

**Entropy Foundations for Stabilized Finite Element Isogeometric Methods  
Energy Dissipation, Variational Multiscale Analysis, Variation Entropy, Discontinuity  
Capturing and Free Surface Flows**

ten Eikelder, M.F.P.

**DOI**

[10.4233/uuid:04152eb5-52f3-49e9-acc9-2a88ed2ac98c](https://doi.org/10.4233/uuid:04152eb5-52f3-49e9-acc9-2a88ed2ac98c)

**Publication date**

2020

**Document Version**

Final published version

**Citation (APA)**

ten Eikelder, M. F. P. (2020). *Entropy Foundations for Stabilized Finite Element Isogeometric Methods: Energy Dissipation, Variational Multiscale Analysis, Variation Entropy, Discontinuity Capturing and Free Surface Flows*. [Dissertation (TU Delft), Delft University of Technology].  
<https://doi.org/10.4233/uuid:04152eb5-52f3-49e9-acc9-2a88ed2ac98c>

**Important note**

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.

**ENTROPY FOUNDATIONS FOR STABILIZED  
FINITE ELEMENT ISOGEOMETRIC METHODS**

*Energy Dissipation, Variational Multiscale Analysis, Variation Entropy,  
Discontinuity Capturing and Free Surface Flows*

MARCO FREDERIK PETRUS TEN EIKELDER



**ENTROPY FOUNDATIONS FOR STABILIZED  
FINITE ELEMENT ISOGEOMETRIC METHODS**

*Energy Dissipation, Variational Multiscale Analysis, Variation Entropy,  
Discontinuity Capturing and Free Surface Flows*

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR AAN DE  
TECHNISCHE UNIVERSITEIT DELFT,  
OP GEZAG VAN DE RECTOR MAGNIFICUS  
PROF.DR.IR. T.H.J.J. VAN DER HAGEN,  
VOORZITTER VAN HET COLLEGE VOOR PROMOTIES,  
IN HET OPENBAAR TE VERDEDIGEN OP  
WOENSDAG, 7 OKTOBER, 2020 OM 15:00 UUR

DOOR

MARCO FREDERIK PETRUS TEN EIKELDER

MASTER OF SCIENCE IN INDUSTRIAL AND APPLIED MATHEMATICS  
EINDHOVEN UNIVERSITY OF TECHNOLOGY, THE NETHERLANDS  
GEBOREN TE EINDHOVEN

Dit proefschrift is goedgekeurd door de promotoren.

Samenstelling promotiecommissie bestaat uit:

voorzitter: Rector Magnificus  
promotor: prof. dr. ir. R.H.M. Huijsmans  
copromotor: dr. ir. I. Akkerman

onafhankelijke leden:

prof. dr. ir.	Y. Bazilevs	Brown University
prof. dr. ir.	D. Schillinger	Leibniz Universität Hannover
dr. ir.	K.G. van der Zee	University of Nottingham
prof. dr. ir.	E.H. van Brummelen	Eindhoven University of Technology
prof. dr. ir.	C.W. Oosterlee	Delft University of Technology
prof. dr. ir.	B.J. Boersma	Delft University of Technology (reservelid)

Het onderzoek dat in dit proefschrift wordt beschreven is uitgevoerd aan het departement Maritime and Transport Technology van de Technische Universiteit Delft.

**ENTROPY FOUNDATIONS FOR STABILIZED  
FINITE ELEMENT ISOGEOMETRIC METHODS**

*Energy Dissipation, Variational Multiscale Analysis, Variation Entropy,  
Discontinuity Capturing and Free Surface Flows*

MARCO FREDERIK PETRUS  
TEN EIKELDER

Printed by: Ipskamp printing  
Cover design: Marco ten Eikelder

A catalogue record is available from the Delft University of Technology  
Library.

ISBN/EAN: 978-94-6384-164-1  
Copyright © 2020 by Marco ten Eikelder

*Voor mijn ouders*



---

## PREFACE

---

Before diving into the wonderful world of stabilized finite element methods, I would like to point out the following. A large part of my dissertation is reproduced from scientific articles I have written over the last four years. Each chapter in the body of this work contains a reference to the corresponding article. The different chapters, organized in three parts, deal with various aspects of the overarching theme of the dissertation. Enforcing a consistent notation would upset conventional notation at some places. To avoid this I have chosen to only use a coherent notation *within* each chapter. It goes without saying that all employed definitions are clarified within their own context.

*Marco ten Eikelder*  
*Delft, May 2020*



---

# CONTENTS

---

PREFACE	1
1 INTRODUCTION	7
1.1 Finite element methods	8
1.2 Stabilized finite element methods	10
1.3 The variational multiscale method	11
1.4 Discontinuity capturing	12
1.5 Free-surface flow modeling	14
1.6 Research objective	15
1.7 Outline	16
<b>I ENERGY-DISSIPATIVE MULTISCALE FORMULATIONS</b>	
2 CORRECT ENERGY EVOLUTION OF STABILIZED CONVECTION-DIFFUSION	21
2.1 Introduction	22
2.1.1 Dynamic small-scales	22
2.1.2 Orthogonal small-scales in VMS	22
2.1.3 Isogeometric analysis framework	22
2.1.4 Context	23
2.1.5 Outline	23
2.2 The continuous convection-diffusion equation	24
2.2.1 Strong formulation	24
2.2.2 Weak formulation	25
2.2.3 Global energy evolution	25
2.2.4 Localized energy evolution	26
2.3 Energy evolution of the variational multiscale approach	28
2.3.1 The multiscale split	28
2.3.2 The VMS numerical formulation	29
2.3.3 Local energy evolution of the VMS formulation	31
2.4 Toward a stabilized formulation with correct-energy behavior	32
2.4.1 Design condition	33
2.4.2 The variational multiscale method with dynamic small-scales	33
2.4.3 Orthogonality between the large-scales and the small-scales	34
2.4.4 Consistent SUPG with dynamic small-scales	34
2.4.5 Inconsistent SUPG with dynamic small-scales	34
2.4.6 GLS with dynamic small-scales (GLSD)	35
2.5 Back to a variational multiscale formulation	35
2.5.1 The small-scale solution space	36
2.5.2 Enforced orthogonality with a Lagrange multiplier (DO formulation)	36

2.5.3	Local energy evolution of the formulation with enforced orthogonality . . . . .	37
2.6	Temporal-integration . . . . .	37
2.6.1	The generalized- $\alpha$ time integrator . . . . .	37
2.6.2	Time-integration of the small-scales . . . . .	38
2.6.3	Proper energy evolution . . . . .	39
2.7	Numerical verification . . . . .	41
2.7.1	Model problem description . . . . .	41
2.7.2	Overall energy behavior . . . . .	42
2.7.3	Energy dissipation by the small-scales . . . . .	43
2.7.4	Temporal-term . . . . .	44
2.7.5	Orthogonality-term . . . . .	45
2.8	Conclusions . . . . .	48
3	CORRECT ENERGY EVOLUTION OF STABILIZED INCOMPRESSIBLE NAVIER-STOKES . . . . .	49
3.1	Introduction . . . . .	50
3.1.1	Contributions of this work . . . . .	50
3.1.2	Context . . . . .	50
3.1.3	Outline . . . . .	51
3.2	The continuous incompressible Navier–Stokes equations . . . . .	52
3.2.1	Strong formulation . . . . .	52
3.2.2	Weak formulation . . . . .	53
3.2.3	Global energy evolution . . . . .	54
3.2.4	Local energy evolution . . . . .	55
3.3	Energy evolution of VMS with dynamic small-scales . . . . .	56
3.3.1	The multiscale split . . . . .	57
3.3.2	Dynamic small-scales . . . . .	57
3.3.3	Local energy evolution of the VMSE form . . . . .	58
3.4	Toward a stabilized formulation with correct energy behavior . . . . .	60
3.4.1	Design condition . . . . .	60
3.4.2	Skew-symmetric form . . . . .	60
3.4.3	Stokes projector . . . . .	61
3.4.4	Divergence-free small-scales . . . . .	63
3.4.5	$H_0^1$ -orthogonal small-scales . . . . .	64
3.4.6	Local energy backscatter . . . . .	65
3.4.7	Time-discrete energy behavior . . . . .	65
3.5	Conservation properties . . . . .	66
3.5.1	Continuity . . . . .	66
3.5.2	Linear momentum . . . . .	66
3.5.3	Angular momentum . . . . .	67
3.6	Numerical test case . . . . .	67
3.7	Conclusions . . . . .	71
3.A	GLS formulation with dynamic divergence-free small-scales . . . . .	73
3.B	Definition dynamic stabilization parameter . . . . .	73

## II A FRAMEWORK FOR DISCONTINUITY CAPTURING METHODS

4	VARIATION ENTROPY: A CONTINUOUS LOCAL GENERALIZATION OF THE TVD PROPERTY USING ENTROPY PRINCIPLES.	77
4.1	Introduction . . . . .	78
4.2	Entropy solutions in the classical sense . . . . .	81
4.3	Variation entropy solutions . . . . .	82
4.3.1	The concept . . . . .	82
4.3.2	Characterization of variation entropies . . . . .	84
4.3.3	Variation entropy-variation entropy flux pairs . . . . .	88
4.3.4	Augmented conservation laws . . . . .	90
4.4	Selection of the variation entropy . . . . .	91
4.4.1	Some examples . . . . .	91
4.4.2	Objectivity . . . . .	93
4.4.3	Regularization of 2-norm variation entropy . . . . .	95
4.5	Conclusion and discussion . . . . .	100
4.A	The 3-dimensional version of the convexity condition . . . . .	103
5	A THEORETICAL FRAMEWORK FOR DISCONTINUITY CAPTURING	107
5.1	Introduction . . . . .	108
5.2	Entropy solutions . . . . .	110
5.2.1	The classical entropy . . . . .	110
5.2.2	The variation entropy . . . . .	112
5.3	The VMS-variation entropy framework for discontinuity capturing methods . . . . .	114
5.3.1	Starting point . . . . .	114
5.3.2	Mesh representation and geometrical mapping . . . . .	115
5.3.3	The multiscale split . . . . .	116
5.3.4	A standard optimality projector . . . . .	117
5.3.5	A variation entropy optimality projector . . . . .	119
5.3.6	Small-scale model variation entropy . . . . .	122
5.3.7	Variation entropy viscosity . . . . .	123
5.3.8	Diffusion matrices . . . . .	126
5.3.9	Complete semi-discrete formulations . . . . .	127
5.3.10	The convection-diffusion problem . . . . .	128
5.3.11	Connection to the $YZ\beta$ method . . . . .	129
5.4	Numerical comparison . . . . .	131
5.4.1	Convergence for smooth solutions . . . . .	131
5.4.2	Buckley-Leverett with gravity . . . . .	132
5.4.3	KPP rotating wave . . . . .	136
5.5	Conclusions . . . . .	140
5.A	An alternative optimality projector . . . . .	143

## III AN ENERGY-DISSIPATIVE METHOD FOR FREE-SURFACE FLOW

6	AN ENERGY-DISSIPATIVE LEVEL-SET METHOD FOR THE INCOMPRESSIBLE TWO-PHASE NAVIER-STOKES EQUATIONS WITH SURFACE TENSION	149
6.1	Introduction . . . . .	150
6.1.1	Free-surface flow modeling . . . . .	150
6.1.2	Surface tension . . . . .	150

6.1.3	Energetic stability . . . . .	151
6.1.4	This work . . . . .	151
6.1.5	Structure of this chapter . . . . .	152
6.2	Sharp-interface formulation . . . . .	152
6.2.1	Governing equations . . . . .	152
6.2.2	Energy evolution . . . . .	154
6.2.3	Weak formulation . . . . .	156
6.3	Regularized-interface level-set model . . . . .	156
6.3.1	Sharp-interface level-set formulation . . . . .	157
6.3.2	Non-dimensionalization . . . . .	158
6.3.3	Regularization . . . . .	159
6.3.4	Energy evolution . . . . .	160
6.4	Energy-dissipative formulation . . . . .	165
6.4.1	Functional entropy variables . . . . .	165
6.4.2	Modified formulation . . . . .	168
6.5	Energy-dissipative spatial discretization . . . . .	171
6.5.1	Notation . . . . .	172
6.5.2	Stabilization . . . . .	172
6.5.3	Semi-discrete formulation . . . . .	174
6.6	Energy-dissipative temporal discretization . . . . .	176
6.6.1	Notation . . . . .	177
6.6.2	Identification energy evolution terms . . . . .	177
6.6.3	Discretization other terms . . . . .	182
6.6.4	Fully-discrete energy-dissipative method . . . . .	183
6.7	Numerical experiments . . . . .	186
6.7.1	Static spherical droplet . . . . .	186
6.7.2	Droplet coalescence 2D . . . . .	190
6.7.3	Droplet coalescence 3D . . . . .	195
6.8	Conclusion . . . . .	195
6.A	Equivalence surface tension models . . . . .	199
6.A.1	Sharp-interface model . . . . .	199
6.A.2	Regularized-interface level-set model . . . . .	200
6.B	Energy evolution midpoint level-set discretization . . . . .	201
7	CONCLUSIONS AND FUTURE WORK . . . . .	207
7.1	Conclusions . . . . .	207
7.1.1	Part I: Energy-dissipative multiscale formulations . . . . .	207
7.1.2	Part II: A framework for discontinuity capturing methods . . . . .	209
7.1.3	Part III: An energy-dissipative method for free surface flow . . . . .	210
7.2	Future work . . . . .	211
	BIBLIOGRAPHY . . . . .	213
	SUMMARY . . . . .	229
	SAMENVATTING . . . . .	231
	PUBLICATIONS . . . . .	235
	ACKNOWLEDGMENTS . . . . .	239
	CURRICULUM VITAE . . . . .	241

---

## INTRODUCTION

---

It is widely recognized that computer modeling and simulation form the third pillar of science, alongside with theory and experimentation. The value of numerical methods lies in the fact that they are generally faster and cheaper to execute than physical experiments. Equally important, simulation allows to obtain the whole picture of the problem, whereas experimental techniques are typically restricted to a certain piece. To perform a numerical computation one needs to identify all basic physical phenomena *a priori*. This directly validates the understanding of the problem. Still, experiments are required to draw our attention to these phenomena.

The quality of simulations of complex physical phenomena in fluid mechanics, such as free surface flows, has significantly improved over the last decades. This growth comes from the increase and improved quality of computational resources and the advancement of the numerical algorithms. Yet, in order to tackle large and complex problems in science and engineering, significant progress needs to be made on both the computational resources and numerical methods side.

This thesis concerns the construction of new numerical methods for the simulation of complex flow phenomena. When numerically solving problems in fluid mechanics one wishes to find the quantities (velocity, pressure, density, etc.) that have a meaningful physical interpretation. These physically relevant solutions are obtained by imposing a certain stability condition, often known as the entropy condition. This condition ensures that small numerical errors in the physical quantities do not magnify. A numerical method that solves the corresponding mathematical model does unfortunately not always inherit the stability property of the physical system. In such algorithms a small numerical error or perturbation may lead to a wildly different outcome. This unwanted behavior was the motivation for the design of so-called *stabilized methods*. As the name suggests, these methods aim to restore stability. Even though the quality of numerical solutions obtained by using stabilized methods greatly improves upon standard methods, there is still room for improvement. In the construction of stabilized methods for complex problems typically several approximations occur that are not consistent with the underlying physics. As a result the obtained numerical solutions may not be (provably) stable. This thesis is centered around this issue; it focuses on the development of stable

numerical methods that *originate* from the physical system such that the obtained numerical solutions closely *resemble* the physics.

Among the most challenging problems for numerical methods in fluid mechanics are those that involve a free surface. A typical example in the maritime field is a water-air flow such as the dam break problem, see e.g., Figure 1.1.

