A THEORETICAL FRAMEWORK FOR A MOVING GRID FINITE ELEMENT METHOD *

XIANMIN XU †

Abstract. In this paper, we present a novel theoretical framework for a moving finite element method and prove the convergence of the method under some mild conditions. We introduce a regularized metric to the finite element space with free knots. This leads to a smooth Riemann manifold and a metric space when considering geodesic distance. We show that the moving finite element discretization for a nonlinear reaction-diffusion equation can be viewed as a curve of maximal slope in the discrete metric space. This inspires us to propose a JKO scheme and an explicit stabilized numerical scheme for the moving finite element method. We further prove the convergence of the moving finite element method in a general case using the gradient flow theory in metric spaces. Numerical examples are given to show that the discrete schemes work efficiently for both one dimensional and two dimensional problems.

Key words. Moving finite element method, gradient flows, metric spaces, moving meshes

1. Introduction.

Adaptive finite element methods have been widely used in solving partial differential equations (PDEs) in various scientific and engineering fields, particularly when dealing with singularities or sharp inner/boundary layers in the solution. There are mainly three types of adaptive methods: the h-type method, the p-type method, and the r-type method. The h-type method, widely acknowledged and extensively studied in the literature, involves adaptively refining or coarsening local meshes based on a posteriori error estimates [3, 1, 40]. The method has been proven to achieve optimal convergence [8, 37]. The p-type method, on the other hand, adjusts the local order of the finite element basis to increase or decrease it according to the solution errors. The p-type method can be combined with the h-type method, resulting in the hp method, where both the meshes and the polynomial orders are altered locally based on a posteriori error estimates [1]. The r-type method, often referred to as moving mesh (grid) finite element methods in the literature [10, 39], redistributes the mesh locations according to the solution and is typically utilized for time-dependent problems.

Compared to methods based on local mesh or polynomial order refinement, the theoretical analysis of the moving mesh methods is currently very limited [7, 20, 29, 25, 15], despite their widespread use in various problems (e.g., [32, 31, 22, 35, 26, 11, 38, 19, 5, 33, 43, 6], among many others). One important result in this field is the work done by N. Kopteva in [25], where the first order convergence of the moving mesh finite element method is proved for a stationary linear singularly perturbed equation in one dimension.

Recently, there exists much interest in developing Lagrangian type methods for gradient flow systems [28, 27, 14]. The methods have very close relations with the moving finite element method (MFEM) [32, 4, 21]. In particular, it is found that the moving finite element method for gradient flow systems can be derived naturally from the Onsager variational principle [42]. Based on the variational principle, it is possible to provide optimal error estimates for the stationary solution of the system, which improves the previous results [20, 23]. However, analyzing the dynamic solutions still poses a significant challenge, as it requires a more complex analysis of the discrete

*The work was partially supported by NSFC 11971469 and 12371415.

† LSEC, ICMSEC, NCMIS, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China, xmxu@lsec.cc.ac.cn.

gradient flows.

The main objective of this paper is to establish a theoretical framework to the moving finite element method for gradient flow systems. To accomplish this, we leverage the deep connections between the Onsager variational principle and gradient flows in metric spaces [30, 2]. We introduce a novel mathematical framework for the moving finite element method applied to dissipative systems, formulating it as a discrete gradient flow system in metric spaces. Based on the framework, we develop a Jordan-Kinderlehrer-Otto (JKO) scheme [24] and a new explicit stabilized scheme for the discrete gradient flow system. By utilizing the analysis tools from the theory of gradient flows in metric spaces [2, 34, 36], we are able to prove the convergence of the JKO scheme for a nonlinear gradient flow system.

More precisely, we consider a model problem which is a nonlinear reaction diffusion equation and can be viewed as a (continuous) gradient flow in a metric space with a L^2-distance. To discretize this problem, we first introduce a natural metric on the nonlinear approximation space for polynomials with free knots. We prove that this nonlinear space, equipped with the metric, forms a finite-dimensional Riemann manifold. However, there is a possibility of degeneracy in the manifold, so we add a regularized term to the metric. This regularization also allows us to define a distance measure by considering the geodesic distance within the manifold. We prove that the discrete manifold, equipped with the geodesic distance, results in a discrete metric space. Next, we study the approximation of the continuous gradient flow in the discrete metric spaces, which precisely corresponds to the moving finite element method. We then propose a fully discrete JKO scheme for the discrete gradient flows and prove its convergence to the solution of the continuous problem under mild assumptions. In the implementations, we develop a new explicit stabilized scheme by approximation to the optimization problem in the JKO scheme. Numerical experiments show that the method works well for both one dimensional and two dimensional problems.

The remaining sections of this paper are organized as follows. In Section 2, we briefly introduce some definitions in gradient flow theory in metric spaces. Section 3 introduces the continuous convection-diffusion equation and reformulates it as a gradient flow in a metric space in the L^2-distance. To transition to the discrete setting, Section 4 defines the discrete metric space for the finite element functions with free knots. Subsequently, in Section 5, we present the approximation of the convection-diffusion equation in the discrete metric space while also illustrate the main theoretical results. Section 6 is dedicated to presenting the proof of the existence and convergence results. In section 7, we present an explicit stabilized scheme and some numerical examples. In the final section, we present some concluding remarks.

2. Preliminary: Gradient flows in metric spaces.

We recall some definitions on gradient flows in metric spaces in [2]. Let $(\mathscr{S}, \mathsf{d})$ be a given complete metric space equipped with the distance d. We first introduce the definition of absolutely continuous curves in $(\mathscr{S}, \mathsf{d})$.

DEFINITION 2.1 (absolutely continuous curve). *Let (a, b) be an interval of $\mathbb{R}$. We say a curve $v : (a, b) \mapsto \mathscr{S}$ is a p-absolutely continuous curve or belongs to $AC^p(a, b; \mathscr{S})$ for $p \geq 1$, if there exists a function $m \in L^p(a, b)$, such that*

$$\mathsf{d}(v(s), v(t)) \leq \int_s^t m(r)dr, \qquad \forall a < s \leq t < b. \tag{2.1}$$

In the case $p = 1$, v is called an absolutely continuous curve and the corresponding space is simply denoted as $AC(a, b; \mathscr{S})$.

For absolutely continuous curves, the metric derivative is defined as follows.

Definition 2.2 (metric derivative). *For any curve v in $AC^p(a,b;\mathscr{S})$ with $p \geq 1$, the limit*

$$|v'|(t) := \lim_{s\to t} \frac{\mathsf{d}(v(s), v(t))}{|s-t|}, \tag{2.2}$$

exists for a.e. $t \in (a,b)$ and is called the metric derivative of v. Moreover, $|v'|(t) \in L^p(a,b)$ is the smallest admissible function m in Definition 2.1.

Let $\mathcal{E} : \mathscr{S} \mapsto (-\infty, +\infty]$ be a functional defined on $\mathscr{S}$. Define the admissible set

$$D(\mathcal{E}) := \{v \in \mathscr{S} | \mathcal{E}(v) < +\infty\}.$$

Then the strong upper gradient of $\mathcal{E}$ can be defined as follows

Definition 2.3 (strong upper gradient). *A function $g : \mathscr{S} \mapsto [0, +\infty]$ is a strong upper gradient of $\mathcal{E}$ if for every curve $v \in AC[a,b;\mathscr{S}]$, the function $g \circ v(t) := g(v(t))$ is Borel and satisfies*

$$|\mathcal{E}(v(t)) - \mathcal{E}(v(s))| \leq \int_s^t g \circ v(r)|v'|(r)dr, \qquad \forall a < s \leq t < b. \tag{2.3}$$

In particular, if $g \circ v|v'| \in L^1(a,b)$, then $\mathcal{E}(v(t))$ is absolutely continuous and

$$|(\mathcal{E} \circ v)'|(t) \leq g \circ v(t)|v'|(t), \qquad a.e.\ t \in (a,b). \tag{2.4}$$

A natural candidate for the upper gradient is the local slope of the functional.

Definition 2.4 (local slope). *The local slope of $\mathcal{E}$ at $v \in D(\mathcal{E})$ is defined by*

$$|\partial \mathcal{E}|(v) := \limsup_{w \to v} \frac{(\mathcal{E}(v) - \mathcal{E}(w))^+}{\mathsf{d}(v,w)}, \tag{2.5}$$

where $f^+ = \max(f, 0)$ is the positive part of a function f.

Notice that the local slope is not a strong upper gradient in general and some extra conditions are needed [2].

The following definition is a generalization of the standard gradient flow in metric spaces.

Definition 2.5 (p-curve of maximal slope). *We say a locally absolutely continuous map $u : (a,b) \mapsto \mathscr{S}$ is a p-curve of maximal slope with respect to its upper gradient g, if $\mathcal{E} \circ u(t)$ is a.e. equal to a non-increasing map ψ, and*

$$\psi'(t) \leq -\frac{1}{p}|u'|(t) - \frac{1}{q}(g \circ u(t))^q, \qquad \text{for a.e. } t \in (a,b), \tag{2.6}$$

where $\frac{1}{p} + \frac{1}{q} = 1$. If $p = 2$, we call u a curve of maximal slope.

The existence of a curve of maximal slope for a functional is established in [2] by investigating the JKO (implicit Euler) scheme in time [24], under appropriate conditions. Building upon this analysis, [34] explores the approximation of the energy functional $\mathcal{E}$ using a similar approach. In the subsequent sections, we adopt this framework to investigate the convergence of a moving finite element method for reaction-diffusion equations with gradient flow structures.

3. The continuous problem. We consider the following reaction diffusion equation,

$$\partial_t u = \alpha \Delta u - f'(u), \qquad \text{in } \Omega, \tag{3.1}$$

$$u = 0, \qquad \text{on } \partial\Omega, \tag{3.2}$$

where $\alpha > 0$ is the diffusion coefficient and $f \geq 0$ is an energy density function. We assume that $\Omega \subset \mathbb{R}^2$ is a simple connected domain with smooth boundary, and $f(u)$ is a smooth function such that $f''(u) \geq \lambda_0$ for some $\lambda_0 \in \mathbb{R}$.

It is well known that the equation (3.1) can be viewed as a L^2 gradient flow for the energy functional

$$\mathcal{E}(u) = \int_\Omega \frac{\alpha}{2}|\nabla u|^2 + f(u) dx.$$

For later applications, we rewrite the energy as

$$\mathcal{E}(u) = \begin{cases} \int_\Omega \frac{\alpha}{2}|\nabla u|^2 + f(u)dx, & \text{if } u \in H_0^1(\Omega); \\ +\infty, & \text{otherwise.} \end{cases} \tag{3.3}$$

We will formulate the problem as a curve of maximal slope in a metric space with a L^2-distance by using the definitions in the previous section.

Consider the Lebesgue space $L^2(\Omega)$. Let $(u, v)_0 = \int_\Omega uv dx$ be the inner product of two functions u and v in $L^2(\Omega)$; and $\|u\|_0 = \left(\int (u(x))^2 dx\right)^{\frac{1}{2}}$ be the L^2 norm of u. Then the distance between the two functions u and v in $L^2(\Omega)$ can be defined as

$$\mathsf{d}(u, v) = \|u - v\|_0. \tag{3.4}$$

It is easy to see that $(L^2(\Omega), \mathsf{d})$ is a complete metric space.

For an absolutely continuous curve $u(t, \cdot) : (0, T) \to L^2(\Omega)$, the metric derivative is given as $|u'|$. When u is differentiable with respect to t, we denote the "time" partial derivative as $\partial_t u$. When $\partial_t u \in L^2(\Omega)$, the metric derivative of u is equivalent to the L^2 norm of the time derivative, i.e. $|u'| = \|\partial_t u\|_0$.

For the energy functional $\mathcal{E}$ defined in (3.3), the local slope at $u(t_1)$ is defined as

$$|\partial \mathcal{E}(u)| = \limsup_{v \to u} \frac{|\mathcal{E}(v) - \mathcal{E}(u)|}{\mathsf{d}(u, v)}.$$

If $u \in H^2(\Omega)$, the local slope can be computed analytically,

$$|\partial \mathcal{E}(u)| = \| - \alpha \Delta u + f'(u)\|_0.$$

In this case, a curve of maximal slope is an absolutely continuous curve $u(t) \in AC(L^2(\Omega))$, satisfying

$$\frac{d}{dt}\mathcal{E}(u(t)) \leq -\frac{1}{2}|u'|^2(t) - \frac{1}{2}|\partial \mathcal{E}|^2(u(t)). \tag{3.5}$$

One can easily show that the curve of maximal slope is a solution of the partial differential equation (3.1). Actually, by the equation (3.5), we can derive that

$$\begin{aligned} -\frac{1}{2}|u'|^2(t) - \frac{1}{2}|\partial \mathcal{E}|^2(u(t)) &\geq \frac{d}{dt}\mathcal{E}(u(t)) = (-\alpha \Delta u + f'(u), \dot{u})_0 \\ &\geq -\| - \alpha \Delta u + f'(u)\|_0 \|\dot{u}\|_0 \\ &\geq -\frac{1}{2}|u'|^2(t) - \frac{1}{2}|\partial \mathcal{E}|^2(u(t)). \end{aligned}$$

Therefore, all the inequalities in above equation are equalities and we have

$$\partial_t u = \alpha \Delta u - f'(u),$$

which is exactly the equation (3.1).

