Numerische Mathematik (2022) 152:631–662
https://doi.org/10.1007/s00211-022-01320-0

Numerische Mathematik

A proximal-gradient algorithm for crystal surface evolution

Katy Craig[1] · Jian-Guo Liu[2] · Jianfeng Lu[3] · Jeremy L. Marzuola[4] · Li Wang[5]

Received: 2 September 2020 / Revised: 21 October 2021 / Accepted: 8 September 2022 /
Published online: 29 September 2022

Abstract

As a counterpoint to recent numerical methods for crystal surface evolution, which agree well with microscopic dynamics but suffer from significant stiffness that prevents simulation on fine spatial grids, we develop a new numerical method based on the macroscopic partial differential equation, leveraging its formal structure as the gradient flow of the total variation energy, with respect to a weighted H^{-1} norm. This gradient flow structure relates to several metric space gradient flows of recent interest, including 2-Wasserstein flows and their generalizations to nonlinear mobilities. We develop a novel semi-implicit time discretization of the gradient flow, inspired by the classical minimizing movements scheme (known as the JKO scheme in the 2-Wasserstein case). We then use a primal dual hybrid gradient (PDHG) method to compute each element of the semi-implicit scheme. In one dimension, we prove convergence of the PDHG method to the semi-implicit scheme, under general integrability assumptions on the mobility and its reciprocal. Finally, by taking finite difference approximations of our PDHG method, we arrive at a fully discrete numerical algorithm, with iterations that converge at a rate independent of the spatial discretization: in particular, the convergence properties do not deteriorate as we refine our spatial grid. We close with several numerical examples illustrating the properties of our method, including facet formation at local maxima, pinning at local minima, and convergence as the spatial and temporal discretizations are refined.

KC was supported by NSF DMS grant 1811012 and a Hellman Faculty Fellowship. KC also gratefully acknowledges the support from the Simons Center for Theory of Computing, at which part of this work was completed. JGL was supported in part by NSF DMS-2106988. JL was supported in part by NSF under award DMS-1454939. The research of the The research of JLM was supported by NSF Grant DMS-1312874 and NSF CAREER Grant DMS-1352353. LW was supported in part by NSF DMS-1903425 and DMS-1846854. This collaboration is made possible thanks to the NSF Grant RNMS-1107444 (KI-Net).

✉ Jeremy L. Marzuola
marzuola@math.unc.edu

Extended author information available on the last page of the article

Mathematics Subject Classification 35A15 · 47J25 · 47J35 · 49J45 · 49M29 · 65K10 · 82B21 · 82B05

1 Introduction

The evolution of a crystal surface near a fixed crystallographic plane of symmetry is determined by the desire to minimize the surface free energy [19, 39]. In terms of the height $h(x,t)$ of the surface, $x \in \Omega \subseteq \mathbb{R}^d$, $d \geq 1$, $t \geq 0$, the free energy is given by the well-known total variation energy,

$$\mathcal{E}(h) = \int_\Omega |\nabla h(x,t)|\, \mathrm{d}x. \tag{1.1}$$

Facets on the crystal surface are identified with the regions $\{x : \nabla h(x,t) = 0\}$.

To formally obtain a PDE describing the surface dynamics, we briefly recall some tools from hydrodynamic flows in statistical mechanics. Setting the atomic volume equal to one, the step chemical potential is given by first variation of the energy [48],

$$\mu_s = \frac{\delta E}{\delta h} = -\Delta_1 h, \quad \text{with} \quad \Delta_1 h := \nabla \cdot \left(\frac{\nabla h}{|\nabla h|} \right).$$

By the Gibbs–Thomson relation [28, 29, 37, 43] (which is related to an ideal gas law approximation), the corresponding local-equilibrium density of adatoms is formally $\varrho_s = \varrho^0 \exp[\mu_s/(k_B T)]$, where ϱ^0 is a constant reference density [21, 50], T is a temperature, and k_B is the Boltzmann constant. An application of Fick's law then predicts that the flux is

$$\mathbf{J} = -D_s \nabla \varrho_s = -D_s \varrho^0 \nabla e^{\mu_s/(k_B T)},$$

where D_s is the surface diffusion constant [37]. In this way, we obtain the hydrodynamic equation

$$\partial_t h + \nabla \cdot \mathbf{J} = 0.$$

Normalizing all constants to be one by rescaling in space and time, we formally arrive at the following PDE for the evolution of the crystal surface height:

$$\partial_t h = \Delta e^{-\Delta_1 h}. \tag{1.2}$$

A detailed, yet non-rigorous, derivation of (1.2) from microscopic dynamics can be found in [33]. As far as the authors are aware, the exact form of PDE (1.2) only appeared in [33] for the first time.

That being said, PDEs of the type (1.2) have arisen in various settings in statistical mechanics. Away from facets, this equation is consistent with the continuum limit of the Burton–Cabrera–Frank (BCF) theory for moving steps in 2+1 dimensions [4, 37]. See also [3] for a numerical study of $1d$ facet dynamics. This equation also relates to a family of Kinetic Monte Carlo models of crystal surface relaxation, including both

the solid-on-solid (SOS) and discrete Gaussian models, in which the 1-Laplacian is replaced by a p-Laplacian, $p > 1$ [13, 29, 38]. Progress towards a rigorous understanding of the convergence of related models in statistical mechanics to their limiting PDEs has been made recently in the work of Katsevich [24, 25].

Note that, even in one dimension and for $h(x,t)$ is smooth, the 1-Laplacian $\Delta_1 h$ is a linear combination of positive and negative Dirac masses, so $e^{-\Delta_1 h}$ is not well-defined. Consequently, Eq. (1.2) must be interpreted in a generalized sense. One avenue considered in previous work is to take a first order approximation of the exponential in the Gibbs–Thomson relation, replacing e^x with $1+x$, which leads to the H^{-1} total variation flow studied by Giga et al., [14–16, 26, 41]

$$\partial_t h = \Delta(-\Delta_1 h) \ . \tag{1.3}$$

A limitation of this approach is that it treats local maxima and minima of h symmetrically, in contrast to the original Eq. (1.2), which causes local maxima to form expanding facets, while local minima remain stationary. Ultimately, determining an appropriate notion of weak solution for Eq. (1.2) and proving existence of solutions remains a challenging open problem.

In spite of these gaps in the underlying theory of the crystal surface evolution equation, we seek to develop a computationally efficient numerical method for accurate simulation of its solutions, while respecting the inherent asymmetry between facet formation at local maxima and pinning at local minima. Recent work by the middle three authors and Margetis [33] and the fourth author and Weare [38], numerically explored the crystal surface evolution equation, using various regularizations. On one hand, these simulations compared well with the ensemble averages of the Kinetic Monte Carlo models described in those works and respected the different dynamics occurring near the local maxima and minima of the crystal surface. The work [33] studied precisely the Eq. (1.2), however [38] only studied the dynamics of equations of the form $\partial_t h = \Delta e^{-\Delta_p h}$ for $p > 1$. On the other hand, these studies were not motivated by a strong notion of convergence to the macrosopic PDE dynamics, and due to the inherent stiffness of the model, were only effective on coarse spatial grids, with serious numerical convergence issues arising on fine grids, even in one dimension.

In contrast, we construct our numerical method for crystal surface evolution by starting with the macroscopic PDE (1.2) and leveraging the formal gradient flow structure of the equation, with respect to weighted H^{-1} norms. To see this structure, note that Eq. (1.2) may be rewritten in the following conservative form,

$$\partial_t h + \nabla \cdot \left(M(h) \nabla \frac{\partial \mathcal{E}}{\partial h} \right) = 0, \tag{1.4}$$

where $\mathcal{E}$ is the total variation energy (1.1) and $M(h)$ is the exponential mobility

$$M(h) := e^{-\Delta_1 h}. \tag{1.5}$$

For simplicity in what follows, we suppose that our underlying domain is the d-dimensional torus $\mathbb{T}^d$ and Eq. (1.4) is posed with periodic boundary conditions. We

normalize the initial data $h(x, 0) = h_0(x)$ to have mean zero, $\int h_0 = 0$, a property that is then propagated along the flow (1.4).

Equations of this form (1.4) have a formal gradient flow structure with respect to an H^{-1} norm weighted by the mobility $M(h)$, which we describe in detail in Sect. 2. For example, choosing the constant mobility $M(h) \equiv 1$, one recovers classical H^{-1} gradient flows, and in the case of the linear mobility $M(h) = h + 1$, one recovers 2-Wasserstein gradient flows on the space of probability measures [2, 42]. (Since h has mean zero, $h + 1$ is a probability density as long as $h \geq -1$.). There has also been significant work on equations of this form in the context of reaction diffusion equations [31] and Cahn–Hilliard equations [32], among many others. See also the recent work by Gao [11] who considered the related setting of an exponential mobility determined by a logarithmic correction to the 1 Laplacian.

Again, the problem of exponentiating $-\Delta_1 h$ arises in the definition of the mobility (1.5). In order to circumvent this difficulty and thereby ensure that the weighted H^{-1} gradient flow structure is well-defined, we introduce the following novel approximation: given $\varphi \in C_c^\infty(\mathbb{T}^d)$, $\varphi \geq 0$, $\int_{\mathbb{T}^d} \varphi = 1$, $\varphi_\epsilon(x) := \varphi(x/\epsilon)/\epsilon^d$, we consider

$$M_\epsilon(h) := e^{-\varphi_\epsilon * \Delta_1 h}. \tag{1.6}$$

Unlike previous approximations of $e^{-\Delta_1 h}$ via $1 - \Delta_1 h$, our approximation respects the inherent asymmetry near local maxima and minima of h, becoming large when $-\Delta_1 h \gg 0$ and vanishing when $-\Delta_1 h \ll 0$.

With this approximation in hand, we are able to precisely define the weighted H^{-1} gradient flow of the total variation energy $\mathcal{E}$ with mobility M_ϵ. Then, with the goal of computing this flow numerically, we discretize the gradient flow in time, with a fixed time step $\tau > 0$, via the following semi-implicit method:

$$h^{n+1} \in \arg\min_h \mathcal{E}(h) + \frac{1}{2\tau} \|h - h^n\|^2_{H^{-1}_{h^n}}. \tag{1.7}$$

This approach is inspired by the classical minimizing movements scheme for gradient flows, known as the JKO scheme in the 2-Wasserstein context [2, 23]. In this way, our numerical method can be seen as an extension of recent literature using minimizing movement schemes to simulate nonlinear PDEs as gradient flows on metric spaces; see [5–7, 30] and the references therein. In the context of crystal surface evolution, our work builds upon the well-known literature using implicit Euler time discretizations to simulate Hilbertian gradient flows, including the H^{-1} total variation flow described in Eq. (1.3) [27]. From this perspective, our work is closely related to recent work by Giga and Ueda [17]. As we describe below, we will compute minimizers of Eq. (1.7) via a primal dual algorithm (PDHG), paralleling Giga and Ueda's use of a primal dual algorithm (split Bregman) to compute minimizers of the implicit scheme for their H^{-1} total variation flow . The key difference between our approach and this previous work is the fact that we consider a *weighted* H^{-1} gradient flow, depending on the exponential mobility, instead of a classical H^{-1} gradient flow, in which the exponential mobility is linearized. This allows our approach to capture the asymmetry of the dynamics at local minima and maxima.

We show that the Euler–Lagrange equation characterizing solutions of the semi-implicit scheme is a discrete time version of the conservative PDE (1.4):

$$\frac{h^{n+1} - h^n}{\tau} = -\nabla \cdot \left(M_\epsilon(h^n) \nabla \frac{\partial \mathcal{E}}{\partial h^{n+1}} \right). \tag{1.8}$$

(See Eqs. 2.6 and 2.9 below.) Consequently, interpolating in time,

$$h^\tau(x,t) = h^n, \quad \text{if } t \in [n\tau, (n+1)\tau),$$

and sending our regularization ϵ and time step τ to zero, one formally expects that $h^\tau(x,t)$ approaches a solution of the crystal surface evolution Eq. (1.2). We leave analysis of this convergence to future work, since it directly relates to the challenging open problem of proving existence of solutions to the crystal surface equation. Still, we believe that the success of our numerical method, which is based on this semi-implicit scheme, provides empirical evidence that the gradient flow framework is the appropriate setting for studying generalized solutions to this equation.

In order to translate the semi-implicit scheme (1.7) into a fully discrete numerical method, we use a primal dual hybrid gradient (PDHG) [9] approach, which we describe in detail in Sect. 3. This approach allows us to handle the presence of the 1-Laplacian, as well as preserve the energy decreasing property at the discrete level. Given the nth step of the semi-implicit scheme h^n, our PDHG method iteratively defines a new sequence $h^{(m)}$ that is initialized at h^n and converges to h^{n+1}. A key point in the definition of our PDHG method is that we use different norms to penalize the primal and dual variables. As discovered by Jacobs, Léger, Li, and Osher [22], appropriate selection of the norms is essential to obtaining a scheme that is convergent at the spatially continuous level and leads to a fully discrete numerical method with rate of convergence that does not deteriorate as the spatial discretization is refined; see Remark 3.2. Our PDHG method is well-defined for the total variation energy $\mathcal{E}$ (1.1) and *any* integrable mobility $M(h)$, including the regularized exponential mobility (1.6). Provided that the mobility and its reciprocal remain integrable along the sequence h^n, which holds for the regularized exponential mobility, our main convergence result Theorem 3.4 proves that the inner PDHG interates $h^{(m)}$ converge to a solution of the outer scheme h^{n+1}. We prove this result in one spatial dimension, which coincides with the context of our numerical simulations. Furthermore, if our initialization h^n has the regularity $(h^n)' \in BV(\mathbb{T})$, our theorem provides a rate of convergence for the PDHG method. We remark that existing convergence results for PDHG algorithms do not apply in our context, since our initialization of the primal variable h^n is, in general, infinite $\dot{H}^1$ distance from the optimizer h^{n+1} [9, 47]. We instead build upon the approach introduced by Jacobs, Léger, Li, and Osher for the Rudin-Osher-Fatemi image denoising model [22, 44].

