

# A Proximal-Gradient Algorithm



## **Crystal surface evolution**

- Evolution of a crystal near a fixed crystallographic plane of symmetry
- $h(x,t) = height of crystal; facets on crystal = { x: <math>\nabla h(x,t) = 0$  }
- [Marzuola, Weare '13]: continuum limit of kinetic Monte Carlo models  $E(h) = \frac{1}{p} \int |\nabla h|^p, \ p \ge 1 \qquad \partial_t h = \Delta e^{-\Delta_p h}$
- p=2, existence, uniqueness [Liu, Xu '16-'17, Xu '18, Ambrose '19,...]
- p=1, numerics via microscopic SOS system [Marzuola, Weare '13] finite difference method [Liu, Lu, Margetis, Marzuola '17]

## Our goal:

leverage (very formal) gradient flow structure to design new numerical method

crystal growth PDE

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

"gradient flow"  

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

$$M(h) = e^{-\Delta_1 h}, \quad \mathcal{E}(h) = \|\nabla h\|_1$$

$$\frac{d}{dt} h(t) = -\nabla_M \mathcal{E}(h(t))$$

$$\nabla_M \mathcal{E}(h) = -\nabla \cdot \left( M(h) \nabla \frac{\partial \mathcal{E}}{\partial h} \right)$$

### **Gradient flows**

- h(t) evolves in the direction of steepest descent of  $\mathcal{E}$ , with respect to M
- $\nabla_{M}$  is induced by the underlying metric structure

## Gradient flow

### prof. Mark. A. Peletier, PhD

Centre for Analysis, Scientific Computing, and Applications Department of Mathematics and Computer Science Institute for Complex Molecular Systems

> TUe Technische Universiteit Eindhoven University of Technology Where innovation starts

 $\frac{d}{dt}h(t) = -\nabla_M \mathcal{E}(h(t))$ 

### **Gradient flows**

- h(t) evolves in the direction of steepest descent of  $\mathcal{E}$ , with respect to M
- $\nabla_{M}$  is induced by the underlying metric structure

## Gradient flow

### prof. Mark. A. Peletier, PhD

Centre for Analysis, Scientific Computing, and Applications Department of Mathematics and Computer Science Institute for Complex Molecular Systems

> TUe Technische Universiteit Eindhoven University of Technology Where innovation starts

 $\frac{d}{dt}h(t) = -\nabla_M \mathcal{E}(h(t))$ 

## Our goal:

leverage (very formal) gradient flow structure to design new numerical method

crystal growth PDE

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

"gradient flow"

$$\partial_t h + \nabla \cdot \left( M(h) \nabla \frac{\partial \mathcal{E}}{\partial h} \right) = 0$$
$$M(h) = e^{-\Delta_1 h}, \quad \mathcal{E}(h) = \|\nabla h\|_1$$

Examples:

- M(h) = 1,  $H^{-1}$  gradient flow
- M(h) = h,  $W_2$  gradient flow
- M(h) nonnegative, concave, weighted W<sub>2</sub> gradient flow
   [Carrillo, Lisini, Savaré, Slepčev '09, Dolbeault, Nazaret, Savaré '09, Lisini, Matthew, Savaré '19,...]
- $M(h) \in Lin(\mathbb{R}^{I \times d}, \mathbb{R}^{I \times d})$ , gradient system [Liero, Mielke '13]

Our mobility falls well outside existing, rigorous theory of GF structure.

## Weighted H<sup>-1</sup> GF perspective

[Otto '01, Giga, Giga '10, ...]

Suppose  $M(h) \in L^1(\mathbb{T}^d)$  is nonnegative. Define  $\Delta_h v = \nabla \cdot (M(h) \nabla v)$ 

- Weighted Hilbert space:  $\|v\|_{H^1_h}^2 = \int_{\mathbb{T}^d} M(h) |\nabla v|^2 = \int_{\mathbb{T}^d} v \Delta_h v$
- Dual space:  $\|\psi\|_{H_h^{-1}}^2 = -\int_{\mathbb{T}^d} \psi \Delta_h^{-1} \psi$
- $\bullet$  Subdifferential  $\partial_{H_h^{-1}} \mathcal{E}$  , gradient  $\nabla_{H_h^{-1}} \mathcal{E}$

To go from here to a well-defined gradient flow, need to overcome obstacles:

- time derivative of h(t) with respect to  $\|\cdot\|_{H^{-1}_{h(t)}}$
- M(h(t)) needs to remain integrable and nonnegative along the flow.