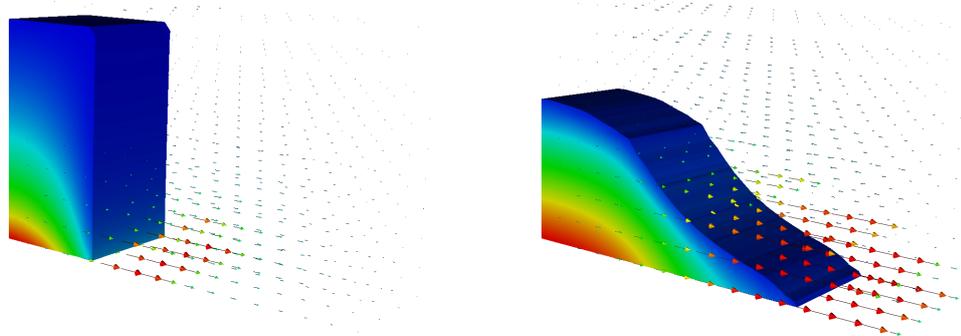


Figure 1.1: Snapshots of a dam break problem. The coloring is based on the pressure field.

In such a problem several aspects come into play: multiple scales (turbulence), violent disturbances and the interface separating the fluids. A numerical method tailored for such a problem ideally treats each of these aspects in a physically correct manner. In this dissertation we look into these three topics.

We choose to employ the (stabilized) finite element and isogeometric analysis method. These are mature approximation methods when it comes to solving mathematical models describing physical phenomena (e.g., fluid and solid mechanics) on complex domains. The methods are particularly suitable when multiscale and multi-physics play a role in the problem. Applying these solution strategies on complex problems is a challenging task. We focus on approximating the mathematical problem in such a way that multiscale effects are properly accounted for and that the numerical approximation satisfies certain stability properties.

In the remainder of this chapter we provide an overview of finite element and isogeometric methods with corresponding stabilization techniques in Sections 1.1-1.4 and provide an overview of free surface modeling techniques in Section 1.5. We present several issues and open problems of the methods. We close with the thesis objective and outline in Section 1.6 and 1.7.

## 1.1 FINITE ELEMENT METHODS

The key ideas behind finite element methods can be traced back to the beginning of the twentieth century. Important contributions of that period include the works of B. Galerkin, W. Ritz, A. Hrennikoff [96] and R. Courant [52]. The method originated in the field of civil and aeronautical engineering where problems in structural analysis appeared.

The finite element method is a methodology to approximate the solution of a partial differential equation (PDE) and may roughly be summarized as follows. As

the name suggests, the key step is to subdivide the corresponding physical domain into a finite number of elements. This is called the discretization of the problem. Then on each of the elements so-called basis functions are defined. The *finite element solution* is an approximation of the solution of the PDE. It is a linear combination of the basis functions (which satisfy certain imposed boundary conditions). This linear combination is determined by solving a system of equations that results when discretizing a weak formulation of the problem. In Figure 1.2 we show a solution profile with its finite element approximation.

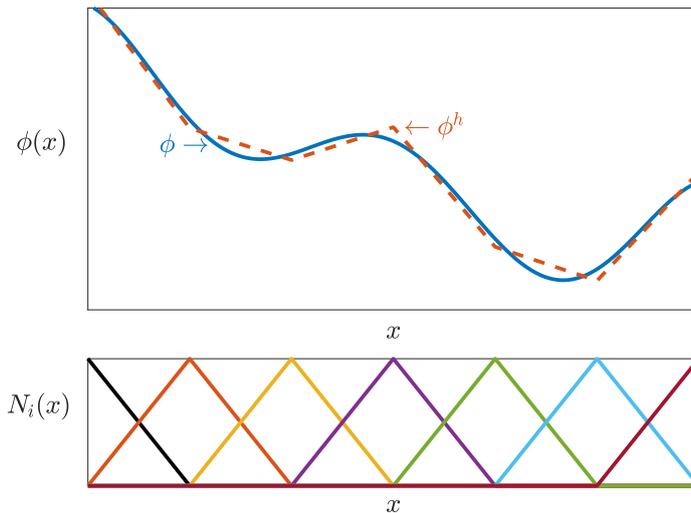


Figure 1.2: Top: A solution profile  $\phi = \phi(x)$  with its finite element approximation  $\phi^h$ . Bottom: The linear basis functions  $N_i = N_i(x)$  that the numerical solution is built from.

The choice of finite element basis functions lies to some extent with the user. For example, one may select basis functions that suit the geometry of the problem. This idea led to the introduction of the isogeometric finite element method. Isogeometric analysis (IGA), introduced by Hughes et al. [101], uses the NURBS (non-uniform rational B-splines) basis functions from Computer-Aided Design (CAD) directly into a finite element framework. In this way the NURBS surfaces in IGA match with the *exact* CAD geometry. As a result, mesh refinement procedures do not require communication with the CAD geometry. Isogeometric methods are efficient and accurate solution strategies [66]. Isogeometric analysis directly became a valuable tool in several fields of engineering, including fluid flow computations [5, 68, 69], fluid-structure interaction [18, 20, 23], shape optimization [42, 162, 198], wind turbines [25, 92], heart-valve flow problems [21, 22, 97, 98, 125, 205], electromagnetics [29, 36], shells [26, 128, 129] and structural mechanics [51, 149, 191]. The reader may consult [50, 152] for more (implementational) details about IGA. We illustrate in Figure 1.3 a solution profile and its isogeometric discretization.

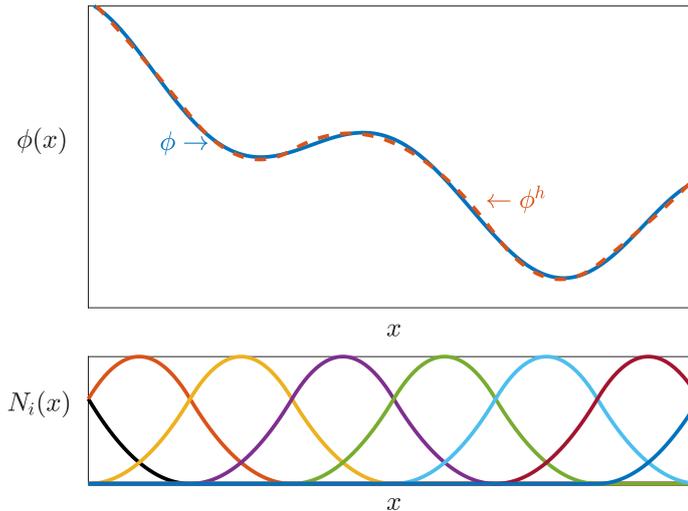


Figure 1.3: Top: A solution profile  $\phi = \phi(x)$  with its isogeometric approximation  $\phi^h$ . Bottom: The second-order NURBS basis functions  $N_i = N_i(x)$  that the numerical solution is built from.

Throughout this thesis we employ continuous finite element methods indicating that the solution is continuous on the entire domain. The natural counterpart are the discontinuous (Galerkin) methods [7, 45, 164] in which the solution is only piecewise continuous.

## 1.2 STABILIZED FINITE ELEMENT METHODS

The procedure to obtain the finite element solution given in Section 1.1 is known as the standard *Galerkin* finite element method. This is an often used approximation method for problems in structural mechanics. For a typical problem in fluid dynamics the standard Galerkin finite element approximation is often not satisfactory; obtained solutions are oscillatory indicating that the method is unstable. Oscillatory solutions typically appear for high Péclet or Reynolds numbers in convection/advection-dominant problems<sup>1</sup> or result from pressure instabilities in mixed problems (e.g., Stokes flow). In the second case the discretization does not satisfy the so-called Banach-Nečas-Babuška (BNB) condition, also known as inf-sup condition, which guarantees uniqueness of the solution. In case of mixed problems, the more general BNB condition simplifies to the Babuška-Brezzi condition [9, 32, 55].

The origin of instabilities in convection-dominant problems is best described in the seminal paper by Brooks and Hughes [33]. These instabilities gave rise to the development of methods that yield non-oscillatory solutions, the so-called *stabilized finite element methods*. These methods are formed by augmenting the standard

<sup>1</sup> In this thesis we do not distinguish between the term ‘convection’ and ‘advection’. We interchangeably use these words to indicate the movement of a quantity, fluid or material.

Galerkin method with consistent, mesh-dependent terms that improve the numerical stability. In [33] the first residual-based stabilized finite element method, known by the name SUPG (Streamline upwind Petrov-Galerkin), is presented. This is the starting point for the development of several stabilized methods. In particular, finite element methods for fluid dynamics experienced a boost with the series of ‘a new finite element method for computational fluid dynamics’ by Hughes and coworkers in 1986-1991 [104–108, 110–112, 170, 171]. In this series the SUPG method is extended to multi-dimensional systems [108, 111]. The concept of stabilized methods is also generalized to Stokes flow [104, 105] which allows to circumvent the Babuška-Brezzi condition. It also includes the construction of stabilized methods for compressible flows [107, 171]. Additionally, a significant contribution was the development of the Galerkin/least-squares method [106, 170]. As the name indicates, the stabilizing term in this method is a least-squares form of the residual of the equation. The method is conceptually easier than SUPG and can be understood as a wider class of interpolation methods compared to standard Galerkin methods.

### 1.3 THE VARIATIONAL MULTISCALE METHOD

The initiation of another crucial stabilized method started in 1989 with the paper of Douglas and Wang [56]. The work is a non-symmetric modification of Hughes and Franca [104] with optimal error estimates. This idea is directly applied to advective-diffuse model problem in Franca et al. [74] and to the incompressible Navier-Stokes equations in [73]. This method was introduced as the *unusually stabilized method*.

Although the stabilized methods introduced up to this point show significantly better behavior than the standard Galerkin methods, they are rather *ad hoc*: a rigorous derivation was virtually absent. Rigorous mathematical proofs show however the validity of the methods. To arrive at a stabilized method the standard Galerkin method is augmented with stabilizing terms which are purely a cure for the appearing oscillations. In fact, even though the Galerkin method is regarded as the father of finite element methods and is optimal in the energy norm, one might argue that Galerkin itself is not a natural way to discretize a PDE. Defining the approximation space via arbitrarily sampling the solution space does not ensure a good representation of the numerical solution. This explains the loss of stability for advection-dominated problems.

The fact that this may feel somewhat dissatisfying has led to the search of a rationale behind stabilized methods. In 1995 Hughes [100] recognized the presence of the oscillations as the result of ‘missing scales’. This induced the *variational multiscale* (VMS) method, for which the bigger picture was presented in 1998 [103] and the rigorous mathematical analysis in 2007 [113]. This method coincides with the unusually stabilized method which seemed to be an *ad hoc* procedure at its introduction. The VMS approach is a procedure of designing numerical methods that are capable of handling multiscale effects. The name multiscale is reflected back into the crux of the method, which is the splitting of the solution space into large- and small-scales (also termed as coarse- and fine-scales). The large-scale component is that part of the solution living on the mesh. On the other hand, the small-scales should be understood as the part that does not fit onto the mesh. These scales are not discretized but an approximate model is employed. We illustrate this in Figure

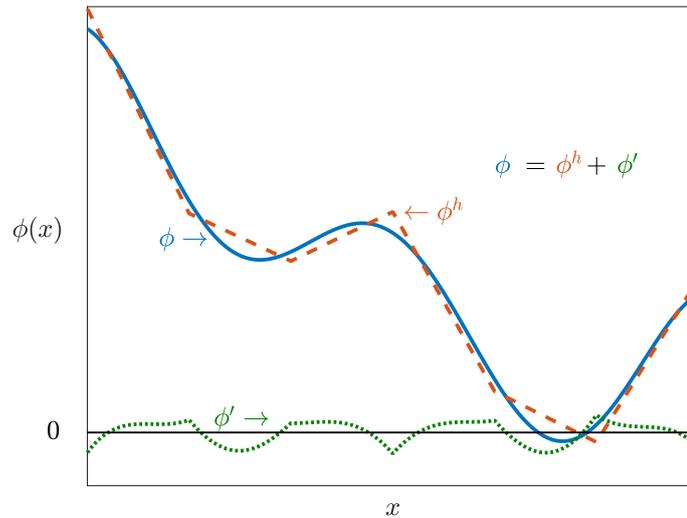


Figure 1.4: The multiscale split of a solution  $\phi = \phi(x)$  into large- and small-scales, denoted  $\phi^h$  and  $\phi'$  respectively.

1.4. The small-scale component is exactly that part of the solution that is discarded when using a Galerkin approximation.

In the situation that the true solution lies in the large-scale solution space, we wish to find *that* solution as our approximate solution. This implies that the splitting procedure is governed by a projection operator. Once the projector has been selected it remains to determine the small-scales. This is formally done via the (fine-scale) Green's function. Unfortunately, for many PDEs no closed form for the Green's function exists. This is where the approximation comes into play. The small-scales are modeled in a residual-based fashion in terms of the large-scales. The resulting expression is then substituted in the large-scale problem and thus the small-scale effect is accounted for when determining the numerical solution.

The rigorous derivation of the VMS method [113] is presented for linear PDEs. When the problem under consideration is non-linear (e.g., the Navier-Stokes equations) the methodology requires several approximation steps [17]. Despite, the resulting method is a powerful way to simulate turbulent flows such as the one depicted in Figure 1.5.

Furthermore, one can show that for symmetric coercive problems the multiscale approach reduces to the standard Galerkin method [178]. This is the underlying reasoning why Galerkin works well for many problems in structural mechanics.

#### 1.4 DISCONTINUITY CAPTURING

The above discussed stabilized methods yield accurate oscillation-free solutions in absence of 'strong shocks'. In presence of discontinuities (e.g., Riemann problems) stabilized methods are not sufficient to capture the sharp layers. Additional to the diffusion in the streamwind direction provided by SUPG, crosswind diffusion is also

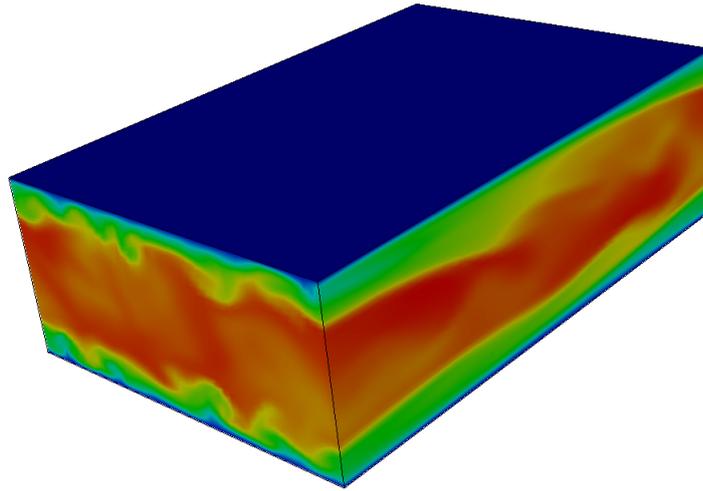


Figure 1.5: A snapshot of a VMS simulation of turbulent channel flow.

required when dealing with sharp layers. As a result, researchers have attempted to rectify this discrepancy via the use of so-called discontinuity capturing (DC) operators [121, 122]. The discontinuity capturing operator is a consistent mesh-dependent term that is added to the stabilized method. It can be understood as an artificial viscosity of which its specific form is to a large extent undetermined. Thus, this solution strategy has a large *ad hoc* component. As a result, a large number of solutions has been proposed, see e.g., [19, 54, 77, 110, 112, 138, 139, 185, 189, 190]. Some of these methods present some motivation of the proposed DC but none is provided with a rigorous derivation. Despite that a foundation is lacking, discontinuity capturing is a useful technique that allows the practitioner to study more realistic problems. In Figure 1.6 we demonstrate that the overall solution quality of the problem may significantly increase when a discontinuity capturing device is employed.