4. The discrete metric space.

For simplicity in notations, we consider only the two-dimensional case.

4.1. The finite element space with free knots.

We introduce a nonlinear approximation space of piecewisely linear functions with free-knots as follows ([18]). Let $\hat{\Omega}$ be a reference domain, $\hat{\mathcal{T}}_N$ be a regular triangulation of $\hat{\Omega}$, and $\hat{\mathcal{N}}$ be the set of vertexes of $\hat{\mathcal{T}}_N$. We suppose that $\#\hat{\mathcal{N}} = N$, i.e. N is the total number of the vertexes in the triangulation $\hat{\mathcal{T}}_N$. Let $F : \hat{\mathrm{x}} \mapsto \mathrm{x}$ be a bijection from $\hat{\Omega}$ to Ω such that $\mathcal{T}_N := F\hat{\mathcal{T}}_N$ be a triangulation of Ω in the following sense:

- The vertex set $\mathcal{N}$ of $\mathcal{T}_N$ is given by

$$\mathcal{N} := \{\mathrm{x} : \mathrm{x} = F(\hat{\mathrm{x}}), \hat{\mathrm{x}} \in \hat{\mathcal{N}}\}.$$

- Each element in $\mathcal{T}_N$ is a triangle formed by three vertexes in $\mathcal{N}$. The elements do not overlap and have positive areas.
- $\mathcal{T}_N$ has the same topology as $\hat{\mathcal{T}}_N$.

If F satisfies all the above conditions, we say F is *admissible* and denote the admissible set

$$\mathcal{A}_{ad} := \{F : F \text{ is admissible}\}.$$

By the above definition, for $F \in \mathcal{A}_{ad}$, we have $F(\hat{x})$ is on the boundary of Ω whenever $\hat{x}$ is on the boundary of $\hat{\Omega}$. Notice that F is a nonlinear mapping from $\hat{\Omega}$ to Ω in general.

Let $\mathcal{T}_N$ be a triangulation of Ω. We define the standard finite element space on $\mathcal{T}_N$ as follows

$$V_h(\mathcal{T}_N) := \{v_h \in C(\Omega) : v_h \text{ is piecewisely linear on } \mathcal{T}_N, v_h = 0 \text{ on } \partial\Omega.\}$$

Then the nonlinear approximation space associated to $\hat{\mathcal{T}}_N$ is defined as follows,

$$V_N := \bigcup_{F \in \mathcal{A}_{ad}} V_h(F\hat{\mathcal{T}}_N).$$

Notice that V_N is not a linear space, since the summation of two functions in the space may not belong to V_N if they correspond to different partitions of Ω. Instead V_N forms a finite dimensional manifold. We can determine the dimension of the manifold. Denote by $\hat{\mathcal{N}}_b$ the set of nodes on the boundary of $\partial\hat{\Omega}$ and by $\hat{\mathcal{N}}_{in}$ the set of inner nodes. We assume that the number of vertexes in $\hat{\mathcal{N}}_b$ is M, i.e. $\#\hat{\mathcal{N}}_b = M < N$. Then we have $\#\hat{\mathcal{N}}_{in} = N - M$. It is easy to see that the number of freedoms in $V_h(\mathcal{T}_N)$ is $N - M$ for any fixed triangulation $\mathcal{T}_N$. Notice that the nodes in the triangulation for functions in V_N can also change positions. For simplicity, we assume that the nodes on the boundary $\partial\Omega$ are fixed. Each vertex in $\hat{\mathcal{N}}_{in}$ has two freedoms. Then the dimension of the manifold V_N is $3(N - M)$.

4.2. The regularized metric space. We introduce a natural metric for functions in V_N. For any $v_h \in V_N$, suppose the corresponding triangulation for v_h is $\mathcal{T}_N$, then it can be written as $v_h = \sum_{i=1}^{N-M} v_i\phi_i$, where v_i is the value of v_h on the vertex $\mathrm{x}_i \in \mathcal{N}_{in}$ and ϕ_i is the standard linear finite element basis function with respect to x_i. For any $\mathrm{x}_i \in \mathcal{N}$, let ω_i be the patch composed by all elements which includes x_i as a vertex. Let $\mathcal{N}(\omega_i)$ be the set of vortexes which is included in ω_i in the triangulation. We now consider the tangential space of the manifold V_N. Without loss of generality, suppose that $v_h(t)$ is a curve on the manifold given by

$$v_h(t) = \sum_{i=1}^{N-M} v_i(t)\phi_i(t, \mathrm{x}).$$

Notice that $\phi_i(t, \mathrm{x}) = \phi_i(\cdots, \mathrm{x}_j(t), \cdots; \mathrm{x})$ depends on all the vertexes $\mathrm{x}_j(t) \in \mathcal{N}(\omega_i)$. Denote by $(x_j(t), y_j(t))$ the coordinate of $\mathrm{x}_j(t)$. Then we have

$$\frac{dv_h}{dt} = \sum_{i=1}^{N-M} \Big(\frac{dv_i}{dt}\phi_i + \frac{dx_i}{dt} \sum_{\mathrm{x}_j \in \mathcal{N}(\omega_i)} v_j \partial_{x_i}\phi_j + \frac{dy_i}{dt} \sum_{\mathrm{x}_j \in \mathcal{N}(\omega_i)} v_j \partial_{y_i}\phi_j \Big).$$

Hereinafter, we denote by $\mathcal{N}_{in} = \{\mathrm{x}_i : 1 \le i \le N-M\}$ the set of inner vertexes in $\mathcal{T}_N$.

Denote by

$$\begin{aligned} \beta_i &:= \sum_{\mathrm{x}_j \in \mathcal{N}(\omega_i)} v_j \partial_{x_i}\phi_j, \qquad && 1 \le i \le N-M \\ \gamma_i &:= \sum_{\mathrm{x}_j \in \mathcal{N}(\omega_i)} v_j \partial_{y_i}\phi_j, \qquad && 1 \le i \le N-M. \end{aligned}$$

Then *the tangential space* of V_N corresponding to a function $v_h \in V_N$ can be given by

$$T_{v_h}V_N = \mathrm{span}\{\phi_i, \beta_i, \gamma_i : 1 \le i \le N-M\}. \tag{4.1}$$

We say V_N is *non-degenerate* at v_h whenever $T_{v_h}V_N$ is a $3(N-M)$ dimensional linear space. In this case, we say v_h is non-degenerate in V_N. When v_h is non-degenerate, the functions $\{\phi_i, \beta_i, \gamma_i : 1 \le i \le N-M\}$ forms a basis of $T_{v_h}V_N$.

In this case, we can introduce a metric on $T_{v_h}V_N$ as follows,

$$g(v_h) = \begin{pmatrix} A & B & C \\ B^T & D & E \\ C^T & E^T & F \end{pmatrix},$$

where

$$\begin{array}{ll} A \in \mathbb{R}^{(N-M)\times(N-M)}, & a_{ij} = \int_\Omega \phi_i\phi_j d\mathrm{x}; \\ B \in \mathbb{R}^{(N-M)\times(N-M)}, & b_{ij} = \int_\Omega \phi_i\beta_j d\mathrm{x}; \\ C \in \mathbb{R}^{(N-M)\times(N-M)}, & c_{ij} = \int_\Omega \phi_i\gamma_j d\mathrm{x}; \\ D \in \mathbb{R}^{(N-M)\times(N-M)}, & d_{ij} = \int_\Omega \beta_i\beta_j d\mathrm{x}; \\ E \in \mathbb{R}^{(N-M)\times(N-M)}, & e_{ij} = \int_\Omega \beta_i\gamma_j d\mathrm{x}; \\ F \in \mathbb{R}^{(N-M)\times(N-M)}, & f_{ij} = \int_\Omega \gamma_i\gamma_j d\mathrm{x}. \end{array}$$

When v_h is non-degenerate, we can easily see that $g(v_h)$ is a positive definite symmetric matrix.

Notice that there may exist degenerate functions in V_N. One trivial example is the zero function (when $v_i = 0$, for all $i = 1, \cdots, N-M$), for which we have $\beta_i = 0$ and $\gamma_j = 0$. In this case, we easily see that B, C, $\cdots$ and F in $g(v_h)$ are all zero matrices. There also exist other type of degenerate functions in V_N. For example, if v_h is a constant function in a patch ω_i, we easily have $\beta_i = \gamma_i = 0$ so that $g(v_h)$ will be degenerate.

When $v_h \in V_h$ is degenerate, $g(v_h)$ is a degenerate matrix. We may introduce a regularized metric on $T_{v_h}V_N$ as

$$g_\delta(v_h) = \begin{pmatrix} A & B & C \\ B^T & D+\delta I & E \\ C^T & E^T & F+\delta I \end{pmatrix}, \tag{4.2}$$

where $\delta > 0$, I is the unit matrix in $\mathbb{R}^{(N-M)\times(N-M)}$. With the above defined regularized metric, (V_N, g_δ) gives a smooth Riemann manifold, as stated in the following lemma.

PROPOSITION 4.1. *(V_N, g_δ) forms a smooth Riemann manifold.*

Proof. We need to prove that $g_\delta(v_h)$ is a Riemann metric on $T_{v_h}V_N$ for all $v_h \in V_N$. It is easy to check that $g_\delta(v_h)$ is semi-positive definite for all $v_h \in V_N$. We will only need to show it is non-degenerate as follows. Let $(\mathbf{a}^T, \mathbf{b}^T, \mathbf{c}^T)$ be a vector in $\mathbb{R}^{3(N-M)}$, such that $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^{N-M}$. Suppose that

$$(\mathbf{a}^T, \mathbf{b}^T, \mathbf{c}^T) g_\delta(v_h) \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \end{pmatrix} = 0.$$

This implies that

$$(\mathbf{a}^T, \mathbf{b}^T, \mathbf{c}^T) g(v_h) \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \end{pmatrix} + \delta(|\mathbf{b}|^2 + |\mathbf{c}|^2) = 0.$$

Notice that $g(v_h)$ is a semi-positive matrix. We have $|\mathbf{b}| = |\mathbf{c}| = 0$. Then we are led to

$$\mathbf{a}^T A \mathbf{a} = 0.$$

Since A is a mass matrix for standard linear finite element space on $\mathcal{T}_N$, we easily have $\mathbf{a} = \mathbf{0}$. ∎

Consider a curve in V_N given by

$$\Gamma(t) := \{v_h(t) = \sum_{i=1}^{N-M} v_i(t)\phi_i(t, \mathrm{x})\},$$

with differentiable coefficients in $t \in (a, b)$. The arc length of $\Gamma(t)$ is given by

$$L(\Gamma) = \int_a^b \left[(\dot{\mathbf{v}}^T, \dot{\mathbf{x}}^T, \dot{\mathbf{y}}^T) g_\delta(u_h) \begin{pmatrix} \dot{\mathbf{v}} \\ \dot{\mathbf{x}} \\ \dot{\mathbf{y}} \end{pmatrix} \right]^{\frac{1}{2}} ds, \tag{4.3}$$

where $\mathbf{v} = (v_1, \cdots, v_{N-M})^T$, $\mathbf{x} = (x_1, \cdots, x_{N-M})^T$ and $\mathbf{y} = (y_1, \cdots, y_{N-M})^T$. $\dot{\mathbf{v}}$ is the derivative of $\mathbf{v}$ with respect to t and similar definitions are for $\dot{\mathbf{x}}$ and $\dot{\mathbf{y}}$. Then, the geodesic distance between two points $v_h^{(1)}$ and $v_h^{(2)}$ in V_N can be defined as follows

$$\mathsf{d}_{N,\delta}(v_h^{(1)}, v_h^{(2)}) := \inf_{\Gamma} \Big\{ L(\Gamma) : \Gamma \ \text{ is a piecewise differentiable curve} \\ \text{connecting } v_h^{(1)} \text{ and } v_h^{(2)} \Big\}.$$

By the theory in Riemann geometry, $(V_N, \mathsf{d}_{N,\delta})$ gives a metric space, as shown in the following lemma.

Proposition 4.2. *$(V_N, \mathsf{d}_{N,\delta})$ is a metric space.*

Proof. We need only to prove $\mathsf{d}_{N,\delta}$ is a metric. By the definition of $\mathsf{d}_{N,\delta}$, we easily see the symmetry $\mathsf{d}_{N,\delta}(v_h^{(1)}, v_h^{(2)}) = \mathsf{d}_{N,\delta}(v_h^{(2)}, v_h^{(1)})$, and the inequality

$$\mathsf{d}_{N,\delta}(v_h^{(1)}, v_h^{(2)}) \le \mathsf{d}_{N,\delta}(v_h^{(1)}, v_h^{(3)}) + \mathsf{d}_{N,\delta}(v_h^{(3)}, v_h^{(2)}),$$

for all $v_h^{(1)}, v_h^{(2)}, v_h^{(3)} \in V_N$. Meanwhile, we also have $\mathsf{d}_{N,\delta}(v_h^{(1)}, v_h^{(2)}) \ge 0$.