While our convergence proof does not extend naturally to higher dimensions, due to the fact that we use the one dimensional Sobolev embedding $BV(\mathbb{T}) \subseteq L^\infty(\mathbb{T})$, our numerical method does extend to higher dimensions. In particular, the definition of our method at the spatially continuous level in Sect. 3.1 is independent of the dimension, and our Algorithm 4.1 can be extended to higher dimensions simply by replacing

the matrix approximation of the operators ∇ and $-\Delta_h$ by their higher dimensional counterparts.

Finally, in Sect. 4, we use this time discretization of the weighted H^{-1} gradient flow as the basis for a fully discrete numerical scheme, replacing the spatially continuous operators in our PDHG method with their finite difference counterparts. In Remark 4.1, we describe how the convergence of this fully discrete scheme, with a rate independent of the spatial discretization, follows from similar arguments as given in Theorem 3.4. Achieving convergence rates that are uniform in the spatial discretization is essential in the context of our problem, since it allows us to choose our inner primal time step in the PDHG method to be relatively large, even as we refine the spatial discretization. Note that this would not be permitted by classical PDHG methods, in which the inner time steps are required to shrink quickly as the spatial discretization is refined. The importance of choosing large inner primal time steps arises due to the fact that convexity properties of $\|\cdot\|^2_{H^{-1}_{h^n}}$ depend on lower bounds on the eigenvalues of the weighted Laplacian $\Delta^{-1}_{h^n}$ (see Eq. 2.3), which may deteriorate along the flow. Choosing large inner primal time steps allow us to overcome this deterioration of convexity, in agreement with our estimates for the optimal choice of time steps in Theorem 3.4.

We conclude, in Sect. 5, with several numerical examples that illustrate properties of our method. Our scheme accurately captures facet formation at local maxima and pinning at local minima. Unlike previous numerical methods, which required a coarse spatial discretization, we observe near first order convergence in both space and time as the spatial discretization and time step τ are refined. Finally, we also illustrate the importance of norm selection in our PDHG method, showing that selecting norms following the classical L^2 approach can cause the number of iterations required for convergence to increase dramatically as the spatial discretization is refined.

There are several directions for future work. As mentioned above, we believe that the strength of our numerical method gives hope that the weighted gradient flow setting is the appropriate context in which to define and prove existence of generalized solutions to the crystal surface evolution equation, by analyzing the convergence of the semi-implicit method as $\tau \to 0$ and $\epsilon \to 0$. Our semi-implicit time discretization and PDHG algorithm can also be naturally extended to related crystal evolution PDEs: see Remark 3.3, where we describe how the 1-Laplacian in Eq. (1.2) can be replaced by the standard Laplacian. Finally, our convergence result for the PDHG scheme holds for general, integrable mobilities $M(h)$. Consequently, it would be natural to extend our approach to simulate related gradient flows for other choices of nonlinear mobilities, such as $M(h) = (1+h)(1-h)$ [8, 10, 32].

2 Crystal height evolution as a weighted H^{-1} gradient flow

We now describe the weighted H^{-1} gradient flow structure of the crystal height evolution PDE (1.2). In Sect. 2.1, we define the weighted H^{-1} spaces and the corresponding notions of gradient flow. In Sect. 2.2, we introduce the semi-implicit time discretization of the gradient flow, which is the basis of our numerical scheme. In Sect. 2.3, we discuss how to apply this framework to the crystal height evolution equation.

2.1 Weighted H^{-1} gradient flow

For any $h : \mathbb{T}^d \to \mathbb{R}$, let $M(h) \in L^1(\mathbb{T}^d)$ denote a nonnegative mobility. Using this mobility, we define the weighted Hilbert space $H_h^1(\mathbb{T}^d)$ as the completion of $C^\infty(\mathbb{T}^d)$ functions with mean zero, under the weighted norm or inner product

$$\|v\|_{H_h^1}^2 = \int_{\mathbb{T}^d} M(h)|\nabla v|^2 \,\mathrm{d}x \,, \tag{2.1}$$

$$(u, v)_1 = \int_{\mathbb{T}^d} M(h)\nabla u \cdot \nabla v \,\mathrm{d}x \,. \tag{2.2}$$

We define $H_h^{-1}(\mathbb{T}^d) := \left(H_h^1(\mathbb{T}^d)\right)^*$ to be the dual space of H_h^1 and let $\langle \cdot, \cdot \rangle$ denote the duality pairing.

By the Riesz-Fréchet representation theorem, the duality mapping $J : H_h^1 \to H_h^{-1}$ given by

$$\langle J(v), u\rangle = (v, u)_1, \quad \forall u \in H_h^1$$

is surjective. Now, consider the weighted Laplacian operator

$$\Delta_h u = \nabla \cdot (M(h)\nabla u) \,,$$

which is well defined for $u \in C^\infty(\mathbb{T}^d)$, in the sense of distributions. For $u, v \in C^\infty(\mathbb{T}^d)$ with mean zero, by definition of $(\cdot, \cdot)_1$ and integration by parts, we have

$$\langle J(v), u\rangle = (v, u)_1 = -\int_{\mathbb{T}^d} u\Delta_h v \,\mathrm{d}x \,.$$

Hence, we identify $J(v) = -\Delta_h v$.

The inverse map $J^{-1} : H_h^{-1} \to H_h^1, \phi \mapsto J^{-1}(\phi)$ is then given by

$$\langle \psi, J^{-1}(\phi)\rangle = (\psi, \phi)_{-1} = -\int_{\mathbb{T}^d} \psi(\Delta_h^{-1}\phi) \,\mathrm{d}x \,, \quad \forall \psi \in H_h^{-1} \,,$$

where $(\cdot, \cdot)_{-1}$ denotes the inner product for H_h^{-1} and Δ_h^{-1} denotes the inverse operator of Δ_h with mean zero. Consequently, we obtain,

$$\|\psi\|_{H_h^{-1}}^2 = -\int_{\mathbb{T}^d} \psi\Delta_h^{-1}\psi \,\mathrm{d}x \,. \tag{2.3}$$

We now turn to the differential structure induced by the H_h^{-1} norm. Given a convex functional $\mathcal{E} : H_h^{-1} \to \mathbb{R} \cup \{+\infty\}$, its subdifferential is

$$\partial_{H_h^{-1}}\mathcal{E}(\psi) = \left\{\xi \in H_h^{-1}(\mathbb{T}^d) : \mathcal{E}(\varphi) \geq \mathcal{E}(\psi) + (\varphi - \psi, \xi)_{-1} \quad \forall \varphi \in H_h^{-1}\right\} \,.$$

For example, the identity mapping $\psi \mapsto \{\psi\}$ is the subdifferential of the convex functional $\mathcal{E}(\psi) = \frac{1}{2}\|\psi\|_{H_h^{-1}}^2$.

Using this notion of subdifferential, we may define H_h^{-1} gradient flows. In order for our construction of the weighted Hilbert spaces to remain valid, we require that $M(h)$ remains nonnegative and integrable along the flow, that is, the flow remains in the space

$$L^1_M = \left\{ h : \mathbb{T}^d \to \mathbb{R} : M(h) \in L^1 \text{ and } M(h) \geq 0 \right\}.$$

Next, we introduce a notion of time derivative for a flow $h(t)$ evolving through the Hilbert spaces $H^{-1}_{h(t)}$.

Definition 2.1 Given $h : [0, T] \to L^1_M$ such that $h(t) \in H^{-1}_{h(t)}$ for all $t \in [0, T]$, we say that $h(t)$ *is differentiable with respect to* $\| \cdot \|_{H^{-1}_{h(t)}}$ in case, for all $t \in [0, T]$, there exists $\epsilon > 0$ so that, for all $s \in (t-\epsilon, t+\epsilon) \cap [0, T]$, $h(s) \in H^{-1}_{h(t)}$ and $h(s)$ is Fréchet differentiable with respect to $\| \cdot \|_{H^{-1}_{h(t)}}$.

With this, we can now define an $H^{-1}_{h(t)}$ gradient flow.

Definition 2.2 Given $h : [0, T] \to L^1_M$ such that $h(t) \in H^{-1}_{h(t)}$ is differentiable, we say h is an $H^{-1}_{h(t)}$ *gradient flow* of an energy $\mathcal{E} : H^{-1}_{h(t)} \to \mathbb{R} \cup \{+\infty\}$ with initial condition h_0 in case

$$\begin{cases} \partial_t h(t) \in -\partial_{H^{-1}_{h(t)}} \mathcal{E}(h(t)) \text{ for all } t \in [0, T], \\ h(0) = h_0. \end{cases} \tag{2.4}$$

In particular, given an energy $\mathcal{E} : H^{-1}_h \to \mathbb{R} \cup \{+\infty\}$, we formally obtain the following expression for its gradient with respect to $H^{-1}_h(\mathbb{T}^d)$,

$$\lim_{\epsilon \to 0} \frac{\mathcal{E}(\psi + \varepsilon\xi) - \mathcal{E}(\psi)}{\varepsilon} = \int_{\mathbb{T}^d} \frac{\partial E}{\partial \psi} \xi = \int_{\mathbb{T}^d} \Delta_h^{-1} \Delta_h \frac{\partial E}{\partial \psi} \xi = \left(\Delta_h \frac{\partial E}{\partial \psi}, \xi \right)_{-1}$$

Therefore,

$$\nabla_{H^{-1}_h} \mathcal{E}(\psi) = \Delta_h \frac{\partial E}{\partial \psi}. \tag{2.5}$$

Consequently, under sufficient regularity of the energy functional $\mathcal{E}$ and under the assumption that the mobility remains integrable and nonnegative along the flow, $H^{-1}_h(\mathbb{T}^d)$ gradient flows correspond to solutions of the conservative PDE (1.4),

$$\partial_t h = -\nabla_{H^{-1}_h} \mathcal{E}(h) \iff \partial_t h + \Delta_h \frac{\partial E}{\partial h} = 0 \iff \partial_t h + \nabla \cdot \left(M(h) \nabla \frac{\partial E}{\partial h} \right) = 0. \tag{2.6}$$

2.2 Semi-implicit scheme for H_h^{-1} gradient flows

We now describe a semi-implicit analogue of the classical minimizing movement scheme to discretize our H_h^{-1} gradient flows in time: given $h^n \in L^1_M \cap H^{-1}_{h^n}$, solve

$$h^{n+1} \in \operatorname*{arg\,min}_{h \in H^{-1}_{h^n}} \mathcal{E}(h) + \frac{1}{2\tau}\|h - h^n\|^2_{H^{-1}_{h^n}}. \tag{2.7}$$

In the particular case $M(h) = h + 1$, $h \geq -1$, H_h^{-1} gradient flows are 2-Wasserstein gradient flows, and the above method can be interpreted as a semi-implicit variant of the Jordan Kinderlehrer Otto (JKO) scheme [23], in which the Wasserstein distance is approximated by the corresponding weighted H^{-1} norm at the previous time step [5].

We begin by showing that, as long as the energy $\mathcal{E}$ is convex, lower semicontinuous, and has compact sublevels with respect to an appropriate topology and $\mathcal{E}(h^n) < +\infty$, then there exists a unique solution to this semi-implicit scheme.

Proposition 2.3 *Fix $h^n \in L^1_M \cap H^{-1}_{h^n}$ and consider an energy $\mathcal{E} : H^{-1}_{h^n} \to \mathbb{R} \cup \{+\infty\}$. Suppose $\mathcal{E}$ is convex and that there exists a topology σ so that $\mathcal{E}$ and $H^{-1}_{h^n}$ are both lower semicontinuous with respect to σ and the sublevel sets of $\mathcal{E}$ are relatively σ-compact in $H^{-1}_{h^n}$. Then, if $\mathcal{E}(h^n) < +\infty$, there exists a unique h^{n+1} so that*

$$h^{n+1} \in \operatorname*{arg\,min}_{h \in H^{-1}_{h^n}} \Phi(h), \quad \text{for} \quad \Phi(h) := \mathcal{E}(h) + \frac{1}{2\tau}\|h - h^n\|^2_{H^{-1}_{h^n}}. \tag{2.8}$$

Remark 2.4 Our assumption that $h^n \in L^1_M$, or equivalently, that the mobility $M(h^n)$ is integrable and nonnegative, is necessary for the weighted Hilbert spaces H^1_h to be well-defined and, hence, for their dual spaces H^{-1}_h to be well-defined. In particular, we must have $h^n \in L^1_M$ in order to define the discrete scheme in Eq. (2.8), and in particular, this is necessary for a solution h^{n+1} of the scheme to exist. Analogous requirements on the mobility have arisen in recent work by Cancés, Gallouët, and Todeschi [5], in which they consider a fully-implicit time discretization, in the special case that $M(h) = h + 1$ and $h > -1$.

Remark 2.5 We choose to introduce the additional topology σ in Proposition 2.3 due to the fact that, in general, the topology induced by $H^{-1}_{h_n}$ may not be strong enough to ensure lower semicontinuity of the energy. In particular, this is the case for the exponential mobility and total variation energy we consider in the next section.