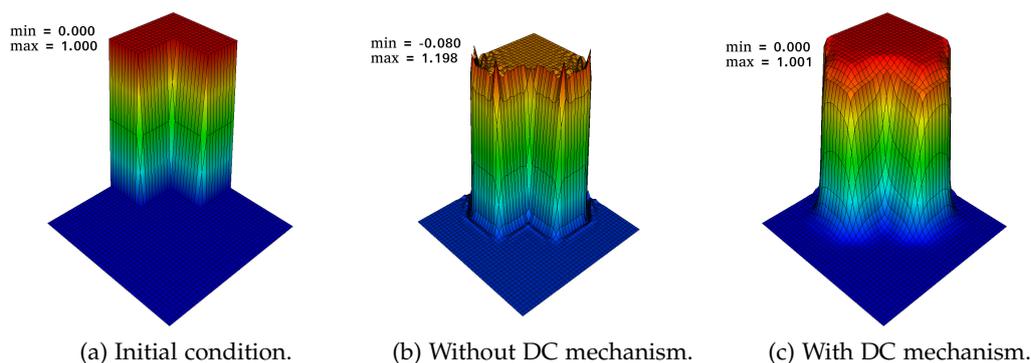


Figure 1.6: Convection skew to the mesh for an internal layer problem: (a) depicts the initial configuration, (b) a snapshot of the solution obtained with standard stabilized method without discontinuity capturing and (c) a snapshot of the solution when a discontinuity capturing device is employed.

An important observation is that the pure stabilized methods such as SUPG, GLS and VMS are linear when the underlying problem is linear. In contrast, the discontinuity capturing term is always non-linear, even for linear problems. As a consequence, the resulting system of equations is more time-consuming to solve.

### 1.5 FREE-SURFACE FLOW MODELING

Representing the motion of an interface is a significant component when dealing with free surface flow problems. Apart from the previously mentioned dam-break problem, free surfaces occur in bubbles, drops, liquefied natural gas (LNG) tanks, combustion, crystal growth, spinodal decomposition, solidification and many more. At each of these interfaces, certain boundary conditions apply which specify the interface location. Typically, these conditions describe the relation of the quantities of interest across the interface. Examples include an equilibrium of velocities or pressures, or a certain jump in the stress prescribed by surface tension. An important aspect of the interface is the density jump, which is large for water-air flow problems. This indicates that small over- or undershoots in a simulations can result in a negative density which renders the results useless. Thus, it is essential to either allow only for very small oscillations or, better, guarantee the satisfaction of the maximum principle to *ab initio* exclude oscillations.

Numerical methods built to simulate these phenomena basically come in two flavors. The first category is composed of methods in which the mesh takes care of the interface location. These *interface tracking* methods [123, 161, 193, 194] provide a way to explicitly handle the interface evolution. Although these methods are generally accurate, topological changes such as merging or pinching require substantial extra effort. As a result, interface tracking methods are not suitable for problems in which a large amount of topological changes occurs. For these problems the so-called *interface capturing* methods are often employed.

Interface capturing methods provide the means to naturally handle interface motion. In these procedures the numerical model contains a variable that describes the interface. The interface capturing methods may in their turn roughly be subdivided into phase field methods, volume-of-fluid methods and level-set methods. The relation between these methods is described in [197]. Another class of methods are the particle-based schemes. In this review we exclude these methods as these fall behind in terms in popularity, as compared to the other classes of methods mentioned above [64].

The phase-field models are diffuse-interface models which implies that the interface is described by a thin transition layer. In this region the physical quantities (such as density) smoothly transition from one side of the interface to the other. Sharp-interface models may be recovered for a vanishing interface thickness [78]. Well-known phase-field models are the Navier-Stokes-Korteweg equations [134] and the Cahn-Hilliard model [38, 39]. Applications of numerical methods for phase-field models do typically not appear in the maritime field but are those problems in which capillary effects play a significant role. The advantage of phase field models is their rigorous thermodynamical structure. A drawback is that numerical methods for phase-field models typically do not satisfy the maximum principle for the density.

Problems in the maritime field often have a large density jump and as such the maximum principle is essential there.

The origin of the volume-of-fluid methods may be traced back to the work of Hirt and Nichols in 1981 [93]. The representation of the surface is generally performed via so-called height functions [153], in which the height of the interface is compared to some reference value. This approach faces several issues when curves intersect and extensions to three-dimensional problems are problematic. Based on the Marker-and-Cell Method (MAC) [87] a global field known as the volume fraction was later introduced. This volume fraction is a field which indicates the ratio of the (two) fluids in the domain. This leads to a method in which the conservation of mass is guaranteed (given that the initial mathematical model is conservative) and topological changes no longer cause problems. The drawbacks of this method are the evaluation of curvature (essential for problems with capillary effects) and the fact that the interface smears out in time. Furthermore, we note that volume-of-fluid methods are often employed in combination with finite-volume methods, both for incompressible [146, 167] and compressible flows [62, 63, 150, 166].

The level-set method [156] is an efficient way to handle the moving interface. In the level-set methods the interface is represented by the zero level-set of a higher-dimensional field. Similar to the phase-field and volume-of-fluid approaches, the level set method is not limited by the complexity of the free surface flow. In contrast to the volume-of-fluid method, curvature evaluations relevant for problems with surface tension are natural. Furthermore, level-set methods satisfy the maximum principle for the density and have proven to be a viable approach for water-air flows in marine applications, see e.g., [3, 5, 6, 151]. The drawbacks of level-set method are the fact that mass-conservation is not guaranteed and that certain re-initialization procedures are required. Furthermore, the underlying algorithm in level-set formulation are never provably thermodynamically stable. This can cause instabilities; Akkerman et al. [3] show that viscous air-water flow level-set simulations may create artificial energy in some cases.

## 1.6 RESEARCH OBJECTIVE

The overall objective of this thesis is the development of finite element isogeometric methods for flow problems that *inherit the stability properties* of the underlying physical system. The constructed numerical techniques should be generally applicable and in particular be useful elements for free surface flow simulations. The challenges discussed in the previous sections indicate that stability issues in free surface flow simulations often stem either from (i) the multiscale formulation, (ii) sharp layers or (iii) the interface separating the fluid. We address each of these challenges in this dissertation and formulate the sub-objectives as:

- (i) Develop an energy-stable finite element method for turbulent flow.
- (ii) Derive a discontinuity capturing mechanism from the underlying physical system.
- (iii) Construct an energy-dissipative, maximum-principle satisfying numerical method for the simulation of free surface flows.

## 1.7 OUTLINE

This thesis is composed of three parts, each linked to one of the sub-objectives of the thesis. In Part I we employ variational multiscale analysis to derive energy-dissipative stabilized methods. First we present the methodology for convection-diffusion equations and subsequently for the incompressible Navier-Stokes equations. Part II focuses on a theoretical framework for discontinuity capturing methods. We develop what we call variation entropy theory, and subsequently use that in combination with variational multiscale analysis to derive the framework. Lastly, in Part III we construct an energy-dissipative level-set method to simulate two-fluid flows.

Below we outline each part in more detail.

*Part I. Energy-dissipative multiscale formulations*

Chapter 2 initiates the design stabilized finite element methods that exhibit correct-energy behavior. The convection-diffusion equation serves as a model problem. We demonstrate that the classical stabilized formulations may create artificial energy. Based on the variational multiscale method we propose a solution strategy to rectify this situation.

In Chapter 3 we extend the developed methodology to the incompressible Navier-Stokes equations. We illustrate that, also here, classical stabilized terms may be the source of instabilities and amend this in the variational multiscale context. An essential ingredient is here the isogeometric discretization method. This allows to obtain pointwise divergence-free solutions which are key for an energy-stable method.

*Part II. A framework for discontinuity capturing methods*

In order to derive a discontinuity capturing method we need to identify the origin of the instabilities. This is the purpose of Chapter 4 which presents what we call variation entropy theory. In this chapter we explain the source of the instabilities in an entropy context. This forms the basis of the derivation of discontinuity capturing methods.

Chapter 5 provides the actual derivation of discontinuity capturing methods. Besides the variation entropy theory it relies on the variational multiscale method. This provides a means to naturally incorporate the discontinuity mechanism into the finite element method. We present a new discontinuity capturing method that naturally emerges from the framework. Additionally, we illustrate that some existing discontinuity capturing devices may be viewed as approximations of this new method and show that these are overly diffusive.

*Part III. An energy-dissipative method for free surface flow*

Chapter 6 considers the construction of an energy-dissipative numerical method for the incompressible Navier-Stokes equations with surface tension. Our methodology employs the level-set method to capture the fluid interface. It appears that standard discretization techniques for the well-known diffuse-interface level-set model are potentially unstable. We use functional entropy variables to overcome this discrepancy.

We close with Chapter 7 which discusses the presented results and outlines possible future research directions.



## Part I

### ENERGY-DISSIPATIVE MULTISCALE FORMULATIONS

This part presents the construction of stabilized methods with an energy-dissipative property that closely resembles that of the continuous model. The methods are presented for the convection-diffusion model problem and the incompressible Navier-Stokes equations.



---

## CORRECT ENERGY EVOLUTION OF STABILIZED CONVECTION-DIFFUSION

---

This chapter is reproduced from [57]:

M.F.P. ten Eikelder and I. Akkerman, *Correct energy evolution of stabilized formulations: The relation between VMS, SUPG and GLS via dynamic orthogonal small-scales and isogeometric analysis. I: The convective-diffusive context*, *Comput. Meth. Appl. Mech. Engrg.* 331 (2018) 259-280.

### ABSTRACT

*This work presents the construction of novel stabilized finite element methods in the convective–diffusive context that exhibit correct-energy behavior. Classical stabilized formulations can create unwanted artificial energy. Our contribution corrects this undesired property by employing the concepts of dynamic as well as orthogonal small-scales within the variational multiscale framework (VMS). The desire for correct energy indicates that the large- and small-scales should be  $H_0^1$ -orthogonal. Using this orthogonality the VMS method can be converted into the streamline-upwind Petrov-Galerkin (SUPG) or the Galerkin/least-squares (GLS) method. Incorporating both large- and small-scales in the energy definition asks for dynamic behavior of the small-scales. Therefore, the large- and small-scales are treated as separate equations.*

*Two consistent variational formulations which depict correct-energy behavior are proposed: (i) the Galerkin/least-squares method with dynamic small-scales (GLSD) and (ii) the dynamic orthogonal formulation (DO). The methods are presented in combination with an energy-decaying generalized- $\alpha$  time-integrator. Numerical verification shows that dissipation due to the small-scales in classical stabilized methods can become negative, on both a local and a global scale. The results show that without loss of accuracy the correct-energy behavior can be recovered by the proposed methods. The computations employ NURBS-based isogeometric analysis for the spatial discretization.*

## 2.1 INTRODUCTION

Stabilized methods and multiscale formulations form an auspicious, versatile and fundamental class of methodologies for finite element computations. The classical Galerkin variational formulation depicts correct-energy behavior although it has limitations concerning accuracy and stability. The popular stabilized methods, i.e. the Streamline upwind Petrov-Galerkin method (SUPG) [33], the Galerkin/least-squares method (GLS) [110], and the variational multiscale method (VMS) [100, 103], overcome these issues, however show incorrect-energy behavior. In this work we focus on convection-diffusion which serves as a model problem for more complex flow problems and turbulence.

This work is devoted to the construction of a new stabilized finite element method displaying *correct-energy behavior*. Correct-energy behavior (or evolution) in a numerical method represents here that the method (i) does not create artificial energy and (ii) closely resembles the energy evolution of the continuous setting. A precise definition is included in Section 2.4.

Our contribution fixes the incorrect energy deficiency by combining several ingredients. These are the *dynamic* and *orthogonal* behavior of the small-scales emerging from the stabilized methods, also referred to as *dynamic orthogonal small-scales*, within the framework of isogeometric analysis.

### 2.1.1 *Dynamic small-scales*

In our quest for a correct-energy displaying formulation we learn that it is essential to use the so-called dynamic small-scales (also referred to as transient small-scales). This approach models the small-scales dynamically, i.e. with an ordinary differential equation in time, and takes its temporal contribution to the large-scale equation into account. This concept has originally been proposed by Codina in [46] and has been further analyzed in [48].

### 2.1.2 *Orthogonal small-scales in VMS*

The multiscale stabilization method based on orthogonal small-scales serves as the next key ingredient of our approach. We link our choice of orthogonal small-scales to an optimality projector induced by the  $H_0^1$ -seminorm. This produces a highly attenuated and localized small-scale Green's function, which is very desirable property [113]. We combine this methodology with residual-based variational multiscale modeling, a concept which emanates from VMS. The VMS approach finds many applications in incompressible turbulence, see e.g., [4, 16, 17, 94, 131, 141], and free surface flow [5, 141]. Possible new directions in stabilized and multiscale methods are suggested in [24].

### 2.1.3 *Isogeometric analysis framework*

In addition, we employ the isogeometric analysis (IGA) methodology, proposed by Hughes et al. in [101], which finds recent applications in various fields of science,

see e.g., [5, 20, 23]. IGA is an effort to close the gap between on one hand Computer-Aided Design (CAD) and on the other Computer-Aided Engineering (CAE). Finite element analysis (FEA) and CAD use a different representation for the geometry which makes a geometry update unpleasant and time-consuming. IGA corrects this deficiency by employing the same NURBS (non-uniform rational B-splines) geometry description as in CAD. This means that the NURBS surfaces in IGA match with the *exact* CAD geometry, in contrast to FEA where the basis functions form an approximation of the CAD geometry. IGA leads to higher-order and higher-continuity discretizations on complex domains. Our choice for IGA is additionally motivated by the second derivatives in the weak formulations. This requires  $C^1$ -continuity of the basis functions. Furthermore, one of the main advantages of using the IGA methodology is that it guarantees the incompressibility constraint to hold exactly [68, 69]. This is a highly favorable property when the velocity field is not given, e.g., in case of the incompressible Navier-Stokes equations.

#### 2.1.4 Context

The methodology is presented in the convective–diffusive model context which serves as a first step of this novel approach. The procedure is developed with the incompressible Navier-Stokes equations in mind which is the next step of this approach and is in itself presented in the sequel work. In the context of stabilized methods a two-step approach, development for linear convection-diffusion followed by incompressible Navier-Stokes, is more common, see e.g., [73, 74, 110, 111, 171].

In the context of two-fluid flow phenomena, many numerical methodologies can unfortunately artificially create energy at the two-fluid interface. Even a small energy-inconsistency at the fluid surface can already lead to highly unstable behavior as is demonstrated in [3]. To rectify this discrepancy, each of the components of the algorithm requires correct-energy behavior. When numerically solving air–water flow usually the components are (i) a standard incompressible Navier-Stokes solver and (ii) an algorithm describing the evolution of the air–water interface. Apart from its shared features with the incompressible Navier-Stokes equations, the convective–diffusive context is also required for the (level set) algorithm describing the evolution of the two-fluid interface.

#### 2.1.5 Outline

The remainder of this chapter is dedicated to the actual construction of a stabilized variational formulation which depicts correct-energy behavior and is summarized as follows. Section 2.2 states the continuous form of the governing convection-diffusion equations, both in the strong form and in the weak form. The energy evolution linked to this formulation follows from the corresponding mixed-formulation which is derived with a Lagrange multiplier approach. Before proposing changes to existing stabilized methods, we introduce and discuss the energy evolution of the existing stabilized methods. Therefore, Section 2.3 discusses the energy evolution in the standard VMS stabilized method with static small-scales. Section 2.4 presents our quest towards a stabilized formulation depicting correct-energy behavior. It

adds the two concepts (i) the dynamic behavior of the small-scales and (ii) the optimality projector yielding the vital orthogonality of the small-scales to the VMS formulation. Invoking the optimality projector in different ways leads to the other well-known stabilized formulations, namely SUPG and GLS. In Section 2.5 the demanded orthogonality between the small- and large-scales is enforced by the proper  $H_0^1$ -optimality projector. This yields an alternative variational multiscale stabilized formulation with correct-energy evolution. Furthermore, the methods demand a time-integrator which is correctly linked to an energy. Therefore, we re-address the generalized- $\alpha$  time-integration method. The energy demand leads to a certain parameter family of the time-stepping parameters. Section 2.6 discusses this temporal-integration method. Section 2.7 presents numerical verification while employing NURBS basis functions for the computations. In Section 2.8, we draw conclusions and outline avenues for future research.