Now suppose that $\mathsf{d}_{N,\delta}(v_h^{(1)}, v_h^{(2)}) = 0$, we need to prove that $v_h^{(1)} = v_h^{(2)}$. In this case we have $\inf_\Gamma L(\Gamma) = 0$ for all curves Γ connecting $v_h^{(1)}$ and $v_h^{(2)}$. If the infimum is obtained for a curve Γ_0, we have

$$L(\Gamma_0) = \int_a^b \left[(\dot{\mathbf{v}}^T, \dot{\mathbf{x}}^T, \dot{\mathbf{y}}^T) g_\delta(u_h) \begin{pmatrix} \dot{\mathbf{v}} \\ \dot{\mathbf{x}} \\ \dot{\mathbf{y}} \end{pmatrix} \right]^{\frac{1}{2}} dt = 0.$$

This leads to $(\dot{\mathbf{v}}^T, \dot{\mathbf{x}}^T, \dot{\mathbf{y}}^T) = \mathbf{0}$, a.e. $t \in (a,b)$. We thus have $v_h^{(1)} = v_h^{(2)}$. If the minimum is not obtained, we can prove the equality by taking limits of a minimizing sequence. ∎

Notice that the space V_N might not be complete, since the limit of a Cauchy sequence of the functions in the space might correspond to a triangulation which is not admissible. For example, suppose a sequence $v_h^{(k)}$ is equal to a constant in a large patch which includes at least one triangle T not intersecting with the boundary. Suppose the triangle shrinks to a point when k goes to infinity and the other vertexes do not change. We also assume that the value of the functions (i.e. $v_h^{(k)}(\mathrm{x}_i)$) on the vertexes does not change. We can easily show that $v_h^{(k)}$ is a Cauchy sequence in $(V_N, \mathsf{d}_{N,\delta})$ since $\mathsf{d}^2_{N,\delta}(v_h^{(k_1)}, v_h^{(k_2)}) = \delta \sum_{i=1}^3 |\mathrm{x}_{i,T}^{(k_1)} - \mathrm{x}_{i,T}^{(k_2)}|^2$. Here $\mathrm{x}_{i,T}^{(k_1)}$ and $\mathrm{x}_{i,T}^{(k_2)}$ (i=1,2,3) are the vertexes of the triangle T corresponding to $v_h^{(k_1)}$ and $v_h^{(k_2)}$, respectively. However, the limit of the meshes corresponds to a triangulation of Ω with different topology, which is not admissible by the definition of $\mathcal{A}_{ad}$. The incompleteness of the discrete space may cause troubles to a numerical method. The degeneracy of the triangulation can be avoided by adding a penalty term to the discrete energy in next section.

5. The discrete problem and the convergence results.

5.1. The discrete gradient flow. In the metric space $(V_N, \mathsf{d}_{N,\delta})$, the discrete energy corresponding to the energy $\mathcal{E}$ in (3.3) is defined as

$$\mathcal{E}_N^{\tilde{\delta}}(u_h) := \int_\Omega \frac{\alpha}{2} |\nabla u_h|^2 + f(u_h) d\mathrm{x} + \tilde{\delta} \int_{\hat{\Omega}} W(\nabla_{\hat{\mathrm{x}}} F(\hat{\mathrm{x}})) d\hat{\mathrm{x}} \tag{5.1}$$

where $\tilde{\delta}$ is a positive parameter and the last term is a penalty term to make sure no degenerate triangles in real simulations. In this paper, we assume $\hat{\Omega} = \Omega$ and $\hat{\mathcal{T}}_h$ is a quasi-uniform partition of Ω. We suppose that $W(\cdot) \geq 0$ is a given function such that

$$W(I) = 0; \quad W(\nabla_{\hat{\mathrm{x}}} F(\hat{\mathrm{x}})) \to \infty \text{ as } \det(\nabla_{\hat{\mathrm{x}}} F(\hat{\mathrm{x}})) \to 0^+. \tag{5.2}$$

The local upper gradient of $\mathcal{E}_N^{\tilde{\delta}}$ in $(V_N, \mathsf{d}_{N,\delta})$ is calculated as

$$|\partial \mathcal{E}_N^{\tilde{\delta}}(u_h)| = \left[(\mathbf{f}_1^T, \mathbf{f}_2^T, \mathbf{f}_3^T) g_\delta(u_h)^{-1} \begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \mathbf{f}_3 \end{pmatrix} \right]^{\frac{1}{2}} \tag{5.3}$$

where

$$f_{1,i} = \frac{\partial \mathcal{E}_N^{\tilde{\delta}}}{\partial u_i}, \quad f_{2,i} = \frac{\partial \mathcal{E}_N^{\tilde{\delta}}}{\partial x_i}, \quad f_{3,i} = \frac{\partial \mathcal{E}_N^{\tilde{\delta}}}{\partial y_i}, \qquad i = 1, \cdots, N - M.$$

Notice that the metric derivative of an absolutely continuous curve $\Gamma = \{u_h(t)\}$ in $(V_N, \mathsf{d}_{N,\delta})$ is given by

$$|\Gamma'(t)| = \left[(\dot{\mathbf{u}}^T, \dot{\mathbf{x}}^T, \dot{\mathbf{y}}^T) g_\delta(u_h) \begin{pmatrix} \dot{\mathbf{u}} \\ \dot{\mathbf{x}} \\ \dot{\mathbf{y}} \end{pmatrix} \right]^{\frac{1}{2}}. \tag{5.4}$$

By definition and direct calculations, the gradient flow in metric space, i.e. the curve of maximal slope is given by

$$g_\delta(u_h) \begin{pmatrix} \dot{\mathbf{u}} \\ \dot{\mathbf{x}} \\ \dot{\mathbf{y}} \end{pmatrix} = - \begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \mathbf{f}_3 \end{pmatrix}. \tag{5.5}$$

This is exactly the semi-discrete formula derived in the moving finite element scheme [42], which originates from the seminal work in [32, 31].

We consider the problem (5.5) in an interval $[0, T]$. In addition, we assume that $\hat{\Omega} = \Omega$ and the mapping $F(\hat{x}) = \hat{x}$ at initial time $t = 0$. With these notations, we can derive a fully discrete problem by the so-called Jordan-Kinderlehrer-Otto(JKO) scheme [24]. We partition the time interval $[0, T]$ by

$$0 = t_0 < t_1 < \cdots < t_K = T.$$

Let $\Delta t = T/K$ be the time step size. A proximal functional related to $\mathcal{E}_N^{\tilde{\delta}}$ is defined as

$$\Phi_N^{\delta,\tilde{\delta}}(\Delta t, u_h; v_h) := \frac{\mathsf{d}_{N,\delta}^2(u_h, v_h)}{2\Delta t} + \mathcal{E}_N^{\tilde{\delta}}(v_h), \tag{5.6}$$

for all $u_h, v_h \in V_N$. Then the JKO scheme of the semi-discrete problem (5.5) is defined as follows. For a given discrete solution u_h^{n-1} at the time t_{n-1}, the solution at t_n is computed by

$$u_h^n \in \operatorname{argmin}_{v_h} \Phi_N^{\delta,\tilde{\delta}}(\Delta t, u_h^{n-1}; v_h) \tag{5.7}$$

5.2. The theoretical results. We define the piecewisely constant functions in time

$$\bar{u}_h(t) = \begin{cases} u_h^0, & t = 0, \\ u_h^n, & t \in (t^{n-1}, t^n], \end{cases} \tag{5.8}$$

which is an approximated solution to the model problem (3.1). We also define the discrete metric gradient for the discrete solution as

$$|u_h'|(t) := \frac{\mathsf{d}_{N,\delta}(u_h^n, u_h^{n-1})}{t_n - t_{n-1}}, \quad t \in (t^n, t^{n-1}).$$

Then, we have the following existence and convergence theorems.

THEOREM 5.1 (Existence of the discrete solution). *For an initial function $u_h^0 \in V_N$ and a partition of the time interval $(0,T)$ with time step Δt, there exists at least one discrete solution $\bar{u}_h(t)$ defined as in (5.8).*

THEOREM 5.2 (Convergence). *Suppose that $u_h^0 \xrightarrow{L^2} u_0$, $\mathcal{E}_N^{\tilde{\delta}}(u_h^0) \to \mathcal{E}(u_0)$, as $N \to \infty$; and $u_0 \in D(\mathcal{E})$. We also assume that the parameters $\Delta t = o(1)$, $\delta = o(N^{-1})$ and $\tilde{\delta} = o(1)$. Let $\bar{u}_h(t)$ be a sequence of discrete solutions of the discrete problem (5.8). Then there exists a subsequence, still denoted as $\bar{u}_h(t)$ and a curve $u(t)$ belongs to $\mathrm{AC}^2_{\mathrm{loc}}([0,\infty), L^2(\Omega))$, such that $u(0) = u_0$,*

$$\lim_{N\to\infty} \|\bar{u}_h(t) - u(t)\|_{L^2} = 0, \qquad \forall t \in (0,T],$$

$$\lim_{N\to\infty} \mathcal{E}_N^{\tilde{\delta}}(\bar{u}_h(t)) = \mathcal{E}(u(t)), \qquad \forall t \in (0,T],$$

$$\lim_{N\to\infty} |\partial\mathcal{E}_N^{\tilde{\delta}}|(\bar{u}_h) = |\partial\mathcal{E}|(u), \quad \text{in } L^p_{\mathrm{loc}}[0,T],$$

$$\lim_{N\to\infty} |u_h'| = |u'|, \qquad \text{in } L^p_{\mathrm{loc}}[0,T].$$

Furthermore, $u(t)$ is a curve of maximal slope for $\mathcal{E}$ with respect to $|\partial\mathcal{E}|$, which is a strong upper gradient for $\mathcal{E}$. We also have the following energy identity

$$\mathcal{E}(u(t)) = \mathcal{E}(u(0)) - \frac{1}{2}\int_0^t |u'|^2(s)ds - \frac{1}{2}\int_0^t |\partial\mathcal{E}|^2(u(s))ds, \quad \forall t \in (0,T]$$

The two theorems will be proved in next section.

6. Proof of the main results.

In this section, we present the proof of the main results. The proof follows the approach in [2] and also in [34].

6.1. Moreau-Yosida approximation.

We first introduce the Moreau-Yosida approximation of the discrete energy $\mathcal{E}_N^{\tilde{\delta}}$ and its properties. Let $s > 0$, the Moreau-Yosida approximation of $\mathcal{E}_N^{\tilde{\delta}}$ at $v_h \in V_N$ is defined as

$$\mathcal{E}_{N,s}(v_h) := \inf_{w_h \in V_N} \Phi_N^{\delta,\tilde{\delta}}(s, v_h; w_h) = \inf_{w_h \in V_N} \frac{\mathsf{d}^2_{N,\delta}(v_h, w_h)}{2s} + \mathcal{E}_N^{\tilde{\delta}}(w_h). \tag{6.1}$$

The set of minimizes is denoted as

$$J_{N,s}[v_h] := \mathrm{argmin}_{w_h \in V_N} \Phi_N^{\delta,\tilde{\delta}}(s, v_h; w_h).$$

With the above notations, it is easy to see that the discrete solution $u_h^n \in J_{N,\Delta t}[u_h^{n-1}]$.

The following theorem show the existence of the Moreau-Yosida approximation, which covers the results in Theorem 5.1.

PROPOSITION 6.1 (Existence of the Moreau-Yosida approximation). *For any $s > 0$ and $v_h \in V_N$, we have*

$$J_{h,s}[v_h] \neq \emptyset. \tag{6.2}$$

In particular, for every choice of $u_h^0 \in V_N$ and a partition of the time with step size Δt, there exists at least one discrete solution $\bar{u}_h(t)$ defined as in (5.8).

