Structure-preserving variational schemes for fourth order nonlinear partial differential equations with a Wasserstein gradient flow structure

Blake Ashworth

Submitted for the degree of Doctor of Philosophy

University of Sussex

August 2020

Declaration

I hereby declare that this thesis has not been and will not be submitted in whole or in part to another University for the award of any other degree. Furthermore, expect for where references are specified, this thesis is my own work under supervision by Prof. Bertram Düring.

Signature:

Blake Ashworth

UNIVERSITY OF SUSSEX

Blake Ashworth, Doctor of Philosophy

Summary

There is a growing interest in studying nonlinear partial differential equations which constitute gradient flows in the Wasserstein metric and related structure preserving variational discretisations. In this thesis, we focus on the fourth order Derrida-Lebowitz-Speer-Spohn (DLSS) equation, the thin film equation, as well as other fourth order examples. We adapt the minimising movement schemes from implicit Euler (BDF1) to higher order schemes, i.e. backward difference formulae and diagonally implicit Runge-Kutta (DIRK) methods.

We prove numerical convergence of discrete solutions of the DIRK2 scheme using a comparison principle type approach with semi-convex based conditions. With basic assumptions including semi-convexity of our energy, verifying that the energy is monotonic in time normally yields convergence of its discrete solution for decreasing time step. However, as in the BDF2 example, for the DIRK2 scheme considered here the energy was not verified to be monotonic (it might be), yet with additional assumptions, convergence is obtained as well as other basic properties of gradient flows.

We propose fully discrete schemes which preserve positivity for the DLSS equation, the Thin Film equation and other nonlinear partial differential equations. We present results of numerical experiments confirming improved rates of convergence for higher order schemes. Furthermore, numerical results with non-constant time steps are presented, improving the efficiency of the proposed schemes.

Acknowledgements

I would like to dedicate this thesis to my supervisor Prof. Bertram Düring, whose continued support, expertise and regular discussions made a big difference during the times that were needed the most and enabled me to progress successfully throughout, thus to this successful publication. Also, I would like to thank my second supervisor Dr David McCormick for his input and support, during the critical few months of the project and also to the Leverhulme Trust for the funding/studentship which made this opportunity possible.

I wish to thank the University of Sussex School of Mathematics and Physical Sciences for always providing a positive, friendly place to work and am grateful for all opportunities and skills I had been able to acquire. Firstly to the faculty and for the years during my MMath degree, where I have been able to build successfully towards my PhD: thank you to all including my senior tutor Prof. James Hirschfeld, academic advisor Dr Konstantin Blyuss, MMath project supervisor Prof. Peter Giesl and annual reviewer Dr. Max Jensen. Also to the admin staff including Becky, Tom, Amelia, Lily, Gemma and Jodie.

I would also not be here without my fellow peers and researchers at Sussex who I have had the pleasure to know and share our journeys with on a daily basis over the last few years, and kept me going whenever times were tough. I am sincerely grateful to my fellow peers those who were with me from the start of my PhD to the very end: these include Aleksandra, for her encouragement and support in the last weeks before submission, office mate Rosie for her close support throughout and during my MMath degree, as well as James, Lam and Stelios. Also, a big thank you to all my office companions over the years, Awss, Francesco, Malian, Panagiota, and Shane, plus other peers including Wakil, George, Benard Andreas, Steven, Laura, Davide, Farzad, Rabab and Laura.

And last but not least, my special regards go to all family and friends in my home region, parents and my sister Estella, who always believed that I could thrive at the challenge and had succeeded.

Contents

1 Introduction

The aim of this thesis is to propose and discuss variational discretisations of nonlinear partial differential equations (PDEs) which constitute gradient flows in the Wasserstein metric [17, 19].

As demonstrated in recent articles, [12, 26, 30, 32, 39], mathematicians are particularly interested in studying evolution equations (in particular diffusion equations) with underlying gradient flow structures [2, 19]. There have been many breakthroughs over the last quarter of a century on how to analyse the behaviour of the dynamics or solutions of evolution equations.

Although analysing the variational forms of evolution PDEs are of interest at present, the main focus nowadays are the study of higher order nonlinear PDEs, which are difficult to compute, due to range restrictions and the inability to linearise the system.

1.1 Gradient Flow Problems of PDEs

We introduce the literature and how the variational form (gradient flow) can help solve numerical solutions of high order nonlinear PDEs.

Over recent years, mathematicians are interested in nonlinear fourth order equations that provide non-negative solutions, i.e. the solution can be interpreted as some physical quantity. Most notably the Derrida-Lebowitz-Speer-Spohn (DLSS) equation, introduced in 1991 (see [15, 16] from Derrida Lebowitz et al.):

$$\partial_t u(x,t) = -2\partial_x\Big(u(x,t)\partial_x\Big(\frac{\partial_x^2(\sqrt{u(x,t)})}{\sqrt{u(x,t)}}\Big)\Big), \quad t > 0, \tag{1}$$

and the Thin Film equation, introduced in 1995 (see [5]):

$$\partial_t u(x,t) = -\partial_x(u(x,t)\partial_x^3 u(x,t)), \quad t > 0, \tag{2}$$

which can be shown to be variationally formulated as gradient flows with respect to the L^2-Wasserstein metric.

There is a growing interest on identifying an underlying structure to these equations, where many have been identified as having underlying gradient flow structures (see [2, 19] for some examples). The discretisation approach to PDEs with gradient flow structures, called the minimising movement scheme, originally for the finite-dimensional case, was introduced back in 1990 by E. De Giorgi, at a conference in Lecce, Italy before being formally published in 1993 [14].

Gradient flow problems can be expressed with respect to the Euclidean space or the L^2 (Lebesgue 2-norm) space.

But since the linear Fokker-Planck equation was first applied twenty one years ago by Jordan, Kinderlehrer and Otto [26] in 1998, the main interest involves the variational formulation of evolution equations as gradient flows with respect to the L^2-Wasserstein metric, also defined as the optimal quadratic transportation distance.

The global existence of non-negative weak solutions to the DLSS equation was proved by Jüngel and Pinnau in [27]. Otto applied this variational formulation as Wasserstein gradient flows, that same year [39], for the porous media equation. Carrillo et al. [12] applied this to conservation equations for interacting gases.

From 2006 onwards, a Lagrangian approach which explores the spatial behaviour of the density, firstly demonstrated by Gosse et al. [21], then Carillo and Moll [6, 13] where the authors applied "optimise-then-discretise" and "discretise-then-optimise" approaches, respectively.

Jüngel and Violet [29] constructed a semi-discrete in time scheme for the DLSS equation, which is also known as the Quantum drift diffusion equation or the "nonlinear logarithmic" equation due to the logarithmic term in the vector field part of the equation. This was shown to preserve non-negativity of the solution for all initial states commencing as this.

The variational form of the DLSS equation was proved by Gianazza et al. [20]. They not only showed the Wasserstein gradient flow structure of the energy functional (Fisher information) but also verified that taking the limit of the temporal discrete scheme preserves non-negativity of the solution, which is essential for ensuring global existence of the solution.

Our aim is to construct new schemes which obey the gradient flow structure. Gradient flows travel as to minimise the associated internal energy functional, hence we wish for the energy at the time discrete level to be monotonically decreasing. The scheme, defined as the minimising movement scheme, has been studied in the Euclidean case, in the probability space, and now by Carillo and Moll [13] in the Lagrangian case. Indeed, they introduced a scheme for various nonlinear evolution equations, rewritten in terms of Lagrangian coordinates which simplify the calculation of the transportation distance between two densities (Wasserstein distance) and guarantees mass preservation and non-negative solutions [3, 18].

This work was considering second order nonlinear equations, but with Gianazza and Toscani et al. [20], this prompted Düring and Matthes et al. [17] to investigate fourth order equations. In this case, they successfully applied a full discretisation to the fourth order DLSS equation (we will formally introduce it in Section 3). They not only created a numerical scheme, derived in line with the structure (Wasserstein gradient flow (WGF) formulation) of the already established PDE from [26], they derived a fully discrete scheme, in space as well as time, where its spatially discrete form covers limitations on the solution space e.g. singularities exist, only non-negative solutions exist. In addition to this the scheme, which is equivalent to the implicit Euler case, guarantees conservation of mass and is unconditionally stable.

The benefits of variational forms, for example, evolution equations with respect to gradient flow structures have been discussed, but more also are other important properties of the solutions, see [17], including the dissipation of internal energies or preservation of mass. The nonlinearity of these equations have been confronted with a construction of a numerical solution, with these built in properties. As well as just analysing the dynamics of evolution equations, we are interested in equations with high nonlinear structures and of high order due to their rich and interesting behaviour, as well as their well known use in theoretical physics and mathematical analysis.

1.2 Higher Order Discretisations of Fourth-Order PDEs

Now an extension of the skeleton of what we want to achieve in this thesis i.e. analysis of gradient flows have been applied only to first order BDF schemes (vaguely for BDF2) and now we adapt the temporal discretisation for BDF3 to 6 then DIRK schemes of orders 2,3 and 4.

We investigate and propose new discretisations for evolution PDEs that carry an underlying L^2-Wasserstein gradient flow structure. Examples of other well known equations are shown in [19, 40, 41], including the Heat equation and the Porous Media equation. In this thesis, the focus will be towards evolution equations of fourth order, rather than second order. With articles [17, 20], introducing the fully discrete scheme for the fourth order DLSS equation with built in constraints, which is well posed, this can be implemented to other fourth order equations with underlying Wasserstein gradient flow structures, by altering the internal energies, and will also be the forefront of this thesis. The challenge is to investigate numerical schemes providing an improved level of accuracy, at the time discrete level. As a consequence of the strong non-linearity of these equations, obtaining an analytic solution is difficult, but our investigations, especially with motivation from previously published results, help provide an improved approximation (numerical solution), shown by a higher numerical order of convergence in time. As from the implicit Euler scheme, we implement a well known, iterative approach called the minimising movement scheme.

Variational formulations provide us with a deep, clearer understanding of the qualitative behaviour of solutions and within the probability space of measures. But more importantly, the minimising movement scheme is a popular tool for implementing a temporal (semi-discrete) approach for approximating the evolution of solutions, [14, 17, 19, 35, 41, 42], called the semi-discrete gradient flow problem. Some basic assumptions and mathematical tools [35, p. 11] can verify the uniqueness of a minimiser of a functional, a necessary condition for a well-posed problem, derived from our numerical schemes that we discuss in this thesis.

The numerical schemes applied for the PDEs we consider were of first order in time originally and little consideration was given for schemes of high order, mainly of second order. We will investigate and find that this work can be generalised to wider classes of schemes and equations. Additional barriers have and will create challenges going forward, but the publications [17] in 2010, and [35] in 2017, have been the main ingredients for creating good approximations of higher order equations with gradient flow structures and tackling limited monotonicity of our internal energies from higher order temporal schemes respectively. But now we are interested in other higher order variants of the scheme and other fourth order equations.

We first adapt the minimising movement schemes from the implicit Euler scheme (BDF1) to higher order schemes, e.g. backward difference formulas (BDF) of orders 2 to 6 (BDF2 to BDF6) and diagonally implicit Runge-Kutta (DIRK) methods up to stage five.

The main problem in this thesis is an extension of recent work by Matthes and Plazotta, [35]. They have successfully shown the well-posedness of the BDF2 type scheme (in a variational form by deriving a time discrete evolution variational inequality (EVI), that is a new variational form of the BDF2 method), formulated as a consequence of the semi-convexity assumption on the energy functional $\mathcal{E}(\cdot)$, guaranteeing a well-posed BDF scheme. They also shown the convergence of the discrete solution with similar initial data for a decreasing time step size by using a comparison principle approach, constructing an error estimate between two similar discrete solutions, for example an estimate on $\mathcal{W}_2[u_\tau^k, v_\eta^l]^2 - \mathcal{W}_2[u_\tau^0, v_\eta^0]^2$ where $(u_\tau^k)_{k\in\mathbb{N}}$ and $(v_\eta^l)_{l\in\mathbb{N}}$ are two discrete solutions with time steps τ and η, respectively. Our main contribution is to extend this to higher order, and multistage, schemes where the energy functional is not shown to be monotonically decreasing, in comparison to the demonstrated BDF1 (implicit Euler) scheme. This was achieved by a comparison principle approach, also demonstrated by Matthes, Plazotta [35], which involves an alternative variational form of the BDF2 method. We adapt this approach, for an "appropriate" two stage diagonally implicit Runge-Kutta (DIRK2) scheme with a high order of accuracy.

The approach characterises the gradient flow which assumes λ-convexity (semi-convexity of modulus λ) of the energy functional (see Santambrogio, from [42, Sect. 2.2] or [41, Sect. 8.1]), and shows how it generates a unique stable solution as well as giving an equivalent form of the gradient flow problem with metric counterparts. This is defined as the evolution variational inequality (EVI) and we shall prove this discretely in the non-trivial case when λ is negative, where convexity is only satisfied for a sufficiently small time step size.

Remark 1.1. In terms of "appropriate", the DIRK2 scheme we select only has a maximum order of accuracy of two. Therefore, there is no guarantee that the scheme is of order two, hence carries any benefit to the BDF1 scheme.

In addition, we carefully select and discretise in space, via the Galerkin and Lagrangian transformation approach for other, some well known, equations. After some careful consideration of initial datums to select, we look to see whether our fully discrete approach is indeed compatible for other fourth order nonlinear diffusion equations. We analyse the numerical convergence (L^2-error) rates.

From [35], the general energy functional was considered for the whole real line $\mathbb{R}$, due to the fact that they were analysing for general metric spaces and hence, for general evolution PDEs. In other words, the associated energy functionals are not always an integral of a squared term, hence may not be positive definite. However, for our evolution equations investigated in this thesis, the respective energies are integrals of a squared (non-negative) term, guaranteeing a non-negative output. For example, we are formulating the DLSS Equation as a L^2-Wasserstein gradient flow with the associated Fisher Information term which integrates a squared term, hence non-negativity i.e. $\mathcal{E}(\cdot) \in \mathbb{R}_+$ for all reference/observer points $u \in \mathcal{P}_M(\Omega)$. This provides us a lower bound for our energy functional term which is zero at least, unlike the case from [35, Thm. 4.4] for general metric cases, where an inductive result from the minimising movement scheme estimates had to be shown for general cases. Hence this shall simplify our final estimate proof worked on in Sections 5 to 6.

The next part of our introduction brings us to a detailed structure of our thesis that enables us to construct our new numerical methods for finding a discrete approximate solution for our selected PDE in the case of Wasserstein spaces.

1.3 Outline of the Thesis

The outline of what each section of this thesis covers is shown here. The final part summarises the aims/main contributions in bullet points. With the background of our project set out, we briefly explain the purposes of each of the investigation to be carried out in this thesis:

The basic theory, which we build in this thesis is first mentioned, before our contribution. We commence with the traditional optimal transport problem in Section 2, including the Wasserstein metric. Then Section 3, in the Wasserstein space aspect, introduces the basic facts on the variational structure of the investigated PDEs and gradient flows, including its discrete portrayal, which will be a major tool for our contribution.

Section 4 introduces the schemes to be worked with. Starting with the basic backward difference formulas (BDF), explaining how this generates the orders of accuracy as concerned. Runge-Kutta methods are then introduced proving how the general form of schemes, for second and third stage cases, has a maximum order of accuracy. We also propose the minimising movement schemes for DIRK scheme cases, including the DIRK2 scheme we worked with earlier, as well as an example from both DIRK3 and DIRK5.

For example, we shall begin considering the minimising movement scheme, in relation to the implicit Euler, stage one, scheme on the probability space of measures with mass M, $\mathcal{P}_M(\Omega)$, i.e. the Wasserstein case:

$$u_\tau^n := \operatorname*{argmin}_{u \in \mathcal{P}_M(\Omega)} \frac{1}{2\tau} \mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u). \tag{3}$$

Section 5 provides the ingredients required for our investigation. We shall summarise the classical minimising movement scheme for the backward difference formula (BDF) methods, highlighting how the monotonicity of the energy functional is not guaranteed and that more work is to be done for higher order and stage numerical schemes. We shall also introduce the diagonally implicit Runge-Kutta two stage (DIRK2) scheme that carries a suitably high order of accuracy. Article [35] proved the numerical convergence of discrete solutions for the BDF2 scheme, with a *not* fully diminished energy functional (Fisher Information) term. We shall adopt a similar result for the DIRK2 scheme, although we shall not generalise the numerical scheme as they did, i.e. the domain and metric form is fixed, as well as the PDE and energy functional term $\mathcal{E}(\cdot)$. Estimating the energy functional terms concludes the section. In addition to the existence of a minimiser result from the previous section, we shall start adopting a similar result, from the BDF2 scheme, for the DIRK2 scheme. Specifically, we shall construct the variational form of the minimising movement scheme, where we show how to adapt the discrete forms of the evolution variational inequalities for both stages, as a result of the basic assumptions on our energy functionals, i.e. lower semi-continuity, coercivity and semi-convexity. Note again that we shall not generalise the numerical scheme as they did in [35], by Matthes and Plazotta.

The numerical convergence proof is conducted in Section 6, adopting parts of the comparison principle approach from [35]. Furthermore, semi-convexity on the energy functional and the L^2-Wasserstein metric terms are assumed and will be the main ingredient for verifying well-posedness for gradient flow problems.

In Section 7 we carry out the process of constructing the numerical results for the BDF and DIRK schemes. First, we implement the full discretisation [17, 48], applied to the DLSS equation, and three other equations, including the Thin Film equation. We will also show the brief outline for the optimisation problem for all schemes.

The simulations for the L^2-error numerical convergence on the time step size τ, as well as verifying, numerically, the monotonicity of the energy functional over time are carried out in Section 8. This concludes whether higher order, and multi-stage, schemes are more accurate numerical approximations to the DLSS and Thin Film equations, plus other fourth order nonlinear partial differential equations, carrying similar structures.

In this thesis, our main contributions are as follows:

- we derive temporally higher order minimising movement schemes, based on BDF and DIRK schemes with higher order of accuracy (Section 4).

- we prove the convergence of discrete solutions for the DIRK2 scheme with an arbitrary intermediate time step (Section 5 to 6).

- we derive fully discrete "discretise-then-optimise" schemes for several fourth order nonlinear partial differential equations, e.g. Thin Film equation, with underlying gradient flow structure (Section 7).

- we implement the new temporally higher order BDF and DIRK schemes for various nonlinear fourth order partial differential equations and assess their numerical convergence rates, as well as their relation with the smoothness of the initial conditions, altering their built-in parameters (Section 8).

We shall provide an introduction to gradient flows in the Wasserstein space, but first an introduction to optimal transport, referring to original problems proposed by Monge and Kantorovich [42]. Firstly, we look at two main problems that concern finding the best possible plan for transporting a quantity from one space to another.

2 Optimal Transport Problem

We give the idea of the optimal transport problem, hence an interpretation of the Wasserstein metrics, which we work on/use for the initial value problems.

The Optimal Transport problem was introduced a few years before the French revolution by Gaspard Monge [37]. Monge proposed the problem originally, where the density of a given mass is transported to another location to form a new density but the mass is preserved, i.e. the problem is to find the minimum cost of transporting one density into another. For example, we turn over a pile of sand into a hole of the same volume and we find the most efficient way of transporting a physical quantity from one place to another?

We consider two probability measures on two probability spaces X and Y, which are $\mathcal{P}(X)$ and $\mathcal{P}(Y)$ respectively, and the cost function $c : X \times Y \to \mathbb{R}_+$. The distribution of the sand in the initial pile has density $\mu(x)$, and the density after transportation is denoted by $\nu(y)$. The movement is described by the transport map $T : X \to Y$; $X, Y \in \Omega$, where Ω is a metric space, e.g. Euclidean or Wasserstein. The map T describes the movement of the particles from X to Y, x denotes the original location of the particle of sand and $T(x)$ denotes the destination of the corresponding particle x. By introducing $d : X \times Y \to \mathbb{R}_+$ as the distance function (is strictly positive unless $x = y$), its cost of transportation is defined as $c(x, y) = d(x, y)$ for some $x \in X$ and $y \in Y$. The actual measure spaces considered are

(i) (X, μ) for the initial pile of sand.

(ii) (Y, ν) for the hole the sand is transported to.

2.1 Monge Problem

Here is the basic introduction to the optimal transport problem, where the particles from one starting point all map to the same destination (not merged or separated), as discussed from [41, 42], by Santambrogio.

Rather than saying we are mapping content from some space X directly to some space Y, we can specify some space X equivalently to the pre-image of Y (the inverse mapping), i.e. for $X = T^{-1}(Y)$ and $Y \subset \mathbb{R}^d$:

$$\int_{T^{-1}(Y)} \mu(x)dx = \int_Y \nu(y)dy,$$

which interprets the conservation of mass, that is density $\mu(x)$ is considered in $T^{-1}(Y)$ and its newly formed density, after transportation, $\nu(y)$ is considered in Y (see Figure 1 as an illustration).

The aim is to minimise the cost of overall transportation, i.e. we minimise

$$\int_X |T(x) - x|\mu(x)dx, \tag{4}$$

whereby this integral represents the work created by transporting the original density μ.

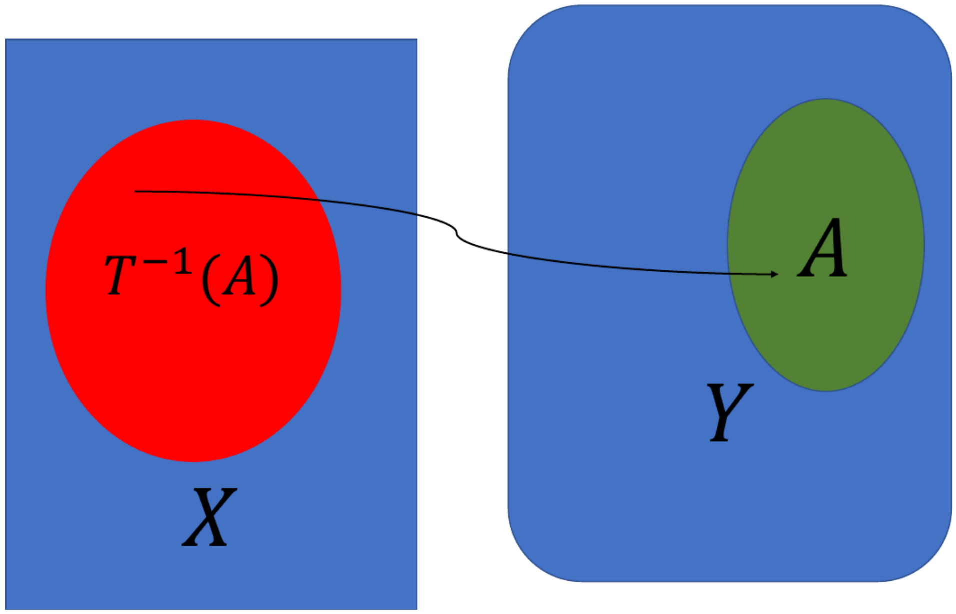


Figure 1: The optimal transport mapping from X to Y i.e. $T : A \to A$ where $A \subset Y$.

Remark 2.1. Transport map may not exist for some Monge problems, e.g. from a Dirac measure to a non-Dirac measure.

Due to this problem, this leaves us with inconclusive evidence on whether a minimiser exists [41, 42],

.

2.2 Kantorovich Problem

The Kanotorovich problem is a relaxation of the Monge problem: the problem is now considered as transport plans (here masses can be split during transportation) and also consists of generalised cost functions, i.e. it is not necessarily related to the Euclidean distance $|x - y|$. It is formulated with two probability measures $\mu(x) \in \mathcal{P}(X)$ and $\nu(y) \in \mathcal{P}(Y)$.

We consider the space X endowed with the original distribution, e.g. $X := \{x_1, x_2, \ldots, x_m\}$, e.g. the quantities of our pile of sand are originally distributed at points $x_i \in X$, $i = 1, 2 \ldots m$. This gives us quantities at each point with mass $\overline{m}_i = \overline{m}(x_i)$.

After transportation, we have a new distribution of sand particles $Y := \{y_1, y_2, \ldots y_n\}$, i.e. the quantities of our pile of sand are now relocated to points $y_j \in Y$, $j = 1, 2, \ldots, n$. During transportation, the sand particles are distributed from the original distribution X and allocated to either of the n points

$y_j \in Y$. The mass at these points is defined as $\overline{n}_j := \overline{n}(y_j)$. Note that we consider

$$X = \bigcup_{i=1}^{m} x_i, \quad \text{and} \quad Y = \bigcup_{j=1}^{n} y_j.$$

As a result of transportation, we have a cost function defined as $c(x_i, y_j)$, for transporting a unit mass from x_i to y_j. The process is described more explicitly as follows:

- A map $\gamma(x_i, y_j)$ moves distributed quantities of sand from x_i to y_j.
- Quantities of sand must be preserved during transportation i.e. $\gamma(x, y) \geq 0$ and

$$\sum_{j=1}^{n} \gamma(x_i, y_j) = \overline{m}_i, \qquad \sum_{i=1}^{m} \gamma(x_i, y_j) = \overline{n}_j. \tag{5}$$

- As a result, the total cost of transportation plans (also called transference plans) is

$$C(\gamma) = \sum_{i=1}^{m} \sum_{j=1}^{n} c(x_i, y_j)\gamma(x_i, y_j). \tag{6}$$

Here, we have portrayed the transportation problem as a linear programming structure. In this case, the optimal transport problem involves minimising the linear cost functional over the set of possible transport plans with constraint (5), that is

$$\min(C(\gamma)) := \min \Big(\sum_{i=1}^{m} \sum_{j=1}^{n} c(x_i, y_j)\gamma(x_i, y_j) \Big), \tag{7}$$

where the linear constraints ensure that the mass is preserved during transportation.

Furthermore, a transport plan (**not a map which would transport to one fixed location *only***) is considered, alternatively to the linear programming format. Here a function $\gamma(\mu, \nu)$ moves an amount of sand from distribution μ to ν. Again we say that masses can be split or an amount transported can be split and located to multiple destinations in a manner such that the plan is *optimal* (see Figure 2 as an illustration).

Definition 2.2. (Kantorovich Problem) Given two densities $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, on the probability spaces X and Y, respectively, the cost function $c : X \times Y \to \mathbb{R}_+$ and a set of transport plans from one density to the other:

$$\Gamma(\mu, \nu) := \{\gamma \in \mathcal{P}(X \times Y) : \quad (\pi_x)\gamma = \mu, (\pi_y)\gamma = \nu\}. \tag{8}$$

Here (π_x) and (π_y) are the projections from $X \times Y$ to X and to Y respectively, these measure how the particles are distributed from x to y for each pair $(x, y) \in X \times Y$. Hence the problem is to calculate

$$\min_{\gamma \in \Gamma(\mu,\nu)} \Big\{ \int_{X \times Y} c(x, y) d\gamma(x, y) : \gamma \in \Gamma(\mu, \nu) \Big\}.$$

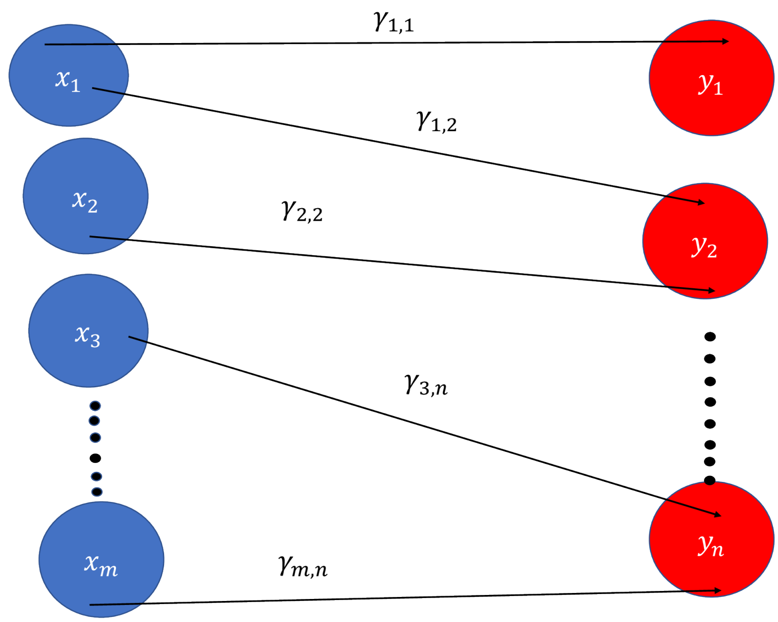


Figure 2: A transport plan splitting particles from $x_1, x_2, \ldots x_m \subset X$ to multiple destinations $y_1, y_2, \ldots y_n \subset Y$ and merging particles from within X to one destination within Y.

We note that since this allows for various destinations, the transport map T (this maps from one place to one destination *only*) from the Monge problem cannot exist for the Kantorovich problem [41, 42].

Let us discuss in more detail to the transport plan defined in (8). It consists of a measure γ defined on the product space $X \times Y$ and X is located within a set of transport plans. It takes some mass in set $A \subset X$ which is transported and distributed to a set $B \subset Y$. γ must be fixed, i.e. it is a coupling between the densities μ and ν with the two sets $A \subset X$ and $B \subset Y$ respectively. In simple terms

(i) $\gamma(A \times Y) := \mu(A)-$ mass in $A \subset X$ distributed in the whole space Y.

(ii) $\gamma(X \times B) := \nu(B)-$ mass in $B \subset Y$ distributed from the whole space X.

Therefore, as already mentioned, the cost of transport plans γ is

$$C(\gamma) := \sum_{x,y} c(x,y)\gamma(x,y) \approx \int_{X \times Y} c(x,y)d\gamma(x,y).$$

2.3 Dual Problem

The duality problem is an approach to interpreting the optimal transport problem, introduced in the previous two sections. Two scenarios are shown: ensure the cost of having a product shipped is more efficient to doing it ourselves; ensure profit is maximised, but not to over-weigh the transport cost, in order to minimise overpricing and hence maintain ourselves in competition.

2.3.1 Scenario One

Here, we briefly discuss a more deeper understanding/structure to the optimal transport problems. We briefly summarise efficient approaches for finding the optimal cost, for example in economic scenarios, starting with a scenario which was published by L. Caffarelli [9]:

Assume, rather than transporting ourselves, we hire a shipper to transport goods for us, where we just pay for loading and unloading (see [47, Thm. 1.3] by Villani). Initially, a product is charged for loading at point $x \in X$ at a price of $\phi(x)$ per unit. After transporting these purchased goods, the goods are charged for unloading at point $y \in Y$ at a price of $\psi(y)$ per unit. The idea is by hiring a shipper to not charge as much as transporting it ourselves, hence we wish for the total loading and unloading costs to not exceed transport cost (by ourselves), that is

$$\phi(x) + \psi(y) \leq c(x, y), \quad x \in X, y \in Y. \tag{9}$$

Considering all integrable functions ϕ and ψ over measures μ and ν respectively, constraint (9) and the infimum-supremum argument, the relation between the minimum Kantorovich problem and the maximum dual problem is

$$\min_{\gamma \in \Gamma(\mu,\nu)} \int_{X \times Y} c(x, y) d\gamma(x, y) = \sup_{(\phi,\psi)\in\Phi} \int_X \phi(x) d\mu + \int_Y \psi(y) d\nu, \tag{10}$$

where $\Phi = \{(\phi, \psi) \in L^1(\mu) \times L^1(\nu) : \phi(x) + \psi(y) \leq c(x, y)\}$.

2.3.2 Scenario Two

Alternatively, we can consider an alternative view of the problem (see [43] by Savaré). We assume a lorry company is contracted to transport goods. Initially, a product is purchased at point $x \in X$ at a charge of $\phi(x)$ per unit. After transporting these purchased goods, the goods are sold at point $y \in Y$ at a price of $\psi(y)$ per unit. As a company, you wish for prices to be as efficient as possible, in order to help beat competition. In this case, we can take profits to not exceed transport cost, i.e.

$$\psi(y) \leq c(x, y) + \phi(x), \quad x \in X, y \in Y. \tag{11}$$

The total profit $P(\phi, \psi)$ from selling the purchased goods after transportation is given as

$$P(\phi, \psi) := \sum_{j=1}^{n} \overline{n}(y_j)\psi(y_j) - \sum_{x=1}^{m} \overline{m}(x_i)\phi(x_i),$$

which is also classed as your gross profit, in business sense, that is profit before company overheads.

The problem now is to maximise profits, i.e. we wish to maximise $P(\phi, \psi)$ among all pairings (ϕ, ψ) satisfying the constraint (9). From the constraint (9), the total transport cost exceeds the total profit, i.e.

$$C(\gamma) \geq P(\phi, \psi).$$

To justify the last inequality, we consider a shipper who offers to charge us a loading and an unloading fee for transportation, but we would like this to be more cost efficient than transporting it ourselves, that is the cost $C(\gamma)$.

For competitiveness reasons, as mentioned above, the optimal transport plan involves finding a pair of competitive prices (ϕ, ψ) such that

$$C(\gamma) := P(\phi, \psi),$$

where ϕ and ψ represents the sales and cost of purchase (raw materials) respectively. Here this gives us maximised profits, but keeps themselves in good competition at the same time.

Let us denote the "slacking" $s(x, y)$ which measures the difference between the transport cost and profit:

$$s(x, y) = c(x, y) - (\phi(x) - \psi(y)).$$

The pairing (x, y), the starting point to the final point, is connected by optimal transport when it maximises the profit i.e.

$$c(x, y) = \psi(y) - \phi(x).$$

By the Von-Neumann Minimax Theorem (see [46] by Neumann for details), we summarise the optimal transport problem, in line with the duality problem, via a linear programming structure, i.e. we wish to find, for $\gamma_{i,j} \geq 0$:

$$\min_{\gamma} \sum_{i,j} c(x_i, y_j)\gamma(x_i, y_j) \quad \text{s.t.} \quad \sum_{j} \gamma(x_i, y_j) = \overline{m}_i, \quad \sum_{i} \gamma(x_i, y_j) = \overline{n}_j,$$

$$\min_{\gamma} \sum_{i,j} c(x_i, y_j)\gamma(x_i, y_j) = \max_{\phi,\psi} \sum_{i,j} (\psi(y_j)\overline{n}(y_j) - \phi(x_i)\overline{m}(x_i)); \quad c(x_i, y_j) - \phi(x_i) - \psi(y_j) \geq 0,$$

where we minimise the cumulative transportation cost, which is equivalent to maximising the total profit after transportation.

2.4 L^2-Wasserstein Distance $\mathcal{W}_2$

The Wasserstein distance is given, which is the minimum of the Kantorovich potential.

Now we consider a specialised form of the Kantorovich problem, which aids us with the variational formulation of higher order nonlinear partial differential equations.

Definition 2.3. (Probability Space of Measures with Mass M): We have a set of probabilities when considering some density of mass M in some metric space Ω, which is convex, and some point $x_0 \in \Omega$:

$$\mathcal{P}_M(\Omega) := \Big\{\mu \in \mathcal{P}(\Omega) : \quad \int_{\Omega} d^2(x, x_0) d\mu(x) = M\Big\}. \tag{12}$$

Here we consider the minimum value of transport problems between two probabilities. The L^2-Wasserstein distance, $\mathcal{W}_2$, also called the Monge-Kantorovich distance from Mendivil [36], is related to the Kantorovich problem, where the cost function $c(x, y)$ is defined as the second power of the distance in the metric space, i.e. it describes the distance between two densities of equal mass, distributed on the same probability space.

Definition 2.4. (Wasserstein Distance): For two measures $\mu, \nu \in \mathcal{P}_2(\Omega)$, the L^2-Wasserstein distance $\mathcal{W}_2$ is defined as

$$\mathcal{W}_2[\mu, \nu] := \inf_{\gamma \in \Gamma(\mu,\nu)} \Big(\int_{X \times Y} d(x, y)^2 d\gamma(x, y) \Big)^{\frac{1}{2}}, \tag{13}$$

with $\Gamma(\mu, \nu)$ being the collection of transference plans (8).

Remark 2.5. The finiteness of the Wasserstein distance is guaranteed as a result of (12) containing the two measures $\mu, \nu \in \mathcal{P}(\Omega)$, in other words, via the Young's inequality (see [41, p. 159-161] by Santambrogio):

$$d^2(x, y) \leq 2(|x|^2 + |y|^2) \Rightarrow \mathcal{W}_2[\mu, \nu]^2 \leq \int_{X \times Y} d^2(x, y) d\gamma \leq 2\Big(\int_X |x|^2 d\mu + \int_Y |y|^2 d\nu \Big) < \infty.$$

To be more specific, the Wasserstein distance is computed by finding the (horizontal) distance between x and y (the minimal cost of transferring a unit mass at x to destination y), where its start and finish points, x and y, respectively, are such that the areas under the graphs of densities μ and ν, as defined from the Monge's problem, are equal at x and y.

Lemma 2.6. ***[41, Lem. 5.3-5.4]****: The L^2-Wasserstein distance satisfies the triangle inequality.*

The next part of our introduction brings us towards the scheme that enables us to construct our new numerical methods for finding a discrete approximation solution for our selected PDEs in the case of Wasserstein spaces.

3 Partial Differential Equations as Gradient Flows

We introduce the main useful concepts for the variational form of solving higher order nonlinear PDEs, the Wasserstein gradient flow problem. We consider within the probability space of measures, but a brief introduction is first given within the finite dimensional space with an example of L^2-gradient flows for the well known Heat equation.

Before we discuss our discretisation approach for partial differential equations (PDEs), we first consider the simple, finite-dimensional gradient flow problem in the Euclidean Space of smooth functionals $F : \mathbb{R}^d \to \mathbb{R}$ (see [42, Sect. 2] by Santambrogio), i.e. solving the problem

$$\frac{dx(t)}{dt} = -\nabla F(x(t)), \quad t > 0, \quad x(0) = x_0, \tag{14}$$

where $F \in C^{1,1}(\mathbb{R}^d)$, which means that ∇F is Lipschitz continuous, i.e. there exists a constant $C > 0$ such that (see [38] by Searcóid)

$$|\nabla F(x) - \nabla F(y)| \leq C|x - y| \quad \forall x, y \subset \mathbb{R}^d.$$

Lipschitz continuity guarantees a unique solution to the problem in question (see e.g. [41, Chpt. 8]).

Example 3.1. *In the Euclidean space $(\mathbb{R}^d, |\cdot|)$, for the energy functional $F_1(x) = \frac{1}{2}x^2$, we have that $\nabla F_1 = x(t)$ giving us the gradient flow of F_1 as*

$$\frac{dx(t)}{dt} = -x(t),$$

and the unique solution of the gradient flow as

$$x(t) = x_0 e^{-t}.$$

Example 3.2. *As discussed in [42, p. 56], the Heat equation $\partial_t u(x,t) = \Delta u(x,t)$ is the gradient flow in the L^2 space (L^2-gradient flow) for the Dirichlet energy $\int_{L^2(\mathbb{R}^d)} |\nabla u(x,t)|^2 dx$. The gradient of $\mathcal{E}(u)$ is $-\Delta u$. In this case, this considers equations of the form $\partial_t u = -\frac{\delta \mathcal{E}(u)}{\delta u}$ in general.*

The minimising movement scheme was originally considered on the finite-dimensional Euclidean space $\mathbb{R}^d$ (finite dimensional terms) by De Georgi [14]. Jordan, Kinderlehrer and Otto [19] and then [41, Chpt. 7] extended the minimising movement scheme to the Wasserstein case, where certain functionals $\mathcal{E} : \mathcal{P}_M(\Omega) \to \mathbb{R} \cup \{\infty\}$ are defined on a space of probability measures and with mass M. The scheme in this case is also referred to as the JKO scheme, in recognition of the authors' work.

We now discuss how several well-known evolution PDEs can be interpreted as curves of steepest descent on the L^2-Wasserstein space. This is one of the most spectacular applications of optimal transport and Wasserstein distances.

The main idea is that we consider an energy functional $\mathcal{E}(\cdot)$ on the probability space $\mathcal{P}_M(\Omega)$ of mass M. Gradient flows are also defined as steepest descent curves, where we investigate the trajectory $u(x,t)$ at time points that minimises the energy $\mathcal{E}(\cdot)$ (this is roughly the similar idea to the traditional method, for calculating critical points, where we find u such that $\nabla_{\mathcal{W}_2}\mathcal{E}(u) = 0$). However, we consider a curve that commences at u_0, the initial point, and travels in the direction of maximum decrease of $\mathcal{E}(\cdot)$ (it minimises $\mathcal{E}(\cdot)$ as fast as possible). It can be shown that the gradient $\nabla_{\mathcal{W}_2}\mathcal{E}(\cdot)$ travels in the direction of the maximum increase of $\mathcal{E}(\cdot)$, therefore in simple terms, the gradient flow solves equation (14).

Before moving on to Wasserstein gradient flows, we emphasise that diffusion or evolution equations help assist us with explaining motions of things over time e.g. particles, transport, objects or us.

3.1 L^2-Wasserstein Gradient Flows - Introduction

We derive the L^2-Wasserstein gradient flows, which are interpreted in relation to the continuity equation.

We now consider the gradient flow theory for general metric spaces, and specifically for the Wasserstein space. Its main application nowadays is to characterise evolution PDEs in the probability space. Jordan, Kinderlehrer and Otto characterised well-known equations including the Heat and Fokker-Planck equations with respect to the L^2-Wasserstein metric on probability spaces. The theory has been further developed by [2].

Ambrosio, Gigli and Savaré [2] explored gradient flows in general metric spaces, particularly in the probability space as above. The theory of gradient flows and optimal transport are linked to the study of metric spaces in $\mathcal{P}_M(\Omega)$ supplied by the L^2-Wasserstein metric $\mathcal{W}_2$.

A distance of two measures are considered on a probability space of measures with mass M, defined as $\mathcal{P}_M(\Omega)$, i.e. Wasserstein distance, which is a key output from optimal transport theory, explained in Section 2. Hence we can consider gradient flows of various energy functionals on the Wasserstein (metric) space.

Indeed, by the end of the 1990s, the new century saw an interesting adaptation to gradient flows, introduced by Jordan, Kinderlehrer and Otto [26]. They derived minimising movement schemes, semi-discrete form of gradient flows, which we will discuss shortly. They obtained an alternative formulation of continuity equations (see Peletier's article [40] also):

$$\frac{\partial u(x,t)}{\partial t} = \nabla \cdot \Big(u(x,t)\nabla\Big(\frac{\delta\mathcal{E}(u)}{\delta u}\Big)\Big), \tag{15}$$

whereby the density $u(x,t)$ is considered over a probability space with given domain Ω and mass M, i.e. $\mathcal{P}_M(\Omega)$.

In other words, (15) represents the gradient flow of $\mathcal{E}(\cdot)$ with respect to the L^2-Wasserstein metric. We will explain shortly how the energy comes into play.

From optimal transport theory, the Wasserstein distance supplies a nonlinear metric structure. Indeed, from the introduction of optimal transport, evolution equations (PDEs) with the underlying gradient flow structure (15) have been linked to the continuity equation:

$$\partial_t u(x,t) = -\nabla \cdot (uv).$$

Here v is denoted as the velocity, where Savaré [44, p. 9] shows the link to (15), by selecting a form such that it minimises the energy as fast as possible, i.e. they select $v = -\nabla\left(\frac{\delta\mathcal{E}(u)}{\delta u}\right)$ as well as providing the nonlinear part of the equation.

Remark 3.3. The Wasserstein gradient is denoted by

$$\nabla_{\mathcal{W}_2}\cdot = \nabla \cdot \left(u\nabla\left(\frac{\delta\cdot}{\delta u}\right)\right). \tag{16}$$

3.2 L^2-Wasserstein Space Theory

Along with a Wasserstein gradient flow example for the well known Heat equation, the subsection briefly mentions the theory, in relation to gradient flows, which will be applied to our convergence proof contribution (Sections 5 to 7). A proposition is given, showing how convex and semi-convex assumptions to the associated energy for gradient flow problems provide us with a well-posed problem.

We discuss the extension on the gradient flow characterisations, for example, evolution variational inequalities (EVI) and energy dissipation equality (EDE), where we explore the general theory of our metric space: we will shortly derive the metric derivative, which considers a curve $u : \Omega \times [0,T] \to \mathcal{P}_M(\Omega)$ defined in a Wasserstein space. The metric derivative illustrates the velocity $\partial_t u(x,t)$ in terms of a Wasserstein space i.e. the speed, rather than just a vector. The formula for this is as follows:

$$|u'|(t) = \lim_{\tau\to 0} \frac{\mathcal{W}_2[u(x,t), u(x,t+\tau)]}{|\tau|}.$$

Definition 3.4. (λ-Convex Functionals, see [35, p. 7]): The functional $\mathcal{E}(\cdot)$ is λ-convex or semi-convex of modulus λ if, for all $s \in (0,1)$

$$\mathcal{E}(u(x,t)) \leq (1-s)\mathcal{E}(u(x,0)) + s\mathcal{E}(u(x,1)) - \frac{\lambda}{2}s(1-s)\mathcal{W}_2[u(x,0), u(x,1)]^2,$$

where the curve $u(x,t)$ connects the two end points $u(x,0)$ and $u(x,1)$.

Many PDEs, ranging from first to fourth order in time, can be represented as gradient flows. Examples include the Heat equation and the Porous Media equation, as well as the Cahn Hilliard equation and further fourth order equations. The energy $\mathcal{E}(\cdot)$ will be considered for gradient flows in probability space.

Aim: To minimise $\mathcal{E}(\cdot)$ from the initial point u_0 of a curve $u(x,t)$ as quickly as possible (see [42]).

We first mention an L^2-Wasserstein gradient flow example for the Heat equation, but in slightly more detail:

Example 3.5. *The Heat equation:*

$$\partial_t u(x,t) = \Delta u(x,t),$$

is the Wasserstein gradient flow of the energy functional $\mathcal{E}_2(u) = \int_{\mathcal{P}_M(\Omega)} u(x,t)\log(u(x,t))dx$*, see [42, p. 56].*

It is important to note that, with respect to (14), convexity of the energy functional $\mathcal{E}(\cdot)$ is a key property for generating a well-posed gradient flow problem:

$$\partial_t u(x,t) = -\nabla_{\mathcal{W}_2}\mathcal{E}(u(x,t)), \quad u(x,0) = u_0. \tag{17}$$

We demonstrate how convexity can generate a well-posed problem. In fact, well-posedness of the Wasserstein gradient flow problem (17) is shown by convexity and semi-convexity of $\mathcal{E}(\cdot)$, from [41, Prop. 8.1], and generating a unique solution of (17).

Proposition 3.6. ***[41, Prop. 2.1]****: Let $\mathcal{E}(\cdot)$ be a convex energy functional. Then the problem (17) has a unique solution.*

Proof. Let u_1, u_2 be two solutions to the problem (17) with the same initial data. Consider the function $g(t) = \frac{1}{2}(u_1(t) - u_2(t))^2$. Using (17) we have

$$\begin{aligned} \frac{dg(t)}{dt} &= (u_1(t) - u_2(t)) \cdot \Big(\partial_t u_1(t) - \partial_t u_2(t)\Big) \\ &= -\, (u_1(t) - u_2(t)) \cdot (\nabla_{\mathcal{W}_2}\mathcal{E}(u_1(t)) - \nabla_{\mathcal{W}_2}\mathcal{E}(u_2(t))). \end{aligned} \tag{18}$$

From the fact that $\mathcal{E}(\cdot)$ is convex, the basic property of convex functions give

$$(u_1(t) - u_2(t)) \cdot (\nabla_{\mathcal{W}_2}\mathcal{E}(u_1(t)) - \nabla_{\mathcal{W}_2}\mathcal{E}(u_2(t))) \geq 0. \tag{19}$$

Substituting (19) into (18) gives

$$\frac{dg(t)}{dt} \leq 0 \Rightarrow g(t) \leq g(0).$$

By considering two solutions of (17), denoted as $u_1(x,t)$ and $u_2(x,t)$, we have

$$u_1(x,0) - u_2(x,0) = u_0 - u_0 = 0,$$

giving us $g(0) = 0$ and hence

$$g(t) = \frac{1}{2}(u_1(t) - u_2(t))^2 \leq 0.$$

Thus, $u_1(t) = u_2(t)$ for all $t \geq 0$ since a metric cannot be non-positive, thus we have uniqueness of the solution. $\square$

Proposition 3.7. *[41, Rmk. 2.1]: Uniqueness and stability can also be derived if the weaker condition of semi-convexity of $\mathcal{E}(\cdot)$ is considered.*

Proof. Semi-convexity is defined as λ-convex for some $\lambda \in \mathbb{R}$, i.e. $\mathcal{E}(u) - \frac{\lambda}{2}$ is convex.

This gives us an adapted version of (22):

$$\begin{aligned} &(u_1(t) - u_2(t)) \cdot (\nabla_{\mathcal{W}_2}\mathcal{E}(u_1(t)) - \nabla_{\mathcal{W}_2}\mathcal{E}(u_2(t)) - \lambda(u_1(t) - u_2(t))) \geq 0 \\ \Rightarrow &(u_1(t) - u_2(t)) \cdot (\nabla_{\mathcal{W}_2}\mathcal{E}(u_1(t)) - \nabla_{\mathcal{W}_2}\mathcal{E}(u_2(t))) \geq \lambda(u_1(t) - u_2(t))^2, \end{aligned}$$

and via the Gromwall's lemma and square rooting both sides:

$$\begin{aligned} &\frac{dg(t)}{dt} \leq -2\lambda g(t) \quad \Rightarrow \quad g(t) \leq g(0)\exp(-2\lambda t) \\ \Rightarrow \quad &(u_1(x,t) - u_2(x,t))^2 \leq (u_1(x,0) - u_2(x,0))^2 \exp(-2\lambda\tau). \end{aligned}$$

If $\lambda > 0$, this gives us $|u_1(x,t) - u_2(x,t)| \leq |u_1(x,0) - u_2(x,0)|$ implying convergence to the unique minimiser of $\mathcal{E}(\cdot)$ over time. $\square$

A semi-discretisation in time for these L^2-Wasserstein gradient flow problems is given through the minimising movement scheme, as explained in the next section.

3.3 Minimising Movement Scheme

The minimising movement scheme is introduced, where a key result is shown. Here, the monotonicty of the energy at the time-discrete level for BDF1 is shown, giving us a finite velocity and thus a well-posed problem (particularly uniqueness).

The semi-discrete in time form, of gradient flows was introduced as the minimising movement scheme from De Georgi [14]. The evolution equations, which can be variationally formulated, in order to solve as a Wasserstein gradient flow problem (17), the Cauchy problem, brings out some interesting features. Introduced by De. Georgi, the authors Düring, Matthes et al. [17, p. 260], as well as Jordan, Kinderlehrer et al. worked with the minimising movement scheme, i.e. for a small enough time step size $\tau > 0$, we construct a sequence of points $(u_\tau^n)_{n\in\mathbb{N}}$ such that

$$u_\tau^n \in \operatorname*{argmin}_{u\in\mathcal{P}_M(\Omega)} \Phi^\tau(u_\tau^{n-1}; u), \qquad \Phi^\tau(u_\tau^{n-1}; u) = \frac{1}{2\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u). \tag{20}$$

By generating a discrete solution at the next time step u_τ^n, which minimises $\frac{1}{2\tau}\mathcal{W}_2[u^{n-1}, u]^2 + \mathcal{E}(u)$ as much as possible, we have that the minimiser u_τ^n satisfies

$$\nabla_{\mathcal{W}_2}\left(\mathcal{E}(u) + \frac{\mathcal{W}_2[u_\tau^{n-1}, u]^2}{2\tau}\right)\Big|_{u=u_\tau^n} = 0,$$

which is equivalent to the implicit Euler scheme (or the first backward difference formula (BDF1) scheme), see [42, p. 6] for details. This scheme guarantees strong stability properties.

The fact that $\mathcal{E}(\cdot)$ is assumed to be convex is not necessary, as milder assumptions on $\mathcal{E}(\cdot)$ (lower semi-continuous, semi-convexity and some lower bounds e.g. coercivity) are sufficient enough to give a well-posed problem. For this thesis, we will assume λ-convexity of $\mathcal{E}(\cdot)$, a weaker property of convexity, and will clearly show how this delivers the convergence results of our discrete solutions to (17).

Remark 3.8. We will discuss the benefit of assuming semi-convex energy functionals, but first we mention about subdifferential functions. In the case when $\mathcal{E}(\cdot)$ is not differentiable [42, p. 5], we consider the subdifferential of λ-convex functions i.e. $p \in \overline{\partial}\mathcal{E}(u)$ such that

$$\mathcal{E}(u_*) \geq \mathcal{E}(u) + p \cdot (u_* - u) + \frac{\lambda}{2}\mathcal{W}_2[u, u_*]^2. \tag{21}$$

Note that not all λ-convex functions are differentiable e.g. $f(x) = |x|$ is convex but not differentiable at 0:

- If $\lambda > 0$, the semi-convex property is strengthened to convex, since $\frac{\lambda}{2}\mathcal{W}_2[u, u_*]^2 > 0$ as a result. Hence we revise the subdifferential definition as a result i.e. there exists $p \in \overline{\partial}\mathcal{E}(u)$ such that

  $$\mathcal{E}(u_*) \geq \mathcal{E}(u) + p \cdot (u_* - u), \tag{22}$$

  implying strict convexity of $\mathcal{E}(\cdot)$ for all time step sizes $\tau > 0$.

- If $\lambda < 0$ however, we are only guaranteed the weaker condition, although a sufficiently small τ can give strict convexity. This will be clear later on that $u \to u_*$ at the time discrete level as $\tau \to 0$ hence the latter term of (21) can be neglected as a result.

Let the curve $u(x, t)$ evaluated at time points $t = 0, \tau, 2\tau, \ldots, k\tau, \ldots$ be defined as the sequence of points from (20).

There is a connection between the optimality conditions of the minimisation problem and $\mathcal{E}(\cdot)$. We easily note that the BDF1 penalisation (20) is $\left(\lambda + \frac{1}{2\tau}\right)$-convex, and the scheme (20) is equivalent to the fully stable implicit Euler scheme:

$$\frac{u_\tau^n - u_\tau^{n-1}}{\tau} = -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n),$$

whereby the left hand side of (20) is the time-discrete form of $\partial_t u(x, t)$.

Let us discuss a bit more about the Euler schemes. They are temporal discretisations where the derivative $\partial_t u(x, t)$ is approximated in line with the finite difference method.

See Santambrogio [42, p. 7-10] for details of the following.

Example 3.9. *We refer back to the Wasserstein gradient flow problem (17), with the initial data* $u(x, 0) = u_0$ *and the associated implicit Euler scheme:*

$$\frac{u_\tau^n - u_\tau^{n-1}}{\tau} = -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n) \quad \Rightarrow \quad u_\tau^n = u_\tau^{n-1} - \tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n). \tag{23}$$

The implicit Euler method enjoys strong stability properties (see [23, 24] from Hairer, Wanner). In addition, from earlier discussion by [43, p. 4], we know that $\mathcal{E}(u(x,t))$ decreases over time.

For the main contribution of the thesis, we can prove that the sequence of solutions to (20) $(u_\tau^n)_{n\in\mathbb{N}}$ converges to an interpolated solution over time $\overline{u}_\tau$ for sufficiently small time step size τ i.e. as $\tau \to 0$. Note that we consider the curve $\overline{u}_\tau$ mapped from $\Omega \times [0,T]$ to $\mathcal{P}_M(\Omega)$.

We define the curve $u_\tau(t) = u_\tau^{n+1}$ and a piecewise interpolated solution as $\overline{u}_\tau(t) = u_\tau^{n-1} - (t-(n-1)\tau)\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n)$ for $t \in ((n-1)\tau, n\tau]$.

By the fact that u_τ^n minimises $\Phi^\tau(u_\tau^n; u)$, we have that

$$\mathcal{E}(u_\tau^n) + \frac{\mathcal{W}_2[u_\tau^n, u_\tau^{n-1}]^2}{2\tau} \leq \mathcal{E}(u_\tau^{n-1}),$$

and provided $\mathcal{E}(\cdot)$ is bounded, summing from $n=1$ to $n=N$ (N is the maximum number of time grid intervals) gives us

$$\sum_{n=1}^{N} \frac{\mathcal{W}_2[u_\tau^n, u_\tau^{n-1}]^2}{2\tau} \leq \left(\mathcal{E}(u_\tau^0) - \mathcal{E}(u_\tau^N)\right) \leq C. \tag{24}$$

We can rearrange the metric in terms of a metric derivative as follows:

$$\frac{\mathcal{W}_2[u_\tau^n, u_\tau^{n-1}]^2}{2\tau} = \tau\Big(\frac{\mathcal{W}_2[u_\tau^n, u_\tau^{n-1}]}{2\tau}\Big)^2 = \int_{(n-1)\tau}^{n\tau} |(\overline{u}_\tau)'|(t)^2 dt, \tag{25}$$

and summing (25) from $n=1$ to $n=N$ gives

$$\frac{1}{2}\int_0^T |(\overline{u}_\tau)'|(t)^2 dt \leq C. \tag{26}$$

Remark 3.10. The inequality (26) comes from the result (25).

By considering the Wasserstein distance (you could consider the basic properties of metric spaces), and the Cauchy-Schwartz inequality, the Wasserstein distance between a solution at different time steps is bounded and converges for decreasing time step τ, when taking the result from (24) (for $s < t$, $s \in ((m-1)\tau, m\tau]$ and $t \in ((n-1)\tau, n\tau]$). Furthermore, we consider piecewise constant solutions as

$$\overline{u}_\tau(t) = u_\tau^n, \qquad t \in ((n-1)\tau, n\tau]\,.$$

Therefore, we have as a result of (24) as well as the triangle and Cauchy-Schwarz inequalities:

$$\begin{aligned}
&\mathcal{W}_2[\overline{u}_\tau(t), \overline{u}_\tau(s)] \\
\leq\;& \mathcal{W}_2[\overline{u}_\tau(t), \overline{u}_\tau(t-\tau)] + \mathcal{W}_2[\overline{u}_\tau(t-\tau), \overline{u}_\tau(t-2\tau)] + \cdots + \mathcal{W}_2[\overline{u}_\tau(s+\tau), \overline{u}_\tau(s)] \\
=\;& \sum_{j=m+1}^{n} \mathcal{W}_2[u_\tau^{j-1}, u_\tau^j] \leq \left(\sum_{j=m+1}^{n} \mathcal{W}_2[u_\tau^{j-1}, u_\tau^j]^2\right)^{1/2} |n-m+1|^{1/2} \\
\leq\;& \left(\sum_{j=m}^{n} \mathcal{W}_2[u_\tau^{j-1}, u_\tau^j]^2\right)^{1/2} \left(\frac{t-s}{\tau} + 2\right)^{1/2} \leq (C\tau)^{1/2}\left(\frac{t-s}{\tau} + 2\right)^{1/2} \\
=\;& C^{1/2}\left(t-s+2\tau\right)^{1/2}.
\end{aligned} \tag{27}$$

Therefore, as the time step $\tau \to 0$, the two time points converge i.e. $|s-t| \to 0$ and thus $u_\tau(t)$ converges to the unique solution of problem (17).

3.4 Characterising Limit Curves

Characterisation forms of the problem are discussed, particularly the evolution variational inequality, as a result of semi-convexity assumptions. This shows us how the formulation of this should provide us with a well-posed problem.