#### Then,

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

$$\partial_t h = -\nabla_{H_h^{-1}} \mathcal{E}(h)$$
$$M(h) = e^{-\Delta_1 h}, \quad \mathcal{E}(h) = \|\nabla h\|_1$$

#### "gradient flow"

$$\partial_t h = -\nabla_{H_h^{-1}} \mathcal{E}(h)$$
$$M(h) = e^{-\Delta_1 h}, \quad \mathcal{E}(h) = \|\nabla h\|_1$$

• how to interpret mobility?

$$\partial_t h = -\nabla_{H_h^{-1}} \mathcal{E}(h)$$
$$M(h) = e^{-\Delta_1 h}, \quad \mathcal{E}(h) = \|\nabla h\|_1$$

- how to interpret mobility?
- how to discretize in time?

$$\partial_t h = -\nabla_{H_h^{-1}} \mathcal{E}(h)$$
$$M(h) = e^{-\Delta_1 h}, \quad \mathcal{E}(h) = \|\nabla h\|_1$$

- how to interpret mobility?
- how to discretize in time?
- how to discretize in space?

$$\partial_t h = -\nabla_{H_h^{-1}} \mathcal{E}(h)$$
$$M(h) = e^{-\Delta_1 h}, \quad \mathcal{E}(h) = \|\nabla h\|_1$$

- how to interpret mobility?
- how to discretize in time?
- how to discretize in space?

#### "gradient flow"

$$\partial_t h = -\nabla_{H_h^{-1}} \mathcal{E}(h)$$
$$M(h) = e^{-\Delta_1 h}, \quad \mathcal{E}(h) = \|\nabla h\|_1$$

- how to interpret mobility?
- how to discretize in time?
- how to discretize in space?

- the mobility  $M(h) = e^{-\Delta_1 h}$  doesn't make sense, even for d=1

$$\Delta_1 h = \nabla \cdot \left(\frac{\nabla h}{|\nabla h|}\right) = \nabla \cdot \operatorname{sgn}(\nabla h)$$



#### "gradient flow"

$$\partial_t h = -\nabla_{H_h^{-1}} \mathcal{E}(h)$$
$$M(h) = e^{-\Delta_1 h}, \quad \mathcal{E}(h) = \|\nabla h\|_1$$

- how to interpret mobility?
- how to discretize in time?
- how to discretize in space?
- the mobility  $M(h) = e^{-\Delta_1 h}$  doesn't make sense, even for d=1
- previous work considered linearization of exponential [Giga, Giga '10,...]

$$e^x \approx 1 + x, \quad M(h) \approx 1 - \Delta_1 h$$

- we consider mollified mobility, which respects asymmetry of curvature

$$\varphi \in C_c^{\infty}(\mathbb{T}^d), \ \varphi \ge 0, \ \int_{\mathbb{T}^d} \varphi = 1, \ \varphi_{\epsilon}(x) := \varphi(x/\epsilon)/\epsilon^d,$$
$$M_{\epsilon}(h) := e^{-\varphi_{\epsilon} * \Delta_1 h}$$



- we consider mollified mobility, which respects asymmetry of curvature

$$\varphi \in C_c^{\infty}(\mathbb{T}^d), \ \varphi \ge 0, \ \int_{\mathbb{T}^d} \varphi = 1, \ \varphi_{\epsilon}(x) := \varphi(x/\epsilon)/\epsilon^d,$$
$$M_{\epsilon}(h) := e^{-\varphi_{\epsilon} * \Delta_1 h}$$

$$\partial_t h = -\nabla_{H_h^{-1}} \mathcal{E}(h)$$
$$M_{\epsilon}(h) = e^{-\varphi_{\epsilon} * \Delta_1 h}, \ \mathcal{E}(h) = \|\nabla h\|_1$$

- how to interpret mobility?
- how to discretize in time?
- how to discretize in space?