Proof. By the assumption that $f \geq 0$ in Section 3, we have $\Phi_N^{\delta,\tilde{\delta}} \geq 0$. For given $v_h \in V_N$, we can find a minimizing sequence $w_h^{(k)} \in V_N$ such that

$$\lim_{k\to\infty} \Phi_N^{\delta,\tilde{\delta}}(s, v_h; w_h^{(k)}) = \inf_{w_h \in V_N} \Phi_N^{\delta,\tilde{\delta}}(s, v_h; w_h) \geq 0.$$

Notice that

$$\inf_{w_h \in V_N} \Phi_N^{\delta,\tilde{\delta}}(s, v_h; w_h) \leq \Phi_N^{\delta,\tilde{\delta}}(s, v_h; v_h) = \mathcal{E}_N^{\tilde{\delta}}(v_h) =: C_0 < \infty.$$

Then there exists a positive number $K \in \mathbb{N}^+$, such that

$$\frac{\mathsf{d}^2_{N,\delta}(v_h, w_h^{(k)})}{2s} + \mathcal{E}_N^{\tilde{\delta}}(w_h^{(k)}) < 2C_0, \qquad \text{when } k > K.$$

We then have

$$\mathsf{d}^2_{N,\delta}(v_h, w_h^{(k)}) < 2C_0 s, \quad \text{and} \quad \mathcal{E}_N^{\tilde{\delta}}(w_h^{(k)}) < 2C_0, \qquad \text{when } k > K.$$

Notice that $w_h^{(k)} = \sum w_i^{(k)} \phi_i(\mathrm{x})$ has finite dimensions. We can find a subsequence, which is still denoted as $w_h^{(k)}$ without loss of generality, such that $\mathbf{w}^{(k)}$, $\mathbf{x}^{(k)}$ and $\mathbf{y}^{(k)}$ converge in Eulerian distance. Here $\mathbf{w}^{(k)} \in \mathbb{R}^{N-M}$ is the vector of $w_h^{(k)}(\mathrm{x}_i)$ for $\mathrm{x}_i \in \mathcal{N}_{in}$, $\mathbf{x}^{(k)} \in \mathbb{R}^{N-M}$ and $\mathbf{y}^{(k)} \in \mathbb{R}^{N-M}$ are respectively the coordinates of inner vertexes in the triangulation corresponding to $w_h^{(k)}$. By the definition of $\mathcal{E}_N^{\tilde{\delta}}$ and the property (5.2), we know the area of each element in the triangulation corresponding to $w_h^{(k)}$ has a lower bound independent of k. Therefore, the limit of $\mathbf{x}^{(k)}$and $\mathbf{y}^{(k)}$ will correspond to an admissible triangulation of Ω. We then have $\tilde{w}_h = \lim_{k\to\infty} w_h^{(k)}$ is still in V_N.

By the continuity of $\Phi_N^{\delta,\tilde{\delta}}$, we also have

$$\lim_{k\to\infty} \Phi_N^{\delta,\tilde{\delta}}(s, v_h; w_h^{(k)}) = \Phi_N^{\delta,\tilde{\delta}}(s, v_h; \tilde{w}_h) = \inf\nolimits_{w_h \in V_N} \Phi_h(s, v_h; w_h).$$

This ends the proof of the theorem. ∎

By the above theorem, we know that the Moreau-Yosida approximation of $\mathcal{E}_N^{\tilde{\delta}}$ is well defined for any $s > 0$ and $v_h \in V_N$. This means $J_{N,s}[v_h]$ is not empty. However, the minimizes might not be unique. Thus we can define

$$d^+_{N,s}(v_h) := \sup_{w_h \in J_{N,s}[v_h]} \mathsf{d}_{N,\delta}(v_h, w_h), \quad d^-_{N,s}(v_h) := \inf_{w_h \in J_{N,s}[v_h]} \mathsf{d}_{N,\delta}(v_h, w_h).$$

The following properties of the Moreau-Yosida approximation are direct applications of some known results in literature (e.g. Section 3.1 in [2]).

Lemma 6.1 (Properties of the Moreau-Yosida approximation). *For the Moreau-Yosida approximation of $\mathcal{E}_N^{\tilde{\delta}}$ defined above, the following properties hold.*

(1). The map $(s, v_h) \to \mathcal{E}_{N,s}(v_h)$ is constiuous.

(2). If $0 < s_0 < s_1$ and $v_h^{(i)} \in J_{N,s_i}[v_h]$, we have

$$\mathcal{E}_N^{\tilde{\delta}}(v_h) \geq \mathcal{E}_{N,s_0}(v_h) \geq \mathcal{E}_{N,s_1}(v_h),\ \mathsf{d}_{N,\delta}(v_h^{(1)}, v_h) \leq \mathsf{d}_{N,\delta}(v_h^{(0)}, v_h),$$
$$\mathcal{E}_N^{\tilde{\delta}}(v_h) \geq \mathcal{E}_N^{\tilde{\delta}}(v_h^{(0)}) \geq \mathcal{E}_N^{\tilde{\delta}}(v_h^{(1)}),\ d^+_{N,s_0}(v_h) \leq d^-_{N,s_1}(v_h) \leq d^+_{N,s_1}(v_h).$$

(3). It holds that

$$\lim_{s\downarrow 0} \mathcal{E}_{N,s}(v_h) = \lim_{s\downarrow 0} \inf_{w_h \in J_{N,s}[v_h]} \mathcal{E}_N^{\tilde{\delta}}(v_h) = \mathcal{E}_N^{\tilde{\delta}}(v_h).$$

(4). There exists at most countable set $\mathcal{N}_{v_h} \subset \mathbb{R}^+$, such that

$$d^+_{N,s}(v_h) = d^-_{N,s}(v_h), \quad \forall s \in \mathbb{R}^+ \setminus \mathcal{N}_{v_h}.$$

and $\lim_{s\downarrow 0} d^+_{N,s}(v_h) = 0$.

(5). If $v_{h,s} \in J_{N,s}[v_h]$, then we have

$$|\partial \mathcal{E}_N^{\tilde{\delta}}|(v_{h,s}) \leq \frac{\mathsf{d}_{N,\delta}(v_{h,s}, v_h)}{s}.$$

(6). For every $v_h \in V_N$, the map $s \mapsto \mathcal{E}_{N,s}(v_h)$ is Lipschitz and satisfies

$$\frac{d\mathcal{E}_{N,s}(v_h)}{ds} = -\frac{(d^\pm_{N,s}(v_h))^2}{2s^2}, \quad \forall s \in \mathbb{R}^+ \setminus \mathcal{N}_{v_h},$$

and

$$\frac{\mathsf{d}^2_{N,\delta}(v_{h,s}, v_h)}{2s} + \int_0^s \frac{(d^\pm_{N,r}(v_h))^2}{2r^2} dr = \mathcal{E}_N^{\tilde{\delta}}(v_h) - \mathcal{E}_N^{\tilde{\delta}}(v_{h,s}).$$

(7). There exists a sequence $s_n \downarrow 0$ such that

$$|\partial \mathcal{E}_N^{\tilde{\delta}}|^2(v_h) = \lim_{k\to\infty} \frac{\mathsf{d}^2_{N,\delta}(v_{h,s_k}, v_h)}{s_k^2} = \lim_{k\to\infty} \frac{\mathcal{E}_N^{\tilde{\delta}}(v_h) - \mathcal{E}_N^{\tilde{\delta}}(v_{h,s_k})}{s_k} \geq \liminf_{s\downarrow 0} |\partial \mathcal{E}_N^{\tilde{\delta}}|^2(v_{h,s}).$$

6.2. A priori error estimate. To prove convergence of the discrete gradient flow, it is necessary to introduce the De-Giorgi interpolation defined below.

Definition 6.1 (De Giorgi variational interpolation). *Let $\{u_h^n\}$ be a solution of the variational scheme* (5.7)*, we denote by $\tilde{u}_h : [0, \infty) \to V_N$ an interpolant of the discrete values satisfying*

$$\tilde{u}_h(t) = \tilde{u}_h(t_{n-1} + s) \in J_{N,s}[u_h^{n-1}], \quad \text{if } t = t_{n-1} + s \in (t_{n-1}, t_n). \tag{6.3}$$

Introduce a notation

$$G_N(t) := \frac{d^+_{N,s}(u_h^{n-1})}{s} \geq \frac{\mathsf{d}_{N,\delta}(\tilde{u}_h(t), u_h^{n-1})}{t - t_{n-1}}, \quad t = t_{n-1} + s \in (t_{n-1}, t_n].$$

By the property (5) of the Moreau-Yosida approximation in Lemma 6.1, we have

$$|\partial \mathcal{E}_h^{\tilde{\delta}}|(\tilde{u}_h) \leq G_N(t), \quad \forall t \in (0, T].$$

The following lemma gives some a priori estimates of the fully discrete problem, which are useful in the proof the convergence theorem.

LEMMA 6.2 (A priori estimates). *For each couple of integers $1 \le i \le j \le K$, we have*

$$\frac{1}{2}\int_{t_i}^{t_j} |u_h'|^2(t)dt + \frac{1}{2}\int_{t_i}^{t_j} G_N^2(t)dt + \mathcal{E}_N^{\tilde{\delta}}(u_h^j) = \mathcal{E}_N^{\tilde{\delta}}(u_h^i). \tag{6.4}$$

Moreover for any $1 \le n \le K$, there exists a constant C independent of N and K such that

$$\sum_{j=1}^{n} \frac{\mathsf{d}_{N,\delta}^2(u_h^j, u_h^{j-1})}{2\Delta t} \le \mathcal{E}_N^{\tilde{\delta}}(u_h^0) - \mathcal{E}_N^{\tilde{\delta}}(u_h^{n-1}) \le C$$

$$\mathsf{d}_{N,\delta}^2(\tilde{u}_h, \bar{u}_h) \le C\Delta t.$$

Proof. We use the property (6) of the Moreau-Yosida approximation,

$$\frac{\mathsf{d}_{N,\delta}^2(u_h^j, u_h^{j-1})}{2\Delta t} + \frac{1}{2}\int_{t^{j-1}}^{t^j} \frac{d_s^+(u_h^{n-1})}{2r^2}dr = \mathcal{E}_h^{\tilde{\delta}}(u_h^{j-1}) - \mathcal{E}_h^{\tilde{\delta}}(u_h^j).$$

Summarizing the above equation from $j = 1$ to n and use the definitions of $|u_h'|$ and G_N, we get the first equation of the lemma. Ignoring the second term of the equation, we get the second inequality of the lemma. For the last inequality, we have

$$\begin{aligned}
\mathsf{d}_{N,\delta}^2(\tilde{u}_h(t), \bar{u}_h(t)) &= \mathsf{d}_{N,\delta}^2(\tilde{u}_h(t), u_h^n) \\
&\le (\mathsf{d}_{N,\delta}(\tilde{u}_h(t), u_h^{n-1}) + \mathsf{d}_{N,\delta}(u_h^{n-1}, u_h^n))^2 \\
&\le 2\mathsf{d}_{N,\delta}^2(\tilde{u}_h(t), u_h^{n-1}) + 2\mathsf{d}_{N,\delta}^2(u_h^{n-1}, u_h^n) \\
&\le 4\mathsf{d}_{N,\delta}^2(u_h^{n-1}, u_h^n) \le 4\Delta t \textstyle\sum_{j=1}^n \frac{\mathsf{d}_{N,\delta}^2(u_h^{j-1}, u_h^j)}{\Delta t} \le 4C\Delta t.
\end{aligned}$$

Here in the fourth inequality, we use the property (2) of the Moreau-Yosida approximation.

6.3. Gamma-Convergence of the discrete energy. We now show the convergence of the discrete energy $\mathcal{E}_N^{\tilde{\delta}}$ to the continuous energy $\mathcal{N}$ under mild conditions.

LEMMA 6.3 (Γ-convergence of the energy). *Suppose that $\tilde{\delta} = o(1)$, and f in the definition of $\mathcal{E}$ is Lipschitz continuous, then we have $\mathcal{E}_N^{\tilde{\delta}}$ Γ-converges to $\mathcal{E}$ in $L^2(\Omega)$ norm as $N \to \infty$.*

Proof. We prove the results by the definition of Γ-convergence [9, 16]. We need prove the liminf inequality and limsup inequality, respectively.

i). *Liminf inequality.* For any $v_h \in V_N$ s.t. $v_h \to v$ in $L^2(\Omega)$, we need to prove

$$\mathcal{E}(v) \le \liminf_{N\to\infty} \mathcal{E}_N^{\tilde{\delta}}(v_h).$$

If $\lim_{N\to\infty} \mathcal{E}_N^{\tilde{\delta}}(v_h) = +\infty$, the above equality holds surely. Otherwise, we assume that $\liminf_N \mathcal{E}_N^{\tilde{\delta}}(v_h) < +\infty$. There exists a $N_0 > 0$ and $C > 0$, such that $\mathcal{E}_N^{\tilde{\delta}}(v_h) \le C$, for all $N > N_0$. Notice that $f(v_h) \ge 0$ and the regularized term is also positive, we then have

$$\frac{\alpha}{2}\|\nabla v_h\|^2 \le \mathcal{E}_N^{\tilde{\delta}}(v_h) < C.$$

Notice that $v_h \in H_0^1(\Omega)$, there exists a subsequence, still denoted as v_h, and a function $\tilde{v} \in H^1(\Omega)$, such that $v_h \rightharpoonup \tilde{v}$ in $H^1(\Omega)$ and $v_h \to \tilde{v}$ in L^2. By the condition that $v_h \xrightarrow{L^2} v$, we have $v = \tilde{v} \in H^1(\Omega)$.

