similarly, the integrals in (15) are computed as

$$\int_{\Gamma^\theta_{t_n}} \psi \ \approx\ \frac{|\Omega|}{J} \sum_{j=1}^{J} q^{\theta,\varepsilon}_{n,0,j}\ \psi(U^j), \qquad q^{\theta,\varepsilon}_{n,0,j} = (\chi^\epsilon \circ \rho)(t_n, U^j; \theta), \quad U^j \overset{\text{i.i.d.}}{\sim} \text{Unif}(\Omega), \tag{17}$$

where $\text{Unif}(\Omega)$ is the uniform distribution on Ω. Notice that the terms $\int_{\Gamma_{0-}} \psi_k$ do not depend on θ so they can be computed only once, e.g., in an offline phase. Because of the above relaxation, the gradient of $\mathcal{L}(\theta)$ does not vanish anymore. We can therefore apply gradient descent to minimize the loss function, thus finding an approximation of the solid region Γ.

Remark 3. For solid particles, the phase indicator variables are defined analogously by

$$q_n^{\theta,\varepsilon} = 1 - (\chi^\epsilon \circ \rho)(t_n, X^2_{t_n}; \theta).$$

We can thereafter follow the steps above to approximate the stopped values $\psi(X^2_{\tau^{2,\theta}})\, \mathbb{1}_{\{\tau^{2,\theta} \le t_n\}}$.

3.3. Jump penalty

So far, there is no control on the change in volume of the solid region. Although jumps can be observed [12], we may find a solution which satisfies the growth condition but exhibits non-physical jumps, i.e., jumps of non-physically large size. To deal with this issue, we consider the more general loss function $\underline{\mathcal{L}}(\theta) = \mathcal{L}(\theta) + \lambda \mathcal{P}(\theta)$, with the penalty term

$$\mathcal{P}(\theta) = \mathcal{R}\left(\max_{n=1,\dots,N} |\Gamma^\theta_{t_n} \triangle \Gamma^\theta_{t_{n-1}}| - C \right), \quad C > 0, \quad \mathcal{R}(\Delta) = \Delta^+ - 0.01\Delta^-, \tag{18}$$

where $\triangle$ is the symmetric difference operation, and $C > 0$ is a constant describing the "allowed" volume increment between time steps. We note that $\mathcal{R}$ is the so-called *Leaky ReLU* activation function, whose choice is justified as follows. If the maximum change in volume exceeds the threshold C, the deep level-set function is severely penalized by the "ReLU" part of $\mathcal{R}$ (namely, Δ^+). On the other hand, the "leaky" part of $\mathcal{R}$, i.e., $0.01\Delta^-$, is introduced to slightly incentivize the solid to shrink/grow smoothly over time.

The Lagrange multiplier λ is decomposed into the product of a fixed ratio $\lambda_0 > 0$ (typically less than one) and a dynamically updated factor λ^{scale} such that $\lambda^{\text{scale}} \nabla_\theta \mathcal{P}(\theta)$ matches the scale of the gradient $\nabla_\theta \mathcal{L}(\theta)$. For this task, we employ the "learning rate annealing algorithm" outlined in Wang et al. [47]. It is a popular normalization technique for physics-informed neural networks when the loss function comprises heterogeneous components. Finally, the volume of $\Gamma^\theta_{t_n} \triangle \Gamma^\theta_{t_{n-1}}$ is estimated using a Monte Carlo simulation as in (17).

Throughout the numerical experiments, we set the jump threshold to $C = \frac{|\Omega|}{2}$. In other words, the solid region is told to grow/shrink by at most half the size of the domain per time step. This is a loose constraint, which nevertheless rules out undesirable time discontinuities as we shall see in the numerical experiments. At the same time, we shall see in Section 4.1.4 that the regularized algorithm is still capable of producing *physical jumps*, whose definition is now recalled. We also refer the interested reader to [26]. Let us focus on the radial case and consider for $r \geq 0$, $\Delta \in \mathbb{R}$, the annulus

$$A_{r,\Delta} := \begin{cases} B_{r+\Delta} \setminus B_r, & \Delta \geq 0, \\ B_r \setminus B_{r+\Delta}, & -r \leq \Delta < 0, \\ B_r, & \Delta < -r. \end{cases} \tag{19}$$

A radial solution $(\Gamma_t)_{t\in[0,T]} = (B_{r(t)})_{t\in[0,T]}$ to the Stefan problem is *physical* if for all $t \in [0,T]$, the jump $\Delta r(t) = r(t) - r(t-)$, if positive, satisfies

$$\Delta r(t) = \inf\left\{ \Delta > 0 \ : \ |A_{r(t-),\Delta}| > -\frac{1}{L} \int_{A_{r(t-),\Delta}} u^1(t-,x) \right\}. \tag{20}$$

If $\Delta r(t) = r(t) - r(t-) > 0$, we gather that for physical solutions, $\Delta r(t)$ is the smallest solution of

$$|A_{r(t-),\Delta}| = -\frac{1}{L} \int_{A_{r(t-),\Delta}} u^1(t-,x). \tag{21}$$

The left-hand side in (21) is the volume absorbed by the solid region Γ at time t. The right-hand side in (21) is the aggregate change in temperature upon freezing of $A_{r(t-),\Delta}$.

Remark 4. In the non-radial case, it is not sufficient to consider the volume increments only. The formulation of an appropriate physicality condition is in fact subject of ongoing research.

3.4. Algorithm

The deep level-set method is summarized in Algorithm 1. The learning rate process $(\zeta_m)_{m=1}^M$ is computed using the Adam optimizer [21]. For the Monte Carlo simulation of particles, we use antithetic sampling to explore the domain in a symmetric fashion and to speed up computation. More specifically, we simulate two antithetic particles for each generated initial point x_0^i, namely

$$\begin{aligned} X^i_\cdot &= x^i_0 + \sqrt{\alpha_i}\, W^i_{\cdot\wedge\tau^i} + l^i_{\cdot\wedge\tau^i}, \quad \tau^i = \inf\{t \in [0,T] : x^i_0 + \sqrt{\alpha_i}\, W^i_t + l^i_t \notin \Gamma^i_t\}, \\ \widetilde{X}^i_\cdot &= x^i_0 - \sqrt{\alpha_i}\, W^i_{\cdot\wedge\widetilde{\tau}^i} + \widetilde{l}^i_{\cdot\wedge\widetilde{\tau}^i}, \quad \widetilde{\tau}^i = \inf\{t \in [0,T] : x^i_0 - \sqrt{\alpha_i}\, W^i_t + \widetilde{l}^i_t \notin \Gamma^i_t\}. \end{aligned}$$

3.5. Adding surface tension

3.5.1. Growth condition revisited

The numerical verification of the growth condition given in Definition 2 is not straightforward because of the additional term $\mathcal{K}_t$. It turns out that an alternative formulation, which we now outline, greatly simplifies the implementation. In what follows, we assume that $d \geq 2$. For $\delta > 0$, define the one-sided mushy regions,

$$\partial\Gamma^{\delta,i}_t = \{x \in \Gamma^i_t : \ dist(x, \partial\Gamma_t) \leq \delta\}, \quad i = 1,2.$$

Algorithm 1 (Deep level-set method).

Given: Φ_0 = initial level-set function, J = batch size, M = # training iterations, Ψ = set of test functions, C = jump constant, ϵ = mushy region width

I. **Initialize** $\theta_0 \in \Theta$

II. For $m = 0, \dots, M-1$:

1. **Simulate** $X^{i,j} = (X^{i,j}_{t_n})_{n=0}^N$ ($i=1$: liquid, $i=2$: solid), $U^j \overset{\text{i.i.d.}}{\sim} \text{Unif}(\Omega)$, $j = 1, \dots, J$
2. **Approximate** the signed distance function of $\partial\Gamma$ as $\rho(\cdot;\theta_m) = \frac{\Phi(\cdot;\theta_m)}{|\nabla_x \Phi(\cdot;\theta_m)|}$
3. For $n = 0, \dots, N$, $\psi \in \Psi$, **compute**:
   - **Phase indicator variables:** $q^{\theta_m,\epsilon}_{n,i,j} = \begin{cases} (\chi^\epsilon \circ \rho)(t_n, U^j; \theta_m), & i=0, \\ (\chi^\epsilon \circ \rho)(t_n, X^{1,j}_{t_n}; \theta_m), & i=1, \\ 1 - (\chi^\epsilon \circ \rho)(t_n, X^{2,j}_{t_n}; \theta_m), & i=2 \end{cases}$
   - **Symmetric Differences:** $|\Gamma^{\theta_m}_{t_n} \triangle \Gamma^{\theta_m}_{t_{n-1}}| \approx \frac{|\Omega|}{J} \sum_{j=1}^J |q^{\theta_m,\epsilon}_{n,0,j} - q^{\theta_m,\epsilon}_{n-1,0,j}|$
   - **Integrals:** $\int_{\Gamma^{\theta_m}_{t_n}} \psi \approx \frac{|\Omega|}{J} \sum_{j=1}^J q^{\theta_m,\epsilon}_{n,0,j} \psi(U^j)$
   - **Stopping probabilities:** $Q^{\theta_m,\epsilon}_{n,i,j} = q^{\theta_m,\epsilon}_{n,i,j}(1 - \sum_{l<n} Q^{\theta_m,\epsilon}_{l,i,j})$, $i = 1,2$
   - **Stopped values:** $\psi(X^{i,j}_{\tau^{\theta_m}}) \mathbb{1}_{\{\tau^{\theta_m} \le t_n\}} \approx \sum_{l=1}^n Q^{\theta_m,\epsilon}_{l,i,j} \psi(X^{i,j}_{t_l})$
4. **Loss**: $\mathcal{L}(\theta_m) = \sum_{\psi\in\Psi} \sum_{n=1}^N \ell_n(\theta_m,\psi)^2$, with

$$\ell_n(\theta_m,\psi) = \int_{\Gamma_{0-}} \psi - \int_{\Gamma^{\theta_m}_{t_n}} \psi - \frac{1}{JL} \sum_{j=1}^J \left(\eta\, \psi(X^{1,j}_{\tau^{1,\theta_{m,j}}}) \mathbb{1}_{\{\tau^{1,\theta_{m,j}} \le t_n\}} - \psi(X^{2,j}_{\tau^{2,\theta_{m,j}}}) \mathbb{1}_{\{\tau^{2,\theta_{m,j}} \le t_n\}} \right)$$

5. **Jump penalty:** $\mathcal{P}(\theta_m) = \mathcal{R}\left(\max_{n=1,\dots,N} |\Gamma^{\theta_m}_{t_n} \triangle \Gamma^{\theta_m}_{t_{n-1}}| - C \right)$
6. **Gradient step:** $\theta_{m+1} = \theta_m - \zeta_m \nabla_\theta \underline{\mathcal{L}}(\theta_m)$, $\underline{\mathcal{L}}(\theta_m) = \mathcal{L}(\theta_m) + \lambda_0\, \lambda_m^{\text{scale}}\, \mathcal{P}(\theta_m)$

III. **Return** $\Gamma^{\theta_M} = \{\Phi(\cdot\,;\theta_M) \le 0\}$

Let $(T^{\delta,i}_l)_{l\ge1}$ be the arrival times of a time-space Poisson point process with intensity

$$\lambda^i(t,y) = \frac{\gamma}{\delta} |\kappa_{\partial\Gamma_t}(\hat{y})|^{2-d} \frac{\mathbb{1}_{\partial\Gamma^{\delta,i}_t}(y)}{|\partial\Gamma^{\delta,i}_t|}, \quad \hat{y} = \text{proj}_{\partial\Gamma_t}(y), \quad d \ge 2, \tag{22}$$

where $\text{proj}_{\partial\Gamma_t}(y)$ is the projection of y onto $\partial\Gamma_t$, i.e., the point on $\partial\Gamma_t$ closest to y. For each $l \ge 1$ and $i = 1,2$, we simulate a Brownian particle $Y^{l,i} = (Y^{l,i}_s)_{s\in[T^{\delta,i}_l, T]}$ where, according to (22), the initial position of $Y^{l,i}$ is uniformly distributed in $\partial\Gamma^{\delta,i}_{T^{\delta,i}_l}$ if $d=2$ and inversely proportional to the absolute mean curvature of $\partial\Gamma_{T^{\delta,i}_l}$ otherwise. We also define the exit times $\tau^{\delta,i}_l = \inf\{s \ge T^{\delta,i}_l : Y^{l,i}_s \notin \Gamma^i_s\}$.