#### "gradient flow"

$$\partial_t h = -\nabla_{H_h^{-1}} \mathcal{E}(h)$$

$$M_{\epsilon}(h) = e^{-\varphi_{\epsilon} * \Delta_1 h}, \ \mathcal{E}(h) = \|\nabla h\|_1$$

- how to interpret mobility?
- how to discretize in time?
- how to discretize in space?
- we consider a semi-implicit scheme (c.f. [Murphy, Walkington '19] for PME)

$$h^{n+1} \in \arg\min_{h} \mathcal{E}(h) + \frac{1}{2\tau} \|h - h^{n}\|_{H^{-1}_{h^{n}}}^{2}.$$
$$\frac{h^{n+1} - h^{n}}{\tau} = -\nabla \cdot \left(M(h^{n})\nabla \frac{\partial \mathcal{E}}{\partial h^{n+1}}\right)$$

- for the TV energy E, if  $h^n \in D(E)$  and  $M(h^n)$  is integrable and nonnegative, there exists a unique solution  $h^{n+1}$  to our semi-implicit scheme

#### "gradient flow"

$$\partial_t h = -\nabla_{H_h^{-1}} \mathcal{E}(h)$$

$$M_{\epsilon}(h) = e^{-\varphi_{\epsilon} * \Delta_1 h}, \ \mathcal{E}(h) = \|\nabla h\|_1$$

- how to interpret mobility?
- how to discretize in time?
- how to discretize in space?
- we consider a semi-implicit scheme (c.f. [Murphy, Walkington '19] for PME)

$$h^{n+1} \in \underset{h}{\operatorname{arg\,min}} \mathcal{E}(h) + \frac{1}{2\tau} \|h - h^n\|_{H_{h^n}}^2.$$
  
$$\frac{h^{n+1} - h^n}{\tau} = -\nabla \cdot \left( M(h^n) \nabla \frac{\partial \mathcal{E}}{\partial h^{n+1}} \right) \qquad \frac{\partial E}{\partial h} = -\Delta_1 h$$

 for the TV energy E, if h<sup>n</sup> ∈ D(E) and M(h<sup>n</sup>) is integrable and nonnegative, there exists a unique solution h<sup>n+1</sup> to our semi-implicit scheme

$$\partial_t h = -\nabla_{H_h^{-1}} \mathcal{E}(h)$$
$$M_{\epsilon}(h) = e^{-\varphi_{\epsilon} * \Delta_1 h}, \ \mathcal{E}(h) = \|\nabla h\|_1$$

- how to interpret mobility?
- how to discretize in time?
- how to discretize in space?

#### "gradient flow"

$$\partial_t h = -\nabla_{H_h^{-1}} \mathcal{E}(h)$$
$$M_{\epsilon}(h) = e^{-\varphi_{\epsilon} * \Delta_1 h}, \ \mathcal{E}(h) = \|\nabla h\|_1$$

- how to interpret mobility?
- how to discretize in time?
- how to discretize in space?

- we want to solve

$$h^{n+1} \in \underset{h}{\operatorname{arg\,min}} \underbrace{\mathcal{E}(h) + \frac{1}{2\tau} \|h - h^n\|_{H_{h^n}}^2}_{\|\nabla h\|_1 + \frac{1}{2\tau} \|h - h^n\|_{H_{h^n}}^2} = f(Kh) + g(h)$$

 $f: \mathcal{Z} \to \mathbb{R}$  convex,  $g: \mathcal{H} \to \mathbb{R}$  convex,  $K: \mathcal{H} \to \mathcal{Z}$  bounded, linear

- primal-dual algorithm! (c.f. [Laborde, Benamou, Carlier '16], [Carrillo, C., Wang, Wei '19],...)
- what is the role of the Hilbert spaces  $\mathcal{Z}, \mathcal{H}$  ?