Notice that $\frac{\alpha}{2}\int_\Omega |\nabla v|^2 dx$ is convex with respect to ∇v. $f(v)$ is a Lipschitz continuous function with respect to v. We can derive that $\mathcal{E}(v)$ is lower semi-continuous with respect to the weak H^1 norm. This leads to

$$\mathcal{E}(v) \le \liminf_{N\to\infty} \mathcal{E}(v_h) \le \liminf_{N\to\infty} \mathcal{E}_N^{\tilde{\delta}}(v_h).$$

ii) *Limsup inequality.* For any $v \in L^2(\Omega)$, we need to prove that there exists a sequence $\tilde{v}_h \in V_N$, such that $\tilde{v}_h \xrightarrow{L^2} v$ and

$$\mathcal{E}(v) \ge \limsup_{N\to\infty} \mathcal{E}_N^{\tilde{\delta}}(\tilde{v}_h).$$

If $v \notin H_0^1(\Omega)$, we have $\mathcal{E}(v) = \infty$. There is nothing to prove. Otherwise, we can assume $v \in H_0^1(\Omega)$ such that $\mathcal{E}(v) < \infty$. In this case, we can choose a fixed triangulation $\tilde{\mathcal{T}}$ of Ω such that $F(\hat{\mathrm{x}}) = \hat{\mathrm{x}}$. Then by the assumption for W(Eq. (5.2)), we know that the regularized term is zero in $\mathcal{E}_N^{\tilde{\delta}}$. Let $\tilde{v}_h = \pi_h v \in V_h(\tilde{\mathcal{T}})$ to be the projection of v on such a mesh. By the density of the piecewise continuous functional space in $H_0^1(\Omega)$, we have $\|v - \tilde{v}_h\|_{H^1} \to 0$. Notice further that f is Lipschitz continuous. This leads to $\mathcal{E}_N^{\tilde{\delta}}(\tilde{v}_h) \to \mathcal{E}(v)$. ◻

6.4. Continuity of distances. In the discrete spaces, we have introduced a distance which is different from that for continuous problems. We need the following results on continuity of distances.

LEMMA 6.4 (Continuity of the distances). *Suppose that $\delta = o(N^{-1})$. Let u and v be two functions in $H_0^1(\Omega)$, and $u_h, v_h \in V_N$ be two sequences of functions satisfying $u_h \xrightarrow{L^2} u$ and $v_h \xrightarrow{L^2} v$ as $N \to \infty$, then we have*

$$\mathsf{d}_{N,\delta}(u_h, v_h) \to \mathsf{d}(u,v) := \|u - v\|_{L^2(\Omega)}, \quad \text{as } N \to \infty.$$

Proof. We first assume $u = v = 0$. This implies that u_h and v_h converge to the same zero function in $L^2(\Omega)$ when $N \to \infty$. We will show that $\mathsf{d}_{N,\delta}(u_h, v_h) \to 0$ in this case. Actually, notice that

$$\begin{aligned}\mathsf{d}_{N,\delta}(u_h, v_h) &\le \mathsf{d}_{N,\delta}(u_h, 0(\mathcal{T}_N)) + \mathsf{d}_{N,\delta}(0(\mathcal{T}_N), 0(\tilde{\mathcal{T}}_N)) + \mathsf{d}_{N,\delta}(0(\tilde{\mathcal{T}}_h), v_h) \\ &= \mathsf{d}(u_h, 0) + \left(\delta \sum_{\hat{\mathrm{x}}_i \in \hat{\mathcal{N}}} |F_1(\hat{\mathrm{x}}_i) - F_2(\hat{x}_i)|^2\right)^{1/2} + \mathsf{d}(0, v_h) \to 0,\end{aligned}$$

where $\mathcal{T}_N = F_1 \hat{\mathcal{T}}_N$ and $\tilde{\mathcal{T}}_N = F_2 \hat{\mathcal{T}}_N$ (with $F_i \in \mathcal{A}_{\mathrm{ad}}$) are respectively partitions of Ω with respect to u_h and v_h, $0(\mathcal{T}_N)$ and $0(\tilde{\mathcal{T}}_N)$ are the corresponding zero functions in $V_h(\mathcal{T}_N)$ and $V_h(\tilde{\mathcal{T}}_N)$. Here in the last limit, we have used the condition that $u_h \xrightarrow{L^2} 0$, $v_h \xrightarrow{L^2} 0$ and $\lim_{N\to\infty} N\delta = 0$.

Then, we consider a (quasi-)uniform partition $\widetilde{\mathcal{T}}$ and let $\pi_h u$ be the projection of u on $V_h(\widetilde{\mathcal{T}})$. Notice that both u_h and $\pi_h u$ converges to u in $L^2(\Omega)$. By similar arguments as above, this leads to the fact that $\mathsf{d}_{N,\delta}(u_h, \pi_h u) \to 0$. Similarly, we also have $\mathsf{d}_{N,\delta}(v_h, \pi_h v) \to 0$. Notice that

$$|\mathsf{d}_{N,\delta}(u_h, v_h) - \mathsf{d}_{N,\delta}(\pi_h u, \pi_h v)| \le \mathsf{d}_{N,\delta}(u_h, \pi_h u) + \mathsf{d}_{N,\delta}(\pi_h v, v_h).$$

and $\mathsf{d}_{N,\delta}(\pi_h u, \pi_h v) = \mathsf{d}(\pi_h u, \pi_h v)$. This gives

$$|\mathsf{d}_{N,\delta}(u_h, v_h) - \mathsf{d}(\pi_h u, \pi_h v)| \to 0,$$

This implies that $\mathsf{d}(\pi_h u, \pi_h v) \to \mathsf{d}(u, v)$. This ends the proof of the lemma. ◻

6.5. λ-convexity. The following lemma states that λ-convexity of the proximal functional $\Phi_N^{\delta,\tilde{\delta}}$ and Φ. The results are essential for the proof of the convergence theorem.

LEMMA 6.5 (Uniform λ-convexity). *Suppose that f in the definition of $\mathcal{E}$ is a smooth function satisfying $f''(\cdot) \geq \lambda_0$ for a constant $\lambda_0 \in \mathbb{R}$. Let*

$$\Phi(\tau, u; v) := \frac{\mathsf{d}^2(u,v)}{\tau} + \mathcal{E}.$$

There exists a $\lambda \in \mathbb{R}$ and $N_0 \in \mathbb{N}^+$ such that the following conclusions hold. For any $v_0, v_1 \in H_0^1(\Omega)$, there exists a curve $v(t)$, $t \in (0,1)$, such that $v(0) = v_0$, $v(1) = v_1$, and for all $0 < \tau < \frac{1}{\lambda^-}$ with $\lambda^- = \max(0, -\lambda)$,

$$\begin{aligned} \Phi(\tau, v_0; v(s)) &\leq (1-s)\Phi(\tau, v_0; v_0) + s\Phi(\tau, v_0; v_1) \\ &\quad -\frac{1}{2}(\lambda + \frac{1}{\tau})s(1-s)\mathsf{d}^2(v_0, v_1). \end{aligned} \tag{6.5}$$

A similar inequality also holds for $\Phi_N^{\delta,\tilde{\delta}}$ uniformly. For any $v_h^{(0)}, v_h^{(1)} \in V_N$ with $N \geq N_0$, there exists a curve $v_h(s)$, $s \in (0,1)$, such that $v_h(0) = v_h^{(0)}$, $v_h(1) = v_h^{(1)}$ and for all $0 < \tau < \frac{1}{\lambda^-}$,

$$\begin{aligned} \Phi_N^{\delta,\tilde{\delta}}(\tau, v_h^{(0)}; v_h(s)) &\leq (1-s)\Phi_N^{\delta,\tilde{\delta}}(\tau, v_h^{(0)}; v_h^{(0)}) + s\Phi_N^{\delta,\tilde{\delta}}(\tau, v_h^{(0)}; v_h^{(1)}) \\ &\quad -\frac{1}{2}(\lambda + \frac{1}{\tau})s(1-s)\mathsf{d}_{N,\delta}^2(v_h^{(0)}, v_h^{(1)}). \end{aligned} \tag{6.6}$$

Proof. We first prove the inequality (6.5) for the continuous problem. For given v_0 and v_1, we simply set $v(t) = (1-t)v_0 + tv_1$. Notice that the second order derivative $f(\cdot)$ is bounded from below. We easily have the following inequality

$$f(v(t)) \leq (1-t)f(v_0) + tv_1 - \frac{\lambda_0}{2}t(1-t)|v_0 - v_1|^2.$$

Then by the convexity of the first term in $\mathcal{E}$, we have

$$\mathcal{E}(v(t)) \leq (1-t)\mathcal{E}(v_0) + t\mathcal{E}(v_1) - \frac{\lambda_0}{2}t(1-t)\|v_1 - v_0\|_0^2. \tag{6.7}$$

Notice again that

$$\frac{\|v(t) - v_0\|_0^2}{2\tau} = \frac{t^2\|v_1 - v_0\|_0^2}{2\tau} = \frac{t}{2\tau}\|v_1 - v_0\|_0^2 - \frac{t(1-t)}{2\tau}\|v_1 - v_0\|_0^2.$$

This together with the equation (6.7) gives the inequality that

$$\Phi(\tau, v_0; v(t)) \leq (1-t)\Phi(\tau, v_0; v_0) + t\Phi(\tau, v_0; v_1) - \frac{1}{2}(\lambda_0 + \frac{1}{\tau})t(1-t)\mathsf{d}^2(v_0, v_1).$$

This leads to (6.5) directly.

By the similar arguments, we can prove the inequality (6.6) with $\lambda \leq \lambda_0$ when both $v_h^{(0)}, v_h^{(1)}$ are piecewise linear functions on the same partition $\tilde{\mathcal{T}}$. In this case, both the penalty term in $\mathcal{E}_N^{\tilde{\delta}}$ and the stabilized term in $\mathsf{d}_{N,\delta}$ are constant or even zero on the curve linearly connecting $v_h^{(0)}$ and $v_h^{(1)}$. For general cases, the result (6.6) can be proved by taking limit for $N \to \infty$ and using the result (6.5) noticing that $\delta = o(N^{-1})$ and $\tilde{\delta} = o(1)$. The rigorous proof is given in the appendix. ∎

The following lemma show the property of the strong upper gradient under the condition of λ-convexity.

Lemma 6.6. *If the λ-convexity property holds as stated in the previous Lemma, then the local slope can be represented as*

$$|\partial \mathcal{E}|(v) = \sup_{w \neq v} \left(\frac{\mathcal{E}(v) - \mathcal{E}(w)}{\mathsf{d}(v,w)} + \frac{\lambda}{2} \mathsf{d}(v,w) \right)^{+}.$$

If, in addition, $\mathcal{E}$ is d-lower semicontinuous, then $|\partial\mathcal{E}|$ is also a strong upper gradient of $\mathcal{E}$ and is d-lower semicontinuous.

Remark 6.1. *The proof of the lemma is given in Theorem 2.4.9 and Corollary 2.4.10 in [2]. The generalization of the lemma to the general p-curves can be found in [34].*

With the previous results, we can prove the following lemma easily.