## 2.2 THE CONTINUOUS CONVECTION-DIFFUSION EQUATION

### 2.2.1 Strong formulation

Let  $\Omega$  denote the spatial domain with boundary  $\Gamma = \Gamma_g \cup \Gamma_h$ , see Figure 2.1.

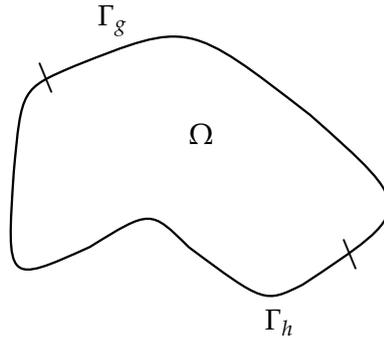


Figure 2.1: Spatial domain  $\Omega$  with its boundaries  $\Gamma = \Gamma_g \cup \Gamma_h$ .

The governing equations of the convection-diffusion problem in strong form read

$$\partial_t \phi + \mathbf{a} \cdot \nabla \phi - \nabla \cdot \kappa \nabla \phi = f \quad \text{in } \Omega \times \mathcal{I}, \quad (2.2.1a)$$

$$\phi = g \quad \text{in } \Gamma_g \times \mathcal{I}, \quad (2.2.1b)$$

$$-a_n^- \phi + \kappa \partial_n \phi = h \quad \text{in } \Gamma_h \times \mathcal{I}, \quad (2.2.1c)$$

$$\phi(\mathbf{x}, 0) = \phi_0(\mathbf{x}) \quad \text{in } \Omega, \quad (2.2.1d)$$

where  $t \in \mathcal{I} = (0, T)$  is the time with final time  $T > 0$ ,  $x \in \Omega$  the spatial coordinate,  $\phi = \phi(x, t) : \Omega \times \mathcal{I}$  the dependent variable with time derivative  $\partial_t \phi$ , normal flux  $\partial_n \phi = \mathbf{n} \cdot \nabla \phi$  and  $f : \Omega \times \mathcal{I} \rightarrow \mathbb{R}$ ,  $g : \Gamma_g \times \mathcal{I} \rightarrow \mathbb{R}$ ,  $h : \Gamma_h \times \mathcal{I} \rightarrow \mathbb{R}$  and  $\phi_0 : \Omega \rightarrow \mathbb{R}$  are prescribed data. The convective velocity  $\mathbf{a} = \mathbf{a}(x)$  is a given solenoidal vector field ( $\nabla \cdot \mathbf{a} = 0$ ) and  $\kappa \geq 0$  denotes the given diffusivity. The outward unit normal to  $\Gamma$  is  $\mathbf{n}$  and the normal velocity component denotes  $a_n = \mathbf{a} \cdot \mathbf{n}$  with positive and negative parts  $a_n^\pm = \frac{1}{2}(a_n \pm |a_n|)$ .

### 2.2.2 Weak formulation

Let  $\mathcal{W}^0$  and  $\mathcal{W}^g$  denote suitable function-spaces satisfying the homogeneous and non-homogeneous Dirichlet conditions, respectively. Using these spaces the standard weak formulation of the problem reads as follows:

Find  $\phi \in \mathcal{W}^g$  such that for all  $w \in \mathcal{W}^0$ ,

$$\begin{aligned} (w, \partial_t \phi)_\Omega + (w, \mathbf{a} \cdot \nabla \phi)_\Omega - (w, a_n^- \phi)_{\Gamma_h} + (\nabla w, \kappa \nabla \phi)_\Omega &= (w, f)_\Omega \\ &+ (w, h)_{\Gamma_h}. \end{aligned} \quad (2.2.2)$$

Here  $(\cdot, \cdot)_D$  denotes the  $L^2(D)$  inner product over  $D$ . Consistency of the strong (2.2.1) and the weak formulation (2.2.2) easily follows from applying integration by parts on the diffusive term.

Instead of enforcing the Dirichlet boundary conditions *a priori*, it is also possible to relax this condition in the function space by employing a Lagrange multiplier setting. The weak statement translates into a *mixed formulation*:

Find  $(\phi, \lambda_\Omega) \in \mathcal{W} \times \mathcal{V}$  such that for all  $(w, q) \in \mathcal{W} \times \mathcal{V}$ ,

$$\begin{aligned} (w, \lambda_\Omega)_{\Gamma_g} &= (w, \partial_t \phi)_\Omega + (w, \mathbf{a} \cdot \nabla \phi)_\Omega - (w, a_n^- \phi)_{\Gamma_h} \\ &+ (\nabla w, \kappa \nabla \phi)_\Omega - (w, f)_\Omega - (w, h)_{\Gamma_h}, \end{aligned} \quad (2.2.3a)$$

$$(q, \phi)_{\Gamma_g} = (q, g)_{\Gamma_g}. \quad (2.2.3b)$$

Here  $\mathcal{W}$  represents the unrestricted function space and  $\mathcal{V}$  is a suitable Lagrange multiplier space. Consult [102, 117] for the appropriate construction of the spaces. The following section employs this formulation to derive energy statements.

Applying an appropriate choice of weighting functions  $w$  and  $q$  in (2.2.3) and subsequently performing a partial integration step recovers the strong form (2.2.1). Additionally, the expression for the Lagrange multiplier follows as a complimentary result

$$\lambda_\Omega = \kappa \partial_n \phi, \quad (2.2.4)$$

and equals the diffusive flux. Note that the continuous setting allows us to provide a closed form of the Lagrange multiplier. This does not hold in a discrete setting. Furthermore, the subscript in the notation of the Lagrange multiplier is added for consistency with Section 2.2.4.

### 2.2.3 Global energy evolution

The evolution of the energy linked to the strong form (2.2.1) follows from choosing  $w = \phi$  and  $q = \lambda$  in the mixed formulation (2.2.3). Subtracting the resulting equations yields

$$\begin{aligned} (\phi, \partial_t \phi)_\Omega + (\nabla \phi, \kappa \nabla \phi)_\Omega + (\phi, \mathbf{a} \cdot \nabla \phi)_\Omega - (\phi, a_n^- \phi)_{\Gamma_h} &= (g, \lambda_\Omega)_{\Gamma_g} + (\phi, f)_\Omega \\ &+ (\phi, h)_{\Gamma_h}. \end{aligned} \quad (2.2.5)$$

By performing integration by parts on the interior convective term and employing the divergence-free constraint, the convective term turns into a boundary term. The energy, defined as  $\mathcal{E}_\Omega = \frac{1}{2}(\phi, \phi)_\Omega$ , evolves as

$$\frac{d}{dt}\mathcal{E}_\Omega = -\|\kappa^{1/2}\nabla\phi\|_\Omega^2 + (\phi, f)_\Omega - (1, \mathcal{F}_\Omega)_\Gamma, \quad (2.2.6)$$

where  $\|\cdot\|_D$  defines the standard  $L^2$ -norm over  $D$ . The conservative energy flux provides a different contribution on each segment of the boundary:

$$\mathcal{F}_\Omega = \begin{cases} a_n e - g\lambda_\Omega & \text{on } \Gamma_g, \\ |a_n|e - \phi h & \text{on } \Gamma_h, \\ 0 & \text{elsewhere,} \end{cases} \quad (2.2.7)$$

with  $e := \frac{1}{2}\phi^2$  the pointwise energy. The terms on the Dirichlet boundary are (i) the amount of energy flowing in and out by convection and (ii) the energy gained and lost by diffusion through the boundary, respectively. On the other boundary, the terms represent (i) the energy loss by the strict convective outflow and (ii) the energy change by the flux boundary condition. The energy can only increase as a result of the prescribed body force or the boundary conditions, represented by the last two terms on the right-hand side of (2.2.6). The diffusive term, when active, contributes to a decay of the energy. The last term on the right-hand side of (2.2.6) represents the convective and diffusive fluxes of energy across the interface. Substitution of the boundary condition and the Lagrange multiplier (again possible because of the continuous setting) and a partial integration step leads to the alternative expression of the flux

$$\mathcal{F}_\Omega = a_n e - \kappa \partial_n e. \quad (2.2.8)$$

The two terms respectively describe the convective and viscous-driven flow of energy.

#### 2.2.4 Localized energy evolution

This section presents a localized version of (2.2.6) for arbitrary subdomains  $\omega \subset \Omega$  with boundary  $\partial\omega$ . The complement domain denotes  $\Omega - \omega$  with boundary  $\partial(\Omega - \omega)$  and the shared boundary of the both subdomains is  $\chi_\omega := \partial\omega \cap \partial(\Omega - \omega)$ . Figure 2.2 shows the domain with its boundaries.

The variational statement consists of the weak formulation (2.2.3) enforced on the subdomain  $\omega$  and is again augmented with a Lagrange multiplier that enforces the Dirichlet boundary condition. The unrestricted solution space  $\mathcal{W}_\omega$  allows discontinuities across the subdomain interface and the space of suitable Lagrange multipliers denotes  $\mathcal{V}_\omega$ . The weak statement reads:

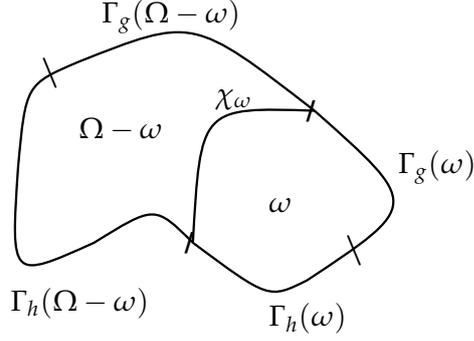


Figure 2.2: Spatial domain  $\Omega$  with a subdomain  $\omega \subset \Omega$ . The shared boundary of  $\omega$  and its complement is  $\chi_\omega$ . The boundaries  $\Gamma_g$  and  $\Gamma_h$  split according to  $\omega$ .

Find  $(\phi, \lambda_\omega) \in \mathcal{W}_\omega \times \mathcal{V}_\omega$  such that for all  $(w, q) \in \mathcal{W}_\omega \times \mathcal{V}_\omega$ ,

$$\begin{aligned} (w, \lambda_\omega)_{\chi_\omega} + (w, \lambda_\omega)_{\Gamma_g(\omega)} &= (w, \partial_t \phi)_\omega + (\nabla w, \kappa \nabla \phi)_\omega + (w, \mathbf{a} \cdot \nabla \phi)_\omega \\ &\quad - (w, a_n^- \phi)_{\Gamma_h(\omega)} - (w, f)_\omega - (w, h)_{\Gamma_h(\omega)}, \end{aligned} \quad (2.2.9a)$$

$$(q^h, [\phi^h])_{\chi_\omega} + (q, \phi)_{\Gamma_g(\omega)} = (q, g)_{\Gamma_g(\omega)}, \quad (2.2.9b)$$

where  $\Gamma_g(D) := \Gamma_g \cap \partial D$  and  $\Gamma_h(D) := \Gamma_h \cap \partial D$  for domain  $D$ . The jump term  $[\phi^h]$  is given by

$$[\phi^h] := \phi_\omega^h - \phi_{\Omega-\omega}^h, \quad (2.2.10)$$

where the terms are defined on  $\omega$  and  $\Omega - \omega$ , respectively. The Lagrange multiplier can be interpreted as an auxiliary flux across the interface  $\chi_\omega$ , it represents the flow outward  $\omega$ . The weak form of the complement domain easily follows by replacing  $\omega$  by  $\Omega - \omega$  in (2.2.9). The superposition of the both formulations leads to the balance

$$\lambda_\omega + \lambda_{\Omega-\omega} = 0. \quad (2.2.11)$$

Thus that what flows out of  $\omega$  enters in  $\Omega - \omega$ . See [102] for the formal details of such a derivation. Again, a partial integration step provides the expression for the Lagrange multipliers:

$$\lambda_\omega = \kappa \partial_{n_\omega} \phi, \quad (2.2.12a)$$

$$\lambda_{\Omega-\omega} = \kappa \partial_{n_{\Omega-\omega}} \phi, \quad (2.2.12b)$$

with  $\partial_{n_D}$  the directional derivative outward of a domain  $D$ . The local energy statement follows when choosing  $w = \phi$  and  $q = \lambda_\omega$ , this yields

$$\begin{aligned} (\phi, \lambda_\omega)_{\chi_\omega} + (g, \lambda_\omega)_{\Gamma_g(\omega)} &= (\phi, \partial_t \phi)_\omega + \|\kappa^{1/2} \nabla \phi\|_\omega^2 + (\phi, \mathbf{a} \cdot \nabla \phi)_\omega \\ &\quad - (w, a_n^- \phi)_{\Gamma_h(\omega)} - (\phi, f)_\omega - (\phi, h)_{\Gamma_h(\omega)}. \end{aligned} \quad (2.2.13)$$

By applying integration by parts on the convective term we find that the energy on subdomain  $\omega$  evolves as

$$\frac{d}{dt} \mathcal{E}_\omega = -\|\kappa^{1/2} \nabla \phi\|_\omega^2 + (\phi, f)_\omega - (1, \mathcal{F}_\omega)_{\partial\omega}, \quad (2.2.14)$$

where the outward energy flux now splits into three parts

$$\mathcal{F}_\omega = \begin{cases} a_{n_\omega} e - g \lambda_\omega & \text{on } \Gamma_g(\omega), \\ |a_{n_\omega}| e - \phi h & \text{on } \Gamma_h(\omega), \\ a_{n_\omega} e - \phi \lambda_\omega & \text{on } \chi_\omega, \\ 0 & \text{elsewhere.} \end{cases} \quad (2.2.15)$$

In comparison with global energy behavior, the additional last term represents an energy flux, with a convective and diffusive component, across the subdomain interface  $\chi_\omega$ . Again, the substitution of the boundary condition and the Lagrange multiplier, and subsequently performing a partial integration step lead to

$$\mathcal{F}_\omega = a_{n_\omega} e - \kappa \partial_{n_\omega} e \quad \text{on } \partial\omega. \quad (2.2.16)$$

**Remark 2.2.1.** *This section provides all the statements in a continuous form. A direct consequence is that the standard discrete setting, the Galerkin method, displays correct-energy behavior.*

This chapter now presents the energy evolution of standard stabilized methods and subsequently constructs a methodology that closely resembles the local energy evolution of the continuous equation. In particular, the design of the method precludes artificial local energy creation.

**Remark 2.2.2.** *To increase the readability of this chapter we now restrict ourselves to boundary conditions precluding the energy flux  $\mathcal{F}$  on  $\Gamma$ . This occurs for example when employing homogeneous Dirichlet and periodic boundary conditions. The proposed methodology can easily be generalized to domains with non-homogeneous boundaries.*

## 2.3 ENERGY EVOLUTION OF THE VARIATIONAL MULTISCALE APPROACH

This section concerns the energy evolution in the stabilized residual-based variational multiscale (RBVMS) formulation. Therefore we start off with a brief recap of the canonical VMS method.