Once we have shown that a sequence of discrete solutions $(u_\tau^n)_{n\in\mathbb{N}}$ converge to a limit curve, an approach is required to describe the resulting limit curve. As we have recently explained, the semi-discrete form of a gradient flow is equipped with metrics. However, quite clearly, the Cauchy problem (14) has no meaning in terms of a Wasserstein (metric) space.

We discuss a couple of approaches, (see [42, p. 12-13]), which make the assumption of convexity (and/or semi-convexity), to derive an alternative, but equivalent form of (14). This time however as a result, it possesses metric counterparts but also the latter maintains the well-posed state of our problem.

The authors in [42, Prop. 3.1] prove how two curves satisfying the energy dissipation equality (EDE) or the evolution variational inequality (EVI) generate a well-posed problem for almost identical initial data. This is shown now:

3.4.1 Energy Dissipation Equality (EDE)

Young's inequality and the chain rule gives us the difference between the energy functional terms $\mathcal{E}$ at two arbitrary time points t, s, i.e.

$$\begin{aligned}\mathcal{E}(u(x,s)) - \mathcal{E}(u(x,t)) &= -\int_s^t \partial_r |\nabla_{\mathcal{W}_2}\mathcal{E}(u(x,r))| dr \\ &= \int_s^t -\nabla_{\mathcal{W}_2}\mathcal{E}(u(x,r)) \cdot \partial_r u(x,r) dr \\ &\leq \frac{1}{2}\int_s^t \Big(|\partial_r u(x,r)|^2 + |\nabla_{\mathcal{W}_2}\mathcal{E}(u(x,r))|^2 \Big) dr.\end{aligned} \tag{28}$$

If this is considered for (17) this gives us the energy dissipation equality (EDE), that is the inequality in (28) is an equality:

$$\mathcal{E}(u(x,s)) - \mathcal{E}(u(x,t)) = \int_s^t |\nabla_{\mathcal{W}_2}\mathcal{E}(u(x,r))|^2 dr = \frac{1}{2}\int_s^t \Big(|\partial_r u(x,r)|^2 + |\nabla_{\mathcal{W}_2}\mathcal{E}(u(x,r))|^2 \Big) dr.$$

Furthermore the right hand term of (28) is equivalent to $\mathcal{E}(u(x,s)) - \mathcal{E}(u(x,t))$ and the middle term is equivalent to $\frac{1}{2}\int_s^t \Big(|\partial_r u(x,r)|^2 + |\nabla_{\mathcal{W}_2}\mathcal{E}(u(x,r))|^2 \Big) dr$. Hence the inequality (28) can be reversed, providing us with an equivalent definition of the gradient flow:

$$\mathcal{E}(u(x,s)) - \mathcal{E}(u(x,t)) \geq \frac{1}{2}\int_s^t \Big(|\partial_r u(x,r)|^2 + |\nabla_{\mathcal{W}_2}\mathcal{E}(u(x,r))|^2 \Big) dr = \int_s^t |\nabla_{\mathcal{W}_2}\mathcal{E}(u(x,r))|^2 dr.$$

3.4.2 Evolution Variational Inequality (EVI)

Another characterisation to gradient flows is the inequality when $\mathcal{E}$ is λ-convex, see (22) for all $u_* \in \mathcal{P}_M(\Omega)$ and the characterised vector $p \in \overline{\partial}\mathcal{E}(u)$ (or $\nabla_{\mathcal{W}_2}\mathcal{E}(u)$ if $\mathcal{E} \in C^1$).

This is a main ingredient for a unique and stable solution which we will work to in this thesis.

If $\mathcal{E}(\cdot)$ is λ-convex, but not convex, then the gradient p is portrayed from the inequality (21).

By selecting an arbitrary curve $u(x,t)$, since

$$\frac{1}{2}\partial_t \mathcal{W}_2[u_*, u(x,t)]^2 = \mathcal{W}_2[u_*, u(x,t)] \cdot (-\partial_t u(x,t)),$$

we have by substituting into (21):

$$\frac{1}{2}\partial_t \mathcal{W}_2[u(x,t), u_*]^2 \leq \mathcal{E}(u_*) - \mathcal{E}(u) - \frac{\lambda}{2}\mathcal{W}_2[u(x,t), u_*]^2. \tag{29}$$

This gives us a second alternative definition of our original gradient flow problem (17). The inequality (29) is called the evolution variational inequality (EVI).

We now show uniqueness and stability, obtained by (29): for two different curves $u_1(x,t)$ and $u_2(x,s)$ satisfying (17) with similar initial data, we reverse the order of the energy functional terms $\mathcal{E}(u_1(x,t))$ and $\mathcal{E}(u_2(x,s))$ (by the definition of metric spaces, it is clear that $\mathcal{W}_2[u_1(x,t), u_2(x,s)]^2 = \mathcal{W}_2[u_2(x,s), u_1(x,t)]^2$), retrieving two similar looking inequalities:

$$\frac{1}{2}\partial_t \mathcal{W}_2[u_1(x,t), u_2(x,s)]^2 \leq \mathcal{E}(u_2(x,s)) - \mathcal{E}(u_1(x,t)) - \frac{\lambda}{2}\mathcal{W}_2[u_1(x,t), u_2(x,s)]^2, \tag{30a}$$

$$\frac{1}{2}\partial_s \mathcal{W}_2[u_1(x,t), u_2(x,s)]^2 \leq \mathcal{E}(u_1(x,t)) - \mathcal{E}(u_2(x,s)) - \frac{\lambda}{2}\mathcal{W}_2[u_1(x,t), u_2(x,s)]^2. \tag{30b}$$

By considering $(s,t) \to \mathcal{W}_2[u_1(x,t), u_2(x,s)]^2$ and restricting along the curve (s,t) when $s = t$, this gives us from (30a)+(30b):

$$\frac{d\mathcal{E}_E(t)}{dt} \leq -2\lambda \mathcal{E}_E(t),$$

where $\mathcal{E}_E(t) := \mathcal{W}_2[u_1(x,t), u_2(x,t)]^2$.

From Gronwall's lemma, this gives us

$$\mathcal{W}_2[u_1(x,t), u_2(x,t)]^2 \leq C \exp(-2\lambda t),$$

for some $C > 0$. This yields that, since $\mathcal{E}_E(t)$ is non-negative, it converges to zero as time tends to infinity. But in fact C is $\mathcal{W}_2[u_1(x,0), u_2(x,0)]^2$ which is zero due to the similarity of initial data. Hence $u_1(x,t)$ is a unique and stable solution of the gradient flow problem.

We expect that convexity should be enough to prove uniqueness of the gradient flow problem (17), from what we witnessed from the last section. Some of the main points on the uniqueness theory for general metric spaces include that some gradient flows with the EVI characterisation is also for the energy dissipation inequality (EDE), but the latter is not enough to guarantee uniqueness of the problem i.e. the EDE characterisation does not consider λ-convexity of the energy.

Showing the existence of curves satisfying the EVI (under some additional assumptions) is then the main issue for gradient flow theory with uniqueness, and hence part of our main thesis contribution.

3.5 Nonlinear Diffusion Equations as L^2-Wasserstein Gradient Flows

This section is concluded by highlighting the nonlinear PDEs we will be implementing in this thesis, as well as some detailed deduction of how the Wasserstein gradient flow of various energies relate to these, from the lemmas.

In this part, we recall the scheme that enables us to construct our new numerical methods for finding a discrete approximation for solutions of our selected PDE in the case of Wasserstein spaces.

We consider the minimising movement scheme on the probability space of measures with mass M, $\mathcal{P}_M(\Omega)$, from Düring and Matthes [17]:

$$u_\tau^n := \underset{u \in \mathcal{P}_M(\Omega)}{\operatorname{argmin}} \frac{1}{2\tau} \mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u),$$

where $\mathcal{E} : \mathcal{P}_M(\Omega) \to \mathbb{R}$ is the energy functional associated to the gradient flow on the probability space $\mathcal{P}_M(\Omega)$.

Remark 3.11. As you probably noticed, it is the similar layout to the Euclidean case [14, 17, 26, 41, 42], when we are in the Euclidean space but the metric term is adjusted for the Wasserstein case here.

We introduce notation for certain functionals $\mathcal{E} : \mathcal{P}_M(\Omega) \to \mathbb{R} \cup \{\infty\}$ on probability spaces with mass M:

- **Integral of function of density**: $\mathcal{F}(u) = \int_\Omega f(u(x))dx$ from [17] as the energy functional ("Fisher Information" for the DLSS equation).
- **Integral of potential energy**: $\mathcal{V}(u) = \int_\Omega V(x)du$ - $\mathcal{V}$ is used in addition for constructing the scheme of the linear Fokker-Planck equations.

The interest of studying evolution equations of the form (14), with underlying gradient flow structures, was ignited by recent articles, particularly on linear Fokker-Planck equations, [19], which was put into gradient flow form via the variational formulation with respect to the L^2-Wasserstein metric [19, 40, 41]:

(i) **Heat equation**: $\partial_t u(x,t) := \Delta u(x,t)$ is the gradient flow of $\mathcal{E}_h(u) := \int_\Omega u(x,t)\log(u(x,t))dx$.

(ii) **Linear Fokker-Planck equation**: $\partial_t u(x,t) - \Delta u(x,t) - \nabla \cdot (u(x,t)\nabla V(x)) = 0$ is the gradient flow of $\mathcal{E}_l(u) := \int_\Omega u(x,t)\log(u(x,t))dx + V(x)\partial_x u(x,t)dx$.

(iii) **Porous Medium equation**: $\partial_t u(x,t) - \Delta((u(x,t))^m) - \nabla \cdot (u(x,t)\nabla V(x,t)) = 0$ is the gradient flow of $\mathcal{E}_p(u) := \dfrac{1}{m-1}\int_\Omega (u(x,t))^m dx + \int_\Omega V(x,t)u(x,t)dx$ for some exponent $m > 1$ and given potential V.

Remark 3.12. For the Heat equation $\partial_t u(x,t) = \Delta u(x,t)$ (refer back to Example 3.6), the L^2-gradient flow for the Dirichlet energy $\int_{L^2(\mathbb{R}^d)} |\nabla u(x,t)|^2 dx$ is equivalent to the L^2-Wasserstein gradient flow of the relative entropy:

$$\int_\Omega u(x,t)\log(u(x,t))dx.$$

In addition, higher order nonlinear equations are of interest, like the DLSS equation, which we aim to fully discretise [17, 34].

For our thesis, we consider four different fourth order nonlinear partial differential equations with underlying Wasserstein gradient flow structures. As a first example, consider the Derrida-Lebowitz-Speer-Spohn (DLSS) equation:

$$\partial_t u(x,t) = -2\partial_x\Big(u(x,t)\partial_x\Big(\frac{\partial_x^2(\sqrt{u(x,t)})}{\sqrt{u(x,t)}}\Big)\Big), \quad t>0, \quad x\in(0,1). \tag{31}$$

The DLSS equation allows a variational formulation with respect to the L^2-Wasserstein metric, i.e. it has been generated [21, 40] (and [41, Sect. 8] for calculation) as the Wasserstein gradient flow of the Fisher Information $\mathcal{E}_f$:

$$\mathcal{E}_f(u) := \frac{1}{2}\int_\Omega u(x,t)\partial_x(\log u(x,t))^2 dx.$$

The following lemma explains the computation that leads from the gradient flow (see (15) for this) to the actual DLSS equation (31):

Lemma 3.13. *The variational formulation i.e. gradient flow formulation, involves the construction of the "minimising movement scheme" as seen in [17]. Furthermore, the variational formulation of the DLSS equation is equivalent to the L^2-Wasserstein gradient flow of the energy functional which is, for $\Omega\subset\mathbb{R}^d$:*

$$\mathcal{E}_f(u) = \frac{1}{2}\int_\Omega u(x,t)\partial_x(\log(u(x,t)))^2 dx = 2\int_\Omega\Big(\partial_x\big(\sqrt{u(x,t)}\big)\Big)^2 dx. \tag{32}$$

Proof. We consider the Wasserstein gradient $\nabla_{\mathcal{W}_2}(\cdot)$, which is defined as from the energy functional (information for this equation) $\mathcal{E}_f(\cdot)$ [40], [41, Sect. 8.2]:

$$\nabla_{\mathcal{W}_2}\mathcal{E}_f(u) = -\nabla\cdot\Big(u\nabla\frac{\delta\mathcal{E}_f}{\delta u}\Big).$$

For simplicity, we are considering the one-dimensional case i.e.

$$\nabla_{\mathcal{W}_2}\mathcal{E}_f(u) = -\partial_x\Big(u\partial_x\frac{\delta\mathcal{E}_f}{\delta u}\Big). \tag{33}$$

Firstly, we have for $\dfrac{\delta\mathcal{E}_f}{\delta u}$:

$$\frac{\delta\mathcal{E}_f}{\delta u} = 4\Big(\frac{\partial\sqrt{u}}{\partial x}\Big)\frac{\partial}{\partial u}\Big(\frac{\partial\sqrt{u}}{\partial x}\Big) = 4\Big(\frac{\partial\sqrt{u}}{\partial x}\Big)\frac{\partial^2\sqrt{u}}{\partial x^2}\frac{\partial x}{\partial u} = 4\frac{\partial\sqrt{u}}{\partial u}\frac{\partial^2\sqrt{u}}{\partial x^2} = 2\frac{\partial_x^2\sqrt{u}}{\sqrt{u}}.$$

Hence substituting into (33) gives us (31). $\square$

Can other higher order nonlinear partial differential equations also be variationally formulated?

We also have other fourth order nonlinear equations, for example, we extend our work to the full discretisation of other equations of fourth order.

We also ask the same question of other fourth order equations such as the Cahn Hilliard equation or the Thin Film equation, whereby results are in progress. We present such equations here:

- **Thin Film equation**: As explained at the beginning of [10, 33] and [30, Thm. 3.9, 3.10], the Thin Film equation is

$$\partial_t u(x,t) = -\partial_x(u(x,t)\partial_x^3 u(x,t)), \tag{34}$$

which is the Wasserstein gradient flow of

$$\mathcal{E}_t(u) := \frac{1}{2}\int_\Omega \Big(\partial_x u(x,t)\Big)^2 dx.$$

- **Nonlinear Diffusion Equation 1** As explained also, by Kamalinejad in [30, Thm. 3.9, 3.10], a PDE of the form for some $\alpha \in \mathbb{R}$:

$$\partial_t u(x,t) = -2\alpha\partial_x(u(x,t)\partial_x((u(x,t))^{\alpha-1}\partial_x^2(u(x,t))^\alpha)), \tag{35}$$

which is the Wasserstein gradient flow of

$$\mathcal{E}_v(u) := \int_\Omega (\partial_x(u(x,t))^a))^2 dx.$$

- **Nonlinear Diffusion Equation 2**: As given in [30, Thm. 3.11, p. 561],

$$\partial_t u(x,t) = -\partial_x\Big(u(x,t)\partial_x^2\Big(\frac{\partial_x u(x,t)}{(u(x,t))^2}\Big)\Big), \tag{36}$$

which is the Wasserstein gradient flow of

$$\mathcal{E}_f(u) := \frac{1}{2}\int_\Omega (\partial_x \log(u(x,t)))^2 dx.$$

Remark 3.14. The Thin Film equation (34) is a special case of (35) when $\alpha = 1$ and the DLSS equation (31) when $\alpha = \frac{1}{2}$.

Lemma 3.15. *For any $\alpha \in \mathbb{R}$, we have, for general cases, that the evolution equation:*

$$\partial_t u(x,t) = -2\alpha\partial_x(u(x,t)\partial_x((u(x,t))^{\alpha-1}\partial_x^2(u(x,t))^\alpha)), \tag{37}$$

is a Wasserstein gradient flow of a discrete energy functional:

$$\mathcal{E}_g(u) := \int_\Omega (\partial_x(u(x,t))^\alpha)^2 dx.$$

Proof. We have that the functional derivative is

$$\frac{\delta\mathcal{E}_g}{\delta u} = 2\frac{\partial u^\alpha}{\partial x}\frac{\partial}{\partial u}\frac{\partial u^\alpha}{\partial x} = 2\frac{\partial u^\alpha}{\partial x}\frac{\partial^2 u^\alpha}{\partial x^2}\frac{\partial x}{\partial u} = 2\frac{\partial u^\alpha}{\partial u}\frac{\partial^2 u^\alpha}{\partial x^2} = 2\alpha u^{\alpha-1}\partial_x^2 u^\alpha,$$

then substituting into (33) gives us the result (37). ∎

4 Minimising Movement Schemes of Higher Order of Accuracy

The aim is to extend the minimising movement schemes (3) from the implicit Euler case to higher order/stage cases, e.g. backward difference formula, diagonally implicit Runge-Kutta 2-5 stage (DIRK 2-5) schemes. Simply speaking, we would generate a sequence of discrete solutions from an adapted form of above. However, the more complex structure of these schemes is going to create some difficulties along the way which we have to unlock, like guaranteeing that the gradient flow features are preserved.

Before we progress to our main problem, from Section 5, we recall the minimising movement schemes introduced in Section 3, a semi-discrete (time discrete) form of the gradient flow problem (17), equivalent to the implicit Euler case in the last section. We now adapt these schemes which generate high orders of accuracy. We shall derive our schemes in our thesis from this section, using a Taylor approximation approach, however this only assumes smoothness of an arbirtary solution in question and hence our numerical convergence results shown later in Section 9 are expected to deteriorate in relation to here, but this is to be discussed further later.

Furthermore, in this thesis, we define our numerical solution at time point t^n as $u_\tau(t^n)$, where τ is the time step size and its discrete solution as u_τ^n i.e. $t^n = n\tau$ and $u_\tau^n \approx u_\tau(t^n)$.

4.1 Backward Difference Formula (BDF) Schemes

The BDF1-6 schemes are proven, showing why they are of order of accuracy one to six, in time, respectively.

- **BDF1 Scheme**: This scheme gives a first order approximation to (17). Taylor expanding $u(t_\tau^{n-1})$ about $t = t_\tau^n$ gives

$$u_\tau(t^{n-1}) = u_\tau(t^n) - \tau\partial_t u_\tau(t^n) + O(\tau^2) \Rightarrow \partial_t u_\tau(t^n) = \frac{u_\tau(t^n) - u_\tau(t^{n-1})}{\tau} + O(\tau).$$

  Thus, replacing $u(t^n)$ by its approximate u_τ^n and similarly for other time points, gives us the BDF1 scheme:

$$u_\tau^n - u_\tau^{n-1} = -\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n). \tag{38}$$

- **BDF2 Scheme**: This scheme gives a second order approximation to (17). Taylor expanding $u_\tau(t^{n-1})$ and $u_\tau(t^{n-2})$ about $t = t^n$ gives

$$u_\tau(t^{n-1}) = u_\tau(t^n) - \tau\partial_t u_\tau(t^n) + \frac{\tau^2}{2}\partial_t^2 u_\tau(t^n) + O(\tau^3), \tag{39a}$$

$$u_\tau(t^{n-2}) = u_\tau(t^n) - 2\tau\partial_t u_\tau(t^n) + 2\tau^2\partial_t^2 u_\tau(t^n) + O(\tau^3). \tag{39b}$$

For this to be second order, we wish to eliminate the τ^2 terms, which is possible by calculating $4(39a) - (39b)$, giving us

$$3u_\tau(t^n) - 4u_\tau(t^{n-1}) + u_\tau(t^{n-2}) = -2\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau(t^n)) + O(\tau^3).$$

Thus, replacing $u(t^n)$ by its approximate u_τ^n and similarly for the other time point, gives us the BDF2 scheme:

$$3u_\tau^n - 4u_\tau^{n-1} + u_\tau^{n-2} = -2\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n). \tag{40}$$

- **BDF3 Scheme**: This scheme gives a third order approximation to (17), see Appendix A for derivation:

$$11u_\tau^n - 18u_\tau^{n-1} + 9u_\tau^{n-2} - 2u_\tau^{n-3} = -6\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n). \tag{41}$$

- **BDF4 Scheme**: This scheme gives a fourth order approximation to (17), see Appendix A for the derivation:

$$25u_\tau^n - 48u_\tau^{n-1} + 36u_\tau^{n-2} - 16u_\tau^{n-3} + 3u_\tau^{n-4} = -12\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n). \tag{42}$$

- **BDF5 Scheme**: This scheme gives a fifth order approximation to (17), see Appendix A for the derivation:

$$137u_\tau^n - 300u_\tau^{n-1} + 300u_\tau^{n-2} - 200u_\tau^{n-3} + 75u_\tau^{n-4} - 12u_\tau^{n-5} = -60\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n). \tag{43}$$

- **BDF6 Scheme**: This scheme gives a sixth order approximation to (17), see Appendix A for the derivation:

$$147u_\tau^n - 360u_\tau^{n-1} + 450u_\tau^{n-2} - 400u_\tau^{n-3} + 225u_\tau^{n-4} - 72u_\tau^{n-5} + 10u_\tau^{n-6} = -60\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n). \tag{44}$$

4.2 Construction of our Higher Order BDF Minimising Movement Schemes

Detailed construction of the minimising movement schemes for BDF1-6 (38)-(44) are given from Section 4.1.

As illustrated in [17] for the first order scheme (BDF1), we shall consider the Wasserstein gradient flow (17) for $u(x,t) \in \mathcal{P}_M(\Omega)$ with a smooth potential $\mathcal{E} : \mathcal{P}_M(\Omega) \to \mathbb{R} \cup \{\infty\}$.

- **BDF1 Scheme**: This is fully shown from article [17], but here is the implementation before we do the same for higher order schemes:

Firstly we consider

$$u_\tau^n := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}}\, \Phi_1^\tau(u_\tau^{n-1};u),$$

$$\Phi_1^\tau(u_\tau^{n-1};u) := \frac{\rho}{\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2 + \mathcal{E}(u),$$

and the minimiser u_τ^n gives us

$$\frac{2\rho}{\tau}\left(u_\tau^n - u_\tau^{n-1}\right) = -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n),$$

which satisfies the BDF1 formula (38) if $\rho = \frac{1}{2}$. Thus our scheme is

$$\begin{cases} u_\tau^n := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}}\, \Phi_1^\tau(u_\tau^{n-1};u), \\ \Phi_1^\tau(u_\tau^{n-1};u) := \frac{1}{2\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2 + \mathcal{E}(u). \end{cases}$$

- **BDF2 Scheme**: For the second order inductive scheme, we have

$$u_\tau^n := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}}\, \Phi_2^\tau(u_\tau^{n-1},u_\tau^{n-2};u),$$

$$\Phi_2^\tau(u_\tau^{n-1},u_\tau^{n-2};u) := \frac{a}{\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2 + \frac{b}{\tau}\mathcal{W}_2[u_\tau^{n-2},u]^2 + \mathcal{E}(u).$$

The minimiser u_τ^n satisfies the condition of a critical point if

$$\frac{2}{\tau}\Big(a(u_\tau^n - u_\tau^{n-1}) + b(u_\tau^n - u_\tau^{n-2})\Big) = -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n),$$

which satisfies the BDF2 formula (40) if $a = 1$ and $b = -\frac{1}{4}$. Thus our scheme is

$$\begin{cases} u_\tau^n := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}}\, \Phi_2^\tau(u_\tau^{n-1},u_\tau^{n-2};u), \\ \Phi^\tau(u_\tau^{n-1},u_\tau^{n-2};u) := \frac{1}{\tau}\mathcal{W}_2[u^{n-1},u]^2 - \frac{1}{4\tau}\mathcal{W}_2[u^{n-2},u]^2 + \mathcal{E}(u). \end{cases}$$

- **BDF3 Scheme**: We implement the same idea from BDF2 for the BDF3 scheme, see Appendix A for the derivation:

$$\begin{cases} u_\tau^n := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}}\, \Phi_3^\tau(u_\tau^{n-1},u_\tau^{n-2},u_\tau^{n-3};u), \\ \Phi_3^\tau(u_\tau^{n-1},u_\tau^{n-2},u_\tau^{n-3};u) := \frac{3}{2\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2 - \frac{3}{4\tau}\mathcal{W}_2[u_\tau^{n-2},u]^2 + \frac{1}{6\tau}\mathcal{W}_2[u_\tau^{n-3},u]^2 + \mathcal{E}(u). \end{cases} \tag{45}$$

- **BDF4 Scheme**: We implement the same idea from BDF2 to 3 for the BDF4 scheme, see Appendix A for the derivation:

$$\begin{cases} u_\tau^n := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}}\, \Phi_4^\tau(u_\tau^{n-1},u_\tau^{n-2},u_\tau^{n-3},u_\tau^{n-4};u), \\ \Phi_4^\tau(u_\tau^{n-1},u_\tau^{n-2},u_\tau^{n-3},u_\tau^{n-4};u) := \frac{2}{\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2 - \frac{3}{2\tau}\mathcal{W}_2[u_\tau^{n-2},u]^2 \\ \qquad\qquad + \frac{2}{3\tau}\mathcal{W}_2[u_\tau^{n-3},u]^2 - \frac{1}{8\tau}\mathcal{W}_2[u_\tau^{n-4},u]^2 + \mathcal{E}(u). \end{cases} \tag{46}$$

- **BDF5 Scheme**: We implement the same idea from BDF2 to 4 for the BDF5 scheme, see Appendix A for the derivation:

$$\begin{cases} u_\tau^n := \underset{u \in \mathcal{P}_M(\Omega)}{\operatorname{argmin}} \Phi_5^\tau(u_\tau^{n-1}, u_\tau^{n-2}, u_\tau^{n-3}, u_\tau^{n-4}, u_\tau^{n-5}; u), \\ \\ \quad \Phi_5^\tau(u_\tau^{n-1}, u_\tau^{n-2}, u_\tau^{n-3}, u_\tau^{n-4}, u_\tau^{n-5}; u) \\ \quad := \frac{5}{2\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 - \frac{5}{2\tau}\mathcal{W}_2[u_\tau^{n-2}, u]^2 + \frac{5}{3\tau}\mathcal{W}_2[u_\tau^{n-3}, u]^2 - \frac{5}{8\tau}\mathcal{W}_2[u_\tau^{n-4}, u]^2 \\ \qquad + \frac{1}{10\tau}\mathcal{W}_2[u_\tau^{n-5}, u]^2 + \mathcal{E}(u). \end{cases} \tag{47}$$

- **BDF6 Scheme**: We implement the same idea from BDF2 to 5 for the BDF3 scheme, see Appendix A for the derivation:

$$\begin{cases} u_\tau^n := \underset{u \in \mathcal{P}_M(\Omega)}{\operatorname{argmin}} \Phi_6^\tau(u_\tau^{n-1}, u_\tau^{n-2}, u_\tau^{n-3}, u_\tau^{n-4}, u_\tau^{n-5}, u_\tau^{n-6}; u), \\ \\ \quad \Phi_6^\tau(u_\tau^{n-1}, u_\tau^{n-2}, u_\tau^{n-3}, u_\tau^{n-4}, u_\tau^{n-5}, u_\tau^{n-6}; u) \\ \quad := \frac{3}{\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 - \frac{15}{4\tau}\mathcal{W}_2[u_\tau^{n-2}, u]^2 + \frac{10}{3\tau}\mathcal{W}_2[u_\tau^{n-3}, u]^2 - \frac{15}{8\tau}\mathcal{W}_2[u_\tau^{n-4}, u]^2 \\ \qquad + \frac{3}{5\tau}\mathcal{W}_2[u_\tau^{n-5}, u]^2 - \frac{1}{12\tau}\mathcal{W}_2[u_\tau^{n-6}, u]^2 + \mathcal{E}(u). \end{cases} \tag{48}$$

With the BDF schemes sorted for now, we go further by defining multistep schemes e.g. Runge-Kutta schemes.

4.3 Runge-Kutta Stage Two Scheme

The multistage schemes commence here with the introduction to the Runge-Kutta stage two scheme, and wish to construct a general Butcher array giving us an overall order of accuracy of two.

Before we commence our discussion about multistep (Runge-Kutta) methods, the Butcher array is an array illustrating the parameters for the equations of the discrete solution. The vertical array $c_i; \quad i = 1, \dots, s$ is the node vector, where we take c_i to lie between $[0, 1]$, and the horizontal array $b_i; \quad i = 1, \dots, s$ is the weight vector for the slopes at each time points:

$$\begin{array}{c|cccc} c_1 & a_{11} & a_{12} & \dots & a_{1s} \\ c_2 & a_{21} & a_{22} & \dots & a_{2s} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ c_s & a_{s1} & a_{s2} & \dots & a_{ss} \\ \hline & b_1 & b_2 & \dots & b_s. \end{array} \tag{49}$$

Proposition 4.1. *If we let $c_1 = a$, $c_2 = 1$ and $a_{11} + a_{12} = a$, a Butcher array for a second order two stage Runge-Kutta scheme can be given as:*

$$\begin{array}{c|cc} a & a_{11} & a_{12} \\ 1 & \frac{1}{2(1-a)} & \frac{1-2a}{2(1-a)} \\ \hline & \frac{1}{2(1-a)} & \frac{1-2a}{2(1-a)}. \end{array}$$

Proof. The Butcher array for the two stage Runge-Kutta scheme is as below:

$$\begin{array}{c|cc} a & a_{11} & a_{12} \\ 1 & a_{21} & a_{22} \\ \hline & b_1 & b_2, \end{array}$$

where we let $c_1 = a$ and $c_2 = 1$, which considers the generation of a sequence of discrete solutions of a time step of τ, with an intermediate time step of $a\tau$. This gives the following schemes for each stage:

$$u_\tau^{n+a-1} = u_\tau^{n-1} - a_{11}\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+a-1}) - a_{12}\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n}), \tag{50a}$$

$$u_\tau^{n} = u_\tau^{n-1} - a_{21}\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+a-1}) - a_{22}\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n}). \tag{50b}$$

Expanding the actual solution $u(t_\tau^n)$ at time $t = t_\tau^n$ about $t = t_\tau^{n-1}$ gives

$$u(t_\tau^n) = u_\tau(t^{n-1}) + \tau\partial_t u_\tau(t^{n-1}) + \frac{\tau^2}{2}\partial_t^2 u_\tau(t^{n-1}) + O(\tau^3). \tag{51}$$

Taking the initial value problem (17):

$$\partial_t u_\tau(t^{n-1}) = -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau(t))|_{t=t^{n-1}} = -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau(t^{n-1})), \tag{52a}$$

$$\begin{aligned} \partial_t^2 u_\tau(t^{n-1}) &= -\partial_t(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau(t)))|_{t=t_\tau^{n-1}} \\ &= \partial_u(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau(t^{n-1})))\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau(t^{n-1})), \end{aligned} \tag{52b}$$

and similarly for $t = t^{n+a-1}$ and $t = t^n$, substituting (52a) and (52b) into (51) gives

$$\begin{aligned} u_\tau(t^n) = \quad & u_\tau(t^{n-1}) - \tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau(t^{n-1})) \\ & + \frac{\tau^2}{2}\partial_u(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau(t^{n-1})))\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau(t^{n-1})) + O(\tau^3). \end{aligned} \tag{53}$$

Expanding $\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+a-1})$ about u_τ^{n+a-1} gives us, with assistance from (50a):

$$\begin{aligned} -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+a-1}) &= -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}) - (u_\tau^{n+a-1} - u_\tau^{n-1})\partial_u(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1})) \\ &= -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}) + \tau\big(a_{11}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}) \\ &\quad + a_{12}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1})\big)\partial_u(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1})) \\ &= -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}) + (a_{11} + a_{12})\tau\partial_u(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}))\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}). \end{aligned} \tag{54}$$

Similarly, expanding $\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n)$ about u_τ^{n-1} gives us, with assistance from (50a):

$$\begin{aligned} -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n) &= -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}) - (u_\tau^n - u_\tau^{n-1})\partial_u(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1})) \\ &= -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}) + \tau(a_{21}+a_{22})\partial_u(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}))\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}). \end{aligned} \tag{55}$$

Substituting these into the second equation from (50a) gives

$$\begin{aligned} u_\tau^n = \;& u_\tau^{n-1} - a_{21}\tau\big[\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}) - (a_{11}+a_{12})\tau\partial_u(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}))\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1})\big] \\ & - a_{22}\tau\big[\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}) - (a_{21}+a_{22})\tau\partial_u(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}))\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1})\big] \\ = \;& u_\tau^{n-1} - (a_{21}+a_{22})\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}) \\ & + \tau^2[a_{21}(a_{11}+a_{12}) + a_{22}(a_{21}+a_{22})]\partial_u(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}))\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}). \end{aligned} \tag{56}$$

We wish for $u(t_\tau^n) - u_\tau^n = O(\tau^3)$ therefore comparing (53) and (56) gives as follows:

- From the coefficients of $-\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1})$, we set

$$a_{21} + a_{22} = 1. \tag{57}$$

- From the coefficients of $\partial_u(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}))\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1})$ and part (i), we set

$$a_{21}(a_{11}+a_{12}) + a_{22}(a_{21}+a_{22}) = (a_{11}+a_{12})a_{21} + a_{22} = \frac{1}{2}. \tag{58}$$

Solving simultaneously and subtracting (58) from (57) gives us the Butcher array entries

$$(1 - a_{11} - a_{12})a_{21} = \frac{1}{2},$$

and we let $a_{11} + a_{12} = a$, the first row of the Butcher array.

Hence we have $a_{21} = \frac{1}{2(1-a)}$ and thus $a_{22} = 1 - \frac{1}{2(1-a)} = \frac{1-2a}{2(1-a)}$ in order for this scheme to be of second order. We also have that $a_{21} + a_{22} = 1$. □

There are many examples of this scheme which have theoretical order of accuracy of two, the intermediate time point could be altered flexibly. We mentioned an example of a second order, in time, Runge-Kutta scheme, but we wish to investigate these schemes of a diagonal structure:

Example 4.2. *Taking the gradient flow problem (17) and $c_1 = \frac{1}{4}$, we have that*

$$a_{21} = \frac{2}{3}; \qquad a_{22} = \frac{1}{3},$$

and that since $a_{11} + a_{12} = \frac{1}{4}$, and hence can take $a_{11} = \frac{1}{8}$ and $a_{12} = \frac{1}{8}$, giving the final scheme:

$$\begin{cases} u_\tau^{n-3/4} = u_\tau^{n-1} - \dfrac{\tau}{8}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-3/4}) - \dfrac{\tau}{8}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n), \\ \quad u_\tau^n = u_\tau^{n-1} - \dfrac{2\tau}{3}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-3/4}) - \dfrac{\tau}{3}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n). \end{cases}$$

We mentioned an example of a second order, in time, Runge-Kutta scheme, but we wish to investigate these scheme of a diagonal structure, also considered by Westdickenberg, Wilkening [48].

4.4 Minimising Movement Scheme: Diagonally Implicit Runge-Kutta 2 (DIRK2) Scheme

The scheme from the previous section is modified slightly, such that the matrix of the Butcher array is of diagonal form. We highlight and prove at the end of this subsection, an important result, such that we wish for scheme to be L-stable and hence also A-stable. In other words the last row of the matrix and the row vector should match in order to guarantee L-stability.

We have introduced multistage schemes i.e. Runge-Kutta schemes. However we will be discussing this for the diagonal case, the Runge-Kutta two stage scheme, but $a_{12} = 0$. A DIRK2 scheme can also be easily understood within a Butcher array, but the consideration of a DIRK2 scheme is that a general Runge-Kutta scheme can include discrete solutions at earlier time steps, dependent on terms with later time points (refer back to (50a)) but the DIRK scheme does not, which simplifies the layout but maintains the second order of accuracy in time.

Definition 4.3. We have the general form of the diagonally implicit Runge-Kutta stage q methods which have the form

$$\begin{cases} u_\tau^{n,i} := u_\tau^{n-1} - \tau \sum_{j=1}^{i} a_{ij} \nabla_{\mathcal{W}_2} \mathcal{E}(u_\tau^{n,j}), \\ u_\tau^{n} := u_\tau^{n-1} - \tau \sum_{i=1}^{q} b_i \nabla_{\mathcal{W}_2} \mathcal{E}(u_\tau^{n,i}). \end{cases} \tag{59}$$

Hence, the Butcher array, for the DIRK2 schemes ($q = 2$) is

$$\begin{array}{c|cc} c_1 & a_{11} & \\ c_2 & a_{21} & a_{22} \\ \hline & b_1 & b_2. \end{array}$$

Now lets convert this form to second order of accuracy in time:

Proposition 4.4. *The Butcher array, from [48] in this case will be*

$$\begin{array}{c|cc} a & a & \\ 1 & \frac{1}{2(1-a)} & \frac{1-2a}{2(1-a)} \\ \hline & \frac{1}{2(1-a)} & \frac{1-2a}{2(1-a)}, \end{array} \tag{60}$$

giving us the following system:

$$u_\tau^{n+a-1} := u_\tau^{n-1} - a\tau \nabla_{\mathcal{W}_2} \mathcal{E}(u_\tau^{n+a-1}), \tag{61a}$$

$$u_\tau^{n} := u_\tau^{n-1} - \frac{\tau}{2(1-a)} \nabla_{\mathcal{W}_2} \mathcal{E}(u_\tau^{n+a-1}) - \frac{(1-2a)\tau}{2(1-a)} \nabla_{\mathcal{W}_2} \mathcal{E}(u_\tau^{n}). \tag{61b}$$

Proof. By the simple alteration of the Runge-Kutta two stage scheme, from Section 4.3, with $a_{12} = 0$ and $a_{11} + a_{12} = a$, the DIRK2 scheme is of maximum order of two if

$$a_{11} = a, \quad a_{21} = \frac{1}{2(1-a)}, \quad a_{22} = \frac{1-2a}{2(1-a)}. \tag{62}$$

Indeed, for simplicity, we wish for the scheme to generate a sequence of solutions in time steps of τ i.e. $u_\tau^{n-1} \to u_\tau^n$, via the support of intermediate solution u_τ^{n+a-1}. Hence, we take $c_1 := a$, $c_2 := 1$, $a_{11} := a$, $a_{12} := 0$, $a_{21} := \frac{1}{2(1-a)}$, $a_{22} := \frac{1-2a}{2(1-a)}$, $b_1 := \frac{1}{2(1-a)}$, $b_2 := \frac{1-2a}{2(1-a)}$ and we obtain the system from (59). $\square$

Now we immediately construct the minimising movement scheme with respect to DIRK2 of order two:

Corollary 4.5. *From the system (61a)-(61b), the minimising movement scheme for the DIRK2 scheme is as follows for each stage:*

Stage One

$$\begin{cases} u_\tau^{n+a-1} := \underset{u \in \mathcal{P}_M(\Omega)}{\operatorname{argmin}} \Phi_{2,1}^\tau(u_\tau^{n-1}; u), \\ \Phi_{2,1}^\tau(u_\tau^{n-1}; u) := \dfrac{1}{2a\tau} \mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u), \end{cases} \tag{63}$$

Stage Two

$$\begin{cases} u_\tau^n := \underset{u \in \mathcal{P}_M(\Omega)}{\operatorname{argmin}} \Phi_{2,2}^\tau(u_\tau^{n+a-1}, u_\tau^{n-1}; u), \\ \Phi_{2,2}^\tau(u_\tau^{n+a-1}, u_\tau^{n-1}; u) := \dfrac{1}{2a(1-2a)\tau} \mathcal{W}_2[u_\tau^{n+a-1}, u]^2 \\ \qquad - \dfrac{1-2a(1-a)}{2a(1-2a)\tau} \mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u). \end{cases} \tag{64}$$

Proof. In order to transform the system (61a)-(61b) into one equation, from which we can go forward to constructing a minimising movement scheme for the latter, we shall follow a procedure by removing the $\nabla_{\mathcal{W}_2} \mathcal{E}(u_\tau^{n+a-1})$ term by calculating $2a(1-a)(61b) - (61a)$. This gives us as a result:

$$\frac{1}{\tau}\Big(\frac{2(1-a)}{1-2a} u_\tau^n - \frac{1}{a(1-2a)} u_\tau^{n+a-1} + \frac{1-2a(1-a)}{a(1-2a)} u_\tau^{n-1}\Big) = -\nabla_{\mathcal{W}_2} \mathcal{E}(u_\tau^n), \tag{65}$$

with $-\nabla_{\mathcal{W}_2} \mathcal{E}(u_\tau^n)$ being our Wasserstein gradient flow for $u(x,t)$ (17), from [17, Sect. 2.2] i.e. $\partial_t u = -\nabla_{\mathcal{W}_2} \mathcal{E}(u)$.

Since we are working with discrete solutions at two time steps u_τ^n, u_τ^{n+a-1} we shall need to construct a second order minimising movement scheme, similar for BDF-type cases, but for the first intermediate time step u_τ^{n+a-1}. Then we shall go on to create another scheme for the second intermediate time step u_τ^n, but dependent on the recently calculated intermediate time step solutions u_τ^{n+a-1} and u_τ^{n-1}.

Now we are in a position to construct minimising movement schemes for each stage, (61a) and (61b). For simplicity, we wish to remove the $\nabla_{\mathcal{W}_2} \mathcal{E}(u_\tau^{n+a-1})$ term, achieved by equation (65).

Remark 4.6. A SDIRK2 case could be considered. In this case, the leading diagonal elements are equal i.e. $a_{11} = a_{22}$. Therefore from the above, a must be such that

$$a = \frac{1-2a}{2(1-a)} \Rightarrow a = 1 \pm \frac{1}{2}\sqrt{2}.$$

We now commence the constructions of our numerical schemes. By setting up our diagonally implicit Runge-Kutta (DIRK2) scheme for general intermediate time steps, with an underlying gradient flow problem, we shall define our minimising movement schemes for each stage, starting with stage one: For the next two subsections, we hire three constants $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$, for this part *only*, which we solve in comparison to (61a) for stage one, then (65) for stage two. We work with stage one first, then stage two afterwards:

- **Stage One**: From $u_\tau^{n-1} \to u_\tau^{n+a-1}$: Here we aim that u_τ^{n+a-1} minimises the Yosida-regularised functional $\Phi_{2,1}^\tau(u_\tau^{n-1}; u)$ i.e. the minimising movement scheme is

$$u_\tau^{n+a-1} := \underset{u \in \mathcal{P}_M(\Omega)}{\operatorname{argmin}}\, \Phi_{2,1}^\tau(u_\tau^{n-1}; u), \qquad \Phi_{2,1}^\tau(u_\tau^{n-1}; u) := \frac{\lambda_1}{\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u).$$

Hence the minimiser u_τ^{n+a-1} (the critical point) satisfies

$$\frac{2\lambda_1}{\tau}\left(u_\tau^{n+a-1} - u_\tau^{n-1}\right) = -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+a-1}),$$

satisfying (61a) if $2\lambda_1 = \dfrac{1}{a} \Rightarrow \lambda_1 = \dfrac{1}{2a}$.

With the minimising movement scheme defined for stage one, we carry out the similar implementation for stage two:

- **Stage Two**: Now we construct the scheme for u_τ^n, dependent on u_τ^{n-1} and intermediate time step u_τ^{n+a-1}. We aim to minimise the Yosida-regularised functional $\Phi_{2,2}^\tau(u_\tau^{n-1}, u_\tau^{n+a-1}; u)$ i.e. the minimising movement scheme is

$$u_\tau^n := \underset{u \in \mathcal{P}_M(\Omega)}{\operatorname{argmin}}\, \Phi_{2,2}^\tau(u_\tau^{n+a-1}, u_\tau^{n-1}; u),$$
$$\Phi_{2,2}^\tau(u_\tau^{n+a-1}, u_\tau^{n-1}; u) := \frac{\lambda_2}{\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \frac{\lambda_3}{\tau}\mathcal{W}_2[u_\tau^{n+a-1}, u]^2 + \mathcal{E}(u).$$

By taking u_τ^n as the minimiser of $\Phi_{2,2}$, we have that

$$\frac{2(\lambda_2+\lambda_3)}{\tau}u_\tau^n - \frac{2\lambda_2}{\tau}u_\tau^{n-1} - \frac{2\lambda_3}{\tau}u_\tau^{n+a-1} = -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n), \tag{66}$$

satisfying (65) if

$$\lambda_2 = -\frac{1-2a(1-a)}{2a(1-2a)}, \qquad \lambda_3 = \frac{1}{2a(1-2a)}.$$

Alternatively, substituting (61a) into (66) gives

$$2(\lambda_2+\lambda_3)u_\tau^n - 2(\lambda_2+\lambda_3)u_\tau^{n-1} = -\tau(2a\lambda_3\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+a-1}) + \nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n)),$$

equivalent to

$$u_\tau^n - u_\tau^{n-1} = -\tau\Big(\frac{a\lambda_3}{\lambda_2+\lambda_3}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+a-1}) + \frac{1}{2(\lambda_2+\lambda_3)}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n)\Big),$$

satisfying equation (61b) instead if $\lambda_2 = -\dfrac{1-2a(1-a)}{2a(1-2a)}$ and $\lambda_3 = \dfrac{1}{2a(1-2a)}$ again.

Thus this gives us the schemes per stage (63) and (64) and the proof is complete. □

Remark 4.7. Clearly, you can observe that (64) is undefined (no minimising movement scheme) when considering a half time step $t^{n-1/2}$ i.e. $a = \dfrac{1}{2}$.

Example 4.8. *Whilst we aim to work on the comparison principle for any DIRK2 scheme with a reasonably high order of accuracy, we will consider applying this for the quarter time step i.e.* $a = \dfrac{1}{4}$, *from Westdickenburg and Wilkening [48]. We will apply this example for our investigation as we progress:*

$$\begin{array}{c|cc} 1/4 & 1/4 & \\ 1 & 2/3 & 1/3 \\ \hline & 2/3 & 1/3. \end{array}$$

This gives us the following minimising movement scheme for

- ***Stage One:***

$$\begin{cases} u^{n-3/4} := \underset{x\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}}\, \Phi_{2,1}^\tau(u_\tau^{n-1};u), \\ \Phi_{2,1}^\tau(u_\tau^{n-1};u) := \dfrac{2}{\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2 + \mathcal{E}(u). \end{cases}$$

- ***Stage Two:***

$$\begin{cases} u_\tau^{n} := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}}\, \Phi_{2,2}^\tau(u_\tau^{n-1},u_\tau^{n-3/4};u), \\ \Phi_{2,2}^\tau(u_\tau^{n-1},u_\tau^{n-3/4};u) := \dfrac{4}{\tau}\mathcal{W}_2[u_\tau^{n-3/4},u]^2 - \dfrac{5}{2\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2 + \mathcal{E}(u). \end{cases}$$

Before moving onto construction of third order schemes, we highlight how this scheme of second order of accuracy guarantees strong stability properties, as shown by Hairer and Wanner, [25, Prop. 3.1, p.40]):

Lemma 4.9. *All DIRK2 schemes with the last step being equivalent to the last intermediate step i.e.* $b_i = a_{2i}$ *for* $i = 1, 2$, *and second order of accuracy are A-stable and L-stable.*

Proof. We break the proof into three parts: (i) For deriving the stability function for Runge-Kutta (this includes DIRK methods obviously); (ii) For verifying A-stability; (iii) For verifying L-stability.

(i) We have the stability function as

$$R(z) := 1 + zb^T(\mathbb{I} - Az)^{-1}\mathbf{1}, \tag{67}$$

where $u_\tau^n = R(z)u_\tau^{n-1}$.

To verify this, the intermediate stages of the Runge-Kutta method has the following matrix representation for an s-stage method, where the scalar test problem $u'(t) = \lambda u(t)$ is applied and $z = \tau\lambda$:

$$\begin{bmatrix} u_\tau^{n,1} \\ u_\tau^{n,2} \\ \vdots \\ u_\tau^{n,s} \end{bmatrix} = u_\tau^{n-1}\mathbf{1} + z \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1s} \\ a_{21} & a_{22} & \dots & a_{2s} \\ \vdots & \vdots & & \vdots \\ a_{s1} & a_{s2} & \dots & a_{ss} \end{bmatrix} \begin{bmatrix} u_\tau^{n,1} \\ u_\tau^{n,2} \\ \vdots \\ u_\tau^{n,s} \end{bmatrix} \tag{68}$$

$$\Leftrightarrow \quad U_\tau^n(\mathbb{I} - Az) = u_\tau^{n-1}\mathbf{1} \Leftrightarrow U_\tau^n = (\mathbb{I} - Az)^{-1}\mathbf{1}u_\tau^{n-1},$$

where $U_\tau^n = [u_\tau^{n,1}, u_\tau^{n,2}, \dots, u_\tau^{n,s}]^T$ and A is the $s \times s$ matrix.

The final stage is represented as

$$u_\tau^n = u_\tau^{n-1} + zb^T U_\tau^n. \tag{69}$$

As a result, substituting (68) into (69) gives us

$$u_\tau^n = u_\tau^{n-1} + zb^T(\mathbb{I} - Az)^{-1}\mathbf{1}u_\tau^{n-1},$$

and since matrix/vector multiplication gives us a result on the real line $\mathbb{R}$, factorising out u_τ^{n-1} gives us (67).

(ii) Moving onto the A-stable and L-stable proof now: As a result of equation (62), in relation to the corresponding Butcher array (60), we manipulate and simplify the stability function as follows:

$$\begin{aligned} R(z) &= 1 + z \begin{bmatrix} \frac{1}{2(1-a)} & \frac{1-2a}{2(1-a)} \end{bmatrix} \left(\begin{bmatrix} 1 - az & 0 \\ \frac{z}{2(a-1)} & 1 + \frac{(1-2a)z}{2(a-1)} \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= 1 + z \begin{bmatrix} \frac{1}{2(1-a)} & \frac{1-2a}{2(1-a)} \end{bmatrix} \frac{1}{(1-az)\left(\frac{2(a-1)+(1-2a)z}{2(a-1)}\right)} \begin{bmatrix} 1 + \frac{1-2a}{2(a-1)}z & 0 \\ \frac{z}{2(1-a)} & 1 - az \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= 1 - \frac{2(1-a)z}{(1-az)\,(2(a-1) + (1-2a)z)} \begin{bmatrix} \frac{1}{2(1-a)} & \frac{(1-2a)(1-az)}{2(1-a)} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= 1 - \frac{1 + (1-2a)(1-az)}{(1-az)\,(2(a-1) + (1-2a)z)} z = \frac{2(1-az)(a-1) - z}{(1-az)\,(2(a-1) + (1-az)z)}. \end{aligned} \tag{70}$$

We know that if $R(z) \leq 1$, then the discrete solution u_τ^n converges to zero i.e. the stability region is z such that $R(z) = \frac{P(z)}{Q(z)} \leq 1$. Considering the fact that, from [25, Def. 3.3], we have that this method is A-stable if

$$E(y) = Q(iy)Q(-iy) - P(iy)P(-iy) \geq 0, \quad i \in \mathbb{C},$$

for all $y \in \mathbb{R}$.

Therefore from our derived function $R(z)$, we have that

$$\begin{aligned}
E(y) =&(1-iay)\left(2(a-1)+(1-iay)iy\right)(1+iay)\left(2(a-1)-(1+iay)iy\right)\\
&-\left(2(1-iay)(a-1)-iy\right)\left(2(1+iay)(a-1)+iy\right)\\
=&(1+a^2y^2)\left(4(1-a)^2+(a+a^2y^2)y^2\right)-\left(2(1+a^2y^2)(1-a)^2+y^2\right)\\
=&4(1-a)^2(1+a^2y^2)+(a+a^2y^2)y^2-2(1-a)^2(1+a^2y^2)-y^2\\
=&2(1-a)^2(1+a^2y^2)+(2a^2y^2+a^4y^4)y^2,
\end{aligned}$$

which is clearly non-negative for all $y \in \mathbb{R}$ and thus the A-stability hypothesis holds.

(iii) We know that an A-stable scheme is furthermore L-stable if the stability function satisfies

$$R(z) \to 0 \quad \text{as} \quad z \to \infty. \tag{71}$$

From (70), we clearly observe that the degree of the polynomial of numerator $P(z)$ and denominator $Q(z)$ is one and two respectively, i.e. $deg(P(z)) = 1$ and $deg(Q(z)) = 3$ and thus, via the L'Hôpital's rule approach, the result (71) holds.

□

4.5 Diagonally Implicit Runge-Kutta Three Stage (DIRK3) Scheme

We move on from two stage and construct a scheme of three stages and gives a system of properties which gurantee a maximum order of accuracy of three.

4.5.1 Scheme One

There are two schemes of third order which we show, with the latter for our numerical simulation. But here is the first one:

Definition 4.10. The Butcher array for the scheme is as below (Note: We consider that the last two rows are equal to guarantee L-stability, as for our DIRK2 scheme i.e. $a_{3i} = b_i$; $(i = 1, 2, 3)$):

$$\begin{array}{c|ccc}
c_1 & a_{11} & & \\
c_2 & a_{21} & a_{22} & \\
c_3 & a_{31} & a_{32} & a_{33} \\
\hline
 & b_1 & b_2 & b_3,
\end{array} \tag{72}$$

which gives the following schemes for each stage (NOTE: $c_1 < c_2 < 1$):

$$u_\tau^{n+c_1-1} := u_\tau^{n-1} - a_{11}\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+c_1-1}), \tag{73a}$$

$$u_\tau^{n+c_2-1} := u_\tau^{n-1} - a_{21}\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+c_1-1}) - a_{22}\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+c_2-1}), \tag{73b}$$

$$u_\tau^{n} := u_\tau^{n-1} - a_{31}\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+c_1-1}) - a_{32}\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+c_2-1}) - a_{33}\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n}). \tag{73c}$$

Proposition 4.11. *If we let $c_3 = 1$, $a_{11} = c_1$ and $a_{21} + a_{22} = c_2$, then the Butcher array for a third order DIRK3 scheme can be given as:*

$$\begin{array}{c|ccc}
c_1 & c_1 & & \\
c_2 & \dfrac{c_1 + c_2 - 4c_1c_2}{2(1-3c_1)(1-c_1)} & \dfrac{6c_1^2c_2 - 4c_1c_2 - c_1 + c_2}{2(1-3c_1)(1-c_1)} & \\
1 & \dfrac{1-3c_2}{6(1-c_1)(c_1-c_2)} & \dfrac{1-3c_1}{6(1-c_2)(c_2-c_1)} & \dfrac{2-3(c_1-2c_1c_2+c_2)}{6(1-c_1)(1-c_2)} \\
\hline
 & \dfrac{1-3c_2}{6(1-c_1)(c_1-c_2)} & \dfrac{1-3c_1}{6(1-c_2)(c_2-c_1)} & \dfrac{2-3(c_1-2c_1c_2+c_2)}{6(1-c_1)(1-c_2)}.
\end{array}$$

By re-using the the system of equations (73a) for (72), we obtain the system:

$$u_\tau^{n+c_1-1} := \quad u_\tau^{n-1} - c_1\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+c_1-1}), \tag{74a}$$

$$\begin{aligned} u_\tau^{n+c_2-1} := \quad & u_\tau^{n-1} - \frac{(c_1+c_2-4c_1c_2)\tau}{2(1-3c_1)(1-c_1)}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+c_1-1}) \\ & - \frac{(6c_1^2c_2 - 4c_1c_2 - c_1 + c_2)\tau}{2(1-3c_1)(1-c_1)}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+c_2-1}), \end{aligned} \tag{74b}$$

$$\begin{aligned} u_\tau^{n} := \quad & u_\tau^{n-1} - \frac{(1-3c_2)\tau}{6(1-c_1)(c_1-c_2)}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+c_1-1}) \\ & - \frac{(1-3c_1)\tau}{6(1-c_2)(c_2-c_1)}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+c_2-1}) - \frac{[2-3(c_1-2c_1c_2+c_2)]\tau}{6(1-c_1)(1-c_2)}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n}). \end{aligned} \tag{74c}$$

Proof. Expanding the actual solution $u_\tau(t^n)$ at time $t = t^n$ about $t = t^{n-1}$ gives

$$u_\tau(t^n) = u_\tau(t^{n-1}) + \tau\partial_t u_\tau(t^{n-1}) + \frac{\tau^2}{2}\partial_t^2 u_\tau(t^{n-1}) + \frac{\tau^3}{6}\partial_t^3 u_\tau(t^{n-1}) + O(\tau^4).$$

Taking the initial value problem $\partial_t u(x,t) = -\nabla_{\mathcal{W}_2}\mathcal{E}(u(x,t))$ and referencing the same chain rule approach (52a,52b) from the DIRK2 scheme, we have

$$\partial_t u_\tau(t^{n-1}) = -\nabla_{\mathcal{W}_2}\mathcal{E}(u(t))|_{t=t_\tau^{n-1}} = -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}), \tag{75a}$$

$$\partial_t^2 u_\tau(t^{n-1}) = -\partial_t(\nabla_{\mathcal{W}_2}\mathcal{E}(u(t)))|_{t=t_\tau^{n-1}} = \partial_u(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}))\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}), \tag{75b}$$

$$\begin{aligned} \partial_t^3 u_\tau(t^{n-1}) &= -\partial_t^2(\nabla_{\mathcal{W}_2}\mathcal{E}(u(t)))|_{t=t_\tau^{n-1}} \\ &= -\partial_u^2\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}))[\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1})]^2 - [\partial_u(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}))]^2\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}), \end{aligned} \tag{75c}$$

and similarly for $t = t^{n+a-1}$ and $t = t^n$, we have

$$\begin{aligned} u(t_\tau^n) =& u(t_\tau^{n-1}) - \tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}) + \frac{\tau^2}{2}\partial_u(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}))\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}) \\ & - \frac{\tau^3}{6}\left\{\partial_u^2(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}))[\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1})]^2 + [\partial_u(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}))]^2\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1})\right\} + O(\tau^4). \end{aligned} \tag{76}$$

In line with (54)-(56), expanding $\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+c_1-1})$ about u_τ^{n-1} gives us, with assistance from (73a):

$$\begin{aligned} -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+c_1-1}) =& -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}) + a_{11}\tau\partial_u(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}))\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}) \\ & - a_{11}^2\tau^2[\partial_u(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}))]^2\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}) \\ & - \frac{1}{2}a_{11}^2\tau^2\partial_u^2(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}))[\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1})]^2. \end{aligned} \tag{77}$$

Similarly, expanding $\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+c_2-1})$ about u_τ^{n-1} gives us, with assistance from equation (73b):

$$\begin{aligned} -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+c_2-1}) =& -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}) + (a_{21}+a_{22})\tau\partial_u(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}))\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}) \\ & - \big[a_{21}a_{11} + a_{22}(a_{21}+a_{22})\big]\tau^2[\partial_u(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}))]^2\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}) \\ & - \frac{1}{2}(a_{21}+a_{22})^2\tau^2\partial_u^2(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}))[\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1})]^2. \end{aligned} \tag{78}$$

Similarly, expanding $\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n})$ about u_τ^{n-1} gives us, with assistance from equation (73c):

$$\begin{aligned} -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n}) =& -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}) + (a_{31}+a_{32}+a_{33})\tau\partial_u(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}))\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}) \\ & - \big[a_{11}a_{31} + a_{32}(a_{21}+a_{22}) + a_{33}(a_{31}+a_{32}+a_{33})\big]\tau^2[\partial_u(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}))]^2\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}) \\ & - \frac{1}{2}(a_{31}+a_{32}+a_{33})^2\tau^2\partial_u^2(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}))[\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1})]^2 + O(\tau^3). \end{aligned} \tag{79}$$

By substituting the expansions (77), (78) and (79) into the final equation of (73a), we have

$$\begin{aligned} u_\tau^n =& u_\tau^{n-1} - [a_{31}+a_{32}+a_{33}]\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}) + \big\{a_{31}a_{11} + a_{32}(a_{21}+a_{22}) \\ & + a_{33}(a_{31}+a_{32}+a_{33})\big\}\tau^2\partial_u(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}))\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}) \\ & - \big\{a_{31}a_{11}^2 + a_{32}[a_{11}a_{21} + a_{21}(a_{21}+a_{22})] \\ & + a_{33}[a_{11}a_{31} + a_{32}(a_{21}+a_{22}) + a_{33}(a_{31}+a_{32}+a_{33})]\big\}\tau^3[\partial_u(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}))]^2\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}) \\ & - \frac{1}{2}\big\{a_{31}a_{11}^2 + a_{32}(a_{21}+a_{22})^2 + a_{33}(a_{31}+a_{32}+a_{33})^2\big\}\tau^3\partial_u^2(\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1}))[\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1})]^2 \\ & + O(\tau^4). \end{aligned} \tag{80}$$

We wish for $u_\tau(t^n) - u_\tau^n = O(\tau^4)$. Hence if we let $a_{11} = c_1$ and $a_{21}+a_{22} = c_2$, as well as comparing (76) and (80), we wish for the following system to be satisfied, see [1, Thm. 3, p. 1009]:

$$a_{31}+a_{32}+a_{33} = 1, \tag{81a}$$

$$a_{31}c_1 + a_{32}c_2 + a_{33} = \frac{1}{2}, \tag{81b}$$

$$a_{31}c_1^2 + a_{32}c_2^2 + a_{33} = \frac{1}{3}, \tag{81c}$$

$$a_{11}a_{31}c_1 + (a_{21}c_1 + a_{22}c_2)a_{32} + (a_{31}c_1 + a_{32}c_2 + a_{33})a_{33} = \frac{1}{6}, \tag{81d}$$

thus also giving us for consistency and hence a simplified system of equations.