Lemma 6.7 (liminf condition for the slope). *For any $v_h \in V_N$, such that $v_h \to v$ in $L^2(\Omega)$ when $N \to \infty$, we have*

$$|\partial \mathcal{E}|(v) \leq \liminf_{N\to\infty} |\partial \mathcal{E}_N^{\tilde{\delta}}|(v_N).$$

Proof. By the uniform λ-convexity of Lemma 6.5, the Γ-convergence of Lemma 6.3, and the continuity of the distances of Lemma 6.4, the result in this lemma is a direct conclusion of Proposition 13 in [34]. ◻

6.6. Compactness. The following lemma states some important compactness results.

Lemma 6.8. *Suppose that $\lim_{N\to\infty} \Delta t = 0$, $u_h^0 \xrightarrow{L^2} u_0$, $\mathcal{E}_N^{\tilde{\delta}}(u_h^0) \to \mathcal{E}(u_0)$, as $N \to \infty$, $u_0 \in D(\mathcal{E})$. Let $u_h(t_k)$, $k = 1, \cdots, K$ be the solutions of the discrete problem* (5.7)*. Let $\bar{u}_h$ and $\tilde{u}_h$ are the piecewisely constant approximation and the De Giorgi interpolation defined in* (5.8) *and* (6.3)*, respectively. Then there exists a subsequence, still denoted as u_h and a curve $u(t)$ belongs to $AC^2_{\mathrm{loc}}([0,\infty),\mathcal{E})$, a non-increasing function $\varphi : [0,\infty) \mapsto \mathbb{R}$ and a function $A \in L^2_{\mathrm{loc}}[0,\infty)$, such that*

$$\begin{aligned}
&\bar{u}_h(t) \xrightarrow{L^2} u(t), \quad \tilde{u}_h(t) \xrightarrow{L^2} u(t), && \text{as } N \to \infty, \forall t \in [0,T],\\
&\varphi(t) := \lim_{N\to\infty} \mathcal{E}_h^{\tilde{\delta}}(\bar{u}_h) \geq \mathcal{E}(u), && \forall t \in [0,T],\\
&\mathcal{E}(u(0)) = \mathcal{E}(u_0), &&\\
&|u_h'| \rightharpoonup A \text{ in } L^2_{\mathrm{loc}}([0,\infty)), \quad A(t) \geq |u'|(t), && \text{for } L^1 - a.e.\ t \in (0,\infty),\\
&\liminf_{N\to\infty} G_N(t) \geq |\partial\mathcal{E}|(u(t)). &&
\end{aligned}$$

Proof. Notice that $\mathcal{E}_N^{\tilde{\delta}}(u_h^0) \to \mathcal{E}(u_0)$ as $N \to \infty$. Without loss of generality, we can assume that $\mathcal{E}_N^{\tilde{\delta}}(u_h^0) < \mathcal{E}(u_0) + C_0$ for some $C_0 > 0$. Therefore, we easily have

$$\mathcal{E}_N^{\tilde{\delta}}(u_h^n) \leq \mathcal{E}_N^{\tilde{\delta}}(u_h^0) < \mathcal{E}(u_0) + C_0 < \infty.$$

By the definition of $\mathcal{E}_N^{\tilde{\delta}}$ and the positivity of the function f, we have $\frac{\alpha}{2}|u_h^n|_1^2 \leq \mathcal{E}_h^{\tilde{\delta}}(u_h^n) < \mathcal{E}(u_0) + C_0$. By the Rellich compact embedding theorem, we know that

$$u_h(t) \subset K := \{\mathcal{E}_h^{\tilde{\delta}}(v_h) < C\} \text{ is compact in } L^2.$$

By the energy estimate that,

$$\sum_{j=1}^{n} \frac{\mathsf{d}^2_{N,\delta}(u_h^j,u_h^{j-1})}{2\Delta t} \quad\le\quad \mathcal{E}_h^{\tilde{\delta}}(u_h^0)-\mathcal{E}_h^{\tilde{\delta}}(u_h^{n-1}) \le \mathcal{E}_h^{\tilde{\delta}}(u_h^0), \quad \forall n>0.$$

This leads to $\int_0^T |u_h'|^2(t)\,\mathrm{dt} \le C$. Then there exists a subsequence, still denoted as $|u_h'|$, which converges weakly in $L^2(0,T)$ to a function A as $N \to \infty$. This is $|u_h'| \rightharpoonup A \,\text{in}\, L^2_{\text{loc}}(0,T)$.

For any fixed $0 \le s < t$, let us define $s(n) = \lfloor s/\Delta t \rfloor$ and $t(n) = \lceil t/\Delta t \rceil$. Then by $\lim_{N\to\infty} \Delta t = 0$, we have

$$s(n) \le s < t \le t(n), \quad \lim_{N\to\infty} s(n) = s, \lim_{N\to\infty} t(n) = t.$$

By the inequality

$$\|\bar{u}_h(s) - \bar{u}_h(t)\|_0 \le \mathsf{d}_{N,\delta}(\bar{u}_h(s), \bar{u}_h(t)) \le \int_{s(n)}^{t(n)} |u_h'|(r)dr,$$

we have

$$\limsup_{N\to\infty} \|\bar{u}_h(s) - \bar{u}_h(t)\|_0 \le \int_s^t A(r)dr.$$

Then we could apply the Ascoli-Arzela theorem(c.f. Proposition 3.3.1 in [2]) to obtain $\bar{u}_h(t) \xrightarrow{L^2} u(t)$, $\forall t \in [0,T]$, where $u(t) \in L^2(\Omega)$ is continuous with respect to t. By the estimate in Lemma 6.2, we also have $\tilde{u}_h(t) \xrightarrow{L^2} u(t)$. Furthermore, the limit implies that

$$\|u(s) - u(t)\|_0 \le \int_s^t A(r)dr.$$

By the definition of the metric gradient, we have $u(t) \in AC([0,t], L^2(\Omega))$ and satisfies $|u'|(t) \le A(t)$, $a.e. t \in (0,T)$.

Notice that $\mathcal{E}_N^{\tilde{\delta}}(\bar{u}_h(t))$ is a non-increasing function for any given solution $\bar{u}_h(t)$ in V_N. By Helly's lemma (c.f. Lemma 3.3.3 in [2]) , there exists a subsequence of the discrete solution, still denoted as $\bar{u}_h$, and a function $\varphi(t)$, such that for all $\tilde{T} > 0$,

$$\varphi(t) = \lim_{N\to\infty} \mathcal{E}_N^{\tilde{\delta}}(\bar{u}_h(t)), \qquad \forall t \in (0, \tilde{T}).$$

By Lemma 6.3, we have $\varphi(t) = \lim_{N\to\infty} \mathcal{E}_N^{\tilde{\delta}}(\bar{u}_h(t)) \ge \mathcal{E}(u(t))$. In particular, by the well-preparedness of the intial condition, we have $\varphi(0) = \mathcal{E}(u_0) = \mathcal{E}(u(0))$.

Finally, by the estimate $|\partial \mathcal{E}_N^{\tilde{\delta}}|(\tilde{u}_h) \le G_N(t)$, we have

$$\liminf_{N\to\infty} G_N(t) \ge \liminf_{N\to\infty} |\partial \mathcal{E}_N^{\tilde{\delta}}|(\tilde{u}_h(t)).$$

Notice that by Lemma 6.7, we have

$$\liminf_{N\to\infty} |\partial \mathcal{E}_N^{\tilde{\delta}}|(\tilde{u}_h(t)) \ge |\partial \mathcal{E}|(u(t)).$$

This end the proof of the lemma. ∎

6.7. Proof of the convergence theorem. We are ready to prove the main convergence result as follows.

Proof of Theorem 5.2. By the compact results in Lemma 6.8, we have the following relation up to a subsequence,

$$\begin{aligned}
&\mathcal{E}(u(t)) + \frac{1}{2}\int_0^t |u'|^2(t)\,\mathrm{dt} + \frac{1}{2}\int_0^t |\partial\mathcal{E}|^2(u(t))\,\mathrm{dt}\\
&\le \lim_{N\to\infty} \mathcal{E}_N^{\tilde{\delta}}(\bar{u}_h) + \frac{1}{2}\int_0^t A(t)^2\,\mathrm{dt} + \frac{1}{2}\int_0^t \liminf_{N\to\infty} G_N(t)^2\,\mathrm{dt}\\
&\le \liminf_{N\to\infty}\left(\mathcal{E}_N^{\tilde{\delta}}(\bar{u}_h) + \frac{1}{2}\int_0^t G_N(t)^2\,\mathrm{dt} + \frac{1}{2}\int_0^t |u_h'|^2(t)dt\right)\\
&\le \mathcal{E}(u(0)).
\end{aligned}$$

In the last inequality, we have used the equation (6.4) in Lemma 6.2 and the assumptions on $u_{0,h}$. On the other hand, since $|\partial\mathcal{E}|$ is a strong upper gradient (i.e. Lemma 6.6), we have

$$\mathcal{E}(u_0) \le \mathcal{E}(u(t)) + \int_0^t |\partial\mathcal{E}|(u(s))|u'|(s)\,\mathrm{ds}\,.$$

Therefore, we have

$$\begin{aligned}
&|u'|(t) = |\partial\mathcal{E}|(u(s)) \qquad a.e. t \in (0,\infty),\\
&\mathcal{E}(u_0) = \mathcal{E}(u(t)) + \int_0^t |u'|(t)|\partial\mathcal{E}|(u(t))\,\mathrm{dt}\,.
\end{aligned}$$

This implies the energy identity in Theorem 5.2 and also the relation

$$\frac{d}{dt}\mathcal{E}(u(t)) = -|\partial\mathcal{E}|(u(t))|u'|(t),$$

i.e. $u(t)$ is a curve of maximal slope for $\mathcal{E}$ with respect to $|\partial\mathcal{E}|$. This ends the proof.

7. Numerical experiments.

7.1. Implementations. The JKO scheme (5.7) gives a fully implicit scheme. Notice that $\Phi_N^{\delta,\tilde{\delta}}(\Delta t, u_h; u_h^{n-1})$ is a nonlinear and nonconvex functional with respect to u_h. To solve the optimization problem (5.7) is usually very difficult. We will consider some simplified schemes below.

Firstly, we will do quadralization and compute the distance $\mathsf{d}_{N,\delta}(u_h^{n-1}, v_h)$ approximately. Let $\mathbf{v}$, $\mathbf{x}$ and $\mathbf{y}$ be the coordinates with respect to v_h. Suppose also that $\mathbf{u}^{(n-1)}$, $\mathbf{x}^{(n-1)}$ and $\mathbf{y}^{(n-1)}$ are coordinates with respect to u_h. Then we set

$$\widetilde{\mathsf{d}}^2_{N,\delta}(u_h^{n-1}, v_h) = \begin{pmatrix} \mathbf{v}-\mathbf{u}^{(n-1)}\\ \mathbf{x}-\mathbf{x}^{(n-1)}\\ \mathbf{y}-\mathbf{y}^{(n-1)}\end{pmatrix}^T g_\delta(u_h^{n-1}) \begin{pmatrix} \mathbf{v}-\mathbf{u}^{(n-1)}\\ \mathbf{x}-\mathbf{x}^{(n-1)}\\ \mathbf{y}-\mathbf{y}^{(n-1)}\end{pmatrix}.$$

In each step, we minimize the following functional

$$\inf_{\mathbf{v},\mathbf{x},\mathbf{y}} \widetilde{\Phi}_N^{\delta,\tilde{\delta}}(\mathbf{v},\mathbf{x},\mathbf{y}) := \frac{\widetilde{\mathsf{d}}^2_{N,\delta}(u_h^{n-1}, v_h)}{2\Delta t} + \mathcal{E}_N^{\tilde{\delta}}(v_h). \tag{7.1}$$

This is still a nonlinear optimization problem. We can further simplify it by doing $\mathcal{E}_N^{\tilde{\delta}}(v_h)$ and let

$$\mathcal{E}_N^{\tilde{\delta}}(v_h) \approx \mathcal{E}_N^{\tilde{\delta}}(u_h^{(n-1)}) + \begin{pmatrix} \mathbf{f}_1^{(n-1)} \\ \mathbf{f}_2^{(n-1)} \\ \mathbf{f}_3^{(n-1)} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{v} - \mathbf{u}^{(n-1)} \\ \mathbf{x} - \mathbf{x}^{(n-1)} \\ \mathbf{y} - \mathbf{y}^{(n-1)} \end{pmatrix},$$

where $f_{1,i}^{(n-1)} = \frac{\partial \mathcal{E}_N^{\tilde{\delta}}}{\partial v_i}(u_h^{(n-1)})$, $f_{2,i}^{(n-1)} = \frac{\partial \mathcal{E}_N^{\tilde{\delta}}}{\partial x_i}(u_h^{(n-1)})$ and $f_{3,i}^{(n-1)} = \frac{\partial \mathcal{E}_N^{\tilde{\delta}}}{\partial y_i}(u_h^{(n-1)})$. Then the solution of (7.1) can be approximated by

$$g_\delta(u_h^{n-1}) \begin{pmatrix} \mathbf{u}^{(n)} - \mathbf{u}^{(n-1)} \\ \mathbf{x}^{(n)} - \mathbf{x}^{(n-1)} \\ \mathbf{y}^{(n)} - \mathbf{y}^{(n-1)} \end{pmatrix} = -\Delta t \begin{pmatrix} \mathbf{f}_1^{(n-1)} \\ \mathbf{f}_2^{(n-1)} \\ \mathbf{f}_3^{(n-1)} \end{pmatrix}. \tag{7.2}$$

This is an forward Euler scheme for (5.5) used in the standard MFEM [32, 42]. It is known that very small time step must be chosen for the forward Euler scheme since the corresponding ODE system (5.5) is very stiff in general cases.