### 2.3.1 The multiscale split

The residual-based variational multiscale approach emanates, as the name suggests, from the theory of the variational multiscale methods. This approach explicitly treats the solution component not be represented by the discretization in an approximate sense. This is done as follows. The trial solution and weighting function spaces split as

$$\mathcal{W} = \mathcal{W}^h \oplus \mathcal{W}', \quad (2.3.1)$$

where  $\mathcal{W}^h$  is the space spanned by the finite-dimensional discretization and  $\mathcal{W}'$  is its infinite-dimensional complement in  $\mathcal{W}$ . Based on the multiscale split in the space the components of the solution and weight decouple as

$$\begin{aligned} \phi &= \phi^h + \phi', \\ w &= w^h + w', \end{aligned} \quad (2.3.2)$$

with  $\phi^h, w^h \in \mathcal{W}^h$  and  $\phi', w' \in \mathcal{W}'$  the large-scales and the small-scales solution and weight, respectively. The small-scale space  $\mathcal{W}'$  represents the component of  $\mathcal{W}$  not reproduced by the grid and is therefore also called space of fine-scales, sub-scales or subgrid-scales<sup>1</sup>. In order to obtain a well-defined decomposition for a given  $v \in \mathcal{W}$ , the elements  $v^h \in \mathcal{W}^h, v' \in \mathcal{W}'$  with  $v = v^h + v'$  require a unique definition. Employing an optimality projector  $\mathcal{P}^h : \mathcal{W} \rightarrow \mathcal{W}^h$  for the decoupling as<sup>2</sup>:

$$\begin{aligned} v^h &= \mathcal{P}^h v, \\ v' &= (\mathcal{I} - \mathcal{P}^h) v, \end{aligned} \quad (2.3.3)$$

achieves uniqueness. Here  $\mathcal{I} : \mathcal{W} \rightarrow \mathcal{W}$  is the identity operator. Using this multi-scale split we arrive at the following alternative – equivalent – weak statement:

Find  $\phi^h \in \mathcal{W}^h, \phi' \in \mathcal{W}'$  such that for all  $w^h \in \mathcal{W}^h, w' \in \mathcal{W}'$ ,

$$\begin{aligned} (w^h, \partial_t \phi^h + \mathbf{a} \cdot \nabla \phi^h)_\Omega + (\nabla w^h, \kappa \nabla \phi^h)_\Omega + (w^h, \partial_t \phi' + \mathbf{a} \cdot \nabla \phi')_\Omega \\ + (\nabla w^h, \kappa \nabla \phi')_\Omega = (w^h, f)_\Omega, \end{aligned} \quad (2.3.4a)$$

$$\begin{aligned} (w', \partial_t \phi^h + \mathbf{a} \cdot \nabla \phi^h)_\Omega + (\nabla w', \kappa \nabla \phi^h)_\Omega + (w', \partial_t \phi' + \mathbf{a} \cdot \nabla \phi')_\Omega \\ + (\nabla w', \kappa \nabla \phi')_\Omega = (w', f)_\Omega. \end{aligned} \quad (2.3.4b)$$

Note that this formulation is still exact. However, the space  $\mathcal{W}'$  is infinite-dimensional and as such not amenable for a discrete implementation.

### 2.3.2 The VMS numerical formulation

Let us now subdivide the domain  $\Omega$  into elements  $\Omega_e$ . The domain of element interiors  $\tilde{\Omega}$  does not include the element boundaries and reads

$$\tilde{\Omega} = \bigcup_e \Omega_e. \quad (2.3.5)$$

The weak formulation (2.3.4) converts into a numerical formulation when the small-scale equation (2.3.4b) is replaced by an approximation for the small-scale solution  $\phi'$ . The small-scale equation can be written in the form

$$(w', \partial_t \phi' + \mathbf{a} \cdot \nabla \phi' - \kappa \Delta \phi')_\Omega = -(w', \mathcal{R} \phi^h)_\Omega, \quad (2.3.6)$$

where the large-scale residual reads

$$\mathcal{R} \phi^h = \partial_t \phi^h + \mathbf{a} \cdot \nabla \phi^h - \kappa \Delta \phi^h - f. \quad (2.3.7)$$

This implies that the small-scales are driven by the residuals of the large-scales. The corresponding Euler-Lagrange form of the small-scale equation reads

$$\partial_t \phi' + \mathbf{a} \cdot \nabla \phi' - \kappa \Delta \phi' = -\mathcal{R} \phi^h. \quad (2.3.8)$$

- <sup>1</sup> The terms sub-scales or subgrid-scales could be linked to a turbulence modeling character of the approach. The current method does not fit in that framework. To emphasize this difference we use the terminology small-scales.
- <sup>2</sup> There are infinitely many choices for the projector  $\mathcal{P}^h$ . Linear projectors suffice for the current problem. More details can be found in Hughes [100].

Note that the Euler-Lagrange equations are in strong form, i.e. the weight  $w'$  in (2.3.6) is ignored. This pertains to both the small-scale solution as well as the residual forcing.

Employing a Green's function provides an explicit expression for the small-scales. In this expression the integral operator is approximated by an algebraic stabilization parameter  $\tau_{\text{static}}$ . This step is necessary to arrive at an implementable method. Details of this approximation can be found in [100]. To obtain a stabilized formulation, the small-scales are modeled as:

$$\hat{\phi}' = -\tau_{\text{stat}}\mathcal{R}\phi^h, \quad (2.3.9a)$$

$$\partial_t \hat{\phi}' = 0, \quad (2.3.9b)$$

where  $\tau_{\text{stat}}$  is a positive stabilization parameter. In the following we ignore the hat-sign. This algebraic operator depends on both the physics and the discretization. More details can be found in Hughes and Sangalli [113].

The definition of the stabilization parameter is inspired by the theory of stabilized methods for convection-diffusion equations (see e.g., [111, 171]), and reads:

$$\tau_{\text{stat}} = (\tau_{\text{conv}}^{-2} + \tau_{\text{diff}}^{-2} + \tau_{\text{time}}^{-2})^{-1/2}, \quad (2.3.10)$$

where

$$\tau_{\text{conv}}^{-2} = \mathbf{a} \cdot \mathbf{G} \mathbf{a}, \quad (2.3.11a)$$

$$\tau_{\text{diff}}^{-2} = C_I \kappa^2 \mathbf{G} : \mathbf{G}, \quad (2.3.11b)$$

$$\tau_{\text{time}}^{-2} = \left( \frac{\alpha_m}{\alpha_f \gamma \Delta t} \right)^2. \quad (2.3.11c)$$

Here  $\mathbf{G}$  is the second-rank metric tensor given by

$$\mathbf{G} = \frac{\partial \xi^T}{\partial x} \frac{\partial \xi}{\partial x'}, \quad (2.3.12)$$

where  $\partial \xi / \partial x$  is the inverse Jacobian of the map between the elements in the reference and physical domain. The stabilization parameter treats deformed and curved domains naturally due to its direct dependence on  $\mathbf{G}$ . The metric tensor  $\mathbf{G}$  scales as  $h^{-2}$  where  $h$  is the mesh size. The positive constant  $C_I$  is defined by an inverse estimate. It is independent of the mesh size and can be computed from an element-wise eigenvalue problem [86].

The definition of  $\tau_{\text{time}}$  is based on the generalized- $\alpha$  time-integrator given in Section 2.6. Here  $\alpha_f, \alpha_m, \gamma$  are algorithmic time-stepping coefficients and  $\Delta t$  is the time step. It reduces to the commonly used value of  $4/\Delta t^2$  when employing  $\rho_\infty = 1$ , see e.g., [4, 16, 17] (consult these references for the definition of  $\rho_\infty$ ). This choice results in the Crank-Nicolson time-integrator, see Section 2.6.2.

Employing integration by parts on the stabilized terms, the small-scales appear without derivatives. The resulting form is the VMS stabilized statement

Find  $\phi^h \in \mathcal{W}^h$  such that for all  $w^h \in \mathcal{W}^h$ ,

$$\begin{aligned} (w^h, \partial_t \phi^h)_\Omega + (w^h, \mathbf{a} \cdot \nabla \phi^h)_\Omega + (\nabla w^h, \kappa \nabla \phi^h)_\Omega \\ - (\mathbf{a} \cdot \nabla w^h + \kappa \Delta w^h, \phi^h)_{\hat{\Omega}} = (w^h, f)_\Omega, \end{aligned} \quad (2.3.13a)$$

$$\tau_{\text{stat}}^{-1} \phi^h = -\mathcal{R}\phi^h. \quad (2.3.13b)$$

It is important to emphasize that we treat the small-scale expression (2.3.13b) as a separate equation. At this stage a straightforward substitution is certainly possible, however when the small-scales are modeled dynamically, this is not the case anymore. In line with the analysis in later sections we therefore omit substitution here.

### 2.3.3 Local energy evolution of the VMS formulation

To arrive at local energy evolution, we augment the weak formulation in Lagrange multiplier setting form to allow discontinuous functions across subdomains, similar to (2.2.9). The weak statement reads:

Find  $(\phi^h, \lambda_\omega^h) \in \mathcal{W}_\omega^h \times \mathcal{V}_\omega^h$  such that for all  $(w^h, q^h) \in \mathcal{W}_\omega^h \times \mathcal{V}_\omega^h$ ,

$$\begin{aligned} (w^h, \lambda_\omega^h)_{\chi_\omega} &= (w^h, \partial_t \phi^h)_\omega + (w^h, \mathbf{a} \cdot \nabla \phi^h)_\omega + (\nabla w^h, \kappa \nabla \phi^h)_\omega - (w^h, f)_\omega \\ &\quad - (\mathbf{a} \cdot \nabla w^h + \kappa \Delta w^h, \phi')_{\tilde{\omega}}, \end{aligned} \quad (2.3.14a)$$

$$(q^h, [\phi^h])_{\chi_\omega} = 0, \quad (2.3.14b)$$

$$\tau_{\text{stat}}^{-1} \phi' = -\mathcal{R} \phi^h. \quad (2.3.14c)$$

Here  $\tilde{\omega}$  represents the domain of element interiors of  $\omega$ . The discretization does not allow explicit evaluation of the Lagrange multiplier. We select  $w^h = \phi^h$  in the large-scale equation (2.3.14a) and add the small-scale equation multiplied by  $\phi'$  and integrate. The resulting statement is:

$$\begin{aligned} (\phi^h, \partial_t \phi^h)_\omega + (\phi', \partial_t \phi^h)_{\tilde{\omega}} + \|\tau_{\text{stat}}^{-1/2} \phi'\|_{\tilde{\omega}}^2 + \|\kappa^{1/2} \nabla \phi^h\|_\omega^2 - (\phi^h, f)_\omega \\ - (\phi', f)_{\tilde{\omega}} - 2(\kappa \Delta \phi^h, \phi')_{\tilde{\omega}} + \frac{1}{2}(\phi^h, a_n \phi^h)_{\chi_\omega} - (\phi^h, \lambda_\omega^h)_{\chi_\omega} = 0. \end{aligned} \quad (2.3.15)$$

Here we have employed the incompressibility constraint to convert the interior convective term to a boundary term.

**Remark 2.3.1.** *When the velocity field is obtained by a numerical method the incompressibility constraint is often not exactly fulfilled though. However, by appropriately employing isogeometric analysis this can be achieved exactly [69]. Our implementation already employs the proper IGA spaces to allow a smooth transition to the incompressible Navier-Stokes equations.*

The local large-scale energy is the energy of resolved solution:  $\mathcal{E}_\omega^h = \frac{1}{2} (\phi^h, \phi^h)_\omega$  and evolves by (2.3.15) as:

$$\begin{aligned} \frac{d}{dt} \mathcal{E}_\omega^h &= -\|\kappa^{1/2} \nabla \phi^h\|_\omega^2 + (\phi^h, f)_\omega - (1, \mathcal{F}_\omega^h)_{\chi_\omega} \\ &\quad - \|\tau_{\text{stat}}^{-1/2} \phi'\|_{\tilde{\omega}}^2 + (\phi', f)_{\tilde{\omega}} + 2(\kappa \Delta \phi^h, \phi')_{\tilde{\omega}} - (\phi', \partial_t \phi^h)_{\tilde{\omega}}, \end{aligned} \quad (2.3.16)$$

with the energy flux

$$\mathcal{F}_\omega^h = a_n e^h - \lambda_\omega^h \phi^h. \quad (2.3.17)$$

where the *pointwise large-scale energy* is  $e^h := \frac{1}{2}\phi^h\phi^h$ . The *local total energy* is defined using the superposition of the small-scales and large-scales as:  $\mathcal{E}_\omega = \frac{1}{2}(\phi^h + \phi', \phi^h + \phi')_{\tilde{\omega}}$ . Its evolution directly follows:

$$\begin{aligned} \frac{d}{dt}\mathcal{E}_\omega &= -\|\kappa^{1/2}\nabla\phi^h\|_\omega^2 + (\phi^h, f)_\omega - (1, \mathcal{F}_\omega^h)_{\chi_\omega} \\ &\quad - \|\tau_{\text{stat}}^{-1/2}\phi'\|_{\tilde{\omega}}^2 + (\phi', f)_{\tilde{\omega}} + 2(\kappa\Delta\phi^h, \phi')_{\tilde{\omega}} + (\partial_t\phi', \phi^h + \phi')_{\tilde{\omega}}. \end{aligned} \quad (2.3.18)$$

We observe from this relation that the standard static VMS formulation does not possess a desirable energy behavior. The first line closely resembles the continuous energy evolution relation. No explicit expression for  $\lambda_\omega^h$  exists. The second line appears as a result of the stabilization terms. Its first term contributes to a decay of the energy, which is favorable from a stability argument. It can be interpreted as the diffusive energy decay of the missing small-scales. The last two terms are problematic. These unsymmetric terms can be bounded by both the physical diffusion  $\|\kappa^{1/2}\nabla\phi^h\|_\omega^2$  and the numerical diffusion  $\|\tau_{\text{stat}}^{-1/2}\phi'\|_{\tilde{\omega}}^2$ . The procedure is analogous to the standard coercivity analysis: apply Cauchy-Schwarz and Young's inequality subsequently. This leads to restrictions on the stabilization parameter  $\tau_{\text{stat}}$  depending on the diffusivity and the time step. More importantly, the overall diffusion of the method can be less than the physical diffusion. Hence, the small-scales can artificially create energy, which we numerically show in Section 2.7, and are therefore both numerically and physically undesirable. The next section corrects this deficiency.

**Remark 2.3.2.** *The global energy evolution easily follows when substituting  $\omega = \Omega$  and  $\tilde{\omega} = \tilde{\Omega}$  into (2.3.18):*

$$\begin{aligned} \frac{d}{dt}\mathcal{E}_\Omega &= -\|\kappa^{1/2}\nabla\phi^h\|_\Omega^2 + (\phi^h, f)_\Omega \\ &\quad - \|\tau_{\text{stat}}^{-1/2}\phi'\|_{\tilde{\Omega}}^2 + (\phi', f)_{\tilde{\Omega}} + 2(\kappa\Delta\phi^h, \phi')_{\tilde{\Omega}} + (\partial_t\phi', \phi^h + \phi')_{\tilde{\Omega}}. \end{aligned} \quad (2.3.19)$$

*Note the cancellation of the local energy flux.*

## 2.4 TOWARD A STABILIZED FORMULATION WITH CORRECT-ENERGY BEHAVIOR

This section presents a path with alternative stabilized formulations towards rectification of the discrepancy indicated in the previous section. First we adopt the concept of dynamic small-scales to eliminate the unwanted terms containing the temporal derivatives. Next, the undesirable diffusive term vanishes when employing orthogonal small-scales with the optimality projector. This leads to other well-known stabilization formulations, namely SUPG and GLS.

### 2.4.1 Design condition

To clarify, let us explicitly mention the design condition of the stabilized formulation which emerges from (2.3.18). We seek for a stabilized formulation corresponding to (2.2.1) which displays local energy behavior as:

$$\begin{aligned} \frac{d}{dt} \mathcal{E}_\omega &= - \|\kappa^{1/2} \nabla \phi^h\|_\omega^2 + (\phi', f)_{\tilde{\omega}} - (1, \mathcal{F}_\omega^h)_{\chi_\omega} \\ &\quad - \|\tau^{-1/2} \phi'\|_\omega^2 + (\phi^h, f)_\omega. \end{aligned} \quad (2.4.1)$$

In this chapter we call this *correct-energy behavior*. The positive scalar  $\tau$  represents the stabilization parameter of the small-scale equation and equals  $\tau = \tau_{\text{static}}$  when using static small-scales as in (2.3.9).