Proof of Proposition 2.3 First, we consider existence. Since $\Phi(h^n) = \mathcal{E}(h^n) < +\infty$,

$$\inf_{h \in H^{-1}_{h^n}(\mathbb{T})} \Phi(h) < +\infty,$$

and we may choose a minimizing sequence $h^k \in H^{-1}_{h^n}(\mathbb{T})$ so that $\lim_{k\to+\infty} \Phi(h^k) = \inf_h \Phi(h)$. Since $\Phi(h) \geq \mathcal{E}(h)$, $\{h^k\}$ belongs to a sublevel set of $\mathcal{E}$, so up to a sub-

sequence, there exists $\bar{h}$ so that $h^k \xrightarrow{\sigma} \bar{h} \in H_h^{-1}$. By lower semicontinuity of $\mathcal{E}$ and $\|\cdot\|_{H_{h^n}^{-1}(\mathbb{T})}$ with respect to σ, $\liminf_{k\to+\infty} \Phi(h^k) = \Phi(\bar{h})$. Thus, $\bar{h}$ is a solution of (2.7), so a solution exists.

It remains to show uniqueness. Suppose h^{n+1} and $\bar{h}$ are distinct solutions of (2.7). Define $h_\alpha = (1-\alpha)h^{n+1} + \alpha\bar{h}$. Then, by the convexity of $\mathcal{E}$ and the strict convexity of $h \mapsto \|h - h^n\|^2_{H_{h^n}^{-1}}$,

$$\Phi(h_\alpha) < (1-\alpha)\Phi(h^{n+1}) + \alpha\Phi(\bar{h}) = \inf_{h\in H_h^{-1}(\mathbb{T}^d)} \Phi(h),$$

which is a contradiction. Therefore $h^{n+1} = \bar{h}$, so solutions of (2.7) are unique. □

Given a discrete sequence $\{h^n\}$ defined by our semi-implicit scheme (2.7), the convexity of Φ and the fact that h^{n+1} is a global minimum imply that we have the following Euler–Lagrange equation characterizing h^{n+1},

$$0 \in \partial_{H_{h^n}^{-1}} \Phi(h^{n+1}) \iff \frac{h^{n+1} - h^n}{\tau} \in -\partial_{H_{h^n}^{-1}} \mathcal{E}(h^{n+1}). \tag{2.9}$$

Consequently, interpolating in time, $h^\tau(x,t) = h^n$, if $t \in [n\tau, (n+1)\tau)$, one formally expects that, under sufficient regularity of $\mathcal{E}$ and h^n, as $\tau \to 0$, $h^\tau(x,t)$ approaches a solution of the $H_{h^n}^{-1}$ gradient flow, in the sense of Definition 2.2.

A key difficulty in the analysis of this limit is proving that the discrete time solutions h^n remain in the space L^1_M, so that the weighted dual Sobolev spaces $H_{h^n}^{-1}$ remain well-defined. For example, in the present work, for our regularization of the mobility (1.6), a sufficient condition is that $\Delta_1 h^n \in \mathcal{M}(\mathbb{T}^d)$, the space of finite Borel measures. Developing sufficient conditions on the initial data h^0 so that $h^n \in L^1_M$ for all $n \geq 1$ can be thought of as a type of a priori estimate and would be an important first step toward solving the challenging open problem of proving existence of solutions to the PDE (1.2). While the requirement $h^n \in L^1_M$ is independent of the choice of energy $\mathcal{E}$, the energy $\mathcal{E}$ does provide an important tool for proving this type of estimates, since it can be shown to decrease along solutions of the discrete time scheme (see Lemma 3.7 (1)), providing a weak notion of regularity, $E(h^n) < +\infty$.

We leave the rigorous study of the convergence of the discrete time scheme to the continuum PDE to future work. Our hope is that the framework developed in the present paper will provide the first steps toward the rigorous study of this limit and, ultimately, a proof of existence for solutions to the crystal height evolution equation.

2.3 H_h^{-1} gradient flow for crystal surface evolution

We now describe how our crystal surface evolution equation fits into this gradient flow framework. As discussed in the introduction, the crystal surface evolution PDE may be formally rewritten in conservative form (1.4) for an exponential mobility (1.5) and the total variation energy. However, in order for this formal description to coincide with a well-defined H_h^{-1} gradient flow, we must extend the energy to a functional defined on

all of H_h^{-1}, satisfying the hypotheses of Proposition 2.3, and the mobility must remain nonnegative and integrable along the flow. We now consider each of these issues.

2.3.1 Total variation energy on H_h^{-1}

Since H_h^1 is defined as the completion of C^∞ functions with mean zero under the H_h^1 norm, any element $\psi \in H_h^{-1}$ is uniquely defined by its action on such smooth functions. In particular, if there exists $f \in L^1$ with mean zero so that $\langle \psi, v \rangle = \int_{\mathbb{T}^d} f v \, \mathrm{d}x$ for all $v \in C^\infty$ with mean zero, we will identify ψ with f and say $\psi \in L^1$. (We restrict to f with mean zero, since such an f is only determined up to a constant.)

In this way, we extend the definition of the total variation energy to H_h^{-1},

$$\mathcal{E}(\psi) := \begin{cases} \|\psi\|_{TV} & \text{if } \psi \in L^1(\mathbb{T}^d) \text{ and } \int \psi = 0, \\ +\infty & \text{otherwise,} \end{cases} \tag{2.10}$$

where, for any $\psi \in L^1(\mathbb{T}^d)$,

$$\|\psi\|_{TV} := \sup_{\phi} \left\{ -\int_{\mathbb{T}^d} \psi \nabla \cdot \phi : \phi \in C^\infty(\mathbb{T}^d), \ \|\phi\|_\infty \le 1 \right\}. \tag{2.11}$$

Furthermore, if $\|\psi\|_{TV} < +\infty$, then the distributional derivative $\nabla\psi$ is a signed measure and

$$\|\psi\|_{TV} = \sup_{\phi} \left\{ \int_{\mathbb{T}^d} \nabla\psi \cdot \phi : \phi \in L^\infty(\mathbb{T}^d), \ \|\phi\|_\infty \le 1 \right\}. \tag{2.12}$$

We now show that $\mathcal{E}$ satisfies the hypotheses of Proposition 2.3, so that the semi-implicit scheme is well defined.

Proposition 2.6 *Consider $h \in L^1_M$, so that $M(h)$ is nonnegative and integrable, and consider the total variation energy $\mathcal{E} : H_h^{-1} \to \mathbb{R} \cup \{+\infty\}$ defined in Eq. (2.10). Then $\mathcal{E}$ is convex, and, letting σ denote the topology of convergence in distribution, $\mathcal{E}$ and $\|\cdot\|_{H_h^{-1}}$ are both lower semicontinuous with respect to σ and the sublevel sets of $\mathcal{E}$ are relatively σ-compact in H_h^{-1}.*

Proof The convexity of $\mathcal{E}$ follows immediately from the fact that $L^1 \cap H_h^{-1}$ is convex and $\|\cdot\|_{TV}$ is a convex functional on L^1.

Next, we show that $\mathcal{E}$ and $\|\cdot\|_{H_h^{-1}}$ are lower semicontinuous with respect to convergence in distribution. We begin with $\|\cdot\|_{H_h^{-1}}$. Suppose $\psi^k \in H_h^{-1}$ converges to

$\psi \in H_h^{-1}$ in distribution. Then,

$$\begin{aligned}\liminf_{k\to\infty} \|\psi^k\|_{H_h^{-1}} &= \liminf_{k\to\infty} \sup_{\phi\in C^\infty, \|\phi\|_{H_h^1}\le 1} \langle \psi^k, \phi\rangle \ge \sup_{\phi\in C^\infty, \|\phi\|_{H_h^1}\le 1} \liminf_{k\to+\infty} \langle \psi^k, \phi\rangle \\ &= \sup_{\phi\in C^\infty, \|\phi\|_{H_h^1}\le 1} \langle \psi, \phi\rangle = \|\psi\|_{H_h^{-1}}\end{aligned}$$

We now show lower semicontinuity of $\mathcal{E}$ with respect to a sequence $\psi^k \in H_h^{-1}$ converging to $\psi \in H_h^{-1}$ in distribution. Without loss of generality, we may assume that $\liminf_{k\to\infty} \mathcal{E}(\psi^k) < +\infty$, or the result is trivially true. Consider a subsequence, ψ^{k^l} that attains the limit, i.e. $\liminf_{k\to\infty} \mathcal{E}(\psi^k) = \lim_{l\to_\infty} \mathcal{E}(\psi^{k^l})$ and for which $\mathcal{E}(\psi^{k^l}) < +\infty$. For simplicity of notation, we identify this subsequence with the original sequence ψ^k. Since $\mathcal{E}(\psi^k) < +\infty$ for all k, along this sequence, the energy coincides with $\|\cdot\|_{TV}$. Thus,

$$\begin{aligned}\lim_{k\to\infty} \mathcal{E}(\psi^k) &= \lim_{k\to\infty} \|\psi_k\|_{TV} = \lim_{k\to+\infty} \sup_{\phi\in C^\infty, \|\phi\|_\infty\le 1} -\int \psi^k \nabla\cdot\phi \\ &\ge \sup_{\phi\in C^\infty, \|\phi\|_\infty\le 1} \lim_{k\to+\infty} -\int \psi^k \nabla\cdot\phi = \sup_{\phi\in C^\infty, \|\phi\|_\infty\le 1} \\ &\quad -\int \psi\nabla\cdot\phi = \|\psi\|_{TV} = \mathcal{E}(\psi).\end{aligned}$$

We now show relative compactness of the sublevel sets of $\mathcal{E}$. Suppose $\sup_k \mathcal{E}(\psi^k) \le C$ for some $C \in \mathbb{R}$. By classical results, there exists $\psi \in L^1(\mathbb{T}^d)$ such that $\mathcal{E}(\psi) = \|\psi\|_{TV} \le C$ and $\psi^k \to \psi$ in $L^1(\mathbb{T}^d)$ [51, Corollary 5.3.4]. Since convergence in $L^1(\Omega)$ implies convergence in distribution and $\|\cdot\|_{H_h^{-1}}$ is lower semicontinuous in distribution, we conclude $\psi \in H_h^{-1}$, which gives the result. □

2.3.2 Regularization of mobility

While we require that the mobility remain nonnegative and integrable along the flow, in order for the spaces H_h^{-1} to remain well defined, this fails for the exponential mobility (1.5), even for smooth functions in one dimension. For example, for any $h \in C^\infty(\mathbb{T})$, the function $h'/|h'| : \mathbb{T} \to \{-1, 0, 1\}$ is piecewise constant, and $-\Delta_1 h = -(h'/|h'|)'$ is signed measure, consisting of a linear combination of positive and negative Dirac masses, corresponding to local maxima and minima of h. Thus, $e^{-\Delta_1 h}$ is not well-defined.

Consequently, we instead approximate the mobility by convolving $-\Delta_1 h$ with a mollifier. Given $\varphi \in C_c^\infty(\mathbb{T}^d)$, $\varphi \ge 0$, $\int_{\mathbb{T}^d} \varphi = 1$, define the mollifier $\varphi_\epsilon(x) = \varphi(x/\epsilon)/\epsilon$. We then consider the mobility

$$M_\epsilon(h) := e^{-\varphi_\epsilon * \Delta_1 h}, \tag{2.13}$$

which well defined for $h \in C^1(\mathbb{T}^d)$, since $\varphi_\epsilon * \Delta_1 h = \nabla\varphi_\epsilon * (\nabla h/|\nabla h|) \in C^\infty(\mathbb{T}^d)$ for all $\epsilon > 0$. In one dimension, this regularization replaces each Dirac mass in $\Delta_1 h$ with an appropriately weighted mollifier φ_ϵ. Since $\varphi_\epsilon * \Delta_1 h \to \Delta_1 h$ in the narrow topology as $\epsilon \to 0$, for any $v \in C^\infty(\mathbb{T}^d)$, we have

$$\lim_{\epsilon\to 0} \int_{\mathbb{T}^d} M_\epsilon(h)|\nabla v|^2 \,\mathrm{d}x \;= \int_{\mathbb{T}^d} M(h)|\nabla v|^2 \,\mathrm{d}x \;,$$

so that the weighted Hilbert norms defines in Eq. (2.1) converge along smooth functions. For this reason, there is hope that the gradient flow structure induced by the dual norms also converge as $\epsilon \to 0$, in the sense of Sandier and Serfaty [45, 46], but we leave the rigorous analysis of this limit to future work.

3 A PDHG method for computing the semi-implicit scheme

In the previous section, we defined the following semi-implicit scheme for approximating H_h^{-1} gradient flows,

$$h^{n+1} \in \operatorname*{arg\,min}_{h\in H^{-1}_{h^n}} \Phi(h), \quad \text{for} \quad \Phi(h) := \mathcal{E}(h) + \frac{1}{2\tau}\|h - h^n\|^2_{H^{-1}_{h^n}}. \tag{3.1}$$

In order to use this scheme as a numerical method for simulating solutions of the crystal growth equation, we need an approach to compute the minimizer h^{n+1} of Φ.

In this section, we reformulate the above minimization problem as a saddle-point problem, so that solutions can be computed via operator splitting methods. In particular, given an element of the discrete time sequence h^n we apply a primal dual hybrid gradient (PDHG) method [9] to compute the next element in the sequence h^{n+1}. The PDHG method is essentially composed of alternating implicit Euler steps in the primal and dual variables, subject to appropriate averaging; see Remark 3.1. An important aspect of our method is that the implicit Euler step in the primal variables is taken with respect to an $\dot{H}^1$ norm, while the implicit Euler step in the dual variables is with respect to an L^2 norm. Appropriate selection of the norms is essential to proving convergence of the scheme; see Remark 3.2. This also leads to a fully discrete numerical method that converges with a rate independent of the spatial discretization; see Remark 4.1.