In the three-dimensional, radially symmetric case (namely $\Gamma_t = B_{r(t)} \subseteq \mathbb{R}^3$), the Poisson intensities simply read

$$\lambda^i(t,\cdot) = \frac{\gamma}{\delta} r(t) \frac{\mathbb{1}_{A_{r(t),\Delta_i}}}{|A_{r(t),\Delta_i}|}, \quad \Delta_i = (-1)^{i+1}\delta, \tag{23}$$

with the annuli $A_{r,\Delta}$ defined in (19). We therefore gather that the expected number of simulated particles is directly proportional to the radius of the solid region.

Proposition 3. *Assume that $\alpha_1 = \alpha_2$ and consider a classical solution of the radially symmetric Stefan problem with surface tension. Moreover, let $N^{\delta,i}_t := \sum_{l=1}^\infty \mathbb{1}_{\{T^{\delta,i}_l \le t\}}$ be the counting process associated with $(T^{\delta,i}_l)$, $i = 1,2$. Then for all $\psi \in C^\infty_c(\Omega)$,*

$$\int_{\Gamma_{0-}} \psi - \int_{\Gamma_t} \psi = \frac{1}{L}\left(\eta\, \mathbb{E}^{\mu^1}[\psi(X^1_{\tau^1}) \mathbb{1}_{\{\tau^1 \le t\}}] - \mathbb{E}^{\mu^2}[\psi(X^2_{\tau^2}) \mathbb{1}_{\{\tau^2 \le t\}}] + \mathcal{K}_t \right), \tag{24}$$

$$\mathcal{K}_t = -\mathcal{K}^1_t - \mathcal{K}^2_t, \tag{25}$$

$$\mathcal{K}^i_t = \lim_{\delta\downarrow0} \mathbb{E}\left[\sum_{l=1}^{N^{\delta,i}_t} \left(\psi(Y^{l,i}_{\tau^{\delta,i}_l}) \mathbb{1}_{\{\tau^{\delta,i}_l \le t\}} - \psi(Y^{l,i}_{T^{\delta,i}_l}) \right) \right]. \tag{26}$$

Proof. See Appendix A.3. □

Remark 5. When the diffusivities of the solid and the liquid phase are different ($\alpha_1 \neq \alpha_2$), the right-hand side of the growth condition (24) contains an additional term, namely

$$\widetilde{\mathcal{K}}_t = -\gamma \frac{\alpha_2 - \alpha_1}{2} \int_0^t \int_{\partial\Gamma_s} \kappa_{\partial\Gamma_s}\, \partial_\nu \psi.$$

Recalling that $u^1 = u^2 = -\gamma \kappa_{\partial\Gamma_s}$, the above term can be approximated via

$$\widetilde{\mathcal{K}}_t \approx \frac{\alpha_2}{2\delta} \int_0^t \int_{\partial\Gamma_s^{\delta,2}} u^2\, \partial_\nu \psi - \frac{\alpha_1}{2\delta} \int_0^t \int_{\partial\Gamma_s^{\delta,1}} u^1\, \partial_\nu \psi.$$

The integrals $\int_{\partial\Gamma_s^{\delta,i}} u^i\, \partial_\nu \psi$, in turn, can be computed using the stochastic representation of u^i given in (41), below. Note that $\partial_\nu \psi$ is easy to compute since the outward normal ν is available from the deep level-set function.

Remark 6. It is conjectured that the representation (26) of the curvature terms holds beyond the radial case. The only difference is that the effect of boundary particles is reversed in *concave* sections of the boundary. Specifically, we expect that, in general,

$$\mathcal{K}_t^i = \lim_{\delta \downarrow 0} \mathbb{E}\left[\sum_{l=1}^{N_t^{\delta,i}} \operatorname{sign}\big(\kappa_{\partial\Gamma_{T_l^{\delta,i}}}(\widehat{Y^{l,i}_{T_l^{\delta,i}}})\big) \big(\psi(Y^{l,i}_{\tau_l^{\delta,i}})\, \mathbb{1}_{\{\tau_l^{\delta,i} \leq t\}} - \psi(Y^{l,i}_{T_l^{\delta,i}})\big) \right]. \tag{27}$$

We show how to approximate $\mathcal{K}_t^i$ of (27). First, the mushy regions can be expressed in terms of the level-set function as follows:

$$\partial\Gamma_t^{\delta,1} \approx \left\{ x \in \Omega : \frac{\Phi(t,x)}{|\nabla_x \Phi(t,x)|} \in (0,\delta] \right\}, \quad \partial\Gamma_t^{\delta,2} \approx \left\{ x \in \Omega : \frac{\Phi(t,x)}{|\nabla_x \Phi(t,x)|} \in [-\delta, 0] \right\}.$$

Without loss of generality, we focus on the term $\mathcal{K}_t^1$ and drop the superscripts ^{1} throughout. Using the time discretization $\{t_n = n\Delta t : n = 0, \dots, N\}$, $\Delta t = \frac{T}{N}$, $N \in \mathbb{N}$, we obtain for $\delta > 0$ sufficiently small that

$$\mathcal{K}_{t_n} \approx \sum_{m=0}^{n-1} \mathbb{E}\left[\sum_{l=N_{t_m}^{\delta}+1}^{N_{t_{m+1}}^{\delta}} \mathbb{E}\left[\operatorname{sign}\big(\kappa_{\partial\Gamma_{t_m}}(\widehat{Y^l_{t_m}})\big) \big(\psi(Y^l_{\tau_l^\delta})\, \mathbb{1}_{\{\tau_l^\delta \leq t_n\}} - \psi(Y^l_{t_m})\big) \;\middle|\; T_l^\delta = t_m \right] \right].$$

Writing $\mathbb{E}_{t_m,y}[\cdot] = \mathbb{E}[\,\cdot \mid T_l^\delta = t_m,\ Y^l_{t_m} = y]$ and recalling the Poisson intensity in (22) we find that

$$\begin{aligned} \mathcal{K}_{t_n} &\approx \sum_{m=0}^{n-1} \frac{\gamma \Delta t}{\delta\, |\partial\Gamma_{t_m}^\delta|} \int_{\partial\Gamma_{t_m}^\delta} |\kappa_{\partial\Gamma_{t_m}}(\widehat{y})|^{2-d}\, \mathbb{E}_{t_m,y}\left[\operatorname{sign}\big(\kappa_{\partial\Gamma_{t_m}}(\widehat{y})\big) \big(\psi(Y_{\tau^\delta})\, \mathbb{1}_{\{\tau^\delta \leq t_n\}} - \psi(y)\big) \right] \mathrm{d}y \\ &= \frac{\gamma \Delta t}{\delta} \sum_{m=0}^{n-1} \mathbb{E}\left[\operatorname{sign}\big(\kappa_{\partial\Gamma_{t_m}}(\widehat{Y_{t_m}})\big) \Big|\kappa_{\partial\Gamma_{t_m}}(\widehat{Y_{t_m}})\Big|^{2-d}\, \mathbb{E}\left[\psi(Y_{\tau^\delta})\, \mathbb{1}_{\{\tau^\delta \leq t_n\}} - \psi(Y_{t_m}) \;\middle|\; T^\delta = t_m,\ Y_{t_m} \right] \right]. \end{aligned} \tag{28}$$

The latter is estimated through Monte Carlo simulation where, for simplicity, the curvature is evaluated at Y_{t_m} instead of its projection $\widehat{Y_{t_m}}$ onto $\partial\Gamma_{t_m}$. This is an accurate approximation of the "true" curvature when the width δ of the mushy region is small. What remains is to approximate the mean curvature, which we address in the next section.

3.5.2. Mean curvature approximation

In this section, we present simple algorithms to locally approximate the mean curvature of a manifold $\Sigma \subset \mathbb{R}^d$, $d = 2, 3$, of codimension 1. Other techniques, e.g., using finite differences, are discussed in [32]. If $\Sigma = \{\phi = 0\}$ for some level-set function $\phi \in C^2(\mathbb{R}^d)$, the mean curvature can be expressed as

$$\kappa_\Sigma(y) = \frac{\operatorname{div}(\nu(y))}{d-1}, \quad y \in \Sigma, \tag{29}$$

with the outward normal vector field $\nu = \frac{\nabla\phi}{|\nabla\phi|}$. If ϕ is parameterized by a neural network, one could apply automatic differentiation to compute κ_Σ. But computing second order derivatives through automatic differentiation is costly since it entails nested gradient tapes. Indeed, the machine would need to keep track of the first order derivatives as well. Alternative approaches have been proposed to speed up computation, e.g., using Monte Carlo techniques [37]. Here, we propose a dilation approach echoing the geometric nature of curvature and solely exploiting the gradient of ϕ.

(a) $d = 2$ (b) $d = 3$

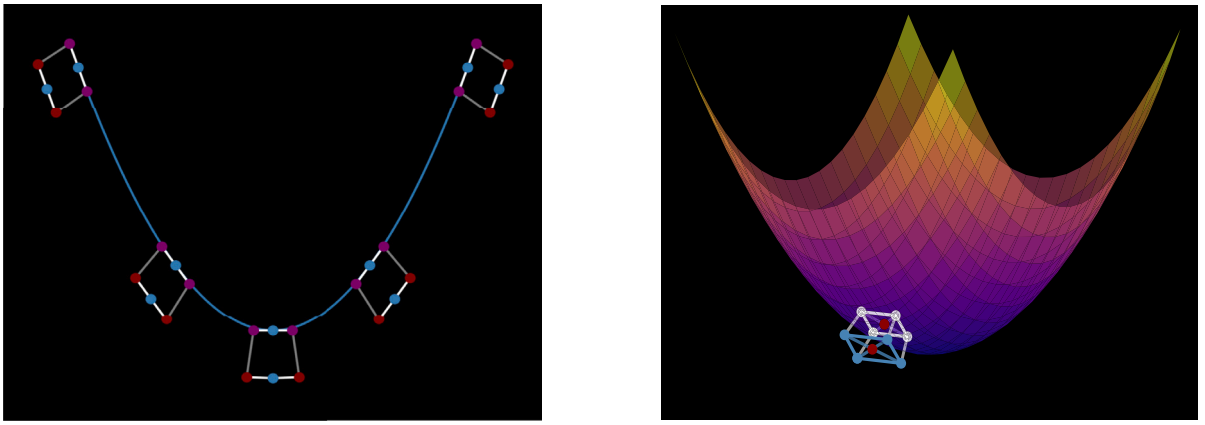


Fig. 4. Dilation technique to capture the mean curvature of codimension 1 interfaces in $\mathbb{R}^d$, $d = 2, 3$.

As a motivation, consider a circle $\Sigma_r = \partial B_r \subseteq \mathbb{R}^2$ of radius r whose curvature is $\kappa_{\Sigma_r} = \frac{1}{r}$. One way to approximate the curvature is to look at the ratio between arc lengths of Σ_r and the dilated circle $\Sigma_{r+\epsilon}$ for small $\epsilon > 0$. Indeed, writing $\Sigma_r \supset \mathcal{A}_{r,\epsilon_0} = \{(r\cos(t\epsilon_0), r\sin(t\epsilon_0)) : t \in [0,1]\}$ for an arc with radius r and angle $\epsilon_0 \in [0, 2\pi)$ we see that

$$\frac{\mathcal{H}^1(\mathcal{A}_{r+\epsilon,\epsilon_0})}{\mathcal{H}^1(\mathcal{A}_{r,\epsilon_0})} = \frac{\epsilon_0(r+\epsilon)}{\epsilon_0 r} = 1 + \epsilon\, \kappa_{\Sigma_r} \implies \kappa_{\Sigma_r} = \frac{1}{\epsilon}\left(\frac{\mathcal{H}^1(\mathcal{A}_{r+\epsilon,\epsilon_0})}{\mathcal{H}^1(\mathcal{A}_{r,\epsilon_0})} - 1\right). \tag{30}$$

We can mimic this procedure for general subsets $\Sigma \subset \mathbb{R}^2$, $\dim_{\mathcal{H}}(\Sigma) = 1$, by approximating Σ locally by a circle. The procedure is summarized in Algorithm 2 where Σ is the zero level set of some function $\phi \in C^1(\mathbb{R}^2)$. An illustration is also given in Fig. 4a. When $\epsilon_0 > 0$ is small, the arc of length ϵ_0 in (30) is accurately replaced by a segment tangent to Σ. We recall that the outward normal vector field of Σ is readily available from the level-set function. The tangent vector field is therefore available as well.