#### "gradient flow"

$$\partial_t h = -\nabla_{H_h^{-1}} \mathcal{E}(h)$$
$$M_{\epsilon}(h) = e^{-\varphi_{\epsilon} * \Delta_1 h}, \ \mathcal{E}(h) = \|\nabla h\|_1$$

- how to interpret mobility?
- how to discretize in time?
- how to discretize in space?

#### - we want to solve

- primal-dual algoritmed (c.f. [Laborde, Benamou, Carlier '16], [Carrillo, C., Wang, Wei '19],...)

- what is the role of the Hilbert spaces  $\mathcal{Z},\mathcal{H}$  ?

## Which Hilbert space?

1) 
$$\mathcal{Z} = \mathcal{H} = L^2(\mathbb{R}^d)$$
  
2)  $\mathcal{Z} = L^2(\mathbb{R}^d)$ ,  $\mathcal{H} = \dot{H}^1(\mathbb{R}^d)$ 

- [Jacobs, Léger, Li, Osher '19]
- Consider gradient descent of a smooth, convex function w/ unique min  $u_*$ .

$$F(u) = f(Ku) + g(u)$$

• Convergence rate:  $F(u_n) \leq F(u^*) + 2L_{\mathcal{H}} \frac{\|u^* - u_0\|_{\mathcal{H}}^2}{n+4}$ 

## Which Hilbert space?

1) 
$$\mathcal{Z} = \mathcal{H} = L^2(\mathbb{R}^d)$$
  
2)  $\mathcal{Z} = L^2(\mathbb{R}^d)$ ,  $\mathcal{H} = \dot{H}^1(\mathbb{R}^d)$ 

- [Jacobs, Léger, Li, Osher '19]
- Consider gradient descent of a smooth, convex function w/ unique min  $u_*$ .

$$F(u) = f(Ku) + g(u)$$

• Convergence rate:  $F(u_n) \leq F(u^*) + 2L_{\mathcal{H}} \frac{\|u^* - u_0\|_{\mathcal{H}}^2}{n+4}$ 

• Nesterov:  

$$F(u_n) \leq \min_{u} \left[ F(u) + 4L_{\mathcal{H}} \frac{\|u - u_0\|_{\mathcal{H}}^2}{(n+2)^2} \right]$$

$$\leq \min_{\|u - u_0\|_{\mathcal{H}} \leq R} \left[ F(u) + 4L_{\mathcal{H}} \frac{\|u - u_0\|_{\mathcal{H}}^2}{(n+2)^2} \right]$$

$$\leq F(u_*) + 4L_{\mathcal{H}} \frac{R^2}{(n+2)^2} + \underbrace{\min_{\|u - u_0\|_{\mathcal{H}} \leq R} F(u) - F(u_*)}_{\delta_F(R)}$$
Thus, one can get around  $\|u_* - u_0\|_{\mathcal{H}} = +\infty$ , as long as  $\delta_F(R) \to 0$ .

## Which Hilbert space?

1) 
$$\mathcal{Z} = \mathcal{H} = L^2(\mathbb{R}^d)$$
  
2)  $\mathcal{Z} = L^2(\mathbb{R}^d)$ ,  $\mathcal{H} = \dot{H}^1(\mathbb{R}^d)$ 

- [Jacobs, Léger, Li, Osher '19]
- Analogous result holds for Chambolle-Pock's PDHG method (nonsmooth):

$$u_{n+1} = \underset{u \in \mathcal{H}}{\arg\min} \ g(u) + (u, K^T \bar{p}_n)_{\mathcal{H}} + \frac{1}{2\lambda} \|u - u_n\|_{\mathcal{H}}^2,$$
$$p_{n+1} = \underset{p \in \mathcal{Z}}{\arg\max} - f^*(p) + (K u_{n+1}, p)_{\mathcal{Z}} - \frac{1}{2\sigma} \|(p - p_n)\|_{\mathcal{Z}}^2,$$

$$\bar{p}_{n+1} = 2p_{n+1} - p_n.$$

for  $u^N = \frac{1}{N} \sum_{n=1}^N u_n$  and  $\lambda \sigma \| K^T K \|_{\mathcal{H}}^2 < 1$ , we have

$$F(u_N) \le F(u_*) + C\frac{R}{N} + \underbrace{\min_{\|u-u_0\|_{\mathcal{H}} \le R} F(u) - F(u_*)}_{\delta_F(R)}$$