Solving equations (81a) to (81c) for a_{31}, a_{32}, a_{33} gives us

$$a_{31} = \frac{1-3c_2}{6(1-c_1)(c_1-c_2)}, \quad a_{32} = \frac{3c_1-1}{6(1-c_2)(c_1-c_2)}, \quad a_{33} = \frac{6c_1c_2 - 3(c_1+c_2)+2}{6(1-c_1)(1-c_2)}.$$

Furthermore, by substituting into (81d) and knowing that $a_{21} + a_{22} = c_2$, we can solve simultaneously and obtain unique solutions also for a_{21} and a_{22} in terms of c_1, c_2:

$$a_{21} = \frac{c_1 + c_2 - 4c_1c_2}{2(1-c_1)(1-3c_1)}, \quad a_{22} = \frac{6c_1^2c_2 - 4c_1c_2 - c_1 + c_2}{2(1-c_1)(1-3c_1)}.$$

Thus the Butcher array and its corresponding system (100) is verified and the proof is complete. $\square$

Now we immediately construct the minimising movement scheme with respect to DIRK2 of order two:

Corollary 4.12. *From the system (61a)-(61b), the minimising movement scheme for the DIRK2 scheme is as follows for each stage:*

- ***Stage One***

$$\begin{cases} u^{n+c_1-1} := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}}\, \Phi_1^\tau(u^{n-1};u), \\ \Phi_1^\tau := \dfrac{1}{2c_1\tau}\mathcal{W}_2[u^{n-1},u]^2 + \mathcal{E}(u). \end{cases} \tag{82}$$

- ***Stage Two***

$$\begin{cases} u_\tau^{n+c_2-1} := & \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}}\, \Phi_2^\tau(u_\tau^{n+c_1-1}, u_\tau^{n-1};u), \\ \Phi_2^\tau := & \dfrac{c_1+c_2-4c_1c_2}{2c_1(6c_1^2c_2 - 4c_1c_2 - c_1 + c_2)\tau}\mathcal{W}_2[u_\tau^{n+c_1-1},u]^2 \\ & -\dfrac{c_1+c_2-4c_1c_2-2c_1(1-3c_1)(1-c_1)}{2c_1(6c_1^2c_2 - 4c_1c_2 - c_1 + c_2)\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2 + \mathcal{E}(u). \end{cases} \tag{83}$$

- ***Stage Three***

$$\begin{cases} u_\tau^n := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}}\, \Phi_3^\tau(u_\tau^{n-3/8}, u_\tau^{n-3/4}, u_\tau^{n-1};u), \\ \Phi_3^\tau := \dfrac{y_6}{\tau}\mathcal{W}_2[u_\tau^{n+c_2-1},u]^2 + \dfrac{y_7}{\tau}\mathcal{W}_2[u_\tau^{n+c_1-1},u]^2 + \dfrac{y_8}{\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2 + \mathcal{E}(u), \end{cases} \tag{84}$$

where the prefactors are

$$\begin{aligned} y_6 =& \frac{c_1(1-c_1)^2(1-3c_1)^2}{(6c_1^2c_2 - 4c_1c_2 - c_1 + c_2)\left(3(c_1 - 2c_1c_2 + c_2) - 2\right)(c_1 - c_2)}, \\ y_7 =& \frac{(18c_1^2 - 12c_1 + 3)c_2^3 - (24c_1^2 - 13c_1 + 4)c_2^2 - (12c_1^3 - 25c_1^2 + 8c_1 - 2) + 3c_1^3}{2\left(6c_1^2c_2 - 4c_1c_2 - c_1 + c_2\right)\left(2 - 3(c_1 - 2c_1c_2 + c_2)\right)}, \\ y_8 =& \frac{(c_1 - c_2)[(36c_1^4 - 60c_1^3 + 48c_1^2 - 18c_1 + 3)c_2^2 - (36c_1^4 - 72c_1^3 + 60c_1^2 - 22c_1 + 4)c_2}{2c_1(c_1 - c_2)(6c_1^2 - 4c_1c_2 - c_1 + c_2)} \\ &+ \frac{18c_1^4 - 42c_1^3 + 35c_1^2 - 12c_1 + 2}{2c_1(c_1 - c_2)(6c_1^2 - 4c_1c_2 - c_1 + c_2)}. \end{aligned} \tag{85}$$

Proof. As we did for DIRK2 and to assist us in constructing this system (100) into one equation we apply the following:

- Stage two: Eliminate the $\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+c_1-1})$ term by calculating $2c_1(1-3c_1)(1-c_1)\cdot$(74b) - $(c_1 + c_2 - 4c_1c_2)\cdot$(74a), which gives us

$$\begin{aligned}&\frac{1}{c_1(6c_1^2c_2-4c_1c_2-c_1+c_2)\tau}\Big\{2c_1(1-3c_1)(1-c_1)u_\tau^{n+c_2-1}-(c_1+c_2-4c_1c_2)u_\tau^{n+c_1-1}\\&+(c_1+c_2-4c_1c_2-2c_1(1-3c_1)(1-c_1))\,u_\tau^{n-1}\Big\}=-\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+c_2-1}).\end{aligned}\tag{86}$$

- Stage three: Eliminate the $\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+c_2-1})$ term by substituting (86) into equation (74c), before substituting with equation (74a). This gives us

$$\begin{aligned}&\frac{1}{(2-3(c_1-2c_1c_2+c_2))\,\tau}\Bigg(6c_1(1-c_1)(1-c_2)u_\tau^n+\frac{2c_1(1-c_1)^2(1-3c_1)^2}{(6c_1^2c_2-4c_1c_2-c_1+c_2)(c_1-c_2)}u_\tau^{n+c_2-1}\\&-\frac{(18c_1^2-12c_1+3)c_2^3-(24c_1^2-13c_1+4)c_2^2-(12c_1^3-25c_1^2+8c_1-2)c_2+3c_13}{6c_1^2c_2-4c_1c_2-c_1+c_2}u_\tau^{n+c_1-1}\\&+\Big\{\frac{(c_1-c_2)[(36c_1^4-60c_1^3+48c_1^2-18c_1+3)c_2^2-(36c_1^4-72c_1^3+60c_1^2-22c_1+4)c_2}{c_1(c_1-c_2)(6c_1^2-4c_1c_2-c_1+c_2)}\\&+\frac{18c_1^4-42c_1^3+35c_1^2-12c_1+2}{c_1(c_1-c_2)(6c_1^2-4c_1c_2-c_1+c_2)}\Big\}u_\tau^{n-1}\Big\}\Bigg)=-\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n).\end{aligned}\tag{87}$$

We now start to construct our minimising movement, per stage, schemes as we did for the second order case, but with an additional step to carry out this time:

- **Stage One**: Firstly from $u_\tau^{n-1}\to u_\tau^{n+c_1-1}$ which is immediately defined as (82), similarly as BDF1 and DIRK2 stage one schemes.

- **Stage Two**: So we move straight onto the scheme for minimiser $u_\tau^{n+c_2-1}$ dependent from intermediate time steps $u_\tau^{n+c_1-1}$ and u_τ^{n-1}, which is

$$\begin{aligned}u_\tau^{n+c_2-1}&:=\operatorname*{argmin}_{u\in\mathcal{P}_M(\Omega)}\Phi_2^\tau(u_\tau^{n+c_1-1},u_\tau^{n-1};u),\\ \Phi_2^\tau&:=\frac{y_4}{\tau}\mathcal{W}_2[u_\tau^{n+c_1-1},u]^2+\frac{y_5}{\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2+\mathcal{E}(u).\end{aligned}$$

  From similar calculations from the already implemented schemes, the minimiser $u_\tau^{n+c_2-1}$ satisfies (86) when

$$y_4=\frac{c_1+c_2-4c_1c_2}{2c_1(6c_1^2c_2-4c_1c_2-c_1+c_2)},\quad y_5=-\frac{c_1+c_2-4c_1c_2-2c_1(1-3c_1)(1-c_1)}{2c_1(6c_1^2c_2-4c_1c_2-c_1+c_2)},$$

  hence the final scheme here is as (83).

- **Stage Three**: Finally, moving to the final stage for minimiser u_τ^n dependent from intermediate time steps $u_\tau^{n+c_2-1}$, $u_\tau^{n+c_1-1}$ and u_τ^{n-1}, which is

$$u_\tau^n:=\operatorname*{argmin}_{u\in\mathcal{P}_M(\Omega)}\Phi_3^\tau(u_\tau^{n-3/8},u_\tau^{n-3/4},u_\tau^{n-1};u),$$

$$\Phi_3^\tau(u_\tau^{n-3/8},u_\tau^{n-3/4},u_\tau^{n-1};u):=\frac{y_6}{\tau}\mathcal{W}_2[u_\tau^{n+c_2-1},u]^2+\frac{y_7}{\tau}\mathcal{W}_2[u_\tau^{n+c_1-1},u]^2+\frac{y_8}{\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2+\mathcal{E}(u).$$

  Again, from similar calculations from already implemented schemes, the minimiser u_τ^n satisfies (84) and (85).

Thus the proof for the schemes per stage are complete. $\square$

Example 4.13. *By selecting for our intermediate time steps $c_1 = \frac{1}{4}$ and $c_2 = \frac{3}{4}$, this gives us our final Butchers array for a third order DIRK3 scheme:*

$$\begin{array}{c|ccc} 1/4 & 1/4 & & \\ 3/4 & 2/3 & 1/12 & \\ 1 & 5/9 & 1/3 & 1/9 \\ \hline & 5/9 & 1/3 & 1/9, \end{array}$$

providing us with the system of equations:

$$u_\tau^{n-3/4} := u_\tau^{n-1} - \frac{\tau}{4}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-3/4}), \tag{88a}$$

$$u_\tau^{n-1/4} := u_\tau^{n-1} - \frac{2\tau}{3}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-3/4}) - \frac{\tau}{12}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1/4}), \tag{88b}$$

$$u_\tau^{n} := u_\tau^{n-1} - \frac{5\tau}{9}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-3/4}) - \frac{\tau}{3}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1/4}) - \frac{\tau}{9}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n}). \tag{88c}$$

- *The stage one minimising movement scheme generates $u_\tau^{n-3/4}$, given by*

$$u_\tau^{n-3/4} := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}} \left\{ \frac{2}{\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u) \right\}.$$

- *The stage two minimising movement scheme generates $u_\tau^{n-1/4}$: Substituting (88a) for $\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-3/4})$ into (88b) gives*

$$12u_\tau^{n-1/4} - 32u_\tau^{n-3/4} + 20u_\tau^{n-1} = -\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1/4}). \tag{89}$$

The discrete solution at $t = t^{n-1/4}$, $u_\tau^{n-1/4}$ satisfies

$$u_\tau^{n-1/4} := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}} \left\{ \frac{a}{\tau}\mathcal{W}_2[u_\tau^{n-3/4}, u]^2 + \frac{b}{\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u) \right\},$$

if

$$\frac{2(a+b)}{\tau}u_\tau^{n-1/4} - \frac{2a}{\tau}u_\tau^{n-3/4} - \frac{2b}{\tau}u_\tau^{n-1} = -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1/4}),$$

satisfying (89) if $a = 16$ and $b = -10$. We hence have the stage two minimising movement scheme:

$$u_\tau^{n-1/4} := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}} \left\{ \frac{16}{\tau}\mathcal{W}_2[u_\tau^{n-3/4}, u]^2 - \frac{10}{\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u) \right\}.$$

- *The stage three minimising movement scheme generates u_τ^n: Calculating $9(88c) - 36(88b) + 76(88a)$ gives us*

$$9u_\tau^n - 36u_\tau^{n-1/4} + 76u_\tau^{n-3/4} - 49u_\tau^{n-1} = -\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n).$$

The discrete solution at $t = t^n$, u_τ^n satisfies

$$u_\tau^n := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}} \left\{\frac{\alpha_1}{\tau}\mathcal{W}_2[u_\tau^{n-1/4}, u]^2 + \frac{\alpha_2}{\tau}\mathcal{W}_2[u_\tau^{n-3/4}, u]^2 + \frac{\alpha_3}{\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u)\right\},$$

satisfying (92) if $\alpha_1 = 18$, $\alpha_2 = -38$ and $\alpha_3 = \frac{49}{2}$. We hence have the stage three minimising movement scheme:

$$u_\tau^n := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}} \left\{\frac{18}{\tau}\mathcal{W}_2[u_\tau^{n-1/4}, u]^2 - \frac{38}{\tau}\mathcal{W}_2[u_\tau^{n-3/4}, u]^2 + \frac{49}{2\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u)\right\}.$$

Hence we have solved to find the following minimising movement schemes per stage:

- For stage one:

$$\begin{cases} u_\tau^{n-3/4} := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}}\, \Phi_1^\tau(u_\tau^{n-1}; u), \\ \quad \Phi_1^\tau := \dfrac{2}{\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u). \end{cases}$$

- For stage two:

$$\begin{cases} u_\tau^{n-1/4} := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}}\, \Phi_2^\tau(u_\tau^{n-3/4}, u_\tau^{n-1}; u), \\ \quad \Phi_2^\tau := \dfrac{16}{\tau}\mathcal{W}_2[u_\tau^{n-3/4}, u]^2 - \dfrac{10}{\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u). \end{cases}$$

- For stage three:

$$\begin{cases} u_\tau^n := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}}\, \Phi_3^\tau(u_\tau^{n-1/4}, u_\tau^{n-3/4}, u_\tau^{n-1}; u), \\ \Phi_3^\tau := \dfrac{18}{\tau}\mathcal{W}_2[u_\tau^{n-1/4}, u]^2 - \dfrac{38}{\tau}\mathcal{W}_2[u_\tau^{n-3/4}, u]^2 + \dfrac{49}{2\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u). \end{cases}$$

Example 4.14. *By selecting for our intermediate time steps $c_1 = \frac{1}{4}$ and $c_2 = \frac{1}{2}$, this gives us our final Butcher array for a third order DIRK3 scheme:*

$$\begin{array}{c|ccc} 1/4 & 1/4 & & \\ 1/2 & 2/3 & -1/6 & \\ 1 & 4/9 & 1/3 & 1/9 \\ \hline & 5/9 & 1/3 & 1/9, \end{array}$$

providing us with the system of equations:

$$u_\tau^{n-3/4} = u_\tau^{n-1} - \frac{\tau}{4}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-3/4}), \tag{90a}$$

$$u_\tau^{n-1/2} = u_\tau^{n-1} - \frac{2\tau}{3}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-3/4}) + \frac{\tau}{6}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1/2}), \tag{90b}$$

$$u_\tau^n = u_\tau^{n-1} - \frac{4\tau}{9}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-3/4}) - \frac{\tau}{3}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1/2}) - \frac{2\tau}{9}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n). \tag{90c}$$

- *The stage one minimising movement scheme generates $u_\tau^{n-3/4}$, given by*

$$u_\tau^{n-3/4} := \underset{u \in \mathcal{P}_M(\Omega)}{\operatorname{argmin}} \left\{ \frac{2}{\tau} \mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u) \right\}.$$

- *The stage two minimisation scheme generates $u_\tau^{n-1/2}$: Calculating $3(90b) - 8(90a)$ gives*

$$-6u_\tau^{n-1/2} + 16u_\tau^{n-3/4} - 10u_\tau^{n-1} = -\tau \nabla_{\mathcal{W}_2} \mathcal{E}(u_\tau^{n-1/2}). \tag{91}$$

  The discrete solution at $t = t^{n-1/2}$, $u_\tau^{n-1/2}$ satisfies

$$u_\tau^{n-1/2} := \underset{u \in \mathcal{P}_M(\Omega)}{\operatorname{argmin}} \left\{ \frac{a}{\tau} \mathcal{W}_2[u_\tau^{n-3/4}, u]^2 + \frac{b}{\tau} \mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u) \right\},$$

  if given that $u_\tau^{n-1/2}$:

$$\frac{2(a+b)}{\tau} u_\tau^{n-1/2} - \frac{2a}{\tau} u_\tau^{n-3/4} - \frac{2b}{\tau} u_\tau^{n-1} = -\nabla_{\mathcal{W}_2} \mathcal{E}(u_\tau^{n-1/2}),$$

  satisfying (91) if $a = -8$ and $b = 5$. We hence have the stage two minimising movement scheme:

$$u_\tau^{n-1/2} := \underset{u \in \mathcal{P}_M(\Omega)}{\operatorname{argmin}} \left\{ -\frac{8}{\tau} \mathcal{W}_2[u_\tau^{n-3/4}, u]^2 + \frac{5}{\tau} \mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u) \right\}.$$

- *The stage three minimising movement scheme generates u_τ^n: Calculating $-9(90c) - 18(90b) + 64(90a)$ gives us*

$$-9u_\tau^n - 18u_\tau^{n-1/2} + 64u_\tau^{n-3/4} - 37u_\tau^{n-1} = 2\tau \nabla_{\mathcal{W}_2} \mathcal{E}(u_\tau^n). \tag{92}$$

  The discrete solution at $t = t^n$, u_τ^n satisfies

$$u_\tau^n := \underset{u \in \mathcal{P}_M(\Omega)}{\operatorname{argmin}} \left\{ \frac{\alpha_1}{\tau} \mathcal{W}_2[u_\tau^{n-1/2}, u]^2 + \frac{\alpha_2}{\tau} \mathcal{W}_2[u_\tau^{n-3/4}, u]^2 + \frac{\alpha_3}{\tau} \mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u) \right\},$$

  satisfying (92) if $\alpha_1 = -\frac{9}{2}$, $\alpha_2 = 16$ and $\alpha_3 = -\frac{37}{4}$. We hence have the stage three minimising movement scheme:

$$u_\tau^n := \underset{u \in \mathcal{P}_M(\Omega)}{\operatorname{argmin}} \left\{ -\frac{9}{2\tau} \mathcal{W}_2[u_\tau^{n-1/2}, u]^2 + \frac{16}{\tau} \mathcal{W}_2[u_\tau^{n-3/4}, u]^2 - \frac{37}{4\tau} \mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u) \right\}.$$

Hence we have solved to find the following minimising movement schemes per stage:

- For stage one:

$$\begin{cases} u_\tau^{n-3/4} := \underset{u \in \mathcal{P}_M(\Omega)}{\operatorname{argmin}} \Phi_1^\tau(u_\tau^{n-1}; u), \\ \qquad \Phi_1^\tau := \frac{2}{\tau} \mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u). \end{cases}$$

- For stage two:

$$\begin{cases} u_\tau^{n-1/2} := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}}\, \Phi_2^\tau(u_\tau^{n-3/4}, u_\tau^{n-1}; u), \\ \Phi_2^\tau := -\dfrac{8}{\tau}\mathcal{W}_2[u_\tau^{n-3/4}, u]^2 + \dfrac{5}{\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u). \end{cases}$$

- For stage three:

$$\begin{cases} u_\tau^{n} := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}}\, \Phi_3^\tau(u_\tau^{n-1/2}, u_\tau^{n-3/4}, u_\tau^{n-1}; u), \\ \Phi_3^\tau := -\dfrac{9}{2\tau}\mathcal{W}_2[u_\tau^{n-1/2}, u]^2 + \dfrac{16}{\tau}\mathcal{W}_2[u_\tau^{n-3/4}, u]^2 - \dfrac{37}{4\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u). \end{cases}$$

4.5.2 Scheme Two

Another key example of the DIRK3 minimising movement scheme is shown, which has been published, and will be used for our numerical experiments in Section 8. We now investigate another three stage, third order DIRK method which is L-stable from Ascher and Ruuth et al. [4, Thm. 5], that is

$$\begin{array}{c|ccc} \alpha_1 & \alpha_1 & & \\ \alpha_2 & \beta_1 & \alpha_1 & \\ 1 & \beta_2 & \beta_3 & \alpha_1 \\ \hline & \beta_2 & \beta_3 & \alpha_1, \end{array} \tag{93}$$

where $\alpha_1 = 0.4358665215$, $\alpha_2 = 0.7179332608$, $\beta_1 = 0.2820667392$, $\beta_2 = 1.208496649$, $\beta_3 = -0.644363171$.

This provides us with the following system of equations:

$$u_\tau^{n+\alpha_1-1} = u_\tau^{n-1} - \alpha_1\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+\alpha_1-1}), \tag{94a}$$

$$u_\tau^{n+\alpha_2-1} = u_\tau^{n-1} - \beta_1\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+\alpha_1-1}) - \alpha_1\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+\alpha_2-1}), \tag{94b}$$

$$u_\tau^{n} = u_\tau^{n-1} - \beta_2\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+\alpha_1-1}) - \beta_3\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+\alpha_2-1}) - \alpha_1\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n}). \tag{94c}$$

Corollary 4.15. *The minimising movement scheme for system (94a)-(94c) is as follows:*

- ***Stage One***

$$\begin{cases} u_\tau^{n+\alpha_1-1} := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}}\, \Phi_1^\tau(u_\tau^{n-1}; u), \\ \Phi_1^\tau := \dfrac{1}{2\alpha_1\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u). \end{cases} \tag{95}$$

- ***Stage Two***

$$\begin{cases} u_\tau^{n+\alpha_2-1} := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}}\, \Phi_2^\tau(u_\tau^{n+\alpha_1-1}, u_\tau^{n-1}; u), \\ \Phi_2^\tau := \dfrac{\beta_1}{2\alpha_1^2\tau}\mathcal{W}_2[u_\tau^{n+\alpha_1-1}, u]^2 + \dfrac{\alpha_1-\beta_1}{2\alpha_1^2\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u). \end{cases} \tag{96}$$

- ***Stage Three***

$$\begin{cases} u_\tau^n := & \operatorname*{argmin}_{u\in\mathcal{P}_M(\Omega)} \Phi_3^\tau(u_\tau^{n+\alpha_2-1}, u_\tau^{n+\alpha_1-1}, u_\tau^{n-1}; u), \\ \Phi_3^\tau := & \dfrac{\beta_3}{2\alpha_1^2\tau}\mathcal{W}_2[u_\tau^{n+\alpha_2-1}, u]^2 + \dfrac{\alpha_1\beta_2-\beta_1\beta_3}{2\alpha_1^3\tau}\mathcal{W}_2[u_\tau^{n+\alpha_1-1}, u]^2 + \\ & \dfrac{\alpha_1^2-\alpha_1(\beta_2+\beta_3)+\beta_1\beta_3}{2\alpha_1^3\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2. \end{cases} \tag{97}$$

Proof. Similarly, from Section 4.5.1, we start by eliminating the $\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+\alpha_2-1})$, $\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+\alpha_1-1})$, $\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n)$ terms respectively to retrieve the set of equations:

- Stage two: α_1·(94b) - β_1·(94a) gives

$$\frac{1}{\alpha_1}u_\tau^{n+\alpha_2-1} - \frac{\beta_1}{\alpha_1^2}u_\tau^{n+\alpha_1-1} + \frac{(\beta_1-\alpha_1)}{\alpha_1^2}u_\tau^{n-1} = -\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+\alpha_2-1}), \tag{98}$$

- Stage three: α_1·(94c) - β_3·(94b) + $\left(\dfrac{\beta_1\beta_3}{\alpha_1} - \beta_2\right)$·(94a) gives

$$\begin{aligned} \frac{1}{\alpha_1}u_\tau^n - \frac{\beta_3}{\alpha_1^2}u_\tau^{n+\alpha_2-1} - \left(\frac{\beta_2}{\alpha_1^2} - \frac{\beta_1\beta_3}{\alpha_1^3}\right)u_\tau^{n+\alpha_1-1} - \frac{1}{\alpha_1^2}\left(\alpha_1-\beta_2-\beta_3+\frac{\beta_1\beta_3}{\alpha_1}\right)u_\tau^{n-1} \\ = -\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n), \end{aligned} \tag{99}$$

where (98) helps determine the coefficients for the L^2 Wasserstein distances in Φ_2^τ, and (99) for Φ_3^τ (see below).

- **Stage One**: The minimising movement scheme for $u_\tau^{n+\alpha_1-1}$ has been demonstrated from earlier examples, i.e. for stage one:

$$u_\tau^{n+\alpha_1-1} := \operatorname*{argmin}_{u\in\mathcal{P}_M(\Omega)} \Phi_1^\tau(u_\tau^{n-1}; u), \qquad \Phi_1^\tau(u_\tau^{n-1}; u) := \frac{1}{2\alpha_1\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u).$$

- **Stage Two**: We propose the minimising movement scheme for $u_\tau^{n+\alpha_2-1}$ which is

$$\begin{aligned} u_\tau^{n+\alpha_2-1} &:= \operatorname*{argmin}_{u\in\mathcal{P}_M(\Omega)} \Phi_2^\tau(u_\tau^{n+\alpha_1-1}, u_\tau^{n-1}; u), \\ \Phi_2^\tau(u_\tau^{n+\alpha_1-1}, u_\tau^{n-1}; u) &:= \frac{(\alpha_2-\alpha_1)b}{\tau}\mathcal{W}_2[u_\tau^{n+\alpha_1-1}, u]^2 + \frac{\alpha_2 c}{\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u). \end{aligned}$$

  The minimiser $u_\tau^{n+\alpha_2-1}$ satisfies

$$\frac{2(\alpha_2-\alpha_1)b + 2\alpha_2 c}{\tau}u_\tau^{n+\alpha_2-1} - \frac{2(\alpha_2-\alpha_1)b}{\tau}u_\tau^{n+\alpha_1-1} - \frac{2\alpha_2 c}{\tau}u_\tau^{n-1} = -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n+\alpha_2-1}),$$

  which is (98) if $b = \dfrac{\beta_1}{2(\alpha_2-\alpha_1)\alpha_1^2}$ and $c = \dfrac{\alpha_1-\beta_1}{2\alpha_1^2\alpha_2}$.

- **Stage Three**: The minimising movement scheme for u_τ^n is also proposed which is

$$\begin{aligned} u_\tau^n := & \operatorname*{argmin}_{u\in\mathcal{P}_M(\Omega)} \Phi_3^\tau(u_\tau^{n+\alpha_2-1}, u_\tau^{n+\alpha_1-1}, u_\tau^{n-1}; u), \\ \Phi_3^\tau(u_\tau^{n+\alpha_2-1}, u_\tau^{n+\alpha_1-1}, u_\tau^{n-1}; u) := & \frac{(1-\alpha_2)p}{\tau}\mathcal{W}_2[u_\tau^{n+\alpha_2-1}, u]^2 + \frac{(1-\alpha_1)q}{\tau}\mathcal{W}_2[u_\tau^{n+\alpha_1-1}, u]^2 \\ & + \frac{r}{\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u). \end{aligned}$$

The minimiser u_τ^n satisfies

$$\begin{aligned} & \frac{2(1-\alpha_2)p + 2(1-\alpha_1)q + 2r}{\tau} u_\tau^n - \frac{2(1-\alpha_2)p}{\tau} u_\tau^{n+\alpha_2-1} - \frac{2(1-\alpha_1)q}{\tau} u_\tau^{n+\alpha_1-1} - \frac{2r}{\tau} u_\tau^{n-1} \\ = & -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n), \end{aligned}$$

which is (99) if the following parameters p, q, r are satisfied:

$$p = \frac{\beta_3}{2\alpha_1^2(1-\alpha_2)}, \quad q = \frac{\alpha_1\beta_2 - \beta_1\beta_3}{2\alpha_1^3(1-\alpha_1)}, \quad r = \frac{\alpha_1^2 - \alpha_1(\beta_2+\beta_3) + \beta_1\beta_3}{2\alpha_1^3}.$$

The proof is complete. □

4.6 Minimising Movement Scheme: Five stage Runge-Kutta (DIRK5) Scheme

The DIRK5 minimising movement scheme is constructed. Again due to the tediousness of the construction, we only go from an example which has been published and has order of accuracy four.

The second example for the third order Runge-Kutta method provides a much improved error and third order numerical convergence, but does a L-stable fourth order Runge-Kutta method provide anything better? We introduce the Butcher array from [25], by Hairer and Wanner, which is

$$\begin{array}{c|ccccc} 1/4 & 1/4 & & & & \\ 3/4 & 1/2 & 1/4 & & & \\ 11/20 & 17/50 & -1/25 & 1/4 & & \\ 1/2 & 371/1360 & -137/2720 & 15/544 & 1/4 & \\ 1 & 25/24 & -49/48 & 125/16 & -85/12 & 1/4 \\ \hline & 25/24 & -49/48 & 125/16 & -85/12 & 1/4, \end{array}$$

providing a system of equations:

$$u_\tau^{n-3/4} = u_\tau^{n-1} - \frac{\tau}{4}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-3/4}), \tag{100a}$$

$$u_\tau^{n-1/4} = u_\tau^{n-1} - \frac{\tau}{2}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-3/4}) - \frac{\tau}{4}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1/4}), \tag{100b}$$

$$u_\tau^{n-9/20} = u_\tau^{n-1} - \frac{17\tau}{50}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-3/4}) + \frac{\tau}{25}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1/4}) - \frac{\tau}{4}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-9/20}), \tag{100c}$$

$$
\begin{aligned}
u_\tau^{n-1/2} = \quad & u_\tau^{n-1} - \frac{371\tau}{1360}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-3/4}) + \frac{137\tau}{2720}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1/4}) \\
& - \frac{15\tau}{544}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-9/20}) - \frac{\tau}{4}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1/2}),
\end{aligned} \tag{100d}
$$

$$
\begin{aligned}
u_\tau^{n} = \quad & u_\tau^{n-1} - \frac{25\tau}{24}\nabla_{\mathcal{W}_2}\mathcal{E}(u^{n-3/4}) + \frac{49\tau}{48}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1/4}) - \frac{125\tau}{16}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-9/20}) \\
& + \frac{85\tau}{12}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1/2}) - \frac{\tau}{4}\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n}).
\end{aligned} \tag{100e}
$$

Corollary 4.16. *From the system (100a)-(100e), the minimising movement scheme for the DIRK5 scheme is as follows for each stage:*

- ***Stage One***

$$
\begin{cases}
u_\tau^{n-3/4} := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}} \Phi_{5,1}^\tau(u_\tau^{n-1};u), \\
\quad \Phi_{5,1}^\tau := \frac{2}{\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2 + \mathcal{E}(u).
\end{cases} \tag{101}
$$

- ***Stage Two***

$$
\begin{cases}
u_\tau^{n-1/4} := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}} \Phi_{5,2}^\tau(u_\tau^{n-1},u_\tau^{n-3/4};u), \\
\quad \Phi_{5,2}^\tau := \frac{4}{\tau}\mathcal{W}_2[u_\tau^{n-3/4},u]^2 - \frac{2}{\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2 + \mathcal{E}(u).
\end{cases} \tag{102}
$$

- ***Stage Three***

$$
\begin{cases}
u_\tau^{n-9/20} := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}} \Phi_{5,3}^\tau(u_\tau^{n-1},u_\tau^{n-3/4},u_\tau^{n-1/4};u), \\
\quad \Phi_{5,3}^\tau := -\frac{8}{25\tau}\mathcal{W}_2[u_\tau^{n-1/4},u]^2 + \frac{84}{25\tau}\mathcal{W}_2[u_\tau^{n-3/4},u]^2 - \frac{26}{25\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2 + \mathcal{E}(u).
\end{cases} \tag{103}
$$

- ***Stage Four***

$$
\begin{cases}
u_\tau^{n-1/2} = \quad \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}} \Phi_{5,4}^\tau(u_\tau^{n-1},u_\tau^{n-3/4},u_\tau^{n-1/4},u_\tau^{n-9/20};u), \\
\quad \Phi_{5,4}^\tau = \frac{15}{68\tau}\mathcal{W}_2[u_\tau^{n-9/20},u]^2 - \frac{25}{68\tau}\mathcal{W}_2[u_\tau^{n-1/4},u]^2 + \frac{89}{34\tau}\mathcal{W}_2[u_\tau^{n-3/4},u]^2 \\
\qquad - \frac{8}{17\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2 + \mathcal{E}(u).
\end{cases} \tag{104}
$$

- ***Stage Five***

$$
\begin{cases}
u_\tau^{n} = \quad \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}} \Phi_{5,5}^\tau(u_\tau^{n-1},u_\tau^{n-3/4},u_\tau^{n-1/4},u_\tau^{n-9/20},u_\tau^{n-1/2};u), \\
\Phi_{5,5}^\tau = -\frac{170}{3\tau}\mathcal{W}_2[u_\tau^{n-1/2},u]^2 + \frac{275}{4\tau}\mathcal{W}_2[u_\tau^{n-9/20},u]^2 - \frac{103}{12\tau}\mathcal{W}_2[u_\tau^{n-1/4},u]^2 \\
\qquad - \frac{37}{6\tau}\mathcal{W}_2[u_\tau^{n-3/4},u]^2 + \frac{14}{3\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2 + \mathcal{E}(u).
\end{cases} \tag{105}
$$

Proof. Similarly, from earlier examples, we eliminate the $\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-3/4})$, $\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-9/20})$, $\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1/4})$, $\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1/2})$ terms respectively to retrieve the set of equations, from which we can construct a set of minimising movement schemes for each stage:

- Stage two: (100b) - 2·(100a) gives

$$4u_\tau^{n-1/4} - 8u_\tau^{n-3/4} + 4u_\tau^{n-1} = -\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1/4}). \tag{106}$$

- Stage three: $\dfrac{25}{2}$·(100c) + 2·(100b) - 21·(100a) gives

$$4u_\tau^{n-9/20} + \frac{16}{25}u_\tau^{n-1/4} - \frac{168}{25}u_\tau^{n-3/4} + \frac{52}{25}u_\tau^{n-1} = -\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-9/20}). \tag{107}$$

- Stage four: 136·(100d) - 15·(100c) + 25·(100b) - 178·(100a) gives

$$4u_\tau^{n-1/2} - \frac{15}{34}u_\tau^{n-9/20} + \frac{25}{34}u_\tau^{n-1/4} - \frac{89}{17}u_\tau^{n-3/4} + \frac{16}{17}u_\tau^{n-1} = -\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1/2}). \tag{108}$$

- Stage five: 24·(100e) + 680·(100d) - 825·(100c) + 103·(100b) + 74·(100a) gives

$$4u_\tau^{n} + \frac{340}{3}u_\tau^{n-1/2} - \frac{275}{2}u_\tau^{n-9/20} + \frac{103}{6}u_\tau^{n-1/4} + \frac{37}{3}u_\tau^{n-3/4} - \frac{28}{3}u_\tau^{n-1} = -\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n}). \tag{109}$$

where (106) helps determine the coefficients for the L^2 Wasserstein distances in Φ_2^τ, (107) for Φ_3^τ, (108) for Φ_4^τ and (109) for Φ_5^τ (see below).

Now we have the tools laid out to construct the minimising movement schemes for each stage:

- **Stage One**: The minimising movement scheme for the minimiser $u_\tau^{n-3/4}$ has been demonstrated from earlier examples hence verifies (101), so we work on the minimising movement schemes for $u_\tau^{n-1/4}$, $u_\tau^{n-9/20}$, $u_\tau^{n-1/2}$ and u_τ^{n}.

- **Stage Two**: We propose the minimising movement scheme for the minimiser $u_\tau^{n-1/4}$ which is

$$u_\tau^{n-1/4} := \operatorname*{argmin}_{u\in\mathcal{P}_M(\Omega)} \Phi_{5,2}^\tau(u_\tau^{n-1}, u_\tau^{n-3/4}; u),$$

$$\Phi_{5,2}^\tau(u_\tau^{n-1}, u_\tau^{n-3/4}; u) := \frac{2b}{\tau}\mathcal{W}_2[u_\tau^{n-3/4}, u]^2 + \frac{4c}{3\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u).$$

The minimiser $u_\tau^{n-1/4}$ satisfies

$$\frac{12b+8c}{3\tau}u_\tau^{n-1/4} - \frac{4b}{\tau}u_\tau^{n-3/4} - \frac{8c}{3\tau}u_\tau^{n-1} = -\nabla\mathcal{E}(u_\tau^{n-1/4}),$$

which is (106) if $b = 2$ and $c = -\dfrac{3}{2}$, hence verifies (102).

- **Stage Three**: We now propose the minimising movement scheme for the minimiser $u_\tau^{n-9/20}$ which is

$$u_\tau^{n-9/20} := \operatorname*{argmin}_{u\in\mathcal{P}_M(\Omega)} \Phi_{5,3}^\tau(u_\tau^{n-1}, u_\tau^{n-3/4}, u_\tau^{n-1/4}; u),$$

$$\begin{aligned}\Phi_{5,3}^\tau(u_\tau^{n-1}, u_\tau^{n-3/4}, u_\tau^{n-1/4}; u) := {}& -\frac{5d}{\tau}\mathcal{W}_2[u_\tau^{n-1/4}, u]^2 + \frac{10e}{3\tau}\mathcal{W}_2[u_\tau^{n-3/4}, u]^2 \\ & + \frac{20f}{11\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u).\end{aligned}$$

The minimiser $u_\tau^{n-9/20}$ satisfies

$$\frac{1}{\tau}\left(\frac{20e}{3}+\frac{40f}{11}-10d\right)u_\tau^{n-9/20}+\frac{10d}{\tau}u_\tau^{n-1/4}-\frac{20e}{3\tau}u_\tau^{n-3/4}-\frac{40f}{11\tau}u_\tau^{n-1}=-\nabla\mathcal{E}(u_\tau^{n-9/20}),$$

which is (107) if $d=\frac{8}{125}$, $e=\frac{126}{125}$ and $f=-\frac{143}{250}$, hence verifies (103).

- **Stage Four**: We now propose the minimising movement scheme for the minimiser $u_\tau^{n-1/2}$ which is

$$u_\tau^{n-1/2}:=\operatorname*{argmin}_{u\in\mathcal{P}_M(\Omega)}\Phi_{5,4}^\tau(u_\tau^{n-1},u_\tau^{n-3/4},u_\tau^{n-1/4},u_\tau^{n-9/20};u),$$

$$\begin{aligned}\Phi_{5,4}^\tau(u_\tau^{n-1},u_\tau^{n-3/4},u_\tau^{n-1/4},u_\tau^{n-9/20};u):=&-\frac{20p}{\tau}\mathcal{W}_2[u_\tau^{n-9/20},u]^2-\frac{4q}{\tau}\mathcal{W}_2[u_\tau^{n-1/4},u]^2\\&+\frac{4r}{\tau}\mathcal{W}_2[u_\tau^{n-3/4},u]^2+\frac{2s}{\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2+\mathcal{E}(u).\end{aligned}$$

The minimiser $u_\tau^{n-1/2}$ satisfies

$$\frac{8r+4s-40p-8q}{\tau}u_\tau^{n-1/2}+\frac{40p}{\tau}u_\tau^{n-9/20}+\frac{8q}{\tau}u_\tau^{n-1/4}-\frac{8r}{\tau}u_\tau^{n-3/4}-\frac{4s}{\tau}u_\tau^{n-1}=-\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^{n-1/2}),$$

which is (108) if $p=-\frac{3}{272}$, $q=\frac{25}{272}$, $r=\frac{89}{136}$ and $s=-\frac{4}{17}$, hence verifies (104).

- **Stage Five**: We finally propose the minimising movement scheme for the minimiser u_τ^n which is

$$u_\tau^n:=\operatorname*{argmin}_{u\in\mathcal{P}_M(\Omega)}\Phi_{5,5}^\tau(u_\tau^{n-1},u_\tau^{n-3/4},u_\tau^{n-1/4},u_\tau^{n-9/20},u_\tau^{n-1/2};u),$$

$$\begin{aligned}&\Phi_{5,5}^\tau(u_\tau^{n-1},u_\tau^{n-3/4},u_\tau^{n-1/4},u_\tau^{n-9/20},u_\tau^{n-1/2};u)\\:=&\ \frac{2a}{\tau}\mathcal{W}_2[u_\tau^{n-1/2},u]^2+\frac{20b}{9\tau}\mathcal{W}_2[u_\tau^{n-9/20},u]^2+\frac{4c}{\tau}\mathcal{W}_2[u_\tau^{n-1/4},u]^2+\frac{4d}{3\tau}\mathcal{W}_2[u_\tau^{n-3/4},u]^2\\&+\frac{e}{\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2+\mathcal{E}(u).\end{aligned}$$

The minimiser u_τ^n satisfies

$$\begin{aligned}&\frac{1}{\tau}\left(4a+\frac{40b}{9}+8c+\frac{8d}{3}+2e\right)u_\tau^n-\frac{4a}{\tau}u_\tau^{n-1/2}-\frac{40b}{9\tau}u_\tau^{n-9/20}-\frac{8c}{\tau}u_\tau^{n-1/4}-\frac{8d}{3\tau}u_\tau^{n-3/4}-\frac{2e}{\tau}u_\tau^{n-1}\\&=-\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n),\end{aligned}$$

which is (109) if $a=-\frac{85}{3}$, $b=\frac{495}{16}$, $c=-\frac{103}{48}$, $d=-\frac{37}{8}$ and $e=\frac{14}{3}$, hence verifies (105).

Thus the proof for the schemes per stage (101)-(105) are complete. □

4.7 A-stabilty and L-stability of SDIRK Methods (See [25])

We finish by showing that the equivalent last two rows of the Butcher array guarantees L-stability for not only DIRK2 but for a general number of stages.

Unlike for the DIRK2 scheme case, a similar version of proving A-stability is too complicated for the third order DIRK3 case and higher stages. But, as explained in [25, Table 6.3, p.103-104], we collect that an SDIRK3 scheme (DIRK3 scheme but with all the leading diagonal elements equal) is A-stable if the leading diagonal elements $a_{11} = a_{22} = a_{33}$ is such that $a_{ii} \in [1/3, 1.07]; \quad i = 1, 2, 3$. Furthermore, a SDIRK5 scheme is also shown there, which again from the same citation is A-stable if the leading diagonal elements $a_{11} = a_{22} = \ldots a_{55}$ is such that

$$a_{ii} \in [0.247, 0.362] \cup [0.421, 0.473]; \quad i = 1, 2, \ldots 5.$$

Now onto the next lemma, produced by Hairer and Wanner, [25, Prop. 3.8, p.45], showing that all "stiffly accurate" DIRK schemes are L-stable.

Lemma 4.17. *All A-stable DIRK schemes with the last step being equivalent to the last intermediate step i.e.* $b_i = a_{si}; \quad i = 1, 2, \ldots s$ *are L-stable.*

Proof. We pay attention to part of (67), where we can transfer the z part into the inverse operation i.e.

$$zb^T(\mathbb{I} - Az)^{-1} = b^T[z^{-1}(\mathbb{I} - Az)]^{-1}.$$

Since we can rewrite in matrix form:

$$z^{-1}(\mathbb{I} - Az) = \frac{1}{z}\begin{bmatrix} 1 - a_{11}z & 0 & \ldots & 0 \\ -a_{21}z & 1 - a_{22}z & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ -a_{s1}z & -a_{s2}z & \ldots & 1 - a_{ss}z \end{bmatrix},$$

and as $z \to \infty$, you can clearly observe that $z^{-1}(\mathbb{I} - Az) \to -A$ and thus

$$\lim_{z\to\infty} zb^T(\mathbb{I} - Az)^{-1} = -b^T A^{-1}.$$

Since $a_{sj} = b_j$ and by denoting $e_s = [0 \quad 0 \quad \cdots \quad 1]^T$, we have that

$$A^T e_s = [a_{s1} \quad a_{s2} \quad \ldots \quad a_{ss}]^T = [b_1 \quad b_2 \quad \ldots \quad b_s]^T.$$

In other words, $A^T e_s = b$. By transposing both sides, giving us $e_s^T A = b^T$ and applying A^{-1} to the right on both sides, this gives us $e_s^T = b^T A^{-1}$.

Thus from earlier we have the final result:

$$\lim_{z\to\infty} 1 + zb^T(I - Az)^{-1} = 1 - b^T A^{-1} = 1 - e_s^T \mathbf{1} = 1 - 1 = 0.$$

∎

5 Higher order generalisations of the Minimising Movement Scheme

We introduce the pitfull for higher order minimising movement schemes, where the energy is not monotonically decreasing. We build in some estimates on our energy functionals and hence verify that the metric dissipates for decreasing time step size.

We start by recalling *gradient flows* in the probability space of smooth energy functionals $\mathcal{E} : \mathcal{P}_M(\Omega) \to \mathbb{R}$, solving the problem

$$\partial_t u(x,t) = -\nabla_{\mathcal{W}_2}\mathcal{E}(u(x,t)), \qquad u(x,0) = u_0, \qquad u \in \mathcal{P}_M(\Omega).$$

This has a unique solution provided that $\nabla_{\mathcal{W}_2}\mathcal{E}(\cdot)$ is Lipschitz continuous in $\mathcal{P}_M(\Omega)$ (i.e. $\mathcal{E}(\cdot) \in C^{1,1}(\mathcal{P}_M(\Omega))$ [35, p. 1]. However, well-posedness also follows from the assumption that $\mathcal{E}(\cdot)$ is uniformly semi-convex [41, Prop. 8.1].

The aim is to find the curve of steepest descent of $\mathcal{E}(\cdot)$ from the initial point u_0 [42]. A semi-discretisation for the problem is achieved by means of the minimising movement scheme, recalled from the previous section.

This section provides us with the tools required for this, before proving the uniqueness result. Various assumptions (semi-continuity, coercivity and semi-convexity from [35]) will be derived later, but beforehand we recall the famously known minimising movement scheme (also known as the JKO scheme) originally proposed by E. De Giorgi [14] and used by Gallouët and Monsaingeon [19]. Also, we begin deriving some estimates with the basic assumption that the energy is finite for initial time.

5.1 Introduction to the Minimising Movement Scheme

The minimising movement scheme is recalled. The evolution equation for solving gradient flows (17), also referred to as the Cauchy problem, can be semi-discretised in time, using the minimising movement scheme, which enables us to find a sequence u_τ^n as follows. For fixed $\tau > 0$:

$$u_\tau^n := \operatorname*{argmin}_{u\in\mathcal{P}_M(\Omega)} \frac{1}{2\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u).$$

By generating a semi-discrete solution at the next time step u_τ^n which minimises $\frac{1}{2\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \mathcal{E}(u)$, we have that the minimiser u_τ^n satisfies

$$\nabla_{\mathcal{W}_2}\left(\mathcal{E}(u) + \frac{\mathcal{W}_2[u_\tau^{n-1}, u]^2}{2\tau}\right)\Big|_{u=u_\tau^n} = 0,$$

which is equivalent to the implicit Euler scheme (or the backward difference formula one (BDF1) scheme), see [42, p.6] for details. This scheme guarantees strong stability properties, that is A-stability and L-stability (we have shown how from Lemmas 4.9 and 4.17).

5.2 Minimising Movement Schemes - Backward Difference Formula 1 (BDF1)

We show the monotonicity of the energy functional at the time-discrete level for the BDF1 scheme. We recall the minimising movement scheme for the BDF1 scheme as the penalisation of the energy functional $\mathcal{E}(\cdot)$ i.e.

$$\Phi_1^\tau(u_\tau^{n-1};\cdot) : \mathcal{P}_M(\Omega) \to \mathbb{R} \cup \{\infty\}, \tag{110a}$$

$$\Phi_1^\tau(u_\tau^{n-1};u) := \frac{1}{2\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2 + \mathcal{E}(u). \tag{110b}$$

From the priori estimates in [34], by Matthes and Osberger, it was shown that the energy functional $\mathcal{E}(\cdot)$ is monotonically decreasing i.e. with u_τ^n being the minimiser of the Yosida-regularised function $\Phi_1^\tau(u_\tau^{n-1};u)$, we have

$$\begin{aligned} &\Phi_1^\tau(u_\tau^{n-1};u_\tau^n) \leq \Phi_1^\tau(u_\tau^{n-1};u_\tau^{n-1}) \\ \Leftrightarrow &\frac{1}{2\tau}\mathcal{W}_2[u_\tau^{n-1},u_\tau^n]^2 + \mathcal{E}(u_\tau^n) \leq \mathcal{E}(u_\tau^{n-1}) \Rightarrow \mathcal{E}(u_\tau^n) \leq \mathcal{E}(u_\tau^{n-1}). \end{aligned}$$

Hence, from this, we have from the semi-discrete form of the gradient flow problem (17) that as we progress over time

$$u_\tau^n - u_\tau^{n-1} \to 0 \quad \text{as} \quad \tau \to 0,$$

implying convergence of the discrete solution to the actual solution with respect to the L^2-Wasserstein metric.

Now we introduce higher order BDF schemes and the limitations of their respective gradient flow structures, which formally leads to the main contribution in this thesis.

5.3 Minimising Movement Schemes - Backward Difference Formulas 2 to 6 (BDF2 to 6)

This part introduces how the energy is not monotonically decreasing at the time-discrete level. Also comments on how the BDF3 to 6 schemes are not A-stable and hence why we fast-track to the DIRK schemes for adapting the variational form of the minimising movement scheme for the BDF2 scheme, shown by Matthes and Plazotta [35].

5.3.1 BDF2 Minimising Movement Scheme

The basic minimising movement schemes were introduced in [14], by De Georgi. This scheme was extended for second order in time situations by G. Legendre et al. [32]. Therefore, as implemented for the simple BDF1 scheme, we recall the minimising movement scheme for the BDF2 scheme as the "penalisation" of the energy functional $\mathcal{E}(\cdot)$:

$$\Phi_2^\tau(u_\tau^{n-2}, u_\tau^{n-1}; \cdot) : \mathcal{P}_M(\Omega) \to \mathbb{R} \cup \{\infty\}, \tag{111a}$$

$$\Phi_2^\tau(u_\tau^{n-2}, u_\tau^{n-1}; u) := \frac{1}{\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 - \frac{1}{4\tau}\mathcal{W}_2[u_\tau^{n-2}, u]^2 + \mathcal{E}(u). \tag{111b}$$

This scheme, as a consequence to their results from [35], is a motivation to extending the analytical convergence approach for multistage schemes e.g. DIRK schemes.

Indeed, we have shown above that the energy functional is dissipative for the BDF1 schemes theoretically in the last subsection. This is shown numerically also for the BDF2 to 6 schemes. However $\mathcal{E}(\cdot)$ is not shown to be theoretically dissipating when applying the higher order BDF schemes. In fact, as shown in [35], we have that the energy functional is *only almost* dissipative for the BDF2 type schemes, i.e. from the following proposition:

Proposition 5.1. *With u_τ^n being the minimiser of $\Phi_2^\tau(u_\tau^{n-2}, u_\tau^{n-1}; u)$, from (111b), we have both*

$$\begin{aligned}
&\Phi_2^\tau(u_\tau^{n-2}, u_\tau^{n-1}; u_\tau^n) \leq \Phi_2^\tau(u_\tau^{n-2}, u_\tau^{n-1}; u_\tau^{n-1}) \\
&\Leftrightarrow \frac{1}{\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2 - \frac{1}{4\tau}\mathcal{W}_2[u_\tau^{n-2}, u_\tau^n]^2 + \mathcal{E}(u_\tau^n) \leq \mathcal{E}(u_\tau^{n-1}) - \frac{1}{4\tau}\mathcal{W}_2[u_\tau^{n-2}, u_\tau^{n-1}]^2 \\
&\Rightarrow \mathcal{E}(u_\tau^n) \leq \mathcal{E}(u_\tau^{n-1}) + \frac{1}{4\tau}\mathcal{W}_2[u_\tau^{n-2}, u_\tau^n]^2,
\end{aligned} \tag{112a}$$

and

$$\begin{aligned}
&\Phi_2^\tau(u_\tau^{n-2}, u_\tau^{n-1}; u_\tau^n) \leq \Phi_2^\tau(u_\tau^{n-2}, u_\tau^{n-1}; u_\tau^{n-2}) \\
&\Leftrightarrow \frac{1}{\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2 - \frac{1}{4\tau}\mathcal{W}_2[u_\tau^{n-2}, u_\tau^n]^2 + \mathcal{E}(u_\tau^n) \leq \mathcal{E}(u_\tau^{n-1}) + \frac{1}{\tau}\mathcal{W}_2[u_\tau^{n-2}, u_\tau^{n-1}]^2 \\
&\Rightarrow \mathcal{E}(u_\tau^n) \leq \mathcal{E}(u_\tau^{n-1}) + \frac{1}{\tau}\mathcal{W}_2[u_\tau^{n-2}, u_\tau^{n-1}]^2 + \frac{1}{4\tau}\mathcal{W}_2[u_\tau^{n-2}, u_\tau^n]^2.
\end{aligned} \tag{112b}$$

With this inconclusive information about the monotonicity of the energy functional for the BDF2 scheme, we aim to conclude convergence of other higher order schemes via a comparison principle. This approach has already been used in [35] for the BDF2 scheme.

5.3.2 BDF3 to 6 Minimising Movement Schemes

However, since the BDF3 to BDF6 schemes are not A-stable and do not demonstrate a clear improvement of the L^2-numerical convergence rate in numerical experiments, in comparison to BDF1-2, it would be impractical to extend this to these schemes. But on the other hand, we will numerically present this in Section 8 to illustrate clearly.

Instead we shall look into extending the comparison principle approach from [35] to DIRK schemes. We start by recalling the DIRK2 scheme, as well as introducing the minimising movement scheme for each stage. Then, in the subsequent subsections, we obtain some estimates on the energy functional $\mathcal{E}(\cdot)$.

5.4 Diagonally Implicit Runge-Kutta Two Stage (DIRK2) Minimising Movement Schemes

We recall the DIRK2 scheme with two diagrams illustrating the idea for each of the two stages. As mentioned above, backward difference formulas of higher order do not have the desirable stability properties in comparison to BDF1 and 2, as we just explained.

From now on in this thesis, we shall investigate diagonally implicit Runge-Kutta schemes, which provide L-stability, as explained in detail from the previous section. In other words, by solving at each new time step individually, these schemes can lead to higher orders of accuracy, and from what you will see from Section 8, improved numerical errors. However, (see the next Remark below), many DIRK schemes that have been published, may have high order of accuracy overall, but *only* have stage order of one for latter stages which may restrict us when it comes to error intolerances.

Furthermore, they have been well used in many practical applications, including fluid dynamics, medicine and gas transmission networks [45]. Before we commence our contribution, we briefly summarise the scheme, from Section 4.4:

For stage one, we recall (see equation (63)) the minimising movement scheme for DIRK2 stage one as the penalisation of the energy functional $\mathcal{E}(\cdot)$:

$$\Phi^{\tau}_{2,1}(u^{n-1}_{\tau};\cdot):\mathcal{P}_M(\Omega)\to\mathbb{R}\cup\{\infty\}, \tag{113a}$$

$$\Phi^{\tau}_{2,1}(u^{n-1}_{\tau};u):=\frac{1}{2a\tau}\mathcal{W}_2[u^{n-1}_{\tau},u]^2+\mathcal{E}(u). \tag{113b}$$

We have that the piecewise constant interpolations of the discrete solutions u^{n+a-1}_{τ}, defined for all time $t>0$, gives us the interpolated solution for $t\in((n-1)\tau,(n+a-1)\tau]$ (see Figure 3) and $n\in\mathbb{N}$:

$$\overline{u}_{\tau}(0):=u_0,\quad \overline{u}_{\tau}(t):=u^{n+a-1}_{\tau}.$$

Then we also recall (see equation (64)) the minimising movement scheme for DIRK2 stage two as the penalisation of the energy functional $\mathcal{E}(\cdot)$ i.e.

$$\Phi^{\tau}_{2,2}(u^{n-1}_{\tau},u^{n+a-1}_{\tau};\cdot):\mathcal{P}_M(\Omega)\to\mathbb{R}\cup\{\infty\}, \tag{114a}$$

$$\Phi^{\tau}_{2,2}(u^{n-1}_{\tau},u^{n+a-1}_{\tau};u):=-\frac{1-2a(1-a)}{2a(1-2a)\tau}\mathcal{W}_2[u^{n-1}_{\tau},u]^2+\frac{1}{2a(1-2a)\tau}\mathcal{W}_2[u^{n+a-1}_{\tau},u]^2+\mathcal{E}(u). \tag{114b}$$

We have that the piecewise constant interpolations of the discrete solutions u^{n}_{τ}, for all time $t>0$, gives us the interpolated solution for $t\in((n+a-1)\tau,n\tau]$ (see Figure 4) and $n\in\mathbb{N}$:

$$\overline{u}_{\tau}(0):=u_0,\quad \overline{u}_{\tau}(t):=u^{n}_{\tau}.$$

This is the method in the DIRK2 case. The next subsection recalls from Section 4.4 how (113b) and (114b) are obtained.

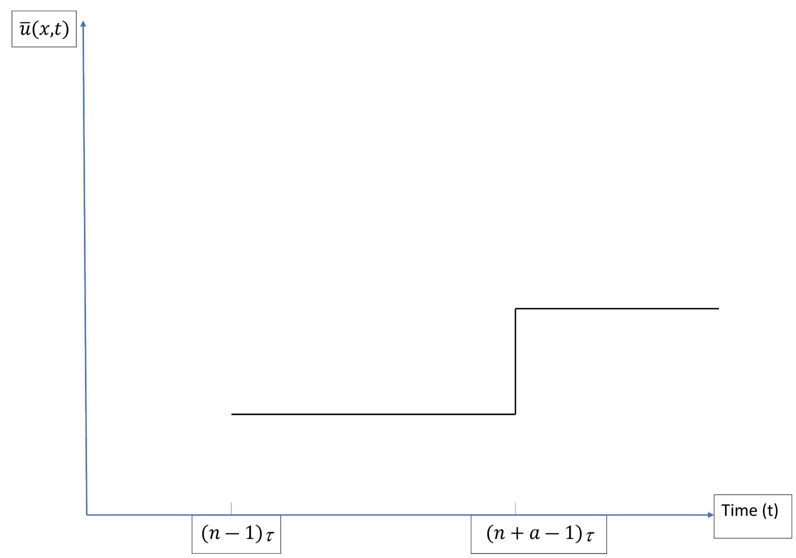


Figure 3: Piecewise constant interpolated solution for stage one of the DIRK2 scheme.