Algorithm 2 (Local approximation of curvature, $d = 2$).

Given: $\phi \in C^1(\Omega)$ $(\nabla\phi \neq 0)$, $\Sigma = \{\phi = 0\}$, $y \in \Sigma$, $\epsilon_0 > 0$, $\epsilon > 0$

I. **Pick** a direction μ of the tangent line at y, i.e., perpendicular to $\nu(y) = \frac{\nabla\phi(y)}{|\nabla\phi(y)|}$
II. **Consider**
  1. the segment $S_y \subseteq \mathbb{R}^2$ with endpoints $y^{\pm} = y \pm \epsilon_0\mu$
  2. the dilated segment $S_y^{\epsilon} \subseteq \mathbb{R}^2$ with endpoints $y^{\pm} + \epsilon\, \nu(y^{\pm})$
III. **Compute** the ratio $\frac{\mathcal{H}^1(S_y^{\epsilon})}{\mathcal{H}^1(S_y)} \approx 1 + \epsilon\, \kappa_{\Sigma}(y)$
IV. **Return** the curvature $\kappa_{\Sigma}(y) \approx \frac{1}{\epsilon}\left(\frac{\mathcal{H}^1(S_y^{\epsilon})}{\mathcal{H}^1(S_y)} - 1\right)$

A similar procedure can be used in the three-dimensional case, as explained in Algorithm 3 and displayed in Fig. 4b. In short, the segments in Algorithm 2 become quadrangles and lengths are replaced by areas. In Step III of Algorithm 3, note that the mean curvature appears naturally from the cross products in $(1 + \epsilon\kappa_{\Sigma,1})(1 + \epsilon\kappa_{\Sigma,2})$, where $(\kappa_{\Sigma,i})_{i=1}^2$ are the principal curvatures of Σ. As the principal directions are unknown in general, the algorithm picks randomly two orthogonal vectors in the plane tangent to Σ at some point $y \in \Sigma$. This randomization has little impact on the accuracy of the obtained, as can be seen in Fig. 5 and Fig. 6.

Algorithm 3 (Local approximation of mean curvature, $d = 3$).

Given: $\phi \in C^1(\Omega)$ $(\nabla\phi \neq 0)$, $\Sigma = \{\phi = 0\}$, $y \in \Sigma$, $\epsilon_0 > 0$, $\epsilon > 0$

I. **Pick** two directions $\mu_1 \perp \mu_2$ in the tangent plane at y, i.e., perpendicular to $\nu(y) = \frac{\nabla\phi(y)}{|\nabla\phi(y)|}$
II. **Consider**
  1. the quadrangle Q_y with vertices $y_i^{\pm} = y \pm \epsilon_0\mu_i$, $i = 1, 2$
  2. the quadrangle Q_y^{ϵ} with vertices $y_i^{\pm} + \epsilon\, \nu(y_i^{\pm})$, $i = 1, 2$
III. **Compute** the ratio $\frac{\mathcal{H}^2(Q_y^{\epsilon})}{\mathcal{H}^2(Q_y)} \approx (1 + \epsilon\, \kappa_{\Sigma,1}(y))(1 + \epsilon\, \kappa_{\Sigma,2}(y)) \approx 1 + 2\epsilon\, \kappa_{\Sigma}(y)$
IV. **Return** the mean curvature $\kappa_{\Sigma}(y) \approx \frac{1}{2\epsilon}\left(\frac{\mathcal{H}^2(Q_y^{\epsilon})}{\mathcal{H}^2(Q_y)} - 1\right)$

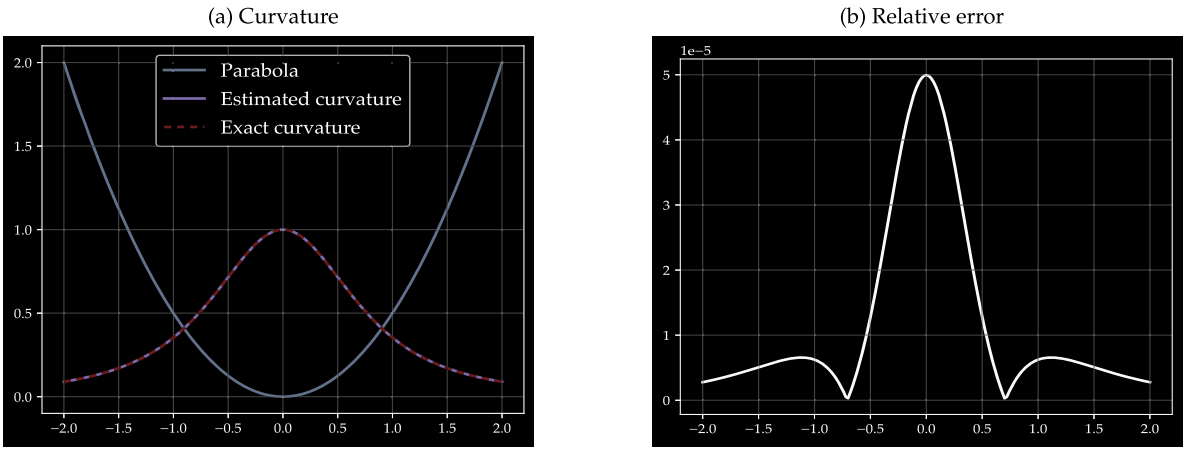


Fig. 5. Estimated curvature (left panel) and relative error (right panel) for a parabola ($d = 2$) with parameter $a = 2$.

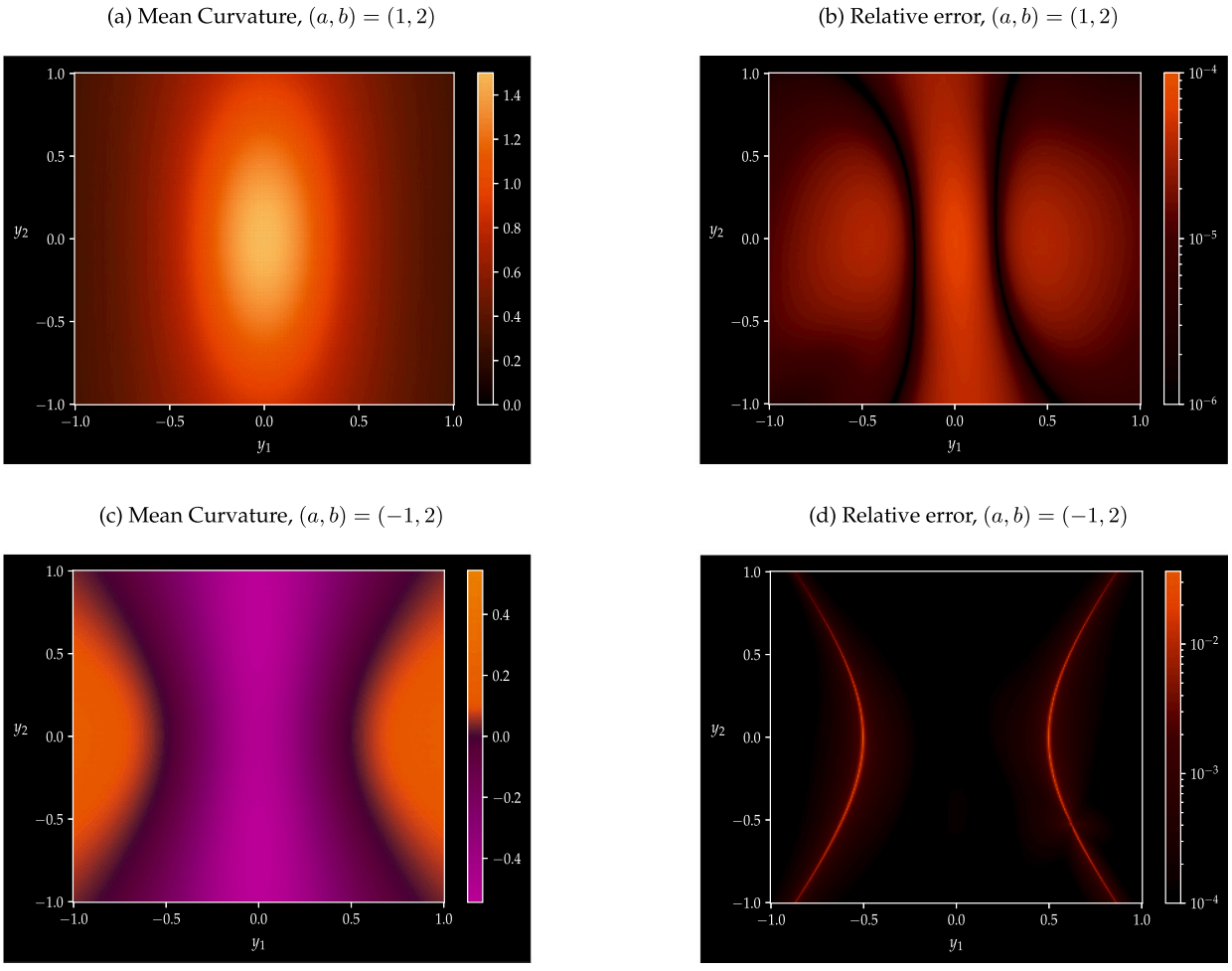


Fig. 6. Estimated mean curvature (left panels) and relative error (right panels) for a paraboloid ($d = 3$) with parameters $(a, b) = (1, 2)$ (top panels) and $(a, b) = (-1, 2)$ (bottom panels).

Back to the Stefan problem, we naturally choose $\Sigma = \partial\Gamma_t$, $t \in [0, T]$, and $\epsilon \ll \epsilon_0 \ll \delta$, where δ is the width of the fuzzy regions $\partial\Gamma_t^{\delta,1}$, $\partial\Gamma_t^{\delta,2}$. Indeed, $\epsilon \ll \epsilon_0$ is imposed to gain accuracy while $\epsilon_0 \ll \delta$ ensures that the points $y^{\pm}$ in Algorithm 2 ($y_i^{\pm}$ in Algorithm 3) belong to the fuzzy region for all $y \in \partial\Gamma_t$.

Table 1
Default Training Parameters (d = number of space dimensions).

Parameter	Definition	Value
J	Batch size	2^{7+d}
M	Number of training iterations	3000
K	Number of test functions ($\|\Psi\|$)	$100d$
N	Number of time steps	100
ϵ_i	Mushy region width of phase $i \in \{1,2\}$	$\sqrt{\alpha_i d \frac{T}{N}}$
λ_0	Lagrange multiplier	0.1

Table 2
Default problem parameters.