We begin, in Sect. 3.1, by defining our PDHG scheme. In Sect. 3.2, we state our main theorem: in one dimension, provided that the reciprocal of the mobility remains integrable, the PDHG scheme converges in the ergodic sense to the solution h^{n+1}. We prove this result in Sect. 3.3. Our results apply to the total variation energy (2.10) and any nonnegative, integrable mobility $M(h)$.

3.1 Definition of PDHG scheme

To place our problem in the framework of the PDHG method, note that, by definition of the total variation energy (2.10–2.12), minimizing Φ is equivalent to solving the

following saddle point problem

$$\inf_{h\in L^1,\int h=0} \Phi(h) = \inf_{h\in L^1,\int h=0} \sup_{\phi\in L^\infty} \mathcal{L}(h,\phi), \tag{3.2}$$

$$\mathcal{L}(h,\phi) := \int \nabla h\cdot\phi + \frac{1}{2\tau}\|h-h^n\|^2_{H^{-1}_{h^n}} - F^*(\phi), \tag{3.3}$$

$$F^*(\phi) := \begin{cases} 0 & \text{if } \|\phi\|_\infty \le 1, \\ +\infty & \text{otherwise.} \end{cases} \tag{3.4}$$

To numerically compute a minimizer of this problem, we apply PDHG, initializing the inner iterations, denoted by $h^{(m)}$, with the value of the semi-implicit sequence at the previous step $h^{(0)} := h^n$ and initializing the dual variables to be zero, $\phi^{(0)} = 0$. The PDHG algorithm [9, equation 11] is then given as follows:

$$h^{(m+1)} = \operatorname*{arg\,min}_{h\in L^1,\int h=0} \frac{1}{2\tau}\|h-h^{(0)}\|^2_{H^{-1}_{h^{(0)}}} + \int \nabla h\cdot\phi^{(m)} + \frac{1}{2\lambda}\|h-h^{(m)}\|^2_{\dot H^1} \tag{3.5}$$

$$\bar h^{(m+1)} = 2h^{(m+1)} - h^{(m)} \tag{3.6}$$

$$\phi^{(m+1)} = \operatorname*{arg\,max}_{\phi\in L^\infty} -F^*(\phi) + \int \nabla \bar h^{(m+1)}\cdot\phi - \frac{1}{2\sigma}\|\phi-\phi^{(m)}\|_2^2, \tag{3.7}$$

where $\lambda, \sigma > 0$ are given parameters. We note that the second step is an extrapolation, while the other two steps are optimization sub-problems in h and ϕ, respectively.

The PDHG iterations are easier to compute than our original minimization problem (3.1), since their optimizers are characterized by the Euler–Lagrange equations:

$$h^{(m+1)} = \left(-\Delta - \frac{\lambda}{\tau}\Delta^{-1}_{h^n}(\cdot - h^n)\right)^{-1}\left(-\Delta h^{(m)} + \lambda\nabla\cdot\phi^{(m)}\right) \tag{3.8}$$

$$\bar h^{(m+1)} = 2h^{(m+1)} - h^{(m)} \tag{3.9}$$

$$\phi^{(m+1)} = (\mathrm{id} + \sigma\,\partial F^*)^{-1}(\phi^{(m)} + \sigma\nabla\bar h^{(m+1)})\,, \tag{3.10}$$

where

$$(\mathrm{id} + \sigma\,\partial F^*)^{-1}(u(x)) = \min(|u(x)|, 1)\,\mathrm{sgn}(u(x)).$$

These have several benefits over the Euler–Lagrange equation for the semi-implicit scheme (1.8), which in the case of the total variation energy is given by

$$h^{n+1} = (\Delta^{-1}_{h^n} + \tau\Delta_1)^{-1}\left(\Delta^{-1}_{h^n}h^n\right).$$

First, our method allows us to avoid inverting the 1-Laplacian, which would require further regularizations. Second, our approach preserves the decrease of the TV energy at the discrete time level: see Remark 3.1 and Fig. 4 below. Third, as predicted in

our main convergence theorem, Theorem 3.4, we are able to choose λ large to ease inversion of Δ_h: see Fig. 6 below.

Remark 3.1 (*Interpretation as proximal point algorithm*) In the special case that $\lambda = \sigma$, the PDHG method can be characterized as a proximal point algorithm on the product space $\dot{H}^1(\mathbb{T}) \times L^2(\mathbb{T})^d$, endowed with the norm $\|\cdot\|_{\mathsf{L}} := \|\mathsf{L}^{1/2}\cdot\|_2$ for

$$\mathsf{L} = \begin{bmatrix} -\Delta & \lambda\nabla\cdot \\ -\sigma\nabla & \mathrm{id} \end{bmatrix}.$$

For further details in a slightly simpler case see, for example, He and Yuan [20].

Remark 3.2 (*Choice of norms*) It is essential to the convergence of the PDHG algorithm that we use a $\dot{H}^1$ norm penalization in our definition of $h^{(m+1)}$, instead of an L^2 penalization, as in our definition of $\phi^{(m+1)}$. As observed by Jacobs, Léger, Li, and Osher [22], this choice of norms ensures that the gradient operator $\nabla : \dot{H}^1 \to (L^2)^d$ is bounded, so Chambolle and Pock's estimate of the partial primal dual gap applies: see Eqs. (3.27) and (3.28) in the proof of our main theorem.

Remark 3.3 (*Extension to the standard Laplacian*) It is possible to extend the above algorithm to the case of crystal evolution equations with alternative surface energy interactions. In particular, when Δ_1 is replaced by $\Delta = \Delta_2$ (see e.g. [1, 12, 13, 18, 34–36]), one would replace $F^*(\phi)$ with $F^*(\phi) = \chi_{\|\phi\|_2 \le 1}$. In this case, $(\mathrm{I}+\sigma\,\partial F^*)^{-1}(u) = u/\|u\|_2$. On the other hand, for general Δ_p, $p \neq 1, 2$, there is no explicit formula for this operator (the proximal map).

3.2 Convergence of PDHG to semi-implicit scheme

We now prove that, in one dimension, if the reciprocal of the mobility is integrable, we have

$$\lim_{M\to+\infty} \Phi(h^{(M)}) = \inf_{h\in L^1(\mathbb{T}^d),\, \int h = 0} \Phi(h) = \Phi(h^{n+1}),$$

where $(h^{(M)}, \phi^{(M)})$ are the ergodic sequences, defined by

$$\left(h^{(M)}, \phi^{(M)}\right) = \left(\frac{1}{M}\sum_{m=1}^{M} h^{(m)}, \frac{1}{M}\sum_{m=1}^{M} \phi^{(m)}\right). \tag{3.11}$$

Furthermore, if the initial condition for our PDHG scheme $h^{(0)} := h^n$ is sufficiently regular, we obtain quantitative estimates on the rate of convergence.

Our main result is the following:

Theorem 3.4 *Suppose the PDHG algorithm is initialized with*

1. $h^{(0)} := h^n \in L^1_M(\mathbb{T}) \cap H^{-1}_{h^n}(\mathbb{T})$ *with* $\mathcal{E}(h^n) < +\infty$ *and* $1/M(h^n) \in L^1(\mathbb{T})$*;*
2. $\phi^{(0)} \in L^\infty(\mathbb{T})$ *with* $\|\phi^{(0)}\|_\infty \le 1$.

Then, for all $\epsilon > 0$, there exist M_*, λ_*, σ_* so that an ϵ-approximate solution may be obtained using the step sizes λ_* and σ_* in at most M_* iterations of our scheme, i.e.

$$\Phi(h^{(M)}) - \Phi(h^{n+1}) \leq \epsilon \,, \quad \forall M \geq M_*,$$

where $h^{(M)}$ is the ergodic sequence and h^{n+1} is the unique minimizer of Φ. The constants M_*, λ_*, σ_* depend on ϵ, $\|h^n\|_{TV}$, $\|M(h^n)\|_1$, $\|1/M(h^n)\|_1$, and the rate at which the function $\delta \mapsto \|h^n * \varphi_\delta - h^n\|_{TV}$ converges to zero, where $\varphi_\delta(x) = \varphi(x/\delta)/\delta$ is a compactly supported mollifier.

If, in addition, the initialization $h^{(0)} := h^n$ satisfies $\nabla h^n \in BV(\mathbb{T})$, then

$$\|h^n * \varphi_\delta - h^n\|_{TV} \leq \delta \|\nabla h^n\|_{TV} M_1(\varphi) \,,$$

so there exists a computable constant c depending on $\|h^n\|_{TV}$, $\|\nabla h^n\|_{TV}$, $\|M(h^n)\|_1$, $\|1/M(h^n)\|_1$ and φ, so that for

$$M_* := 2\pi \frac{16c}{\epsilon^2}, \quad \lambda_* = \frac{c}{\epsilon}, \quad \sigma_* = \frac{\epsilon}{c},$$

we have that $(h^{(M)}, \phi^{(M)})$ is an ϵ-approximate solution for all $M \geq M_*$.

Remark 3.5 The assumption $h^n \in L^1_M \cap H^{-1}_{h^n}, \mathcal{E}(h^n) < +\infty$ ensures sufficient regularity so that the subsequent step of the scheme h^{n+1} is well-defined; see Propositions 2.3 and 2.6.

Remark 3.6 Our assumption that the reciprocal of the mobility is integrable is similar to analogous assumptions in recent work on weighted Hilbert space discretizations for 2-Wasserstein gradient flows. In particular, Cancés, Gallouët, and Todeschi [5] consider a fully implicit scheme for $M(h) = h + 1 > 0$ on a compact domain, which ensures $1/M(h) \in L^\infty$, hence the reciprocal of the mobility is integrable.

In the particular case of the regularized exponential mobility (2.13), the constraint that $h^n \in L^1_M$ and $1/M(h^n) \in L^1$ is equivalent to requiring $M(h^n)$ and $1/M(h^n)$ be integrable. In fact, they are both in $L^\infty(\mathbb{T})$ for all $\epsilon > 0$, due to the estimate

$$|\nabla \varphi_\epsilon * \operatorname{sgn}(h^n(x))| \leq \frac{1}{\epsilon} \|\nabla \varphi\|_1 .$$

The key step in our proof of Theorem 3.4, is to estimate

$$\min_{\|h^{(0)} - h\|_{\dot{H}^1} \leq R} \Phi(h) - \Phi(h^{n+1}), \quad h^{(0)} := h^n, \tag{3.12}$$

by a quantitative bound that goes to zero as $R \to +\infty$. This is the content of Proposition 3.11 below. This estimate shows that, even though the initialization of our PDHG scheme $h^{(0)} = h^n$ will, in general, be an infinite $\dot{H}^1(\mathbb{T})$ distance from the optimizer h^{n+1}, we can still make the objective function Φ arbitrarily close to the optimum while remaining finite $\dot{H}^1(\mathbb{T})$ distance from the initialization.

3.3 Proof of convergence of PDHG

We begin by collecting a few basic estimates for the outer semi-implicit time discretization, which are immediate consequences of the definition of the sequence in Eq. (2.7), since $\Phi(h^{n+1}) \leq \Phi(h^n)$.

Lemma 3.7 (basic estimates for semi-implicit scheme) *Let $\mathcal{E}$ be the total variation energy (2.10–2.11), and suppose $h^n \in L^1_M(\mathbb{T}) \cap H^{-1}_{h^n}(\mathbb{T})\ \forall\, n \in \mathbb{N}$ and $\mathcal{E}(h^0) < +\infty$. Then,*

1. $\|h^{n+1}\|_{TV} \leq \|h^n\|_{TV} \leq \cdots \leq \|h^0\|_{TV} < +\infty$,
2. $\|h^{n+1} - h^n\|_{H^{-1}_{h^n}} \leq 2\tau \|h^n\|_{TV} \leq 2\tau \|h^0\|_{TV}$.

Next, we collect a few elementary properties of the space $H^{-1}_h(\mathbb{T})$.

Lemma 3.8 *Suppose $h \in L^1_M(\mathbb{T})$ and $\psi \in H^{-1}_h(\mathbb{T})$. Then there exists $\eta_\psi \in L^1(\mathbb{T})$ so that $\|\eta_\psi\|_1 \leq \|\psi\|_{H^{-1}_h} \|M(h)\|_1$ satisfying*

$$\langle \psi, f \rangle = \int_{\mathbb{T}} \eta_\psi \cdot \nabla f \text{ for all } f \in C^\infty(\mathbb{T}) \quad \text{and} \quad \|\psi\|^2_{H^{-1}_h} = \int_{\mathbb{T}} \frac{|\eta_\psi|^2}{M(h)}.$$

Proof By the definition of H^{-1}_h as the dual of H^1_h and the Riesz-Fréchet Representation theorem, there exists $\xi_\psi \in H^1_h$ so that

$$\langle \psi, f \rangle = \int_{\mathbb{T}} M(h(x)) \nabla f(x) \cdot \nabla \xi_\psi(x) dx \text{ for all } f \in C^\infty(\mathbb{T}) \text{ with mean zero} \tag{3.13}$$

and

$$\|\psi\|^2_{H^{-1}_h} = \|\xi_\psi\|^2_{H^1_h} = \int_{\mathbb{T}} M(h(x)) |\nabla \xi_\psi(x)|^2 dx. \tag{3.14}$$

Note that, due to the fact that we may add or subtract a constant from f without modifying ∇f, Eq. (3.13) holds for all $f \in C^\infty(\mathbb{T})$.