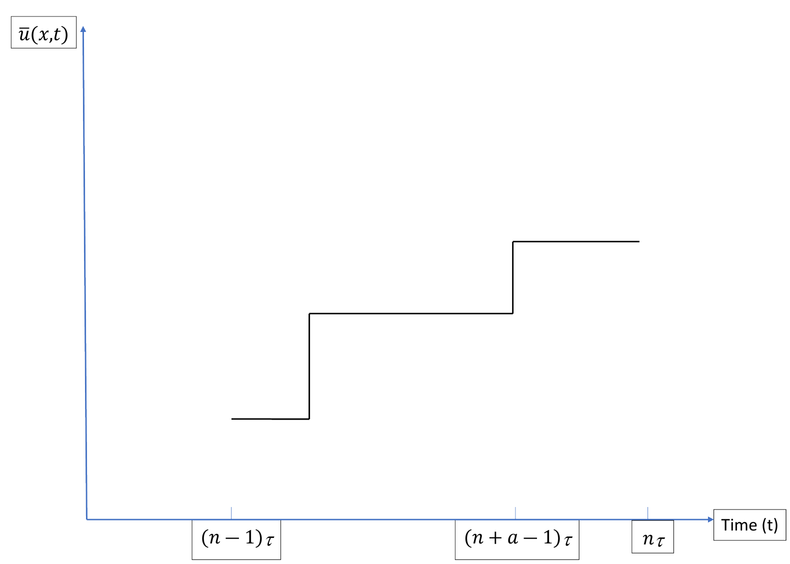


Figure 4: Piecewise constant interpolated solution for stage two of the DIRK2 scheme.

5.5 Derivation of the DIRK2 Scheme

To recap from Section 4.4, for a DIRK2 scheme to have order two, we require a Butcher array of the form (60), where the entries are found such that $\sum_{j=0}^{i} a_{ij} = c_i$ (refer back to the general Runge-Kutta scheme (49)), where i defines the stage of the scheme. Furthermore, $c_i = 1$ when $i = 2$ and $a_{11} = c_1$.

In contradicting the system (60), only a small order of accuracy of one is expected, hence only giving this scheme a similar, unimproved level of accuracy compared to the implicit Euler (BDF1) scheme.

Remark 5.2. Note that, from this scheme, each stage has an order of accuracy of one, despite the entire scheme being of order two.

5.6 Auxillian/Estimates for our Minimising Movement Schemes

We apply some estimate/inequalities from the DIRK2 scheme, given that u_τ^{n+a-1} and u_τ^n are minimisers for stages one and two respectively.

As part of our main contribution, we aim to find an estimate of $\mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - \mathcal{W}_2[u_\tau^0, v_\eta^0]^2$, where u_τ and v_η are assumed to be two different discrete solutions to (17), with N, M and τ, η representing the number of time step intervals and the time step sizes respectively. Hence, when combined with iterations and summations, the minimising movement schemes (63) and (64) brings out some valuable estimates, of our energy functionals, and later, metric terms, that will assist us in proving that

$$\mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - C_1\mathcal{W}_2[u_\tau^0, v_\eta^0]^2 \leq C_2\tau. \tag{115}$$

These estimates will mainly be used in the latter stages of the comparison principle proof. The estimates are mainly a consequence of u_τ^{n+a-1} and u_τ^n being the minimisers for the Yosida-regularised functionals $\Phi_{2,1}^\tau(u_\tau^{n-1}; u)$ and $\Phi_{2,2}^\tau(u_\tau^{n-1}, u_\tau^{n+a-1}; u)$ respectively. We have from substituting $u = u_\tau^{n-1}, u_\tau^{n+a-1}$ into our two minimising movement schemes (113b) and (114b):

$$\Phi_{2,1}^\tau(u_\tau^{n-1}; u_\tau^{n-1}) = \mathcal{E}(u_\tau^{n-1}), \tag{116a}$$

$$\Phi_{2,2}^\tau(u_\tau^{n-1}, u_\tau^{n+a-1}; u_\tau^{n-1}) = \frac{1}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]^2 + \mathcal{E}(u_\tau^{n-1}), \tag{116b}$$

$$\Phi_{2,2}^\tau(u_\tau^{n-1}, u_\tau^{n+a-1}; u_\tau^{n+a-1}) = -\frac{1-2a(1-a)}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]^2 + \mathcal{E}(u_\tau^{n+a-1}), \tag{116c}$$

and hence, referring to our minimisers for each stage:

- **Stage One**: With the estimate, since u_τ^{n+a-1} minimises the potential $\Phi_{2,1}(u_\tau^{n-1}, u)$, that is from (116a),

$$\Phi_{2,1}^\tau(u_\tau^{n-1}; u_\tau^{n+a-1}) \leq \Phi_{2,1}^\tau(u_\tau^{n-1}; u_\tau^{n-1}), \tag{117}$$

  we have the following proposition:

Proposition 5.3. *In equation (117), the sequences of discrete solutions* $(u_\tau^{n+a-1})_{n\in\mathbb{N}}$ *and* $(u_\tau^{n-1})_{n\in\mathbb{N}}$ *satisfy*

$$\frac{1}{2a\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]^2 + \mathcal{E}(u_\tau^{n+a-1}) \leq \mathcal{E}(u_\tau^{n-1}) \quad \Rightarrow \quad \mathcal{E}(u_\tau^{n+a-1}) \leq \mathcal{E}(u_\tau^{n-1}). \tag{118}$$

The result (118) *is achieved by the fact that a squared metric term is non-negative.*

Example 5.4. *We have that the inequality for (118) when* $a = 1/4$ *is*

$$\frac{2}{\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n-3/4}]^2 + \mathcal{E}(u_\tau^{n-3/4}) \leq \mathcal{E}(u_\tau^{n-1}) \quad \Rightarrow \quad \mathcal{E}(u_\tau^{n-3/4}) \leq \mathcal{E}(u_\tau^{n-1}). \tag{119}$$

That is, $u_\tau^{n-3/4}$ *minimises the potential* $\Phi_{2,1}(u_\tau^{n-1}; u)$*.*

- **Stage Two**: With the two following inequalities, since u_τ^n minimises the potential $\Phi_{2,2}^\tau(u_\tau^{n-1}, u_\tau^{n+a-1}; u)$, that is from (116b,116c),

$$\Phi_{2,2}^\tau(u_\tau^{n-1}, u_\tau^{n+a-1}; u_\tau^n) \leq \Phi_{2,2}^\tau(u_\tau^{n-1}, u_\tau^{n+a-1}; u_\tau^{n+a-1}), \tag{120a}$$

$$\text{and} \quad \Phi_{2,2}^\tau(u_\tau^{n-1}; u_\tau^{n+a-1}; u_\tau^n) \leq \Phi_{2,2}^\tau(u_\tau^{n-1}, u_\tau^{n+a-1}; u_\tau^{n-1}), \tag{120b}$$

we have the following propositions:

Proposition 5.5. *From equation (120a), the sequences of discrete solutions* $(u_\tau^n)_{n\in\mathbb{N}}$*,* $(u_\tau^{n+a-1})_{n\in\mathbb{N}}$ *and* $(u_\tau^{n-1})_{n\in\mathbb{N}}$ *satisfy*

$$\begin{aligned} &\frac{1}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n+a-1}, u_\tau^n]^2 - \frac{1-2a(1-a)}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2 + \mathcal{E}(u_\tau^n), \\ \leq &-\frac{1-2a(1-a)}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]^2 + \mathcal{E}(u_\tau^{n+a-1}). \end{aligned} \tag{121}$$

Proof. Inequality (120a) is applied with the left hand side coming from substituting $u = u_\tau^n$ in (114b) and the right hand side directly from (116c). □

Example 5.6. *By substituting* $a = 1/4$ *into (121), we have that*

$$\frac{4}{\tau}\mathcal{W}_2[u_\tau^{n-3/4}, u_\tau^n]^2 - \frac{5}{2\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2 + \mathcal{E}(u_\tau^n) \leq -\frac{5}{2\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n-3/4}]^2 + \mathcal{E}(u_\tau^{n-3/4}). \tag{122}$$

Proposition 5.7. *From equation (120b), the sequences of discrete solutions* $(u_\tau^n)_{n\in\mathbb{N}}$*,* $(u_\tau^{n+a-1})_{n\in\mathbb{N}}$ *and* $(u_\tau^{n-1})_{n\in\mathbb{N}}$ *satisfy*

$$\begin{aligned} &\frac{1}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n+a-1}, u_\tau^n]^2 - \frac{1-2a(1-a)}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2 + \mathcal{E}(u_\tau^n) \\ \leq \quad &\frac{1}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]^2 + \mathcal{E}(u_\tau^{n-1}). \end{aligned} \tag{123}$$

Proof. Inequality (120b) is applied with the left hand side coming from substituting $u = u_\tau^n$ in (114b) and the right hand side directly from (116b). □

Example 5.8. *We have that the inequality for (121) when $a = 1/4$ is*

$$\frac{4}{\tau}\mathcal{W}_2[u_\tau^{n-3/4}, u_\tau^n]^2 - \frac{5}{2\tau}\mathcal{W}_2[u^{n-1}, u^n]^2 + \mathcal{E}(u_\tau^n) \leq \frac{4}{\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n-3/4}]^2 + \mathcal{E}(u_\tau^{n-1}). \qquad (124)$$

That is, $u_\tau^{n-3/4}$ minimises the potential $\Phi_{2,1}(u_\tau^{n-1}; u)$ for both (122) and (124).

Now we remark on how to simplify our later estimates and for our discrete evolution variational inequality (EVI) (we will introduce this in the subsequent section). These will influence whether our prefactors from our final estimate of $\mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - C_1\mathcal{W}_2[u_\tau^0, v_\eta^0]^2$ are positive or negative i.e. can we bound some of our terms of these above by zero (these terms can be omitted as a result, simplifying our target (115))?

For $a \in \left(0, \frac{1}{2}\right)$, we have that $2a(1-2a) \geq 0$. Otherwise, for $a \in \left(\frac{1}{2}, 1\right)$, we have that $2a(1-2a) \leq 0$. By the simple observation that $2a(1-a) \in (0,1)$ for all $a \in (0,1)$, we have that $2a(1-a) - 1 \leq 0$ for all $a \in (0,1)$.

Combining these with the fact that squared metrics are non-negative, we can simplify the inequalities (121) as follows, dependent on the two intervals $a \in (0, 1/2)$ and $a \in (1/2, 1)$:

- For $a \in \left(\frac{1}{2}, 1\right)$:

$$\frac{1}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n+a-1}, u_\tau^n]^2 + \mathcal{E}(u_\tau^n) \leq -\frac{1-2a(1-a)}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]^2 + \mathcal{E}(u_\tau^{n+a-1}).$$

- For $a \in \left(0, \frac{1}{2}\right)$:

$$-\frac{1-2a(1-a)}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2 + \mathcal{E}(u_\tau^n) \leq \mathcal{E}(u_\tau^{n+a-1}). \qquad (125)$$

Example 5.9. *Furthermore, from (125), we have for $a = 1/4$, which is also the simplified version of (124):*

$$-\frac{5}{2\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2 + \mathcal{E}(u_\tau^n) \leq \mathcal{E}(u_\tau^{n-3/4}).$$

Estimates are given for the energy functionals at the time-discrete level at various time points. These are our key ingredients for implementing the convergence proof in Section 6. But before we do, we implement another variational form of the DIRK2 scheme, in line with the BDF2 scheme [35]. Therefore, in the next section, we explain several assumptions, which help construct a discrete form of the evolution variational inequality, an equivalent representation of (17) which considers semi-convex energies, for both stages.

We formally introduce and adapt the variational formulation of the BDF2 scheme, from Matthes and Plazotta [35], to the DIRK2 scheme.

5.7 Main Assumptions for the Evolution Variation Inequality (EVI) - Semi Convexity

The general assumptions are mentioned, first with lower-semi-continuity and coercivity. Furthermore, the semi-convexity assumptions for each stage. We note that the range of λ, that gives us the stronger convexity condition, is restricted to non-positives, for ensuring well-posedness of gradient flow problem, as we discussed in Section 3.

For the PDEs we are investigating, their corresponding energy functionals $\mathcal{E}(\cdot)$ are non-negative, since the energies are integrals of a squared modulus function, which are non-negative. From this, we have constructed a range of estimates, using the assumption $\mathcal{E}(u_\tau^0) \le K_1 < \infty$. These estimates can be generally applied across a wide range of PDEs including the ones that we shall consider in the following.

The minimising movement schemes and the resultant estimates from the last subsections are designed mainly for assistance with the final estimates. The estimates we can derive come from manipulating the functionals Φ^τ and noting that for BDF1 for example, evaluating Φ_1^τ at the minimiser u_τ^n is going to be smaller than the result for evaluating at the intermediate solution $u = u_\tau^{n-1}$.

Now we consider a variational form of these schemes, worked on by Matthes & Plazotta, [35] for BDF2. We know how to explore the dynamics of solutions via the minimising movement scheme, but the idea of our new construction is to verify how light assumptions can help us verify convergence of discrete solutions, working round having to verify strict monotonicity, which we cannot theoretically prove.

Furthermore, the higher order minimising movement schemes (unlike BDF1) do not guarantee that the energy $\mathcal{E}(\cdot)$ monotonically decreases in time and hence obey the structure of a gradient flow. The variational form of the BDF2 scheme successfully shown the numerical convergence of discrete solutions, without proving energy monotonicity, to what we describe as the limit curve u_* in an alternative approach, also satisfying results from gradient flow properties like uniqueness from convexity. There is a basic variational form, called the energy dissipation equation which we introduced in Section 3.4, but also and with respect to convexity, we consider the evolution variational inequality. As we said in the introduction, one of our aims is to verify the limit curve u_*, from our gradient flow, is admissible to the inequality.

To verify that a scheme is well-posed, we must show that a unique minimiser exists, via the two following standard assumptions on the energy functional for Wasserstein gradient flows (see [35, p. 6] or [2, Lem. 2.4.8], by Ambrosio et al.):

Assumption 5.10. *The following assumptions are as follows:*

(i) ***Semi-continuity of*** $\mathcal{E}(\cdot)$: *The energy functional $\mathcal{E}(\cdot)$ is sequentially lower semi-continuous on $(\mathcal{P}_M(\Omega), \mathcal{W}_2)$:*

$$u_k \to u \qquad \Longrightarrow \qquad \mathcal{E}(u) \le \liminf_{k\to\infty} \mathcal{E}(u_k). \tag{126}$$

(ii) ***Coercivity of*** $\mathcal{E}(\cdot)$: *There exist $\tau_* > 0$ and $u_* \in \mathcal{P}_M(\Omega)$ such that*

$$c_* := \inf_{u\in\mathcal{P}_M(\Omega)} \frac{1}{2\tau_*}\mathcal{W}_2[u_*, u] + \mathcal{E}(u) > -\infty. \tag{127}$$

Remark 5.11. With respect to the PDEs considered in this thesis, the energy functionals are non-negative which implies interestingly, that both assumptions are satisfied.

For this next part, we seek and construct an alternative form of the minimising movement scheme, by assuming semi-convexity of the energy functional, $\mathcal{E}(\cdot)$ and hence the Yosida-regularised functionals $\Phi^{\tau}_{2,1}(u^{n-1}_{\tau}; u)$ and $\Phi^{\tau}_{2,2}(u^{n-1}_{\tau}, u^{n+a-1}_{\tau}; u)$ before taking on the stronger condition of strict convexity, that guarantees uniqueness of minimisers, by setting conditions on the modulus of convexity, λ. The semi-convexity of $\mathcal{E}(\cdot)$ is a necessary condition for a well defined sequence of discrete solutions $(u^n_{\tau})_{n\in\mathbb{N}}$.

To do this, we consider the assumptions for $\mathcal{E} : \mathcal{P}_M(\Omega) \to \mathbb{R} \cup \{\infty\}$ in line with [35], including semi-convexity. Alongside the lower semi-continuity and coercivity properties, we also have that $\mathcal{E}(\cdot)$ is semi-convex of modulus $\lambda \in \mathbb{R}$. Considering semi-convexity, it is important to note this controls the downward slope (dissipation): should this occur of our function and will this generate multiple (not unique) minimisers? We will explain this shortly, but by setting the modulus to satisfy that $\lambda \leq 0$ and $(-\lambda)\tau \leq \frac{2(a-1)}{(1-2a)\tau}$, this yields the stronger property of strict convexity, which provides a unique minimiser.

Firstly, by considering the semi-convexity of $\mathcal{E} : \mathcal{P}_M(\Omega) \to \mathbb{R} \cup \{\infty\}$ similar as in [35, p. 6] where in particular, the assumption holds for the L^2-Wasserstein metric [35, Thm. 7], we have that:

- **Stage One**: From [2, Prop. 9.3.12], we have that $\Phi_{2,1}(u^{n-1}_{\tau}, u)$ from (63) is semi-convex of modulus $\dfrac{1}{a\tau} + \lambda$.

  Hence, for all $u^{n-1}_{\tau}, \gamma_0, \gamma_1 \in \mathcal{P}_M(\Omega)$ and every $\tau \in [0, \tau_*)$, where τ_* is chosen as the maximum time step size, there exists a continuous curve $\gamma_s : [0,1] \to \mathcal{P}_M(\Omega)$ joining γ_0, γ_1 along which $\Phi_{2,1}(u^{n-1}_{\tau}; u)$ satisfies

  $$\Phi^{\tau}_{2,1}(u^{n-1}_{\tau}; \gamma_s) \leq (1-s)\Phi^{\tau}_{2,1}(u^{n-1}_{\tau}; \gamma_0) + s\Phi^{\tau}_{2,1}(u^{n-1}_{\tau}; \gamma_1) - \frac{1}{2}\left(\frac{1}{a\tau} + \lambda\right) s(1-s)\mathcal{W}_2[\gamma_1, \gamma_0]^2, \quad (128)$$

  where $s \in (0,1)$ and λ is the modulus of convexity.

  In order for strict convexity to apply, we require that from the last term of (128):

  $$\lambda > -\frac{1}{a\tau}, \quad \text{and} \quad a \neq 0. \quad (129)$$

  Example 5.12. *For $a = 1/4$, we have that (67) is semi-convex of modulus $\dfrac{4}{\tau} + \lambda$ and the semi-convexity condition being*

  $$\Phi^{\tau}_{2,1}(u^{n-1}_{\tau}; \gamma_s) \leq (1-s)\Phi^{\tau}_{2,1}(u^{n-1}_{\tau}; \gamma_0) + s\Phi^{\tau}_{2,1}(u^{n-1}_{\tau}; \gamma_1) - \frac{1}{2}\left(\frac{4}{\tau} + \lambda\right) s(1-s)\mathcal{W}_2[\gamma_1, \gamma_0]^2,$$

  with the strict convexity condition being, which is obtained also by substituting $a = 1/4$ into (129):

  $$\lambda > -\frac{4}{\tau}. \quad (130)$$

Remark 5.13. For $a \in (0,1)$, this gives us $-a\tau \in (-\tau, 0) \Rightarrow -\frac{1}{a\tau} < -\frac{1}{\tau}$. Hence we can assume, without loss of generality, for any $a \in (0,1)$:

$$\lambda > -\frac{1}{\tau}.$$

- **Stage Two**: Again, from [2, Prop. 9.3.12], we have that $\Phi_{2,1}(u_\tau^{n-1}, u)$ from (64) is semi-convex of modulus $\frac{2(1-a)}{(1-2a)\tau} + \lambda$.

  Hence, for all $u_\tau^{n+a-1}, u_\tau^{n-1}, \gamma_0, \gamma_1 \in \mathcal{P}_M(\Omega)$ and every $\tau \in [0, \tau_*)$ there exists a continuous curve $\gamma_s; [0,1] \to \mathcal{P}_M(\Omega)$ joining γ_0, γ_1 along which $\Phi_{2,2}$ satisfies

  $$\begin{aligned} \Phi_{2,2}^\tau(u_\tau^{n-1}, u^{n+a-1}; \gamma_s) \leq \quad & (1-s)\Phi_{2,2}^\tau(u_\tau^{n-1}, u_\tau^{n+a-1}; \gamma_0) + s\Phi_{2,2}^\tau(u_\tau^{n-1}, u_\tau^{n+a-1}; \gamma_1) \\ & -\frac{1}{2}\left(\frac{2(1-a)}{(1-2a)\tau} + \lambda\right) s(1-s)\mathcal{W}_2[\gamma_1, \gamma_0]^2, \end{aligned} \tag{131}$$

  where s and λ are defined as in stage one (see previous page).

  In order for strict convexity to apply, we require that from the last term of (131):

  $$\lambda > \frac{2(a-1)}{(1-2a)\tau} \quad \text{and} \quad a \neq \frac{1}{2}. \tag{132}$$

  Example 5.14. *For $a = 1/4$, we have that (67) is semi-convex of modulus $\frac{3}{\tau} + \lambda$ and the semi-convexity condition as*

  $$\begin{aligned} \Phi_{2,2}^\tau(u_\tau^{n-1}, u_\tau^{n-3/4}; \gamma_s) \leq & (1-s)\Phi_{2,2}(u_\tau^{n-1}, u_\tau^{n-3/4}; \gamma_0) + s\Phi_{2,2}^\tau(u_\tau^{n-1}, u_\tau^{n-3/4}; \gamma_1) \\ & -\frac{1}{2}\left(\frac{3}{\tau} + \lambda\right) s(1-s)\mathcal{W}_2[\gamma_1, \gamma_0]^2, \end{aligned} \tag{133}$$

  with the strict convexity condition being

  $$\lambda > -\frac{3}{\tau}. \tag{134}$$

Remark 5.15. For the next section, where we wish the strict convexity conditions to apply for both stages, we analyse as follows for $a \in (0, 1/2)$:

(i) **Stage One**: $a \in (0, 1/2)$ gives that

$$-\frac{1}{a} \in (-\infty, -2).$$

(ii) **Stage Two**: $2(a-1) \in 2(-1, -1/2)$ and $1 - 2a \in (0,1) \Rightarrow \frac{1}{1-2a} \in (1, \infty)$ gives that

$$\frac{2(a-1)}{1-2a} \in (-\infty, -1).$$

Hence for the condition to apply for both stages, we set for the modulus of convexity:

$$\lambda > \frac{2(a-1)}{(1-2a)\tau}. \tag{135}$$

Example 5.16. *When $a = 1/4$, the strict convexity condition for both stages is*

$$\lambda > -\frac{3}{\tau}.$$

Remark 5.17. Given that both inequalities for λ, (129) and (132) (plus (130) and (134) where $a = 1/4$, respectively), apply, we have that both $\Phi^{\tau}_{2,1}(u^{n-1}_{\tau}; u), \Phi^{\tau}_{2,2}(u^{n-1}_{\tau}, u^{n+a-1}_{\tau}; u)$ are strictly convex, in order to help generate unique minimisers u^{n+a-1}_{τ} and u^{n}_{τ} for each stage, respectively.

However, for the similar construction of [35, eqn. 3.4], in order for a minimiser of each stage to exist, the convexity conditions (129) and (132) must be strengthened in order for a Cauchy sequence of minimisers to exist, that is a distance between two potentially different minimising sequences $\mathcal{W}_2[u_k, u_l]^2$ has a finite upper bound.

Therefore we will assume that

$$\lambda \leq 0 \quad \text{and} \quad (-\lambda)\tau_* < \frac{2(1-a)}{1-2a}. \tag{136}$$

The first equation comes from the arguments by Propositions 3.6 and 3.7, i.e. λ needs to be negative, to guarantee a well-posed problem. The second equation (136) comes from (135), assuming without loss of generality, and since $\tau_* > \tau$ by definition.

5.8 Existence of a Minimiser

We adapt the Matthes, Plazotta proof for BDF2, that ensures the unique existence of a minimiser for both stages.

With our minimising movement schemes introduced in Section 4, we are in position to begin discussing the variational form of the DIRK2 scheme, but beforehand, it is important to conclude whether a unique minimiser exists:

There are assumptions on $\mathcal{E}(\cdot)$, including semi-continuity (126) and coercivity (127). Using these conditions and by applying Young's inequality, we can show that the scheme provides unique minimisers for both $\Phi^{\tau}_{2,1}(u^{n-1}_{\tau}; v)$ and $\Phi^{\tau}_{2,2}(u^{n-1}_{\tau}, u^{n+a-1}_{\tau}; u)$.

Theorem 5.18. *For all $\tau \in (0, \tau_*)$, $a \in (0, 1/2)$ and $u^{n-1}_{\tau}, u^{n+a-1}_{\tau} \in \mathcal{P}_M(\Omega)$, there exist unique minimisers as follows:*

- *Stage One: There exists an unique minimiser v_* of $v \to \Phi^{\tau}_{2,1}(u^{n-1}_{\tau}; v)$.*
- *Stage Two: There exists an unique minimiser u_* of $u \to \Phi^{\tau}_{2,2}(u^{n-1}_{\tau}, u^{n+a-1}_{\tau}; u)$.*

Proof. For stage one, the proof of this theorem is adapted from [35, Thm. 8] (this is the same result for the BDF2 scheme):

Retrieving the functional $\Phi^{\tau}_{2,1}(u^{n-1}_{\tau};u)$ from (63), we wish to show it has a finite lower bound. By using the simple Young's inequality:

$$\mathcal{W}_2[u_*,w]^2 \leq 2\mathcal{W}_2[u^{n-1}_{\tau},u_*]^2 + 2\mathcal{W}_2[u^{n-1}_{\tau},w]^2,$$

before substituting into (63) gives us as a result of the coercivity assumption (127):

$$\begin{aligned}\Phi^{\tau}_{2,1}(u^{n-1}_{\tau};u) =& \frac{1}{2a\tau}\mathcal{W}_2[u^{n-1}_{\tau},u]^2 + \mathcal{E}(u) \geq \frac{1}{4a\tau}\mathcal{W}_2[u_*,u]^2 - \frac{1}{2a\tau}\mathcal{W}_2[u^{n-1}_{\tau},u_*]^2 + \mathcal{E}(u) \\ >& \frac{1}{4a\tau_*}\mathcal{W}_2[u_*,u]^2 + \mathcal{E}(u) - \frac{1}{2a\tau}\mathcal{W}_2[u^{n-1}_{\tau},u_*]^2 > c_* - \frac{1}{2a\tau}\mathcal{W}_2[u^{n-1}_{\tau},u_*]^2.\end{aligned}$$

Since the lower bound is independent of the unknown arbitrary curve $u \in \mathcal{P}_M(\Omega)$, a finite lower bound of $\Phi^{\tau}_{2,1}(\cdot)$ exists i.e.

$$\underline{\phi} = \inf_{u\in\mathcal{P}_M(\Omega)} \Phi^{\tau}_{2,1}(u^{n-1}_{\tau};u) > -\infty.$$

To show that a minimising sequence $(u_k)_{k\in\mathbb{N}}$ is a Cauchy sequence, we select two end points of the curve, which are $\gamma_0 = u_k$ and $\gamma_1 = u_l$, with the curves midpoint being $u_{k,l} = \gamma_{1/2}$.

Substituting this into (128) gives

$$\Phi^{\tau}_{2,1}(u^{n-1}_{\tau};u_{k,l}) \leq \frac{1}{2}\Phi^{\tau}_{2,1}(u^{n-1}_{\tau};u_k) + \frac{1}{2}\Phi^{\tau}_{2,1}(u^{n-1}_{\tau};u_l) - \frac{1}{8}\left(\frac{1}{a\tau}+\lambda\right)\mathcal{W}_2[u_k,u_l]^2.$$

From (136), we have that $1 + a\lambda\tau \geq 1 - 2a$ and not containing zero for $a \in (0, 1/2)$, hence we have a finite upper bound for our Wasserstein distance $\mathcal{W}_2[u_k,u_l]^2$, that is

$$\begin{aligned}\mathcal{W}_2[u_k,u_l]^2 &\leq \frac{4a\tau}{1+a\lambda\tau}\left(\Phi^{\tau}_{2,1}(u^{n-1}_{\tau};u_k) + \Phi^{\tau}_{2,1}(u^{n-1}_{\tau};u_l) - 2\Phi^{\tau}_{2,1}(u^{n-1}_{\tau};u_{k,l})\right) \qquad (137)\\ &\leq \frac{4a\tau}{1+a\lambda\tau}\left(\Phi^{\tau}_{2,1}(u^{n-1}_{\tau};u_k) + \Phi^{\tau}_{2,1}(u^{n-1}_{\tau};u_l) - 2\underline{\phi}\right).\end{aligned}$$

By assuming there are two minimising sequences $(u_k)_{k\in\mathbb{N}}$ and $(u_l)_{l\in\mathbb{N}}$ then as the sequences progress such that they minimise $\Phi_{2,1}(u^{n-1}_{\tau};u)$, the right hand side of (137) progresses to

$$\frac{4a\tau}{1+a\lambda\tau}\left(\underline{\phi}+\underline{\phi}-2\underline{\phi}\right) = 0,$$

which results in the metric $\mathcal{W}_2[u_k,u_l]^2$ becoming sufficiently small, hence the Cauchy property is satisfied.

Since $(\mathcal{P}_M(\Omega),\mathcal{W}_2)$ is complete, then every Cauchy sequence $(u_k)_{k\in\mathbb{N}}$ converges to a limit point $u_* \in \mathcal{P}_M(\Omega)$.

Finally, by the semi-continuity assumption (126) and a distance between two points a continuous function, we have that $\Phi_{2,2}(\cdot)$ is lower semi-continuous and gives us

$$\underline{\phi} \leq \Phi^{\tau}_{2,2}(u^{n-1}_{\tau},u^{n+a-1}_{\tau};u_*) \leq \liminf_{k\to\infty}\Phi^{\tau}_{2,2}(u^{n-1}_{\tau},u^{n+a-1}_{\tau};u_*) = \underline{\phi}.$$

Thus the limit point u_* is a minimiser of $\Phi^{\tau}_{2,2}(u^{n-1}_{\tau},u^{n+a-1}_{\tau};u)$ and uniqueness is satisfied by assumptions (132), as explained in the next subsection. The proof for the first stage is complete. □

Proof. Now for stage two: by retrieving the functional $\Phi^\tau_{2,2}(u^{n-1}_\tau, u^{n+a-1}_\tau; u)$ from (64), we wish to show it also has a finite lower bound i.e. $\Phi_{2,2}(\cdot) > -\infty$. From two different versions of the triangle inequality and binomial theorem, we have that, for $b > \frac{1-2a(1-a)}{2a(1-a)}$:

$$\mathcal{W}_2[u^{n-1}_\tau, u]^2 \leq (1+b)\mathcal{W}_2[u^{n-1}_\tau, u^{n+a-1}_\tau]^2 + \left(1 + \frac{1}{b}\right)\mathcal{W}_2[u^{n+a-1}_\tau, u]^2, \tag{138a}$$

$$\mathcal{W}_2[u_*, u]^2 \leq 2\mathcal{W}_2[u^{n-1}_\tau, u_*]^2 + 2\mathcal{W}_2[u^{n-1}_\tau, u]^2, \tag{138b}$$

and substituting these into (64) gives us, again from assumption (127):

$$\begin{aligned}
& \Phi^\tau_{2,2}(u^{n-1}_\tau, u^{n+a-1}_\tau; u) \\
= \;& \frac{1}{2a(1-2a)\tau}\mathcal{W}_2[u^{n+a-1}_\tau, u]^2 - \frac{1-2a(1-a)}{2a(1-2a)\tau}\mathcal{W}_2[u^{n-1}_\tau, u]^2 + \mathcal{E}(u) \\
> \;& \frac{2a(1-a)(1+b)-1}{2a(1-2a)(1+b)\tau}\mathcal{W}_2[u^{n-1}_\tau, u]^2 - \frac{b}{2a(1-2a)\tau}\mathcal{W}_2[u^{n-1}_\tau, u^{n+a-1}_\tau]^2 + \mathcal{E}(u) \\
\geq \;& \frac{2a(1-a)(1+b)-1}{4a(1-2a)(1+b)\tau}\mathcal{W}_2[u_*, u]^2 + \mathcal{E}(u) - \frac{b}{2a(1-2a)\tau}\mathcal{W}_2[u^{n-1}_\tau, u^{n+a-1}_\tau]^2 \\
& + \frac{1+2a(a-1)(1+b)}{2a(1-2a)(1+b)\tau}\mathcal{W}_2[u^{n-1}_\tau, u_*]^2 \\
\geq \;& \frac{2a(1-a)(1+b)-1}{4a(1-2a)(1+b)\tau_*}\mathcal{W}_2[u_*, u]^2 + \mathcal{E}(u) - \frac{b}{2a(1-2a)\tau}\mathcal{W}_2[u^{n-1}_\tau, u^{n+a-1}_\tau]^2 \\
& + \frac{1+2a(a-1)(1+b)}{2a(1-2a)(1-b)\tau}\mathcal{W}_2[u^{n-1}_\tau, u_*]^2 \\
> \;& c_* - \frac{b}{2a(1-2a)\tau}\mathcal{W}_2[u^{n-1}_\tau, u^{n+a-1}_\tau]^2 + \frac{1+2a(a-1)(a+b)}{2a(1-2a)(1+b)\tau}\mathcal{W}_2[u^{n-1}_\tau, u_*]^2.
\end{aligned}$$

Since the lower bound is independent of the unknown reference point $u \in \mathcal{P}_M(\Omega)$, a finite lower bound of $\Phi^\tau_{2,2}(\cdot)$ exists, that is

$$\underline{\psi} = \inf_{u \in \mathcal{P}_M(\Omega)} \Phi^\tau_{2,2}(u^{n-1}_\tau, u^{n+a-1}_\tau; u) > -\infty.$$

To show that a minimising sequence $(u_k)_{k\in\mathbb{N}}$ is a Cauchy sequence, we select two end points of the curve, which are $\gamma_0 = u_k$ and $\gamma_1 = u_l$, with the curves midpoint being $u_{k,l} = \gamma_{1/2}$.

Substituting this into (128) gives

$$\begin{aligned}
\Phi^\tau_{2,2}(u^{n-1}_\tau, u^{n+a-1}_\tau; u_{k,l}) \leq \;& \frac{1}{2}\Phi^\tau_{2,2}(u^{n-1}_\tau, u^{n+a-1}_\tau; u_k) + \frac{1}{2}\Phi^\tau_{2,2}(u^{n-1}_\tau, u^{n+a-1}_\tau; u_l) \\
& - \frac{1}{8}\left(\frac{2(1-a)}{(1-2a)\tau} + \lambda\right)\mathcal{W}_2[u_k, u_l]^2.
\end{aligned}$$

From (136), we have that $2(1-a) + (1-2a)\lambda\tau > 0$, which does not contain zero for $a \in (0, 1/2)$, hence we have a finite upper bound for our Wasserstein distance $\mathcal{W}_2[u_k, u_l]^2$, that is

$$\begin{aligned}
\mathcal{W}_2[u_k, u_l]^2 \leq \;& \frac{4(1-2a)\tau}{2(1-a)+(1-2a)\lambda\tau}\Big(\Phi^\tau_{2,2}(u^{n-1}_\tau, u^{n+a-1}_\tau; u_k) + \Phi^\tau_{2,2}(u^{n-1}_\tau, u^{n+a-1}_\tau; u_l) \\
& - 2\Phi^\tau_{2,2}(u^{n+a-1}_\tau, u^{n-1}_\tau; u_{k,l})\Big)
\end{aligned} \tag{139}$$

$$\leq \quad \frac{4(1-2a)\tau}{2(1-a)+(1-2a)\lambda\tau}\Big(\Phi^{\tau}_{2,2}(u^{n-1}_{\tau},u^{n+a-1}_{\tau};u_k)+\Phi^{\tau}_{2,2}(u^{n-1}_{\tau},u^{n+a-1}_{\tau};u_l)-2\underline{\phi}\Big).$$

By assuming there are two minimising sequences $(u_k)_{k\in\mathbb{N}}$ and $(u_l)_{l\in\mathbb{N}}$ then as the sequences progress such that they minimise $\Phi_{2,2}(u^{n-1}_{\tau},u^{n+a-1}_{\tau};u)$, the right hand side of (139) progresses to

$$\frac{4(1-2a)\tau}{2(1-a)+(1-2a)\lambda\tau}\Big(\underline{\phi}+\underline{\phi}-2\underline{\phi}\Big)=0,$$

which results in the metric $\mathcal{W}_2[u_k,u_l]^2$ becoming sufficiently small, hence the Cauchy property is satisfied.

Since $(\mathcal{P}_M(\Omega),\mathcal{W}_2)$ is complete, then every Cauchy sequence $(u_k)_{k\in\mathbb{N}}$ converges to a limit point $u_*\in\mathcal{P}_M(\Omega)$.

Finally, by the semi-continuity assumption (126) and a distance between two points a continuous function, we have that $\Phi_{2,2}(\cdot)$ is lower semi-continuous and gives us

$$\underline{\phi}\leq\Phi^{\tau}_{2,2}(u^{n-1}_{\tau},u^{n+a-1}_{\tau};u_*)\leq\liminf_{k\to\infty}\Phi^{\tau}_{2,2}(u^{n-1}_{\tau},u^{n+a-1}_{\tau};u_k)=\underline{\phi}.$$

Thus the limit point u_* is a minimiser of $\Phi^{\tau}_{2,2}(u^{n-1}_{\tau},u^{n+a-1}_{\tau};u)$ and uniqueness is satisfied by (136). The proof for the second stage is complete.

□

5.9 Adapting the Discrete EVI

We adapt the discrete form of the differential evolution variational inequality (EVI) from Matthes, Plazotta [35]. This combines the semi-convexity estimate from Section 5.7 and the estimates from Section 5.6 (u^{n+a-1}_{τ} is a minimiser from the stage one scheme).

We have lined out our semi-convexity conditions for our Yosida-regularised functionals (128), (131) and also our strict convexity conditions, obtained by conditional modulus of convexities (129), (132).

Now we combine these with the minimisers for both stages, to generate a novel inequality in terms of metric and energy functional terms only, defined as the discrete evolution variational inequality (EVI) (see [35, p.14] for a similar construction for BDF2).

5.9.1 Stage One DIRK2 Scheme

By starting with stage one, it is necessary to select two appropriate end points of the curve γ_s, which we say are $\gamma_0,\gamma_1\in\mathcal{P}_M(\Omega)$ in order to generate an inequality based on the fact that u^{n+a-1}_{τ} is the minimiser of $\Phi^{\tau}_{2,1}(u^{n-1}_{\tau};u)$. This will be seen here as the main ingredient for our discrete EVI construction:

Lemma 5.19 (See Lemma 2 of [35] for original idea)**.** *The discrete solution $(u^{n+a-1})_{n\in\mathbb{N}}$ satisfies*

$$\left(\frac{1}{2a\tau}+\frac{\lambda}{2}\right)\mathcal{W}_2[u^{n+a-1}_{\tau},u]^2-\frac{1}{2a\tau}\mathcal{W}_2[u^{n-1}_{\tau},u]^2\leq\mathcal{E}(u)-\mathcal{E}(u^{n+a-1}_{\tau})-\frac{1}{2a\tau}\mathcal{W}_2[u^{n-1}_{\tau},u^{n+a-1}_{\tau}]^2. \quad (140)$$

Proof. The proof is similar as the proof in [35, Lem. 2]. From [35, Ass. E3], for semi-convexity, there exists $\lambda \in \mathbb{R}$ such that for all $u_\tau^{n+a-1}, u_\tau^{n-1}, \gamma_0, \gamma_1 \in \mathcal{D}(\mathcal{E})$ and $\tau \in [0, \tau_*)$, there exists a continuous curve which satisfies

$$\Phi_{2,1}^\tau(u_\tau^{n-1}; \gamma_s) \leq (1-s)\Phi_{2,1}^\tau(u_\tau^{n-1}; \gamma_0) + s\Phi_{2,1}^\tau(u_\tau^{n-1}; \gamma_1) - \frac{1}{2}\left(\frac{1}{a\tau} + \lambda\right) s(1-s)\mathcal{W}_2[\gamma_1, \gamma_0]^2. \quad (141)$$

Let the two end points be $\gamma_0 := u_\tau^{n+a-1}, \gamma_1 := u$ with $(\gamma_s)_{s\in[0,1]}$ the corresponding connecting curve that implies semi-convexity. Then combining (141) with u_τ^{n+a-1} minimising $\Phi_{2,1}^\tau(u_\tau^{n-1}; u)$ for all $s \in (0,1)$ gives us

$$\begin{aligned}
0 \leq\ & \Phi_{2,1}^\tau(u_\tau^{n-1}; \gamma_s) - \Phi_{2,1}^\tau(u_\tau^{n-1}; u_\tau^{n+a-1}) \\
\leq\ & (1-s)\Phi_{2,1}^\tau(u_\tau^{n-1}; u_\tau^{n+a-1}) + s\Phi_{2,1}^\tau(u_\tau^{n-1}; u) - \frac{1}{2}\left(\frac{1}{a\tau} + \lambda\right) s(1-s)\mathcal{W}_2[u_\tau^{n+a-1}, u]^2 \\
& - \Phi_{2,1}^\tau(u_\tau^{n-1}; u_\tau^{n+a-1}) \\
=\ & s\Phi_{2,1}^\tau(u_\tau^{n-1}; u) - s\Phi_{2,1}^\tau(u_\tau^{n-1}; u_\tau^{n+a-1}) - \frac{1}{2}\left(\frac{1}{a\tau} + \lambda\right) s(1-s)\mathcal{W}_2[u_\tau^{n+a-1}, u]^2.
\end{aligned}$$

Dividing both sides of (142) by $s \in (0,1)$ and letting $s \to 0$ gives

$$\begin{aligned}
0 \leq\ & \Phi_{2,1}^\tau(u_\tau^{n-1}; u) - \Phi_{2,1}^\tau(u_\tau^{n-1}; u_\tau^{n+a-1}) - \frac{1}{2}\left(\frac{1}{a\tau} + \lambda\right)\mathcal{W}_2[u^{n+a-1}, u]^2 \\
=\ & \frac{1}{2a\tau}\mathcal{W}_2[u^{n-1}, u]^2 + \mathcal{E}(u) - \frac{1}{2a\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]^2 \\
& - \mathcal{E}(u_\tau^{n+a-1}) - \frac{1}{2}\left(\frac{1}{a\tau} + \lambda\right)\mathcal{W}_2[u^{n+a-1}, u]^2.
\end{aligned} \quad (142)$$

Rearrangement of terms in (142) gives

$$\left(\frac{1}{2a\tau} + \frac{\lambda}{2}\right)\mathcal{W}_2[u^{n+a-1}, u]^2 - \frac{1}{2a\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 \leq \mathcal{E}(u) - \mathcal{E}(u_\tau^{n+a-1}) - \frac{1}{2a\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]^2. \quad (143)$$

before multiplying (143) by $\dfrac{2a\tau}{1+a\lambda\tau}$ completes the proof:

$$\begin{aligned}
& \mathcal{W}_2[u_\tau^{n+a-1}, u]^2 - \frac{1}{1+a\lambda\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 \\
\leq\ & \frac{2a\tau}{1+a\lambda\tau}\left(\mathcal{E}(u) - \mathcal{E}(u_\tau^{n+a-1}) - \frac{1}{2a\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]^2\right).
\end{aligned} \quad (144)$$

□

Example 5.20. *By substituting $a = 1/4$ into (143), we have that*

$$\mathcal{W}_2[u_\tau^{n-3/4}, u]^2 - \frac{4}{4+\lambda\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 \leq \frac{2\tau}{4+\lambda\tau}\left(\mathcal{E}(u) - \mathcal{E}(u_\tau^{n-3/4}) - \frac{2}{\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n-3/4}]^2\right).$$

Remark 5.21. Given our construction for general intermediate time steps, it is crucial to comment that the result (144) is valid *only* if $1 + a\lambda\tau > 0$ which is true always, as a result of the condition (129), that is

$$1 + a\lambda\tau > 1 - a\tau\left(\frac{1}{a\tau}\right) = 1 - 1 = 0.$$

5.9.2 Stage Two DIRK2 Scheme

The same process is applied from stage one, although different end points γ_0 and γ_1 are considered:

Lemma 5.22. *The discrete solution $(u_\tau^n)_{n\in\mathbb{N}}$ satisfies*

$$\begin{aligned}
&\left(\frac{1-a}{(1-2a)\tau}+\frac{\lambda}{2}\right)\mathcal{W}_2[u_\tau^n,u]^2-\frac{1}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n+a-1},u]^2+\frac{1-2a(1-a)}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2\\
\leq\;&\mathcal{E}(u)-\mathcal{E}(u_\tau^n)-\frac{1}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n+a-1},u_\tau^n]^2+\frac{1-2a(1-a)}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1},u_\tau^n]^2.
\end{aligned}\tag{145}$$

Proof. The proof again is similar to the proof in [35, Lem. 2] and the previous lemma (140): Similarly for $\Phi_{2,2}^\tau(u_\tau^{n-1},u_\tau^{n+a-1};u)$, we have (let $\gamma_0=u_\tau^n$)

$$\begin{aligned}
0&\leq\Phi_{2,2}^\tau(u_\tau^{n-1},u_\tau^{n+a-1};\gamma_s)-\Phi_{2,2}^\tau(u_\tau^{n-1},u_\tau^{n+a-1};u_\tau^n) &(146)\\
&\leq s\Phi_{2,2}^\tau(u_\tau^{n-1},u_\tau^{n+a-1};u)-s\Phi_{2,2}^\tau(u_\tau^{n-1},u_\tau^{n+a-1};u_\tau^n)-\frac{1}{2}\left(\frac{2(1-a)}{(1-2a)\tau}+\lambda\right)s(1-s)\mathcal{W}_2[u_\tau^n,u]^2.
\end{aligned}$$

Dividing through (146) by s and letting $s\to0$ gives

$$\begin{aligned}
0\leq\;&\Phi_{2,2}^\tau(u_\tau^{n-1},u_\tau^{n+a-1};u)-\Phi_{2,2}^\tau(u_\tau^{n-1},u_\tau^{n+a-1};u_\tau^n)-\frac{1}{2}\left(\frac{2(1-a)}{(1-2a)\tau}+\lambda\right)\mathcal{W}_2[u_\tau^n,u]^2\\
\Leftrightarrow 0\leq\;&\frac{1}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n+a-1},u]^2-\frac{1-2a(1-a)}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2+\mathcal{E}(u)-\mathcal{E}(u_\tau^n)\\
&-\frac{1}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n+a-1},u_\tau^n]^2+\frac{1-2a(1-a)}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1},u_\tau^n]^2\\
&-\frac{1}{2}\left(\frac{2(1-a)}{(1-2a)\tau}+\lambda\right)\mathcal{W}_2[u_\tau^n,u]^2.
\end{aligned}\tag{147}$$

Rearrangement of terms in (147) gives

$$\begin{aligned}
&\left(\frac{1-a}{(1-2a)\tau}+\frac{\lambda}{2}\right)\mathcal{W}_2[u_\tau^n,u]^2-\frac{1}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n+a-1},u]^2+\frac{1-2a(1-a)}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2\\
\leq\;&\mathcal{E}(u)-\mathcal{E}(u_\tau^n)-\frac{1}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n+a-1},u_\tau^n]^2+\frac{1-2a(1-a)}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1},u_\tau^n]^2,
\end{aligned}\tag{148}$$

and the proof is complete. □

Example 5.23. *Via (133), we have that (148) for $a=1/4$ gives us*

$$\begin{aligned}
&\left(\frac{3}{2\tau}+\frac{\lambda}{2}\right)\mathcal{W}_2[u_\tau^n,u]^2-\frac{4}{\tau}\mathcal{W}_2[u_\tau^{n-3/4},u]^2+\frac{5}{2\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2\\
\leq\;&\mathcal{E}(u)-\mathcal{E}(u_\tau^n)-\frac{4}{\tau}\mathcal{W}_2[u_\tau^{n-3/4},u_\tau^n]^2+\frac{5}{2\tau}\mathcal{W}_2[u_\tau^{n-1},u_\tau^n]^2.
\end{aligned}$$

5.10 Classical Estimates for the Energy Functional - From BDF2 [17] to our DIRK2 Case

When we construct our comparison principle estimate, we shall assume that the energy functional $\mathcal{E}(\cdot)$ (this was the case in [35]) is positive, since the integral in (32) is non-negative. Furthermore, we have that the energy, with respect to the initial data, is finite, that is

$$\mathcal{E}(u_\tau^0) \leq K_1. \tag{149}$$

Therefore, we set up some estimates, for which we can eventually sum and/or iterate on n and will become our main ingredients for the comparison principle and numerical convergence proof.

5.10.1 Comparisons for our DIRK2 Method and the BDF2 Method [35]

The BDF2 approach focused on the telescopic summation which was straightforward to analyse, given that this was not a multistage scheme, for deriving the crucial estimates for the energy and metrics. However, adaptations are necessary due to the intermediate time steps and some difficulties in cancelling out metric terms from telescopic summation. Indeed for the energy terms, we apply the relationship of our intermediate solution u_τ^{n+a-1} to both stages, to show finite energy at various time steps by simple induction. Unlike [35], who proved this for general metric space and internal energies, the result of the proof is restricted for the energy functionals being positive, which is for our selected continuity equations and intermediate time step parameter a, where the latter is explained in Lemma 5.25.

5.10.2 Estimates to be derived

Given the initial assumption (149), we derive some estimates on the energy functional terms, in order to apply this to our final estimate $\mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - C_1\mathcal{W}_2[u_\tau^0, v_\eta^0]^2$, to help assist in bounding metric terms above by zero. Alternatively, does our terms tend to zero as the time step size $\tau \to 0$? We considered $a \in (0, 1)$, but since both intervals will constitute different estimation outcomes later on e.g. whether the prefactors of various terms are non-negative or otherwise, and to align our general contribution to the DIRK2 example from [48], we will *only consider* the interval $a \in (0, 1/2)$ from now on:

In fact, a set of iterations could be applied on n, providing us estimates for $\mathcal{E}(u_\tau^{n+a-1})$, $\mathcal{E}(u_\tau^N)$ and $\mathcal{E}(u_\tau^n)$, with finite upper bounds. Hence, from only considering the properties of our minimising movement schemes, we can actually verify also that $\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]$, $\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]$ and $\mathcal{W}_2[u_\tau^{n+a-1}, u_\tau^n]$ converge to zero for vanishing time step τ.

5.11 Finite Metric and Energy Estimates

The following lemma, ensures the finiteness of the Wasserstein metric and hence energy functionals, regardless of the number of time-based intervals. Indeed, the number of time intervals N diverge to infinity as the time step τ dissipates, hence we have to verify that the final sum can be controlled.

5.11.1 Finite Wasserstein Metric

First we begin by showing that the metric $\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]$ is finite from the Discrite EVI constructions in Section 5.9:

Lemma 5.24. *For all $a \in \left(\overline{a}, 1 - \frac{\sqrt{2}}{2}\right)$, where $\overline{a} \approx 0.12$, we guarantee that the Wasserstein distance, summed for each time point $n = 1, \ldots, N$ is bounded:*

$$\frac{1}{\tau}\sum_{n=1}^{N} \mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2 \leq K, \tag{150}$$

where K is some prefactor dependent on K_1, δ, τ (δ is defined later in the proof). Hence this gives us that

$$C\mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - \mathcal{W}_2[u_\tau^0, v_\eta^0]^2 \leq K\tau.$$

Proof. Rearranging equation (145) gives for an arbitrary curve $u = u_\tau^{n+a-1}$:

$$\begin{aligned}\mathcal{E}(u_\tau^n) \leq \quad & \mathcal{E}(u_\tau^{n+a-1}) + \frac{1-2a(1-a)}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2 - \frac{1-2a(1-a)}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]^2 \\ & - \frac{1+2a(1-a)+a(1-2a)\lambda\tau}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n+a-1}, u_\tau^n]^2.\end{aligned} \tag{151}$$

By applying the reverse triangle inequality, then Young's inequality, to $\mathcal{W}_2[u_\tau^{n+a-1}, u_\tau^n]^2$ i.e.

$$\mathcal{W}_2[u_\tau^{n+a-1}, u_\tau^n]^2 \geq \frac{\epsilon}{1+\epsilon}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2 - \epsilon\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]^2,$$

for positive $\epsilon > 0$ (to be selected shortly), we have when substituting into (151):

$$\begin{aligned}\mathcal{E}(u_\tau^n) \leq \quad & \mathcal{E}(u_\tau^{n+a-1}) + \frac{1-2a(1-a)}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2 - \frac{1-2a(1-a)}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]^2 \\ & + \frac{(1+2a(1-a)+a(1-2a)\lambda\tau)\,\epsilon}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]^2 \\ & - \frac{(1+2a(1-a)+a(1-2a)\lambda\tau)\,\epsilon}{2a(1-2a)(1+\epsilon)\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2.\end{aligned}$$

Collecting terms gives

$$\begin{aligned}\mathcal{E}(u_\tau^n) \leq \quad & \mathcal{E}(u_\tau^{n+a-1}) + \frac{(1+2a(1-a))\,\epsilon - 1 + 2a(1-a) + a(1-2a)\epsilon\lambda\tau}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]^2 \\ & - \frac{4a(1-a)\epsilon + 2a(1-a) - 1 + a(1-2a)\epsilon\lambda\tau}{2a(1-2a)(1+\epsilon)\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2.\end{aligned} \tag{152}$$

Rearranging (140) to find an estimate for $\mathcal{E}(u_\tau^{n+a-1})$, which is

$$\mathcal{E}(u_\tau^{n+a-1}) \leq \mathcal{E}(u_\tau^{n-1}) - \frac{2+a\lambda\tau}{2a\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]^2,$$

before substituting into (152), which gives us

$$\begin{aligned}\mathcal{E}(u_\tau^n) \leq \quad & \mathcal{E}(u_\tau^{n-1}) - \frac{2a(1-a)(1+2\epsilon) - 1 + a(1-2a)\epsilon\lambda\tau}{2a(1-2a)(1+\epsilon)\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2 \\ & + \frac{(1+2a(1-a))\,\epsilon - \left(2a^2-6a+3\right) + a(1-2a)(\epsilon-1)\lambda\tau}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]^2. \end{aligned} \tag{153}$$

1. For the prefactor of $\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]^2$: By selecting $\epsilon = \frac{2a^2-6a+3}{1+2a(1-a)}$, this simplifies to

$$\frac{2a^2-4a+1}{1-2a(1-a)}\lambda\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]^2, \tag{154}$$

   which is non-positive for all $a \in \left(0, 1-\frac{\sqrt{2}}{2}\right)$.

2. For all $a \in (\overline{a}, 1-\frac{\sqrt{2}}{2}))$, where $\overline{a} \approx 0.12$ (see Appendix B for detailed workings), we can guarantee non-positivity of the $\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2$ term for our selected parameter ϵ.

Thus, simplifying (153) as a result gives

$$\frac{\delta}{\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2 \leq \mathcal{E}(u_\tau^{n-1}) - \mathcal{E}(u_\tau^n), \tag{155}$$

where δ is given as

$$\frac{2a(1-a)(2a^2-10a+6) - 1 + a(1-2a)(2a^2-6a+3)\lambda\tau}{8a(1-a)(1-2a)}, \tag{156}$$

then summing both sides from $n = 1$ to $n = N$, before applying (149) and dividing through by δ gives us our result and the proof is complete! □

Remark 5.25. Note that this holds provided λ and a are *not* such that

$$\lambda\tau = \frac{1-2a(1-a)(2a^2-10a+6)}{a(1-2a)(2a^2-6a+3)}.$$

5.11.2 Finite Energy Functionals

Furthermore, we can apply the result (150) and assumption (149) to prove finiteness of $\mathcal{E}(u_\tau^{n-1})$, $\mathcal{E}(u_\tau^{n+a-1})$ and $\mathcal{E}(u_\tau^n)$:

Lemma 5.26. *(Finiteness for Our Energy Functionals): $\mathcal{E}(u_\tau^{n-1})$, $\mathcal{E}(u_\tau^{n+a-1})$ and $\mathcal{E}(u_\tau^n)$ are finite e.g. $\mathcal{E}(u_\tau^{n-1}) < K$*

Proof. Taking the base case $n = 1$, then (149) gives $\mathcal{E}(u_\tau^a) < K_1$.

Now taking the basis $n = 1$ for (121) gives the following as a result of (150):

$$\mathcal{E}(u_\tau^1) \leq \frac{1-2a(1-a)}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^0, u_\tau^1]^2 + \mathcal{E}(u_\tau^a) \leq K.$$

Since the result for $\mathcal{E}(u_\tau^1)$ holds and $\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2 \leq K_1\tau$ holds for all $n = 2, \ldots, N$ due to (150) and non-negativity of the left hand side, then inductively, $\mathcal{E}(u_\tau^{n+a-1}) \leq K$ holds for all $n = 1, 2, \ldots, N$. □

5.12 Convergence of Metric Terms

We verify that the Wasserstein metric between the discrete solution at different time points converge to zero as the time step decreases, as you would expect!

We provide proofs for the convergence of certain metric terms and hence densities u_τ^n for decreasing time steps:

Lemma 5.27. *The discrete solutions (densities) $u_\tau^{n-1}, u_\tau^{n+a-1}, u_\tau^n$ converge, as the time step size $\tau \to 0$, to each other or to the limit curve (solution of gradient flow problem) u_*.*

Proof. We break this into three parts:

(i) The result $\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2 \to 0$ as $\tau \to 0$, is verified immediately from (150), for any time level $n = 1, 2, \ldots, N$.

(ii) Next up, we show that $\mathcal{W}_2[u_\tau^{n+a-1}, u_\tau^n]^2 \to 0$ as $\tau \to 0$. As a result of the same inductive argument from Lemma 5.26, and the non-negativity of $\mathcal{E}(\cdot)$, we conclude that from (118),

$$\frac{1}{2a\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]^2 \leq K_1.$$

Thus $\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2 \to 0$ as $\tau \to 0$.

(iii) Finally, we show that $\mathcal{W}_2[u_\tau^{n+a-1}, u_\tau^n]^2 \to 0$. To do this, we rearrange (123), which gives

$$\mathcal{W}_2[u_\tau^{n+a-1}, u_\tau^n]^2 \leq 2a(1-2a)\tau(\mathcal{E}(u_\tau^{n-1}) - \mathcal{E}(u_\tau^n)) + \mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]^2 + (1-2a(1-a))\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2,$$

and as a result of parts (i)-(ii), the right hand side is finite and thus gives the result for (iii). The proof for this part is complete.

∎

In addition, the classical estimates for $\mathcal{E}(\cdot)$ are derived on the energy functional terms, by rearranging the inequalities derived from the minimising movement schemes i.e. u_τ^n minimises $\Phi_{2,2}^\tau(u_\tau^{n-1}, u_\tau^{n+a-1}; u)$, and applying iterations and summations which eventually gives finite upper bounds on $\mathcal{E}(\cdot)$, considering the similar initial assumptions (I0)-(I2) from [35].