Parameter	Definition	Value
T	Time horizon	1
R	Radius of spherical domain Ω	1
α_1	Diffusivity of liquid particles	0.5
α_2	Diffusivity of solid particles	0.5
γ	Surface tension coefficient	0

Let us verify Algorithms 2 and 3, respectively, for parabolas $\Sigma_a := \{y \in \mathbb{R}^2 : y_2 = \frac{y_1^2}{a}\}$ and paraboloids $\Sigma_{a,b} := \{y \in \mathbb{R}^3 : y_3 = \frac{y_1^2}{a} + \frac{y_2^2}{b}\}$ with parameters $a, b \in \mathbb{R} \setminus \{0\}$. The mean curvature functions are given, respectively, by

$$\kappa_{\Sigma_a}(y) = \frac{2}{a\left(1 + \frac{4y_1^2}{a^2}\right)^{3/2}} \quad (d=2), \qquad \kappa_{\Sigma_{a,b}}(y) = \frac{\frac{4y_1^2}{a} + \frac{4y_2^2}{b} + a + b}{ab\left(1 + \frac{4y_1^2}{a^2} + \frac{4y_2^2}{b^2}\right)^{3/2}} \quad (d=3). \tag{31}$$

Fig. 5 shows the approximated curvature and the error relative to (31) for a parabola ($d = 2$) with parameter $a = 2$. In Fig. 6, we repeat the exercise for a paraboloid with parameters $(a, b) = (1, 2)$ (top panels) and a hyperbolic paraboloid with $(a, b) = (-1, 2)$ (bottom panels). As expected, we note that the error for the hyperbolic paraboloid in Fig. 6d is only sizeable along the points with zero mean curvature.

4. Numerical results

Table 1 and Table 2 report the default training and problem parameters, respectively, across the examples. The value of M in Table 1 (3000) indicates the *maximum* number of iterations if the loss function has not stabilized yet. For the algorithm with surface tension, we set $M = 1000$ as each training iteration is more costly. Note that the Lagrange multiplier in front of the penalty function is less than one to give more importance to the loss term than the penalty. The deep-level set function in (11) consists of a feedforward neural network with input (t, x) of size $\bar{d} = d + 1$ and two hidden layers with $20 + \bar{d}$ hidden nodes. The code is implemented in Python 3.9 using Tensorflow 2.7 and run on CPU (10 cores) on a 2021 Macbook pro with 64 GB unified memory and Apple M1 Max chip.

4.1. Radial case

We first discuss examples where the solid is radially symmetric. For concreteness, we regard the solid as an ice ball surrounded by (liquid) water.

4.1.1. One-phase, melting regime ($d = 2$)

Consider the one-phase Stefan problem without surface tension as in Example 1. Given $\Gamma_{0-} = B_{1/2} \subset B_1 = \Omega$, and assuming the initial temperature u_0^1 to be radially symmetric, the solid (ice) remains radial as well, i.e., $\Gamma_t = B_{r(t)}$ for some càdlàg function $r : [0, T] \to [0, R]$ (see Fig. 7). We also set $\eta = 1$, and $L = 1/4$. The temperature in the liquid (water) is initially constant, namely $u_0^1 \equiv \frac{1}{|B_R \setminus B_{r_0}|}$.

Fig. 8 compares the behavior of $r(t)$ as t approaches the melting time with the theoretical asymptotic rate given by Hadzic and Raphael [17], namely

$$r(t) \sim \sqrt{\tau_\varnothing - t}\, e^{-\sqrt{\frac{1}{2}|\log(\tau_\varnothing - t)|}}, \quad t \uparrow \tau_\varnothing := \inf\{t \geq 0 : \Gamma_t = \varnothing\}. \tag{32}$$

As can be observed, the melting rate obtained with the deep level-set method is indeed close to the theoretical one.

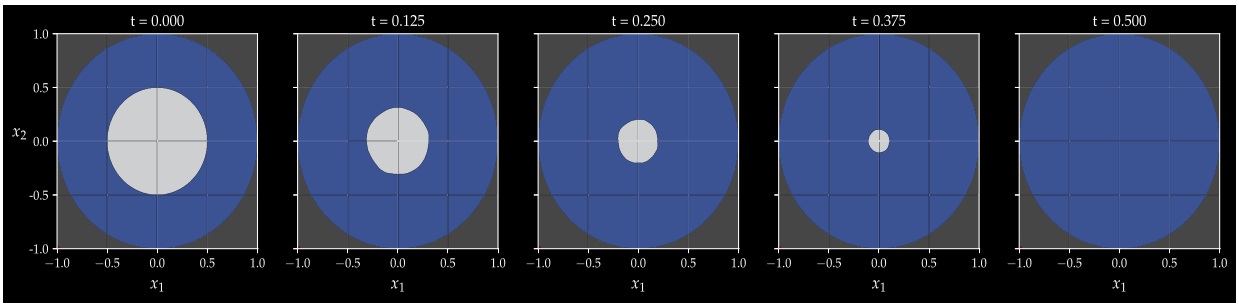


Fig. 7. One-phase Stefan problem in the radially symmetric case ($d = 2$). Evolution of the solid and the liquid regions over time.

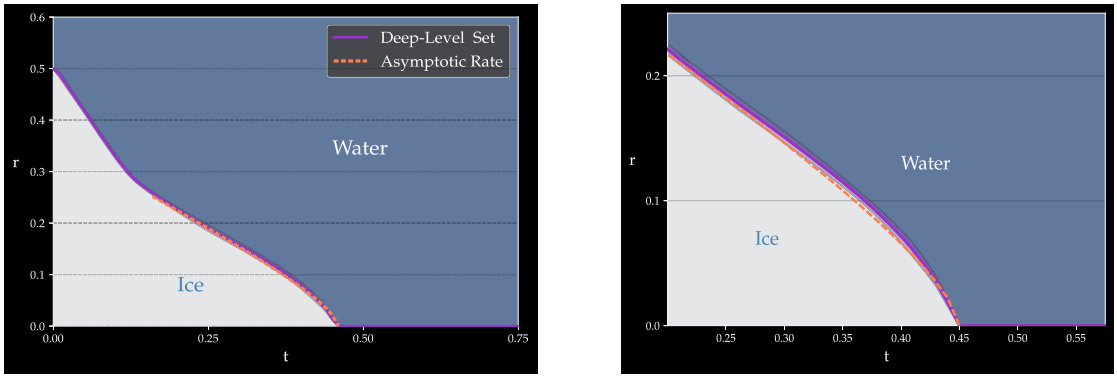


Fig. 8. Mean radius of the solid region over time (left panel) and zoomed in on the melting time (right panel). The light purple curves represent one standard deviation above and below the mean radius for various angles. The orange dashed curve is the theoretical melting rate from [17].

4.1.2. *Two-phase, freezing regime* ($d = 2$)

Let $d = 2$ and consider the two-phase, radially symmetric supercooled Stefan problem without surface tension. Suppose that the temperature is initially constant in the liquid and the solid regions, namely

$$u_0^1 = -\frac{\mathbb{1}_{B_R \setminus B_{r_0}}}{2|B_R \setminus B_{r_0}|}, \qquad u_0^2 = -\frac{1}{10|B_{r_0}|}\mathbb{1}_{B_{r_0}}.$$

We therefore choose the constants $c_1 = 1/2$, $c_2 = 1/10$ in Remark 2. In this example, the radius of the solid region converges to some constant $r_\infty \in (0, R)$. In other words, the solid region neither melts completely nor covers the whole domain in the long run. In light of the Dirichlet condition (2e), the long-term temperature must be constant and equal to $u^1(\infty, \cdot) \equiv u^2(\infty, \cdot) \equiv 0$. In fact, the theoretical value for r_∞ can be derived from the Stefan growth condition. Indeed, choosing an increasing sequence $\psi_m \in C_c^\infty(\Omega)$ such that $\psi_m \uparrow 1$ in Ω and time integrating (38) in the proof of Proposition 1 between $t = 0-$ and $t = T \to \infty$ yields

$$|\Gamma_\infty| - |\Gamma_{0-}| = \frac{1}{L}\sum_{i=1}^{2}\left(\int_{\Gamma_\infty^i} u_\infty^i - \int_{\Gamma_{0-}^i} u_0^i\right) \iff \pi r_\infty^2 - \pi r_0^2 = \frac{c_1 + c_2}{L}. \tag{33}$$

The long-term radius is therefore $r_\infty = \sqrt{r_0^2 + \frac{c_1+c_2}{L\pi}}$. We can thus solve for r_∞ and compare it with the obtained long-term radius. We choose $r_0 = 1/2$, $L = 4$, and $T = 5$. The long-term radius is roughly equal to 0.59, as confirmed in Fig. 9.

Let us now estimate the temperature in both phases as a function of time and radius from the trained moving solid. In view of (5), we here have that $u^i(t, x)\,\mathrm{d}x = -c_i\,\mathbb{P}(X_t^i \in \mathrm{d}x, \tau^i > t)$ for $t \in [0-, T]$, $i = 1, 2$. Writing $v^i(t, |x|) := u^i(t, x)$ and using polar coordinates yields

$$v^i(t, r) = -\frac{c_i}{2\pi r}\underbrace{\frac{\mathbb{P}(|X_t^i| \in \mathrm{d}r \mid \tau^i > t)}{\mathrm{d}r}}_{(a)}\underbrace{\mathbb{P}(\tau^i > t)}_{(b)}.$$

The term (a) is the time t conditional density of the norm of surviving particles which can be approximated using Monte Carlo simulation and kernel density estimation (KDE). The second term (b) is the unconditional survival probability over time and can also be estimated using simulation. Fig. 10 displays the temperature of the solid and the liquid over time. As can be seen, the temperature at time $t = 2$ is already close to its equilibrium $u_\infty^1 = u_\infty^2 = 0$.

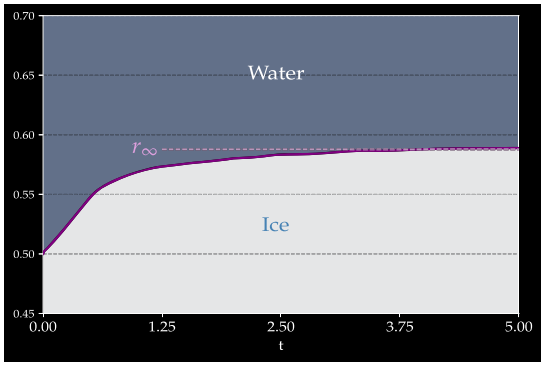


Fig. 9. Two-phase, supercooled Stefan problem without surface tension ($d = 2$) discussed in Section 4.1.2. The radius indeed converges to the theoretical long-term value $r_\infty \approx 0.59$.

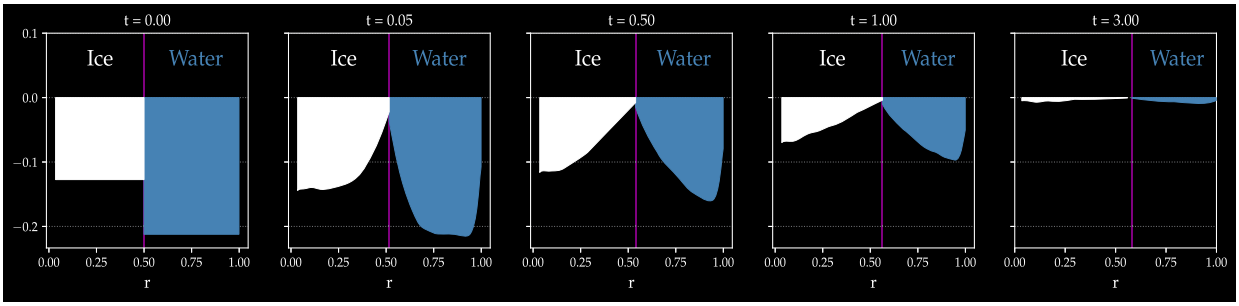


Fig. 10. Two-phase, supercooled Stefan problem without surface tension ($d = 2$) discussed in Section 4.1.2. Temperature (vertical axis) in the liquid (u^1) and the solid (u^2) retrieved using Monte Carlo simulation and kernel density estimation (KDE). At $t = 3$, both phases are already close to their equilibrium temperature $u^1_\infty = u^2_\infty = 0$ and long term radius $r_\infty \approx 0.59$.