Define $\eta_\psi(x) = \nabla \xi_\psi(x) M(h(x))$. Since $\xi_\psi \in H^1_h$ and $M(h) \in L^1$, by Hölder's inequality,

$$\|\eta_\psi\|_1 \leq \|\nabla \xi_\psi \sqrt{M(h)}\|_2 \|\sqrt{M(h)}\|_2 \leq \|\xi_\psi\|_{H^1_h} \|M(h)\|_1^{1/2} = \|\psi\|_{H^{-1}_h} \|M(h)\|_1^{1/2}.$$

Finally, substituting η_ψ in Eqs. (3.13) and (3.14) above gives the result. □

We will also use the following elementary estimate relating the L^∞ and TV norms.

Lemma 3.9 *If $g \in L^1(\mathbb{T})$, $\int g = 0$, and $\|g\|_{TV} < +\infty$, then $\|g\|_\infty \leq \|g\|_{TV}$.*

Proof Since $g \in BV(\mathbb{T})$ with $\int_{\mathbb{T}} g = 0$, there exist $x_0, x_1 \in \mathbb{T}$ such that $g(x_0) \geq 0$ and $g(x_1) \leq 0$. By the characterization of the total variation norm in terms of variations of g over partitions of $\mathbb{T}$, for any such x_0 and x_1, we have

$$|g(x_0)| + |g(x_1)| = g(x_0) - g(x_1) = |g(x_0) - g(x_1)| \leq \|g\|_{TV}.$$

Hence $\|g\|_\infty \leq \|g\|_{TV}$. □

In order to quantify the decay of (3.12), we construct a competitor h_δ that satisfies the constraint $\|h^n - h_\delta\|_{\dot{H}^1} \leq R$ and for which we can estimate $\Phi(h_\delta) - \Phi(h^{n+1})$ by considering the total variation energy $\mathcal{E}$ and the norm $h \mapsto \|h - h^n\|_{H^{-1}_{h^n}}$ separately.

Lemma 3.10 (construction of competitor) *Let h^{n+1} denote the minimizer of Φ. Then, there exists $h_\delta \in BV(\mathbb{T})$ so that*

1. $\|h_\delta - h^n\|_{\dot{H}^1} \leq \frac{2\sqrt{2\pi}}{\delta}\|\varphi\|_\infty \|h^n\|_{TV}$
2. $\|h_\delta - h^n\|^2_{H^{-1}_{h^n}} - \|h^{n+1} - h^n\|^2_{H^{-1}_{h^n}} \leq 16\pi\delta M_1(\varphi)\|h^n\|^2_{TV}\,\|1/M(h^n)\|_1$;
3. $\|h_\delta\|_{TV} - \|h^{n+1}\|_{TV} \leq \|h^n * \varphi_\delta - h^n\|_{TV}$.

Proof In order to construct our approximating sequence h_δ, we first prove some basic properties of $h^{n+1} - h^n$. By Lemma 3.7 (1), we have $\|h^{n+1}\|_{TV} \leq \|h^n\|_{TV} < +\infty$, so $\|h^{n+1} - h^n\|_{TV} \leq 2\|h^n\|_{TV}$. Furthermore, since h^{n+1} and h^n have mean zero, so does $h^{n+1} - h^n$. Thus, by Lemma 3.9, we conclude

$$\|h^{n+1} - h^n\|_\infty \leq \|h^{n+1} - h^n\|_{TV} \leq 2\|h^n\|_{TV}. \tag{3.15}$$

By Lemma 3.7 (2), we also have

$$\|h^{n+1} - h^n\|_{H^{-1}_{h^n}} \leq 2\tau\|h^n\|_{TV}. \tag{3.16}$$

Therefore, by Lemma 3.8, for $\psi = h^{n+1} - h^n$, there exists $\eta \in L^1(\mathbb{T})$ so that

$$\|\eta\|_1 \leq \|h^{n+1} - h^n\|_{H^{-1}_{h^n}} \|M(h^n)\|_1, \tag{3.17}$$

$$\langle h^{n+1} - h^n, f\rangle = \int \eta f' \text{ for all } f \in C^\infty(\mathbb{T}), \tag{3.18}$$

$$\|h^{n+1} - h^n\|^2_{H^{-1}_{h^n}(\mathbb{T})} = \int \frac{|\eta|^2}{M(h^n)}. \tag{3.19}$$

Since $h^{n+1} - h^n \in L^\infty(\mathbb{T})$, Eq. (3.18) implies that the distributional gradient $\eta' \in L^\infty(\mathbb{T})$, so by Poincaré's inequality, $\eta \in W^{1,\infty}(\mathbb{T})$ with

$$\|\eta\|_\infty \leq 2\pi\|\eta'\|_\infty = 2\pi\|h^{n+1} - h^n\|_\infty \leq 4\pi\|h^n\|_{TV} \tag{3.20}$$

We now use η to construct our approximation h_δ. Fix a compactly supported mollifier $\varphi : \mathbb{R} \to [0, +\infty)$, $\operatorname{supp}\varphi \subseteq B_{2\pi}(0)$, and let $\varphi_\delta(x) = \varphi(x/\delta)/\delta$. (This mollifier

does not need to coincide with that used to regularize the mobility.) Define

$$\eta_\delta := \eta * \varphi_\delta. \tag{3.21}$$

so $(h^{n+1} - h^n) * \varphi_\delta = \eta_\delta'$. We then choose our approximation h_δ to be

$$h_\delta = h^n + (h^{n+1} - h^n) * \varphi_\delta \tag{3.22}$$

With this definition of h_δ in hand, we turn to the proof of item (1) above. By inequality (3.15), we have for all $f \in C^\infty(\mathbb{T})$,

$$\begin{aligned}\left|\int f'(h^n - h_\delta)\right| &= \left|\int f'(h^{n+1} - h^n) * \varphi_\delta\right| = \left|\int (\varphi_\delta * f)'(h^{n+1} - h^n)\right| \\ &\le \|\varphi_\delta * f\|_\infty \|h^{n+1} - h^n\|_{TV} \le \frac{2\sqrt{2\pi}}{\delta}\|\varphi\|_\infty \|h^n\|_{TV}\|f\|_2.\end{aligned}$$

This ensures $h^n - h_\delta \in H^1(\mathbb{T})$ and implies the bound in item (1).

Now, we turn to item (2). First, we estimate the rate at which η_δ converges to η. By definition of η_δ, the fact $\|\eta'\|_\infty = \|h^{n+1} - h^n\|_\infty \le 2\|h^n\|_{TV}$, and inequality (3.15),

$$\begin{aligned}|\eta_\delta(x) - \eta(x)| &= \left|\int_{\mathbb{T}} \varphi_\delta(x - y)(\eta(y) - \eta(x))dy\right| \le \|\eta'\|_\infty \int_{\mathbb{T}} \varphi_\delta(x - y)|x - y|dy \\ &\le 2\delta M_1(\varphi)\|h^n\|_{TV}\end{aligned} \tag{3.23}$$

where $M_1(\varphi)$ is the first moment of φ.

Next, we estimate $\|h_\delta - h^n\|_{H^{-1}_{h^n}}$ in term of η_δ. By definition,

$$\begin{aligned}\|h_\delta - h^n\|_{H^{-1}_{h^n}} &= \|(h^{n+1} - h^n) * \varphi_\delta\|_{H^{-1}_{h^n}} = \sup_{f \in C^\infty(\mathbb{T}) \text{ s.t. } \int f = 0} \frac{\int (h^{n+1} - h^n) * \varphi_\delta f}{\|f\|_{H^1_{h^n}}} \\ &= \sup_{f \in C^\infty(\mathbb{T})} \frac{\int (h^{n+1} - h^n) * \varphi_\delta f}{\|f\|_{H^1_{h^n}}},\end{aligned}$$

where in the last equality, we use that $\int h^{n+1} - h^n = \int (h^{n+1} - h^n) * \varphi_\delta = 0$. Using that $(h^{n+1} - h^n) * \varphi_\delta = (\eta_\delta)'$, integrating by parts, and applying Hölder's inequality, we obtain that, for any $f \in C^\infty(\mathbb{T})$,

$$\begin{aligned}\frac{\int (h^{n+1} - h^n) * \varphi_\delta f}{\|f\|_{H^1_{h^n}}} &= -\frac{\int \eta_\delta f'}{\|f\|_{H^1_{h^n}}} = -\frac{\int \eta_\delta M(h^n)^{-1/2} M(h^n)^{1/2} f'}{\left(\int M(h^n)|f'|^2\right)^{1/2}} \\ &\le \left(\int \frac{|\eta_\delta|^2}{M(h^n)}\right)^{1/2},\end{aligned}$$

Thus, $\|h_\delta - h^n\|_{H^{-1}_{h^n}} \le \left(\int |\eta_\delta|^2 / M(h)\right)^{1/2}$.

We apply this to prove item (2). By Eq. (3.19),

$$\begin{aligned}\|h_\delta - h^n\|^2_{H^{-1}_{h^n}} - \|h^{n+1} - h^n\|^2_{H^{-1}_{h^n}} &= \|(h^{n+1} - h^n) * \varphi_\delta\|^2_{H^{-1}_{h^n}} - \|h^{n+1} - h^n\|^2_{H^{-1}_{h^n}} \\ &\leq \int \frac{|\eta_\delta|^2}{M(h^n)} - \int \frac{|\eta|^2}{M(h^n)} \\ &= \int \frac{1}{M(h^n)} (\eta_\delta - \eta)(\eta_\delta + \eta) \\ &\leq \|\eta_\delta - \eta\|_\infty (\|\eta_\delta\|_\infty + \|\eta\|_\infty) \|1/M(h^n)\|_1 \\ &\leq 16\pi\delta M_1(\varphi)\|h^n\|^2_{TV} \left\|1/M(h^n)\right\|_1,\end{aligned}$$

where, in the last inequality, we apply our uniform bound on η, inequality (3.20), and our uniform estimate on the convergence of η_δ to η, inequality (3.23). This completes the proof of (2).

We conclude by showing item (3). By the triangle inequality,

$$\begin{aligned}\|h_\delta\|_{TV} - \|h^{n+1}\|_{TV} &= \|h^n + (h^{n+1} - h^n) * \varphi_\delta\|_{TV} - \|h^{n+1}\|_{TV} \\ &\leq \|h^n * \varphi_\delta - h^n\|_{TV} + \|h^{n+1} * \varphi_\delta\|_{TV} - \|h^{n+1}\|_{TV}. \quad (3.24)\end{aligned}$$

Furthermore, for any $f \in C^\infty(\mathbb{T})$,

$$\begin{aligned}-\int f'(h^{n+1} * \varphi_\delta) = -\int (\varphi_\delta * f)' h^{n+1} &\leq \|\varphi_\delta * f\|_\infty \|h^{n+1}\|_{TV} \\ &\leq \|f\|_\infty \|h^{n+1}\|_{TV}.\end{aligned}$$

Therefore $\|h^{n+1} * \varphi_\delta\|_{TV} \leq \|h^{n+1}\|_{TV}$, which combined with (3.24) gives item (3). □

We now apply this lemma to prove our key estimate, quantifying the rate of convergence of functions h that are a finite distance from the initialization h^n in the $\dot{H}^1$ norm to the optimizer of Φ.

Proposition 3.11 *For any compactly supported mollifier φ_δ, there exists an explicit constant $C > 0$ depending on $\|h^n\|_{TV}$, $\|M(h^n)\|_1$, and $\|1/M(h^n)\|_1$, so that*

$$\min_{\|h^n - h\|_{\dot{H}^1} \leq R} \Phi(h) - \Phi(h^{n+1}) \leq \frac{C}{R} + \|h^n * \varphi_\delta - h^n\|_{TV} \quad (3.25)$$

where $\delta = \frac{2\sqrt{2\pi}}{R}\|\varphi\|_\infty \|h^n\|^2_{TV}$.

Proof Let h_δ be as in Lemma 3.10 and choose $\delta > 0$ so that

$$\frac{2\sqrt{2\pi}}{\delta}\|\varphi\|_\infty \|h^n\|_{TV} = R. \quad (3.26)$$

Then Lemma 3.10 (1) guarantees that $\|h_\delta - h^n\|_{\dot{H}^1} \le R$, so h_δ is a candidate for the minimization problem (3.25). Therefore, it suffices to bound the objective functional when $h = h_\delta$. By Lemma 3.10 (2) and (3), we have

$$\left(\frac{1}{2\tau}\|h_\delta - h^n\|^2_{H^{-1}_{h^n}} + \|h_\delta\|_{TV}\right) - \left(\frac{1}{2\tau}\|h^{n+1} - h^n\|^2_{H^{-1}_{h^n}} + \|h^{n+1}\|_{TV}\right)$$
$$\le 16\pi\delta M_1(\varphi)\|h^n\|^2_{TV}\|1/M(h^n)\|_1 + \|h^n * \varphi_\delta - h^n\|_{TV}$$

which, combined with Eq. (3.26), gives the result. □

We now turn to the proof of our main result, Theorem 3.4, which shows that the PDGH algorithm converges to the optimizer in the ergodic sense: that is, if $h^{(M)}$ is the ergodic sequence (3.11), then $\lim_{M\to+\infty} \Phi(h^{(M)}) = \Phi(h^{n+1})$.