Corollary 5.28. *By bounding and summing Proposition 5.5 (121) from $n = 1$ to $n = N$, as well as applying assumption (149), we have that the sequence of discrete solutions $(u_\tau^n)_{n\in\mathbb{N}}$ with $n = 1, 2, \ldots, N$ provide the following estimate for $\mathcal{E}(u_\tau^N)$:*

$$\mathcal{E}(u_\tau^N) \leq K_1 + \frac{1-2a(1-a)}{2a(1-2a)} \sum_{n=1}^{N} \frac{\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2}{\tau}. \tag{157}$$

Example 5.29. *By letting $a = 1/4$ i.e. applying Example 5.6 (122) gives us*

$$\mathcal{E}(u_\tau^N) \leq K_1 + \frac{5}{2} \sum_{n=1}^{N} \frac{\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2}{\tau}. \tag{158}$$

Corollary 5.30. *By applying Lemma 5.27 to Proposition 5.5 (121), we have that the sequence of discrete solutions* $(u_\tau^n)_{n\in\mathbb{N}}$ *with* $n = 1, 2, \dots, N$ *provide the following estimate for* $\mathcal{E}(u_\tau^n)$*:*

$$\mathcal{E}(u_\tau^n) \leq K_1 + \frac{1-2a(1-a)}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2. \tag{159}$$

Example 5.31. *Letting* $a = 1/4$ *i.e. applying Example 5.8 (122) gives*

$$\mathcal{E}(u_\tau^n) \leq K_1 + \frac{5}{2\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2. \tag{160}$$

6 Convergence Analysis

We show the theoretical convergence of the discrete solutions (at time level) to the limit curve, the solution of the gradient flow problem. The procedure constitutes of the comparison principle, which by assumption verifes a unique solution to the gradient flow problem when we consider two different discrete solutions with similar initial conditions.

6.1 Limit Trajectory

The aim here is to show that the discrete solution $(u_\tau^n)_{n\in\mathbb{N}}$ is well defined. Furthermore, we aim to conclude that the piecewise constant interpolation solution $(\overline{u}_{\tau_k})_{k\in\mathbb{N}}$ locally converges uniformly in time to the curve of steepest descent u_* (the solution of the gradient flow of the energy functional $\mathcal{E}(\cdot)$). This approach is in line with the convergence proof for BDF2 in [35]. Our main theorem is the following:

Theorem 6.1 (Based on Theorem 11 of [35])**.** *(Convergence Result): We take a vanishing sequence $(\tau_n)_{n\in\mathbb{N}}$ of time step sizes $\tau_n \in (0, \tau_*)$, which is strictly decreasing and such that the consecutive terms of the sequence are integers (this is an assumption in order to simplify the technicalities of the proof). Also, given the initial datum, which is u_τ^0 (both stages) and u_τ^a (for stage two produced by stage one) that satisfies the initial assumptions (I0)-(I2) (see the comparison principle theorem below for these), plus the additional assumption:*

$$\mathcal{W}_2[u_{\tau_k}^0, u_{\tau_l}^0]^2 \leq K_2\tau_k, \tag{161}$$

we have a well defined discrete solution $(u_\tau^n)_{n\in\mathbb{N}}$. Furthermore, we have the local uniform convergence of piecewise constant interpolations $(\overline{u}_{\tau_k})_{k\in\mathbb{N}}$ with respect to time to an L^2-absolutely continuous function $u_* \in AC^2([0,\infty), \mathcal{P}_M(\Omega))$ (solution of the gradient flow for $\mathcal{E}(\cdot)$ i.e. the limit u_* satisfies the differential evolution variation inequality (EVI)).

Firstly, we show the comparison principle theorem which will prove the well-defined solution part of the convergence result theorem. We shall focus on the comparison principle theorem in Sections 6.2-6.4, before returning to show that the solution of our gradient flow satisfies the EVI, in Section 6.5.

6.2 Comparison Principle Theorem

The aim of the comparison principle theorem is to help show the numerical convergence and uniqueness of the discrete solution $(u_\tau^n)_{n\in\mathbb{N}}$. Before we begin, here is the detailed explanation of the theorem alongside the general assumptions used, based on our initial data and from [35]. Its proof is given during Sections 6.3 (for preparation) and 6.4:

Theorem 6.2 (Based on Theorem 14 of [35]). *(Comparison Principle): We consider the following:*

- *Two time steps $\tau, \eta \in (0, \tau_*)$, that are related by $R = \dfrac{\tau}{\eta} \in \mathbb{N}$.*
- *Two pairs of initial data (u_τ^0, u_τ^a) and (v_η^0, v_η^a).*
- *Intermediate time step $a \in (\bar{a}, (1 - \sqrt{2})/2)$.*
- *Terminal time $T > 0$.*
- *There exists a constant C, expressible in terms of K_1, K_2, λ and a, from initial assumptions (149) and (161).*

Then a piecewise constant interpolation $\overline{u}_\tau$ and $\overline{v}_\eta$ of discrete solutions $(u_\tau^n)_{n\in\mathbb{N}}$ and $(v_\eta^m)_{m\in\mathbb{N}}$ satisfies

$$\mathcal{W}_2[\overline{u}_\tau(t), \overline{v}_\eta(t)]^2 \leq C(\mathcal{W}_2[u_\tau^0, v_\eta^0]^2 + \tau), \tag{162}$$

for all $t \in [0, T]$.

Remark 6.3. The main aim of theorem is to show that $\mathcal{W}_2[u_\tau^N, v_\eta^M]^2$ converges to $\mathcal{W}_2[u_\tau^0, v_\eta^0]^2$ as the time step size $\tau \to 0$, i.e. we will show that

$$\mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - C_1 \mathcal{W}_2[u_\tau^0, v_\eta^0]^2 \leq C_2 \tau, \tag{163}$$

which satisfies (162), when $C = \max\{C_1, C_2\}$, dependent on K_1, K_2, λ and a. This will be shown *via the telescopic sum, in line with [35, Sect. 4.1].*

The main problem is an extension to recent work by D. Matthes and S. Plazotta [35]: they fully shown the well-posedness of the BDF2 type scheme (in a variational form by deriving a time discrete evolution variation inequality (EVI)), formulated as a consequence of the semi-convexity assumption on the energy functional $\mathcal{E}(\cdot)$, guaranteeing a well-posed BDF scheme.

Indeed they shown a comparison principle, constructing an estimate on the difference between the distance of two similar initial data, and the distance of its corresponding solutions at terminal time, i.e. an estimate on $\mathcal{W}_2[u_\tau^n, v_\eta^m]^2 - C_1 \mathcal{W}_2[u_\tau^0, v_\eta^0]^2$). In other words, we adapt this approach, in an innovative manner, for a two stage diagonally implicit Runge-Kutta (DIRK2) scheme.

6.2.1 Our changes to the BDF2 case [35]

By applying the classical estimates for the metrics and energies, we applied ourselves in Sections 5.11-5.12, We initially focus on substituting out terms with intermediate solution content u_τ^{n+a-1}, which will enable us to complete the comparison principle approach in a similar sequence to the BDF2 case (Lemmas 6.4 and 6.6). The iteration part of the inequality from the differential discrete EVI result from [35, p. 16] is *not* repeated here due to major differentiations of the two schemes (the DIRK2 scheme considers only one τ time step, unlike two lots of τ for BDF2). Indeed, the parameter H_τ (is h_τ from [35, p. 17]) is now selected from parameter comparisons (see Lemma 6.7).

6.2.2 Details of the Proof

The aim is to verify numerical convergence of our discrete solutions, by an alternative variational formulation of our BDF/DIRK schemes, which mainly involves implementing some assumptions including convexity/semi-convexity, and estimates in similar line to the BDF2 scheme.

From Section 5.7, we assumed semi-convexity of the energy functional, which is necessary for a well-posed Wasserstein gradient flow problem (17), which implies that we can apply semi-convexity assumptions for the Yosida-regularised functions given by [2]. Furthermore, we claimed and proved that a minimiser of a functional is expected to be unique if the stronger convexity condition applies. Hence by applying stronger constraints on λ, for this to be possible, we combined this along with the defined minimisers for each intermediate/final stage of our schemes to derive discrete forms of the differential evolution variational inequality (EVI).

This brings us to the comparison principle in relation to the DIRK2 scheme, where we seek to combine the EVIs, convexity assumptions and also the (118)-(123) and (157)-(160) inequalities to verify (163) where C_1 is an exponential prefactor, which from (161) clearly implies convergence to a unique curve of steepest descent for decreasing time step τ.

Indeed, from Section 5, we had used the facts and sums from our minimising movement schemes and the differential EVI estimates in order to set up some classical estimates on the energy functional and metric terms, in line with [35, p. 14-15].

6.3 Outline of the Comparison Principle Proof

We give a more detailed outline of the comparison principle proof, and what we wish to analyse:

- Create an estimate on $\mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - c\mathcal{W}_2[u_\tau^{N-1}, v_\eta^M]^2$.
- Apply the telescopic summation with the estimate to the above point, in order to create an estimate to $C_3\mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - \mathcal{W}_2[u_\tau^0, v_\eta^0]^2$.
- Analyse the prefactors of each term on the right hand side to the inequality. Some terms are bounded above by zero (energy functionals are integrals of squared terms implying non-negativity), otherwise with assistance of estimates from Section 5 and another lemma, concerning a velocity, verifies that all remaining terms are proportional to τ.

Lemma 6.4 gives us the estimate for $\mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - \mathcal{W}_2[u_\tau^{N-1}, v_\eta^M]^2$, using the discrete differential EVI.

We now construct the comparison principle proof, considering the telescopic sum based estimate from [35]. Firstly, by applying the estimate $\mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - \mathcal{W}_2[u_\tau^{N-1}, v_\eta^M]^2$, we are then able to use the actual telescopic summation from [35]. From this, the aim is to conclude that (162) holds for some prefactor C (see last subsection for this) which implies numerical convergence.

The telescopic summation initially provides an estimate dependent on multiple metric and energy functional terms. From this, we aim to achieve the outcome stated on the previous paragraph by using the following:

1. Analysing the prefactors (coefficients of the metric and energy functional terms, dependent of time step size τ) of each term. Are they positive (**this may be true for a restricted range of a *only***), which hence we can bound above by classical estimates for $\mathcal{E}(\cdot)$ (see next point)? Or are they negative, which hence we can bound those, interacting with the metric terms, above by zero.

2. The classical estimates for $\mathcal{E}(\cdot)$ have been derived. We derived a number of inequalities on the energy functional terms, by rearranging the inequalities derived from the minimising movement schemes i.e. u_τ^n minimises $\Phi_{2,2}^\tau(u_\tau^{n-1}, u_\tau^{n+a-1}; u)$. Applying iterations and summations which eventually gives finite upper bounds on $\mathcal{E}(\cdot)$, considering the similar initial assumptions (I0)-(I2) from [35, p. 12].

We commence the main body of the comparison principle proof shortly, but first we have some final preparation steps. To start this off, we create a new estimate, which is a combination of the discrete differential EVI estimates for both stages, but where the $\mathcal{W}_2[u_\tau^{N+a-1}, u]$ term is substituted out. This is in order to enable us to apply the telescopic sum on an estimate $C\mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - \mathcal{W}_2[u_\tau^0, v_\eta^0]^2$, by a simple working estimate of $\mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - \mathcal{W}_2[u_\tau^{N-1}, v_\eta^M]^2$ and summing both sides of the inequality by $C_3\mathcal{W}_2[u_\tau^{N-1}, v_\eta^M]^2$, for some prefactor C_3 and $C = C_1^{-1}$:

Lemma 6.4. *The discrete solutions $(u_\tau^n)_{n\in\mathbb{N}}$ and $(v_\eta^m)_{m\in\mathbb{N}}$ satisfy*

$$\begin{aligned}
&\mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - \mathcal{W}_2[u_\tau^{N-1}, v_\eta^M]^2 \\
\leq\ & \frac{2\left(2(1-a)+a(1-2a)\lambda\tau\right)\tau}{\left(2(1-a)+(1-2a)\lambda\tau\right)(1+a\lambda\tau)}\mathcal{E}(v_\eta^M) - \frac{2(1-2a)\tau}{2(1-a)+(1-2a)\lambda\tau}\mathcal{E}(u_\tau^N) \\
&- \frac{2\tau}{\left(2(1-a)+(1-2a)\lambda\tau\right)(1+a\lambda\tau)}\mathcal{E}(u_\tau^{N+a-1}) + \frac{1-2a(1-a)}{a\left(2(1-a)+(1-2a)\lambda\tau\right)}\mathcal{W}_2[u_\tau^{N-1}, u_\tau^N]^2 \\
&- \frac{\left(2(1-a)+a(1-2a)\lambda\tau\right)\lambda\tau}{\left(2(1-a)+(1-2a)\lambda\tau\right)(1+a\lambda\tau)}\mathcal{W}_2[u_\tau^{N-1}, v_\eta^M]^2 - \frac{1}{a\left(2(1-a)+(1-2a)\lambda\tau\right)}\mathcal{W}_2[u_\tau^{N+a-1}, u_\tau^N]^2 \\
&- \frac{1}{a\left(2(1-a)+(1-2a)\lambda\tau\right)(1+a\lambda\tau)}\mathcal{W}_2[u_\tau^{N-1}, u_\tau^{N+a-1}]^2.
\end{aligned} \tag{164}$$

Proof. We investigate the estimate $q^{N,M} - q^{N-1,M}$, where $q^{n,m} = \mathcal{W}_2[u_\tau^n, v_\eta^m]^2$, with the use of (140) and (145): For the first part we rearrange (145) in order to substitute $\mathcal{W}_2[u_\tau^N, u]^2$ into the estimate, where we let the reference point $u = v_\eta^M$:

$$q^{N,M} - q^{N-1,M} = \mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - \mathcal{W}_2[u_\tau^{N-1}, v_\eta^M]^2$$

$$\leq \quad \frac{2(1-2a)\tau}{2(1-a)+(1-2a)\lambda\tau}\Big(\mathcal{E}(v_\eta^M)-\mathcal{E}(u_\tau^N)\Big)+\frac{1-2a(1-a)}{a\,(2(1-a)+(1-2a)\lambda\tau)}\mathcal{W}_2[u_\tau^{N-1},u_\tau^N]^2$$
$$-\frac{1+a(1-2a)\lambda\tau}{a\,(2(1-a)+(1-2a)\lambda\tau)}\mathcal{W}_2[u_\tau^{N-1},v_\eta^M]^2-\frac{1}{a\,(2(1-a)+(1-2a)\lambda\tau)}\mathcal{W}_2[u_\tau^{N+a-1},u_\tau^N]^2$$
$$+\frac{1}{a\,(2(1-a)+(1-2a)\lambda\tau)}\mathcal{W}_2[u_\tau^{N+a-1},v_\eta^M]^2. \tag{165}$$

For the next part, we rearrange (140) in order to substitute for $\mathcal{W}_2[u_\tau^{N+a-1},u]^2$ into the estimate $q^{N,M}-q^{N-1,M}$ for (165), where we let the reference point $u=v_\eta^M$. Furthermore, we have simplified the prefactor of $\mathcal{W}_2[u_\tau^{N-1},v_\eta^M]^2$ and have bounded the $\mathcal{W}_2[u_\tau^{N+a-1},u_\tau^N]^2$ term above by zero since via (136):

$$2(1-a)+(1-2a)\lambda\tau>2(1-a)+\frac{2(1-2a)(a-1)}{1-2a}=0.$$

Hence, we have that, after expansion and simplification:

$$q^{N,M}-q^{N-1,M}=\mathcal{W}_2[u_\tau^N,v_\eta^M]^2-\mathcal{W}_2[u_\tau^{N-1},v_\eta^M]^2$$
$$\leq \quad \frac{2(1-2a)\tau}{2(1-a)+(1-2a)\lambda\tau}(\mathcal{E}(v_\eta^M)-\mathcal{E}(u_\tau^N))+\frac{1-2a(1-a)}{a\,(2(1-a)+(1-2a)\lambda\tau)}\mathcal{W}_2[u_\tau^{N-1},u_\tau^N]^2$$
$$-\frac{1+a(1-2a)\lambda\tau}{a\,(2(1-a)+(1-2a)\lambda\tau)}\mathcal{W}_2[u_\tau^{N-1},v_\eta^M]^2-\frac{1}{a\,(2(1-a)+(1-2a)\lambda\tau)}\mathcal{W}_2[u_\tau^{N+a-1},u_\tau^N]^2$$
$$+\frac{1}{a\,(2(1-a)+(1-2a)\lambda\tau)}\Big\{\frac{2a\tau}{1+a\lambda\tau}(\mathcal{E}(v_\eta^M)-\mathcal{E}(u_\tau^{N+a-1}))$$
$$-\frac{1}{1+a\lambda\tau}(\mathcal{W}_2[u_\tau^{N-1},u_\tau^{N+a-1}]^2-\mathcal{W}_2[u_\tau^{N-1},v_\eta^M]^2)\Big\} \tag{166}$$
$$= \quad \frac{2\,(2(1-a)+a(1-2a)\lambda\tau)\,\tau}{(2(1-a)+(1-2a)\lambda\tau)\,(1+a\lambda\tau)}\mathcal{E}(v_\eta^M)-\frac{2(1-2a)\tau}{2(1-a)+(1-2a)\lambda\tau}\mathcal{E}(u_\tau^N)$$
$$-\frac{2\tau}{(2(1-a)+(1-2a)\lambda\tau)\,(1+a\lambda\tau)}\mathcal{E}(u_\tau^{N+a-1})+\frac{1-2a(1-a)}{a\,(2(1-a)+(1-2a)\lambda\tau)}\mathcal{W}_2[u_\tau^{N-1},u_\tau^N]^2$$
$$-\frac{(2(1-a)+a(1-2a)\lambda\tau)\,\lambda\tau}{(2(1-a)+(1-2a)\lambda\tau)\,(1+a\lambda\tau)}\mathcal{W}_2[u_\tau^{N-1},v_\eta^M]^2-\frac{1}{a\,(2(1-a)+(1-2a)\lambda\tau)}\mathcal{W}_2[u_\tau^{N+a-1},u_\tau^N]^2$$
$$-\frac{1}{a\,(2(1-a)+(1-2a)\lambda\tau)\,(1+a\lambda\tau)}\mathcal{W}_2[u_\tau^{N-1},u_\tau^{N+a-1}]^2,$$

and the result (164) is proved. □

Example 6.5. *For when $a=1/4$, this gives us the estimate,*

$$\mathcal{W}_2[u_\tau^N,v_\eta^M]^2-\mathcal{W}_2[u_\tau^{N-1},v_\eta^M]^2$$
$$\leq\frac{2(12+\lambda\tau)\tau}{(3+\lambda\tau)(4+\lambda\tau)}\mathcal{E}(v_\eta^M)-\frac{2\tau}{3+\lambda\tau}\mathcal{E}(u_\tau^N)-\frac{16\tau}{(3+\lambda\tau)(4+\lambda\tau)}\mathcal{E}(u_\tau^{N-3/4})+\frac{5}{3+\lambda\tau}\mathcal{W}_2[u_\tau^{N-1},u_\tau^N]^2$$
$$-\frac{(12+\lambda\tau)\lambda\tau}{(3+\lambda\tau)(4+\lambda\tau)}\mathcal{W}_2[u_\tau^{N-1},v_\eta^M]^2-\frac{8}{3+\lambda\tau}\mathcal{W}_2[u_\tau^{N-3/4},u_\tau^N]^2$$
$$-\frac{32}{(3+\lambda\tau)(4+\lambda\tau)}\mathcal{W}_2[u_\tau^{N-1},u_\tau^{N-3/4}]^2.$$

With this main first ingredient constructed, we now go ahead with the proof of estimating the comparison of two different solutions, via the comparison principle:

6.4 Comparison Principle Proof

The proof is lengthy with multiple steps, but the idea of the proof is as follows in more detail:

- For comparing the two solutions, which have time step sizes τ and η for each corresponding one, we introduce a parameter $R = \frac{\tau}{\eta}$. Working with rationals can be complicated, hence the term R is restricted to the set of natural numbers $\mathbb{N}$ to simplify the proof.

- We substitute in the discrete EVI estimates from stage one into stage two. This provides an estimate on $C\mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - \mathcal{W}_2[u_\tau^0, v_\eta^0]^2$, with $C = C_1^{-1}$ representing an exponential prefactor, which tends to $\exp(2\lambda T)$ for decreasing time step τ. Then afterwards, we adopt the telescopic summation as in [35].

- Finally, by applying the fact that some of our energy functional and metric terms, as discussed from earlier, are non-negative, we have that some of our right hand terms are bounded by zero. We need to multiply the estimate of $C\mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - \mathcal{W}_2[u_\tau^0, v_\eta^0]^2$ by C_1 to the form given in the theorem. Here, the exponential form of C comes into play nicely, since it is strictly positive.

With the fact that metric terms are non-negative and some of the energy functional terms, with non-positive prefactors, can be bounded above by zero, in view of (32) and (136), along with the fact that the distance between discrete solutions within the same family e.g. $\mathcal{W}_2[u_\tau^{N-1}, u_\tau^N]$ tends towards zero for decreasing time step size τ, the only main issue with proving numerical convergence is the mixed discrete solution metric term, which has a non-negative prefactor.

The first lemma rewrites the estimate (166) in a simpler form, with the introduction of two variables, which are found later to lie between -1 and 1. This will be useful when we attempt to construct convergent geometric summations for our final estimate.

Lemma 6.6. *From estimate (164), the discrete solutions $(u_\tau^n)_{n\in\mathbb{N}}$ and $(v_\eta^m)_{m\in\mathbb{N}}$ satisfy*

$$\begin{aligned} &\mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - (2(1-a) - (1-a(1-a))\lambda\tau)\, g_\tau h_\tau^{-1} \mathcal{W}_2[u_\tau^{N-1}, v_\eta^M]^2 \\ \leq\ & 2\,(2-a+a(1-a)\lambda\tau)\, g_\tau h_\tau^{-1} \tau \mathcal{E}(v_\eta^M) - 2(1-a) g_\tau \tau \mathcal{E}(u_\tau^N) - 2 g_\tau h_\tau^{-1} \mathcal{E}(u_\tau^{N+a-1}) \\ &+ \frac{1-2a(1-a)}{a} g_\tau \mathcal{W}_2[u_\tau^{N-1}, u_\tau^N]^2 - \frac{1}{a} g_\tau \mathcal{W}_2[u_\tau^{N+a-1}, u_\tau^N]^2 - \frac{1}{a} g_\tau h_\tau^{-1} \mathcal{W}_2[u_\tau^{N-1}, u_\tau^{N+a-1}]^2, \end{aligned} \tag{167}$$

where $g_\tau := \dfrac{1}{2(1-a) + (1-2a)\lambda\tau}$ *and* $h_\tau := 1 + a\lambda\tau$.

Proof. The approach is to rearrange (166) in order to derive an estimate on $\mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - C_3 \mathcal{W}_2[u_\tau^{N-1}, v_\eta^M]^2$, where C_3 is also a prefactor of the metric, dependent on a, λ and τ, which tends to one for decreasing time step τ. Using this as the main ingredient, similarly to as seen in [35, p. 17], via telescopic summation.

Firstly, by rearranging (166) we have the estimate for $q^{N,M} - C_3 q^{N-1,M} = \mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - C_3 \mathcal{W}_2[u_\tau^{N-1}, v_\eta^M]^2$ (we shall retire the $q^{\cdot}$ notation from now on):

$$\mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - \frac{2(1-a) - (1-2a(1-a))\,\lambda\tau}{(2(1-a) + (1-2a)\lambda\tau)\,(1+a\lambda\tau)} \mathcal{W}_2[u_\tau^{N-1}, v_\eta^M]^2$$

$$
\begin{aligned}
\leq \quad & \frac{2\left(2(1-a)+a(1-2a)\lambda\tau\right)\tau}{\left(2(1-a)+(1-2a)\lambda\tau\right)(1+a\lambda\tau)}\mathcal{E}(v_\eta^M) - \frac{2(1-2a)\tau}{2(1-a)+(1-2a)\lambda\tau}\mathcal{E}(u_\tau^N) \qquad (168)\\
& - \frac{2\tau}{\left(2(1-a)+(1-2a)\lambda\tau\right)(1+a\lambda\tau)}\mathcal{E}(u_\tau^{N+a-1}) + \frac{1-2a(1-a)}{a\left(2(1-a)+(1-2a)\lambda\tau\right)}\mathcal{W}_2[u_\tau^{N-1},u_\tau^N]^2\\
& - \frac{1}{a\left(2(1-a)+(1-2a)\lambda\tau\right)}\mathcal{W}_2[u_\tau^{N+a-1},u_\tau^N]^2\\
& - \frac{1}{a\left(2(1-a)+(1-2a)\lambda\tau\right)(1+a\lambda\tau)}\mathcal{W}_2[u_\tau^{N-1},u_\tau^{N+a-1}]^2,
\end{aligned}
$$

which can be rewritten as

$$
\begin{aligned}
& \mathcal{W}_2[u_\tau^N,v_\eta^M]^2 - \left(2(1-a)-(1-2a(1-a))\lambda\tau\right)g_\tau h_\tau^{-1}\mathcal{W}_2[u_\tau^{N-1},v_\eta^M]^2\\
\leq \quad & 2\left(2(1-a)+a(1-2a)\lambda\tau\right)g_\tau h_\tau^{-1}\tau\mathcal{E}(v_\eta^M) - 2(1-2a)g_\tau\tau\mathcal{E}(u_\tau^N) - 2g_\tau h_\tau^{-1}\tau\mathcal{E}(u_\tau^{N+a-1}) \qquad (169)\\
& + \frac{1-2a(1-a)}{a}g_\tau\mathcal{W}_2[u_\tau^{N-1},u_\tau^N]^2 - \frac{1}{a}g_\tau\mathcal{W}_2[u_\tau^{N+a-1},u_\tau^N]^2 - \frac{1}{a}g_\tau h_\tau^{-1}\mathcal{W}_2[u_\tau^{N-1},u_\tau^{N+a-1}]^2,
\end{aligned}
$$

where $g_\tau = \dfrac{1}{2(1-a)+(1-2a)\lambda\tau}$, $h_\tau = 1+a\lambda\tau$, and the result (167) is proved. ☐

By introducting a new notation $Q^{n,m} = H_\tau^n H_\eta^m \mathcal{W}_2[u_\tau^n, v_\eta^m]^2$, with $n \in \{\overline{n} \in \mathbb{N}_0 : \overline{n} \leq N\}$ and $m \in \{\overline{m} \in \mathbb{N}_0 : \overline{m} \leq M\}$ for our Wasserstein metrics, with its corresponding prefactors H_τ^n, an estimate on the difference between two final solutions and its two intial datum can now be prepared i.e. an estimate on $H_\tau^n H_\eta^m \mathcal{W}_2[u_\tau^n, v_\eta^m] - \mathcal{W}_2[u_\tau^0, v_\eta^0]$ can be constructed in comparison to the prefactor of $\mathcal{W}_2[u_\tau^{N-1}, v_\eta^M]^2$ from (167). Furthermore, by the same argument as [35], the notation H_τ^n is shown to be an exponential, time dependent prefactor:

Lemma 6.7. *The inequality (167) can be rewritten as a resulting estimate for* $C\mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - \mathcal{W}_2[u_\tau^0, v_\eta^0]^2$ *where* C *is an exponential prefactor.*

Proof. By defining $Q^{n,m} = H_\tau^n H_\eta^m \mathcal{W}_2[u_\tau^n, v_\eta^m]^2$, we have that $Q^{N,M} - Q^{0,0}$ gives the desired form where $C = H_\tau^n H_\eta^m$. By comparing prefactors of $\mathcal{W}_2$ from (167), we can define a sufficient expression for our functional H_τ^n etc. as follows:

$$
\begin{aligned}
Q^{N,M} - Q^{N-1,M} &= H_\tau^N H_\eta^M \mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - H_\tau^{N-1} H_\eta^M \mathcal{W}_2[u_\tau^{N-1}, v_\eta^M]^2\\
&= H_\tau^N H_\eta^M (\mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - H_\tau^{-1}\mathcal{W}_2[u_\tau^{N-1}, v_\eta^M]^2) \qquad (170)\\
&= H_\tau^N H_\eta^M \left(\mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - \left(2(1-a)-(1-2a(1-a))\lambda\tau\right)g_\tau h_\tau^{-1}\mathcal{W}_2[u_\tau^{N-1}, v_\eta^M]^2\right).
\end{aligned}
$$

The idea of the last two terms, is to bound the scaled metric term $Q^{n,m} - Q^{n-1,m}$ via the estimate (167). To achieve this, by comparing terms from the last two lines of (170), we have the prefactor H_τ^n, H_η^m set as

$$
H_\tau^{-1} = \frac{\left(2(1-a)-(1-2a(1-a))\lambda\tau\right)g_\tau}{h_\tau} \Rightarrow H_\tau = \frac{h_\tau}{\left(2(1-a)-(1-2a(1-a))\lambda\tau\right)g_\tau}, \qquad (171)
$$

and from the left hand side of (170), we have an exponentially dependent form of the Wasserstein metric term i.e. when $n = N$ and $m = M$,

$$
Q^{N,M} = \exp\Big(T\lambda_\tau + T\lambda_\eta\Big)\mathcal{W}_2[u_\tau^N, v_\eta^m]^2, \qquad (172)
$$

where $\lambda_\tau := \dfrac{\log(H_\tau)}{\tau}$.

Since N, M constitutes of the number of time step intervals, which of course increases as the time step τ decreases, we have that

$$\begin{aligned}\lim_{\tau\to 0} H_\tau^N &= \lim_{\tau\to 0}\left(\frac{(1+a\lambda\tau)\,(2(1-a)+(1-2a)\lambda\tau)}{2(1-a)-(1-2a(1-a))\lambda\tau}\right)^{T/\tau} \\ &= \lim_{\tau\to 0}\exp\left(\frac{T}{\tau}\log\left(\frac{(1+a\lambda\tau)\,(2(1-a)+(1-2a)\lambda\tau)}{2(1-a)-(1-2a(1-a))\lambda\tau}\right)\right),\end{aligned} \tag{173}$$

Before the next step, we evaluate from (136) that $\lim_{\tau\to 0}\lambda\tau = 0$, verified as a result of the Squeezing Theorem: since

$$\frac{2(a-1)}{(1-2a)\tau_*} \leq \lambda \leq 0 \Rightarrow -\frac{2(a-1)\tau}{(1-2a)\tau_*} \leq \lambda\tau \leq 0, \tag{174}$$

where τ_* is fixed, we have that

$$\lim_{\tau\to 0}\frac{2(a-1)\tau}{(1-2a)\tau_*} = \lim_{\tau\to 0} 0 = 0, \tag{175}$$

and by the L-Hôpital's rule, we have that (see Appendix C for workings)

$$\lim_{\tau\to 0}\frac{T}{\tau}\log\left(\frac{(1+a\lambda\tau)\,(2(1-a)+(1-2a)\lambda\tau)}{2(1-a)-(1-2a(1-a))\lambda\tau}\right) = \lambda T. \tag{176}$$

Hence by the limit chain rule, we have that

$$\lim_{\tau\to 0} H_\tau^N = \lim_{\tau\to 0}\left(\frac{(1+a\lambda\tau)\,(2(1-a)+(1-2a)\lambda\tau)}{2(1-a)-(1-2a(1-a))\lambda\tau}\right)^{T/\tau} = \lim_{u\to\lambda T}\exp(u) = \exp(\lambda T), \tag{177}$$

and the result of the lemma is complete. □

Now that we have selected the notation H_τ^n to be in line with the left hand side of our estimate (167), the difference $Q^{N,M} - Q^{0,0}$ can be expanded into the telescopic summation, from which we can begin substituting the estimate (167) for each summation term (the time point indexes will obviously differ for each one). The omission of non-positive terms, after verifying the prefactors are finite, can be shown for simplification.

Lemma 6.8. *The sequence of discrete solutions $(u_\tau^n)_{n\in\mathbb{N}}$ and $(v_\eta^m)_{m\in\mathbb{N}}$ satisfy*

$$\begin{aligned} Q^{N,M} - Q^{0,0} \leq\ & 2\left(2(1-a)+a(1-2a)\lambda\eta\right)\eta g_\eta h_\eta^{-1} H_\tau^N \sum_{m=RN+1}^{M} H_\eta^m \mathcal{E}(u_\tau^N) \\ & + 2\left(2(1-a)+a(1-2a)+\lambda\tau\right)\tau g_\tau h_\tau^{-1}\sum_{n=1}^{N} H_\tau^n H_\eta^{R(n-1)}\mathcal{E}(v_\eta^{R(n-1)}) \\ & + 2\left(2(1-a)+a(1-2a)\lambda\eta\right)\eta g_\eta h_\eta^{-1}\sum_{n=1}^{N}\sum_{m=R(n-1)+1}^{Rn} H_\eta^m H_\tau^n \mathcal{E}(u_\tau^n) \\ & + \frac{1-2a(1-a)}{a} g_\tau \sum_{n=1}^{N} H_\tau^n H_\eta^{R(n-1)}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2 \\ & + \frac{1-2a(1-a)}{a} g_\eta\left(H_\tau^N\sum_{m=RN+1}^{M} + \sum_{n=1}^{N}\sum_{m=R(n-1)+1}^{Rn} H_\tau^n\right) H_\eta^m \mathcal{W}_2[v_\eta^{m-1}, v_\eta^m]^2, \end{aligned} \tag{178}$$

where N, M represent the maximum number of time intervals.

Proof. With the preliminaries set up, the aim is to derive a estimate on $Q^{N,M} - Q^{0,0} = H_\tau^N H_\eta^M \mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - \mathcal{W}_2[u_\tau^0, v_\eta^0]^2$ via the telescopic summation from [35, p. 17], where we now bound (170), of various indexes, above by (167), which after collecting terms gives us

$$
\begin{aligned}
& Q^{N,M} - Q^{0,0} = H_\tau^N H_\eta^M \mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - \mathcal{W}_2[u_\tau^0, v_\eta^0]^2 \\
= \quad & \sum_{m=RN+1}^{M} (Q^{N,m} - Q^{N,m-1}) + \sum_{n=1}^{N} \left(\left(Q^{n,R(n-1)} - Q^{n-1,R(n-1)} \right) + \sum_{m=R(n-1)+1}^{Rn} (Q^{n,m} - Q^{n,m-1}) \right) \\
\leq \quad & 2H_\tau^N \left(2(1-a) + a(1-2a)\lambda\eta\right) \eta g_\eta h_\eta^{-1} \sum_{m=RN+1}^{M} H_\eta^m \mathcal{E}(u_\tau^N) \\
& + \frac{1-2a(1-a)}{a} g_\tau \sum_{n=1}^{N} H_\tau^n H_\eta^{R(n-1)} \mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2 - 2\tau g_\tau h_\tau^{-1} \sum_{n=1}^{N} H_\tau^n H_\eta^{R(n-1)} \mathcal{E}(u_\tau^{n+a-1}) \\
& - 2(1-2a)\eta g_\eta \left(H_\tau^N \sum_{m=RN+1}^{M} + \sum_{n=1}^{N} \sum_{m=R(n-1)+1}^{Rn} H_\tau^n \right) H_\eta^m \mathcal{E}(v_\eta^m) \qquad (179) \\
& - 2\eta g_\eta h_\eta^{-1} \left(H_\tau^N \sum_{m=RN+1}^{M} + \sum_{n=1}^{N} \sum_{m=R(n-1)+1}^{Rn} H_\tau^n \right) H_\eta^m \mathcal{E}(v_\eta^{m+a-1}) \\
& + 2 \left(\left(2(1-a) + a(1-2a)\lambda\eta\right) \eta g_\eta h_\eta^{-1} \sum_{n=1}^{N} \sum_{m=R(n-1)+1}^{Rn} H_\eta^m - (1-2a)\tau g_\tau \sum_{n=1}^{N} H_\eta^{R(n-1)} \right) H_\tau^n \mathcal{E}(u_\tau^n) \\
& + 2\left(2(1-a) + a(1-2a)\lambda\tau\right) \tau g_\tau h_\tau^{-1} \sum_{n=1}^{N} H_\tau^n H_\eta^{R(n-1)} \mathcal{E}(v_\eta^{R(n-1)}) \\
& + \frac{1-2a(1-a)}{a} g_\eta \left(H_\tau^N \sum_{m=RN+1}^{M} + \sum_{n=1}^{N} \sum_{m=R(n-1)+1}^{Rn} H_\tau^n \right) H_\eta^m \mathcal{W}_2[v_\eta^{m-1}, v_\eta^m]^2 \\
& - \frac{1}{a} g_\tau \sum_{n=1}^{N} H_\tau^n H_\eta^{R(n-1)} \mathcal{W}_2[u_\tau^{n+a-1}, u_\tau^n]^2 - \frac{1}{a} g_\tau h_\tau^{-1} \sum_{n=1}^{N} H_\tau^n H_\eta^{R(n-1)} \mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]^2 \\
& - \frac{1}{a} g_\eta h_\eta^{-1} \left(H_\tau^N \sum_{m=RN+1}^{M} + \sum_{n=1}^{N} \sum_{m=R(n-1)+1}^{Rn} H_\tau^n \right) H_\eta^m \mathcal{W}_2[v_\eta^{m-1}, v_\eta^{m+a-1}]^2 \\
& - \frac{1}{a} g_\eta \left(H_\tau^N \sum_{m=RN+1}^{M} + \sum_{n=1}^{N} \sum_{m=R(n-1)+1}^{Rn} H_\tau^n \right) H_\eta^m \mathcal{W}_2[v_\eta^{m+a-1}, v_\eta^m]^2.
\end{aligned}
$$

We can simplify this estimate significantly, shortly, but beforehand we shall verify whether the prefactors are finite with respect to (136): For $a \in (0, (1-\sqrt{2})/2)$, we have that $2(1-a) + (1-2a)\lambda\tau$ and $1 + a\lambda\tau$ are both non-zero hence

$$0 < g_\tau, h_\tau^{-1} < \infty,$$

holds for all λ satisfying (136) and similarly for when τ is replaced by η (τ is directly proportionate to η).

From (171), $H_\tau = \dfrac{h_\tau}{[2(1-a) - (1-2a(1-a))\lambda\tau] g_\tau}$, since we have from the assumption (136) that:

(i) $1+a\lambda\tau \in \left(\frac{2a^2-4a+1}{1-2a}, 1\right)$,

(ii) $2(1-a)+(1-2a)\lambda\tau \in (0, 2(1-a))$,

(iii) $2(1-a)-(1-2a(1-a))\lambda\tau \in \left(2(1-a), \frac{4(1-a)^3}{1-2a}\right)$,

the numerator of H_τ hence lies between $[0, 2(1-a)]$ and the denominator of H_τ, as from (iii), giving us that $H_\tau \in (0,1)$.

Hence, we have that $H_\tau, H_\eta \in (0,1)$ and from our earlier statements regarding $\mathcal{E}(\cdot)$ and metric terms, that is these are both non-negative terms that we bound above by zero, thus simplifies our estimate to

$$\begin{aligned}
& Q^{N,M} - Q^{0,0} \\
\leq\ & 2\left(2(1-a)+a(1-2a)\lambda\eta\right)\eta g_\eta h_\eta^{-1} H_\tau^N \sum_{m=RN+1}^{M} H_\eta^m \mathcal{E}(u_\tau^N) \\
& + \frac{1-2a(1-a)}{a} g_\tau \sum_{n=1}^{N} H_\tau^n H_\eta^{R(n-1)} \mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2 \\
& + 2\left(2(1-a)+a(1-2a)+\lambda\tau\right)\tau g_\tau h_\tau^{-1} \sum_{n=1}^{N} H_\tau^n H_\eta^{R(n-1)} \mathcal{E}(v_\eta^{R(n-1)}) \\
& + 2\left(2(1-a)+a(1-2a)\lambda\eta\right)\eta g_\eta h_\eta^{-1} \sum_{n=1}^{N} \sum_{m=R(n-1)+1}^{Rn} H_\eta^m H_\tau^n \mathcal{E}(u_\tau^n) \\
& + \frac{1-2a(1-a)}{a} g_\eta \left(H_\tau^N \sum_{m=RN+1}^{M} + \sum_{n=1}^{N} \sum_{m=R(n-1)+1}^{Rn} H_\tau^n \right) H_\eta^m \mathcal{W}_2[v_\eta^{m-1}, v_\eta^m]^2,
\end{aligned}$$

and the result (178) is proved. □

Before moving to the next step, we mention the following corollary concerning the number of time step intervals N, M for $(u_\tau^n)_{n\in\mathbb{N}}$ and $(v_\eta^m)_{m\in\mathbb{N}}$, respectively. This assists in simplifying our final proof.

Corollary 6.9. *The expression $M - RN \leq R$ for maximum time grid intervals N and M for discrete solutions $(u_\tau^n)_{n\in\mathbb{N}}$ and $(v_\eta^m)_{m\in\mathbb{N}}$. Indeed, these are defined as:*

$$N := max\{n : n\tau \leq T\}, \quad M := max\{m : m\eta \leq T\}.$$

Proof. The result $M - RN \leq R$ from [35, p. 17] is shown first. This gives us, since via definition, adding another interval to N exceeds the terminal time T:

$$M \leq \frac{T}{\eta}, \qquad N+1 \geq \frac{T}{\tau} \Rightarrow N \geq \frac{T}{\tau} - 1.$$

Hence we have that

$$M - RN \leq \frac{T}{\eta} - \frac{\tau}{\eta}\left(\frac{T}{\tau} - 1\right) = \frac{T}{\eta} - \frac{T}{\eta} + \frac{\tau}{\eta} = R,$$

and the result is proved. □

And now we proceed to the final part of the comparison principle proof. In addition to proving that the Wasserstein metric is finite i.e. $\frac{1}{\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2 < \infty$, our job now is to show that our estimates on our energy functional terms (see (157)-(159) from Section 5.13) give us an estimate on $Q^{N,M} - Q^{0,0}$, which is proportional to τ, thus generating our main ingredient for proving numerical convergence of our discrete solution to the Wasserstein gradient flow problem.

Lemma 6.10. *An application of estimates (157)-(159) give us the final result (162).*

Proof. The two metric terms clearly vanish for sufficiently small τ. But with the energy functional terms non-negative, more work is to be done here. By applying some of the estimates on the energy functional terms, term by term derived from Section 5.12, convergence towards zero is easily proved:

(i) Applying (157) gives us

$$
\begin{aligned}
& 2\left(2(1-a)+a(1-2a)+\lambda\eta\right)\eta g_\eta h_\eta^{-1} H_\tau^N \sum_{m=RN+1}^{M} H_\eta^m \mathcal{E}(u_\tau^N) \\
\leq\; & 2\left(2(1-a)+a(1-2a)\lambda\eta\right)\eta g_\eta h_\eta^{-1} H_\tau^N \sum_{m=RN+1}^{M} H_\eta^m \left(K_1 + \frac{1-2a(1-a)}{2a(1-2a)\tau}\sum_{n=1}^{N}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2\right) \\
=\; & 2\left(2(1-a)+a(1-2a)\lambda\eta\right)\eta K_1 g_\eta h_\eta^{-1} H_\tau^N \sum_{m=RN+1}^{M} H_\eta^m \\
& + \frac{1-2a(1-a)}{a(1-2a)R}\left(2(1-a)+a(1-2a)\lambda\eta\right) g_\eta h_\eta^{-1} H_\tau^N \sum_{m=RN+1}^{M} H_\eta^m \sum_{n=1}^{N}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2 \\
\leq\; & 2\left(2(1-a)+a(1-2a)\lambda\eta\right)\tau K_1 g_\eta h_\eta^{-1} H_\tau^N \\
& + \frac{1-2a(1-a)}{a(1-2a)}\left(2(1-a)+a(1-2a)\lambda\eta\right) g_\eta h_\eta^{-1} H_\tau^N H_\eta^{RN+1} \sum_{n=1}^{N}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2 \\
\leq\; & 2\left(2(1-a)+a(1-2a)\lambda\eta\right)\tau K_1 g_\eta h_\eta^{-1} H_\tau^N \\
& + \frac{1-2a(1-a)}{a(1-2a)}\left(2(1-a)+a(1-2a)\lambda\eta\right) g_\eta h_\eta^{-1} \sum_{n=1}^{N}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2.
\end{aligned} \tag{180}
$$

where we applied corollary 6.9.

(ii) Applying (159) gives us, when computing the sum to infinity on H_τ since this lies between $(0,1)$:

$$
\begin{aligned}
& 2\left(2(1-a)+a(1-2a)\lambda\tau\right)\tau g_\tau h_\tau^{-1}\sum_{n=1}^{N} H_\tau^n H_\eta^{R(n-1)}\mathcal{E}(v_\eta^{R(n-1)}) \\
\leq\; & 2\left(2(1-a)+a(1-2a)\lambda\tau\right)\tau g_\tau h_\tau^{-1}\sum_{n=1}^{N} H_\eta^{R(n-1)+n}\left(K_1 + \frac{1-2a(1-a)}{2a(1-2a)\eta}\mathcal{W}_2[v_\eta^{R(n-1)-1}, v_\eta^{R(n-1)}]^2\right) \\
=\; & 2K_1\left(2(1-a)+a(1-2a)\lambda\tau\right)\tau g_\tau h_\tau^{-1}\sum_{n=1}^{N} H_\eta^{R(n-1)+n} \\
& + \frac{1-2a(1-a)}{a(1-2a)} R\left(2(1-a)+a(1-2a)\lambda\tau\right) g_\tau h_\tau^{-1}\sum_{n=1}^{N} H_\eta^{R(n-1)+1}\mathcal{W}_2[v_\eta^{R(n-1)-1}, v_\eta^{R(n-1)}]^2
\end{aligned}
$$

$$
\begin{aligned}
&\leq \quad 2K_1\left(2(1-a)+a(1-2a)\lambda\tau\right)\tau g_\tau h_\tau^{-1}\left(\frac{H_\eta}{1-H_\eta^{R+1}}\right) \\
&\qquad + \frac{1-2a(1-a)}{a(1-2a)} R\left(2(1-a)+a(1-2a)\lambda\tau\right) g_\tau h_\tau^{-1}\sum_{n=1}^{N}\mathcal{W}_2[v_\eta^{R(n-1)-1}, v_\eta^{R(n-1)}]^2. \qquad (181)
\end{aligned}
$$

The rational expression from term one in (181) simplifies as follows:

$$
\begin{aligned}
\frac{H_\eta}{1-H_\eta^{R+1}} &= \frac{h_\eta}{\left(2(1-a)-(1-2a(1-a))\lambda\eta\right)g_\eta}\cdot\frac{\left(2(1-a)-(1-2a(1-a))\lambda\eta\right)^{R+1}g_\eta^{R+1}}{\left(2(1-a)-(1-2a(1-a))\lambda\eta\right)^{R+1}g_\eta^{R+1}-h_\eta^{R+1}} \\
&= \frac{\left(2(1-a)-(1-2a(1-a))\lambda\eta\right)^{R}g_\eta^{R}h_\eta}{\left(2(1-a)-(1-2a(1-a))\lambda\eta\right)^{R+1}g_\eta^{R+1}-h_\eta^{R+1}} \\
&= \frac{\left(\frac{(2(1-a)-(1-2a(1-a))\lambda\eta)}{2(1-a)+(1-2a)\lambda\eta}\right)^{R}(1+a\lambda\eta)}{\left(\frac{(2(1-a)-(1-2a(1-a))\lambda\eta)}{2(1-a)+(1-2a)\lambda\eta}\right)^{R+1}-(1+a\lambda\eta)^{R+1}} \qquad (182) \\
&= \frac{(1+a\lambda\eta)\left(2(1-a)-(1-2a(1-a))\lambda\eta\right)^{R}\left(2(1-a)+(1-2a)\lambda\eta\right)}{\left(2(1-a)-(1-2a(1-a))\lambda\eta\right)^{R+1}-(1+a\lambda\eta)^{R+1}\left(2(1-a)+(1-2a)\lambda\eta\right)^{R+1}},
\end{aligned}
$$

since the denominator after simplification is non-zero for all λ satisfying (136) (see Appendix D), hence the first term of (181) converges to zero as $\tau \to 0$.

In part (ii), we applied the fact that $H_\tau \leq H_\eta$ since, from (171) and that $\lambda < 0$, $\eta \leq \tau$:

$$
H_\tau = \frac{(1+a\lambda\tau)\left(2(1-a)+(1-2a)\lambda\tau\right)}{\left(2(1-a)-(1-2a(1-a))\lambda\tau\right)} \leq \frac{(1+a\lambda\eta)\left(2(1-a)+(1-2a)\lambda\eta\right)}{\left(2(1-a)-(1-2a(1-a))\lambda\eta\right)} = H_\eta. \qquad (183)
$$

We already know that $H_\tau \in (0,1)$.

Also, since we deduced that $H_\tau, H_\eta \in (0,1)$, we can construct its finite sum to infinity.

Note that the denominator in the last term of (181) is non-zero for assumption (136).

The remarks are applied to part (iii) below:

(iii) Applying (160) gives us, when computing the sum to infinity on H_τ since this lies between $(0,1)$:

$$
\begin{aligned}
&2\left(2(1-a)+a(1-2a)\lambda\eta\right)g_\eta h_\eta^{-1}\sum_{n=1}^{N}\sum_{m=R(n-1)+1}^{Rn}H_\eta^m H_\tau^n\mathcal{E}(u_\tau^n) \\
&\leq \quad 2\left(2(1-a)+a(1-2a)\lambda\eta\right)\eta g_\eta h_\eta^{-1}\sum_{n=1}^{N}\sum_{m=R(n-1)+1}^{Rn}H_\eta^{m+n}\left(K_1+\frac{1-2a(1-a)}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1},u_\tau^n]^2\right) \\
&\leq \quad 2K_1\left(2(1-a)+a(1-2a)\lambda\eta\right)\eta g_\eta h_\eta^{-1}\sum_{n=1}^{N}\sum_{m=R(n-1)+1}^{Rn}H_\eta^{m+n} \\
&\qquad + \frac{1-2a(1-a)}{a(1-2a)R}\left(2(1-a)+a(1-2a)\lambda\eta\right)g_\eta h_\eta^{-1}\sum_{n=1}^{N}\sum_{m=R(n-1)+1}^{Rn}H_\eta^{m+n}\mathcal{W}_2[u_\tau^{n-1},u_\tau^n]^2 \\
&\leq \quad 2K_1\left(2(1-a)+a(1-2a)\lambda\eta\right)\eta g_\eta h_\eta^{-1}\sum_{n=1}^{N}\sum_{m=R(n-1)+1}^{Rn}H_\eta^{n+R(n-1)+1}
\end{aligned}
$$

$$+\frac{1-2a(1-a)}{a(1-2a)R}\left(2(1-a)+a(1-2a)\lambda\eta\right)g_\eta h_\eta^{-1}\sum_{n=1}^{N}\sum_{m=R(n-1)+1}^{Rn}\mathcal{W}_2[u_\tau^{n-1},u_\tau^n]^2$$

$$= 2K_1\left(2(1-a)+a(1-2a)\lambda\eta\right)\tau g_\eta h_\eta^{-1}\sum_{n=1}^{N}H_\eta^{n+R(n-1)+1}$$

$$+\frac{1-2a(1-a)}{a(1-2a)}\left(2(1-a)+a(1-2a)\lambda\eta\right)g_\eta h_\eta^{-1}\sum_{n=1}^{N}\mathcal{W}_2[u_\tau^{n-1},u_\tau^n]^2$$

$$\leq 2K_1\left(2(1-a)+a(1-2a)\lambda\eta\right)\tau g_\eta h_\eta^{-1}\left(\frac{H_\eta^2}{1-H_\eta^{R+1}}\right) \tag{184}$$

$$+\frac{1-2a(1-a)}{a(1-2a)}\left(2(1-a)+a(1-2a)\lambda\eta\right)g_\eta h_\eta^{-1}\sum_{n=1}^{N}\mathcal{W}_2[u_\tau^{n-1},u_\tau^n]^2.$$

Since the rational expression from term one simplifies as follows, using (182):

$$\frac{H_\eta^2}{1-H_\eta^{R+1}}=\frac{(1+a\lambda\eta)\left(2(1-a)-(1-2a(1-a))\lambda\eta\right)^R\left(2(1-a)+(1-2a)\lambda\eta\right)}{\left(2(1-a)-(1-2a(1-a))\lambda\eta\right)^{R+1}-(1+a\lambda\eta)^{R+1}\left(2(1-a)+(1-2a)\lambda\eta\right)^{R+1}}, \tag{185}$$

which is finite since the denominator after simplification is non-zero for all λ satisfying (136). To prove by contradiction, The denominator is zero if and only if (see Appendix D for workings):

$$\lambda\tau=\frac{2(a-1)}{a(1-2a)}, \tag{186}$$

contradicting (136) for all $a\in(0,(1-\sqrt{2})/2)$. Thus the first term converges to 0 as $\tau\to 0$.

For the below estimate, we apply that

$$H_\tau^n H_\eta^{R(n-1)}\leq 1,$$

as a consequence of our definition of this functional.

Thus this gives us when substituting in parts (180), (181) and (184), alongside estimate (183) into (178):

$$\begin{aligned}
Q^{N,M}-Q^{0,0}=&H_\tau^N H_\eta^M\mathcal{W}_2[u_\tau^N,v_\eta^M]^2-\mathcal{W}_2[u_\tau^0,v_\eta^0]^2\\
\leq&2\left(2(1-a)+a(1-2a)\lambda\eta\right)K_1\tau g_\eta h_\eta^{-1}H_\tau^N\\
&+2K_1\left(2(1-a)+a(1-2a)\lambda\eta\right)\tau g_\eta h_\eta^{-1}\left(\frac{H_\eta^2}{1-H_\eta^{R+1}}\right)\\
&+\frac{1-2a(1-a)}{a(1-2a)}\left(2[2(1-a)+a(1-2a)\lambda\eta]g_\eta h_\eta^{-1}+g_\tau\right)\sum_{n=1}^{N}\mathcal{W}_2[u_\tau^{n-1},u_\tau^n]^2\\
&+2K_1[2(1-a)+a(1-2a)\lambda\tau]\tau g_\tau h_\tau^{-1}\left(\frac{H_\eta}{1-H_\eta^{R+1}}\right)\\
&+\frac{1-2a(1-a)}{a(1-2a)}R\left(2(1-a)+a(1-2a)\lambda\eta\right)g_\tau h_\eta^{-1}\tau\sum_{n=1}^{N}\mathcal{W}_2[v_\eta^{R(n-1)-1},v_\eta^{R(n-1)}]^2\\
&+\frac{1-2a(1-a)}{a(1-2a)}g_\eta\left(H_\tau^N\sum_{m=RN+1}^{M}+\sum_{n=1}^{N}\sum_{m=R(n-1)+1}^{Rn}H_\tau^n\right)\mathcal{W}_2[v_\eta^{m-1},v_\eta^m]^2.
\end{aligned}$$

Our energy estimates are substituted in, but the final barrier to negotiate is whether the $\sum_{n=1}^{N} \mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2$ terms are finite i.e. can we verify that the total velocity does not diverge to infinity. We have verified that the metric terms vanish for decreasing time step, but the summation dependent on the number of time grid points is not guaranteed, due to the number of grid points diverging as the time step decreases. However, Lemma 5.24 (150) shows this still is true, as a consequence of the Young's inequality, semi-convexity conditions and the differential EVI.

Thus we conclude that the estimate of $Q^{N,M} - Q^{0,0}$ is proportional to τ which, as a result, tends to zero for dissipating time steps (see (172)-(177) for how C approaches an exponential constant), and the comparison principle for the DIRK2 type scheme is proved.

In other words, dividing through by $C = H_\tau^N H_\eta^M = \exp(\lambda_\tau T + \lambda_\eta T)$ gives

$$
\begin{aligned}
&\mathcal{W}_2[u_\tau^N, v_\eta^M]^2 - C_1 \mathcal{W}_2[u_\tau^0, v_\eta^0]^2 \\
\leq & C_1 \Big\{ 2K_1 \left(2(1-a) + a(1-2a)\lambda\eta\right) \tau g_\eta h_\eta^{-1} H_\tau^N + \frac{1-2a(1-a)}{a(1-2a)} \Big(\left(2(1-a) + a(1-2a)\lambda\eta\right) (g_\eta h_\eta^{-1} \\
& + g_\tau) + g_\tau \Big) g_\eta h_\eta^{-1} \sum_{n=1}^{N} \mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2 + 2K_1 \left(2(1-a) + a(1-2a)\lambda\tau\right) \tau g_\tau h_\tau^{-1} \left(\frac{H_\eta}{1 - H_\eta^{R+1}} \right) \\
& + \frac{1-2a(1-a)}{a(1-2a)} R \left(2(1-a) + a(1-2a)\lambda\tau\right) g_\tau h_\eta^{-1} \sum_{n=1}^{N} \mathcal{W}_2[v_\eta^{R(n-1)-1}, v_\eta^{R(n-1)}]^2 \\
& + 2K_1 \left(2(1-a) + a(1-2a)\lambda\eta\right) \tau g_\eta h_\eta^{-1} \left(\frac{H_\eta^2}{1 - H_\eta^{R+1}} \right) \\
& + \frac{1-2a(1-a)}{a(1-2a)} g_\eta \left(H_\tau^N \sum_{m=RN+1}^{M} + \sum_{n=1}^{N} \sum_{m=R(n-1)+1}^{Rn} H_\tau^n \right) \mathcal{W}_2[v_\eta^{m-1}, v_\eta^m]^2 \Big\} \\
\leq & C_2 \tau,
\end{aligned}
\tag{187}
$$

where C_2 is the max of all prefactors.

Furthermore C_1 is such that

$$
\frac{1}{\exp(\lambda_\tau t + \lambda_\eta t)} \leq \frac{1}{\exp((\lambda_\tau + \lambda_\eta)T)} \leq \frac{1}{\exp(2\lambda_\tau T)},
$$

since $\lambda_\tau, \lambda_\eta \leq 0$ and $\lambda_\tau \leq \lambda_\eta = C$ via (183). Furthermore, this tends to $\exp(-2\lambda T)$ (see Lemma 6.7 and (177)). This completes the proof of the comparison principle theorem (result (162)).

□

6.5 Numerical Convergence Proof

With the comparison principle taken care of, we can now proceed to finish the numerical convergence proof, based on [35, p.18-19]. In other words we now complete the proof of Theorem 6.1.

Again there are several steps to the proof. The main steps of the proof are as follows:

- Using the initial assumption we have convergence of the discrete solution to the limit curve u_*. Since the estimate, from the comparison principle was $C\mathcal{W}_2[u_\tau^n, v_\eta^m]^2 - \mathcal{W}_2[u_\tau^0, v_\eta^0]^2 \leq K\tau$, where

$K = C_1^{-1}C_2$ then rearranging the inequality and setting $\tau \to 0$ gives the simple result here. Note that we now replace u_τ by u_{τ_k} and another solution v_η by u_{τ_l} (τ is replaced by τ_k and η is replaced by τ_l).

- Taking our known estimates on the energy and the metrics, plus the Young's inequality, this shows that the metric derivative was uniformly bounded in $L^2(0,T)$.
- Substituting in the estimate EVI from stage one into the EVI from stage two. Hence, by some manipulation, this showed that the limit curve u_* satisfied the EVI in continuous form.

Lemma 6.11. *From the comparison principle theorem and as a consequence of assumptions (161), the values $(\overline{u}_{\tau_k}(t))_{k\in\mathbb{N}}$ converge in the complete probability space to the limit curve $u_*(t)$.*

Proof. We can bound the distance between two interpolated solutions by the same value of order t^α, where $\alpha \geq 0$ i.e. when $\alpha = 1$ and assumption (161):

$$\mathcal{W}_2[\overline{u}_{\tau_k}(t), \overline{u}_{\tau_l}(t)]^2 \leq C\mathcal{W}_2[u^0_{\tau_k}, u^0_{\tau_l}]^2 + C\tau_k \leq C(1+K_3)\tau_k = C_*\tau_k,$$

where $C = (H^N_\tau H^M_\eta)^{-1}$ and $C_* = C(1+K_3)$.