4.1.3. *Two-phase, with surface tension* ($d = 3$)

Consider the three-dimensional Stefan problem with surface tension. The initial temperatures in the solid and the liquid are respectively given by

$$u_0^1 = -\frac{\mathbb{1}_{B_R \setminus B_{r_0}}}{2|B_R \setminus B_{r_0}|}, \quad u_0^2 = -\frac{\mathbb{1}_{B_{r_0}}}{|B_{r_0}|}.$$

We therefore have $\eta = -1$ (supercooled liquid), $c_1 = 1/2$, and $c_2 = 1$. Again, the solution remains radially symmetric, i.e., $\Gamma_t = B_{r(t)}$ for some function $r : [0, T] \to [0, R]$. The temperature at the interface is therefore $u^1 \equiv u^2 \equiv -\frac{\gamma}{r(t)}$. We use the growth condition in Proposition 3 to train the parameters and estimate the mean curvature via Algorithm 3. To benchmark our method, we apply a useful trick in the three-dimensional radial case to get rid of surface tension. A similar argument is given in [20, Section 4]. For $i = 1, 2$, define

$$\widetilde{u}^i(t, x) = u^i(t, x) + \frac{\gamma}{|x|}. \tag{34}$$

Then $\widetilde{u}^1 \equiv \widetilde{u}^2 \equiv 0$ on $\partial\Gamma$, so the Gibbs-Thomson condition (2e') is effectively "absorbed" by the transformation in (34). In addition, $\Delta\left(|x|^{-1}\right) = (3 - d)|x|^{-3}$, hence $x \mapsto |x|^{-1}$ is harmonic if $d = 3$ and $\widetilde{u}^i$ still solves the heat equation (2a). We can therefore apply the deep level-set method without surface tension by changing the initial condition to $\widetilde{u}^i_0(x) = u^i_0(x) + \frac{\gamma}{|x|}$. It is worth noting that the Stefan condition (2d) is also affected. As $u^i = \widetilde{u}^i - \frac{\gamma}{|x|}$ and $\partial_\nu(\frac{1}{|x|}) = -\frac{1}{|x|^2}$, we obtain

$$V = \frac{\alpha_2}{2L}\partial_\nu u^2 - \frac{\alpha_1}{2L}\partial_\nu u^1 = \frac{\alpha_2}{2L}\partial_\nu \widetilde{u}^2 - \frac{\alpha_1}{2L}\partial_\nu \widetilde{u}^1 + \gamma\frac{\alpha_2 - \alpha_1}{2L|x|^2}.$$

If $\alpha_1 = \alpha_2$, then the Stefan condition remains unchanged. Otherwise, when $t \mapsto r(t)$ is strictly monotone, the growth condition in (1) becomes

$$\int_{\Gamma_{0-}} \psi - \int_{\Gamma_t} \psi = \frac{1}{L}\left(\mathbb{E}^{\mu^1}[\psi(X^1_{\tau^1})\mathbb{1}_{\{\tau^1 \le t\}}] - \mathbb{E}^{\mu^2}[\psi(X^2_{\tau^2})\mathbb{1}_{\{\tau^2 \le t\}}] + \mathcal{K}_t \right), \quad \psi \in C_c^\infty(\Omega),$$

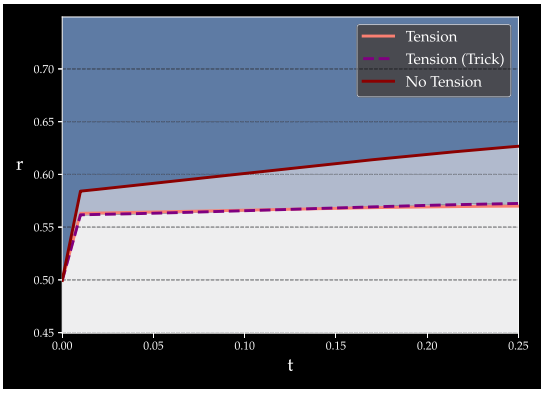


Fig. 11. Two-phase, radially symmetric Stefan problem ($d = 3$) with surface tension ($\gamma = 0.25$) and without. Time evolution of the radius of the solid obtained from the growth condition in Proposition 3 (orange curve), using the radial trick (dashed purple curve), and without surface tension (red curve).

$$\mathcal{K}_t = \gamma \frac{\alpha_1 - \alpha_2}{2} \int_0^t \int_{\partial\Gamma_s} \frac{\psi}{|x|^2} = \gamma \frac{\alpha_1 - \alpha_2}{2} \left(\int_{\Gamma_t} \frac{\psi}{|x|^2} - \int_{\Gamma_0} \frac{\psi}{|x|^2} \right).$$

In this example, we assume that $\alpha_1 = \alpha_2 = \frac{1}{2}$, $L = 2$, $\gamma = 0.25$. The above remedy when $\alpha_1 \neq \alpha_2$ is therefore not needed. We note that after the radial trick the liquid is still supercooled, i.e., $\widetilde{u}_0^1 \leq 0$.

Fig. 11 compares the radius of the solid over time obtained from the growth condition in Proposition 3 (orange curve), using the radial trick (dashed purple curve), and without surface tension (red curve). The algorithm with surface tension and its benchmark indeed gives similar results. As expected, the growth of the solid ball is less pronounced with surface tension as the freezing point becomes negative for convex solid regions.

4.1.4. *Two-phase, jump in the radius* ($d = 2$)

Consider the two-dimensional radial supercooled Stefan problem inside the ball $\Omega = B_1 \subseteq \mathbb{R}^2$. We set the latent heat to $L = 2$ and assume that there is no surface tension at the interface. Given $\Gamma_{0-} = B_{r_0}$, $r_0 \in (0, R)$, and assuming the initial temperature to be radial, the solid region remains radial as well, i.e., $\Gamma_t = B_{r(t)}$ for some càdlàg function $[0-, T] \ni t \mapsto r(t)$. The goal of this experiment is to demonstrate the method's ability to generate jumps in the radius $r(\cdot)$. To this end, consider the following initial condition:

$$u_0^1(x) = -\frac{1}{|A_{r_0,\delta_0}|} \mathbb{1}_{A_{r_0,\delta_0}}(x), \quad u_0^2(x) = -\frac{1}{|B_{r_0}|} \mathbb{1}_{B_{r_0}}(x), \quad A_{r_0,\delta_0} = B_{r_0+\delta_0} \setminus B_{r_0}, \tag{35}$$

where $r_0 = \frac{1}{4}$, $\delta_0 = \frac{1}{8}$. In other words, the liquid is strongly supercooled inside the annulus A_{r_0,δ_0} and at zero temperature elsewhere. We may therefore expect a sizeable liquid region surrounding the solid to freeze immediately, leading to an initial positive jump in the radius. This is indeed the case as observed in Fig. 13. As seen in Section 3.3, we can in fact quantify the magnitude of the jump for physical solutions. In this case, we have from (21) that $\Delta r(0)$ must be the smallest positive value (if any) such that

$$|A_{r_0,\Delta}| = \frac{1}{L} \int_{A_{r_0,\Delta}} \left(-\frac{\gamma}{|x|} - u_0^1(x) \right).$$

Absent of surface tension ($\gamma = 0$) and in light of the initial data in (35), this gives

$$|A_{r_0,\Delta}| = \frac{1}{L} \int_{A_{r_0,\Delta}} |u_0^1| \iff \pi\big((r_0 + \Delta)^2 - r_0^2\big) = \frac{(r_0 + \Delta \wedge \delta_0)^2 - r_0^2}{L((r_0 + \delta_0)^2 - r_0^2)}. \tag{36}$$

Solving numerically, we obtain $\Delta r(0) \approx 0.22$. See also Fig. 12 which compares the area $|A_{r_0,\Delta}|$ with the excess temperature $\frac{1}{L}\int_{A_{r_0,\Delta}} |u_0^1|$ as functions of Δ. We finally observe in Fig. 13 that the initial jump produced by the deep level-set method nearly matches the physical jump size.

4.2. *General case*

4.2.1. *Two-phase, square-shaped solid* ($d = 2$)

Let $\Omega = B_1 \subseteq \mathbb{R}^2$ and suppose that Γ_{0-} is the two-dimensional ℓ_1-ball of radius $r_0 \in (0, 1)$. For the initial level-set function, we consider $\Phi_0(x) = |x|_1 - r_0$, $|x|_1 = \sum_{i=1}^2 |x_i|$. The temperature is initially uniform in both the liquid and the solid, namely

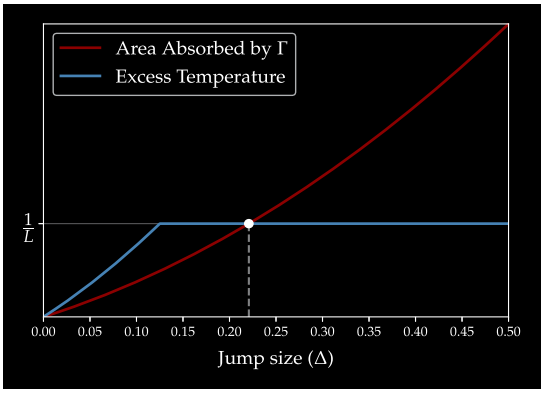


Fig. 12. Magnitude of the jump in the radius obtained by equating the excess temperature (blue curve) to the volume absorbed (red curve), corresponding to the left-hand and the right-hand sides of (36), respectively. As can be seen, the initial jump size is roughly equal to $\Delta r(0) \approx 0.22$ (precisely 0.2208).

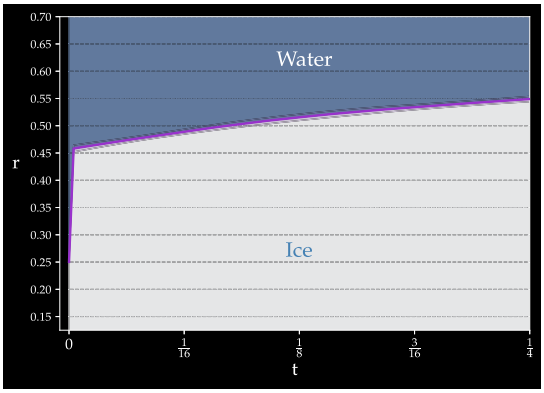


Fig. 13. Two-dimensional, two-phase supercooled Stefan problem without surface tension. The initial jump has size ≈ 0.21 (precisely 0.2098) and is close to the physical jump $\Delta r(0) \approx 0.22$.

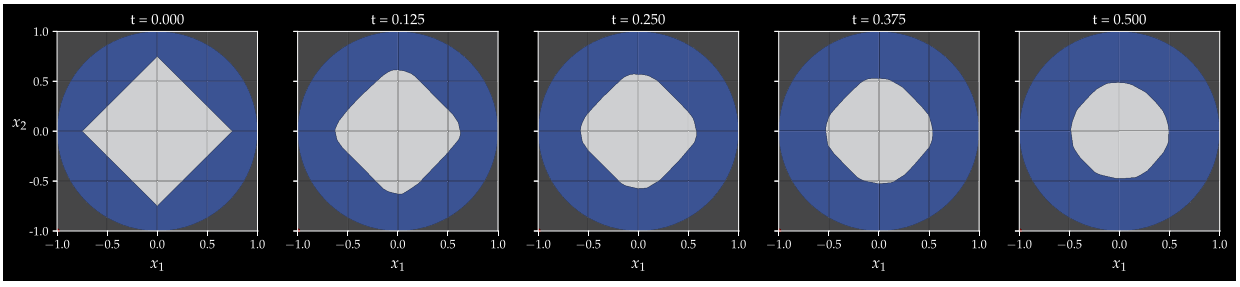


Fig. 14. Two-phase Stefan problem with square-shaped initial solid region ($d = 2$). Evolution of the solid over time (no surface tension).

$$u_0^1 = -c_1 \frac{\mathbb{1}_{\Omega \setminus \Gamma_{0-}}}{|\Omega \setminus \Gamma_{0-}|}, \quad u_0^2 = -c_2 \frac{\mathbb{1}_{\Gamma_{0-}}}{|\Gamma_{0-}|}.$$

We choose $\alpha_1 = 1/2$, $\alpha_2 = 1/20$, $L = 1/100$, and $\gamma = 0$. This example intends to investigate the situation where the liquid particles have a much larger diffusivity than the solid particles. As the liquid is not supercooled, we expect the solid region to melt faster than with similar diffusivity.

As can be seen in Fig. 14, the solid region becomes rounder as the liquid particles are more likely to hit the corners of the initial square-shaped region. Moreover, the particles in the solid typically take longer to hit the interface as their diffusivity is far less than the one of the liquid particles.