Proof of Theorem 3.4 Following Chambolle and Pock [9, equation 17] and Jacobs, Léger, Li, and Osher [22], we consider the partial primal-dual gap

$$\mathcal{G}_{R_1,R_2}(h,\phi) := \sup_{\hat{\phi}:\left\|\hat{\phi}-\phi^{(0)}\right\|_2\le R_1} \mathcal{L}(h,\hat{\phi}) - \inf_{\hat{h}:\left\|\hat{h}-h^0\right\|_{\dot{H}^1}\le R_2} \mathcal{L}(\hat{h},\phi) \tag{3.27}$$

where $\mathcal{L}(h,\phi)$ is the Lagrangian defined in Eq. (3.3). Since the gradient operator $\partial_x : \dot{H}^1(\mathbb{T}) \to L^2(\mathbb{T})$ satisfies

$$\|h'\|_2 = \|h\|_{\dot{H}^1}, \quad \forall h \in \dot{H}^1,$$

the operator norm of the gradient is one. Consequently, by Chambolle and Pock [9, Theorem 1], if $\lambda\sigma \le 1$, then along the ergodic sequences (3.11),

$$\mathcal{G}_{R_1,R_2}(h^{(M)},\phi^{(M)}) \le \frac{1}{M}\left(\frac{R_1^2}{\sigma} + \frac{R_2^2}{\lambda}\right). \tag{3.28}$$

We seek to bound each term in the partial primal-dual gap separately. Since $\|\phi^{(0)}\|_\infty \le 1$ (in fact, in practice we take $\phi^{(0)} = 0$) we have

$$\{\hat{\phi} : \|\hat{\phi}\|_\infty \le 1\} \subseteq \{\hat{\phi} : \|\hat{\phi} - \phi^{(0)}\|_2 \le 2\sqrt{2\pi}\}. \tag{3.29}$$

Since $\hat{\phi} \mapsto \mathcal{L}(h^{(M)},\hat{\phi}) \ne -\infty$ only if $\|\hat{\phi}\|_\infty \le 1$, this implies

$$\sup_{\hat{\phi}:\left\|\hat{\phi}-\phi^{(0)}\right\|_2\le 2\sqrt{2\pi}} \mathcal{L}(h^{(M)},\hat{\phi}) = \sup_{\hat{\phi}:\|\hat{\phi}\|_\infty\le 1} \mathcal{L}(h^{(M)},\hat{\phi}) = \Phi(h^{(M)}). \tag{3.30}$$

By definition of $\phi^{(m+1)}$ in Eq. (3.7), $F^*(\phi^{(m+1)}) < +\infty$, so $\|\phi^{(m+1)}\|_\infty \le 1$ for all $m \in \mathbb{N}$ and the ergodic sequence also satisfies $\|\phi^{(M)}\|_\infty \le 1$ for all $M \in \mathbb{N}$. Thus,

for any $R > 0$,

$$\inf_{\hat{h}:\left\|\hat{h}-h^n\right\|_{\dot{H}^1}\leq R} \mathcal{L}(\hat{h},\phi^{(M)}) \leq \inf_{\hat{h}:\left\|\hat{h}-h^n\right\|_{\dot{H}^1}\leq R} \sup_{\phi:\|\phi\|_\infty\leq 1} \mathcal{L}(\hat{h},\phi) = \inf_{\hat{h}:\left\|\hat{h}-h^n\right\|_{\dot{H}^1}\leq R} \Phi(\hat{h}). \tag{3.31}$$

Combining these estimates, we conclude that for any $R > 0$,

$$\begin{aligned}
\Phi(h^{(M)}) - \Phi(h^{n+1}) &\overset{(3.30)}{=} \sup_{\hat{\phi}:\left\|\hat{\phi}-\phi^n\right\|_2\leq 2\sqrt{2\pi}} \mathcal{L}(h^{(M)},\hat{\phi}) - \Phi(h^{n+1}) \\
&\overset{(3.27)}{=} \mathcal{G}_{2\sqrt{2\pi},R}(h^{(M)},\phi^{(M)}) + \inf_{\hat{h}:\left\|\hat{h}-h^n\right\|_{\dot{H}^1}\leq R} \mathcal{L}(\hat{h},\phi^{(M)}) - \Phi(h^{n+1}) \\
&\overset{(3.31)}{\leq} \mathcal{G}_{2\sqrt{2\pi},R}(h^{(M)},\phi^{(M)}) + \inf_{\hat{h}:\left\|\hat{h}-h^n\right\|_{\dot{H}^1}\leq R} \Phi(\hat{h}) - \Phi(h^{n+1}) \\
&\overset{(3.28)}{\leq} \frac{1}{M}\left(\frac{8\pi}{\sigma} + \frac{R^2}{\lambda}\right) + \inf_{\hat{h}:\left\|\hat{h}-h^n\right\|_{\dot{H}^1}\leq R} \Phi(\hat{h}) - \Phi(h^{n+1}) \\
&\overset{(3.25)}{\leq} \frac{1}{M}\left(\frac{8\pi}{\sigma} + \frac{R^2}{\lambda}\right) + \frac{C}{R} + \|h^n * \varphi_\delta - h^n\|_{TV},
\end{aligned}$$

where $\delta = \frac{2\sqrt{2\pi}}{R}\|\varphi\|_\infty \|h^n\|_{TV}$ and $C > 0$ depends on $\|h^n\|_{TV}$, $\|M(h^n)\|_1$, and $\|1/M(h^n)\|_1$. We may optimize the first term on the right hand side by choosing

$$\sigma = 2\sqrt{2\pi}/R, \quad \lambda = R/2\sqrt{2\pi},$$

in which case we obtain

$$\Phi(h^{(M)}) - \Phi(h^{n+1}) \leq \frac{4\sqrt{2\pi}R}{M} + \frac{C}{R} + \|h^n * \varphi_\delta - h^n\|_{TV}. \tag{3.32}$$

We claim that, since $\|h^n\|_{TV} < +\infty$,

$$\lim_{R\to+\infty} \|h^n * \varphi_\delta - h^n\|_{TV} = 0. \tag{3.33}$$

Thus, we conclude the existence of M_*, λ_*, σ_* such that for all $M \geq M_*$, we have an ϵ-approximate solution.

It remains to prove the claim (3.33). Note that if $\phi \in C^\infty$ satisfies $\|\phi\|_\infty \le 1$, then $\|\varphi_\delta * \phi\|_\infty \le 1$ for all $\delta > 0$. Hence,

$$\begin{aligned}\|h^n * \varphi_\delta\|_{TV} &= \sup_{\|\phi\|_\infty \le 1} \int -(h^n * \varphi_\delta)\phi' = \sup_{\|\phi\|_\infty \le 1} \int -h^n(\phi * \varphi_\delta)' \\ &\le \sup_{\|\phi\|_\infty \le 1} \int -h^n \phi' = \|h^n\|_{TV} < +\infty\end{aligned}$$

This shows $\|h^n * \varphi_\delta - h^n\|_{TV} < +\infty$. Hence, for all $\epsilon > 0$, there exists $\phi \in C^\infty$ so

$$\|h^n * \varphi_\delta - h^n\|_{TV} \le -\int (h^n * \varphi_\delta - h^n)\phi' + \epsilon = -\int h^n(\phi * \varphi_\delta - \phi)' + \epsilon. \tag{3.34}$$

Since ϕ is a smooth function on a compact set, sending $\delta \to 0$, we conclude that $\limsup_{\delta \to 0} \|h^n * \varphi_\delta - h^n\|_{TV} \le \epsilon$. Since $\epsilon > 0$ was arbitrary, this proves our claim, again using Eq. (3.26), relating δ and R.

Now, suppose the function h^n satisfies a higher regularity assumption: $(h^n)' \in BV(\mathbb{T})$. Following the same argument as in Eq. (3.34), we have

$$\|h^n * \varphi_\delta - h^n\|_{TV} = \sup_{\|\phi\|_\infty \le 1} \int (h^n)' \cdot (\phi * \varphi_\delta - \phi).$$

Furthermore,

$$\begin{aligned}\int (h^n)' \cdot (\phi * \varphi_\delta - \phi) &= \iint (h^n)'(x) \cdot (\phi(x - y) - \phi(x))\varphi_\delta(y) dy dx \\ &= -\delta \iint_{\mathbb{T}\times\mathbb{T}} \int_0^1 ((h^n)'(x))^t D\phi(x - sy) y \varphi(y) ds dy dx \\ &\le \delta \|(h^n)'\|_{TV} M_1(\varphi) \|\phi\|_\infty\end{aligned}$$

As a consequence, Eq. (3.32) becomes

$$\Phi(h^{(M)}) - \Phi(h^{n+1}) \le \frac{4\sqrt{2\pi}R}{M} + \frac{C}{R} + \delta \|(h^n)'\|_{TV} M_1(\varphi),$$

where $\delta = C'/R$, for $C' = 2\sqrt{2\pi}\|\varphi\|_\infty \|h^n\|_{TV}$. Thus, to obtain an $\epsilon > 0$ accurate solution, we require

$$M \ge 4\sqrt{2\pi}R \left(\epsilon - \frac{C''}{R}\right)^{-1}, \quad \text{for } C'' = C + C'\|(h^n)'\|_{TV} M_1(\varphi).$$

Optimizing in $R \geq 0$, we obtain that for $R = 2C''/\epsilon$, the choices

$$M_* := \frac{16\sqrt{2\pi}C''}{\epsilon^2}, \quad \lambda_* = \frac{C''}{\sqrt{2\pi}\epsilon}, \quad \sigma_* = \frac{\sqrt{2\pi}\epsilon}{C''},$$

ensure that $(h^{(M)}, \phi^{(M)})$ is an ϵ-approximate solution for all $M \geq M_*$. □

4 Fully discrete numerical method

In this section, we describe how the discrete time, spatially continuous PDHG algorithm introduced in Sect. 3.1 can be implemented as a fully discrete numerical method for simulating crystal surface evolution. In one spatial dimension, let $[0, 2\pi]$ be the computational domain with periodic boundary conditions and Δx and τ be the spatial grid spacing and outer time step, respectively. Choose $0 = x_1 < \cdots < x_{N_x} = 2\pi - \Delta x$, where $x_j = (j-1)\Delta x$, $\Delta x = \frac{2\pi}{N_x}$. For notational simplicity, let h and ϕ denote the discrete vector approximations in $\mathbb{R}^{N_x}$ of the corresponding functions,

$$h = (h_1, \cdots, h_{N_x})^t, \quad \phi = (\phi_1, \cdots \phi_{N_x})^t.$$

Let D and A be the matrix approximations of the operators ∇ and $-\Delta_h$, where D is given by a centered difference method and $\mathsf{A} := \mathsf{D}^t \operatorname{diag}(M(h)_1, \ldots, M(h)_{N_x})\mathsf{D}$.

We discretize our PDHG method (3.8)–(3.10) via a finite difference scheme, replacing the spatially continuous operators with their discrete counterparts. This leads to Algorithm 4.1. Finally, we construct our numerical solution $h(x,t)$ for the crystal surface evolution equation by linearly interpolating between the spatial gridpoints and taking a piecewise constant interpolation between the outer discrete time sequence h^n.

Algorithm 4.1 PDHG for crystal surface evolution

Input: h^0, T, τ, λ, σ, Choose $\delta > 0$ sufficiently small
$n = 0$
while $n\tau \leq T$ **do**
  Let $h^{(0)} = h^n$, $\phi^{(0)} = 0$, and $m = 0$;
  repeat
    $h^{(m+1)} = (\mathsf{D}^t\mathsf{D} + \frac{\lambda}{\tau}\mathsf{A}^{-1}(\cdot - h^n))^{-1}\left(\mathsf{D}^t\mathsf{D}h^{(m)} - \lambda\mathsf{D}^t\phi^{(m)}\right)$,
    $\bar{h}^{(m+1)} = 2h^{(m+1)} - h^{(m)}$,
    $\phi^{(m+1)} = (\mathrm{I} + \sigma\partial F^*)^{-1}(\phi^{(m)} + \sigma D\bar{h}^{(m+1)})$,
    $m = m + 1$,
  until *stopping criteria are achieved:* $|(h^{(m+1)} - h^{(m)}, \phi^{(m+1)} - \phi^{(m)})| < \delta$;
  $h^{n+1} = \bar{h}^{(m+1)}$ and $n = n + 1$,
end

Remark 4.1 (*Convergence of fully discrete algorithm*) Using standard estimates relating finite difference operators to their continuum counterparts, one could adapt our

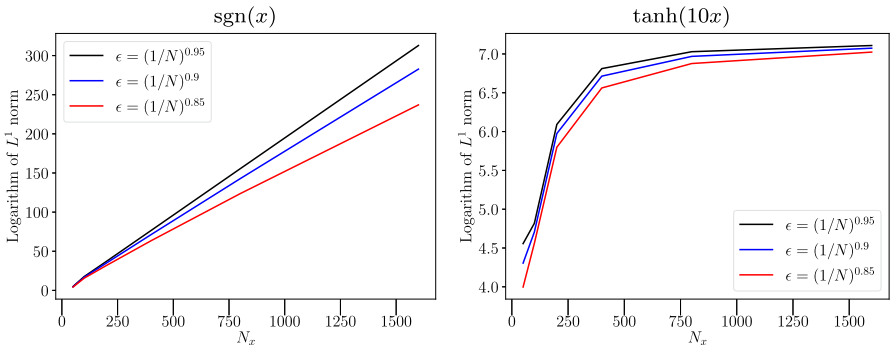


Fig. 1 Choosing the function $\mathrm{sgn}(x)$ or $\tanh(x)$ in the mobility (4.2) leads to different behavior as $\epsilon \to 0$, $N_x \to +\infty$. Above, we consider the spatially discrete mobility for height profile $h(x) = \sin(x)$. Left: For the original mobility, with $\mathrm{sgn}(x)$, even when $\epsilon \to 0$ slowly as $N_x \to +\infty$, the L^1 norm of the mobility diverges. Right: Approximating with $\tanh(10x)$ allows us to send $\epsilon \to 0$ rapidly as $N_x \to +\infty$, while preserving a uniform bound on the L^1 norm of the mobility

main convergence result, Theorem 3.4, to be a convergence result for the fully discrete PDHG method, which comprise the inner iterations of Algorithm 4.1. See, for example, work by Wang and Lucier [49], which considers related estimates for the Rudin-Osher-Fatemi image denoising model.