Hence as $\tau_k \to 0$, this gives us $\mathcal{W}_2[\overline{u}_{\tau_k}(t), \overline{u}_{\tau_l}(t)] \to 0$ i.e. the sequence of discrete solutions $(\overline{u}_{\tau_k}(t))_{k\in\mathbb{N}}$ converges to the limit curve $u_*(t)$ uniformly for $t \in [0,T]$. □

Lemma 6.12. *The metric derivative is uniformly bounded in $L^2(0,T)$ and possesses a $L^2(0,T)$-weakly convergent subsequence with a limit.*

Proof. The time discrete solution derivatives (metric derivatives) are assigned in relation to the interpolated solution as

$$|u'_{\tau_k}|(t) = \frac{\mathcal{W}_2[\overline{u}_{\tau_k}(t-\tau_k), \overline{u}_{\tau_k}(t)]}{\tau_k} = \frac{\mathcal{W}_2[u^{n-1}_{\tau_k}, u^n_{\tau_k}]}{\tau_k},$$

for $t \in ((k-1)\tau_k, k\tau_k]$, to show that the metric derivative is uniformly bounded. We must show that the right hand side is well defined, that is

$$\frac{\mathcal{W}_2[u^{n-1}_{\tau_k}, u^n_{\tau_k}]}{\tau_k} \leq C,$$

verified immediately from Lemma 5.24 (150). The result is achieved also by the output of Lemma 5.27, part (i).

Alternatively, parts (ii) and (iii) of Lemma 5.27 provide uniformly bounded metric derivatives with respect to $\mathcal{W}_2[u^{n-1}_{\tau_k}, u^{n+a-1}_{\tau_k}]$ and $\mathcal{W}_2[u^{n+a-1}_{\tau_k}, u^n_{\tau_k}]$.

Thus, this gives us that the metric derivative is uniformly bounded and this gives us the that L^2-Wasserstein distance between the limit curve at two different time points in $[0,T]$ is bounded i.e. has a $L^2(0,T)$ weakly convergent subsequence.

This gives us the final result i.e. the limiting curve u_* satisfies the evolution variational inequality (EVI). □

Lemma 6.13. *By combining together the discrete EVI estimates for each intermediate stages, we aim to verify that, where A-G are prefactors with respect to a,τ,λ:*

$$\begin{aligned}
&\left(\frac{A}{2\tau}+\frac{\lambda}{2}\right)\mathcal{W}_2[u_\tau^n,u]^2-\frac{A}{2\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2\\
\leq\quad &\mathcal{E}(u)-\mathcal{E}(u_\tau^n)+B\mathcal{W}_2[u_\tau^{n-1},u_\tau^{n+a-1}]^2+C\mathcal{W}_2[v_\eta^{m-1},v_\eta^{m+a-1}]^2+D\mathcal{W}_2[u_\tau^{n-1},u_\tau^n]^2+E\mathcal{W}_2[v_\eta^{m-1},v_\eta^m]^2\\
&+F\mathcal{W}_2[u_\tau^{n+a-1},u_\tau^n]^2+G\mathcal{W}_2[v_\eta^{m+a-1},v_\eta^m]^2,
\end{aligned}$$

given that we have sequence of vanishing time steps $(\tau_k)_{k\in\mathbb{N}}$, and satisfies the integrated form of the EVI (see [35, p. 5], [41, Sect. 6.2]):

$$\frac{1}{2}\mathcal{W}_2[u_*(t),u]^2-\frac{1}{2}\mathcal{W}_2[u_*(s),u]^2\leq\int_s^t\left(\mathcal{E}(u)-\mathcal{E}(u_*(r))-\frac{\lambda}{2}\mathcal{W}_2[u_*(r),u]^2\right)dr.$$

Proof. Firstly, by rearranging (144), we have that

$$\mathcal{W}_2[u_\tau^{n+a-1},u]^2\leq\frac{2a\tau}{1+a\lambda\tau}\left(\mathcal{E}(u)-\mathcal{E}(u_\tau^{n+a-1})\right)+\frac{1}{1+a\lambda\tau}\left(\mathcal{W}_2[u_\tau^{n-1},u]^2-\mathcal{W}_2[u_\tau^{n-1},u_\tau^{n+a-1}]^2\right),$$

then substituting into (148) gives us in simplified form:

$$\begin{aligned}
&\left(\frac{1-a}{(1-2a)\tau}+\frac{\lambda}{2}\right)\mathcal{W}_2[u_\tau^n,u]^2+\left(\frac{1-2a(1-a)}{2a(1-2a)\tau}-\frac{1}{2a(1-2a)(1+a\lambda\tau)\tau}\right)\mathcal{W}_2[u_\tau^{n-1},u]^2\\
\leq\quad &\frac{2(1-a)+a(1-2a)\lambda\tau}{(1-2a)(1+a\lambda\tau)}\mathcal{E}(u)-\mathcal{E}(u_\tau^n)-\frac{1}{(1-2a)(1+a\lambda\tau)}\mathcal{E}(u_\tau^{n+a-1})\\
&-\frac{1}{2a(1-2a)(1+a\lambda\tau)\tau}\mathcal{W}_2[u_\tau^{n-1},u_\tau^{n+a-1}]^2-\frac{1}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n+a-1},u_\tau^n]^2\\
&+\frac{1-2a(1-a)}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1},u_\tau^n]^2.
\end{aligned}\tag{188}$$

For simplicity we rearrange and substitute (121) into (188) to give us

$$\begin{aligned}
&\left(\frac{1-a}{(1-2a)\tau}+\frac{\lambda}{2}\right)\mathcal{W}_2[u_\tau^n,u]^2-\frac{2(1-a)-(1-2a(1-a))\lambda\tau}{2(1-2a)(1+a\lambda\tau)\tau}\mathcal{W}_2[u_\tau^{n-1},u]^2\\
\leq\quad &\frac{2(1-a)+a(1-2a)\lambda\tau}{(1-2a)(1+a\lambda\tau)}\left(\mathcal{E}(u)-\mathcal{E}(u_\tau^n)\right)-\frac{1-2a+a^2}{a(1-2a)^2(1+a\lambda\tau)\tau}\mathcal{W}_2[u_\tau^{n-1},u_\tau^{n+a-1}]^2\\
&-\frac{2(1-a)+a(1-2a)\lambda\tau}{2a(1-2a)^2(1+a\lambda\tau)\tau}\mathcal{W}_2[u_\tau^{n+a-1},u_\tau^n]^2\\
&+\frac{(1-2a(1-a))\left(2(1-a)+a(1-2a)\lambda\tau\right)}{2a(1-2a)^2(1+a\lambda\tau)\tau}\mathcal{W}_2[u_\tau^{n-1},u_\tau^n]^2.
\end{aligned}\tag{189}$$

Multiplying through by $\dfrac{(1-2a)(1+a\lambda\tau)}{2(1-a)+a(1-2a)\lambda\tau}$ gives us for the prefactors of:

- $\mathcal{W}_2[u_\tau^n,u]^2$: $\dfrac{(1-a)(1+a\lambda\tau)}{(2(1-a)+a(1-2a)\lambda\tau)\,\tau}+\dfrac{\lambda\tau(1-2a)(1+a\lambda\tau)}{2\,(2(1-a)+a(1-2a)\lambda\tau)\,\tau}$.

  Furthermore, we wish to rewrite as

  $$\frac{(1-a)(1+a\lambda\tau)}{(2(1-a)+a(1-2a)\lambda\tau)\,\tau}+\overline{a}+\frac{\lambda}{2},$$

  where we work out $\overline{a}$ as

  $$\overline{a}=\frac{\lambda\tau(1-2a)(1+a\lambda\tau)}{2\,(2(1-a)+a(1-2a)\lambda\tau)\,\tau}-\frac{\lambda}{2}=-\frac{\lambda\tau}{2\,(2(1-a)+a(1-2a)\lambda\tau)\,\tau}.$$

Thus the prefactor is written as

$$\begin{aligned}
&\frac{(1-a)(1+a\lambda\tau)}{(2(1-a)+a(1-2a)\lambda\tau)\,\tau} - \frac{\lambda\tau}{2\,(2(1-a)+a(1-2a)\lambda\tau)\,\tau} + \frac{\lambda}{2} \\
=&\frac{2(1-a)-(1-2a(1-a))\,\lambda\tau}{2\,(2(1-a)+a(1-2a)\lambda\tau)\,\tau} + \frac{\lambda}{2}.
\end{aligned}$$

- $\mathcal{W}_2[u_\tau^{n-1}, u]^2$: From simple cancellation:

$$\frac{(1-2a(1-a))\,\lambda\tau - 2(1-a)}{2(1-2a)(1+a\lambda\tau)\tau} \cdot \frac{(1-2a)(1+a\lambda\tau)}{2(1-a)+a(1-2a)\lambda\tau} = \frac{(1-2a(1-a))\,\lambda\tau - 2(1-a)}{2\,(2(1-a)+a(1-2a)\lambda\tau)\,\tau}.$$

Thus multiplying through (189) by $\dfrac{(1-2a)(1+a\lambda\tau)}{2(1-a)+a(1-2a)\lambda\tau}$ gives us

$$\begin{aligned}
&\left(\frac{2(1-a)-(1-2a(1-a))\,\lambda\tau}{2\,(2(1-a)+a(1-2a)\lambda\tau)\,\tau} + \frac{\lambda}{2}\right)\mathcal{W}_2[u_\tau^n, u]^2 - \frac{2(1-a)-(1-2a(1-a))\,\lambda\tau}{2\,(2(1-a)+a(1-2a)\lambda\tau)\,\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 \\
\leq&\mathcal{E}(u) - \mathcal{E}(u_\tau^n) - \frac{1-2a+a^2}{a(1-2a)\,(2(1-a)+a(1-2a)\lambda\tau)\,\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]^2 \\
&- \frac{1}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n+a-1}, u_\tau^n]^2 + \frac{(1-2a(1-a))}{2a(1-2a)\tau}\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2.
\end{aligned}$$

Since $\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2, \mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]^2, \mathcal{W}_2[u_\tau^{n+a-1}, u_\tau^n]^2$ tends towards zero for decreasing time step size τ_k (as $k \to \infty$), gives us, when multiplying through by τ_k (see Lemmas 5.27 and 6.12), summing from $n = n_k(s)+1$ to $n = n_k(t)$ (we define $n_k(r) = \max\{n : n\tau_k \leq r\}$ then evaluating as $k \to \infty$:

$$\begin{aligned}
&\frac{\lambda}{2}\int_s^t \mathcal{W}_2[u_*(r), u]^2 dr + \lim_{k\to\infty} \frac{2(1-a)-(1-2a(1-a))\,\lambda\tau_k}{2\,(2(1-a)+a(1-2a)\lambda\tau_k)} \left(\mathcal{W}_2[u_*(t), u]^2 - \mathcal{W}_2[u_*(s), u]^2\right) \\
\leq\quad&\int_s^t (\mathcal{E}(u) - \mathcal{E}(u_*(r)))\,dr.
\end{aligned}$$

Since $\lambda\tau \to 0$ as $k \to \infty$ ($\tau \to 0$), the prefactor $\dfrac{2(1-a)-(1-2a(1-a))\,b}{2\,(2(1-a)+a(1-2a)b)} \to \dfrac{1}{2}$ and the result is proved. $\square$

7 Fully Discrete Forms of Numerical Schemes

Now that we have analysed the DIRK2 scheme, we move on to set out the analysis of the numerical results of our various BDF and DIRK schemes in the next two sections. We provide a brief explanation on the spatial discretisation process (this was proposed from [17, p. 9+]) of our schemes. Since BDF2 to BDF6 schemes are either already investigated or it turns out that they possess no improved error/convergence rate from the lower order one (only BDF1 and BDF2 schemes are A-stable and have been demonstrated by articles [17, 35]), we will *only* review the discretisation for the DIRK2 Scheme that we have shown theoretical convergence to, plus an example of two higher order DIRK schemes.

We first outline the process, implemented by Düring et al. [17] in brief, before going into more detail later:

- Computing L^2- Wasserstein distances $\mathcal{W}_2$ are complicated, particularly when you are considering a large number of possible redistributions of the first configuration to the second (new) configuration. So in one space dimension, the discretisation relies on reformulating the **minimising movement schemes** into Lagrangian coordinates in terms of the pseudo-inverse distribution function:

  $$G : [0, M] \to \Omega,$$

  and its derivative i.e. $g = \partial_w G$, before discretising it with a Galerkin ansatz with piecewise affine linear basis functions.

  Here, the Lagrangian coordinate $\omega = U(x) \in [0, M]$ was introduced where U was the distribution of density $u(t; x)$ and conjugate to the inverse distribution function $G(t; \omega)$.

- This allows us to compute the Wasserstein Distance as the L^2-norm of G (see [47]), that is

  $$\mathcal{W}_2[u_1, u_2] := \left(\int_0^M |G_1(\omega) - G_2(\omega)|^2 \, d\omega \right)^{\frac{1}{2}}.$$

  During the reformulation of the scheme into Lagrangian coordinates, the Wasserstein distance was transformed as the L^2 norm of the derivative of the inverse distribution function G:

  $$\mathcal{W}_2[u^*, u]^2 := \iint_{[0,M]^2} (M - max(\eta, \eta')) \big(g(\eta) - g^*(\eta)\big) \big(g(\eta') - g^*(\eta')\big) d\eta d\eta'.$$

  The solution is later recovered back into Eulerian coordinates by the formula [17, Lem. 2.5], with k denoting the point on the spatial grid:

  $$u(x) = \sqrt{\frac{(g_k + g_{k-1})\delta_k}{2\left(g_k^2(x - x_{k-1}) + g_{k-1}^2(x_k - x)\right)}} \quad \text{with} \quad x_k = \frac{1}{2} \sum_{i=1}^{k} \delta_j (g_i + g_{i-1}).$$

- We then discretise in mass space using a Galerkin ansatz (order one) to obtain a fully discrete, finite dimensional problem.

- Finally, from the spatial discretisation, the Wasserstein distance in finite-dimensional form becomes

$$\mathcal{W}_2[u^*, u]^2 := \sum_{j,k=1}^{n} a_{j,k}(g_j - g_j^*)(g_k - g_k^*),$$

  with $a_{j,k} = a_{k,j}$ where $a_{j,k}$ are the entries of a symmetric matrix A. See [17, Lem. 2.6] for details of each specific entry, which provides us with a quadratic minimisation problem.

- Existence of a unique, discrete solution (well-posed) of the scheme is shown [17, Thm. 2.7]. In addition, its global minimiser is shown by deriving a priori estimate on the discrete solution.

- With a mass constraint, the Lagrange multiplier λ and its associated Lagrangian functional L^τ is introduced with $\Psi : \mathbb{G}_M^n \times \mathbb{G}_M^n$ such that

$$L^\tau(\mathbf{g}^*, \mathbf{g}, \lambda) := \Psi^\tau(\mathbf{g}^*; \mathbf{g}) - \lambda \left(1 - \sum_{k=1}^{n} \Delta_k g_k\right),$$

  which is transformed in Lagrangian coordinates (see [17, Lem. 2.3]). In other words, we minimise subject to the mass constraint and a depends on the BDF scheme, defined in the grid (see [17, p. 112]). The functional(s) $\Psi(\mathbf{g}^*; \mathbf{g})$ will be derived for each scheme in Section 7.4, but see [16, Lem. 2.3].

  The minimisers $\mathbf{g}$ of the Lagrangian functional are charcterised by its critical points $(\mathbf{g}, \lambda)$, by "classical theory of variations" [17]. To find the zeros, Newton's method is applied.

  In our setting, with higher order BDF/DIRK schemes carrying additional intermediate steps, the optimality conditions for $(\mathbf{g}, \lambda)$ i.e. $\mathbf{G_k} = \dfrac{\partial L^\tau}{\partial g_k}$ have to be derived for each scheme and stages.

7.1 Lagrangian coordinates

We summarise the transformation of the energy functional (Fisher information) from Eulerian coordinates to Lagrangian coordinates for the DLSS equation (Duering and Matthes [17]). Section 7.3 adapts this to other fourth order nonlinear PDEs.

As in [17], the Lagrangian coordinate $\omega = U(x) \in [0, M]$ is introduced where U is the distribution of density $u(x,t)$ and conjugate to the inverse distribution function $G(\omega, t)$. In other words, G is the inverse distribution function of u, with $g = \partial_\omega G : [0, M] \to \mathbb{R}_+$ and $u(x) = \partial_x U(x)$. A change of variable now carried out on the energy functional gives

$$\mathcal{E}(u) = \frac{1}{2}\int_0^M \left(\partial_\omega \left(\frac{1}{g(\omega)}\right)\right)^2 d\omega,$$

with the Wasserstein distance transformed as from (13), representing the L^2-norm of the inverse distribution function G.

Note that this is specific to the Fisher information, for the DLSS equation (31), but we can also consider other examples.

7.2 Spatial Discretization and discretisation of the Wasserstein term

We summarise the ingredients for the spatial discretisation, used by Duering and Matthes [17], to transform the problem into finite-dimensional form, using a Galerkin ansatz/ finite element in one-dimensional approach and weight vectors $\mathbf{g} = (g_1, g_2, ..., g_n)$.

The infinite dimensional variational problems to our inductive schemes are transformed into finite dimensional problems. As defined in [17, Sect. 2.4], we have

- n mesh points on the spatial grid, with $k \in [0, n]$ denoting the specific node.
- Mesh $\Omega_n = \{\omega_0, \omega_1, \ldots, \omega_n\}$ with $\omega_0 = 0$, $\omega_k < \omega_{k+1}$ and $\omega_n = M$.
- Single and double gaps: $\delta_k = \omega_k - \omega_{k-1}$ and $\Delta_k = \dfrac{\omega_{k+1} - \omega_{k-1}}{2}$.
- Mass constraint
$$\sum_{k=1}^{n} \Delta_k g_k = 1. \tag{190}$$
- Piecewise, linear functions $g : [0, M] \to \mathbb{R}_+$ and hat function $\phi_k : \Omega \to \mathbb{R}$ of the form
$$g(\omega) = \sum_{k=1}^{n} g_k \phi_k(\omega), \quad \text{where} \quad \phi_k \implies g(\omega_k) = g_k,$$
with the set of these functions defined as the ansatz space.
- The weight vectors $\mathbf{g} = (g_1, \ldots, g_n)$ with $g(0) = g(M) = g_n$.

Finally, the Wasserstein distance in finite-dimensional form becomes

$$\mathcal{W}_2[u^*, u]^2 = \sum_{j,k=1}^{n} a_{j,k}(g_j - g_j^*)(g_k - g_k^*),$$

with $a_{j,k} = a_{k,j}$ where $a_{j,k}$ are the entries of a symmetric matrix. See [17, Lem. 2.6] for details.

7.3 Discrete Energy Functionals

We briefly mention the Galerkin approach for the Fisher information from the DLSS equation and then for the other equations introduced at the end of Section 3. An alternative approach is shown in detail in Appendix E, motivated by equation (105) where $a = 1/2$. *Note* We work with (105) soon when $a = 1/4$ and $a = 2$.

The function g in the ansatz space of functions satisfies

$$g(\omega) = \frac{g_k(\omega - \omega_{k-1}) + g_{k-1}(\omega_k - \omega)}{\omega_k - \omega_{k-1}}, \qquad \omega \in [\omega_{k-1}, \omega_k]. \tag{191}$$

7.3.1 DLSS Equation

By the representation of the Fisher information (see [17, Sect. 2.5.2]):

$$\mathcal{E}(u) = \frac{1}{2}\int_0^M \left(\partial_\omega\left(\frac{1}{g}\right)\right)^2 d\omega = \frac{1}{2}\sum_{k=1}^n \mathbf{F}_k^d[\mathbf{g}], \tag{192}$$

$$\mathbf{F}_k^d[\mathbf{g}] = \frac{\delta_k^{-1}}{3}\left(\frac{1}{g_k} - \frac{1}{g_{k-1}}\right)^2\left(1 + \frac{g_{k-1}}{g_k} + \frac{g_k}{g_{k-1}}\right).$$

Alternatively, (192) gives us

$$\mathcal{E}_d(u) = -\frac{1}{2}\int_0^M \left(\frac{\partial_\omega g(\omega)}{(g(\omega))^2}\right)^2 d\omega. \tag{193}$$

Since $\partial_\omega g(\omega) = \dfrac{g_k - g_{k-1}}{\delta_k}$, we have for (193), when transforming into Lagrangian coordinates (see Appendix E for workings):

$$\mathcal{E}_d(u) = \frac{1}{6}\sum_{k=1}^n \mathbf{F}_k^d[\mathbf{g}], \qquad \mathbf{F}_k^d[\mathbf{g}] = \frac{1}{\delta_k}(g_k - g_{k-1})\left(\frac{1}{g_{k-1}^3} - \frac{1}{g_k^3}\right). \tag{194}$$

7.3.2 Thin Film Equation (34)

As explained at the beginning of [33], by Matthes and McCann et al. and [30, Thm. 3.9, 3.10] by Kamalinejad, the Thin Film equation (34) is a Wasserstein gradient flow of the Dirichlet energy functional:

$$\mathcal{E}_t(u) = \frac{1}{2}\int_{\mathbb{R}} (\partial_x u(x,t))^2\, dx,$$

and transforming into Lagrangian coordinates $x = G(w)$ gives

$$\mathcal{E}_t(u) = \frac{1}{2}\int_{\mathbb{R}} \left(\frac{\partial}{\partial x}\left(\frac{1}{g(\omega)}\right)\right)^2 dx = \frac{1}{2}\int_0^M \left(-\frac{\partial_\omega g(\omega)}{(g(\omega))^2}\frac{1}{g(\omega)}\right)^2 \frac{\partial x}{\partial \omega} d\omega = \frac{1}{2}\int_0^M \frac{(\partial_\omega g(\omega))^2}{(g(\omega))^5} d\omega.$$

Then, as we applied for the DLSS equation, the spatial discrete form of this gives us (see Appendix E for computation) the discrete form of the energy functional as

$$\mathcal{E}_t(u) = \frac{1}{8}\sum_{k=1}^n \mathbf{F}_k^t[\mathbf{g}], \qquad \mathbf{F}_k^t[\mathbf{g}] = \frac{1}{\delta_k}(g_k - g_{k-1})\left(\frac{1}{g_{k-1}^4} - \frac{1}{g_k^4}\right). \tag{195}$$

7.3.3 Fourth Order Nonlinear Equation (35)

We test/analyse for $a = 1/4$ and $a = 2$.

Again, as explained in [30, Thm. 3.9, 3.10], a PDE of the form for some $a \in \mathbb{R}$,

$$\partial_t u = -2a\partial_x(u(x,t)\partial_x((u(x,t))^{a-1}\partial_x^2(u(x,t))^a)),$$

is a Wasserstein gradient flow of the Dirichlet energy functional:

$$\mathcal{E}_v(u) := \int_{\mathbb{R}} (\partial_x(u(x,t))^a))^2 dx. \tag{196}$$

Transforming into Lagrangian coordinates $x = G(\omega)$ gives

$$\mathcal{E}_v(u) := \int_{\mathbb{R}} \left(\frac{\partial}{\partial x} \left(\frac{1}{g(\omega)} \right)^a \right)^2 dx = \int_0^M \left(-a \frac{\partial_\omega g(\omega)}{(g(\omega))^{a+1}} \frac{1}{g(\omega)} \right)^2 \frac{\partial x}{\partial \omega} d\omega = a^2 \int_0^M \frac{(\partial_\omega g(\omega))^2}{(g(\omega))^{2a+3}} d\omega.$$

Then, as we applied for the DLSS equation, the spatial discrete form of this gives us (see Appendix E for workings)

$$\mathcal{E}_v(u) = \frac{a^2}{2(a+1)} \sum_{k=1}^{n} \mathbf{F}_k^v[\mathbf{g}], \quad \mathbf{F}_k^v[\mathbf{g}] := \frac{1}{\delta_k}(g_k - g_{k-1}) \left(\frac{1}{g_{k-1}^{2(a+1)}} - \frac{1}{g_k^{2(a+1)}} \right). \tag{197}$$

Remark 7.1. DLSS and Thin Film equations are obtained as special cases, if $a = \frac{1}{2}$ and $a = 1$ respectively.

7.3.4 Fourth Order Nonlinear Equation (36)

The process is repated for the nonlinear equation (36).

We recall the fourth order PDE, from [30, Thm. 3.11]:

$$\partial_t u = -\partial_x \left(u \partial_{xx} \left(\frac{\partial_x u(x,t)}{u^2} \right) \right), \tag{198}$$

which is a Wasserstein gradient flow of another energy functional:

$$\begin{aligned}
\mathcal{E}_f(u) :=& \frac{1}{2} \int_{\mathbb{R}} (\partial_x \log(u(x,t)))^2 dx = \frac{1}{2} \int_{\mathbb{R}} \left(\frac{\partial}{\partial x} \left(\log \left(\frac{1}{g(\omega)} \right) \right) \right)^2 dx \\
=& \frac{1}{2} \int_{\mathbb{R}} \left(\partial_\omega \left(\log \left(\frac{1}{g(\omega)} \right) \right) \frac{\partial \omega}{\partial x} \right)^2 dx \\
=& \frac{1}{2} \int_0^M \left(-\frac{\partial_\omega g(\omega)}{(g(\omega))^2} \right)^2 \frac{\partial x}{\partial \omega} d\omega = \frac{1}{2} \int_0^M \frac{(\partial_\omega g(\omega))^2}{(g(\omega))^3} d\omega.
\end{aligned}$$

Transforming into Lagrangian coordinates $x = G(\omega)$ gives us (see Appendix E for workings)

$$\mathcal{E}_f(u) = \frac{1}{4} \sum_{k=1}^{n} \mathbf{F}_k^f[\mathbf{g}], \qquad \mathbf{F}_k^f[\mathbf{g}] := \frac{1}{\delta_k}(g_k - g_{k-1}) \left(\frac{1}{g_{k-1}^2} - \frac{1}{g_k^2} \right). \tag{199}$$

7.4 Fully Discrete Euler-Lagrange Equations for BDF Schemes

We summarise the fully discrete Euler-Lagrange equations for BDF1 to 6 schemes, i.e. BDF1 is already seen by Duering and Matthes et al. [17], but we've adapted here for BDF1 to 6 schemes, easily aligned from the schemes, shown in Section 4. The weight vector $\mathbf{g}$ at new time point $\mathbf{g}^n$ minimises ψ^τ for each scheme.

The second part introduces the Lagrangian functional with the mass constraint, where we start to construct the ingredients needed to numerically construct the weight vector minimising L^τ under constraint for our new time point via the Newton's method. We will briefly summarise this procedure for each scheme in the upcoming subsections.

Appendix F-G provides the workings for $\mathbf{G}_k$ for each equation.

With the mass constraint $\sum_{k=1}^{n} \Delta_k g_k = 1$, a Lagrange multiplier λ and its associated Lagrangian functional L^τ are introduced with $\Psi^\tau : \mathbb{G}_M^n \times \mathbb{G}_M^n$ such that (we derived these schemes in the Euclidean case back in Section 4.2):

The inductive scheme in the Lagrangian case for the BDFi scheme is [17, Sect. 2.6]

$$\mathbf{g}^n \in \underset{\mathbf{g}\in\mathbb{G}_M^n}{\operatorname{argmin}} \Psi_i^\tau(\mathbf{g}^{n-1}, \dots, \mathbf{g}^{n-i}; \mathbf{g}),$$

such that the functional Ψ_i^τ is defined as

$$\textbf{BDF1 Scheme}: \Psi^\tau(\mathbf{g}^{n-1}; \mathbf{g}) := \frac{1}{2\tau} \sum_{j,k=1}^{n} a_{j,k}(g_j - g_j^{n-1})(g_k - g_k^{n-1}) + \frac{1}{\alpha} \sum_{k=1}^{n} \mathbf{F}_k[\mathbf{g}]. \tag{200a}$$

BDF2 Scheme :

$$\Psi^\tau(\mathbf{g}^{n-1}, \mathbf{g}^{n-2}; \mathbf{g}) := \frac{1}{\tau} \sum_{j,k=1}^{n} a_{j,k}(g_j - g_j^{n-1})(g_k - g_k^{n-1}) \tag{200b}$$
$$- \frac{1}{4\tau} \sum_{j,k=1}^{n} a_{j,k}(g_j - g_j^{n-2})(g_k - g_k^{n-2}) + \frac{1}{\alpha} \sum_{k=1}^{n} \mathbf{F}_k[\mathbf{g}].$$

BDF3 Scheme :

$$\Psi^\tau(\mathbf{g}^{n-1}, \mathbf{g}^{n-2}, \mathbf{g}^{n-3}; \mathbf{g}) := \frac{3}{2\tau} \sum_{j,k=1}^{n} a_{j,k}(g_j - g_j^{n-1})(g_k - g_k^{n-1}) \tag{200c}$$
$$- \frac{3}{4\tau} \sum_{j,k=1}^{n} a_{j,k}(g_j - g_j^{n-2})(g_k - g_k^{n-2}) + \frac{1}{6\tau} \sum_{j,k=1}^{n} a_{j,k}(g_j - g_j^{n-3})(g_k - g_k^{n-3}) + \frac{1}{\alpha} \sum_{k=1}^{n} \mathbf{F}_k[\mathbf{g}].$$

BDF4 Scheme :

$$\Psi^\tau(\mathbf{g}^{n-1}, \mathbf{g}^{n-2}, \mathbf{g}^{n-3}, \mathbf{g}^{n-4}; \mathbf{g})$$
$$:= \frac{2}{\tau} \sum_{j,k=1}^{n} a_{j,k}(g_j - g_j^{n-1})(g_k - g_k^{n-1}) - \frac{3}{2\tau} \sum_{j,k=1}^{n} a_{j,k}(g_j - g_j^{n-2})(g_k - g_k^{n-2}) \tag{200d}$$
$$+ \frac{2}{3\tau} \sum_{j,k=1}^{n} a_{j,k}(g_j - g_j^{n-3})(g_k - g_k^{n-3}) - \frac{1}{8\tau} \sum_{j,k=1}^{n} a_{j,k}(g_j - g_j^{n-4})(g_k - g_k^{n-4}) + \frac{1}{\alpha} \sum_{k=1}^{n} \mathbf{F}_k[\mathbf{g}].$$

BDF5 Scheme :

$$\Psi^\tau(\mathbf{g}^{n-1}, \mathbf{g}^{n-2}, \mathbf{g}^{n-3}, \mathbf{g}^{n-4}, \mathbf{g}^{n-5}; \mathbf{g})$$
$$:= \frac{5}{2\tau} \sum_{j,k=1}^{n} a_{j,k}(g_j - g_j^{n-1})(g_k - g_k^{n-1}) - \frac{5}{2\tau} \sum_{j,k=1}^{n} a_{j,k}(g_j - g_j^{n-2})(g_k - g_k^{n-2}) \tag{200e}$$
$$+ \frac{5}{3\tau} \sum_{j,k=1}^{n} a_{j,k}(g_j - g_j^{n-3})(g_k - g_k^{n-3}) - \frac{5}{8\tau} \sum_{j,k=1}^{n} a_{j,k}(g_j - g_j^{n-4})(g_k - g_k^{n-4})$$
$$+ \frac{1}{10\tau} \sum_{j,k=1}^{n} a_{j,k}(g_j - g_j^{n-5})(g_k - g_k^{n-5}) + \frac{1}{\alpha} \sum_{k=1}^{n} \mathbf{F}_k[\mathbf{g}].$$

BDF6 Scheme :

$$\begin{aligned}
&\Psi^\tau(\mathbf{g}^{n-1},\mathbf{g}^{n-2},\mathbf{g}^{n-3},\mathbf{g}^{n-4},\mathbf{g}^{n-5},\mathbf{g}^{n-6};\mathbf{g})\\
:=&\frac{3}{\tau}\sum_{j,k=1}^{n} a_{j,k}(g_j-g_j^{n-1})(g_k-g_k^{n-1})-\frac{15}{4\tau}\sum_{j,k=1}^{n} a_{j,k}(g_j-g_j^{n-2})(g_k-g_k^{n-2}) \qquad (200f)\\
&+\frac{10}{3\tau}\sum_{j,k=1}^{n} a_{j,k}(g_j-g_j^{n-3})(g_k-g_k^{n-3})-\frac{15}{8\tau}\sum_{j,k=1}^{n} a_{j,k}(g_j-g_j^{n-4})(g_k-g_k^{n-4})\\
&+\frac{3}{5\tau}\sum_{j,k=1}^{n} a_{j,k}(g_j-g_j^{n-5})(g_k-g_k^{n-5})-\frac{1}{12\tau}\sum_{j,k=1}^{n} a_{j,k}(g_j-g_j^{n-6})(g_k-g_k^{n-6})+\frac{1}{\alpha}\sum_{k=1}^{n}\mathbf{F}_k[\mathbf{g}],
\end{aligned}$$

where $\alpha=6$ for the DLSS equation, $\alpha=8$ for the Thin Film equation, $\alpha=\dfrac{2(a+1)}{a^2}$ for equation (35) and $\alpha=4$ for equation (36).

Now (200) is minimised subject to the mass constraint [17, Def. 2.4]. To this end we define the Lagrangian functional:

$$L^\tau(\mathbf{g}^*,\mathbf{g},\lambda):=\Psi^\tau(\mathbf{g}^*;\mathbf{g})-\lambda\left(1-\sum_{k=1}^{n}\Delta_k g_k\right).$$

The minimisers $\mathbf{g}$ are given by its critical points $(\mathbf{g},\lambda)$, by "classical theory of variations". In other words, we set $\mathbf{G}_k=0$ where $\mathbf{G}_k:=\dfrac{\partial L^\tau}{\partial g_k}$. From this, we require the identities of $\dfrac{\partial \mathbf{F}_k}{\partial g_k}$ and $\dfrac{\partial \mathbf{F}_{k+1}}{\partial g_k}$, varying for each of our equations.

With higher order BDF schemes carrying additional intermediate steps, the $\mathbf{G}_k$ results for each scheme are given in Appendix F.

Furthermore, we also need the entries $\dfrac{\partial^2\mathbf{F}_k[\mathbf{g}]}{\partial g_k^2}$, $\dfrac{\partial^2\mathbf{F}_k[\mathbf{g}]}{\partial g_k\partial g_{k-1}}$, $\dfrac{\partial^2\mathbf{F}_{k+1}[\mathbf{g}]}{\partial g_k^2}$ and $\dfrac{\partial^2\mathbf{F}_{k+1}[\mathbf{g}]}{\partial g_k\partial g_{k+1}}$ for the Jacobian matrix $(\mathbf{H}_{j,k})_{j,k=1,..,n+1}$.

We now summarise these for each equation (see Appendix G for workings):

- **DLSS equation**: The resulting $\dfrac{\partial \mathbf{F}_k^d[\mathbf{g}]}{\partial g_k}$ and $\dfrac{\partial \mathbf{F}_{k+1}^d[\mathbf{g}]}{\partial g_k}$ are given as in [17, p. 11] from the DLSS equation:

$$\begin{cases}\dfrac{\partial \mathbf{F}_k^d[\mathbf{g}]}{\partial g_k}:=\dfrac{1}{\delta_k}\left(\dfrac{2}{3g_k^3}+\dfrac{1}{3g_{k-1}^3}-\dfrac{g_{k-1}}{g_k^4}\right),\\[2ex] \dfrac{\partial \mathbf{F}_{k+1}^d[\mathbf{g}]}{\partial g_{k+1}}:=\dfrac{1}{\delta_k}\left(\dfrac{2}{3g_k^3}+\dfrac{1}{3g_{k+1}^3}-\dfrac{g_{k+1}}{g_k^4}\right).\end{cases}$$

We take the entries again from the end of [17, p. 11] for the DLSS equation i.e. for $\dfrac{\partial^2\mathbf{F}_k^d[\mathbf{g}]}{\partial g_k^2}$, $\dfrac{\partial^2\mathbf{F}_k^d[\mathbf{g}]}{\partial g_k\partial g_{k\pm1}}$, $\dfrac{\partial^2\mathbf{F}_{k+1}^d[\mathbf{g}]}{\partial g_k^2}$ and $\dfrac{\partial^2\mathbf{F}_{k+1}^d[\mathbf{g}]}{\partial g_k\partial g_{k\pm1}}$, for the Hessian matrix entries:

$$\left\{\begin{array}{ll}\dfrac{\partial^2\mathbf{F}_k^d[\mathbf{g}]}{\partial g_k^2}:=\dfrac{2}{\delta_k}\left(\dfrac{2g_{k-1}}{g_k^5}-\dfrac{1}{g_k^4}\right), & \dfrac{\partial^2\mathbf{F}_{k+1}^d[\mathbf{g}]}{\partial g_k^2}:=\dfrac{2}{\delta_{k+1}}\left(\dfrac{2g_{k+1}}{g_k^5}-\dfrac{1}{g_k^4}\right),\\[2ex] \dfrac{\partial^2\mathbf{F}_k^d[\mathbf{g}]}{\partial g_k g_{k-1}}:=-\dfrac{1}{\delta_k}\left(\dfrac{1}{g_{k-1}^4}+\dfrac{1}{g_k^4}\right), & \dfrac{\partial^2\mathbf{F}_{k+1}^d[\mathbf{g}]}{\partial g_k\partial g_{k+1}}:=-\dfrac{1}{\delta_{k+1}}\left(\dfrac{1}{g_{k+1}^4}+\dfrac{1}{g_k^4}\right).\end{array}\right\} \qquad (201)$$

- **Thin Film equation**: For the Lagrangian critical point $\mathbf{G}_k$, when $k = 1, \ldots, n$:

$$\begin{cases} \dfrac{\partial \mathbf{F}_k^t[\mathbf{g}]}{\partial g_k} := \dfrac{1}{\delta_k}\left(\dfrac{3}{g_k^4} + \dfrac{1}{g_{k-1}^4} - \dfrac{4g_{k-1}}{g_k^5}\right), \\ \dfrac{\partial \mathbf{F}_{k+1}^t[\mathbf{g}]}{\partial g_k} := \dfrac{1}{\delta_{k+1}}\left(\dfrac{3}{g_k^4} + \dfrac{1}{g_{k+1}^4} - \dfrac{4g_{k+1}}{g_k^5}\right), \end{cases}$$

and we have for the Hessian matrix entries, for $k = 1, \ldots, n$:

$$\left\{ \begin{array}{ll} \dfrac{\partial^2 \mathbf{F}_k^t[\mathbf{g}]}{\partial g_k^2} := \dfrac{1}{\delta_k}\left(\dfrac{20g_{k-1}}{g_k^6} - \dfrac{12}{g_k^5}\right), & \dfrac{\partial^2 \mathbf{F}_{k+1}^t[\mathbf{g}]}{\partial g_k^2} := \dfrac{1}{\delta_{k+1}}\left(\dfrac{20g_{k+1}}{g_k^6} - \dfrac{12}{g_k^5}\right), \\ \dfrac{\partial^2 \mathbf{F}_k^t[\mathbf{g}]}{\partial g_k g_{k-1}} := -\dfrac{4}{\delta_k}\left(\dfrac{1}{g_{k-1}^5} + \dfrac{1}{g_k^5}\right), & \dfrac{\partial^2 \mathbf{F}_{k+1}^t[\mathbf{g}]}{\partial g_k \partial g_{k+1}} := -\dfrac{4}{\delta_{k+1}}\left(\dfrac{1}{g_{k+1}^5} + \dfrac{1}{g_k^5}\right). \end{array} \right\} \tag{202}$$

- **Nonlinear fourth order equations (196)**: For the Lagrangian critical point $\mathbf{G}_k$, when $k = 1, \ldots, n$:

$$\begin{cases} \dfrac{\partial \mathbf{F}_k^v[\mathbf{g}]}{\partial g_k} := \dfrac{1}{\delta_k}\left(\dfrac{2a+1}{g_k^{2(a+1)}} + \dfrac{1}{g_{k-1}^{2(a+1)}} - \dfrac{2(a+1)g_{k-1}}{g_k^{2a+3}}\right), \\ \dfrac{\partial \mathbf{F}_{k+1}^t[\mathbf{g}]}{\partial g_k} := \dfrac{1}{\delta_{k+1}}\left(\dfrac{2a+1}{g_k^{2(a+1)}} + \dfrac{1}{g_{k+1}^{2(a+1)}} - \dfrac{2(a+1)g_{k+1}}{g_k^{2a+3}}\right), \end{cases}$$

and for the Hessian matrix entries, for $k = 1, \ldots, n$:

$$\begin{cases} \dfrac{\partial^2 \mathbf{F}_k^v[\mathbf{g}]}{\partial g_k^2} := \dfrac{1}{\delta_k}\left(\dfrac{2(a+1)(2a+3)g_{k-1}}{g_k^{2(a+2)}} - \dfrac{2(a+1)(2a+1)}{g_k^{2a+3}}\right), \\ \dfrac{\partial^2 \mathbf{F}_{k+1}^t[\mathbf{g}]}{\partial g_k^2} := \dfrac{1}{\delta_{k+1}}\left(\dfrac{2(a+1)(2a+3)g_{k+1}}{g_k^{2(a+2)}} - \dfrac{2(a+1)(2a+1)}{g_k^{2a+3}}\right), \\ \dfrac{\partial^2 \mathbf{F}_k^t[\mathbf{g}]}{\partial g_k g_{k-1}} := -\dfrac{2(a+1)}{\delta_k}\left(\dfrac{1}{g_{k-1}^{2a+3}} + \dfrac{1}{g_k^{2a+3}}\right), \\ \dfrac{\partial^2 \mathbf{F}_{k+1}^t[\mathbf{g}]}{\partial g_k \partial g_{k+1}} := -\dfrac{2(a+1)}{\delta_{k+1}}\left(\dfrac{1}{g_{k+1}^{2a+3}} + \dfrac{1}{g_k^{2a+3}}\right). \end{cases} \tag{203}$$

- **Nonlinear fourth order equation (198)**: For the Lagrangian critical point $\mathbf{G}_k$, when $k = 1, \ldots, n$:

$$\begin{cases} \dfrac{\partial \mathbf{F}_k^f[\mathbf{g}]}{\partial g_k} := \dfrac{1}{\delta_k}\left(\dfrac{1}{g_k^2} + \dfrac{1}{g_{k-1}^2} - \dfrac{2g_{k-1}}{g_k^3}\right), \\ \dfrac{\partial \mathbf{F}_{k+1}^f[\mathbf{g}]}{\partial g_k} := \dfrac{1}{\delta_{k+1}}\left(\dfrac{1}{g_k^2} + \dfrac{1}{g_{k+1}^2} - \dfrac{2g_{k+1}}{g_k^3}\right), \end{cases}$$

and for the Hessian matrix entries, for $k = 1, \ldots, n$:

$$\left\{ \begin{array}{ll} \dfrac{\partial^2 \mathbf{F}_k^f[\mathbf{g}]}{\partial g_k^2} := \dfrac{1}{\delta_k}\left(\dfrac{6g_{k-1}}{g_k^4} - \dfrac{2}{g_k^3}\right), & \dfrac{\partial^2 \mathbf{F}_{k+1}^f[\mathbf{g}]}{\partial g_k^2} := \dfrac{1}{\delta_{k+1}}\left(\dfrac{6g_{k+1}}{g_k^4} - \dfrac{2}{g_k^3}\right), \\ \dfrac{\partial^2 \mathbf{F}_k^f[\mathbf{g}]}{\partial g_k \partial g_{k-1}} := -\dfrac{2}{\delta_k}\left(\dfrac{1}{g_{k-1}^3} + \dfrac{1}{g_k^3}\right), & \dfrac{\partial^2 \mathbf{F}_{k+1}^f[\mathbf{g}]}{\partial g_k \partial g_{k+1}} := -\dfrac{2}{\delta_{k+1}}\left(\dfrac{1}{g_{k+1}^3} + \dfrac{1}{g_k^3}\right). \end{array} \right\} \tag{204}$$

Scheme	BDF1	BDF2	BDF3	BDF4	BDF5	BDF6
β	1	$\frac{3}{2}$	$\frac{11}{6}$	$\frac{25}{12}$	$\frac{137}{60}$	$\frac{147}{60}$

Table 1: The coefficient β of the matrix entries $a_{j,k}$ from the Hessian matrices, dependent on each BDF scheme.

In fact, the zeros of $\mathbf{G}_k$ are found numerically by Newton's method. Hence, we compute the entries of the Hessian $(n+1)\times(n+1)$ matrix $(H_{j,k})$ defined as $\mathbf{H}_{j,k}$ which are

$$\begin{cases} \mathbf{H}_{k,k} := \dfrac{\beta}{\tau}a_{k,k} + \dfrac{1}{\alpha}\dfrac{\partial^2\mathbf{F}_k[\mathbf{g}]}{\partial g_k^2} + \dfrac{1}{\alpha}\dfrac{\partial^2\mathbf{F}_{k+1}[\mathbf{g}]}{\partial g_k^2} : \; j = k, \\ \mathbf{H}_{k-1,k} := \dfrac{\beta}{\tau}a_{k-1,k} + \dfrac{1}{\alpha}\dfrac{\partial^2\mathbf{F}_k[\mathbf{g}]}{\partial g_k \partial g_{k-1}} : \; j = k-1, \\ \mathbf{H}_{k+1,k} := \dfrac{\beta}{\tau}a_{k,k+1} + \dfrac{1}{\alpha}\dfrac{\partial^2\mathbf{F}_{k+1}[\mathbf{g}]}{\partial g_k \partial g_{k+1}} : \; j = k+1, \\ \mathbf{H}_{j,k} := \dfrac{\beta}{\tau}a_{j,k} : \quad \text{otherwise.} \end{cases} \tag{205}$$

where $\alpha = 6$ for the DLSS equation, $\alpha = 8$ for the Thin Film equation, $\alpha = \dfrac{2(a+1)}{a^2}$ for equation (35) and $\alpha = 4$ for equation (36). The partial derivative entries are given from [17, p. 11] for the DLSS equation, system (202) for the Thin Film equation, system (203) for equation (35) and system (204) for equation (36). The actual entries are given in Appendix G.

Furthermore, the values of β, dependent on the BDF scheme, are given in Table 1.

7.5 Newton's Method for BDF Schemes

As shown in [17, Sect. 2.7] the Newton's method for the BDF schemes are briefed.

The Newton's method is an iterative procedure for solving the root of a function. In our case, we are looking to solve the finite dimensional minimising movement scheme with constraint (190) i.e. the aim is to apply the iterative method to approximate the weight $\mathbf{g}^n$ at time t_τ^n, see [17, Sect. 2.7] for the approach. For the weight vector $\mathbf{g} := (g_1, g_2, \dots, g_n) \in \mathbb{R}^n$, we solve

$$(\mathbf{g}_\tau^{(s)}, \lambda_\tau^{(s)}) = (\mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}) - (\mathbf{H}[\mathbf{g}_\tau^{n-1}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}])^{-1}\mathbf{G}[\mathbf{g}_\tau^{n-1}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}], \tag{206}$$

where $\mathbf{g}_\tau^{(s)} = \mathbf{g}_\tau^{n-1}$ when $s = 0$. We iterate on s continuously until $||(\delta\mathbf{g}_\tau^{(s)}, \delta\lambda^{(s)})||$, such that

$$\mathbf{g}_\tau^{(s)} = \mathbf{g}_\tau^{(s-1)} + \delta\mathbf{g}_\tau^{(s)}, \quad \lambda_\tau^{(s)} = \lambda_\tau^{(s-1)} + \delta\lambda_\tau^{(s)},$$

is sufficiently small, with the resulting $\mathbf{g}_\tau^{(s)}$ being $\mathbf{g}_\tau^n$. Obviously, the same procedure is repeated for each time step up to the terminal time i.e. starting from $\mathbf{g}_\tau^{(0)} := \mathbf{g}_\tau^{n-1}$, the Newton's iteration eventually gives $\mathbf{g}_\tau^{(s)} := \mathbf{g}_\tau^n$ and so on.

Note that the above is for the BDF1 case. The procedure is easily adapted for BDF2 and so on to BDF6. That is

- **BDF2 Scheme**: The BDF1 computation is first given for earlier time steps $\mathbf{g}_\tau^{n-2} \to \mathbf{g}_\tau^{n-1}$ then

$$\begin{aligned}(\mathbf{g}_\tau^{(s)}, \lambda_\tau^{(s)}) =&(\mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}) \\ &- (\mathbf{H}[\mathbf{g}_\tau^{n-2}, \mathbf{g}_\tau^{n-1}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}])^{-1}\mathbf{G}[\mathbf{g}_\tau^{n-2}, \mathbf{g}_\tau^{n-1}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}].\end{aligned} \tag{207}$$

- **BDF3 Scheme**: The BDF1 and BDF2 computations are given i.e. $\mathbf{g}_\tau^{n-3} \to \mathbf{g}_\tau^{n-2}$ and $\mathbf{g}_\tau^{n-2} \to \mathbf{g}_\tau^{n-1}$, respectively are first given then

$$\begin{aligned}(\mathbf{g}_\tau^{(s)}, \lambda_\tau^{(s)}) =&(\mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}) \\ &- (\mathbf{H}[\mathbf{g}_\tau^{n-3}, \mathbf{g}_\tau^{n-2}, \mathbf{g}_\tau^{n-1}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}])^{-1}\mathbf{G}[\mathbf{g}_\tau^{n-3}, \mathbf{g}_\tau^{n-2}, \mathbf{g}_\tau^{n-1}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}].\end{aligned} \tag{208}$$

- **BDF4 Scheme**: The BDF1 to 3 computations are given for earlier time steps i.e. $\mathbf{g}_\tau^{n-4} \to \mathbf{g}_\tau^{n-3}$, $\mathbf{g}_\tau^{n-3} \to \mathbf{g}_\tau^{n-2}$ and $\mathbf{g}_\tau^{n-2} \to \mathbf{g}_\tau^{n-1}$, respectively are first given then

$$\begin{aligned}&(\mathbf{g}_\tau^{(s)}, \lambda_\tau^{(s)}) - (\mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}) \\ =& - (\mathbf{H}[\mathbf{g}_\tau^{n-4}, \mathbf{g}_\tau^{n-3}, \mathbf{g}_\tau^{n-2}, \mathbf{g}_\tau^{n-1}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}])^{-1}\mathbf{G}[\mathbf{g}_\tau^{n-4}, \mathbf{g}_\tau^{n-3}, \mathbf{g}_\tau^{n-2}, \mathbf{g}_\tau^{n-1}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}].\end{aligned} \tag{209}$$

- **BDF5 Scheme**: The BDF1 to 4 computations are given for earlier time steps i.e. $\mathbf{g}_\tau^{n-5} \to \mathbf{g}_\tau^{n-4}$, $\mathbf{g}_\tau^{n-4} \to \mathbf{g}_\tau^{n-3}$, $\mathbf{g}_\tau^{n-3} \to \mathbf{g}_\tau^{n-2}$ and $\mathbf{g}_\tau^{n-2} \to \mathbf{g}_\tau^{n-1}$, respectively are first given then

$$\begin{aligned}&(\mathbf{g}_\tau^{(s)}, \lambda_\tau^{(s)}) - (\mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}) \\ =& - (\mathbf{H}[\mathbf{g}_\tau^{n-5}, \mathbf{g}_\tau^{n-4} \dots, \mathbf{g}_\tau^{n-1}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}])^{-1}\mathbf{G}[\mathbf{g}_\tau^{n-5}, \mathbf{g}_\tau^{n-4}, \dots, \mathbf{g}_\tau^{n-1}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}].\end{aligned} \tag{210}$$

- **BDF6 Scheme**: The BDF1 to 5 computations are given for earlier time steps i.e. $\mathbf{g}_\tau^{n-5} \to \mathbf{g}_\tau^{n-4}$, $\mathbf{g}_\tau^{n-4} \to \mathbf{g}_\tau^{n-3}$, $\mathbf{g}_\tau^{n-3} \to \mathbf{g}_\tau^{n-2}$ and $\mathbf{g}_\tau^{n-2} \to \mathbf{g}_\tau^{n-1}$, respectively are first given then

$$\begin{aligned}&(\mathbf{g}_\tau^{(s)}, \lambda_\tau^{(s)}) - (\mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}) \\ =& - (\mathbf{H}[\mathbf{g}_\tau^{n-6}, \mathbf{g}_\tau^{n-5}, \dots, \mathbf{g}_\tau^{n-1}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}])^{-1}\mathbf{G}[\mathbf{g}_\tau^{n-6}, \mathbf{g}_\tau^{n-5}, \dots, \mathbf{g}_\tau^{n-1}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}].\end{aligned} \tag{211}$$

7.6 Fully Discrete Euler-Lagrange Equations for the Two Stage Runge-Kutta (DIRK2) Scheme

The process is as in Section 7.4 but for the DIRK2 scheme. We shall work with the DIRK2 scheme (64) from Section 4.4. Firstly, we work with the scheme for the earlier intermediate step u_τ^{n+a-1}.

By applying [17, Lem. 2.3] we have an almost identical inductive scheme in terms of $\mathbf{g}^{n-1}$, for stage one, apart from the fact that $\Psi^\tau(\mathbf{g}^{n-1}, \mathbf{g})$, from (200) is of the form $\frac{1}{a}\mathcal{W}_2[u^*, u]^2 + \mathcal{E}(u)$, due to the time step size being just $1/a$ of the original.

From [17, Sect. 2.6], the inductive scheme in the Lagrangian case for each stage is

$$\textbf{Stage One}: \mathbf{g}^{n+a-1} \in \underset{\mathbf{g}\in\mathbb{G}^n_M}{\operatorname{argmin}} \Psi^\tau_{2,1}(\mathbf{g}^{n-1}; \mathbf{g}); \tag{212a}$$

$$\Psi^\tau_{2,1}(\mathbf{g}^{n-1}; \mathbf{g}) := \frac{1}{2a\tau} \sum_{j,k=1}^{n} a_{j,k}(g_j - g_j^{n-1})(g_k - g_k^{n-1}) + \frac{1}{\alpha} \sum_{k=1}^{n} \mathbf{F}_k[\mathbf{g}],$$

$$\textbf{Stage Two}: \mathbf{g}^{n} \in \underset{\mathbf{g}\in\mathbb{G}^n_M}{\operatorname{argmin}} \Psi^\tau_{2,1}(\mathbf{g}^{n+a-1}, \mathbf{g}^{n-1}; \mathbf{g}); \tag{212b}$$

$$\Psi^\tau_{2,2}(\mathbf{g}^{n+a-1}, \mathbf{g}^{n-1}; \mathbf{g}) := -\frac{1-2a(1-a)}{2a(1-2a)\tau} \sum_{j,k=1}^{n} a_{j,k}(g_j - g_j^{n-1})(g_k - g_k^{n-1})$$
$$+ \frac{1}{2a(1-2a)\tau} \sum_{j,k=1}^{n} a_{j,k}(g_j - g_j^{n+a-1})(g_k - g_k^{n+a-1}) + \frac{1}{\alpha} \sum_{k=1}^{n} \mathbf{F}_k[\mathbf{g}],$$

where $\mathbf{F}_k[\mathbf{g}]$ and α is given, dependent on the PDE in question, with this being distinguished by (194) for the DLSS equation, (195) for the Thin Film equation, (197) for equation (35) and (199) for equation (36).

From [17, Sect. 2.6.2], we can consider the mass constraint $\int_0^M g(\omega)d\omega = 1$ in our minimiser $\mathbf{g}^n$ if we introduce a Lagrange multiplier λ and the Lagrangian functional

$$L^\tau(\mathbf{g}^{n-1}, \mathbf{g}; \lambda) := \Psi(\mathbf{g}^{n-1}, \mathbf{g}) - \lambda\Big(1 - \sum_{k=1}^{n} \Delta_k g_k\Big),$$

with the critical point $(\mathbf{g}, \lambda)$ satisfying $\mathbf{G}_k = \mathbf{G}_{n+1} = 0$ such that, for any intermediate time steps $a \in \Big(0.12, \frac{1-\sqrt{2}}{2}\Big)$:

- Stage one:

$$\mathbf{G}_k := \frac{1}{a\tau} \sum_{j=1}^{n} a_{j,k}(g_j - g_j^{n-1}) + \frac{1}{\alpha}\frac{\partial \mathbf{F}_k[\mathbf{g}]}{\partial g_k} + \frac{1}{\alpha}\frac{\partial \mathbf{F}_{k+1}[\mathbf{g}]}{\partial g_k} - \lambda\Delta_k,$$
$$\mathbf{G}_{n+1} := 1 - \sum_{k=1}^{n} \Delta_k g_k, \qquad k = 1, \ldots, n.$$

- Stage two:

$$\mathbf{G}_k := \frac{2(1-a)}{(1-2a)\tau}\left(\frac{1}{2a(1-a)} \sum_{j=1}^{n} a_{j,k}(g_j - g_j^{n+a-1}) - \frac{1-2a(1-a)}{2a(1-a)} \sum_{j=1}^{n} a_{j,k}(g_j - g_j^{n-1})\right)$$
$$+ \frac{1}{\alpha}\frac{\partial \mathbf{F}_k[\mathbf{g}]}{\partial g_k} + \frac{1}{\alpha}\frac{\partial \mathbf{F}_{k+1}[\mathbf{g}]}{\partial g_k} - \lambda\Delta_k,$$
$$\mathbf{G}_{n+1} := 1 - \sum_{k=1}^{n} \Delta_k g_k, \qquad k = 1, \ldots, n,$$

where $\alpha = 6$ for the DLSS Equation, $\alpha = 8$ for the Thin Film Equation, $\alpha = \frac{2(a+1)}{a^2}$, $(a \in \mathbb{R}\backslash\{0\})$ for equation (35) and $\alpha = 4$ for equation (36).

Remark 7.2. The notation $\mathbf{F}_k[\mathbf{g}]$, $\mathbf{F}_{k+1}[\mathbf{g}]$ above refers to either $\mathbf{F}_k^d[\mathbf{g}]$ (DLSS equation) or $\mathbf{F}_k^d[\mathbf{g}]$ (Thin Film equation).

The Newton's method can be introduced for finding an approximation to $\mathbf{G}[\mathbf{g}^{n-1};\mathbf{g},\lambda]$, achieved by the entries of the Hessian matrix $\mathbf{H}[\mathbf{g}^{n-1};\mathbf{g},\lambda]$, which is as (205), with $\mathbf{H}_{k,n+1}$ and $\mathbf{H}_{n+1,n+1}$, as stated in [17], but $\beta = \frac{1}{a}$ (stage one) or $\beta = \frac{1-a}{1-2a}$ (stage two).

Example 7.3. *By choosing $a = 1/4$ for the second order scheme from (9.2), the critical point of the Lagrangian for stage two is*

$$\mathbf{G}_k := \frac{3}{\tau}\left(\frac{8}{3}\sum_{j=1}^{n} a_{j,k}(g_j - g_j^{n-3/4}) - \frac{5}{3}\sum_{j=1}^{n} a_{j,k}(g_j - g_j^{n-1})\right) + \frac{1}{\alpha}\frac{\partial \mathbf{F}_k[\mathbf{g}]}{\partial g_k} + \frac{1}{\alpha}\frac{\partial \mathbf{F}_{k+1}[\mathbf{g}]}{\partial g_k} - \lambda\Delta_k, \quad k = 1,\dots,n.$$

7.7 Fully Discrete Euler-Lagrange Equation for the Three Stage Runge-Kutta (DIRK3) Scheme

The process is as in Section 7.4 and 7.6 but for the DIRK3 scheme, from Section 4.5.