Remark 7. Recall that the neural network learns the difference between the level-set function $\Phi(t, \cdot)$, $t \in [0, T]$ and its initial value $\Phi_0(\cdot)$. However, Φ_0 is not smooth in the following examples, so the lack of regularity will carry over to $\Phi(t, \cdot)$ for $t \in (0, T]$. To give more flexibility to the deep level-set function, one could consider

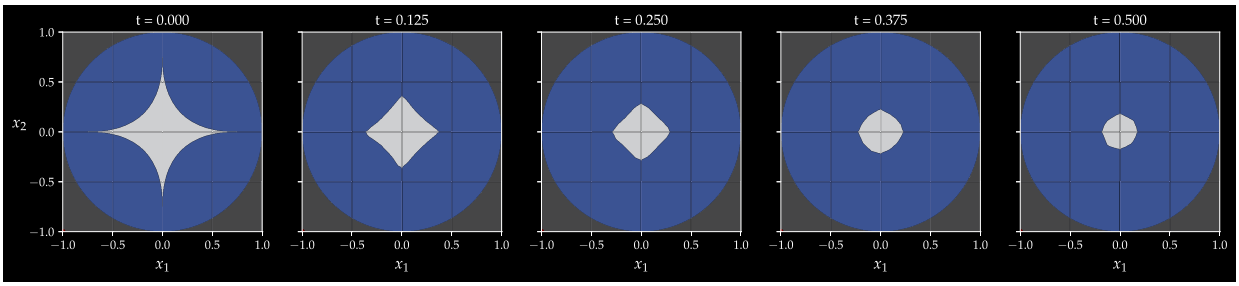


Fig. 15. Two-phase Stefan problem with surface tension ($\gamma = 0.15$). Melting of a diamond-shaped solid region ($d = 2$).

$$\Phi(t,x;\theta) := G(t,x;\theta) + \Phi_0(x)(1 - \widetilde{G}(t;\widetilde{\theta}))_+, \quad t \in [0,T],$$

where $\widetilde{G} : [0,T] \times \widetilde{\Theta} \to \mathbb{R}$ is another feedforward neural network taking only time as an input and capturing the decay of the initial level-set function. However, this additional flexibility did not prove helpful in our numerical experiments. One possible explanation is that the jump penalty prevents the auxiliary neural network from taking significant non-zero values.

4.2.2. *Two-phase, diamond-shaped solid ($d = 2$), melting regime*

We are given a diamond-shaped solid Γ_{0-} defined as the zero sublevel set of

$$\Phi_0(x) = \left(\sum_{i=1}^{2} |x_i|^{1/2} \right)^2 - r_0. \tag{37}$$

See the leftmost panel of Fig. 15 for an illustration. We choose $r_0 = 3/4$ and the temperatures

$$u_0^1 = \frac{\mathbb{1}_{\Omega \setminus \Gamma_{0-}}}{|\Omega \setminus \Gamma_{0-}|}, \qquad u_0^2 = -\frac{\mathbb{1}_{\Gamma_{0-}}}{4|\Gamma_{0-}|}.$$

We also set $\eta = 1$, $L = 1/4$, and $\gamma = 0.15$. Fig. 15 displays the melting of the solid. As can be seen, the spikes of the diamond get rounder and it eventually becomes almost radially symmetric. This is because the liquid particles hit the interface more frequently near the spikes, having more contact points, which drives the melting process there.

4.2.3. *Two-phase, diamond-shaped solid ($d = 2$), freezing regime*

Consider as in Section 4.2.2 a diamond-shaped solid Γ_{0-} with level-set function as in (37). The initial temperature is again uniform but the liquid is supercooled this time, namely

$$u_0^1 = -c_1 \frac{\mathbb{1}_{\Omega \setminus \Gamma_{0-}}}{|\Omega \setminus \Gamma_{0-}|}, \qquad u_0^2 = -c_2 \frac{\mathbb{1}_{\Gamma_{0-}}}{|\Gamma_{0-}|}.$$

We choose $r_0 = 1/2$, $c_1 = 1$, $c_2 = 1/10$. The purpose of this example is to see the impact of surface tension on the growth of the solid. Recall from Remark 6 that for two-dimensional problems, only the sign of the curvature at the interface needs to be estimated (using Algorithm 2).

Figs. 16a and 16b show the evolution of the solid region for the surface tension coefficients $\gamma = 0$ (no tension) and $\gamma = 0.15$, respectively. Observe that the interface is strictly concave almost everywhere (except at the spikes). In light of the Gibbs-Thomson condition (2e'), the temperature at the boundary is *above* zero almost everywhere when surface tension is added, which accelerates the growth of the solid region. This is confirmed in Fig. 16b.

4.2.4. *Two-phase, dumbbell-shaped solid ($d = 2$)*

We finally investigate the melting of a dumbbell-shaped solid as in the top-left panel of Fig. 17. The following parameters are used:

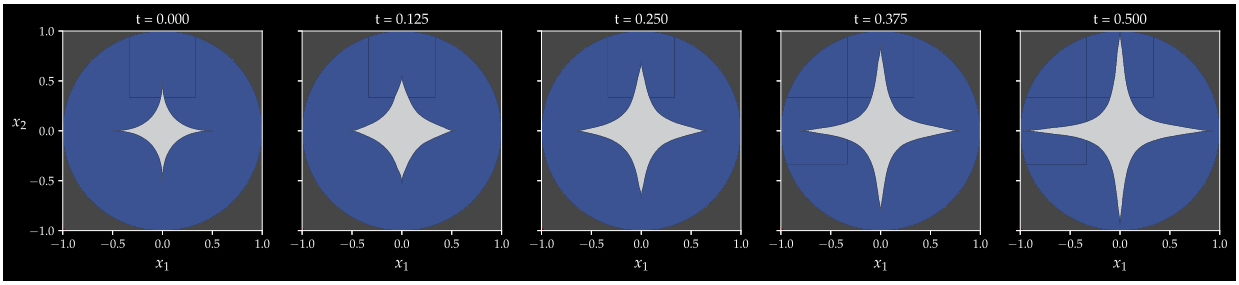
$$\gamma = 0.1, \quad c_1 = 2, \quad c_2 = 0.25, \quad \eta = 1.$$

Again, the initial temperature is constant in both the liquid and the solid. As in the previous section, we use Algorithm 2 to determine the sign of the curvature along the boundary. Fig. 17 describes the short term evolution of the solid region. As can be seen, the solid disconnects in the time interval $[0.06, 0.07]$. We also note that the middle of the "bar" melts faster than its extremities, namely the concave corners. Indeed, concave areas have a melting point *above* zero, which slows down the melting process.

5. Conclusion

We combine the level-set method with deep learning to represent the solid region in two-phase Stefan problems. The growth condition of probabilistic solutions is turned into a loss function and estimated using Monte Carlo simulation. The parameters of

(a) Without surface tension.



(b) With surface tension ($\gamma = 0.15$).

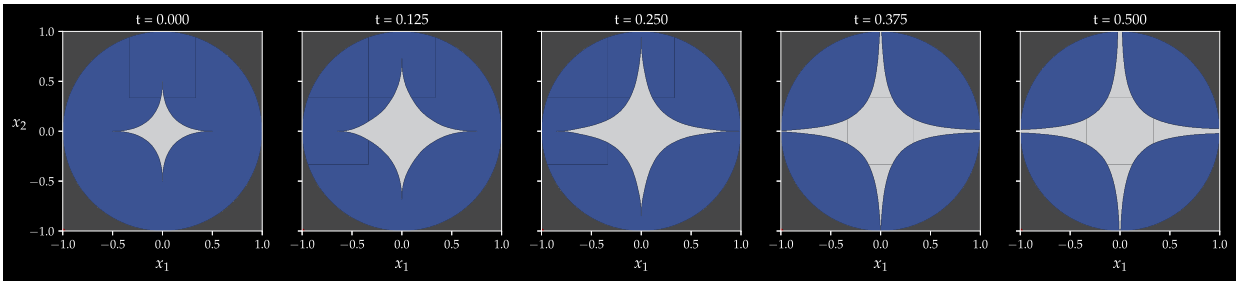


Fig. 16. Two-phase Stefan problem with diamond-shaped initial solid region ($d = 2$) and supercooled liquid. Evolution of the solid region over time.

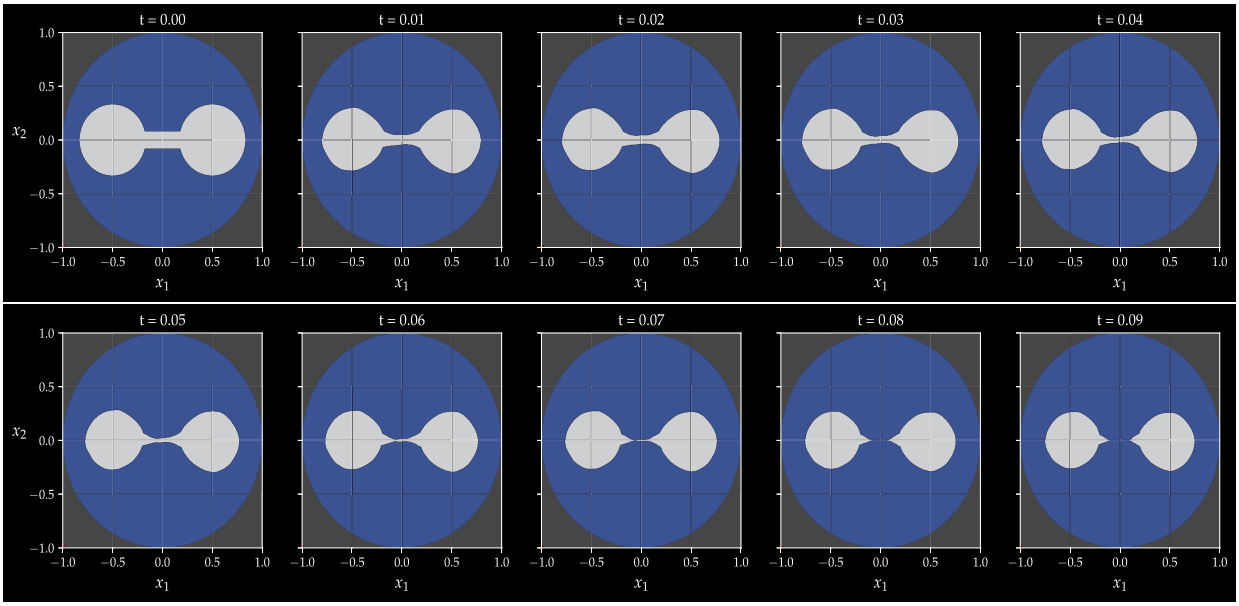


Fig. 17. Two-phase Stefan problem with surface tension and dumbbell-shaped solid region ($d = 2$). The solid disconnects between $t = 0.06$ and $t = 0.07$.

the deep level-set function are then trained using a relaxed formulation. When adding surface tension, the algorithm involves the simulation of particles close to the solid-liquid interface and the computation of the mean curvature using a dilation technique. The latter only requires the outward normal vector to the boundary, which is readily available through automatic differentiation of the deep level-set function.

The numerical examples illustrate the validity and flexibility of the method. In the two-dimensional radial case, the algorithm is capable of accurately capturing known melting rates, long-term radius, and initial jumps in the radius. Further, the three-dimensional radial example of Section 4.1.3 accurately demonstrates the effect of surface tension. We subsequently apply the method to general examples in two dimensions, where the initial solid is square-, diamond-, or dumbbell-shaped. The effect of surface tension is also discussed.

We believe that our findings make a first step towards the computation and understanding of probabilistic solutions for general Stefan problems. Naturally, other examples can be explored. For instance, it would be interesting to see if the method can handle

more complex dynamics with multiple jumps (e.g., Figure 1 in [26]) or capture dendritic solidification [2,15,29]. Another direction of future research is to investigate non-radial three-dimensional examples, such as cubic or dumbbell-shaped solid regions.