### Numerical method for crystal evolution

• Outer time iteration:

$$h^{n+1} = \operatorname{argmin}_{h} \mathcal{E}(h) + \frac{1}{2\tau} \|h - h^{n}\|_{H^{-1}_{h^{n}}}^{2}$$

• Inner time iteration: for  $\lambda \sigma < 1$ ,

$$h^{(m+1)} = \left(\frac{\tau}{\lambda}\Delta_{h^n}\Delta + \mathrm{id}\right)^{-1} \left(\frac{\tau}{\lambda}\Delta_{h^n}\Delta h^{(m)} - \tau\Delta_{h^n}\nabla\cdot\phi^{(m)} + h^n\right)$$
$$\bar{h}^{(m+1)} = 2h^{(m+1)} - h^{(m)}$$
$$\phi^{(m+1)} = (\mathrm{id} + \sigma\partial F^*)^{-1}(\phi^{(m)} + \sigma\nabla\bar{h}^{(m+1)}) ,$$

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

discretize via finite difference method

### Numerical method for crystal evolution

• Outer time iteration:

$$h^{n+1} = \operatorname{argmin}_{h} \mathcal{E}(h) + \frac{1}{2\tau} \|h - h^{n}\|_{H^{-1}_{h^{n}}}^{2}$$

• Inner time iteration: for  $\lambda \sigma < 1$ ,

$$h^{(m+1)} = \left(\frac{\tau}{\lambda}\Delta_{h^n}\Delta + \mathrm{id}\right)^{-1} \left(\frac{\tau}{\lambda}\Delta_{h^n}\Delta h^{(m)} - \tau\Delta_{h^n}\nabla\cdot\phi^{(m)} + h^n\right)$$
  
$$\bar{h}^{(m+1)} = 2h^{(m+1)} - h^{(m)}$$
  
$$\phi^{(m+1)} = (\mathrm{id} + \sigma\partial F^*)^{-1}(\phi^{(m)} + \sigma\nabla\bar{h}^{(m+1)}),$$

discretize via finite difference method

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

#### Benefits:

- avoids inverting 1-Laplacian
- freedom to choose  $\lambda$  large helps with computation of h<sup>(m+1)</sup>

### **Convergence of PDHG**

**Theorem:** [CMLLW '20] Let d=1. Suppose the PDHG algorithm is initialized with

$$h^{(0)} = h^n, \phi^{(0)} = 0,$$

Then for all  $\delta$ >0, there exist  $\tilde{M}$ ,  $\lambda$ ,  $\sigma$ , so that

 $F(h^{(M)}) - F(h^{n+1}) \le \delta, \quad \forall M \ge \tilde{M}, \quad F(h) = \mathcal{E}(h) + \frac{1}{2\tau} \|h - h^n\|_{H^{-1}_{h^n}}^2$ 

### **Convergence of PDHG**

**Theorem:** [CMLLW '20] Let d=1. Suppose the PDHG algorithm is initialized with

$$h^{(0)} = h^n, \phi^{(0)} = 0,$$

Then for all  $\delta$ >0, there exist  $\tilde{M}$ ,  $\lambda$ ,  $\sigma$ , so that

$$F(h^{(M)}) - F(h^{n+1}) \le \delta, \quad \forall M \ge \tilde{M}, \quad F(h) = \mathcal{E}(h) + \frac{1}{2\tau} \|h - h^n\|_{H_{h^n}}^2$$

Remarks:

- Extends to general M(h) provided that M(h<sup>n</sup>) and 1/M(h<sup>n+1</sup>) integrable [c.f. Cancés, Gallouët, and Todeschi '19]
- We require d=1 to conclude  $||h^n||_{\infty} \leq \mathcal{E}(h^n) < +\infty$ . Higher integrability of 1/M(h<sup>n+1</sup>) would be required to weaken this assumption.
- If one has  $\nabla h^n \in BV$ , quantitative estimates:  $\tilde{M} \sim \delta^{-2}$ ,  $\lambda \sim \delta^{-1}$ ,  $\sigma \sim \delta$

- Key step: 
$$\delta_F(R) = \min_{\|h^n - h\|_{\dot{H}^1} \le R} F(h) - F(h^{n+1}) \xrightarrow{R \to +\infty} 0$$



**Observations:** 

- facet formation at local maxima
- pinning at local minima

Nx = 200, Nt = 10,  $\sigma$  = 0.0005,  $\lambda$  = 500,  $\epsilon$  =0.04





**Observations:** 

- facet formation at local maxima
- pinning at local minima

Nx = 200, Nt = 10,  $\sigma$  = 0.0005,  $\lambda$  = 500,  $\epsilon$  =0.04





**Observations:** 

- facet formation at local maxima
- pinning at local minima

Nx = 200, Nt = 10,  $\sigma$  = 0.0005,  $\lambda$  = 500,  $\epsilon$  =0.04





**Observations:** 

- facet formation at local maxima
- pinning at local minima

Nx = 200, Nt = 10,  $\sigma$  = 0.0005,  $\lambda$  = 500,  $\epsilon$  =0.04





Observations:

- facet formation at local maxima
- pinning at local minima

Nx = 200, Nt = 10,  $\sigma$  = 0.0005,  $\lambda$  = 500,  $\epsilon$  =0.04



#### Numerical Desultar an area desuces



### Numerical Results: convergence



Observations:

- Error vs Nx: slightly sublinear convergence (low spatial regularity)
- Error vs Nt: first order (semi-implicit Euler)
- Internal time steps vs Nt: importance of selecting correct Hilbert space

sinusoidal, (Nx = 200), (Nt = 10),  $\sigma$  = 0.0005,  $\lambda$  = 500,  $\epsilon$  =0.05, T = 10<sup>-4</sup>





Today: Convergence of PDHG, E TV energy, gen M(h), d=1



- Convergence as  $\tau \rightarrow 0$ ?
- Convergence as  $\tau, \varepsilon \rightarrow 0$ ?

Today: Convergence of PDHG, E TV energy, gen M(h), d=1

### crystal growth PDE

$$\partial_t h = \Delta e^{-\Delta_1 h} - \dots$$

- Appropriate notion of weak solution?
- Better time discretization/GF formulation to prove existence of wider class of weak solutions? numerics?

#### discrete time scheme

$$h^{n+1} = \operatorname{argmin}_{h} \mathcal{E}(h) + \frac{1}{2\tau} \|h - h^{n}\|_{H^{-1}_{h^{n}}}^{2}$$

- Convergence as  $\tau \rightarrow 0$ ?
- Convergence as  $\tau, \varepsilon \rightarrow 0$ ?

#### "gradient flow"

$$\partial_t h = -\nabla_{H_h^{-1}} \mathcal{E}(h)$$
$$M_{\epsilon}(h) = e^{-\varphi_{\epsilon} * \Delta_1 h}, \ \mathcal{E}(h) = \|\nabla h\|_1$$

#### **PDHG** scheme

$$h^{(m+1)} = \left(\frac{\tau}{\lambda}\Delta_{h^n}\Delta + \mathrm{id}\right)^{-1} \left(\frac{\tau}{\lambda}\Delta_{h^n}\Delta h^{(m)} - \tau\Delta_{h^n}\nabla\cdot\phi^{(m)} + h^n\right)$$
  
$$\bar{h}^{(m+1)} = 2h^{(m+1)} - h^{(m)}$$
  
$$\phi^{(m+1)} = (\mathrm{id} + \sigma\partial F^*)^{-1}(\phi^{(m)} + \sigma\nabla\bar{h}^{(m+1)}) ,$$

Today: Convergence of PDHG, E TV energy, gen M(h), d=1