7.7.1 Scheme One

We move on to working with additional intermediate time steps, using the scheme (82)-(84) from Section 4.5.1. Hence, in our case now, the finite dimensional schemes for each stage are [17, Sect. 2.6]

$$\textbf{Stage One}: \mathbf{g}^{n+c_1-1} \in \underset{\mathbf{g}\in\mathbb{G}_M^n}{\operatorname{argmin}} \Psi_{3,1}^{\tau}(\mathbf{g}^{n-1};\mathbf{g}); \tag{213a}$$

$$\Psi_{3,1}^{\tau}(\mathbf{g}^{n-1};\mathbf{g}) := \frac{1}{2c_1\tau}\sum_{j,k=1}^{n} a_{j,k}(g_j - g_j^{n-1})(g_k - g_k^{n-1}) + \frac{1}{\alpha}\sum_{k=1}^{n}\mathbf{F}_k[\mathbf{g}],$$

$$\textbf{Stage Two}: \mathbf{g}^{n+c_2-1} \in \underset{\mathbf{g}\in\mathbb{G}_M^n}{\operatorname{argmin}} \Psi_{3,2}^{\tau}(\mathbf{g}^{n+c_1-1},\mathbf{g}^{n-1};\mathbf{g}); \tag{213b}$$

$$\begin{aligned}
&\Psi_{3,2}^{\tau}(\mathbf{g}^{n+c_1-1},\mathbf{g}^{n-1};\mathbf{g}) \\
=:&\frac{c_1+c_2-4c_1c_2}{2c_1(6c_1^2c_2-4c_1c_2-c_1+c_2)\tau}\sum_{j,k=1}^{n} a_{j,k}(g_j - g_j^{n+c_1-1})(g_k - g_k^{n+c_1-1}) \\
&+\frac{c_1+c_2-4c_1c_2-2c_1(1-3c_1)(1-c_1)}{2c_1(6c_1^2c_2-4c_1c_2-c_1+c_2)\tau}\sum_{j,k=1}^{n} a_{j,k}(g_j - g_j^{n-1})(g_k - g_k^{n-1}) + \frac{1}{\alpha}\sum_{k=1}^{n}\mathbf{F}_k[\mathbf{g}],
\end{aligned}$$

$$\textbf{Stage Three}: \mathbf{g}^{n} \in \underset{\mathbf{g}\in\mathbb{G}_M^n}{\operatorname{argmin}} \Psi_{3,2}^{\tau}(\mathbf{g}^{n+c_2-1},\mathbf{g}^{n+c_1-1},\mathbf{g}^{n-1};\mathbf{g}); \tag{213c}$$

$$\Psi_{3,3}^{\tau}(\mathbf{g}^{n+c_2-1},\mathbf{g}^{n+c_1-1},\mathbf{g}^{n-1};\mathbf{g})$$

$$
\begin{aligned}
:=&\frac{c_1(1-c_1)(1-3c_1)^2}{(2c_1c_2(3c_1-2)-c_1+c_2)\left(3(c_1-2c_1c_2+c_2)-2\right)\tau}\sum_{j,k=1}^{n} a_{j,k}(g_j-g_j^{n+c_2-1})(g_k-g_k^{n+c_2-1})\\
&+\frac{y_7}{\tau}\sum_{j,k=1}^{n} a_{j,k}(g_j-g_j^{n+c_1-1})(g_k-g_k^{n+c_1-1})+\frac{y_8}{\tau}\sum_{j,k=1}^{n} a_{j,k}(g_j-g_j^{n-1})(g_k-g_k^{n-1})\\
&+\frac{1}{\alpha}\sum_{k=1}^{n}\mathbf{F}_k[\mathbf{g}],
\end{aligned}
$$

where y_7 and y_8 is given as (85) from Section 4.5 and $\mathbf{F}_k[\mathbf{g}]$ and α is given, dependent on the equation in question, with this being distinguished by (194) for the DLSS equation, (195) for the Thin Film equation, (197) for equation (35) and (199) for equation (36).

As explained earlier for already implemented schemes, the critical point of the Lagrangian is

- **Stage One**: For minimiser $u_\tau^{n+c_1-1}$ is

$$
\mathbf{G}_k := \frac{1}{c_1\tau}\sum_{j=1}^{n} a_{j,k}(g_j-g_j^{n-1})+\frac{1}{\alpha}\frac{\partial \mathbf{F}_k[\mathbf{g}]}{\partial g_k}+\frac{1}{\alpha}\frac{\partial \mathbf{F}_{k+1}[\mathbf{g}]}{\partial g_k}-\lambda\Delta_k, \quad (k=1,\dots,n).
$$

- **Stage Two**: For minimiser $u_\tau^{n+c_2-1}$ is

$$
\begin{aligned}
\mathbf{G}_k :=&\frac{c_1+c_2-4c_1c_2}{c_1[6c_1^2c_2-4c_1c_2-c_1+c_2]\tau}\sum_{j=1}^{n} a_{j,k}(g_j-g_j^{n+c_1-1})\\
&+\frac{4c_1c_2+2c_1(1-3c_1)(1-c_1)-c_1-c_2}{c_1(6c_1^2c_2-4c_1c_2-c_1+c_2)\tau}\sum_{j=1}^{n} a_{j,k}(g_j-g_j^{n-1})\\
&+\frac{1}{\alpha}\frac{\partial \mathbf{F}_k[\mathbf{g}]}{\partial g_k}+\frac{1}{\alpha}\frac{\partial \mathbf{F}_{k+1}[\mathbf{g}]}{\partial g_k}-\lambda\Delta_k, \quad (k=1,\dots,n).
\end{aligned}
$$

- **Stage Three**: For minimiser u_τ^{n} is

$$
\begin{aligned}
\mathbf{G}_k :=&\frac{2c_1(1-c_1)^2(1-3c_1)^2}{(6c_1^2c_2-4c_1c_2-c_1+c_2)\left(3(c_1-2c_1c_2+c_2)-2\right)(c_1-c_2)\tau}\sum_{j=1}^{n} a_{j,k}(g_j-g_j^{n+c_2-1})\\
&+\frac{2y_7}{\tau}\sum_{j=1}^{n} a_{j,k}(g_j-g_j^{n+c_1-1})+\frac{2y_8}{\tau}\sum_{j=1}^{n} a_{j,k}(g_j-g_j^{n-1})\\
&+\frac{1}{\alpha}\frac{\partial \mathbf{F}_k[\mathbf{g}]}{\partial g_k}+\frac{1}{\alpha}\frac{\partial \mathbf{F}_{k+1}[\mathbf{g}]}{\partial g_k}-\lambda\Delta_k, \quad (k=1,\dots,n),
\end{aligned}
$$

with the Newton's method carried out on the optimality condition after finding the following entries of the Hessian matrix $\mathbf{H}[\mathbf{g}^n;\mathbf{g},\lambda]$, again given as (205), where β is, for $k=1,\dots,n$:

$$
\textbf{Stage One}: \frac{1}{c_1}, \quad \textbf{Stage Two}: \frac{2(1-3c_1)(1-c_1)}{6c_1^2c_2-4c_1c_2-c_1+c_2}, \quad \textbf{Stage Three}: \frac{6c_1(1-c_1)(1-c_2)}{3(c_1-2c_1c_2+c_2)-2}.
$$

7.7.2 Scheme Two

The process is as in Section 7.4 and 7.6-7.7, but for the main example for DIRK3 which we use only for our numerical results.

Now for the second example of the DIRK3 scheme constructed as (95)-(97) from Section 4.5.2, of which this was formally published in [48] and easier to compute to our minimising movement scheme. Hence, from here now, the functionals $\Psi^{\tau}_{3,i}$ as part of the finite dimensional schemes for each stage i are

$$\textbf{Stage One}: \mathbf{g}^{n+\alpha_1-1} \in \underset{\mathbf{g}\in\mathbb{G}^n_M}{\operatorname{argmin}} \Psi^{\tau}_{3,1}(\mathbf{g}^{n-1};\mathbf{g}); \tag{214a}$$

$$\Psi^{\tau}_{3,1}(\mathbf{g}^{n-1};\mathbf{g}) := \frac{1}{2\alpha_1\tau}\sum_{j,k=1}^{n} a_{j,k}(g_j-g_j^{n-1})(g_k-g_k^{n-1}) + \frac{1}{\alpha}\sum_{k=1}^{n}\mathbf{F}_k[\mathbf{g}],$$

$$\textbf{Stage Two}: \mathbf{g}^{n+\alpha_2-1} \in \underset{\mathbf{g}\in\mathbb{G}^n_M}{\operatorname{argmin}} \Psi^{\tau}_{3,2}(\mathbf{g}^{n+\alpha_1-1},\mathbf{g}^{n-1};\mathbf{g}); \tag{214b}$$

$$\Psi^{\tau}_{3,2}(\mathbf{g}^{n+\alpha_1-1},\mathbf{g}^{n-1};\mathbf{g}) := \frac{\beta_1}{2\alpha_1^2\tau}\sum_{j,k=1}^{n} a_{j,k}(g_j-g_j^{n+\alpha_1-1})(g_k-g_k^{n+\alpha_1-1})$$
$$+\frac{\alpha_1-\beta_1}{2\alpha_1^2\tau}\sum_{j,k=1}^{n} a_{j,k}(g_j-g_j^{n-1})(g_k-g_k^{n-1}) + \frac{1}{\alpha}\sum_{k=1}^{n}\mathbf{F}_k[\mathbf{g}],$$

$$\textbf{Stage Three}: \mathbf{g}^{n} \in \underset{\mathbf{g}\in\mathbb{G}^n_M}{\operatorname{argmin}} \Psi^{\tau}_{3,3}(\mathbf{g}^{n+\alpha_2-1},\mathbf{g}^{n+\alpha_1-1},\mathbf{g}^{n-1};\mathbf{g}); \tag{214c}$$

$$\Psi^{\tau}_{3,3}(\mathbf{g}^{n+\alpha_2-1},\mathbf{g}^{n+\alpha_1-1},\mathbf{g}^{n-1};\mathbf{g}) := \frac{\beta_3}{2\alpha_1^2\tau}\sum_{j,k=1}^{n} a_{j,k}(g_j-g_j^{n+\alpha_2-1})(g_k-g_k^{n+\alpha_2-1})$$
$$+\frac{\alpha_1\beta_2-\beta_1\beta_3}{2\alpha_1^3\tau}\sum_{j,k=1}^{n} a_{j,k}(g_j-g_j^{n+\alpha_1-1})(g_k-g_k^{n+\alpha_1-1})$$
$$+\frac{\alpha_1^2-\alpha_1(\beta_2+\beta_3)+\beta_1\beta_3}{2\alpha_1^3\tau}\sum_{j,k=1}^{n} a_{j,k}(g_j-g_j^{n-1})(g_k-g_k^{n-1}) + \frac{1}{\alpha}\sum_{k=1}^{n}\mathbf{F}_k[\mathbf{g}].$$

As explained earlier for already implemented schemes, the critical points of the Lagrangian concerning the minimising movement schemes are as follows:

- **Stage One**: Minimiser $u_\tau^{n+\alpha_1-1}$ satisfying

$$\mathbf{G}_k := \frac{1}{\alpha_1\tau}\sum_{j=1}^{n} a_{j,k}(g_j-g_j^{n-1}) + \frac{1}{\alpha}\frac{\partial\mathbf{F}_k[\mathbf{g}]}{\partial g_k} + \frac{1}{\alpha}\frac{\partial\mathbf{F}_{k+1}[\mathbf{g}]}{\partial g_k} - \lambda\Delta_k, \quad (k=1,\dots,n),$$

- **Stage Two**: Minimiser $u_\tau^{n+\alpha_2-1}$ satisfying

$$\mathbf{G}_k := \frac{\beta_1}{\alpha_1^2\tau}\sum_{j=1}^{n} a_{j,k}(g_j-g_j^{n+\alpha_1-1}) + \frac{\alpha_1-\beta_1}{\alpha_1^2\tau}\sum_{j=1}^{n} a_{j,k}(g_j-g_j^{n-1})$$
$$+\frac{1}{\alpha}\frac{\partial\mathbf{F}_k[\mathbf{g}]}{\partial g_k} + \frac{1}{\alpha}\frac{\partial\mathbf{F}_{k+1}[\mathbf{g}]}{\partial g_k} - \lambda\Delta_k, \quad (k=1,\dots,n),$$

- **Stage Three**: Minimiser u_τ^n satisfying

$$\begin{aligned}\mathbf{G}_k :=& \frac{\beta_3}{\alpha_1^2\tau}\sum_{j=1}^n a_{j,k}(g_j - g_j^{n+\alpha_2-1}) + \frac{\alpha_1\beta_2-\beta_1\beta_3}{\alpha_1^3\tau}\sum_{j=1}^n a_{j,k}(g_j - g_j^{n+\alpha_1-1})\\ &+\frac{\alpha_1^2-\alpha_1(\beta_2+\beta_3)+\beta_1\beta_3}{\alpha_1^3\tau}\sum_{j=1}^n a_{j,k}(g_j-g_j^{n-1})\\ &+\frac{1}{\alpha}\frac{\partial \mathbf{F}_k[\mathbf{g}]}{\partial g_k}+\frac{1}{\alpha}\frac{\partial \mathbf{F}_{k+1}[\mathbf{g}]}{\partial g_k}-\lambda\Delta_k, \quad (k=1,\dots,n),\end{aligned}$$

with Newton's method carried out on the optimality condition after finding the following entries of the Hessian matrix $\mathbf{H}[\mathbf{g}^{n-1};\mathbf{g},\lambda]$, given again as (205), where $\beta=\frac{1}{\alpha_1}$ for all stages.

7.8 Fully Discrete Euler-Lagrange Equation for the Five Stage Runge-Kutta (DIRK5) Scheme

Last but not least, as in Sections 7.4 and 7.6-7.7 but for the DIRK5 scheme of fourth order.

Next, we move on to working with more additional intermediate time steps, using the scheme (101)-(105) from Section 4.6. Hence, in our case now, the Lagrangian functionals as part of the finite dimensional schemes for each stage are

$$\textbf{Stage One}: \Psi_{5,1}^\tau(\mathbf{g}^{n-1};\mathbf{g}) := \frac{2}{\tau}\sum_{j,k=1}^n a_{j,k}(g_j-g_j^{n-1})(g_k-g_k^{n-1}) + \frac{1}{\alpha}\sum_{k=1}^n \mathbf{F}_k[\mathbf{g}], \tag{215a}$$

$$\begin{aligned}&\textbf{Stage Two}: \Psi_{5,2}^\tau(\mathbf{g}^{n-3/4},\mathbf{g}^{n-1};\mathbf{g}) \qquad (215\text{b})\\ &:= \frac{4}{\tau}\sum_{j,k=1}^n a_{j,k}(g_j-g_j^{n-3/4})(g_k-g_k^{n-3/4}) - \frac{2}{\tau}\sum_{j,k=1}^n a_{j,k}(g_j-g_j^{n-1})(g_k-g_k^{n-1}) + \frac{1}{\alpha}\sum_{k=1}^n \mathbf{F}_k[\mathbf{g}],\end{aligned}$$

$$\begin{aligned}&\textbf{Stage Three}: \Psi_{5,3}^\tau(\mathbf{g}^{n-1/4},\mathbf{g}^{n-3/4},\mathbf{g}^{n-1};\mathbf{g}) \qquad (215\text{c})\\ &:= -\frac{8}{25\tau}\sum_{j,k=1}^n a_{j,k}(g_j-g_j^{n-1/4})(g_k-g_k^{n-1/4}) + \frac{84}{25\tau}\sum_{j,k=1}^n a_{j,k}(g_j-g_j^{n-3/4})(g_k-g_k^{n-3/4})\\ &-\frac{26}{25\tau}\sum_{j,k=1}^n a_{j,k}(g_j-g_j^{n-1})(g_k-g_k^{n-1}) + \frac{1}{\alpha}\sum_{k=1}^n \mathbf{F}_k[\mathbf{g}],\end{aligned}$$

$$\begin{aligned}&\textbf{Stage Four}: \Psi_{5,4}^\tau(\mathbf{g}^{n-9/20},\mathbf{g}^{n-1/4},\mathbf{g}^{n-3/4},\mathbf{g}^{n-1};\mathbf{g}) \qquad (215\text{d})\\ &:= \frac{15}{68\tau}\sum_{j,k=1}^n a_{j,k}(g_j-g_j^{n-9/20})(g_k-g_k^{n-9/20}) - \frac{25}{68\tau}\sum_{j,k=1}^n a_{j,k}(g_j-g_j^{n-1/4})(g_k-g_k^{n-1/4})\\ &+\frac{89}{34\tau}\sum_{j,k=1}^n a_{j,k}(g_j-g_j^{n-3/4})(g_k-g_k^{n-3/4}) - \frac{8}{17\tau}\sum_{j,k=1}^n a_{j,k}(g_j-g_j^{n-1})(g_k-g_k^{n-1}) + \frac{1}{\alpha}\sum_{k=1}^n \mathbf{F}_k[\mathbf{g}],\end{aligned}$$

$$\begin{aligned}&\textbf{Stage Five}: \Psi_{5,5}^\tau(\mathbf{g}^{n-1/2},\mathbf{g}^{n-9/20},\mathbf{g}^{n-1/4},\mathbf{g}^{n-3/4},\mathbf{g}^{n-1};\mathbf{g}) \qquad (215\text{e})\\ &:= -\frac{170}{3\tau}\sum_{j,k=1}^n a_{j,k}(g_j-g_j^{n-1/2})(g_k-g_k^{n-1/2}) + \frac{275}{4\tau}\sum_{j,k=1}^n a_{j,k}(g_j-g_j^{n-9/20})(g_k-g_k^{n-9/20})\end{aligned}$$

$$
-\frac{103}{12\tau}\sum_{j,k=1}^{n}a_{j,k}(g_j-g_j^{n-1/4})(g_k-g_k^{n-1/4})-\frac{37}{6\tau}\sum_{j,k=1}^{n}a_{j,k}(g_j-g_j^{n-3/4})(g_k-g_k^{n-3/4})
+\frac{14}{3\tau}\sum_{j,k=1}^{n}a_{j,k}(g_j-g_j^{n-1})(g_k-g_k^{n-1})+\frac{1}{\alpha}\sum_{k=1}^{n}\mathbf{F}_k[\mathbf{g}].
$$

As explained earlier for already implemented schemes, the critical points of the Lagrangian concerning the minimising movement schemes are as follows:

- **Stage One**: Minimiser $u_\tau^{n-3/4}$ satisfying

$$
\mathbf{G}_k := \frac{4}{\tau}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-1})+\frac{1}{\alpha}\frac{\partial\mathbf{F}_k[\mathbf{g}]}{\partial g_k}+\frac{1}{\alpha}\frac{\partial\mathbf{F}_{k+1}[\mathbf{g}]}{\partial g_k}-\lambda\Delta_k,\quad (k=1,\dots,n),
$$

- **Stage Two**: Minimiser $u_\tau^{n-1/4}$ satisfying

$$
\begin{aligned}
\mathbf{G}_k := \quad & \frac{8}{\tau}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-3/4})-\frac{4}{\tau}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-1})\\
&+\frac{1}{\alpha}\frac{\partial\mathbf{F}_k[\mathbf{g}]}{\partial g_k}+\frac{1}{\alpha}\frac{\partial\mathbf{F}_{k+1}[\mathbf{g}]}{\partial g_k}-\lambda\Delta_k,\quad (k=1,\dots,n),
\end{aligned}
$$

- **Stage Three**: Minimiser $u_\tau^{n-9/20}$ satisfying

$$
\begin{aligned}
\mathbf{G}_k :=& -\frac{16}{25\tau}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n+\alpha_2-1})+\frac{168}{25\tau}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n+\alpha_1-1})-\frac{52}{25\tau}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-1})\\
&+\frac{1}{\alpha}\frac{\partial\mathbf{F}_k[\mathbf{g}]}{\partial g_k}+\frac{1}{\alpha}\frac{\partial\mathbf{F}_{k+1}[\mathbf{g}]}{\partial g_k}-\lambda\Delta_k,\quad (k=1,\dots,n),
\end{aligned}
$$

- **Stage Four**: Minimiser $u_\tau^{n-1/2}$ satisfying

$$
\begin{aligned}
\mathbf{G}_k :=& \frac{15}{34\tau}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-9/20})-\frac{25}{34\tau}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-1/4})+\frac{89}{17\tau}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-3/4})\\
&-\frac{16}{17\tau}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-1})+\frac{1}{\alpha}\frac{\partial\mathbf{F}_k[\mathbf{g}]}{\partial g_k}+\frac{1}{\alpha}\frac{\partial\mathbf{F}_{k+1}[\mathbf{g}]}{\partial g_k}-\lambda\Delta_k,\quad (k=1,\dots,n),
\end{aligned}
$$

- **Stage Five**: Minimiser u_τ^{n+1} satisfying

$$
\begin{aligned}
\mathbf{G}_k :=& -\frac{340}{3\tau}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-1/2})+\frac{275}{2\tau}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-9/20})-\frac{103}{6\tau}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-1/4})\\
&-\frac{37}{3\tau}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-3/4})+\frac{28}{3\tau}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-1})\\
&+\frac{1}{\alpha}\frac{\partial\mathbf{F}_k[\mathbf{g}]}{\partial g_k}+\frac{1}{\alpha}\frac{\partial\mathbf{F}_{k+1}[\mathbf{g}]}{\partial g_k}-\lambda\Delta_k,\quad (k=1,\dots,n),
\end{aligned}
$$

with the Newton's method carried out on the optimality condition after finding the entries of the Hessian matrix $\mathbf{H}[\mathbf{g}^n;\mathbf{g},\lambda]$, from (205) for each scheme, where $\beta=4$ for all stages.

7.9 Newton's Method for DIRK Schemes

The process is as in Section 7.5, but for the DIRK schemes mentioned in Sections 7.6-7.8.

Before we outline the numerical results, here is an outline of the Newton's method, which was already applied for the BDF schemes, but now for the three DIRK schemes shown in this section, again see [17, Sect. 2.7] for the approach:

- **DIRK2 Scheme**: For each time step τ, we apply two iterative processes, one per stage.
  - The first iteration corresponds from stage one, constructing $\mathbf{g}_\tau^{n+a-1}$ dependent on $\mathbf{g}_\tau^{n-1}$, that is

    $$(\mathbf{g}_\tau^{(s)}, \lambda_\tau^{(s)}) = (\mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}) - (\mathbf{H}[\mathbf{g}_\tau^{n-1}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}])^{-1}\mathbf{G}[\mathbf{g}_\tau^{n-1}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}],$$

    where $\mathbf{g}_\tau^{(0)} := \mathbf{g}_\tau^{n-1}$ and the iteration eventually provides $\mathbf{g}_\tau^{(s)} := \mathbf{g}_\tau^{n+a-1}$.
  - Then the stage two Newton iteration provides the discrete solution $\mathbf{g}_\tau^{n}$ dependent on $\mathbf{g}_\tau^{n+a-1}$ and $\mathbf{g}_\tau^{n-1}$, that is

    $$\begin{aligned}(\mathbf{g}_\tau^{(s)}, \lambda_\tau^{(s)}) = \quad & (\mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}) \\ & - (\mathbf{H}[\mathbf{g}_\tau^{n-1}, \mathbf{g}_\tau^{n+a-1}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}])^{-1}\mathbf{G}[\mathbf{g}_\tau^{n-1}, \mathbf{g}_\tau^{n+a-1}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}],\end{aligned}$$

    where $\mathbf{g}_\tau^{(0)} := \mathbf{g}_\tau^{n+a-1}$ and the iteration eventually provides $\mathbf{g}_\tau^{(s)} := \mathbf{g}_\tau^{n}$.
- **DIRK3 Scheme**: For each time step τ, we apply three iterative processes, one per stage.
  - The first iteration corresponds from stage one, constructing $\mathbf{g}_\tau^{n+\alpha_1-1}$ dependent on $\mathbf{g}_\tau^{n-1}$, that is

    $$(\mathbf{g}_\tau^{(s)}, \lambda_\tau^{(s)}) = (\mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}) - (\mathbf{H}[\mathbf{g}_\tau^{n-1}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}])^{-1}\mathbf{G}[\mathbf{g}_\tau^{n-1}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}],$$

    where $\mathbf{g}_\tau^{(0)} := \mathbf{g}_\tau^{n-1}$ and the iteration eventually provides $\mathbf{g}_\tau^{(s)} := \mathbf{g}_\tau^{n+\alpha_1-1}$.
  - Then the stage two Newton iteration provides the discrete solution $\mathbf{g}_\tau^{n+\alpha_2-1}$ dependent on $\mathbf{g}_\tau^{n+\alpha_1-1}$ and $\mathbf{g}_\tau^{n-1}$, that is

    $$\begin{aligned}&(\mathbf{g}_\tau^{(s)}, \lambda_\tau^{(s)}) - (\mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}) \\ =& - (\mathbf{H}[\mathbf{g}_\tau^{n-1}, \mathbf{g}_\tau^{n+\alpha_1-1}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}])^{-1}\mathbf{G}[\mathbf{g}_\tau^{n-1}, \mathbf{g}_\tau^{n+\alpha_1-1}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}],\end{aligned}$$

    where $\mathbf{g}_\tau^{(0)} := \mathbf{g}_\tau^{n+\alpha_1-1}$ and the iteration eventually provides $\mathbf{g}_\tau^{(s)} := \mathbf{g}_\tau^{n+\alpha_2-1}$.
  - Finally the stage three Newton iteration provides the discrete solution $\mathbf{g}_\tau^{n}$ dependent on $\mathbf{g}_\tau^{n+\alpha_1-1}$, $\mathbf{g}_\tau^{n+\alpha_2-1}$ and $\mathbf{g}_\tau^{n-1}$, that is

    $$\begin{aligned}&(\mathbf{g}_\tau^{(s)}, \lambda_\tau^{(s)}) - (\mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}) \\ =& - (\mathbf{H}[\mathbf{g}_\tau^{n-1}, \mathbf{g}_\tau^{n+\alpha_1-1}, \mathbf{g}_\tau^{n+\alpha_2-1}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}])^{-1}\mathbf{G}[\mathbf{g}_\tau^{n-1}, \mathbf{g}_\tau^{n+\alpha_1-1}, \mathbf{g}_\tau^{n+\alpha_2-1}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}],\end{aligned}$$

    where $\mathbf{g}_\tau^{(0)} := \mathbf{g}_\tau^{n+\alpha_2-1}$ and the iteration eventually provides $\mathbf{g}_\tau^{(s)} := \mathbf{g}_\tau^{n}$.

- **DIRK5 Scheme**: For each time step τ, we apply five iterative processes, one per stage.
  - The first iteration corresponds from stage one, constructing $\mathbf{g}_\tau^{n-3/4}$ dependent on $\mathbf{g}_\tau^{n-1}$, that is

$$(\mathbf{g}_\tau^{(s)}, \lambda_\tau^{(s)}) = (\mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}) - (\mathbf{H}[\mathbf{g}_\tau^{n-1}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}])^{-1}\mathbf{G}[\mathbf{g}_\tau^{n-1}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}],$$

    where $\mathbf{g}_\tau^{(0)} := \mathbf{g}_\tau^{n-1}$ and the iteration eventually provides $\mathbf{g}_\tau^{(s)} := \mathbf{g}_\tau^{n-3/4}$.
  - Then the stage two Newton iteration provides the discrete solution $\mathbf{g}_\tau^{n-1/4}$ dependent on $\mathbf{g}_\tau^{n-3/4}$ and $\mathbf{g}_\tau^{n-1}$, that is

$$\begin{aligned}&(\mathbf{g}_\tau^{(s)}, \lambda_\tau^{(s)}) - (\mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)})\\ &= -(\mathbf{H}[\mathbf{g}_\tau^{n-1}, \mathbf{g}_\tau^{n-3/4}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}])^{-1}\mathbf{G}[\mathbf{g}_\tau^{n-1}, \mathbf{g}_\tau^{n-3/4}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}],\end{aligned}$$

    where $\mathbf{g}_\tau^{(0)} := \mathbf{g}_\tau^{n-3/4}$ and the iteration eventually provides $\mathbf{g}_\tau^{(s)} := \mathbf{g}_\tau^{n}$.
  - Then the stage three Newton iteration provides the discrete solution $\mathbf{g}_\tau^{n-11/20}$ dependent on $\mathbf{g}_\tau^{n-1/4}$, $\mathbf{g}_\tau^{n-3/4}$ and $\mathbf{g}_\tau^{n-1}$, that is

$$\begin{aligned}&(\mathbf{g}_\tau^{(s)}, \lambda_\tau^{(s)}) - (\mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)})\\ &= -(\mathbf{H}[\mathbf{g}_\tau^{n-1}, \mathbf{g}_\tau^{n+-3/4}, \mathbf{g}_\tau^{n-1/4}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}])^{-1}\mathbf{G}[\mathbf{g}_\tau^{n-1}, \mathbf{g}_\tau^{n-3/4}, \mathbf{g}_\tau^{n-1/4}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}],\end{aligned}$$

    where $\mathbf{g}_\tau^{(0)} := \mathbf{g}_\tau^{n-1/4}$ and the iteration eventually provides $\mathbf{g}_\tau^{(s)} := \mathbf{g}_\tau^{n-11/20}$.
  - Then the stage four Newton iteration provides the discrete solution $\mathbf{g}_\tau^{n-1/2}$ dependent on $\mathbf{g}_\tau^{n-11/20}$, $\mathbf{g}_\tau^{n-1/4}$, $\mathbf{g}_\tau^{n-3/4}$ and $\mathbf{g}_\tau^{n-1}$, that is

$$\begin{aligned}&(\mathbf{g}_\tau^{(s)}, \lambda_\tau^{(s)}) - (\mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)})\\ &= -(\mathbf{H}[\mathbf{g}_\tau^{n-1}, \ldots, \mathbf{g}_\tau^{n-11/20}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}])^{-1}\mathbf{G}[\mathbf{g}_\tau^{n-1}, \ldots, \mathbf{g}_\tau^{n-11/20}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}],\end{aligned}$$

    where $\mathbf{g}_\tau^{(0)} := \mathbf{g}_\tau^{n-11/20}$ and the iteration eventually provides $\mathbf{g}_\tau^{(s)} := \mathbf{g}_\tau^{n-1/2}$.
  - Finally the stage five Newton iteration provides the discrete solution $\mathbf{g}_\tau^{n}$ dependent on $\mathbf{g}_\tau^{n-1/2}$, $\mathbf{g}_\tau^{n-11/20}$, $\mathbf{g}_\tau^{n-1/4}$, $\mathbf{g}_\tau^{n-3/4}$ and $\mathbf{g}_\tau^{n-1}$, that is

$$\begin{aligned}&(\mathbf{g}_\tau^{(s)}, \lambda_\tau^{(s)}) - (\mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)})\\ &= -(\mathbf{H}[\mathbf{g}_\tau^{n-1}, \ldots, \mathbf{g}_\tau^{n-1/2}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}])^{-1}\mathbf{G}[\mathbf{g}_\tau^{n-1}, \ldots, \mathbf{g}_\tau^{n-1/2}; \mathbf{g}_\tau^{(s-1)}, \lambda_\tau^{(s-1)}],\end{aligned}$$

    where $\mathbf{g}_\tau^{(0)} := \mathbf{g}_\tau^{n-1/2}$ and the iteration eventually provides $\mathbf{g}_\tau^{(s)} := \mathbf{g}_\tau^{n}$.

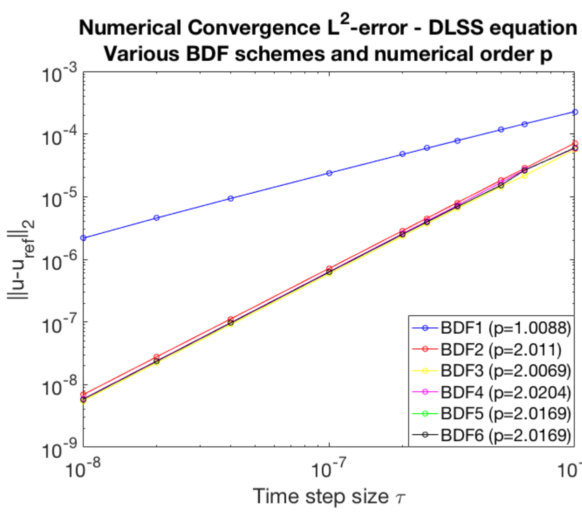
8 Numerical Experiments

In this section, our aim is to conclude on whether we obtain a more accurate, effective scheme based, for example, on higher order BDF and DIRK schemes. Does this create a better approximation to our final solutions $u(x,t)$ of our fourth order nonlinear PDEs and is there any significant effect?

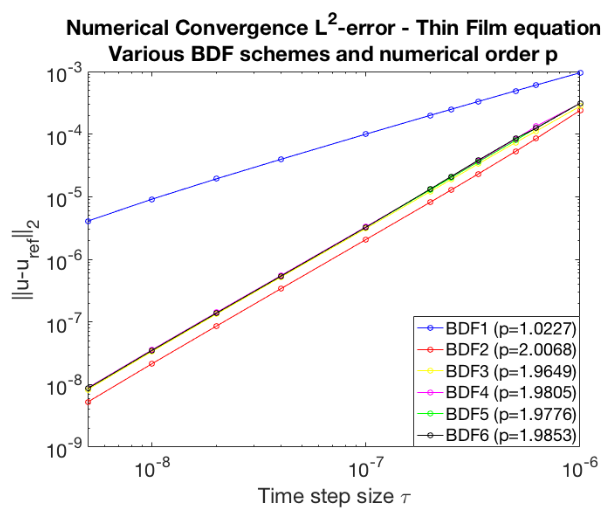
8.1 Analysis for BDF Schemes

Having introduced higher order BDF schemes up to order six, derived with similar features from before, we shall compare the time evolution of our numerical solutions for each scheme before investigating the effects of the convergence rates in the L^2-norm and the error over time step sizes for each scheme.

The plots are first shown for each BDF scheme for the DLSS and Thin Film equations, see Figure 5.



(a) Numerical Convergence L^2 Plot - DLSS equation



(b) Numerical Convergence L^2 Plot - Thin Film equation

Figure 5: Numerical convergence rates for various Wasserstein gradient flow BDF schemes for the DLSS (31)) and Thin Film (34) equations.

Clearly, the difference between the second to sixth order BDF type schemes are small or none i.e. they have the same order of convergence which is two. So despite considering additional information i.e. discrete solution at previous time steps and increased order of accuracy (see Section 4 for details), it does not improve the error and numerical convergence rate compared to the BDF1 and BDF2 schemes, which are A-stable only. In other words, the BDF2 to BDF6 plots have approximately second order convergence only.

8.2 Analysis with DIRK Schemes

Now we consider more challenging multistep schemes, using the diagonally implicit Runge-Kutta (DIRK) schemes of second, third and fourth order, which we have constructed also.

We present numerical results for five higher order BDF schemes plus three examples of diagonally implicit Runge-Kutta schemes consisting of two (63)-(64), three (95)-(97) and five stages (101)-(105), from Sections 4.4-4.6, respectively. Also, despite fixing with one initial condition per equation, we analyse how we could maximise the smoothness of our equation(s). All schemes were implemented in MATLAB.

Before we start, we shall gather some hypothesis for our results: For our error over the time step size, we would expect this to decrease as the step size decreases. With additional solutions at previous time steps to be computed for higher order schemes, as well as higher order of accuracy, the error should be smaller for the higher order BDF and DIRK minimising movement schemes (59) when evaluated over the time step size τ. And finally, smoother initial conditions e.g. continuously differentiable of high order, should lead to increased numerical order of convergence.

The DIRK schemes (59), are of second to fourth order of accuracy respectively compared to the original implicit Euler (BDF1) scheme.

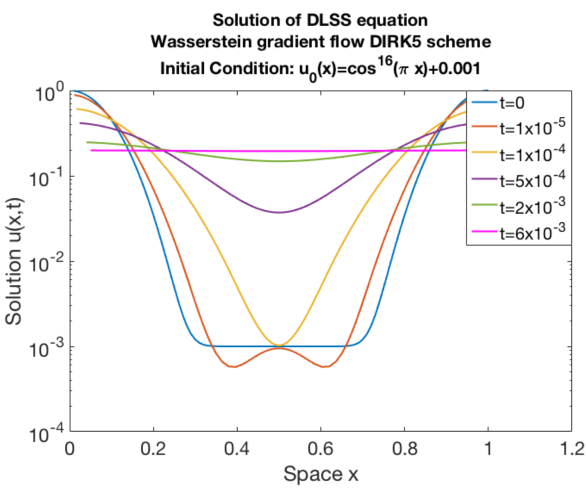
We run several codes before producing the numerical convergence L^2 plots for each scheme on one graph, to compare their rates over a varying time step size τ before we construct several plots for the numerical solutions, with appropriate initial conditions for each, as time t progresses. We have and will be constructing plots for several diffusion equations, as set up in the last section. Furthermore, our plots contain the numerical order of convergence p for each scheme in the legends box.

We also present plots for the Thin Film equation (34) plus equations (35) (where $a = 1/4$ and 2) and (36) but alternative initial datums are considered, in accordance with [22, p. 29], from Grün and Rumpf, and [31, p. 1569], from Kim, respectively. Note that the parameters for the DLSS equation are considered as from [17], but on the other hand, we also summarise additional investigations on how the smoothness of each equation is affected by variance of parameter m and its respective numerical orders of convergence.

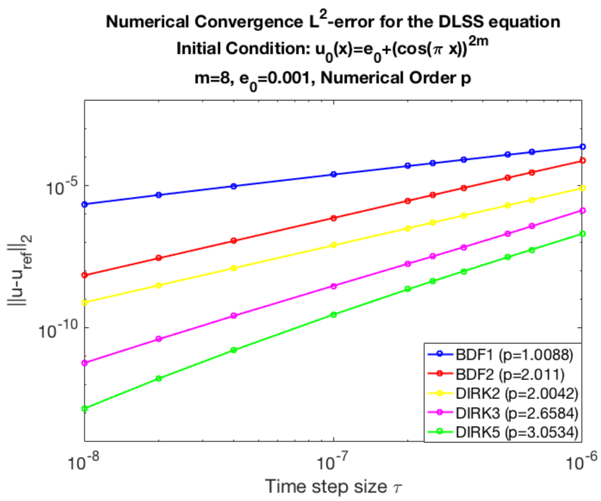
8.3 Results for DLSS Equation

Considering the parameters mentioned in the last paragraph, see Figure 6 for plots, we briefly mention the features of the final solution from the DLSS equation. For our final solution $u(x, t)$ (as seen in the results section in [17] and Figure 6) two local minima develop. Inevitably, each scheme proposed of varying order, constitutes an identical final solution $u(x, t)$ when transformed back from Lagrangian to Eulerian coordinates.

For the DLSS equation, we consider the datum from [7, 17, 28]. We shall use $N = 100$ step intervals with $\tau = 10^{-8}$ (increasing over time, see Figure 13) as the time step size, $T = 5 \times 10^{-6}$ as the terminal time, as used in [17] when plotting the L^2-error plots, with ϵ and m verifying the initial datum and equations considered. We will consider the same terminal time, grid points and initial time step size for the other equations, although with different initial conditions and parameters m and ϵ (see figure titles for these).



(a) Numerical Solution



(b) Numerical Convergence L^2 Plot

Figure 6: Error vs time step τ for various Wasserstein gradient flow BDF plus DIRK schemes and the solution for the DLSS Equation (31)).

Furthermore, we investigate how the smoothness of our initial condition affects the convergence order. From hypothesis, smooth functions should provide a similar order from the Taylor expansion format. The varying initial conditions with respect to parameter to m is shown here along with the table of numerical orders per scheme.

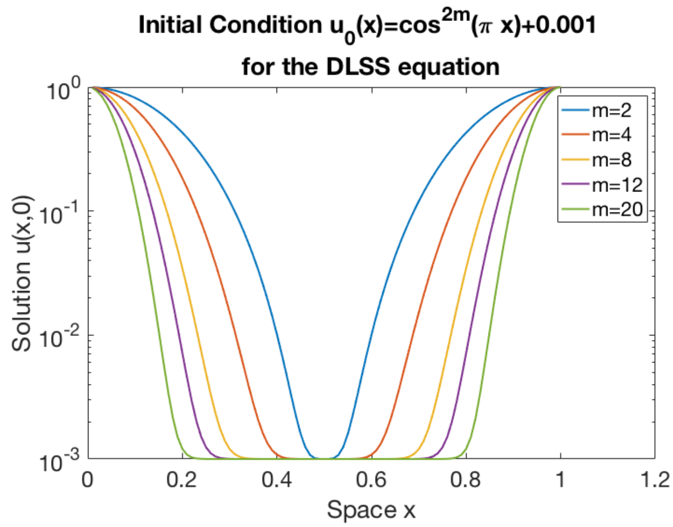


Figure 7: The smoothness comparisons to the initial conditions of the DLSS equation for varying m.

m	BDF1	BDF2	DIRK2	DIRK3	DIRK5
2	1.006	2.0938	1.9965	2.4014	2.384
4	1.0089	2.0165	2.0038	2.687	3.0209
8	1.0088	2.011	2.0042	2.6584	3.0534
12	1.0073	2.0273	1.9899	2.5781	2.9261
20	1.028	1.992	1.9987	2.4635	2.8383

Table 2: The numerical order of convergence for the DLSS equations from variance of m from initial condition.

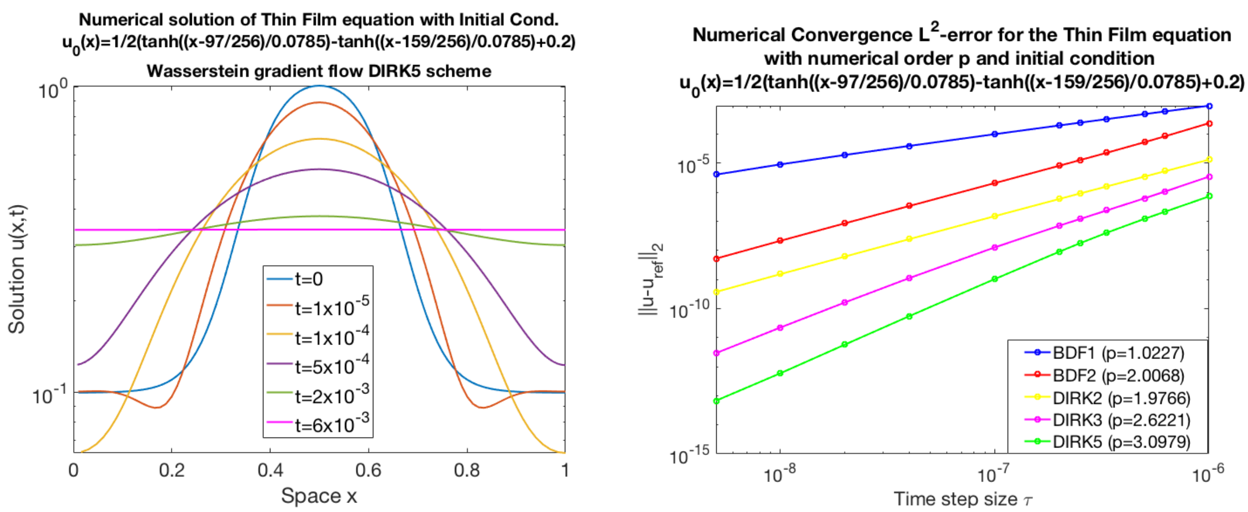
The initial conditions has greater smoothness for the middle parameters and less for the smallest and largest m.

8.4 Results for Thin Film Equation

We consider the initial datum

$$u_0(x) := \frac{1}{2}\left(\tanh\left(\frac{x - 97/256}{m}\right) - \tanh\left(\frac{x - 159/256}{m}\right)\right),$$

an approximation of the piecewise hat function from [22, p. 29], see Figure 8 for plots. From the numerical convergence plot, we observe that the improvement of the numerical convergence error for higher order schemes are more significant for decreasing time step i.e. the plots are converging as the time step increases. Furthermore, the numerical solution violates the minimum principle for small time before the formation of a local minima at time $t = 1 \times 10^{-5}$, until the solution settles to a steady state as time progresses.



(a) Numerical Solution

(b) Numerical Convergence L^2 Plot

Figure 8: Error vs time step τ for various Wasserstein gradient flow BDF plus DIRK schemes and the solution for the Thin Film Equation (34).

Furthermore, we investigate how the smoothness of our initial condition affects the convergence order, as we did for the DLSS equation.

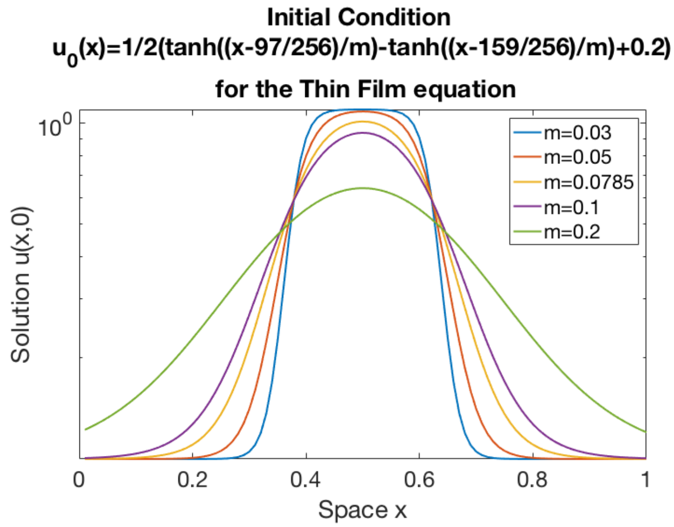


Figure 9: The smoothness comparisons to the initial conditions of the Thin Film equation for varying m.

m	BDF1	BDF2	DIRK2	DIRK3	DIRK5
0.03	0.9917	1.3794	1.487	1.6955	2.2052
0.05	1.0051	1.8084	1.7656	2.1334	2.7444
0.0785	1.0227	2.0068	1.9766	2.6221	3.0979
0.1	1.0078	2.0261	2.007	2.8726	3.4375

Table 3: The numerical order of convergence for the Thin Film equation from variance of m from initial condition.

The initial conditions has greater smoothess for increased parameters, from the convergence error results.

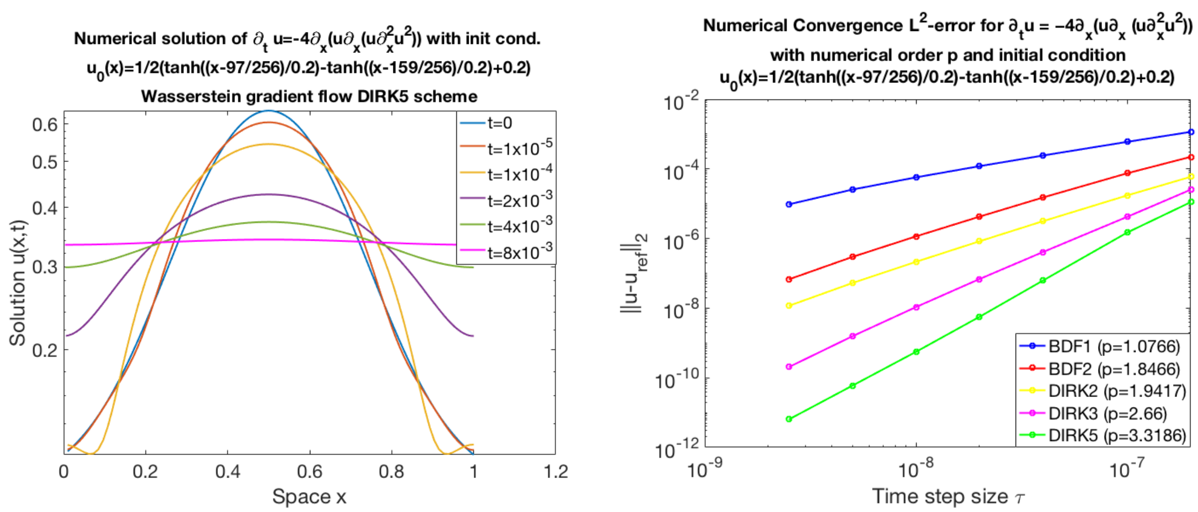
8.5 Results for Nonlinear Equations 1

Now for other unfamiliar equations with the initial datum considered for the best possible approximation (higher numerical order of convergence and lower consistency error). Firstly for (35) when $a = 2$

(see Figure 10 for plots), that is

$$\partial_t u(x,t) = -4\partial_x(u(x,t)\partial_x(u(x,t)\partial_x^2(u(x,t))^2)).$$

Again, the numerical convergence error and rates are more significant for decreasing time step τ. In comparison to the Thin Film equation, by changing a from (35) to $a = 2$, the scheme has a significant numerical convergence rate for the higher order DIRK5 scheme, although the error, e.g. if you carefully observe at time step size $\tau = 1 \times 10^{-8}$, is much higher for the this equation than Thin Film, hence a significantly smaller time step would be required for this fit to be more suitable for this PDE. Here the minimum principle is once again violated with the formation of two local minimas for small enough time, before settling to a steady state as time progresses.



(a) Numerical Solution

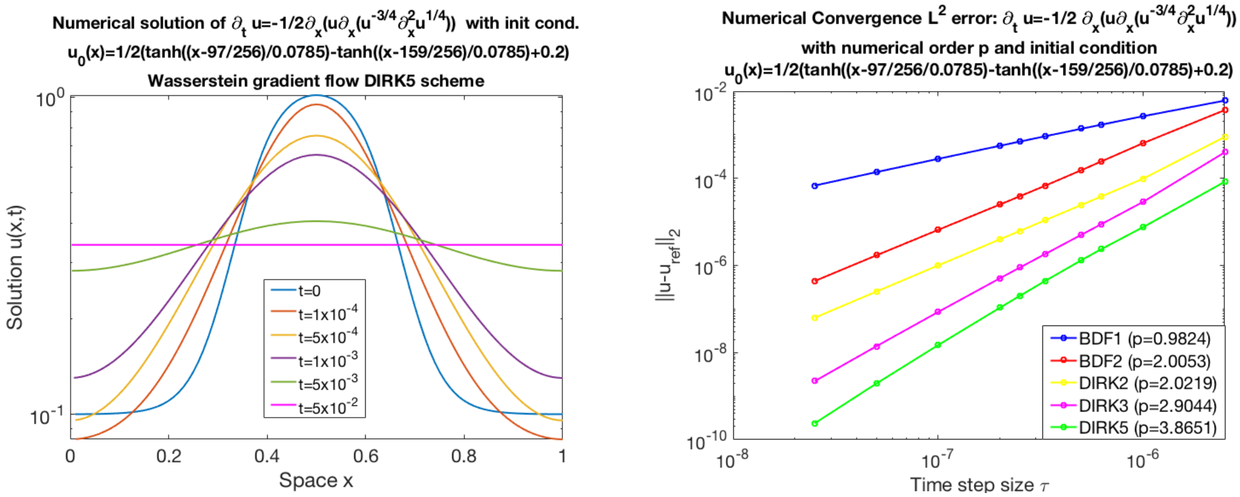
(b) Numerical Convergence L^2 Plot

Figure 10: Error vs time step τ for various Wasserstein gradient flow BDF plus DIRK schemes and the numerical solution for equation (35) when $\alpha = 2$.

8.6 Results for Nonlinear Equation 2

Again for (35) but for $a = \frac{1}{4}$, see Figure 11 for plots:

$$\partial_t u(x,t) = -\frac{1}{2}\partial_x((u(x,t))^{-3/4}\partial_x^2(u(x,t))^{1/4}).$$



(a) Numerical Solution

(b) Numerical Convergence L^2 Plot

Figure 11: Error vs time step τ for various Wasserstein gradient flow BDF plus DIRK schemes and the numerical solution for equation (35) when $\alpha = \frac{1}{4}$.

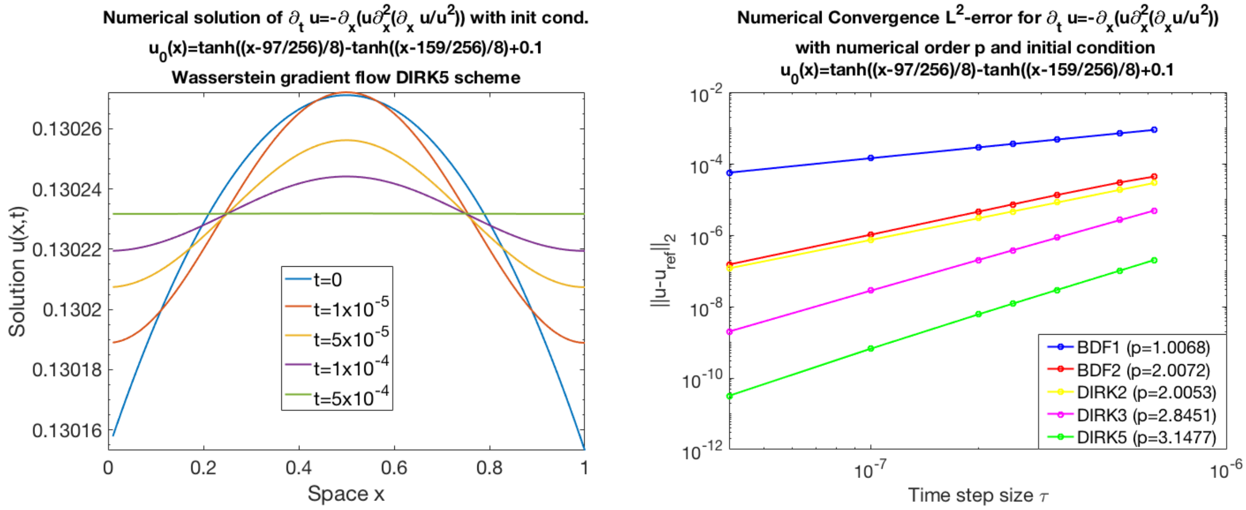
The significant effect on the numerical convergence error/rate is high between the BDF1 scheme and the higher order schemes. The convergence rate is improved for schemes of order two and above, but deteriorates for BDF1, in comparison for the previous equation. Furthermore, the numerical convergence error is improved, but not as efficient for the Thin Film equation. For the numerical convergence plot, the minimum principle is also unsatisfied but no local minima is observed for small time. As usual, the solution settles to a steady state as time progresses.

8.7 Results for Nonlinear Equation 3

Finally, for the equation given in [30, Thm. 3.11, p. 561], see Figure 12 for plots:

$$\partial_t u(x,t) = -\partial_x \left(u(x,t)\partial_{xx} \left(\frac{\partial_x u(x,t)}{(u(x,t))^2} \right) \right).$$

It is inevitable from the numerical convergence plots that there is no benefit to electing the DIRK2 scheme rather than BDF2, in fact the latter has a slightly better rate. The scheme has a significant convergence rate as the time step decreases i.e. the plots slightly converge for increasing time step. On a positive note, the numerical convergence error for this equation is as good as the Thin Film equation, and also the convergence rate is better here also. As for other equations, the numerical solution has no local minimum and the minimum principle is unsatisfied for small time and the solution settles to a steady state over time.



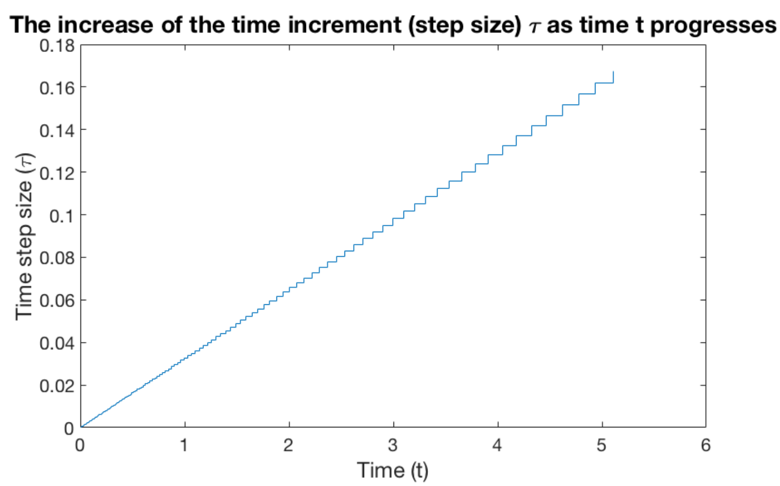
(a) Numerical Solution

(b) Numerical Convergence L^2 Plot

Figure 12: Error vs time step τ for various Wasserstein gradient flow BDF plus DIRK schemes and the numerical solution for equation (36).

8.8 Time Step Variance for Time Progression

Furthermore, the numerical computations are expensive in time, however the significant impact of the approximation occurs at the beginning and the "propagation" of the solution slows as time progresses i.e. the solution tends towards a steady state hence we can gradually increase the time step size τ across every time step. For the numerical solution plots, the relation between the time and the time step size is given as in Figure 13, see [22, p. 29].



(a) Time step size over time

Figure 13: The relationship between the time point and the time steps, applied for each iterative procedure for our numerical solution plots.

In fact the k-th time step T_k is the summation of the time-increments (time step sizes τ_i) evaluated priori to each case i.e. $T_k := \sum_{i=0}^{k} \tau_i$.

Remark 8.1. From the summary of the Newton's method in Section 7.5 (BDF schemes) and Section 7.10 (DIRK schemes), the time step size τ should be generalised in our case now, i.e. should be τ_k for time step T_k and then monotonically increasing for the next time step size i.e. $\tau_k > \tau_{k-1}$.

8.9 Numerical Convergence Summary

The numerical convergence L^2 error plots, where we considered the mesh ratio $\alpha := \frac{\tau}{h^4}$, expectedly gives a higher error for increasing time step size and for the BDF1 scheme, which has a limited order of accuracy. This backs up our initial assumptions that a higher order scheme produces a significantly smaller error than the already constructed BDF1 scheme.

As predicted, it also shows that the error and numerical convergence rate for the BDF2 schemes is improved in comparison to the original BDF1 type scheme. But there is no improvement for BDF3 to 6 schemes of order three to six respectively, which are not A-stable.

On the other hand, the two stage Runge-Kutta type scheme error is significantly smaller in comparison to all the considered BDF schemes. However, considering two schemes of identical order of accuracy of two, the BDF2 and DIRK2 schemes, the numerical order for the DIRK2 scheme is not necessarily superior to BDF2, despite the prior consisting of intermediate stages. Stage one of the DIRK2 scheme only has order of accuracy one, hence ruling out the second stage of being second order, when combining both stages which may be a contributing factor to deteriorating the numerical order result (see Figures 6, 8 and 12). Also note that from the same figures plus Figure 10, see below, the DIRK5 scheme of order four shows only an approximate numerical order of three, which is likely contributed by the recent statement but potentially by weakly chosen parameters. In fact, Tables 2 and 3 with respective initial condition Figures 7 and 9, give an improved numerical order for smoother plots.

Also, from the figure in [17, p. 956], our new schemes show better convergence rates, for the temporal discrete scheme, in comparison to the fully implicit finite difference scheme. With the semi-discretisation for time taken care of, the overlying issue regarding the spatial discretisation is to be investigated i.e. the numerical convergence L^2 error over the mesh size h was significantly worse in comparison to well-known fully implicit finite difference and backward time central space schemes.