In practice, to avoid inverting a near-singular matrix in our computation of $h^{(m+1)}$, we compute the inverse operator in the definition of $h^{(m+1)}$ via

$$\left(\mathsf{D}^t\mathsf{D} + \frac{\lambda}{\tau}\mathsf{A}^{-1}(\cdot - h^n)\right)^{-1} u = \left(\frac{\tau}{\lambda}\mathsf{A}\mathsf{D}^t\mathsf{D} + \mathsf{I}\right)^{-1}\left(\frac{\tau}{\lambda}\mathsf{A}u + h^n\right). \tag{4.1}$$

On the other hand, in order to compute $\phi^{(m+1)}$, we use the explicit formula

$$(\mathrm{I} + \sigma\,\partial F^*)^{-1}(u) = \left[\min(|u_i|, 1)\,\mathrm{sgn}(u_i)\right],$$

where u_i denotes the ith component of the vector u. Note that, while other initializations of the dual variable $\phi^{(0)}$ are possible (for example, initializing $\phi^{(0)}$ to coincide with the last value of $\phi^{(m+1)}$ at the previous outer time step), we observe slightly better performance always initializing $\phi^{(0)} = 0$.

We discretize our regularized mobility as follows:

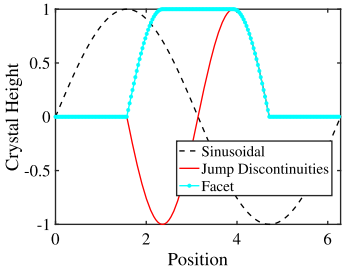
$$M(h) := e^{-\nabla\varphi_\epsilon * \mathrm{sgn}(f)}, \quad f = \mathrm{minmod}\{D_+h, D_-h\}, \tag{4.2}$$

where $D_\pm$ denotes the forward/backward finite difference operators. The minimum modulus limiter of the gradient allows us to respect shock-like objects in the facet formation; see, e.g., [40]. Heuristically, this enforces the property of the original, unregularized mobility (1.5) that once a region of the crystal surface becomes flat at a location x_0, i.e. $\frac{d}{dx}h(x_0, t) = 0$, the surface remains flat at x_0. We compute the convolution in (4.2) via a fast Fourier transform.

Finally, in our simulations, we sometimes approximate $\mathrm{sgn}(x)$ in the definition of the mobility with $\tanh(10x)$. In order to achieve accurate facet formation, we must strike a balance between choosing the spatial discretization N_x large and the mobility regularization parameter $\epsilon > 0$ small. As illustrated in Fig. 1, the original $\mathrm{sgn}(x)$ function is extremely sensitive to small choices of ϵ, which quickly cause the L^1 norm of the mobility to become unbounded as $\epsilon \to 0$, $N_x \to +\infty$, going against the assumption in our convergence result for the PDHG method, Theorem 3.4, which was proved for fixed $\epsilon > 0$. On the other hand, the $\tanh(10x)$ approximation allows us to refine ϵ and N_x simultaneously, while keeping the L^1 norm of the mobility bounded. A thorough analysis of these limits is related to the question of existence of solutions to the crystal surface evolution equation, and we leave a detailed study to future work.

5 Numerical results

In this section, we present a range of numerical examples illustrating the performance of the proposed algorithm. In each test, we consider the stopping criteria $\|(h^{(m+1)} - h^{(m)}, \phi^{(m+1)} - \phi^{(m)})\| < \delta$, where we take the threshold $\delta = 5 \times 10^{-6}$. Unless otherwise specified, the outer time step for the semi-implicit scheme h^n is chosen to be $\tau = T/10$, where T is the final computational time, so that $N_t = 10$. In order to ensure that the matrix inverse in the definition of $h^{(m+1)}$, Eq. (4.1), is well defined, we choose λ sufficiently large so that $\frac{\tau}{\lambda}\|\mathsf{A}\mathsf{D}^t\mathsf{D}\| < 1$. In the following examples, we choose $\sigma = 5 \times 10^{-4}$, $\lambda = 500$ for all N_x. We consider three choices of initial data, as shown in Fig. 2.

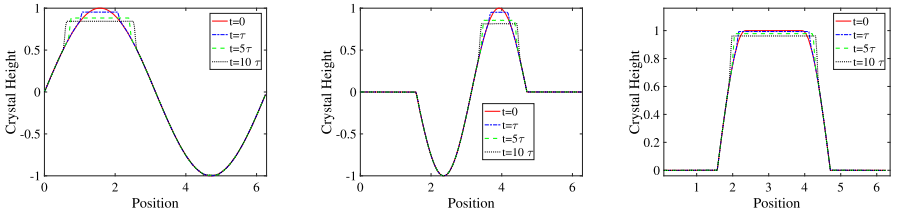


$$h_{\mathrm{sine}}(x) = \sin(x)$$

$$h_{\mathrm{jump}}(x) = \begin{cases} \sin(2x) & x \in (\pi/2, 3\pi/2) \\ 0 & \text{otherwise} \end{cases}$$

$$h_{\mathrm{facet}}(x) = \begin{cases} \sin(2(x-\pi/2)) & x \in \left(\frac{\pi}{2}, \frac{3\pi}{4}\right) \\ 1 & x \in (3\pi/4, 5\pi/4) \\ \cos(2(x-5\pi/4)) & x \in (5\pi/4, 3\pi/2) \\ 0 & \text{otherwise} \end{cases}$$

Fig. 2 Choices of initial data



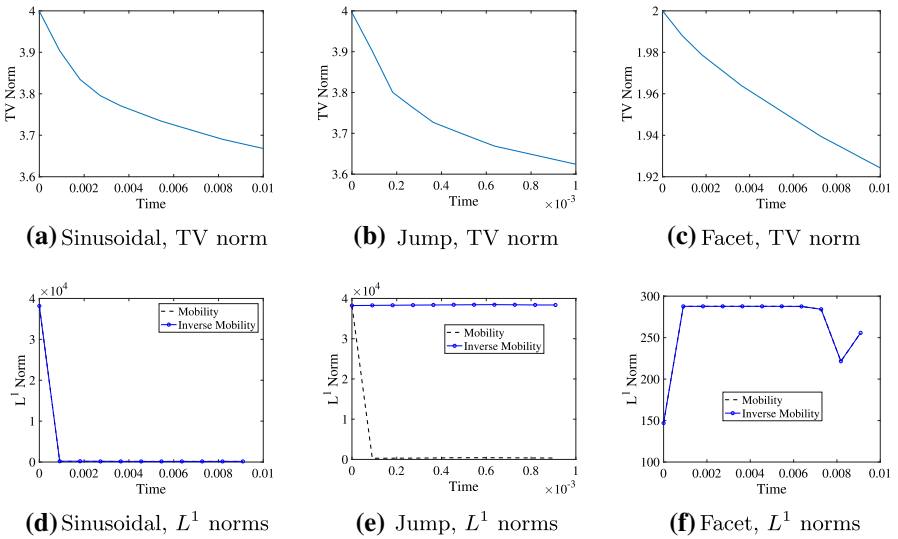
(a) Sinusoidal **(b)** Jump **(c)** Facet

Fig. 3 Dynamics of crystal surface evolution equation for different choices of initial data. Near maxima, flat facets form and expand outward, while minima remain stationary

In Fig. 3, we display the dynamics of the crystal surface evolution equation for each choice of initial data depicted in Fig. 2. We chose $\epsilon = 0.04$, $N_x = 200$ in each of these calculations, letting $T = 10^{-2}$ in the case of the (a) Sinusoidal ($h_0 = h_{\text{sine}}$) and the (c) Facet dynamics ($h_0 = h_{\text{facet}}$) and $T = 10^{-3}$ for the (b) Jump ($h_0 = h_{\text{jump}}$) dynamics. Near the maxima, flat facets expand outward like a free boundary type solution, while the minimum is stationary, as predicted in [33].

Remark 5.1 (*Comparison to related numerical methods*) We note that, qualitatively, the solutions we find using the algorithm we present here match the solutions found in [33] for Eq. (1.2), where the authors implemented a very different numerical scheme. There, a numerical scheme was developed by regularizing the mobility and using stiff solvers in time, with no theoretical basis for convergence to the desired dynamics, though it produced comparable dynamics to those predicted by an exact formula for the facet dynamics, also derived in [33]. Unfortunately, due to the extreme stiffness of the problem, introduced by the mobility, it was impossible to successively refine the grid in the previous numerical scheme to do a proper numerical convergence study. (Indeed, we consider it a notable strength of our present method that we are able to refine the spatial discretization and perform a robust numerical convergence analysis in Fig. 6.) The inability to refine the grid in the previous scheme prevents quantitative comparison with the new method introduced in the present paper. Instead, we simply remark that, on coarse grids, both schemes provide qualitatively similar simulations for each choice of initial data.

In Fig. 4, we analyze properties of the numerical method, under the same choices of parameters as in Fig. 3 and with the same ordering of each type of initial data



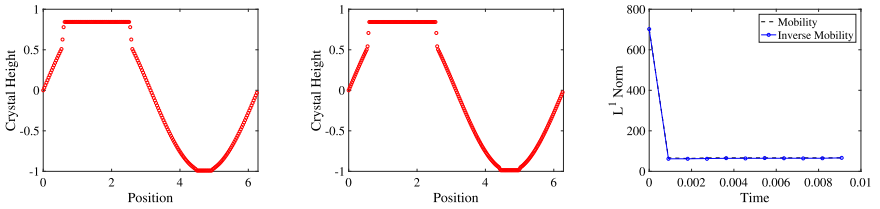
(a) Sinusoidal, TV norm **(b)** Jump, TV norm **(c)** Facet, TV norm

(d) Sinusoidal, L^1 norms **(e)** Jump, L^1 norms **(f)** Facet, L^1 norms

Fig. 4 Top row **a**–**c** The total variation energy decreases in time along numerical solutions for our initial data types, reflecting the underlying gradient flow structure. Bottom row **d**–**f** The L^1 norms of the mobility $M(h)$ and the reciprocal of the mobility $1/M(h)$ are large, but remain bounded along the flow for each type of initial data

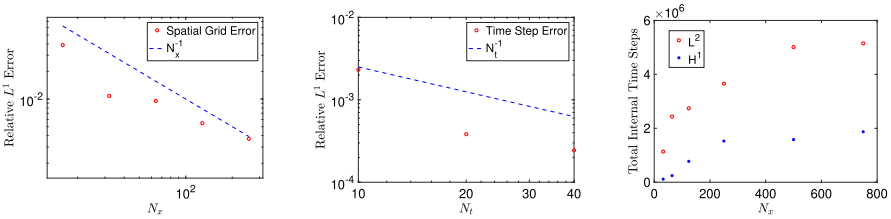
(Sinusoidal, Jump, Facet). In the top row (a-c), we show the decrease in the discrete TV norm $\|Dh\|_1$ in time along solutions of the equation for each of our initial data, reflecting the gradient flow structure of the equation. In the bottom row (d-f), we plot the L^1 norms of the mobility $M(h)$ and its reciprocal $1/M(h)$ for each type of initial data. A key assumption in our convergence result for the PDHG method, Theorem 3.4, is that both remain bounded, uniformly in the spatial discretization. We can see in the above simulations that, while these norms are very large, they indeed remain bounded along the flow.

In Fig. 5, we compare two different choices of mobility for Sinusoidal initial data: (a) Eq. (4.2) and (b) a modified mobility, replacing $\mathrm{sgn}(x)$ with $\tanh(10x)$. In both cases, we take $\epsilon = .04$. As can be seen on the plot (c) containing the corresponding L^1 norms for the modified mobility, on one hand, the modified mobility has the benefit of drastically decreasing the L^1 norm of the mobility and its reciprocal: compare the plot (c) of Fig. 5 to plot (d) of Fig. 4 where the mobility is as in 4.2. The method also requires fewer iterations to meet the stopping criteria. On the other hand, the modified mobility allows for slightly more movement and facet formation at the minimum, which goes against the predicted dynamics of the original equation: compare the plot on the left with the plot in the middle.



(a) Sinusoidal, $\mathrm{sgn}(x)$ Mobility **(b)** Sinusoidal, $\tanh(10x)$ Mobility **(c)** Sinusoidal, $\tanh(10x)$ Mobility, TV norm

Fig. 5 **a**, **b** We compare the dynamics for Sinusoidal initial data for the mobility given by Eq. (4.2) and for a modified mobility, in which $\mathrm{sgn}(x)$ is replaced by $\tanh(10x)$. **c** While the original mobility more accurately prevents facet formation at the local minimum, the modified mobility leads has smaller L^1 norm and requires fewer iterations to converge



(a) Spatial Grid Error Plot **(b)** External Time Error Plot **(c)** Internal time steps

Fig. 6 **a** Log–Log plot of relative L^1 error versus spatial grid size. **b** Log–Log plot of relative L^1 error versus external time step. **c** Comparison of number of time steps required to meet stopping criteria for either $\dot{H}^1$ or L^2 penalization. We observe superior performance for the $\dot{H}^1$ penalization, especially as the spatial grid is refined

Finally, in Fig. 6, we analyze the rate of convergence of our method. We consider sinusoidal initial data with the modified mobility, replacing $\mathrm{sgn}(x)$ with $\tanh(10x)$, $\epsilon = .05$ and $T = 10^{-4}$. In plot (a) of of Fig. 6, we examine how the relative L^1 error depends on the number of spatial gridpoints N_x for a fixed temporal discretization, $N_t = 10$. For $N_x = 16, 32, 64, 128, 256, 512$, we plot $\|h(N_x) - h(2N_x)\|_{L^1}$. We observe slightly sublinear convergence, in line with the low spatial regularity of our solutions.