More broadly, the deep level-set method can be applied to other free boundary problems as well, e.g., the Hele-Shaw and the Stokes flow [11]. In mathematical finance, the method can be adapted to describe the stopping region of American options, thus extending the "neural optimal stopping boundary method" [35] when the geometric structure of the exercise boundary is unknown. Another application would consist in computing the no-trade zone in portfolio choice problems with transaction costs [25].

CRediT authorship contribution statement

Mykhaylo Shkolnikov: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation. **H. Mete Soner:** Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation. **Valentin Tissot-Daguette:** Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Appendix A. Proofs

A.1. Proposition 1

Proof. Let (u^1, u^2, Γ) be a classical solution of (2a)–(2e). Without loss of generality, we prove the claim for $\eta = 1$. First, the Stefan condition and the divergence theorem give

$$\begin{aligned} L \frac{\mathrm{d}}{\mathrm{d}t} \int_{\Gamma_t} \psi &= L \int_{\partial\Gamma_t} V \psi \\ &= \frac{\alpha_2}{2} \int_{\partial\Gamma_t} \partial_\nu u^2 \, \psi - \frac{\alpha_1}{2} \int_{\partial\Gamma_t} \partial_\nu u^1 \, \psi \\ &= \sum_{i=1}^{2} \frac{\alpha_i}{2} \int_{\Gamma_t^i} \operatorname{div}(\nabla u^i \, \psi) \\ &= \sum_{i=1}^{2} \frac{\alpha_i}{2} \int_{\Gamma_t^i} (\Delta u^i \, \psi + \nabla u^i \cdot \nabla \psi). \end{aligned}$$

Observing that the velocity of Γ_t^1 is $-V$, we obtain

$$\int_{\Gamma_t^i} \frac{\alpha_i}{2} \Delta u^i \psi = \int_{\Gamma_t^i} \partial_t u^i \, \psi = \frac{\mathrm{d}}{\mathrm{d}t} \int_{\Gamma_t^i} u^i \psi - (-1)^i \int_{\partial\Gamma_t^i} V u^i \psi.$$

In light of the Dirichlet boundary condition (2e), the last (surface) integral vanishes. Moreover, integration by parts gives

$$\frac{\alpha_i}{2} \int_{\Gamma_t^i} \nabla u^i \cdot \nabla \psi = (-1)^i \frac{\alpha_i}{2} \int_{\partial\Gamma_t^i} u^i \partial_\nu \psi - \frac{\alpha_i}{2} \int_{\Gamma_t^i} u^i \Delta \psi = -\frac{\alpha_i}{2} \int_{\Gamma_t^i} u^i \Delta \psi,$$

invoking again the Dirichlet boundary condition. Thus,

$$L \frac{\mathrm{d}}{\mathrm{d}t} \int_{\Gamma_t} \psi = \sum_{i=1}^{2} \left(\frac{\mathrm{d}}{\mathrm{d}t} \int_{\Gamma_t^i} u^i \psi - \frac{\alpha_i}{2} \int_{\Gamma_t^i} u^i \Delta \psi \right). \tag{38}$$

For $t \geq 0$, define $\overline{\tau}_t^i = \inf\{ s \geq 0 : X_0^i + \sqrt{\alpha_i} \, W_s^i + l_s^i \notin \Gamma_{t-s}^i \}$. Note that, by a Feynman-Kac formula,

$$u^i(t,x) = \mathbb{E}^{\mathbb{P}_x}[u_0^i(X_t^i) \, \mathbb{1}_{\{\overline{\tau}_t^i > t\}}], \quad x \in \Gamma_t^i,$$

where the subscript in $\mathbb{P}_x$ indicates that $X_0^i = x$. Using the time reversibility of Brownian motion, Itô's formula and the definition of τ^i in (3), we obtain

$$\begin{aligned}
(-1)^{i+1}\int_{\Gamma_t^i} u^i\psi &= \int_{\Gamma_t^i} \mathbb{E}^{\mathbb{P}_x}[|u_0^i(X_t^i)|\,\mathbb{1}_{\{\overleftarrow{\tau}_t^i>t\}}]\,\psi(x)\,\mathrm{d}x \\
&= \int_{\Gamma_t^i}\int_{\Gamma_0^i} |u_0^i(y)|\,\mathbb{P}_x(X_t^i\in \mathrm{d}y, \overleftarrow{\tau}_t^i>t)\,\psi(x)\,\mathrm{d}x \\
&= \int_{\Gamma_0^i}\int_{\Gamma_t^i} \psi(x)\,\mathbb{P}_y(X_t^i\in \mathrm{d}x, \tau^i>t)\,|u_0^i(y)|\,\mathrm{d}y \\
&= \mathbb{E}^{\mu^i}[\psi(X_t^i)\,\mathbb{1}_{\{\tau^i>t\}}] \\
&= \mathbb{E}^{\mu^i}[\psi(X^i_{\tau^i\wedge t})] - \mathbb{E}^{\mu^i}[\psi(X^i_{\tau^i})\,\mathbb{1}_{\{\tau^i\le t\}}] \\
&= \mathbb{E}^{\mu^i}[\psi(X_0^i)] + \frac{\alpha_i}{2}\int_0^t \underbrace{\mathbb{E}^{\mu^i}[\Delta\psi(X_s^i)\,\mathbb{1}_{\{\tau^i>s\}}]}_{=(-1)^{i+1}\int_{\Gamma_s^i} u^i\Delta\psi}\,\mathrm{d}s - \mathbb{E}^{\mu^i}[\psi(X^i_{\tau^i})\,\mathbb{1}_{\{\tau^i\le t\}}].
\end{aligned}$$

Hence $\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Gamma_t^i} u^i\psi - \frac{\alpha_i}{2}\int_{\Gamma_t^i} u^i\Delta\psi = (-1)^i\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}^{\mu^i}[\psi(X^i_{\tau^i})\,\mathbb{1}_{\{\tau^i\le t\}}]$ and the result follows. □

A.2. Proposition 2

Proof. It is enough to prove the claim for $\eta = 1$. Looking at the proof of Proposition 1 and using that u^1, u^2 coincide on $\partial\Gamma$, we gather that

$$\begin{aligned}
L\,\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Gamma_t}\psi &= \sum_{i=1}^2\left(\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Gamma_t^i} u^i\psi - (-1)^i\int_{\partial\Gamma_t} Vu^i\psi + \frac{\alpha_i}{2}\int_{\Gamma_t^i}\nabla u^i\cdot\nabla\psi\right) \\
&= \sum_{i=1}^2\left(\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Gamma_t^i} u^i\psi - \frac{\alpha_i}{2}\int_{\Gamma_t^i} u^i\Delta\psi\right) - \dot{\mathcal{K}}_t^I,
\end{aligned}$$

where $\dot{\mathcal{K}}_t^I = \gamma\frac{\alpha_2-\alpha_1}{2}\int_{\partial\Gamma_t}\kappa_{\partial\Gamma_t}\,\partial_\nu\psi$. Moreover, u^i admits the Feynman-Kac representation

$$u^i(t,x) = \mathbb{E}^{\mathbb{P}_x}[u_0^i(X_t^i)\,\mathbb{1}_{\{\overleftarrow{\tau}_t^i>t\}}] - \gamma K^i(t,x), \quad x\in\Gamma_t^i,$$

with K^i and $\overleftarrow{\tau}_t^i$ given in the statement. This implies that

$$L\,\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Gamma_t}\psi = \sum_{i=1}^2(-1)^i\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}^{\mu^i}[\psi(X^i_{\tau^i})\,\mathbb{1}_{\{\tau^i\le t\}}] - \dot{\mathcal{K}}_t^I - \dot{\mathcal{K}}_t^{II} + \dot{\mathcal{K}}_t^{III},$$

with the additional terms

$$\dot{\mathcal{K}}_t^{II} = \gamma\sum_{i=1}^2\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Gamma_t^i} K^i\psi, \quad \dot{\mathcal{K}}_t^{III} = \gamma\sum_{i=1}^2\frac{\alpha_i}{2}\int_{\Gamma_t^i} K^i\Delta\psi. \tag{39}$$

Next, due to (10), K^i is a weak solution of the heat equation in any rectangle $\mathcal{R}\subseteq\operatorname{int}\Gamma^i$, i.e.,

$$\int_{\mathcal{R}}\left(\partial_t + \frac{\alpha_i}{2}\Delta\right)\psi\;K^i = 0, \quad \psi\in C_c^\infty(\mathcal{R}).$$

From Weyl's Lemma, we conclude that K^i is in fact a strong solution of $\partial_t K^i = \frac{\alpha_i}{2}\Delta K^i$ in $\operatorname{int}\Gamma^i$. Noting also that $K^1 = K^2$ on $\partial\Gamma$, we have

$$\frac{1}{\gamma}\dot{\mathcal{K}}_t^{II} = \sum_{i=1}^2\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Gamma_t^i} K^i\psi = \sum_{i=1}^2\left(\int_{\Gamma_t^i}\partial_t K^i\,\psi + (-1)^i\int_{\partial\Gamma_t} VK^i\psi\right) = \sum_{i=1}^2\int_{\Gamma_t^i}\frac{\alpha_i}{2}\Delta K^i\,\psi.$$

Using Green's second identity, note that (again, the outward normal of Γ_t^1 is $-\nu$)

$$\int_{\Gamma_t^i} \Delta K^i \, \psi = (-1)^i \int_{\partial\Gamma_t} (\partial_\nu K^i \, \psi - K^i \, \partial_\nu \psi) + \int_{\Gamma_t^i} K^i \Delta \psi .$$

Given the definition of $\dot{\mathcal{K}}_t^{III}$ in (39) and recalling that $K^1|_{\partial\Gamma} = K^2|_{\partial\Gamma} = \kappa_{\partial\Gamma}$, this implies that

$$\begin{aligned}
\dot{\mathcal{K}}_t^{II} - \dot{\mathcal{K}}_t^{III} &= \gamma \sum_{i=1}^{2} (-1)^i \frac{\alpha_i}{2} \int_{\partial\Gamma_t} (\partial_\nu K^i \, \psi - \kappa_{\partial\Gamma_t} \, \partial_\nu \psi) \\
&= -\dot{\mathcal{K}}_t^{I} + \gamma \int_{\partial\Gamma_t} \left(\frac{\alpha_2}{2} \partial_\nu K^2 - \frac{\alpha_1}{2} \partial_\nu K^1 \right) \psi \\
&= -\dot{\mathcal{K}}_t^{I} + \frac{\mathrm{d}}{\mathrm{d}t} \mathcal{K}_t .
\end{aligned}$$

Integrating in time and rearranging yields the claim. □

A.3. Proposition 3

Proof. Assume again that $\eta = 1$. Using similar arguments to the ones in the proof of Proposition 1, we can show that

$$L\left(\int_{\Gamma_{0-}} \psi - \int_{\Gamma_t} \psi \right) = \sum_{i=1}^{2} \left(\int_{\Gamma_{0-}^i} u^i \psi - \int_{\Gamma_t^i} u^i \psi + \frac{\alpha_i}{2} \int_0^t \int_{\Gamma_s^i} u^i \Delta \psi \right). \tag{40}$$

Expressing $u := -u^1 \mathbb{1}_{\Gamma_t^1} - u^2 \mathbb{1}_{\Gamma_t^2}$ via [26, Lemma 2.4], we find for $\varphi \in \{\psi, \Delta\psi\}$,

$$(-1)^{i+1} \int_{\Gamma_t^i} u^i \varphi = \mathbb{E}[\varphi(X_t^i) \mathbb{1}_{\{\tau^i > t\}}] + (-1)^i \lim_{\delta \downarrow 0} \mathbb{E}\left[\sum_{l=1}^{\infty} \varphi(Y_t^{l,i}) \mathbb{1}_{\{T_l^{\delta,i} \le t < \tau_l^{\delta,i}\}} \right]. \tag{41}$$