8.10 Energy Functional Dissipation

But more importantly, from the numerical approach, the scheme sees the energy functionals (32) dissipate monotonically over time for all our schemes, as shown, for example, the DLSS equation in

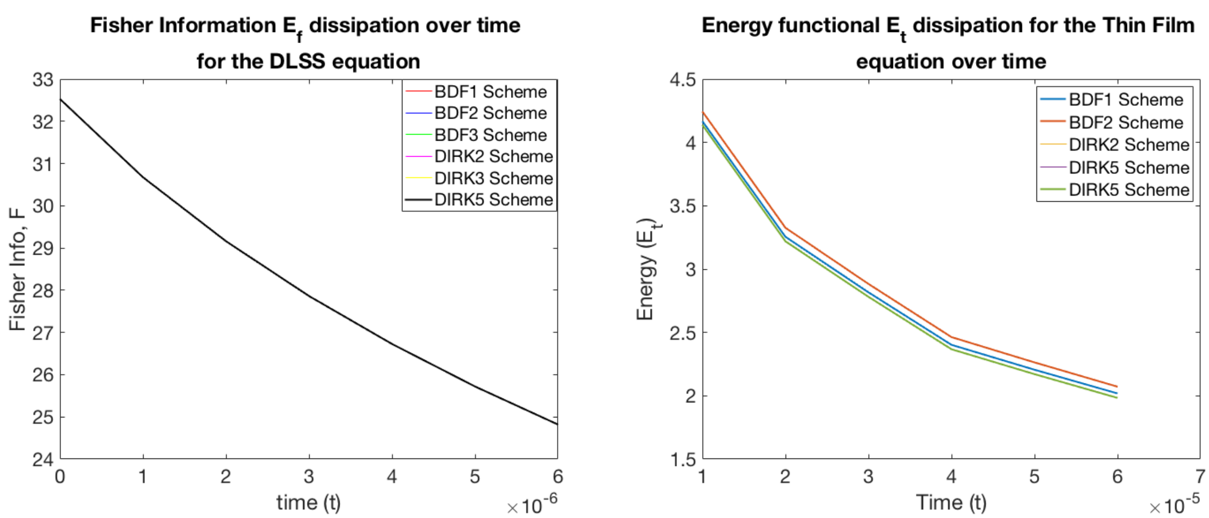
Figure 14. Except for the fact that this is proved theoretically or analytically, for BDF2 (see [35]) and DIRK2, this is despite being unable to directly verify that this was monotonically decreasing:

Also, from our findings, the L-stable Runge-Kutta type schemes of higher order provide the best fit in comparison to the BDF schemes of up to order six.

Hence despite the theoretical challenges of verifying numerical convergence of gradient flow type PDEs, the two plots we have published practically, and numerically, demonstrate not only numerical convergence (see all figures) of our sequence of discrete solutions, but also the geometric behaviour of our solution (the density $u(x,t)$) obeys the gradient flow structure at the discrete level, implying a well-posed solution to our problem for a wide range of higher order nonlinear diffusion equations.

On the other hand, despite our successes with the DLSS equation, there still lies many limitations, when it concerns the intention of investigating other PDEs of similar order, including the fact that our energy functional term admits a limited range only hence further work is needed to consider other equations that could be investigated in future.

In other words, we have only demonstrated the theoretical numerical convergence proof for non-negative energy functionals $\mathcal{E}(\cdot)$. Furthermore, as stated earlier in the subsection, and from the conclusions by Düring et al. [17], weaknesses of our schemes are dominated by the spatial error, which is work to be carried out also.



(a) Energy functional dissipation for the DLSS equation

(b) Energy dissipation for the Thin Film equation

Figure 14: Dissipation of the energy functional $\mathcal{E}(\cdot)$ over time for various BDF and DIRK schemes.

9 Summary

As the three main outcomes from the thesis (see the end of Section 1.3) are covered, we summarise all the key points and findings of the thesis:

9.1 Main Findings

The variational form of the classic minimising movement scheme, with basic assumptions on our energy functionals and Wasserstein metrics, successfully verifies the numerical convergence of our discrete solutions for various forms of diffusion equations of fourth order with strong nonlinearity.

It is clear that the numerical error improves considerably for increased theoretical order and the number of intermediate stages (i.e. from DIRK schemes) of applied schemes.

The numerical order of convergence improves from increasing order and stages, however it is good to point out that this is *not* fully guaranteed for schemes of similar order but various stages (between BDF2 and DIRK2) as we can see clearly, but the improved error from the DIRK2 scheme in comparison is clear to verify, if not in the odd cases (see Figure 12(b)) it is *not* worse. However, we have found that the errors considered across a wider range of orders of our time step τ (we plotted across at least two orders of τ) normally minimises this issue which you would normally expect.

It is worth to point out that we have considered for BDF1 and BDF2 and have constructed the schemes for BDF3 to 6, however as you might have noticed from the final plots, it is *not* practical to investigate for higher order BDF schemes, however results justify that we can improve approximations further from DIRK schemes of higher order with smooth enough initial conditions, hence justifies the selection in our numerical results.

9.2 Future Work to be Carried Out

And finally, there are some limitations to point out. Firstly the psuedo-inverse for the Wasserstein metric is only useful in one space dimension, secondly our spatial discretisation is only applied for a maximum of fourth order, and *not* for sixth order equations of similar structure.

More relevant to our contribution, the convergence proof from the BDF2 scheme is well compatible for the DIRK2 scheme. Although it would be much more complicated, since there would be additional stages to work with, there is potential for this being extended to DIRK3 and DIRK5, particularly that we can work with the similar facts and assumptions on the energy and Wasserstein metric. Possible limitations with the extension would be the ability to generalise the intermediate time steps, where our theoretical order of convergence was only proven for $a \in \left(0.12, \frac{1-\sqrt{2}}{2}\right)$ but the fact that we would work with multiple intermediate steps may balance out this issue.

Finally, this was not considered for this thesis, but some work had already been attempted previously on extending to higher space dimensions. See articles [8], [11] from Carrillo et al. for details.

Appendix

Appendix A: Derivation of BDF3 to 6 Schemes - See Section 4.1

- **BDF3 Scheme**: Taylor expanding $u_\tau(t^{n-1})$, $u_\tau(t^{n-2})$ and $u_\tau(t^{n-3})$ about $t = t^n$ gives

$$u_\tau(t^{n-1}) = u_\tau(t^n) - \tau\partial_t u_\tau(t^n) + \frac{\tau^2}{2}\partial_t^2 u_\tau(t^n) - \frac{\tau^3}{6}\partial_t^3 u_\tau(t^n) + O(\tau^4), \tag{216a}$$

$$u_\tau(t^{n-2}) = u_\tau(t^n) - 2\tau\partial_t u_\tau(t^n) + 2\tau^2\partial_t^2 u_\tau(t^n) - \frac{4\tau^3}{3}\partial_t^3 u_\tau(t^n) + O(\tau^4), \tag{216b}$$

$$u_\tau(t^{n-3}) = u_\tau(t^n) - 3\tau\partial_t u_\tau(t^n) + \frac{9\tau^2}{2}\partial_t^2 u_\tau(t^n) - \frac{9\tau^3}{2}\partial_t^3 u_\tau(t^n) + O(\tau^4). \tag{216c}$$

For this to be third order, we wish to eliminate the τ^2 and τ^3 terms, which are possible by calculating $9(216a) - \frac{9}{2}(216b) + (216c) = 0$, giving us the resulting equation:

$$11u_\tau(t^n) - 18u_\tau(t^{n-1}) + 9u_\tau(t^{n-2}) - 2u_\tau(t^{n-3}) = -6\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau(t^n)) + O(\tau^4).$$

Thus, replacing $u(t^n)$ by its approximate u_τ^n and similarly for other time points, gives us the BDF3 scheme (41).

For the minimising movement scheme, we introduce the inductive scheme, now with another intermediate step to the BDF2 scheme:

$$u_\tau^n := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}}\, \Phi_3^\tau(u_\tau^{n-1}, u_\tau^{n-2}, u_\tau^{n-3}; u),$$
$$\Phi_3^\tau(u_\tau^{n-1}, u_\tau^{n-2}, u_\tau^{n-3}; u) := \frac{a}{\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \frac{b}{\tau}\mathcal{W}_2[u_\tau^{n-2}, u]^2 + \frac{c}{\tau}\mathcal{W}_2[u_\tau^{n-3}, u]^2 + \mathcal{E}(u),$$

and the minimiser u_τ^n gives us

$$\frac{2}{\tau}\Big((a+b+c)u_\tau^n - au_\tau^{n-1} - bu_\tau^{n-2} - cu_\tau^{n-3}\Big) = -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n),$$

which satisfies the BDF3 formula (41) if $a = \frac{3}{2}$, $b = -\frac{3}{4}$ and $c = \frac{1}{6}$. Hence, this gives us our final scheme (45).

- **BDF4 Scheme**: Taylor expanding $u_\tau(t^{n-1})$, $u_\tau(t^{n-2})$, $u_\tau(t^{n-3})$ and $u_\tau(t^{n-4})$ about $t = t^n$ gives

$$u_\tau(t^{n-1}) = u_\tau(t^n) - \tau\partial_t u_\tau(t^n) + \frac{\tau^2}{2}\partial_t^2 u_\tau(t^n) - \frac{\tau^3}{6}\partial_t^3 u_\tau(t^n) + \frac{\tau^4}{2}4\partial_t^4 u_\tau(t^n) + O(\tau^5), \tag{217a}$$

$$u_\tau(t^{n-2}) = u_\tau(t^n) - 2\tau\partial_t u_\tau(t^n) + 2\tau^2\partial_t^2 u_\tau(t^n) - \frac{4\tau^3}{3}\partial_t^3 u_\tau(t^n) + \frac{2\tau^4}{3}\partial_t^4 u_\tau(t^n) + O(\tau^5), \tag{217b}$$

$$\begin{aligned} u_\tau(t^{n-3}) = &u_\tau(t^n) - 3\tau\partial_t u_\tau(t^n) + \frac{9\tau^2}{2}\partial_t^2 u_\tau(t_\tau^n) - \frac{9\tau^3}{2}\partial_t^3 u_\tau(t_\tau^n) \\ &+ \frac{27\tau^4}{8}\partial_t^4 u_\tau(t^n) + O(\tau^5), \end{aligned} \tag{217c}$$

$$u_\tau(t^{n-4}) = u_\tau(t^n) - 4\tau\partial_t u_\tau(t^n) + 8\tau\partial_t^2 u_\tau(t^n) - \frac{32\tau^3}{3}\partial_t^3 u_\tau(t^n) \tag{217d}$$

$$+\frac{32\tau^4}{3}\partial_t^4 u_\tau(t^n) + O(\tau^5).$$

For this to be fourth order, we wish to eliminate the τ^2, τ^3 and τ^4 terms, which are possible by calculating $-16(217a) + 12(217b) - \frac{16}{3}(217c) + (217d) = 0$, giving us the resulting equation:

$$25u_\tau(t^n) - 48u_\tau(t^{n-1}) + 36u_\tau(t^{n-2}) - 16u_\tau(t^{n-3}) + 3u_\tau(t^{n-4}) = -12\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau(t^n)) + O(\tau^5).$$

Thus, replacing $u(t^n)$ by its approximate u_τ^n and similarly for other time points, gives us the BDF4 scheme (42).

Then we introduce for the minimising movement scheme:

$$\begin{aligned} u_\tau^n :=& \operatorname*{argmin}_{u\in\mathcal{P}_M(\Omega)} \Phi_4^\tau(u_\tau^{n-1}, u_\tau^{n-2}, u_\tau^{n-3}, u_\tau^{n-4}; u), \\ \Phi_4^\tau(u_\tau^{n-1}, u_\tau^{n-2}, u_\tau^{n-3}, u_\tau^{n-4}; u) :=& \frac{a}{\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \frac{b}{\tau}\mathcal{W}_2[u_\tau^{n-2}, u]^2 + \frac{c}{\tau}\mathcal{W}_2[u_\tau^{n-3}, u]^2 \\ &+ \frac{d}{\tau}\mathcal{W}_2[u_\tau^{n-4}, u]^2 + \mathcal{E}(u), \end{aligned}$$

and the minimiser u_τ^n gives us

$$\frac{2}{\tau}\Big((a+b+c+d)u_\tau^n - au_\tau^{n-1} - bu_\tau^{n-2} - cu_\tau^{n-3} - du_\tau^{n-4}\Big) = -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n),$$

which satisfies the BDF4 formula (42) if $a = 2$, $b = -\frac{3}{2}$, $c = \frac{2}{3}$ and $d = -\frac{1}{8}$. Hence this gives us our final scheme (46).

- **BDF5 Scheme**: Taylor expanding $u_\tau(t^{n-1})$, $u_\tau(t^{n-2})$, $u_\tau(t^{n-3})$, $u_\tau(t^{n-4})$ and $u_\tau(t^{n-5})$ about $t = t^n$ gives

$$\begin{aligned} u_\tau(t^{n-1}) =& u_\tau(t^n) - \tau\partial_t u_\tau(t^n) + \frac{\tau^2}{2}\partial_t^2 u_\tau(t^n) - \frac{\tau^3}{6}\partial_t^3 u_\tau(t_\tau^n) + \frac{\tau^4}{24}\partial_t^4 u_\tau(t^n) \\ &- \frac{\tau^5}{120}\partial_t^5 u_\tau(t^n) + O(\tau^6), \end{aligned} \tag{218a}$$

$$\begin{aligned} u_\tau(t^{n-2}) =& u_\tau(t^n) - 2\tau\partial_t u_\tau(t^n) + 2\tau^2\partial_t^2 u_\tau(t^n) - \frac{4\tau^3}{3}\partial_t^3 u_\tau(t^n) + \frac{2\tau^4}{3}\partial_t^4 u_\tau(t^n) \\ &- \frac{4\tau^5}{15}\partial_t^5 u_\tau(t^n) + O(\tau^6), \end{aligned} \tag{218b}$$

$$\begin{aligned} u_\tau(t^{n-3}) =& u_\tau(t^n) - 3\tau\partial_t u_\tau(t^n) + \frac{9\tau^2}{2}\partial_t^2 u_\tau(t^n) - \frac{9\tau^3}{2}\partial_t^3 u_\tau(t^n) + \frac{27\tau^4}{8}\partial_t^4 u_\tau(t^n) \\ &- \frac{81\tau^5}{40}\partial_t^5 u_\tau(t^n) + O(\tau^6), \end{aligned} \tag{218c}$$

$$\begin{aligned} u_\tau(t^{n-4}) =& u_\tau(t^n) - 4\tau\partial_t u_\tau(t^n) + 8\tau\partial_t^2 u_\tau(t^n) - \frac{32\tau^3}{3}\partial_t^3 u_\tau(t^n) + \frac{32\tau^4}{3}\partial_t^4 u_\tau(t^n) \\ &- \frac{128\tau^5}{15}\partial_t^5 u_\tau(t^n) + O(\tau^6), \end{aligned} \tag{218d}$$

$$u_\tau(t^{n-5}) = u_\tau(t^n) - 5\tau\partial_t u_\tau(t^n) + \frac{25\tau^2}{2}\partial_t^2 u_\tau(t^n) - \frac{125\tau^3}{6}\partial_t^3 u_\tau(t^n) + \frac{625\tau^4}{24}\partial_t^4 u_\tau(t^n) \tag{218e}$$

$$- \frac{625\tau^5}{24}\partial_t^5 u_\tau(t^n) + O(\tau^6).$$

For this to be fifth order, we wish to eliminate the τ^2, τ^3, τ^4 and τ^5 terms, which are possible by calculating $25(218a) - 25(218b) + \frac{50}{3}(218c) - \frac{25}{4}(218d) + (218e) = 0$, giving us the resulting equation:

$$\begin{aligned}
&137u_\tau(t^n) - 300u_\tau(t^{n-1}) + 300u_\tau(t^{n-2}) - 200u_\tau(t^{n-3}) + 75u_\tau(t^{n-4}) - 12u_\tau(t^{n-5}) \\
=& - 60\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau(t^n)) + O(\tau^6).
\end{aligned}$$

Thus, replacing $u(t^n)$ by its approximate u_τ^n and similarly for other time points, gives us the BDF5 scheme (43).

Then we introduce for the minimising movement scheme:

$$\begin{aligned}
u_\tau^n :=& \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}}\ \Phi_5^\tau(u_\tau^{n-1}, u_\tau^{n-2}, u_\tau^{n-3}, u_\tau^{n-4}, u_\tau^{n-5}; u), \\
\Phi_4^\tau(u_\tau^{n-1}, u_\tau^{n-2}, u_\tau^{n-3}, u_\tau^{n-4}, u_\tau^{n-5}; u) :=& \frac{a}{\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \frac{b}{\tau}\mathcal{W}_2[u_\tau^{n-2}, u]^2 + \frac{c}{\tau}\mathcal{W}_2[u_\tau^{n-3}, u]^2 \\
&+ \frac{d}{\tau}\mathcal{W}_2[u_\tau^{n-4}, u]^2 + \frac{e}{\tau}\mathcal{W}_2[u_\tau^{n-5}, u]^2 + \mathcal{E}(u),
\end{aligned}$$

and the minimiser u_τ^n gives us

$$\frac{2}{\tau}\left((a+b+c+d+e)u_\tau^n - au_\tau^n - bu_\tau^{n-2} - cu_\tau^{n-3} - du_\tau^{n-4} - eu_\tau^{n-5}\right) = -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n),$$

which satisfies the BDF5 formula (43) if $a = \frac{5}{2}$, $b = -\frac{5}{2}$, $c = \frac{5}{3}$, $d = -\frac{5}{8}$ and $e = \frac{1}{10}$. Hence this gives us our final scheme (47).

- **BDF6 Scheme**: Taylor expanding $u_\tau(t^{n-1})$, $u_\tau(t^{n-2})$, $u_\tau(t^{n-3})$, $u_\tau(t^{n-4})$, $u_\tau(t^{n-5})$ and $u_\tau(t^{n-6})$ about $t = t^n$ gives

$$\begin{aligned}
u_\tau(t^{n-1}) =& u_\tau(t^n) - \tau\partial_t u_\tau(t^n) + \frac{\tau^2}{2}\partial_t^2 u_\tau(t^n) - \frac{\tau^3}{6}\partial_t^3 u_\tau(t^n) + \frac{\tau^4}{2}4\partial_t^4 u_\tau(t^n) \\
&- \frac{\tau^5}{120}\partial_t^5 u_\tau(t^n) + \frac{\tau^6}{720}\partial_t^6 u_\tau(t^n) + O(\tau^7),
\end{aligned} \tag{219a}$$

$$\begin{aligned}
u_\tau(t^{n-2}) =& u_\tau(t^n) - 2\tau\partial_t u_\tau(t^n) + 2\tau^2\partial_t^2 u_\tau(t^n) - \frac{4\tau^3}{3}\partial_t^3 u_\tau(t^n) + \frac{2\tau^4}{3}\partial_t^4 u_\tau(t^n) \\
&- \frac{4\tau^5}{15}\partial_t^5 u_\tau(t^n) + \frac{4\tau^6}{45}\partial_t^6 u_\tau(t^n) + O(\tau^7),
\end{aligned} \tag{219b}$$

$$\begin{aligned}
u_\tau(t^{n-3}) =& u_\tau(t^n) - 3\tau\partial_t u_\tau(t^n) + \frac{9\tau^2}{2}\partial_t^2 u_\tau(t^n) - \frac{9\tau^3}{2}\partial_t^3 u_\tau(t^n) + \frac{27\tau^4}{8}\partial_t^4 u_\tau(t^n) \\
&- \frac{81\tau^5}{40}\partial_t^5 u_\tau(t^n) + \frac{81\tau^6}{80}\partial_t^6 u_\tau(t^n) + O(\tau^7),
\end{aligned} \tag{219c}$$

$$\begin{aligned}
u_\tau(t^{n-4}) =& u_\tau(t^n) - 4\tau\partial_t u_\tau(t^n) + 8\tau\partial_t^2 u_\tau(t^n) - \frac{32\tau^3}{3}\partial_t^3 u_\tau(t^n) + \frac{32\tau^4}{3}\partial_t^4 u_\tau(t^n) \\
&- \frac{128\tau^5}{15}\partial_t^5 u_\tau(t^n) + \frac{256\tau^6}{45}\partial_t^6 u_\tau(t^n) + O(\tau^7),
\end{aligned} \tag{219d}$$

$$u_\tau(t^{n-5}) = u_\tau(t^n) - 5\tau\partial_t u_\tau(t^n) + \frac{25\tau^2}{2}\partial_t^2 u_\tau(t^n) - \frac{125\tau^3}{6}\partial_t^3 u_\tau(t^n) + \frac{625\tau^4}{24}\partial_t^4 u_\tau(t^n) - \frac{625\tau^5}{24}\partial_t^5 u_\tau(t^n) + \frac{3125\tau^6}{144}\partial_t^6 u_\tau(t^n) + O(\tau^7), \tag{219e}$$

$$u_\tau(t^{n-6}) = u_\tau(t^n) - 6\tau\partial_t u_\tau(t^n) + 18\tau^2\partial_t^2 u_\tau(t^n) - 36\tau^3\partial_t^3 u_\tau(t^n) + 54\tau^4\partial_t^4 u_\tau(t^n) - \frac{324\tau^5}{5}\partial_t^5 u_\tau(t^n) + \frac{324\tau^6}{5}\partial_t^6 u_\tau(t^n) + O(\tau^7). \tag{219f}$$

For this to be sixth order, we wish to eliminate the τ^2, τ^3, τ^4, τ^5 and τ^6 terms, which are possible by calculating $-36(219a) + 45(219b) - 40(219c) + \frac{45}{2}(219d) - \frac{36}{5}(219e) + (219f) = 0$, giving us the resulting equation:

$$\begin{aligned}&147u_\tau(t^n) - 360u_\tau(t^{n-1}) + 450u_\tau(t^{n-2}) - 400u_\tau(t^{n-3}) + 225u_\tau(t^{n-4}) - 72u_\tau(t^{n-5}) + 10u_\tau(t^{n-6})\\ =& -60\tau\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau(t^n)) + O(\tau^7).\end{aligned}$$

Thus, replacing $u_\tau(t^n)$ by its approximate u_τ^n and similarly for other time points, gives us the BDF6 scheme (44).

Then we introduce for the minimising movement scheme:

$$u_\tau^n := \underset{u\in\mathcal{P}_M(\Omega)}{\operatorname{argmin}}\ \Phi_5^\tau(u_\tau^{n-1}, u_\tau^{n-2}, u_\tau^{n-3}, u_\tau^{n-4}, u_\tau^{n-5}, u_\tau^{n-6}; u),$$

$$\begin{aligned}&\Phi_5^\tau(u_\tau^{n-1}, u_\tau^{n-2}, u_\tau^{n-3}, u_\tau^{n-4}, u_\tau^{n-5}, u_\tau^{n-6}; u)\\ :=&\frac{a}{\tau}\mathcal{W}_2[u_\tau^{n-1}, u]^2 + \frac{b}{\tau}\mathcal{W}_2[u_\tau^{n-2}, u]^2 + \frac{c}{\tau}\mathcal{W}_2[u_\tau^{n-3}, u]^2 + \frac{d}{\tau}\mathcal{W}_2[u_\tau^{n-4}, u]^2\\ &+ \frac{e}{\tau}\mathcal{W}_2[u_\tau^{n-5}, u]^2 + \frac{f}{\tau}\mathcal{W}_2[u_\tau^{n-6}, u]^2 + \mathcal{E}(u),\end{aligned}$$

and the minimiser u_τ^n gives us

$$\frac{2}{\tau}\left((a+b+c+d+e+f)u_\tau^n - au_\tau^{n-1} - bu_\tau^{n-2} - cu_\tau^{n-3} - du_\tau^{n-4} - eu_\tau^{n-5} - fu_\tau^{n-6}\right) = -\nabla_{\mathcal{W}_2}\mathcal{E}(u_\tau^n),$$

which satisfies the BDF6 formula (44) if $a = 3$, $b = -\frac{15}{4}$, $c = \frac{10}{3}$, $d = -\frac{15}{8}$, $e = \frac{3}{5}$ and $f = -\frac{1}{12}$.

Hence this gives us our final scheme (48).

Appendix B: Non-positivity of $\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2$ from Equation (153) in Lemma 5.24

We selected $\epsilon = \frac{2a^2-6a+3}{1+2a(1-a)}$ for proving non-positivity of the $\mathcal{W}_2[u_\tau^{n-1}, u_\tau^{n+a-1}]^2$ term:

By substituting this choice into the $\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2$ term gives

$$\left(2a(1-a)\frac{2a^2-10a+7}{1+2a(1-a)} - 1 + \frac{a(1-2a)(2a^2-6a+3)}{1+2a(1-a)}\lambda\tau\right)\mathcal{W}_2[u_\tau^{n-1}, u_\tau^n]^2, \tag{220}$$

where we wish for this to be non-negative if, by setting an estimate on λ:

$$\frac{a(1-2a)(2a^2-6a+3)}{1+2a(1-a)}\lambda\tau \geq 1 - 2a(1-a)\frac{2a^2-10a+7}{1+2a(1-a)} \tag{221}$$

$$\Leftrightarrow \frac{a(1-2a)(2a^2-6a+3)}{1+2a(1-a)}\lambda\tau \geq \quad \frac{1+2a(1-a)-2a(1-a)(2a^2-10a+7)}{1+2a(1-a)}$$

$$\Rightarrow \lambda\tau \geq \quad \frac{1+2a(1-a)-2a(1-a)(2a^2-10a+7)}{a(1-2a)(2a^2-6a+3)}.$$

However, from the semi-convexity condition on λ, see (136), the inequality (221) holds providing that

$$\frac{1+2a(1-a)-2a(1-a)(2a^2-10a+7)}{a(1-2a)(2a^2-6a+3)} > \frac{2(a-1)}{1-2a} \tag{222}$$

$$\Leftrightarrow 1 > 2a(a-1) - 2a(a-1)(2a^2-10a+7) + 2a(a-1)(2a^2-6a+3)$$

$$\Leftrightarrow 1 > 2a(a-1)(4a-3).$$

which holds for all $a \in \left(0, \frac{1-\sqrt{2}}{2}\right)$.

In addition, from (136) we set $\lambda \leq 0$, thus our approach is only existent if

$$-2a(1-a)(2a^2-10a+6)+1 \leq 0,$$

which is only true in our case for all $a \in \left(\overline{a}, 1-\frac{\sqrt{2}}{2}\right)$, where $\overline{a} \approx 0.12$.

Appendix C: Proof of Exponential Prefactor (176) - See Lemma 6.7

Rewriting as $\dfrac{\log\left(\dfrac{(1+a\lambda\tau)\,(2(1-a)+(1-2a)\lambda\tau)}{2(1-a)-(1-2a(1-a))\lambda\tau}\right)}{\tau/T}$, we have that as $\tau \to 0$, the expression tends to 0/0 (indeterminate form). Therefore, we apply L-Hôpital's rule, which gives us

$$\begin{aligned}
&\lim_{\tau\to 0} T\frac{a\lambda\,(2(1-a)+(1-2a)\lambda\tau)+(1-2a)\lambda(1+a\lambda\tau)}{(1+a\lambda\tau)\,(2(1-a)+(1-2a)\lambda\tau)} \\
&+\lim_{\tau\to 0} T\frac{\lambda(1-2a(1-a))}{2(1-a)-(1-2a(1-a)\lambda\tau)} \\
= \quad & T\frac{(a\lambda\,(2(1-a))+(1-2a)\lambda)\,(2(1-a))+2\lambda(1-a)(1-2a(1-a))}{4(1-a)^2} \\
= \quad & T\frac{2a\lambda(1-a)+(1-2a)\lambda+\lambda(1-2a(1-a))}{2(1-a)} = \lambda T.
\end{aligned}$$

Appendix D: Proof of Result (186) from (185) - See Lemma 6.10

The denominator of (185) is zero if and only if

$$\begin{aligned}
&2(1-a)-(1-2a(1-a))\lambda\eta = (1+a\lambda\eta)[2(1-a)+(1-2a)\lambda\eta] \\
\Rightarrow &2(1-a)-(1-2a(1-a))\lambda\eta = 2(1-a)+(1-2a)\lambda\eta+2a(1-a)\lambda\eta+a(1-2a)(\lambda\eta)^2 \\
\Rightarrow &-\lambda\eta+2a(1-a)\lambda\eta = (1-2a)\lambda\eta+2a(1-a)\lambda\eta+a(1-2a)(\lambda\eta)^2
\end{aligned}$$

$$\Rightarrow 2(a-1)\lambda\eta = a(1-2a)(\lambda\eta)^2$$
$$\Rightarrow \lambda\eta = \frac{2(a-1)}{a(1-2a)}.$$

Appendix E: Finite-Dimensional Form of our Energy Functionals - See Section 7.3

By implementing the discretisation process provided by Düring et al. [17], as we summarised in Section 7.2, we achieve the finite-dimensional forms of our energy functionals for each of our equations, we have provided numerical results for. The computations are given in the order as included in Section 7.3:

Proof of equation (194), Section 7.3.1:

$$\begin{aligned}
\mathcal{E}_d(u) &= \frac{1}{2}\sum_{k=1}^{n} \delta_k^2 (g_k - g_{k-1})^2 \int_{\omega_{k-1}}^{\omega_k} \frac{1}{(g_k(\omega-\omega_{k-1}) + g_{k-1}(\omega_k - \omega))^4} d\omega \\
&= -\frac{1}{6}\sum_{k=1}^{n} \delta_k^2 (g_k - g_{k-1}) \frac{1}{(g_k(\omega-\omega_{k-1}) + g_{k-1}(\omega_k - \omega))^3}\Big|_{\omega_{k-1}}^{\omega_k} \\
&= \frac{1}{6}\sum_{k=1}^{n} \delta_k^2 (g_k - g_{k-1}) \left(\frac{1}{\delta_k^3 g_{k-1}^3} - \frac{1}{\delta_k^3 g_k^3}\right) \\
&= \frac{1}{6}\sum_{k=1}^{n} \frac{1}{\delta_k} (g_k - g_{k-1}) \left(\frac{1}{g_{k-1}^3} - \frac{1}{g_k^3}\right).
\end{aligned}$$

Proof of equation (195), Section 7.3.2:

$$\begin{aligned}
\mathcal{E}_t(u) &= \frac{1}{2}\sum_{k=1}^{n} \int_{\omega_{k-1}}^{\omega_k} \left(\frac{\partial}{\partial\omega}\left(\frac{g_k(\omega-\omega_{k-1}) + g_{k-1}(\omega_k - \omega)}{\delta_k}\right)\right)^2 \cdot \left(\frac{\delta_k}{g_k(\omega-\omega_{k-1}) + g_{k-1}(\omega_k - \omega)}\right)^5 d\omega \\
&= \frac{1}{2}\sum_{k=1}^{n} \int_{\omega_{k-1}}^{\omega_k} \frac{(g_k - g_{k-1})^2}{\delta_k^2} \cdot \frac{\delta_k^5}{(g_k(\omega-\omega_{k-1}) + g_{k-1}(\omega_k - \omega))^5} d\omega \\
&= \frac{1}{2}\sum_{k=1}^{n} \int_{\omega_{k-1}}^{\omega_k} \delta_k^3 (g_k - g_{k-1})^2 \frac{1}{(g_k(\omega-\omega_{k-1}) + g_{k-1}(\omega_k - \omega))^5} d\omega \\
&= -\frac{1}{8}\sum_{k=1}^{n} \delta_k^3 (g_k - g_{k-1}) \cdot \frac{1}{(g_k(\omega-\omega_{k-1}) + g_{k-1}(\omega_k - \omega))^4}\Big|_{\omega_{k-1}}^{\omega_k} \\
&= \frac{1}{8}\sum_{k=1}^{n} \delta_k^3 (g_k - g_{k-1}) \left(\frac{1}{g_{k-1}^4 \delta_k^4} - \frac{1}{g_k^4 \delta_k^4}\right) \\
&= \frac{1}{8}\sum_{k=1}^{n} \frac{1}{\delta_k} (g_k - g_{k-1}) \left(\frac{1}{g_{k-1}^4} - \frac{1}{g_k^4}\right).
\end{aligned}$$

Proof of equation (197), Section 7.3.3:

$$\mathcal{E}_v(u) = a^2 \sum_{k=1}^{n} \int_{\omega_{k-1}}^{\omega_k} \left(\frac{\partial}{\partial\omega}\left(\frac{g_k(\omega-\omega_{k-1}) + g_{k-1}(\omega_k - \omega)}{\delta_k}\right)\right)^2 \cdot \left(\frac{\delta_k}{g_k(\omega-\omega_{k-1}) + g_{k-1}(\omega_k - \omega)}\right)^{2a+3} d\omega$$

$$
\begin{aligned}
&= a^2\sum_{k=1}^{n}\int_{\omega_{k-1}}^{\omega_k}\frac{(g_k-g_{k-1})^2}{\delta_k^2}\cdot\frac{\delta_k^{2a+3}}{(g_k(\omega-\omega_{k-1})+g_{k-1}(\omega_k-\omega))^{2a+3}}d\omega\\
&= a^2\sum_{k=1}^{n}\int_{\omega_{k-1}}^{\omega_k}\delta_k^{2a+1}(g_k-g_{k-1})^2\frac{1}{(g_k(\omega-\omega_{k-1})+g_{k-1}(\omega_k-\omega))^{2a+3}}d\omega\\
&= -\frac{a^2}{2(a+1)}\sum_{k=1}^{n}\delta_k^{2a+1}(g_k-g_{k-1})\cdot\frac{1}{(g_k(\omega-\omega_{k-1})+g_{k-1}(\omega_k-\omega))^{2(a+1)}}\Big|_{\omega_{k-1}}^{\omega_k}\\
&= \frac{a^2}{2(a+1)}\sum_{k=1}^{n}\delta_k^{2a+1}(g_k-g_{k-1})\left(\frac{1}{g_{k-1}^{2(a+1)}\delta_k^{2(a+1)}}-\frac{1}{g_k^{2(a+1)}\delta_k^{2(a+1)}}\right)\\
&= \frac{a^2}{2(a+1)}\sum_{k=1}^{n}\frac{1}{\delta_k}(g_k-g_{k-1})\left(\frac{1}{g_{k-1}^{2(a+1)}}-\frac{1}{g_k^{2(a+1)}}\right).
\end{aligned}
$$

Proof of equation (199), Section 7.3.4:

$$
\begin{aligned}
\mathcal{E}_f(u) &= \frac{1}{2}\sum_{k=1}^{n}\int_{\omega_{k-1}}^{\omega_k}\left(\frac{\partial}{\partial\omega}\left(\frac{g_k(\omega-\omega_{k-1})+g_{k-1}(\omega_k-\omega)}{\delta_k}\right)\right)^2\cdot\left(\frac{\delta_k}{g_k(\omega-\omega_{k-1})+g_{k-1}(\omega_k-\omega)}\right)^3 d\omega\\
&= \frac{1}{2}\sum_{k=1}^{n}\int_{\omega_{k-1}}^{\omega_k}\frac{(g_k-g_{k-1})^2}{\delta_k^2}\cdot\frac{\delta_k^3}{(g_k(\omega-\omega_{k-1})+g_{k-1}(\omega_k\omega))^3}d\omega\\
&= \frac{1}{2}\sum_{k=1}^{n}\int_{\omega_{k-1}}^{\omega_k}\delta_k(g_k-g_{k-1})^2\cdot\frac{1}{(g_k(\omega-\omega_{k-1})+g_{k-1}(\omega_k-\omega))^3}d\omega\\
&= -\frac{1}{4}\sum_{k=1}^{n}\delta_k(g_k-g_{k-1})\frac{1}{(g_k(\omega-\omega_{k-1})+g_{k-1}(\omega_k-\omega))^2}\Big|_{\omega_{k-1}}^{\omega_k}\\
&= \frac{1}{4}\sum_{k=1}^{n}\delta_k(g_k-g_{k-1})\left(\frac{1}{g_{k-1}^2\delta_k^2}-\frac{1}{g_k^2\delta_k^2}\right)=\frac{1}{4}\sum_{k=1}^{n}\frac{1}{\delta_k}(g_k-g_{k-1})\left(\frac{1}{g_{k-1}^2}-\frac{1}{g_k^2}\right).
\end{aligned}
$$

Appendix F: Critical Points of the Lagrangian, $\mathbf{G_k}$ - See Section 7.4

The critical points of the Lagrangian are given as follows for each BDF type scheme:

$$
\textbf{BDF1 Scheme}:\mathbf{G_k}=\frac{1}{\tau}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-1})+\frac{1}{\alpha}\frac{\partial\mathbf{F}_k[\mathbf{g}]}{\partial g_k}+\frac{1}{\alpha}\frac{\partial\mathbf{F}_{k+1}[\mathbf{g}]}{\partial g_k}-\lambda\Delta_k,
$$

$$
\begin{aligned}
\textbf{BDF2 Scheme}:\mathbf{G_k}=&\frac{3}{2\tau}\Big(\sum_{j=1}^{n}\frac{4}{3}a_{j,k}(g_j-g_j^{n-1})-\sum_{j=1}^{n}\frac{1}{3}a_{j,k}(g_j-g_j^{n-2})\Big)\\
&+\frac{1}{\alpha}\frac{\partial\mathbf{F}_k[\mathbf{g}]}{\partial g_k}+\frac{1}{\alpha}\frac{\partial\mathbf{F}_{k+1}[\mathbf{g}]}{\partial g_k}-\lambda\Delta_k,
\end{aligned}
$$

$$
\begin{aligned}
\textbf{BDF3 Scheme}:\mathbf{G_k}=&\frac{11}{6\tau}\Big(\frac{18}{11}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-1})-\frac{9}{11}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-2})\\
&+\frac{2}{11}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-3})\Big)+\frac{1}{\alpha}\frac{\partial\mathbf{F}_k[\mathbf{g}]}{\partial g_k}+\frac{1}{\alpha}\frac{\partial\mathbf{F}_{k+1}[\mathbf{g}]}{\partial g_k}-\lambda\Delta_k,
\end{aligned}
$$

$$
\textbf{BDF4 Scheme}:\mathbf{G_k}=\frac{25}{12\tau}\Big(\frac{48}{25}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-1})-\frac{36}{25}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-2})
$$

$$+\frac{16}{25}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-3})-\frac{3}{25}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-4})\Big)+\frac{1}{\alpha}\frac{\partial \mathbf{F}_k[\mathbf{g}]}{\partial g_k}$$
$$+\frac{1}{\alpha}\frac{\partial \mathbf{F}_{k+1}[\mathbf{g}]}{\partial g_k}-\lambda\Delta_k, \tag{223}$$

$$\begin{aligned}\textbf{BDF5 Scheme}:\mathbf{G_k}=&\frac{137}{60\tau}\Big(\frac{300}{137}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-1})-\frac{300}{137}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-2})\\&+\frac{200}{137}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-3})-\frac{75}{137}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-4})\\&+\frac{12}{137}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-5})\Big)+\frac{1}{\alpha}\frac{\partial \mathbf{F}_k[\mathbf{g}]}{\partial g_k}+\frac{1}{\alpha}\frac{\partial \mathbf{F}_{k+1}[\mathbf{g}]}{\partial g_k}-\lambda\Delta_k,\end{aligned}$$

$$\begin{aligned}\textbf{BDF6 Scheme}:\mathbf{G_k}=&\frac{49}{20\tau}\Big(\frac{120}{49}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-1})-\frac{150}{49\tau}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-2})\\&+\frac{400}{147}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-3})-\frac{75}{49}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-4})\\&+\frac{24}{49}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-5})-\frac{10}{147}\sum_{j=1}^{n}a_{j,k}(g_j-g_j^{n-6})\Big)\\&+\frac{1}{\alpha}\frac{\partial \mathbf{F}_k[\mathbf{g}]}{\partial g_k}+\frac{1}{\alpha}\frac{\partial \mathbf{F}_{k+1}[\mathbf{g}]}{\partial g_k}-\lambda\Delta_k,\end{aligned}$$

where α represents two for the DLSS equation, eight for the Thin Film equation, $\frac{2(a+1)}{a^2}$, for $a\in\mathbb{R}\backslash\{0\}$ for equation (35) and four for equation (36).

Appendix G: Ingredients for the Critical Points of the Lagrangian and Hessian Matrices - See Section 7.4

We have the calculations already for the DLSS equation, from [17, p. 11]. We have the calculations for the following equations:

- **Thin Film equation (34)**: The computations for the Lagrangian critical points, when $k=1,\dots,n$:

$$\begin{aligned}\frac{\partial \mathbf{F}_k^t[\mathbf{g}]}{\partial g_k}&=\frac{1}{\delta_k}\left(\frac{1}{g_{k-1}^4}-\frac{1}{g_k^4}\right)+\frac{4}{\delta_k g_k^5}(g_k-g_{k-1})=\frac{1}{\delta_k g_{k-1}^4}-\frac{1}{\delta_k g_k^4}+\frac{4}{\delta_k g_k^4}-\frac{4g_{k-1}}{\delta_k g_k^5}\\&=\frac{1}{\delta_k}\left(\frac{3}{g_k^4}+\frac{1}{g_{k-1}^4}-\frac{4g_{k-1}}{g_k^5}\right),\end{aligned}$$

$$\begin{aligned}\frac{\partial \mathbf{F}_{k+1}^t[\mathbf{g}]}{\partial g_k}&=\frac{1}{\delta_{k+1}}\left(\frac{1}{g_{k+1}^4}-\frac{1}{g_k^4}\right)-\frac{4}{\delta_{k+1} g_k^5}(g_{k+1}-g_k)=\frac{1}{\delta_{k+1} g_{k+1}^4}-\frac{1}{\delta_{k+1} g_k^4}-\frac{4g_{k+1}}{\delta_{k+1} g_k^5}+\frac{4}{\delta_{k+1} g_k^4}\\&=\frac{1}{\delta_{k+1}}\left(\frac{3}{g_k^4}+\frac{1}{g_{k+1}^4}-\frac{4g_{k+1}}{g_k^5}\right),\end{aligned}$$

and for the Hessian matrix entrries, when $k = 1, \dots, n$:

$$\frac{\partial^2 \mathbf{F}_k^t[\mathbf{g}]}{\partial g_k^2} = \frac{1}{\delta_k} \frac{\partial}{\partial g_k} \left(\frac{3}{g_k^4} + \frac{1}{g_{k-1}^4} - \frac{4g_{k-1}}{g_k^5} \right) = \frac{1}{\delta_k} \left(\frac{20 g_{k-1}}{g_k^6} - \frac{12}{g_k^5} \right),$$

$$\frac{\partial^2 \mathbf{F}_{k+1}^t[\mathbf{g}]}{\partial g_k^2} = \frac{1}{\delta_{k+1}} \frac{\partial}{\partial g_k} \left(\frac{3}{g_k^4} + \frac{1}{g_{k+1}^4} - \frac{4g_{k+1}}{g_k^5} \right) = \frac{1}{\delta_{k+1}} \left(\frac{20 g_{k+1}}{g_k^6} - \frac{12}{g_k^5} \right),$$

$$\frac{\partial^2 \mathbf{F}_k^t[\mathbf{g}]}{\partial g_k g_{k-1}} = \frac{1}{\delta_k} \frac{\partial}{\partial g_{k-1}} \left(\frac{3}{g_k^4} + \frac{1}{g_{k-1}^4} - \frac{4g_{k-1}}{g_k^5} \right) = \frac{1}{\delta_k} \left(-\frac{4}{g_{k-1}^5} - \frac{4}{g_k^5} \right) = -\frac{4}{\delta_k} \left(\frac{1}{g_{k-1}^5} + \frac{1}{g_k^5} \right),$$

$$\frac{\partial^2 \mathbf{F}_{k+1}^t[\mathbf{g}]}{\partial g_k \partial g_{k+1}} = \frac{1}{\delta_{k+1}} \frac{\partial}{\partial g_{k+}} \left(\frac{3}{g_k^4} + \frac{1}{g_{k+1}^4} - \frac{4g_{k+1}}{g_k^5} \right) = \frac{1}{\delta_{k+1}} \left(-\frac{4}{g_{k+1}^5} - \frac{4}{g_k^5} \right) = -\frac{4}{\delta_{k+1}} \left(\frac{1}{g_{k+1}^5} + \frac{1}{g_k^5} \right).$$

- **Nonlinear fourth order equation (196)** The computations for the Lagrangian critical points, when $k = 1, \dots, n$:

$$\frac{\partial \mathbf{F}_k^v[\mathbf{g}]}{\partial g_k} = \frac{1}{\delta_k} \left(\frac{1}{g_{k-1}^{2(a+1)}} - \frac{1}{g_k^{2(a+1)}} \right) + \frac{2(a+1)}{g_k^{2a+3}} \frac{1}{\delta_k} (g_k - g_{k-1})$$
$$= \frac{1}{\delta_k} \left(\frac{2a+1}{g_k^{2(a+1)}} + \frac{1}{g_{k-1}^{2(a+1)}} - \frac{2(a+1) g_{k-1}}{g_k^{2a+3}} \right),$$

$$\frac{\partial \mathbf{F}_{k+1}^v[\mathbf{g}]}{\partial g_k} = \frac{1}{\delta_{k+1}} \left(\frac{2a+1}{g_k^{2(a+1)}} + \frac{1}{g_{k+1}^{2(a+1)}} - \frac{2(a+1) g_{k+1}}{g_k^{2a+3}} \right),$$

and for the Hessian matrix entrries, when $k = 1, \dots, n$:

$$\frac{\partial^2 \mathbf{F}_k^v[\mathbf{g}]}{\partial g_k^2} = \frac{1}{\delta_k} \frac{\partial}{\partial g_k} \left(\frac{2a+1}{g_k^{2(a+1)}} + \frac{1}{g_{k-1}^{2(a+1)}} - \frac{2(a+1) g_{k-1}}{g_k^{2a+3}} \right)$$
$$= \frac{1}{\delta_k} \left(\frac{2(a+1)(2a+3) g_{k-1}}{g_k^{2(a+2)}} - \frac{2(a+1)(2a+1)}{g_k^{2a+3}} \right),$$

$$\frac{\partial^2 \mathbf{F}_{k+1}^v[\mathbf{g}]}{\partial g_k^2} = \frac{1}{\delta_{k+1}} \frac{\partial}{\partial g_k} \left(\frac{2a+1}{g_k^{2(a+1)}} + \frac{1}{g_{k+1}^{2(a+1)}} - \frac{2(a+1) g_{k+1}}{g_k^{2a+3}} \right)$$
$$= \frac{1}{\delta_{k+1}} \left(\frac{2(a+1)(2a+3) g_{k+1}}{g_k^{2(a+2)}} - \frac{2(a+1)(2a+1)}{g_k^{2a+3}} \right),$$

$$\frac{\partial^2 \mathbf{F}_k^v[\mathbf{g}]}{\partial g_k g_{k-1}} = \frac{1}{\delta_k} \frac{\partial}{\partial g_{k-1}} \left(\frac{2a+1}{g_k^{2(a+1)}} + \frac{1}{g_{k-1}^{2(a+1)}} - \frac{2(a+1) g_{k-1}}{g_k^{2a+3}} \right) = -\frac{2(a+1)}{\delta_k} \left(\frac{1}{g_{k-1}^{2a+3}} + \frac{1}{g_k^{2a+3}} \right),$$

$$\frac{\partial^2 \mathbf{F}_{k+1}^v[\mathbf{g}]}{\partial g_k \partial g_{k+1}} = \frac{1}{\delta_{k+1}} \frac{\partial}{\partial g_{k+1}} \left(\frac{2a+1}{g_k^{2(a+1)}} + \frac{1}{g_{k+1}^{2(a+1)}} - \frac{2(a+1) g_{k+1}}{g_k^{2a+3}} \right) = -\frac{2(a+1)}{\delta_{k+1}} \left(\frac{1}{g_{k+1}^{2a+3}} + \frac{1}{g_k^{2a+3}} \right).$$

- **Nonlinear fourth order equation (198)** The computations for the Lagrangian critical points, when $k = 1, \dots, n$:

$$\frac{\partial \mathbf{F}_k^f[\mathbf{g}]}{\partial g_k} = \frac{1}{\delta_k} \left(\frac{1}{g_{k-1}^2} - \frac{1}{g_k^2} \right) + \frac{2}{\delta_k g_k^3} (g_k - g_{k-1}) = \frac{1}{\delta_k g_{k-1}^2} - \frac{1}{\delta_k g_k^2} + \frac{2}{\delta_k g_k^2} - \frac{2 g_{k-1}}{\delta_k g_k^3}$$
$$= \frac{1}{\delta_k} \left(\frac{1}{g_k^2} + \frac{1}{g_{k-1}^2} - \frac{2 g_{k-1}}{g_k^3} \right),$$

$$\frac{\partial \mathbf{F}^f_{k+1}[\mathbf{g}]}{\partial g_k} = \frac{1}{\delta_{k+1}} \left(\frac{1}{g_k^2} + \frac{1}{g_{k+1}^2} - \frac{2g_{k+1}}{g_k^3} \right),$$

and for the Hessian matrix entries, when $k = 1, \ldots, n$:

$$\frac{\partial^2 \mathbf{F}^f_k[\mathbf{g}]}{\partial g_k^2} = \frac{1}{\delta_k} \frac{\partial}{\partial g_k} \left(\frac{1}{g_k^2} + \frac{1}{g_{k-1}^2} - \frac{2g_{k-1}}{g_k^3} \right) = \frac{1}{\delta_k} \left(\frac{6g_{k-1}}{g_k^4} - \frac{2}{g_k^3} \right),$$

$$\frac{\partial^2 \mathbf{F}^f_{k+1}[\mathbf{g}]}{\partial g_k^2} = \frac{1}{\delta_{k+1}} \frac{\partial}{\partial g_k} \left(\frac{1}{g_k^2} + \frac{1}{g_{k+1}^2} - \frac{2g_{k+1}}{g_k^3} \right) = \frac{1}{\delta_{k+1}} \left(\frac{6g_{k+1}}{g_k^4} - \frac{2}{g_k^3} \right),$$

$$\frac{\partial^2 \mathbf{F}^f_k[\mathbf{g}]}{\partial g_k \partial g_{k-1}} = \frac{1}{\delta_k} \frac{\partial}{\partial g_{k-1}} \left(\frac{1}{g_k^2} + \frac{1}{g_{k-1}^2} - \frac{2g_{k-1}}{g_k^3} \right) = -\frac{2}{\delta_k} \left(\frac{1}{g_{k-1}^3} + \frac{1}{g_k^3} \right),$$

$$\frac{\partial^2 \mathbf{F}^f_{k+1}[\mathbf{g}]}{\partial g_k \partial g_{k+1}} = \frac{1}{\delta_{k+1}} \frac{\partial}{\partial g_{k+1}} \left(\frac{1}{g_k^2} + \frac{1}{g_{k+1}^2} - \frac{2g_{k+1}}{g_k^3} \right) = -\frac{2}{\delta_k} \left(\frac{1}{g_{k+1}^3} + \frac{1}{g_k^3} \right).$$

References

[1] R. Alexander. Diagonally implicit Runge-Kutta methods for stiff o.d.e.'s. *SIAM J. Numer. Anal.*, 14(6):1006–1021, 1977.

[2] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[3] L. Ambrosio, S. Lisini, and G. Savaré. Stability of flows associated to gradient vector fields and convergence of iterated transport maps. *Manuscripta Math.*, 121(1):1–50, 2006.

[4] U. M. Ascher, S. J. Ruuth, and R. J. Spiteri. Implicit-explicit Runge-Kutta methods for time-dependent partial differential equations. *Appl. Numer. Math.*, 25(2-3):151–167, 1997. Special issue on time integration (Amsterdam, 1996).

[5] E. Beretta, M. Bertsch, and R. Dal Passo. Nonnegative solutions of a fourth-order nonlinear degenerate parabolic equation. *Arch. Rational Mech. Anal.*, 129(2):175–200, 1995.

[6] A. Blanchet, V. Calvez, and J. A. Carrillo. Convergence of the mass-transport steepest descent scheme for the subcritical Patlak-Keller-Segel model. *SIAM J. Numer. Anal.*, 46(2):691–721, 2008.

[7] P. M. Bleher, J. L. Lebowitz, and Eugene R. Speer. Existence and positivity of solutions of a fourth-order nonlinear PDE describing interface fluctuations. *Comm. Pure Appl. Math.*, 47(7):923–942, 1994.

[8] M. Burger, J. A. Carrillo, and M. Wolfram. A mixed finite element method for nonlinear diffusion equations. *Kinet. Relat. Models*, 3(1):59–83, 2010.

[9] Luis A. Caffarelli and Robert J. McCann. Free boundaries in optimal transport and Monge-Ampère obstacle problems. *Ann. of Math. (2)*, 171(2):673–730, 2010.

[10] E. A. Carlen and S. Ulusoy. Localization, smoothness, and convergence to equilibrium for a thin film equation. *Discrete Contin. Dyn. Syst.*, 34(11):4537–4553, 2014.

[11] J. A. Carrillo, B. Düring, D. Matthes, and D. S. McCormick. A Lagrangian scheme for the solution of nonlinear diffusion equations using moving simplex meshes. *J. Sci. Comput.*, 75(3):1463–1499, 2018.

[12] J. A. Carrillo, R. J. McCann, and C. Villani. Contractions in the 2-Wasserstein length space and thermalization of granular media. *Arch. Ration. Mech. Anal.*, 179(2):217–263, 2006.

[13] J. A. Carrillo and J. S. Moll. Numerical simulation of diffusive and aggregation phenomena in nonlinear continuity equations by evolving diffeomorphisms. *SIAM J. Sci. Comput.*, 31(6):4305–4329, 2009/10.

[14] E. De Giorgi. New problems on minimizing movements. 29:81–98, 1993.

[15] B. Derrida, J. L. Lebowitz, E. R. Speer, and H. Spohn. Dynamics of an anchored Toom interface. *J. Phys. A*, 24(20):4805–4834, 1991.

[16] B. Derrida, J. L. Lebowitz, E. R. Speer, and H. Spohn. Fluctuations of a stationary nonequilibrium interface. *Phys. Rev. Lett.*, 67(2):165–168, 1991.

[17] B. Düring, D. Matthes, and J. P. Milišić. A gradient flow scheme for nonlinear fourth order equations. *Discrete Contin. Dyn. Syst. Ser. B*, 14(3):935–959, 2010.

[18] L. C. Evans, O. Savin, and W. Gangbo. Diffeomorphisms and nonlinear heat flows. *SIAM J. Math. Anal.*, 37(3):737–751, 2005.

[19] T. O. Gallouët and L. Monsaingeon. A JKO splitting scheme for Kantorovich-Fisher-Rao gradient flows. *SIAM J. Math. Anal.*, 49(2):1100–1130, 2017.

[20] U. Gianazza, G. Savaré, and G. Toscani. The Wasserstein gradient flow of the Fisher information and the quantum drift-diffusion equation. *Arch. Ration. Mech. Anal.*, 194(1):133–220, 2009.

[21] L. Gosse and G. Toscani. Identification of asymptotic decay to self-similarity for one-dimensional filtration equations. *SIAM J. Numer. Anal.*, 43(6):2590–2606, 2006.

[22] G. Grün and M. Rumpf. Nonnegativity preserving convergent schemes for the thin film equation. *Numer. Math.*, 87(1):113–152, 2000.

[23] E. Hairer, S. P. Nø rsett, and G. Wanner. *Solving ordinary differential equations. I*, volume 8 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1987. Nonstiff problems.

[24] E. Hairer and G. Wanner. *Solving ordinary differential equations. II*, volume 14 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 1996. Stiff and differential-algebraic problems.

[25] E. Hairer and G. Wanner. *Solving ordinary differential equations. II*, volume 14 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2010. Stiff and differential-algebraic problems, Second revised edition, paperback.

[26] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.

[27] A. Jüngel and R. Pinnau. Global nonnegative solutions of a nonlinear fourth-order parabolic equation for quantum systems. *SIAM J. Math. Anal.*, 32(4):760–777, 2000.

[28] A. Jüngel and R. Pinnau. A positivity-preserving numerical scheme for a nonlinear fourth order parabolic system. *SIAM J. Numer. Anal.*, 39(2):385–406, 2001.

[29] A. Jüngel and I. Violet. First-order entropies for the Derrida-Lebowitz-Speer-Spohn equation. *Discrete Contin. Dyn. Syst. Ser. B*, 8(4):861–877, 2007.

[30] E. Kamalinejad. On local well-posedness of the thin-film equation via the Wasserstein gradient flow. *Calc. Var. Partial Differential Equations*, 52(3-4):547–564, 2015.

[31] J. Kim. A numerical method for the Cahn-Hilliard equation with a variable mobility. *Commun. Nonlinear Sci. Numer. Simul.*, 12(8):1560–1571, 2007.

[32] G. Legendre and G. Turinici. Second-order in time schemes for gradient flows in Wasserstein and geodesic metric spaces. *C. R. Math. Acad. Sci. Paris*, 355(3):345–353, 2017.

[33] D. Matthes, R. J. McCann, and G. Savaré. A family of nonlinear fourth order equations of gradient flow type. *Comm. Partial Differential Equations*, 34(10-12):1352–1397, 2009.

[34] D. Matthes and H. Osberger. A convergent Lagrangian discretization for a nonlinear fourth-order equation. *Found. Comput. Math.*, 17(1):73–126, 2017.

[35] D. Matthes and S. Plazotta. A variational formulation of the BDF2 method for metric gradient flows. *ESAIM Math. Model. Numer. Anal.*, 53(1):145–172, 2019.

[36] F. Mendivil. Computing the Monge-Kantorovich distance. *Comput. Appl. Math.*, 36(3):1389–1402, 2017.

[37] G. Monge. *Géométrie descriptive*. Éditions Jacques Gabay, Sceaux, 1989. Reprint of the 1799 original.

[38] M. Ó. Searcóid. *Metric spaces*. Springer Undergraduate Mathematics Series. Springer-Verlag London, Ltd., London, 2007.

[39] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.

[40] M. A. Peletier. Variational modelling: Energies, gradient flows, and large deviations. Lecture notes available on arXiv:1402.1990.

[41] F. Santambrogio. *Optimal transport for applied mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling.

[42] F. Santambrogio. Euclidean, metric, and Wasserstein gradient flows: an overview. *Bull. Math. Sci.*, 7(1):87–154, 2017.

[43] G. Savaré. Gradient flows, optimal transport, and evolution pde's: a quick introduction to optimal transport, 2010. Slides of talk given at GNFM Summer School, Ravello, September 13–18, 2010.

[44] G. Savaré. Gradient flows, optimal transport, and evolution pde's: Gradient flows and wasserstein distance, 2010. Slides of talk given at GNFM Summer School, Ravello, September 13–18, 2010.

[45] J. Sundnes, G. T. Lines, and A. Tveito. Efficient solution of ordinary differential equations modeling electrical activity in cardiac cells. *Math. Biosci.*, 172(2):55–72, 2001.

[46] J. V. Neumann. Zur Theorie der Gesellschaftsspiele. *Math. Ann.*, 100(1):295–320, 1928.

[47] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[48] M. Westdickenberg and J. Wilkening. Variational particle schemes for the porous medium equation and for the system of isentropic Euler equations. *M2AN Math. Model. Numer. Anal.*, 44(1):133–166, 2010.