In plot (b) of Fig. 6, we examine how the relative L^1 error scales with the external time step, used to define the semi-implicit scheme h^n via $\tau = T/N_t$, for a fixed spatial discretization $N_x = 256$. For $N_t = 5, 10, 20, 40, 80$, we plot $\|h(N_t) - h(2N_t)\|_{L^1}$. We observe approximately first order convergence, in agreement with the interpretation of our scheme as a semi-implicit version of the minimizing movements scheme, which can be thought of as a generalized Euler method.

In plot (c) of Fig. 6, we illustrate the importance of the choice of norms in our PDHG algorithm, as explained in Remark 3.2. This discussion is born out numerically in that compare the number of iterations required for each method as the spatial grid is refined, $N_x = 32, 64, 124, 250, 500, 750$. We consider $T = 10^{-6}$ external time steps, setting $\sigma = 5 \times 10^{-5}$, $\lambda = 5 \times 10^{-5}$ for the L^2 algorithm (the largest we could take to allow convergence for the L^2 Algorithm to still converge at all scales) and $\sigma = 5 \times 10^{-4}$, $\lambda = 500$ for our $\dot{H}^1$ algorithm, Algorithm 4.1. At the fully discrete level, existing work [9] ensures that the PDHG algorithm would converge, even if the norm penalization in the definition of $h^{(m+1)}$ was changed from a $\dot{H}^1$ norm to a L^2 norm. At the level of Algorithm 4.1, this would amount to modifying the computation of $h^{(m+1)}$ as follows:

$$h^{(m+1)} = \left(\mathsf{I} + \frac{\lambda}{\tau}\mathsf{A}^{-1}(\cdot - h^n)\right)^{-1}\left(h^{(m)} - \lambda \mathsf{D}^t \phi^{(m)}\right). \tag{5.1}$$

On one hand, to invert the matrix in the above formula, we need $\frac{\tau}{\lambda}\|\mathsf{A}\| < 1$. On the other hand, existing convergence results on PDHG require $\lambda\sigma\|\mathsf{D}^t\mathsf{D}\| < 1$, where $\|\mathsf{D}^t\mathsf{D}\| \to +\infty$ as the spatial grid is refined. These requirements lead to significant tension regarding the size of λ. In contrast, when choosing the $\dot{H}^1$ norm to penalize the primal variables in our PDHG algorithm, the analogue of the constraint $\lambda\sigma\|\mathsf{D}^t\mathsf{D}\| < 1$ is simply $\lambda\sigma < 1$, since the gradient is a bounded operator on $\dot{H}^1$. Thus, our method avoids this source of tension in the definition of the inner time steps λ, σ.

Remark 5.2 (*Dependence on regularization* $\epsilon > 0$) In the preceding simulations, we consider strictly positive regularization parameters $\epsilon > 0$. Note that our method would not perform well if $\epsilon \to 0$ for a fixed spatial discretization N_x and outer time step τ, since this would cause $\|1/M(h)\|_1$ to become unbounded, violating a key hypothesis in our convergence result, Theorem 3.4. An examination of the proof of Proposition 3.11 and Theorem 3.4 shows that, as $\|1/M(h^n)\|_1 \to +\infty$, we have $\lambda_* \to +\infty$, $\sigma_* \to 0$, and $M_* \to +\infty$: the primal time step becomes arbitrarily large, the dual time step becomes arbitrarily small, and the number of iterations required to obtain a solution of a given accuracy becomes arbitrarily large.

A challenging question that we leave to future work is to find a scaling relationship between ϵ, N_x and τ that allows one to send $\epsilon \to 0$ as $N_x \to +\infty$ and $\tau \to 0$. This is closely related to the question of proving existence of solutions to the PDE and convergence of the outer time discretization to such solutions.

References

1. Ambrose, D.M.: The radius of analyticity for solutions to a problem in epitaxial growth on the torus. Bull. Lond. Math. Soc. **51**(5), 877–886 (2019)
2. Ambrosio, L., Gigli, N., Savaré, G.: Gradient flows in metric spaces and in the space of probability measures, 2nd edn. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel (2008)
3. Bonzel, H., Preuss, E.: Morphology of periodic surface profiles below the roughening temperature: aspects of continuum theory. Surf. Sci. **336**(1–2), 209–224 (1995)
4. Burton, W.-K., Cabrera, N., Frank, F.: The growth of crystals and the equilibrium structure of their surfaces. Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Sci. **243**(866), 299–358 (1951)
5. Cancès, C., Gallouët, T.O., Todeschi, G.: A variational finite volume scheme for Wasserstein gradient flows. arXiv:1907.08305 (2019)
6. Carlier, G., Duval, V., Peyré, G., Schmitzer, B.: Convergence of entropic schemes for optimal transport and gradient flows. SIAM J. Math. Anal. **49**(2), 1385–1418 (2017)
7. Carrillo, J.A., Craig, K., Wang, L., Wei, C.: Primal dual methods for Wasserstein gradient flows. arXiv:1901.08081 (2019)
8. Carrillo, J.A., Laurençot, P., Rosado, J.: Fermi–Dirac–Fokker–Planck equation: well-posedness and long-time asymptotics. J. Differ. Equ. **247**(8), 2209–2234 (2009)
9. Chambolle, A., Pock, T.: On the ergodic convergence rates of a first-order primal-dual algorithm. Math. Program. **159**(1–2), 253–287 (2016)
10. Elliott, C.M., Garcke, H.: On the Cahn–Hilliard equation with degenerate mobility. SIAM J. Math. Anal. **27**(2), 404–423 (1996)
11. Gao, Y.: Global strong solution with BV derivatives to singular solid-on-solid model with exponential nonlinearity. J. Diff. Equ. **267**(7), 4429–4447 (2019)
12. Gao, Y., Katsevich, A.E., Liu, J.-G., Lu, J., Marzuola, J.L.: Analysis of a fourth order exponential pde arising from a crystal surface jump process with Metropolis-type transition rates. Pure Appl. Anal. **3**, 595–612 (2021)
13. Gao, Y., Liu, J.-G., Lu, J., Marzuola, J.L.: Analysis of a continuum theory for broken bond crystal surface models with evaporation and deposition effects. Nonlinearity **33**, 3816–3845 (2020)
14. Giga, M.-H., Giga, Y.: Very singular diffusion equations: second and fourth order problems. Jpn. J. Ind. Appl. Math. **27**(3), 323–345 (2010)
15. Giga, Y., Kohn, R.V.: Scale-invariant extinction time estimates for some singular diffusion equations. Discrete Contin. Dyn. Syst **30**(2), 509–535 (2011)
16. Giga, Y., Kuroda, H., Matsuoka, H.: Fourth-order total variation flow with Dirichlet condition: characterization of evolution and extinction time estimates. Hokkaido Univ. Preprint Ser. Math. **1064**, 1–36 (2015)
17. Giga, Y., Ueda, Y.: Numerical computations of split Bregman method for fourth order total variation flow. J. Comput. Phys. **405**, 109114 (2020)
18. Granero-Belinchón, R., Magliocca, M.: Global existence and decay to equilibrium for some crystal surface models. arXiv:1804.09645 (2018)
19. Gruber, E., Mullins, W.: On the theory of anisotropy of crystalline surface tension. J. Phys. Chem. Solids **28**(5), 875–887 (1967)
20. He, B., Yuan, X.: Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective. SIAM J. Imag. Sci. **5**(1), 119–149 (2012)
21. Ihle, T., Misbah, C., Pierre-Louis, O.: Equilibrium step dynamics on vicinal surfaces revisited. Phys. Rev. B **58**(4), 2289 (1998)
22. Jacobs, M., Léger, F., Li, W., Osher, S.: Solving large-scale optimization problems with a convergence rate independent of grid size. SIAM J. Numer. Anal. **57**(3), 1100–1123 (2019)
23. Jordan, R., Kinderlehrer, D., Otto, F.: The variational formulation of the Fokker-Planck equation. SIAM J. Math. Anal. **29**(1), 1–17 (1998)

24. Katsevich, A.: From local equilibrium to numerical pde: metropolis crystal surface dynamics in the rough scaling limit. arXiv:2108.03527 (2021)
25. Katsevich, A.: The local equilibrium state of a crystal surface jump process in the rough scaling regime. arXiv:2106.04652 (2021)
26. Kobayashi, R., Giga, Y.: Equations with singular diffusivity. J. Stat. Phys. **95**(5–6), 1187–1220 (1999)
27. Kohn, R.V., Versieux, H.M.: Numerical analysis of a steepest-descent pde model for surface relaxation below the roughening temperature. SIAM J. Numer. Anal. **48**(5), 1781–1800 (2010)
28. Krishnamachari, B., McLean, J., Cooper, B., Sethna, J.: Gibbs–Thomson formula for small island sizes: corrections for high vapor densities. Phys. Rev. B **54**(12), 8899 (1996)
29. Krug, J., Dobbs, H., Majaniemi, S.: Adatom mobility for the solid-on-solid model. Zeitschrift für Physik B Condens. Matter **97**(2), 281–291 (1995)
30. Li, W., Lu, J., Wang, L.: Fisher information regularization schemes for Wasserstein gradient flows. J. Comput. Phys. 109449 (2020)
31. Liero, M., Mielke, A.: Gradient structures and geodesic convexity for reaction–diffusion systems. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. **371**(2005), 20120346 (2013)
32. Lisini, S., Matthes, D., Savaré, G.: Cahn–Hilliard and thin film equations with nonlinear mobility as gradient flows in weighted-Wasserstein metrics. J. Differ. Equ. **253**(2), 814–850 (2012)
33. Liu, J.-G., Lu, J., Margetis, D., Marzuola, J.L.: Asymmetry in crystal facet dynamics of homoepitaxy by a continuum model. Physica D **393**, 54–67 (2019)
34. Liu, J.-G., Strain, R.M.: Global stability for solutions to the exponential pde describing epitaxial growth. Interfaces Free Bound. **21**, 51–86 (2019)
35. Liu, J.-G., Xu, X.: Existence theorems for a multidimensional crystal surface model. SIAM J. Math. Anal. **48**(6), 3667–3687 (2016)
36. Liu, J.-G., Xu, X.: Analytical validation of a continuum model for the evolution of a crystal surface in multiple space dimensions. SIAM J. Math. Anal. **49**(3), 2220–2245 (2017)
37. Margetis, D., Kohn, R.V.: Continuum relaxation of interacting steps on crystal surfaces in 2+1 dimensions. Multiscale Model. Simul. **5**(3), 729–758 (2006)
38. Marzuola, J.L., Weare, J.: Relaxation of a family of broken-bond crystal-surface models. Phys. Rev. E **88**(3), 032403 (2013)
39. Najafabadi, R., Srolovitz, D.J.: Elastic step interactions on vicinal surfaces of fcc metals (1994)
40. Nessyahu, H., Tadmor, E.: Non-oscillatory central differencing for hyperbolic conservation laws. J. Comput. Phys. **87**, 408–463 (1990)
41. Odisharia, I.V.: Simulation and Analysis of the Relaxation of a Crystalline Surface. New York University, New York (2006)
42. Otto, F.: The geometry of dissipative evolution equations: the porous medium equation. Commun. Partial Differ. Equ. **26**(1–2), 101–174 (2001)
43. Rowlinson, J.S., Widom, B.: Molecular Theory of Capillarity. Courier Corporation, New York (2013)
44. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**(1–4), 259–268 (1992)
45. Sandier, E., Serfaty, S.: Gamma-convergence of gradient flows with applications to ginzburg-landau. Commun. Pure Appl. Math. A J. Issued Courant Inst. Math. Sci. **57**(12), 1627–1672 (2004)
46. Serfaty, S.: Gamma-convergence of gradient flows on Hilbert and metric spaces and applications. Discrete Contin. Dyn. Syst. **31**(4), 1427–1451 (2011)
47. Shefi, R., Teboulle, M.: Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization. SIAM J. Optim. **24**(1), 269–297 (2014)
48. Shenoy, V., Freund, L.: A continuum description of the energetics and evolution of stepped surfaces in strained nanostructures. J. Mech. Phys. Solids **50**(9), 1817–1841 (2002)
49. Wang, J., Lucier, B.J.: Error bounds for finite-difference methods for Rudin–Osher–Fatemi image smoothing. SIAM J. Numer. Anal. **49**(2), 845–868 (2011)
50. Zangwill, A., Luse, C., Vvedensky, D., Wilby, M.: Equations of motion for epitaxial growth. Surf. Sci. **274**(2), L529–L534 (1992)
51. Ziemer, W.P.: Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation, vol. 120. Springer, Berlin (2012)

Authors and Affiliations

Katy Craig[1] · Jian-Guo Liu[2] · Jianfeng Lu[3] · Jeremy L. Marzuola[4] · Li Wang[5]

Katy Craig
kcraig@math.ucsb.edu

Jian-Guo Liu
jliu@math.duke.edu

Jianfeng Lu
jianfeng@math.duke.edu

Li Wang
wang8818@umn.edu

1 Department of Mathematics, University of California, Santa Barbara, USA

2 Department of Mathematics and Department of Physics, Duke University, Durham, NC 27708, USA

3 Department of Mathematics, Chemistry and Physics, Duke University, Durham, NC 27708, USA

4 Department of Mathematics, University of North Carolina, Chapel Hill, NC 27599, USA

5 School of Mathematics, University of Minnesota, Twin Cities, USA