In our numerical experiments, we use a different scheme. We do Taylor expansions to $\mathcal{E}_N^{\tilde{\delta}}(v_h)$ up to a second order term and set

$$\begin{aligned}\mathcal{E}_N^{\tilde{\delta}}(v_h) \approx \mathcal{E}_N^{\tilde{\delta}}(u_h^{(n-1)}) &+ \begin{pmatrix} \mathbf{f}_1^{(n-1)} \\ \mathbf{f}_2^{(n-1)} \\ \mathbf{f}_3^{(n-1)} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{v} - \mathbf{u}^{(n-1)} \\ \mathbf{x} - \mathbf{x}^{(n-1)} \\ \mathbf{y} - \mathbf{y}^{(n-1)} \end{pmatrix} \\ &+ \frac{1}{2} \begin{pmatrix} \mathbf{v} - \mathbf{u}^{(n-1)} \\ \mathbf{x} - \mathbf{x}^{(n-1)} \\ \mathbf{y} - \mathbf{y}^{(n-1)} \end{pmatrix}^T H(u_h^{(n-1)}) \begin{pmatrix} \mathbf{v} - \mathbf{u}^{(n-1)} \\ \mathbf{x} - \mathbf{x}^{(n-1)} \\ \mathbf{y} - \mathbf{y}^{(n-1)} \end{pmatrix},\end{aligned}$$

where $H(u_h^{(n-1)})$ is the Hessian matrix of $\mathcal{E}_N^{\tilde{\delta}}(v_h)$ with respect to $(\mathbf{v}, \mathbf{x}, \mathbf{y})$ at $u_h^{(n-1)}$. If we use this approximation to replace $\mathcal{E}_N^{\tilde{\delta}}$ in (7.1), we get a new explicit scheme

$$(g_\delta(u_h^{n-1}) + \Delta t H(u_h^{(n-1)})) \begin{pmatrix} \mathbf{u}^{(n)} - \mathbf{u}^{(n-1)} \\ \mathbf{x}^{(n)} - \mathbf{x}^{(n-1)} \\ \mathbf{y}^{(n)} - \mathbf{y}^{(n-1)} \end{pmatrix} = -\Delta t \begin{pmatrix} \mathbf{f}_1^{(n-1)} \\ \mathbf{f}_2^{(n-1)} \\ \mathbf{f}_3^{(n-1)} \end{pmatrix}. \tag{7.3}$$

It turns out the scheme is much more stable than the explicit scheme (7.2). We can choose relatively large time step Δt in numerical simulations.

7.2. Numerical examples. We first test the accuracy of the numerical scheme (7.3) by solving a model problem in one dimension. We consider a linear equation as in [42].

$$\partial_t u = \partial_{xx} u, \qquad x \in (-3, 3). \tag{7.4}$$

The boundary condition is $u(-3) = u(3) = 0$. The initial condition is given by $u_0 = \frac{1}{\sqrt{4\pi\varepsilon}} e^{-\frac{x^2}{4\varepsilon}}$ with $\varepsilon = 0.001$. The corresponding energy is

$$\mathcal{E}(u) = \frac{1}{2} \int_{-3}^{3} (\partial_x u)^2 dx.$$

The analytic solution of the equation is approximately given by $u(x) = \frac{1}{\sqrt{4\pi(\varepsilon+t)}} e^{-\frac{x^2}{4(\varepsilon+t)}}$ when time t is small.

Since the initial value of the solution is a Gauss function concentrated in a narrow interval centered at $x = 0$. We choose the initial partition of I as follows. We separate the interval $[-3, 3]$ into three parts, $[-3, 0.2) \cup [-0.2, 0.2] \cup (0.2, 3]$. We choose uniform meshes in the three intervals, respectively, while we put $2/3$ of the total number of the vertexes in the middle interval $[-0.2, 0.2]$ and put $1/6$ vertexes to each of the other two intervals. In numerical experiments, we take $\delta = \tilde{\delta} = 0.001$ and the regularized term $W = \frac{1}{N_k} \sum_{K \in \mathcal{T}_h} \ln(\frac{N_k |K|}{|\Omega|})$, where $|K|$ is the length of a cell K of the partition, and N_k is the total number of cells. We solve the problem (7.4) until $T = 0.05$.

The error between the discrete solution u_h and the analytic solution u at T in energy norm is computed by

$$err_{H^1} := \Big(\int_I (\partial_x u(x,T) - \partial_x u_h(x,T))^2 dx \Big)^{1/2},$$

$$err_{L^2} := \Big(\int_I (u(x,T) - u_h(x,T))^2 dx \Big)^{1/2}.$$

The numerical errors are shown in Table 1 for various choice of N and Δt. We first test the convergence with respect to the spacial partitions. The convergence order is computed by $s_i := \ln(err_i/err_{i+1})/\ln(N_{i+1}/N_i)$, which implies that the errors decrease with order $O(N^{-s})$. We can see that the H^1-error is of optimal convergence order $O(N^{-1})$ and the L^2-error is of optimal order $O(N^{-2})$. We also compute similarly the convergence order with respect to time Δt. Both the H^1-error and L^2-error are of optimal order $O(\Delta t)$.

In Figure 1, we show the numerical solution of the equation when $N = 30$ and $\Delta t = 10^{-5}$. We could see that the numerical solution agrees with the analytical solution very well. The grid redistributes automatically when time increases. An interesting observation is that the grid seems to concentrate where the second order derivative of the solution is large.

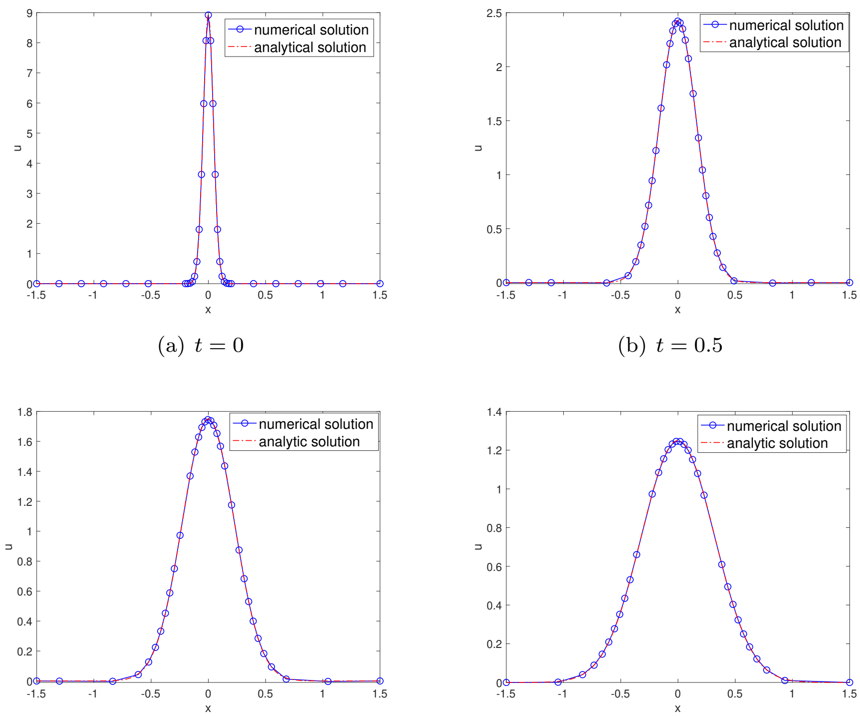
Table 1: The H^1-norm and L^2-norm of the error in Experiment 1.

$\Delta t = 10^{-6}$	err_{H^1}	order	err_{L^2}	order
$N = 16$	0.1837	–	0.0127	–
$N = 30$	0.1136	0.76	0.0038	1.92
$N = 60$	0.0362	1.81	0.000741	2.36
$N = 120$	0.0194	0.99	0.000148	2.32

$N = 120$	err_{H^1}	order	err_{L^2}	order
$\Delta t = 0.01$	0.6013	–	0.0970	–
$\Delta t = 0.005$	0.2513	1.26	0.0430	1.17
$\Delta t = 0.0025$	0.1088	1.21	0.0181	1.25
$\Delta t = 0.00125$	0.0531	1.03	0.0092	0.98

In the second example, we show some numerical results for the two dimensional Allen-Cahn equation. We consider the equation

(7.5) $$\partial_t u = \varepsilon \Delta u + \frac{1}{\varepsilon} f(u), \qquad \text{in } \Omega,$$



(a) $t = 0$ (b) $t = 0.5$

(c) $t = 2$ (d) $t = 4$

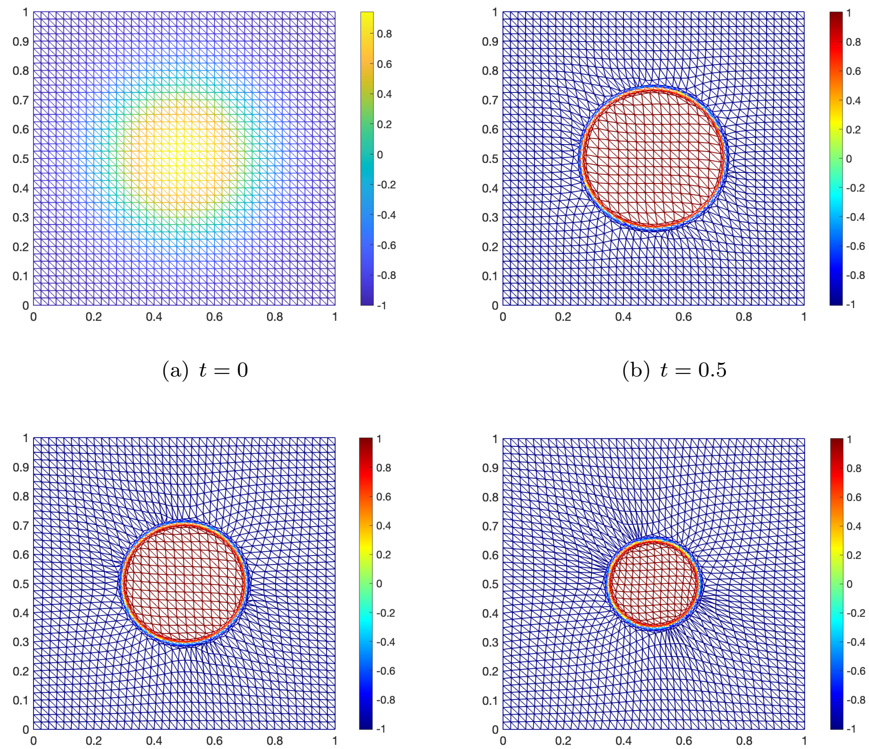
Figure 1: Numerical results of 1D heat equation. The solution and grid of u_h changes with time.

with $\Omega = (0,1) \times (0,1)$ and $f(u) = u^3 - u$. Notice that $f(u) = E'(u)$ with $E(u) = \frac{(1-u^2)^2}{4}$. The corresponding energy is

$$\mathcal{E}(u) = \int_\Omega \frac{\varepsilon}{2}|\nabla u|^2 + \frac{1}{\varepsilon}E(u)dx.$$

For the initial condition, we set $u_0 = (1 + \tanh(\frac{0.25 - \sqrt{(x-0.5)^2+(y-0.5)^2}}{0.1\sqrt{2}})) - 1$ inside the domain and set $u(x) = -1$ for the boundary condition.

In our simulations, we set $\varepsilon = 0.005$, $\delta = \tilde{\delta} = 0.001$ and $\Delta t = 0.0001$. The regularized term is given by $W(\nabla_{\hat{\mathrm{x}}} F) = \int_{\hat{\Omega}} \frac{1}{2}|\nabla_{\hat{\mathrm{x}}} F|^2 d\hat{\mathrm{x}}$. The initial mesh is uniform with mesh size $h = 0.025$. The numerical results are shown in Figure 2. Then the solution evolves and the mesh changes accordingly. The mesh concentrates near the sharp inner layer of the solution. Across the inner layer, the solution of u_h changes dramatically from 1 to -1. The inner layer has a circular shape and the radius decreases gradually with time. This is consistent with the asymptotic analysis result that the solution of the Allen-Cahn equation approximates to a mean curvature flow. We can also see that the meshes also change with time and capture well the evolution of the inner layers. This implies that the MFEM method works quite well for the Allen-Cahn equation.



(a) $t = 0$ (b) $t = 0.5$

(c) $t = 2$ (d) $t = 4$

Figure 2: Numerical results of 2D Allen-Cahn equation. The solution and meshes of u_h changes with time.

8. Conclusions. In conclusion, we present a novel mathematical framework and new numerical schemes for the moving finite element method. A key contribution is the introduction of a regularized metric that enables the formation of a non-degenerate Riemann manifold and a discrete metric space for the finite element space with free knots. To demonstrate the main idea of our approach, we consider a nonlinear reaction diffusion equation as a model problem, which can be interpreted as a curve of maximal slope in the L^2-space. By employing the moving finite element method for numerical discretization, we show that the resulting approximation also follows a curve of maximal slope within the discrete metric space. Importantly, we are able to derive some new numerical schemes, such as the JKO scheme and an explicit stabilized numerical scheme. We establish the existence and convergence of the JKO scheme by utilizing the theory for gradient flows in metric spaces, under mild assumptions. Numerical experiments show that the method works well in both one dimensional and two dimensional problems.

There are some other work need to be done in the future. Firstly, we prove the convergence JKO scheme which is a fully implicit scheme. It is also interesting to further study the convergence of the explicit stabilized scheme, which is much simpler than the JKO scheme and can be used in applications. Secondly, numerical results show that the moving finite element method has optimal convergence rate with respect to space grids and time step size. It is very interesting to establish a priori error estimates for convergence order of the method. Finally, it is worth noting that the framework may be used for understanding other methods that involve moving meshes. Examples include the moving mesh methods [10], the Lagrangian type methods [28, 27, 14], and the arbitrary Lagrangian Eulerian method.