### 2.4.2 The variational multiscale method with dynamic small-scales

An alternative for replacing the small-scale equation with an algebraic relation is to retain the time-integration and only model the spatial part of the operator. This leads to so-called *dynamic small-scales*, as introduced in [48]. The model equation

$$\partial_t \hat{\phi}' + \tau_{\text{dyn}}^{-1} \hat{\phi}' = -\mathcal{R}\phi^h, \quad (2.4.2)$$

is now an ordinary differential equation in time. Again, we ignore the  $\hat{\phantom{x}}$  sign in the following. The time derivative in (2.4.2) eliminates the first unwanted temporal part in the energy evolution (2.3.18). Naturally, the stabilization parameter now omits a temporal part<sup>3</sup>, since it is explicitly handled, therefore:

$$\tau_{\text{dyn}} = (\tau_{\text{conv}}^{-2} + \tau_{\text{diff}}^{-2})^{-1/2}. \quad (2.4.3)$$

Clearly, the static small-scale equation (2.3.9b) does not apply anymore. Therefore, the term  $\partial_t \phi'$  is active in the large-scale equation. The VMS stabilized formulation with dynamic small-scales now reads:

Find  $\phi^h \in \mathcal{W}^h$  such that for all  $w^h \in \mathcal{W}^h$ ,

$$\begin{aligned} (w^h, \partial_t \phi^h + \partial_t \phi')_\Omega + (w^h, \mathbf{a} \cdot \nabla \phi^h)_\Omega + (\nabla w^h, \kappa \nabla \phi^h)_\Omega \\ - (\mathbf{a} \cdot \nabla w^h + \kappa \Delta w^h, \phi')_{\tilde{\Omega}} = (w^h, f)_\Omega, \end{aligned} \quad (2.4.4a)$$

$$\partial_t \phi' + \tau_{\text{dyn}}^{-1} \phi' = -\mathcal{R}\phi^h. \quad (2.4.4b)$$

To arrive at an energy relation we adopt the same procedure as before. The total local energy linked to this variational form evolves as:

$$\begin{aligned} \frac{d}{dt} \mathcal{E}_\omega &= - \|\kappa^{1/2} \nabla \phi^h\|_\omega^2 + (\phi^h, f)_\omega - (1, \mathcal{F}_\omega^h)_{\chi_\omega} \\ &\quad - \|\tau_{\text{dyn}}^{-1/2} \phi'\|_{\tilde{\omega}}^2 + (\phi', f)_{\tilde{\omega}} + 2(\kappa \Delta \phi^h, \phi')_{\tilde{\omega}}, \end{aligned} \quad (2.4.5)$$

with  $\mathcal{F}_\omega^h$  defined in (2.3.17). We observe that adopting dynamic small-scales indeed eliminates the undesired temporal terms.

<sup>3</sup> This explains our notation  $\tau_{\text{static}}$  in Section 2.3 where *static* represents *static small-scales*.

### 2.4.3 Orthogonality between the large-scales and the small-scales

The other unwanted term vanishes when the large-scales and small-scales are appropriately orthogonal with respect to each other, namely

$$(\kappa\Delta\phi^h, \phi')_\Omega = 0. \quad (2.4.6)$$

This defines the optimality projector (2.3.3) which links the stabilized formulation with the desired energy behavior. Therefore we employ the natural choice for the optimality projector:

$$\begin{aligned} \mathcal{P}^h : \phi \in \mathcal{W} \rightarrow \phi^h \in \mathcal{W}^h : \text{Find } \phi^h \in \mathcal{W}^h \text{ such that for all } w^h \in \mathcal{W}^h, \\ (\kappa\Delta w^h, \phi^h)_\Omega = (\kappa\Delta w^h, \phi)_\Omega. \end{aligned} \quad (2.4.7)$$

This yields the required orthogonality.

### 2.4.4 Consistent SUPG with dynamic small-scales

Employing the orthogonality (2.4.6) directly in the large-scale equation, leads to the dynamic small-scale version of the well-known SUPG formulation:

$$\begin{aligned} \text{Find } \phi^h \in \mathcal{W}^h \text{ such that for all } w^h \in \mathcal{W}^h, \\ (w^h, \partial_t \phi^h + \partial_t \phi')_\Omega + (w^h, \mathbf{a} \cdot \nabla \phi^h)_\Omega + (\nabla w^h, \kappa \nabla \phi^h)_\Omega \\ - (\mathbf{a} \cdot \nabla w^h, \phi')_{\tilde{\Omega}} = (w^h, f)_\Omega \end{aligned} \quad (2.4.8a)$$

$$\partial_t \phi' + \tau_{\text{dyn}}^{-1} \phi' = -\mathcal{R}\phi^h. \quad (2.4.8b)$$

Unfortunately, this removes only the contribution from the large-scale equation (2.4.4a); the contribution of the undesirable term from the small-scale equation (2.4.4b) remains:

$$\begin{aligned} \frac{d}{dt} \mathcal{E}_\omega = - \|\kappa^{1/2} \nabla \phi^h\|_\omega^2 + (\phi^h, f)_\omega - (1, \mathcal{F}_\omega^h)_{\chi_\omega} \\ - \|\tau_{\text{dyn}}^{-1/2} \phi'\|_{\tilde{\omega}}^2 + (\phi', f)_{\tilde{\omega}} + (\kappa\Delta\phi^h, \phi')_{\tilde{\omega}}. \end{aligned} \quad (2.4.9)$$

The undetermined sign of the last term indicates that the formulation can still create artificial energy locally.

### 2.4.5 Inconsistent SUPG with dynamic small-scales

Instead of using the orthogonality (2.4.6) only in the large-scale equation, one can adopt it as well on the small-scales. The resulting SUPG-like formulation with dynamic small-scales reads:

$$\begin{aligned} \text{Find } \phi^h \in \mathcal{W}^h \text{ such that for all } w^h \in \mathcal{W}^h, \\ (w^h, \partial_t \phi^h + \partial_t \phi')_\Omega + (w^h, \mathbf{a} \cdot \nabla \phi^h)_\Omega + (\nabla w^h, \kappa \nabla \phi^h)_\Omega \\ - (\mathbf{a} \cdot \nabla w^h, \phi')_{\tilde{\Omega}} = (w^h, f)_\Omega \end{aligned} \quad (2.4.10a)$$

$$\partial_t \phi' + \tau_{\text{dyn}}^{-1} \phi' = -\partial_t \phi^h - \mathbf{a} \cdot \nabla \phi^h + f. \quad (2.4.10b)$$

The energy evolution linked to this formulation,

$$\frac{d}{dt} \mathcal{E}_\omega = - \|\kappa^{1/2} \nabla \phi^h\|_\omega^2 + (\phi^h, f)_\omega - (1, \mathcal{F}_\omega^h)_{\chi_\omega} + (\phi', f)_{\tilde{\omega}} - \|\tau_{\text{dyn}}^{-1/2} \phi'\|_{\tilde{\omega}}^2, \quad (2.4.11)$$

has the desired form. However, this formulation is inconsistent because the small-scales are not forced by a full residual: the diffusive term is removed from the residual. The inconsistent character of the formulation limits the adequacy of this formulation.

#### 2.4.6 GLS with dynamic small-scales (GLSD)

Another alternative is to use the orthogonality only on the large-scale equation, now with double the magnitude. The diffusive stabilized term does not vanish but flips sign instead. In other words the VMS formulation does not convert to a SUPG formulation but to a GLS one. Hence, the VMS approach with the diffusive optimality projection (2.4.6) leads to the so-called GLSD-statement (the  $D$  stands for *dynamic*) which reads

Find  $\phi^h \in \mathcal{W}^h$  such that for all  $w^h \in \mathcal{W}^h$ ,

$$\begin{aligned} (w^h, \partial_t \phi^h + \partial_t \phi')_\Omega + (w^h, \mathbf{a} \cdot \nabla \phi^h)_\Omega + (\nabla w^h, \kappa \nabla \phi^h)_\Omega \\ - (\mathbf{a} \cdot \nabla w^h - \kappa \Delta w^h, \phi')_\Omega = (w^h, f)_\Omega \end{aligned} \quad (2.4.12a)$$

$$\partial_t \phi' + \tau_{\text{dyn}}^{-1} \phi' = - \mathcal{R} \phi^h. \quad (2.4.12b)$$

This formulation possesses the desired energy evolution:

$$\begin{aligned} \frac{d}{dt} \mathcal{E}_\omega = - \|\kappa^{1/2} \nabla \phi^h\|_\omega^2 + (\phi^h, f)_\omega - (1, \mathcal{F}_\omega^h)_{\chi_\omega} \\ - \|\tau_{\text{dyn}}^{-1/2} \phi'\|_{\tilde{\omega}}^2 + (\phi', f)_{\tilde{\omega}}. \end{aligned} \quad (2.4.13)$$

Comparing with the inconsistent SUPG formulation (2.4.10), both variational forms possess the correct-energy behavior. However, this formulation distinguishes itself by its consistent character, i.e. the forcing term in the small-scale equation is driven by the full residual.

**Remark 2.4.1.** *An important observation is that the GLS formulation is justified in a VMS context by invoking the orthogonality demanded for correct-energy behavior.*

## 2.5 BACK TO A VARIATIONAL MULTISCALE FORMULATION: A STABILIZED FORM WITH CORRECT-ENERGY EVOLUTION

Section 2.4 justifies with orthogonality arguments a GLS-based formulation depicting correct-energy behavior. That methodology *assumes* orthogonality between the large-scales and small-scales but does not actually *enforce* it. This section devises an alternative VMS stabilization approach that explicitly *enforces* the required orthogonality.

### 2.5.1 The small-scale solution space

The weak statements of Section 2.4 do not explicitly mention the solution space of the small-scales. The small-scales are effectively pointwise values, i.e.  $\phi' : \Omega \times \mathcal{I} \rightarrow \mathbb{R}$ . Section 2.3 reveals that the small-scales live in a properly defined space, that is  $\phi' \in \mathcal{W}'$ . The orthogonality projector (2.4.6) leads to the following definition of the small-scale space:

$$\mathcal{W}' = \mathcal{W}'_{H_0^1} := \left\{ \phi \in \mathcal{W}; \left( \kappa \Delta \eta^h, \phi \right)_\Omega = 0 \text{ for all } \eta^h \in \mathcal{W}^h \right\}. \quad (2.5.1)$$

Note that the projector is induced by the  $H_0^1$ -seminorm. This function space enjoys good properties, as indicated in [113]. The discretization dependence of the stabilization parameter  $\tau$  originates from the corresponding restricted Green's function. Consult that paper for details.

Employing the restricted solution space  $\mathcal{W}'$  the dynamic VMS formulation (2.4.4) subtly modifies to

Find  $\phi^h \in \mathcal{W}^h, \phi' \in \mathcal{W}'$  such that for all  $w^h \in \mathcal{W}^h$ ,

$$\begin{aligned} & (w^h, \partial_t \phi^h)_\Omega + (\nabla w^h, \kappa \nabla \phi^h)_\Omega + (w^h, \mathbf{a} \cdot \nabla \phi^h)_\Omega \\ & + (w^h, \partial_t \phi')_{\bar{\Omega}} - (\mathbf{a} \cdot \nabla w^h + \kappa \Delta w^h, \phi')_{\bar{\Omega}} - (w^h, f)_\Omega = 0, \end{aligned} \quad (2.5.2a)$$

$$\partial_t \phi' + \tau_{\text{dyn}}^{-1} \phi' + \mathcal{R} \phi^h = 0. \quad (2.5.2b)$$

The small-scale solution possesses the correct orthogonality by construction which implies the correct-energy behavior (2.4.13).

However, the restriction of the small-scale solution in the weak form (2.5.2) is troublesome to directly convert the weak statement into a working numerical method. This is mainly due to the infinite dimensionality of the small-scale space  $\mathcal{W}'$ .

### 2.5.2 Enforced orthogonality with a Lagrange multiplier (DO formulation)

In order to avoid dealing with the restricted subspace (2.5.1), we adopt a Lagrange multiplier setting to reformulate the problem into a mixed formulation. This opens up the search space for  $\phi'$ , while an explicit constraint is added to satisfy the required orthogonality. A Lagrange multiplier enforces this constraint. This formulation reads as follows:

Find  $(\phi^h, \sigma^h) \in \mathcal{W}^h \times \mathcal{W}^h, \phi' : \Omega \times \mathcal{I} \rightarrow \mathbb{R}$  such that for all  $(w^h, \eta^h) \in \mathcal{W}^h \times \mathcal{W}^h$ ,

$$\begin{aligned} & (w^h, \partial_t \phi^h)_\Omega + (\nabla w^h, \kappa \nabla \phi^h)_\Omega + (w^h, \mathbf{a} \cdot \nabla \phi^h)_\Omega \\ & + (w^h, \partial_t \phi')_{\bar{\Omega}} - (\mathbf{a} \cdot \nabla w^h + \kappa \Delta w^h, \phi')_{\bar{\Omega}} - (w^h, f)_\Omega = 0, \end{aligned} \quad (2.5.3a)$$

$$\partial_t \phi' + \tau_{\text{dyn}}^{-1} \phi' - \kappa \Delta \sigma^h + \mathcal{R} \phi^h = 0, \quad (2.5.3b)$$

$$\left( \kappa \Delta \eta^h, \phi' \right)_{\bar{\Omega}} = 0. \quad (2.5.3c)$$

We refer to it as *DO* where the *D* and *O* stand for *dynamic* and *orthogonal*, respectively. Here denotes  $\sigma^h$  the Lagrange multiplier and  $\eta^h$  its associated weighting function.

Note that this formulation asks for  $C^1$ -continuous basis functions because of the use of second derivatives. This additionally motivates our choice of employing IGA.

### 2.5.3 Local energy evolution of the formulation with enforced orthogonality

We obtain the energy evolution of the proposed method in a similar fashion as before. Hence, select  $w^h = \phi^h$  in the large-scale equation,  $\eta^h = \sigma^h + \phi^h$  in the third equation and multiply the small-scale equation by  $\phi'$ . Summation of the three equations and reordering leads to:

$$\begin{aligned} \frac{d}{dt} \mathcal{E}_\omega &= - \|\kappa^{1/2} \nabla \phi^h\|_\omega^2 + (\phi^h, f)_\omega - (1, \mathcal{F}_\omega^h)_{\chi_\omega} \\ &\quad - \|\tau_{\text{dyn}}^{-1/2} \phi'\|_{\bar{\omega}}^2 + (\phi', f)_{\bar{\omega}}. \end{aligned} \quad (2.5.4)$$

Note that the terms originating from (2.5.3c) exactly cancel the undesired orthogonality terms and the small-scale Lagrange multiplier term.

**Remark 2.5.1.** *The separate energy evolution of the large-scales and small-scales deduces in a similar fashion as above. The energies  $\mathcal{E}_\omega^h = \frac{1}{2}(\phi^h, \phi^h)_\omega$  and  $\mathcal{E}'_\omega = \frac{1}{2}(\phi', \phi')_{\bar{\omega}}$  evolve respectively as*

$$\begin{aligned} \frac{d}{dt} \mathcal{E}_\omega^h &= - \|\kappa^{1/2} \nabla \phi^h\|_\omega^2 + (\phi^h, f)_\omega + (\mathbf{a} \cdot \nabla \phi^h, \phi')_{\bar{\omega}} - (\phi^h, \partial_t \phi')_{\bar{\omega}} \\ &\quad - (1, \mathcal{F}_\omega^h)_{\chi_\omega}, \end{aligned} \quad (2.5.5a)$$

$$\frac{d}{dt} \mathcal{E}'_\omega = - \|\tau_{\text{dyn}}^{-1/2} \phi'\|_{\bar{\omega}}^2 + (\phi', f)_{\bar{\omega}} - (\mathbf{a} \cdot \nabla \phi^h, \phi')_{\bar{\omega}} - (\phi', \partial_t \phi^h)_{\bar{\omega}}. \quad (2.5.5b)$$

The first term of (2.5.5b) may be viewed as diffusion of the small-scales. The convective contributions exchange energy between the large-scales and small-scales. It is important to emphasize that these energies do not add up to the total local energy  $\mathcal{E}_\omega$ : the cross terms are missing. Their contributions appear in both (2.5.5a)-(2.5.5b).