Next, let us temporarily write $(Y, T_l, \tau_l, \alpha) = (Y^{l,i}, T_l^{\delta,i}, \tau_l^{\delta,i}, \alpha_i)$ to simplify notation. Using Itô's lemma and Fubini's theorem, observe that

$$\begin{aligned}
\psi(Y_t) \mathbb{1}_{\{T_l \le t < \tau_l\}} &= \psi(Y_{\tau_l \wedge t}) \mathbb{1}_{\{T_l \le t\}} - \psi(Y_{\tau_l}) \mathbb{1}_{\{\tau_l \le t\}} \\
&= \psi(Y_{T_l}) \mathbb{1}_{\{T_l \le t\}} + \frac{\alpha}{2} \int_0^t \Delta\psi(Y_s) \mathbb{1}_{\{T_l \le s < \tau_l\}} \, \mathrm{d}s \\
&\quad + \int_{T_l \wedge t}^{\tau_l \wedge t} \nabla\psi(Y_s) \cdot \mathrm{d}Y_s - \psi(Y_{\tau_l}) \mathbb{1}_{\{\tau_l \le t\}} \\
\implies \mathbb{E}[\psi(Y_t) \mathbb{1}_{\{T_l \le t < \tau_l\}}] &= \mathbb{E}[\psi(Y_{T_l}) \mathbb{1}_{\{T_l \le t\}} - \psi(Y_{\tau_l}) \mathbb{1}_{\{\tau_l \le t\}}] + \frac{\alpha}{2} \int_0^t \mathbb{E}[\Delta\psi(Y_s) \mathbb{1}_{\{T_l \le s < \tau_l\}}] \, \mathrm{d}s .
\end{aligned}$$

Similarly, for X_t^i and τ^i, we have

$$\mathbb{E}[\psi(X_t^i) \mathbb{1}_{\{\tau^i > t\}}] = \mathbb{E}[\psi(X_0^i) - \psi(X_{\tau^i}^i) \mathbb{1}_{\{\tau^i \le t\}}] + \frac{\alpha_i}{2} \int_0^t \mathbb{E}[\Delta\psi(X_s^i) \mathbb{1}_{\{\tau^i > s\}}] \, \mathrm{d}s .$$

Plugging the above expressions into (41), with $\varphi = \psi$, and invoking Fubini's theorem, we obtain

$$\begin{aligned}
(-1)^{i+1} \left(\int_{\Gamma_{0-}^i} u^i \psi - \int_{\Gamma_t^i} u^i \psi \right) &= -\frac{\alpha_i}{2} \int_0^t \mathbb{E}[\Delta\psi(X_s^i) \mathbb{1}_{\{\tau^i > s\}}] \, \mathrm{d}s + \mathbb{E}[\psi(X_{\tau^i}^i) \mathbb{1}_{\{\tau^i \le t\}}] \\
&\quad - (-1)^i \lim_{\delta \downarrow 0} \mathbb{E}\left[\sum_{l=1}^{\infty} \left(\psi(Y_{T_l^{\delta,i}}^{l,i}) \mathbb{1}_{\{T_l^{\delta,i} \le t\}} - \psi(Y_{\tau_l^{\delta,i}}^{l,i}) \mathbb{1}_{\{\tau_l^{\delta,i} \le t\}} \right) \right] \\
&\quad - (-1)^i \lim_{\delta \downarrow 0} \frac{\alpha_i}{2} \int_0^t \mathbb{E}\left[\sum_{l=1}^{\infty} \Delta\psi(Y_s^{l,i}) \mathbb{1}_{\{T_l^{\delta,i} \le s < \tau_l^{\delta,i}\}} \right] \mathrm{d}s
\end{aligned}$$

$$= (-1)^i \frac{\alpha_i}{2} \int_0^t \int_{\Gamma_s^i} u^i \Delta\psi \ \mathrm{d}s + \mathbb{E}[\psi(X^i_{\tau^i}) \mathbb{1}_{\{\tau^i \le t\}}] + (-1)^i \mathcal{K}^i_t,$$

using (41) with $\varphi = \Delta\psi$ for the last equality. In view of (40) and the definition of $\mathcal{K}$, $(N^{\delta,i})_{i=1}^2$ in the statement, we indeed find that

$$\int_{\Gamma_{0-}} \psi - \int_{\Gamma_t} \psi = \frac{1}{L} \left(\mathbb{E}^{\mu^1}[\psi(X^1_{\tau^1}) \mathbb{1}_{\{\tau^1 \le t\}}] - \mathbb{E}^{\mu^2}[\psi(X^2_{\tau^2}) \mathbb{1}_{\{\tau^2 \le t\}}] + \mathcal{K}_t \right). \quad \square$$

References

[1] A. Alessandri, P. Bagnerini, M. Gaggero, Optimal control of propagating fronts by using level set methods and neural approximations, IEEE Trans. Neural Netw. Learn. Syst. 30 (3) (2019) 902–912.
[2] R. Almgren, Variational algorithms and pattern formation in dendritic solidification, J. Comput. Phys. 106 (2) (1993) 337–354.
[3] L. Ambrosio, H.M. Soner, Level set approach to mean curvature flow in arbitrary codimension, J. Differ. Geom. 43 (4) (1996) 693–737.
[4] A. Bachouch, C. Huré, N. Langrené, H. Pham, Deep neural networks algorithms for stochastic control problems on finite horizon: numerical applications, Methodol. Comput. Appl. Probab. 24 (1) (2022) 143–178.
[5] G. Barles, H.M. Soner, P.E. Souganidis, Front propagation and phase field theory, SIAM J. Control Optim. 31 (2) (1993) 439–469.
[6] S. Becker, P. Cheridito, A. Jentzen, Deep optimal stopping, J. Mach. Learn. Res. 74 (2019) 1–25.
[7] S. Becker, P. Cheridito, A. Jentzen, T. Welti, Solving high-dimensional optimal stopping problems using deep learning, Eur. J. Appl. Math. 32 (3) (2021) 470–514.
[8] W.J. Boettinger, J.A. Warren, C. Beckermann, A. Karma, Phase-field simulation of solidification, Annu. Rev. Mater. Sci. 32 (1) (2002) 163–194.
[9] S. Chen, B. Merriman, S. Osher, P. Smereka, A simple level set method for solving Stefan problems, J. Comput. Phys. 135 (1) (1997) 8–29.
[10] C. Cuchiero, S. Rigger, S. Svaluto-Ferro, Propagation of minimality in the supercooled Stefan problem, Ann. Appl. Probab. 33 (2) (2023) 1588–1618.
[11] L.J. Cummings, S.D. Howison, J.R. King, Two-dimensional Stokes and Hele-Shaw flows with free surfaces, Eur. J. Appl. Math. 10 (6) (1999) 635–680.
[12] F. Delarue, S. Nadtochiy, M. Shkolnikov, Global solutions to the supercooled Stefan problem with blow-ups: regularity and uniqueness, Probab. Math. Phys. 3 (1) (2022) 171–213.
[13] A. Figalli, Regularity of interfaces in phase transitions via obstacle problems, in: Proceedings of the International Congress of Mathematicians, 2018.
[14] M. Freidlin, Functional Integration and Partial Differential Equations (AM-109), Princeton University Press, 1985.
[15] F. Gibou, R. Fedkiw, R. Caflisch, S. Osher, A level set approach for the numerical simulation of dendritic growth, J. Sci. Comput. 19 (2002).
[16] Y. Guo, S. Nadtochiy, M. Shkolnikov, Stefan problem with surface tension: uniqueness of physical solutions under radial symmetry, arXiv:2306.02969, 2023.
[17] M. Hadzic, P. Raphael, On melting and freezing for the 2D radial Stefan problem, J. Eur. Math. Soc. 21 (2015).
[18] J. Han, W. E, Deep learning approximation for stochastic control problems, Adv. Neural Inf. Process. Syst. (2016).
[19] J. Han, A. Jentzen, W. E, Solving high-dimensional partial differential equations using deep learning, Proc. Natl. Acad. Sci. 115 (34) (2018) 8505–8510.
[20] L.A. Herraiz, M.A. Herrero, J.J.L. Velázquez, A note on the dissolution of spherical crystals, Proc. R. Soc. Edinb., Sect. A, Math. 131 (2) (2001) 371–389.
[21] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, in: Proceedings of the International Conference on Learning Representations (ICLR), 2015.
[22] G. Lamé, B.P. Clapeyron, Mémoire sur la solidification par réfroidissement d'un globe liquide, Ann. Chim. Phys. 47 (1831) 250–256.
[23] S. Luckhaus, Solutions for the two-phase Stefan problem with the Gibbs–Thomson law for the melting temperature, Eur. J. Appl. Math. 1 (2) (1990) 101–111.
[24] L.M. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, A. Geiger, Occupancy networks: learning 3d reconstruction in function space, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 4455–4465.
[25] J. Muhle-Karbe, M. Reppen, H.M. Soner, A primer on portfolio choice with small transaction costs, Annu. Rev. Financ. Econ. 9 (1) (2017) 301–331.
[26] S. Nadtochiy, M. Shkolnikov, Stefan problem with surface tension: global existence of physical solutions under radial symmetry, arXiv:2203.15113, 2022.
[27] S. Nadtochiy, M. Shkolnikov, X. Zhang, Scaling limits of external multi-particle DLA on the plane and the supercooled Stefan problem, arXiv:2102.09040, 2021.
[28] S. Osher, J.A. Sethian, Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations, J. Comput. Phys. 79 (1) (1988) 12–49.
[29] S.J. Osher, R. Fedkiw, Level Set Methods and Dynamic Implicit Surfaces, Applied Mathematical Sciences, vol. 153, Springer, 2003.
[30] J.J. Park, P.R. Florence, J. Straub, R.A. Newcombe, S. Lovegrove Deepsdf, Learning continuous signed distance functions for shape representation, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 165–174.
[31] H. Pham, X. Warin, M. Germain, Neural networks-based backward scheme for fully nonlinear PDEs, SN Part. Differ. Equ. Appl. 2 (2021).
[32] S. Popinet, Numerical models of surface tension, Annu. Rev. Fluid Mech. 50 (1) (2018) 49–75.
[33] M. Raissi, P. Perdikaris, G. Karniadakis, Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. 378 (2019) 686–707.
[34] A.M. Reppen, H.M. Soner, V. Tissot-Daguette, Deep stochastic optimization in finance, Digit. Finance (2022).
[35] A.M. Reppen, H.M. Soner, V. Tissot-Daguette, Neural optimal stopping boundary, arXiv:2205.04595, 2022.
[36] J.A. Sethian, Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science, Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, 1999.
[37] J. Sirignano, K. Spiliopoulos Dgm, A deep learning algorithm for solving partial differential equations, J. Comput. Phys. 375 (2018) 1339–1364.
[38] H.M. Soner, Motion of a set by the curvature of its boundary, J. Differ. Equ. 101 (2) (1993) 313–372.
[39] H.M. Soner, Convergence of the phase-field equations to the Mullins-Sekerka problem with kinetic undercooling, Arch. Ration. Mech. Anal. 131 (2) (1995) 139–197.
[40] H.M. Soner, V. Tissot-Daguette, Stopping times of boundaries: relaxation and continuity, arXiv:2305.09766, 2023.
[41] J. Stefan, Über einige Probleme der Theorie der Wärmeleitung, Sitzungsber. Akad. Wiss. Wien, Math.-Naturwiss. Kl. 98 (1889) 473–484.
[42] J. Stefan, Über die Theorie der Eisbildung, Monatshefte Math. Phys. 1 (1) (1890) 1–6.
[43] J. Stefan, Über die Verdampfung und die Auflösung als Vorgänge der Diffusion, Ann. Phys. 277 (1890) 725–747.
[44] J. Stefan, Über die Theorie der Eisbildung, insbesondere über die Eisbildung im Polarmeere, Ann. Phys. Chem. 42 (1891) 269–286.
[45] A. Visintin, Stefan problem with surface tension, in: Mathematical Models for Phase Change Problems, Springer, 1989, pp. 191–213.
[46] S. Wang, P. Perdikaris, Deep learning of free boundary and Stefan problems, J. Comput. Phys. 428 (2021) 109914.
[47] S. Wang, Y. Teng, P. Perdikaris, Understanding and mitigating gradient flow pathologies in physics-informed neural networks, SIAM J. Sci. Comput. 43 (5) (2021) A3055–A3081.