Appendix: Details of the proof for the λ-convexity of the discrete proximal functional. We prove the inequality (6.6) for general cases by contradiction. If the inequality does not hold, we can find a series of functions $v_h^{(k)}, \tilde{v}_h^{(k)} \in V_{N_k}$ with $N_k \to +\infty$, and an increasing series of numbers $c_k \to \infty$ and $\tau_k > 0$, such that the following inequality holds for any curve $v_h^{(k)}(t)$ connecting $v_h^{(k)}$ and $\tilde{v}_h^{(k)}$ in V_{N_k},

$$\begin{aligned} \Phi_{N_k}^{\delta,\tilde{\delta}}(\tau_k, v_h^{(k)}; v_h^{(k)}(t_k)) >&(1-t_k)\Phi_{N_k}^{\delta,\tilde{\delta}}(\tau_k, v_h^{(k)}; v_h^{(k)}) + t_k \Phi_{N_k}^{\delta,\tilde{\delta}}(\tau_k, v_h^{(k)}; \tilde{v}_h^{(k)}) \\ &+ \frac{1}{2} c_k t_k (1-t_k) \mathsf{d}_{N,\delta}^2(v_h^{(k)}, \tilde{v}_h^{(k)}), \end{aligned} \tag{8.1}$$

for some $t_k \in (0,1)$. Clearly 0 or 1 can not be an accumulation point of t_k. Otherwise, the above inequality will be an equality. Also 0 can not be an accumulation point of τ_k. Otherwise, both sides of the inequality will go to infinity. Without loss of generality, we can assume that $t_k \to \hat{t}$ and $\tau_k \to \hat{\tau}$ as $k \to \infty$. Then we have $\hat{t} \in (0,1)$ and $\hat{\tau} \in (0, \infty]$.

Then we can see that both $\Phi_{N_k}^{\delta,\tilde{\delta}}(\tau_k, v_h^{(k)}; v_h^{(k)})$ and $\Phi_{N_k}^{\delta,\tilde{\delta}}(\tau_k, v_h^{(k)}; \tilde{v}_h^{(k)})$ are bounded from above. This leads to the fact that $\mathcal{E}_N^{\tilde{\delta}}(v_h^{(k)})$ and $\mathcal{E}_N^{\tilde{\delta}}(\tilde{v}_h^{(k)})$ are bounded. By the Sobolev compact embedding theorem, there exist subsequences of $v_h^{(k)}$ and $\tilde{v}_h^{(k)}$, still denoted as the same notation, such that

$$v_h^{(k)} \to \tilde{v}_0 \text{ in } L^2, \quad v_h^{(k)} \rightharpoonup \tilde{v}_0 \text{ in } H^1$$

and

$$\tilde{v}_h^{(k)} \to \tilde{v}_1 \text{ in } L^2, \quad \tilde{v}_h^{(k)} \rightharpoonup \tilde{v}_1 \text{ in } H^1.$$

We choose the curve connecting $v_h^{(k)}$ and $\tilde{v}_h^{(k)}$ as follows. Let $\mathcal{T}_{N_k}$ be the reference triangulation of Ω such that $F(\hat{x}) = \hat{x}$. Let $\pi_h^k \tilde{v}_0$ and $\pi_h^k \tilde{v}_1$ be the projection of $\tilde{v}_0$ and $\tilde{v}_1$ on $\mathcal{T}_{N_k}$. By the Lemma 6.4, we have $d_0^{(k)} := \mathsf{d}_{N_k}(v_h^{(k)}, \pi_h^k \tilde{v}_0) \to 0$ and $d_1^{(k)} := \mathsf{d}_{N_k}(\tilde{v}_h^{(k)}, \pi_h^k \tilde{v}_1) \to 0$. We choose a curve which includes three parts, the geodesic curve $\gamma_0^k(s)$ with $s \in (0, d_0^{(k)})$ between $v_h^{(k)}$ and $\pi_h^k \tilde{v}_0$ such that $\mathsf{d}_{N_k}(\gamma_0^k(s), v_h^{(k)}) = s$ (the existence of such a curve can be seen in Lemma 1.1.4 in [2]), the linear combination between $\pi_h^k \tilde{v}_0$ and $\pi_h^k \tilde{v}_1$, and the geodesic curve $\gamma_1^k(s)$ with $s \in (1 - d_1^{(k)}, 1)$ between $\tilde{v}_h^{(k)}$ and $\pi_h^k \tilde{v}_1$ such that $\mathsf{d}_{N_k}(\gamma_1^k(s), \tilde{v}_1^{(k)}) = (1-s)$. Defined as follows,

$$v_h^{(k)}(t) = \begin{cases} \gamma_0^k(t) & \text{if } 0 < t < d_1^{(k)} \\ \frac{\pi_h^k \tilde{v}_0 (1 - d_2^{(k)} - t) + \pi_h^k \tilde{v}_1 (t - d_1^{(k)})}{1 - d_2^{(k)} - d_1^{(k)}} & \text{if } d_1^{(k)} < t < 1 - d_2^{(k)} \\ \gamma_1^k(t) & \text{if } 1 - d_2^{(k)} < t < 1. \end{cases}$$

Then by taking limit of the equation (8.1) when $k \to \infty$, (noticing that $c_k \to \infty$, $\delta = o(N^{-1})$ and $\tilde{\delta} = o(1)$), we have

$$\Phi(\hat{\tau}, \tilde{v}_0; \hat{v}(\hat{t})) > (1-\hat{t})\Phi(\hat{\tau}, \tilde{v}_0; \tilde{v}_0) + \hat{t}\Phi(\hat{\tau}; \tilde{v}_0, \tilde{v}_1) + \infty.$$

This contradicts with the inequality (6.5).

REFERENCES

[1] M. Ainsworth and J. T. Oden. A Posteriori Error Estimation in Finite Element Analysis, volume 37. John Wiley & Sons, 2000.

[2] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. Gradient flows: in metric spaces and in the space of probability measures. Springer Science & Business Media, 2005.

[3] I. Babuvška and W. C. Rheinboldt. Error estimates for adaptive finite element computations. SIAM Journal on Numerical Analysis, 15(4):736–754, 1978.

[4] M. J. Baines. Moving Finite Elements. Clarendon Press, Oxford, 1994.

[5] M.J. Baines, M.E. Hubbard, and P. K. Jimack. Velocity-based moving mesh methods for nonlinear partial differential equations. Communications in Computational Physics, 10(3):509–576, 2011.

[6] R. E. Bank and M. S. Metti. A diagonally-implicit time integration scheme for space-time moving finite elements. Journal of Computational Mathematics, 37(3), 2019.

[7] R. E. Bank and R. F. Santos. Analysis of some moving space-time finite element methods. SIAM Journal on Numerical Analysis, 30(1):1–18, 1993.

[8] Peter Binev, Wolfgang Dahmen, and Ron DeVore. Adaptive finite element methods with convergence rates. Numerische Mathematik, 97(2):219–268, 2004.

[9] A. Braides. Gamma-convergence for Beginners, volume 22. Clarendon Press, 2002.

[10] C. J. Budd, W. Huang, and R. D. Russell. Adaptivity with moving grids. Acta Numerica, 18:111–241, 2009.

[11] W. Cao, W. Huang, and R. D. Russell. A moving mesh method based on the geometric conservation law. SIAM Journal on Scientific Computing, 24(1):118–142, 2002.

[12] N. N. Carlson and K. Miller. Design and application of a gradient-weighted moving finite element code I: in one dimension. SIAM Journal on Scientific Computing, 19(3):728–765, 1998.

[13] N. N. Carlson and K. Miller. Design and application of a gradient-weighted moving finite element code II: in two dimensions. SIAM Journal on Scientific Computing, 19(3):766–798, 1998.

[14] J. A. Carrillo, D. Matthes, and M.-T. Wolfram. Lagrangian schemes for wasserstein gradient flows. Handbook of Numerical Analysis, 22:271–311, 2021.

[15] N. M. Chadha and N. Kopteva. A robust grid equidistribution method for a one-dimensional singularly perturbed semilinear reaction–diffusion problem. IMA Journal of Numerical Analysis, 31(1):188–211, 2011.

[16] G. Dal Maso. An introduction to Γ-convergence, volume 8. Springer Science & Business Media, 2012.

[17] C. de Boor. Good approximation by splines with variable knots. In Spline Functions and Approximation Theory: Proceedings of the Symposium held at the University of Alberta, Edmonton May 29 to June 1, 1972, pages 57–72. Springer, 1973.

[18] R. A. DeVore. Nonlinear approximation. Acta Numerica, 7:51–150, 1998.

[19] Y. Di, R. Li, T. Tang, and P. Zhang. Moving mesh finite element methods for the incompressible navier–stokes equations. SIAM Journal on Scientific Computing, 26(3):1036–1056, 2005.

[20] T. F. Dupont and Y. Liu. Symmetric error estimates for moving mesh galerkin methods for advection-diffusion equations. SIAM Journal on Numerical Analysis, 40(3):914–927, 2002.

[21] D. F. Hawken, J. J. Gottlieb, and J. S. Hansen. Review of some adaptive node-movement techniques in finite-element and finite-difference solutions of partial differential equations. Journal of Computational Physics, 95(2):254–302, 1991.

[22] W. Huang, Y. Ren, and R. D. Russell. Moving mesh partial differential equations (MMPDES) based on the equidistribution principle. SIAM Journal on Numerical Analysis, 31(3):709–730, 1994.

[23] P. K. Jimack. Optimal eigenvalue and asymtotic large-time approximations using the moving finite-element method. IMA journal of numerical analysis, 16(3):381–398, 1996.

[24] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the fokker–planck equation. SIAM Journal on Mathematical Analysis, 29(1):1–17, 1998.

[25] N. Kopteva. Convergence theory of moving grid methods. In Adaptive Computations: Theory and Algorithms, pages 159–210. Science Press, Beijing, 2007.

[26] R. Li, T. Tang, and P. Zhang. Moving mesh methods in multiple dimensions based on harmonic maps. Journal of Computational Physics, 170(2):562–588, 2001.
[27] C. Liu and Y. Wang. On Lagrangian schemes for porous medium type generalized diffusion equations: A discrete energetic variational approach. Journal of Computational Physics, 417:109566, 2020.
[28] C. Liu and Y. Wang. A variational Lagrangian scheme for a phase-field model: A discrete energetic variational approach. SIAM Journal on Scientific Computing, 42(6):B1541–B1569, 2020.
[29] Y. Liu, R. E. Bank, T. F. Dupont, S. Garcia, and R. F. Santos. Symmetric error estimates for moving mesh mixed methods for advection-diffusion equations. SIAM Journal on Numerical Analysis, 40(6):2270–2291, 2003.
[30] A. Mielke, R. Rossi, and G. Savaré. Variational convergence of gradient flows and rate-independent evolutions in metric spaces. Milan Journal of Mathematics, 80:381–410, 2012.
[31] K. Miller. Moving finite elements. II. SIAM Journal on Numerical Analysis, 18(6):1033–1057, 1981.
[32] K. Miller and R. N. Miller. Moving finite elements. I. SIAM Journal on Numerical Analysis, 18(6):1019–1032, 1981.
[33] C. Ngo and W. Huang. A study on moving mesh finite element solution of the porous medium equation. Journal of Computational Physics, 331:357–380, 2017.
[34] C. Ortner. Two variational techniques for the approximation of curves of maximal slope. 2005.
[35] W. Ren and X.-P. Wang. An iterative grid redistribution method for singular problems in multiple dimensions. Journal of Computational Physics, 159(2):246–273, 2000.
[36] S. Serfaty. Gamma-convergence of gradient flows on hilbert and metric spaces and applications. Discrete Contin. Dyn. Syst, 31(4):1427–1451, 2011.
[37] R. Stevenson. Optimality of a standard adaptive finite element method. Foundations of Computational Mathematics, 7:245–269, 2007.
[38] H. Tang and T. Tang. Adaptive mesh methods for one-and two-dimensional hyperbolic conservation laws. SIAM Journal on Numerical Analysis, 41(2):487–515, 2003.
[39] T. Tang. Moving mesh methods for computational fluid dynamics. Contemporary mathematics, 383(8):141–173, 2005.
[40] R. Verfürth. A posteriori error estimation techniques for finite element methods. OUP Oxford, 2013.
[41] A. J. Wathen and M. J. Baines. On the structure of the moving finite-element equations. IMA Journal of Numerical Analysis, 5(2):161–182, 1985.
[42] X. Xu. A variational analysis for the moving finite element method for gradient flows. Journal of Computational Mathematics, 41(2), 2023.
[43] F. Zhang, W. Huang, X. Li, and S. Zhang. Moving mesh finite element simulation for phase-field modeling of brittle fracture and convergence of newton's iteration. Journal of Computational Physics, 356:127–149, 2018.