## 2.6 TEMPORAL-INTEGRATION

This section is devoted to the time-integration for which we adopt the generalized- $\alpha$  time integrator. We start off with a brief recap of the method in a general setting, after which we discuss the use of this method for the small-scales particularly. The remainder presents the collection of time-integrators within this framework which yields a concrete energy evolution statement of consecutive time levels.

### 2.6.1 The generalized- $\alpha$ time integrator

We employ the generalized- $\alpha$  method for the temporal-integration [44]. The algorithm reads:

Given the data  $\phi_n, \dot{\phi}_n$ , find  $\phi_{n+\alpha_f}, \dot{\phi}_{n+\alpha_m}, \phi_{n+1}, \dot{\phi}_{n+1}$  such that

$$\dot{\phi}_{n+\alpha_m} = \mathcal{G}(\phi_{n+\alpha_f}), \quad (2.6.1a)$$

$$\phi_{n+1} = \phi_n + \Delta t ((1 - \gamma)\dot{\phi}_n + \gamma\dot{\phi}_{n+1}), \quad (2.6.1b)$$

$$\dot{\phi}_{n+\alpha_m} = (1 - \alpha_m)\dot{\phi}_n + \alpha_m\dot{\phi}_{n+1}, \quad (2.6.1c)$$

$$\phi_{n+\alpha_f} = (1 - \alpha_f)\phi_n + \alpha_f\phi_{n+1}. \quad (2.6.1d)$$

Here  $\partial_t \phi = \mathcal{G}(\phi)$  is the governing ordinary differential equation,  $\dot{\phi}$  is the discretized time derivative and the time step size is  $\Delta t = t_{n+1} - t_n$ . The scalars  $\alpha_f, \alpha_m, \gamma$  are algorithmic coefficients that need to be properly selected. The methods reduce to some of the classical time integrators for specific choices of the time-step parameters, e.g., for  $\alpha_f = \alpha_m = \gamma = 1$  to backward Euler and for  $\alpha_f = \alpha_m = \gamma = \frac{1}{2}$  to Crank-Nicolson. It is unconditionally stable if  $\alpha_m \geq \alpha_f \geq \frac{1}{2}$  (i.e. when it is more implicit than explicit). The second-order accuracy requirement reads [44, 200]:

$$\gamma = \frac{1}{2} + \alpha_m - \alpha_f. \quad (2.6.2)$$

### 2.6.2 Time-integration of the small-scales

The small-scale equations are ordinary differential equations. Employing (2.4.2) for (2.6.1a) an explicit solution of system (2.6.1) directly follows

$$\dot{\phi}'_{n+1} = C \left( -\frac{1}{\gamma\Delta t} \dot{\phi}'_n \left( 1 - \alpha_m + (1 - \gamma)\alpha_f \frac{\Delta t}{\tau_{\text{dyn}}} \right) - \frac{1}{\tau_{\text{dyn}}\gamma\Delta t} \dot{\phi}'_n - \frac{\mathcal{R}_{n+\alpha}^h}{\gamma\Delta t} \right), \quad (2.6.3)$$

with constant  $C = \alpha_f^{-1} (\tau_{\text{time}}^{-1} + \tau_{\text{dyn}}^{-1})^{-1}$ .

When using dynamic small-scales, the stabilizing properties of the weak formulation depend on the relation between the small- and the large-scales. This relation also enters in the Jacobian of the weak formulation. To this purpose we now explore this link. Let us define the so-called *effective stabilization parameter* as follows

$$\tau_{\text{eff}} = -\frac{\partial \dot{\phi}'_{n+\alpha_f}}{\partial \mathcal{R}_{n+\alpha}^h}, \quad (2.6.4)$$

inspired by (2.3.9a). In the case of static small-scales, depicted in (2.3.9), the trivial expression yields

$$\tau_{\text{eff}} = \tau_{\text{stat}} = \left( \tau_{\text{time}}^{-2} + \tau_{\text{dyn}}^{-2} \right)^{-1/2}. \quad (2.6.5)$$

When employing dynamic subscales as in (2.6.3), we get

$$\begin{aligned}
\tau_{\text{eff}} &= -\frac{\partial \phi'_{n+\alpha_f}}{\partial \mathcal{R}_{n+\alpha}^h} \\
&= -\frac{\partial \phi'_{n+\alpha_f}}{\partial \phi'_{n+1}} \cdot \frac{\partial \phi'_{n+1}}{\partial \phi'_{n+1}} \cdot \frac{\partial \phi'_{n+1}}{\partial \mathcal{R}_{n+\alpha}^h} \\
&= \alpha_f \cdot \gamma \Delta t \cdot \frac{C}{\gamma \Delta t} \\
&= \left( \alpha_m \alpha_f^{-1} \gamma^{-1} \Delta t^{-1} + \tau_{\text{dyn}}^{-1} \right)^{-1} \\
&= \left( \tau_{\text{time}}^{-1} + \tau_{\text{dyn}}^{-1} \right)^{-1}, \tag{2.6.6}
\end{aligned}$$

from which our definition of  $\tau_{\text{time}}$  is inspired:

$$\tau_{\text{time}} := \frac{\alpha_f \gamma \Delta t}{\alpha_m}. \tag{2.6.7}$$

The effective stabilization parameter  $\tau_{\text{eff}}$  is very similar to  $\tau_{\text{stat}}$  and shows the same asymptotic behavior. This modification of stabilization parameter effectively constitutes a change in the so-called *r-switch* [188] from  $r = 2$  to  $r = 1$ . The *r-switch* is a smooth approximation of the minimum operator. A high value of the integer  $r$  indicates a sharp switch. In fact, when the stabilization parameters are defined with the *r-switch* of  $r = 1$ :

$$\tilde{\tau}_{\text{stat}} = \left( \tau_{\text{conv}}^{-1} + \tau_{\text{diff}}^{-1} + \tau_{\text{time}}^{-1} \right)^{-1}, \tag{2.6.8a}$$

$$\tilde{\tau}_{\text{dyn}} = \left( \tau_{\text{conv}}^{-1} + \tau_{\text{diff}}^{-1} \right)^{-1}, \tag{2.6.8b}$$

the effective stabilization parameters would be identical.

### 2.6.3 Proper energy evolution

The energy evolution equations (2.4.13) or (2.5.4) reveal a (global) guaranteed energy decay in the absence of external forcing and boundaries, namely,

$$\frac{d}{dt} \mathcal{E}_\Omega = -\|\kappa^{1/2} \nabla \phi^h\|_\Omega^2 - \|\tau_{\text{dyn}}^{-1/2} \phi'\|_\Omega^2. \tag{2.6.9}$$

The time-integrator should obey this decaying property. Moreover, ideally it leads to a guaranteed decay of energy for consecutive time levels, that is,

$$\mathcal{E}_{n+1} \leq \mathcal{E}_n. \tag{2.6.10}$$

To arrive at an appropriate energy statement when employing the generalized- $\alpha$  method the procedure reads as follows. Multiply the small-scale equation with  $\phi'_{n+\alpha_f}$ , integrate the result and add it to the weak form in which  $w^h = \phi_{n+\alpha_f}^h$  is selected. The continuous form of this approach has been demonstrated earlier in this chapter, see e.g., Section 2.3.3. This leads to the correct symmetric inner products

for the spatial terms, and proper norms therefore. Additionally, the temporal terms, leading to the energy derivatives yield,

$$\Delta \mathcal{E} = \Delta t \dot{\mathcal{E}}_{n+\alpha} = \Delta t (\phi_{n+\alpha_f}, \dot{\phi}_{n+\alpha_m})_{\Omega}, \quad (2.6.11)$$

where  $\dot{\mathcal{E}}_{n+\alpha}$  is the temporal derivative of the energy at time level  $n + \alpha$  and  $\phi = \phi^h + \phi'$ . The abuse of notation demands the integration to be interpreted on  $\tilde{\Omega}$  for terms containing the small-scales. In the following we derive time-stepping parameters within the generalized- $\alpha$  time integrator framework which link this temporal term to a proper energy behavior.

Substitution of (2.6.1b)-(2.6.1d) into (2.6.11) yields:

$$\begin{aligned} \Delta t \dot{\mathcal{E}}_{n+\alpha} &= \Delta t (\phi_{n+\alpha_f}, \dot{\phi}_{n+\alpha_m})_{\Omega} \\ &= \Delta t ((1 - \alpha_f)\phi_n + \alpha_f\phi_{n+1}, (1 - \alpha_m)\dot{\phi}_n + \alpha_m\dot{\phi}_{n+1})_{\Omega} \\ &= ((1 - \alpha_f)\phi_n + \alpha_f\phi_{n+1}, \left(1 - \frac{\alpha_m}{\gamma}\right) \Delta t \dot{\phi}_n + \frac{\alpha_m}{\gamma} (\phi_{n+1} - \phi_n))_{\Omega} \\ &= -\frac{(1 - \alpha_f)\alpha_m}{\gamma} (\phi_n, \phi_n)_{\Omega} + \frac{\alpha_f\alpha_m}{\gamma} (\phi_{n+1}, \phi_{n+1})_{\Omega} + (1 - 2\alpha_f) \frac{\alpha_m}{\gamma} (\phi_n, \phi_{n+1})_{\Omega} \\ &\quad + \left( (1 - \alpha_f)\phi_n + \alpha_f\phi_{n+1}, \left(1 - \frac{\alpha_m}{\gamma}\right) \Delta t \dot{\phi}_n \right)_{\Omega}, \end{aligned} \quad (2.6.12)$$

The last term is precarious. The sign of the temporal derivative  $\dot{\phi}_n$  is not determined. It appears without  $(n + 1)$ -counterpart which leads to an uncontrollable last term. We remedy this issue by requiring the last term to vanish. This occurs when  $\alpha_m = \gamma$ . Following this road, the temporal term becomes

$$\begin{aligned} \Delta t \dot{\mathcal{E}}_{n+\alpha} &= \alpha_f (\phi_{n+1}, \phi_{n+1})_{\Omega} - (1 - \alpha_f) (\phi_n, \phi_n)_{\Omega} + (1 - 2\alpha_f) (\phi_n, \phi_{n+1})_{\Omega} \\ &= \mathcal{E}_{n+1} - \mathcal{E}_n + (\alpha_f - \frac{1}{2}) [(\phi_{n+1}, \phi_{n+1})_{\Omega} - 2(\phi_{n+1}, \phi_n)_{\Omega} + (\phi_n, \phi_n)_{\Omega}] \\ &= \mathcal{E}_{n+1} - \mathcal{E}_n + (\alpha_f - \frac{1}{2}) \|\phi_{n+1} - \phi_n\|_{\Omega}^2 \\ &= \mathcal{E}_{n+1} - \mathcal{E}_n + \Delta t^2 (\alpha_f - \frac{1}{2}) \|\dot{\phi}_{n+\alpha_m}\|_{\Omega}^2. \end{aligned} \quad (2.6.13)$$

where the last equality is a direct consequence of (2.6.1) with  $\alpha_m = \gamma$ . Henceforth, by combining this equation with (2.5.4) the discretized energy (of the DO form) satisfies

$$\begin{aligned} \frac{\mathcal{E}_{n+1} - \mathcal{E}_n}{\Delta t} + \Delta t (\alpha_f - \frac{1}{2}) \|\dot{\phi}_{n+\alpha_m}\|_{\Omega}^2 &= -\|\kappa^{1/2} \nabla \phi_{n+\alpha_f}^h\|_{\Omega}^2 + (\phi_{n+\alpha_f}^h, f)_{\Omega} \\ &\quad - \|\tau_{\text{dyn}}^{-1/2} \phi'_{n+\alpha_f}\|_{\tilde{\Omega}}^2 + (\phi'_{n+\alpha_f}, f)_{\tilde{\Omega}}. \end{aligned} \quad (2.6.14)$$

The trivially equivalent form

$$\begin{aligned} \mathcal{E}_{n+1} &= \mathcal{E}_n - \Delta t^2 (\alpha_f - \frac{1}{2}) \|\dot{\phi}_{n+\alpha_m}\|_{\Omega}^2 - \Delta t \|\kappa^{1/2} \nabla \phi_{n+\alpha_f}^h\|_{\Omega}^2 - \Delta t \|\tau_{\text{dyn}}^{-1/2} \phi'_{n+\alpha_f}\|_{\tilde{\Omega}}^2 \\ &\quad + \Delta t (\phi_{n+\alpha_f}^h, f)_{\Omega} + \Delta t (\phi'_{n+\alpha_f}, f)_{\tilde{\Omega}} \end{aligned} \quad (2.6.15)$$

reveals that a decay of the discretized energy is guaranteed when, in absence of forcing, additionally  $\alpha_f \geq \frac{1}{2}$ . The first term on the right-hand side, which again should be interpreted on  $\tilde{\Omega}$  for the small-scales, is numerical diffusion which vanishes for  $\alpha_f = \frac{1}{2}$ . Hence, the parameter family  $\alpha_f \geq \frac{1}{2}, \alpha_m = \gamma$ , which includes the Crank-Nicolson time-integrator, can be linked to a proper energy decay. Notice that for  $\alpha_f = \frac{1}{2}$  the stability constraint is fulfilled and the second-order accuracy condition (2.6.2) is not harmed.

## 2.7 NUMERICAL VERIFICATION

This section provides the numerical verification of the proposed variational formulations of the Sections 2.3-2.5 on a model problem. We focus on the energy behavior on both a global and a local level. First, the energy behavior is assessed verifying the overall performance of the newly proposed methods. Next, we zoom in on the effect of the small-scales on the energy dissipation. We assess its global evolution and local distribution and examine the contributions of the unwanted terms.

## 2.7.1 Model problem description

The problem under consideration is convection skew to the mesh on a  $1 \times 1$ -domain with periodic boundaries. The velocity is  $\mathbf{a} = (1, 1)$ , therefore the profile loops once through the mesh and arrives at its start position at  $t = 1.0$ . The diffusivity is set to  $\kappa = 5 \times 10^{-4}$ . No external forcing is applied. The initial condition is a sharp block of the form:

$$\phi_0(\mathbf{x}) = H(|x - \frac{1}{2}|)H(|y - \frac{1}{2}|), \quad (2.7.1a)$$

$$H(z) = \begin{cases} 1 & z < l \\ 0 & l < z, \end{cases} \quad (2.7.1b)$$

where  $l$  is a specified length. For the discretization we employ NURBS<sup>4</sup>. Linear NURBS are not considered as they would eliminate the diffusive stabilization term  $(\kappa \Delta \phi^h, \phi')_{\Omega}$  and hence the stabilized forms (SUPG, VMS and GLS) coincide. All our implementations use quadratic NURBS to bypass this effect. The sharpest profile that can be exactly represented on the mesh has the form:

$$\phi_0(\mathbf{x}) = \hat{H}(|x - \frac{1}{2}|)\hat{H}(|y - \frac{1}{2}|), \quad (2.7.2a)$$

$$\hat{H}(z) = \begin{cases} 1 & z < l_0 \\ 1 - \frac{(z-l_0)^2}{2h_c^2} & l_0 < z < l_1 \\ \frac{(l_2-z)^2}{2h_c^2} & l_1 < z < l_2 \\ 0 & l_2 < z, \end{cases} \quad (2.7.2b)$$

where  $l_0$ ,  $l_1$  and  $l_2$  are specified lengths of the different segments that have to coincide with mesh lines. We will use  $16 \times 16$ ,  $32 \times 32$  and  $64 \times 64$ -element meshes. As we want to verify the behavior of the method itself and not consider the error in representing the initial condition we use the exact same initial condition on all meshes. This is in this case the initial condition of the  $16 \times 16$ -element mesh. Therefore we choose  $l_0 = nh_c$ ,  $l_1 = (n+1)h_c$  and  $l_2 = (n+2)h_c$  with  $n = 2$  and  $h_c = \frac{1}{16}$ .

The implementations use the energy-conserving time-integrator of Section 2.6 with  $\alpha_f = \frac{1}{2}$  motivated by both the second-order temporal accuracy and the stability. All computations use a CFL number of 0.5.

<sup>4</sup> Note that the quadratic NURBS reduce to B-splines on our uniform Cartesian mesh.

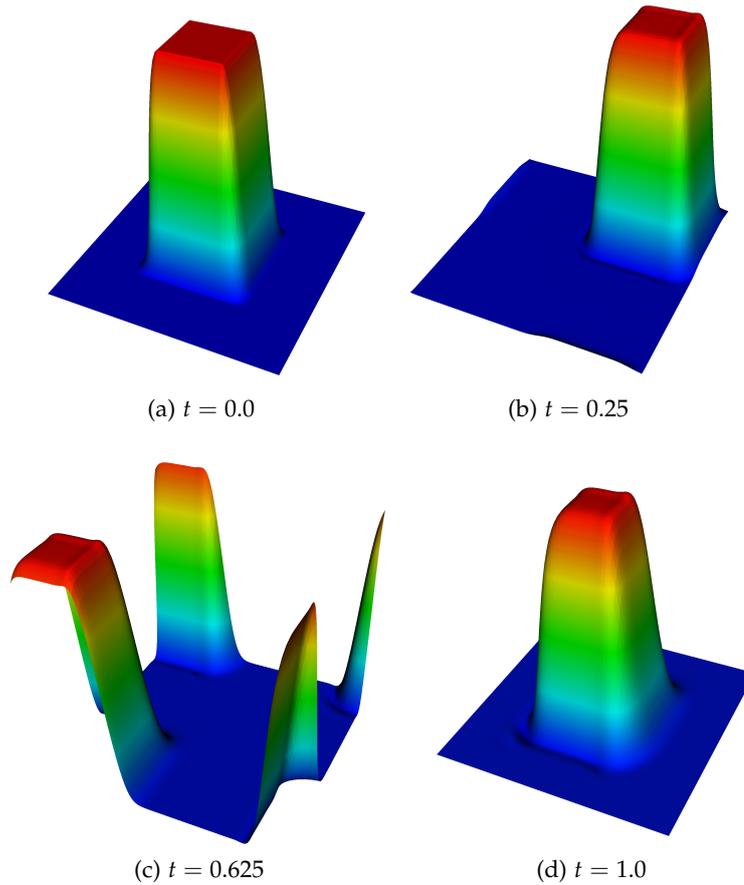


Figure 2.3: Time evolution of the block profile convected through the mesh.

Figure 2.3 shows the profile traveling through the mesh from  $t = 0$  until  $t = 1.0$ . The profile exits the mesh approximately halfway during the simulation (at  $t = 0.5$ ) and enters at the opposite corner due to the periodic boundary conditions.

In the following we present energy evolution results for three different methods: (i) the SUPG method with static small-scales (SUPGS), (ii) the GLS method with dynamic small-scales (GLSD) and (iii) the dynamic orthogonal formulation (DO). These were chosen because the last two exhibit the correct-energy behavior, while SUPG with static small-scales is the classical approach and serves as a reference. It turns out that all methods with static small-scales show very similar behavior.

### 2.7.2 Overall energy behavior

Figure 2.4 displays the energy behavior for various methods on different meshes. It shows convergence of the energy evolution for each one of the methods. For the SUPGS we have two alternative energy definitions, i.e. one based on only the large-scales and one based on both large- and small-scales (denoted as total energy). The energy behavior on the  $16 \times 16$ -mesh is not converged yet whereas the energy on the  $32 \times 32$ -mesh already closely follows that of the finer meshes. In the following we

study in more detail the energy evolution on a  $32 \times 32$ -mesh. At this stage there is no visible difference between these solutions.

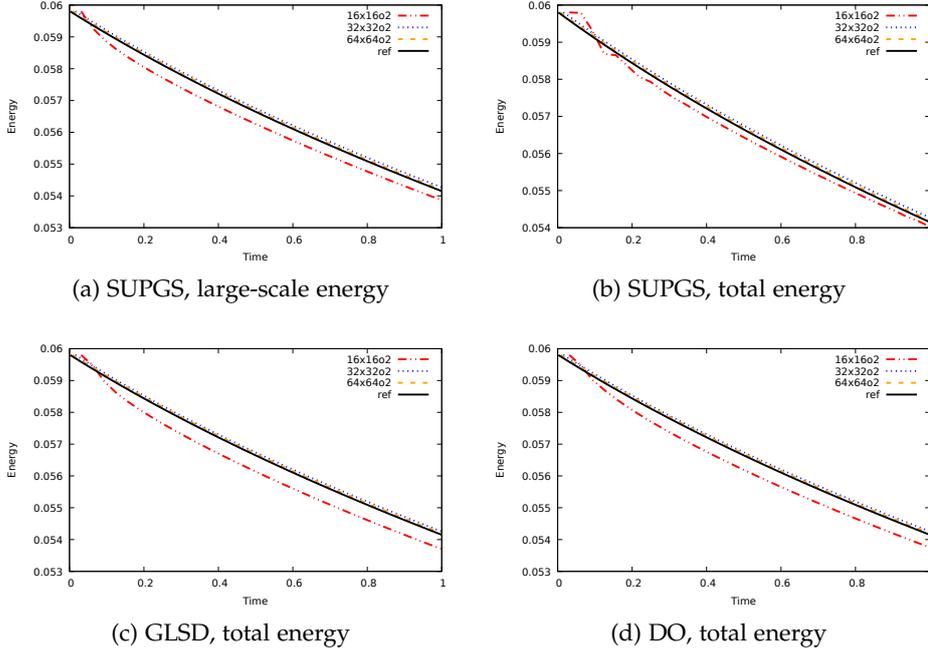


Figure 2.4: Energy evolution for various meshes: (a) the large-scale energy for SUPGS and the total energy for (b) SUPGS, (c) GLSD and (d) DO. An overkill reference solution is added (the continuous black line).

### 2.7.3 Energy dissipation by the small-scales

Here we study the effect of the small-scales on the energy dissipation. The choice  $\alpha_f = \alpha_m = \gamma = \frac{1}{2}$  removes the effect of the time-integrator on the energy dissipation. The energy evolution for SUPGS takes the form:

$$\begin{aligned} \frac{d}{dt} \mathcal{E}_\omega^h &= - \|\kappa^{1/2} \nabla \phi^h\|_\omega^2 - (1, \mathcal{F}_\omega^h)_{\chi_\omega} - \|\tau_{\text{stat}}^{-1/2} \phi'\|_{\tilde{\omega}}^2 \\ &\quad + (\kappa \Delta \phi^h, \phi')_{\tilde{\omega}} - (\phi', \partial_t \phi^h)_{\tilde{\omega}}, \end{aligned} \quad (2.7.3a)$$

$$\begin{aligned} \frac{d}{dt} \mathcal{E}_\omega &= - \|\kappa^{1/2} \nabla \phi^h\|_\omega^2 - (1, \mathcal{F}_\omega^h)_{\chi_\omega} \\ &\quad - \|\tau_{\text{stat}}^{-1/2} \phi'\|_{\tilde{\omega}}^2 + (\kappa \Delta \phi^h, \phi')_{\tilde{\omega}} + (\partial_t \phi', \phi^h + \phi')_{\tilde{\omega}} \end{aligned} \quad (2.7.3b)$$

for the large-scale energy and the total energy respectively. The GLSD method and the DO formulation show correct-energy evolution:

$$\frac{d}{dt} \mathcal{E}_\omega = - \|\kappa^{1/2} \nabla \phi^h\|_\omega^2 - (1, \mathcal{F}_\omega^h)_{\chi_\omega} - \|\tau_{\text{dyn}}^{-1/2} \phi'\|_{\tilde{\omega}}^2. \quad (2.7.4)$$

The right-hand side terms are evaluated at time level  $n + 1/2$ . The last three terms of each of (2.7.3) and the last term of (2.7.4) represent the small-scale contribution to the energy dissipation. Figures 2.5 displays the evolution of the small-scale contribution to energy dissipation on a global scale.

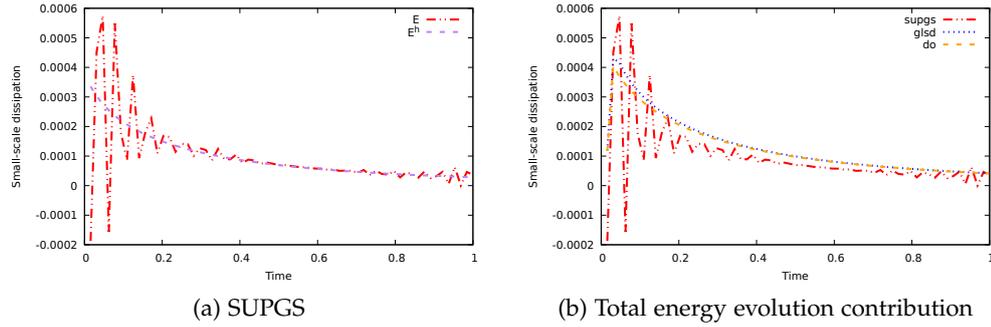


Figure 2.5: Evolution of the small-scale contribution to energy dissipation on a global scale: (a) large-scale ( $E^h$ ) and total energy ( $E$ ) for SUPG with static small-scales, (b) total energy for the three methods.

As anticipated GLSD and DO show positive energy dissipation. On the other hand it is clear that, when considering the total energy, SUPGS has problematic dissipation behavior. It shows severe wiggles resulting in undershoots with negative dissipation. However, when considering large-scale energy there seems to be no problem.

Figure 2.6 shows a typical local distribution of the small-scale dissipation. These largely confirm the findings from Figure 2.5. GLSD and DO show strictly positive energy dissipation throughout the domain. For SUPGS now both energy definitions show problems, as the dissipation becomes negative in certain parts of the domain. Hence, *despite global energy decay, local energy creation cannot be precluded*.

In the following we further analyze the energy dissipation by considering the contribution of (i) the temporal terms (the last terms on the right-hand side of (2.7.3)) and (ii) the orthogonality term ( $\kappa \Delta \phi^h, \phi'$ ).

#### 2.7.4 Temporal-term

Figures 2.7 and 2.8 show the magnitude of the temporal terms for SUPGS on both global and local level, respectively. The temporal term of the total energy has larger values than that of the global one. Both energy definitions show negative dissipation, globally as well as locally. Hence contributions of these terms are undesirable. Comparing with Figure 2.6 we observe that the temporal has a major contribution to the small-scale dissipation in this case.

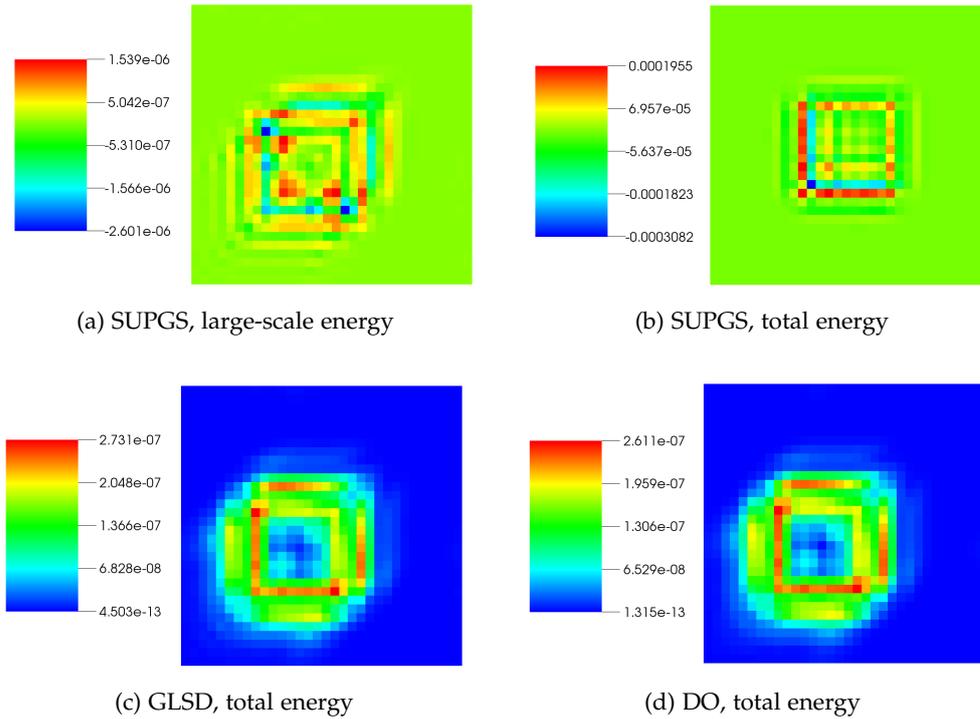


Figure 2.6: Small-scale contribution to energy dissipation on a local scale (at  $t = 1.0$ ): (a) the large-scale energy for SUPGS and the total energy for (b) SUPGS, (c) GLSD and (d) DO.

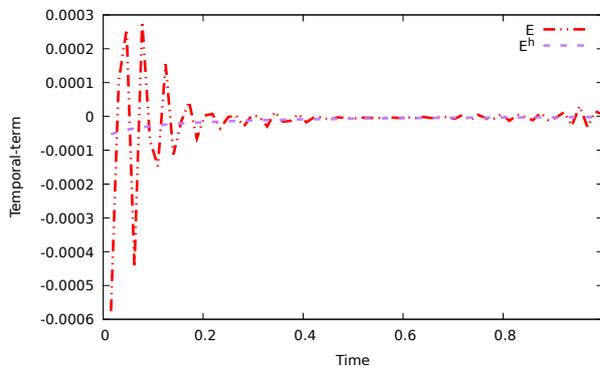


Figure 2.7: Temporal contribution to small-scale energy dissipation on a global scale for the SUPGS method. The contributions to both the large-scale ( $E^h$ ) and total energy ( $E$ ) are displayed.

### 2.7.5 Orthogonality-term

Before continuing we would like to stress that the orthogonality term  $(\kappa \Delta \phi^h, \phi')$  plays different roles in the formulations. In case of SUPGS it is directly an error in the energy behavior, while for GLSD this is an error in the assumed scale separation projector that leads to the correct behavior. Obviously, for DO the orthogonality term should vanish.

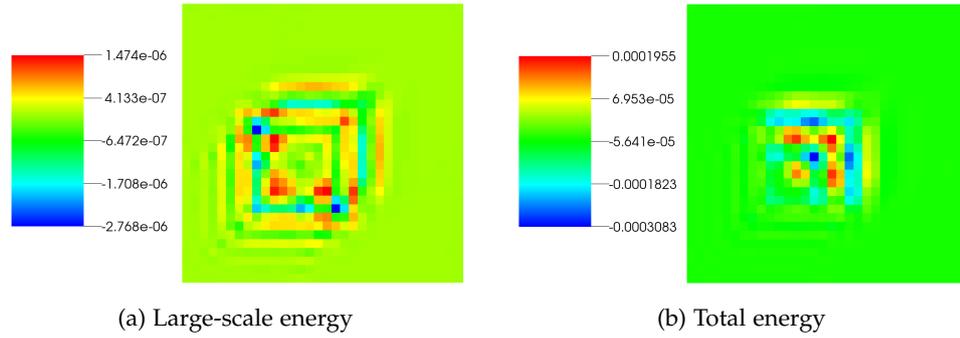


Figure 2.8: Temporal contribution to small-scale energy dissipation on a local scale for the SUPGS method (at  $t = 1.0$ ). The contributions to both the large-scale and total energy are displayed.

The global and local behavior of the orthogonality term is displayed in the Figures 2.9 and 2.10 respectively. These confirm that the orthogonality term vanishes for the DO formulation. For the other methods this is not the case. The global orthogonality has an undetermined sign. Moreover, locally the contribution can be negative while the overall contribution is positive.

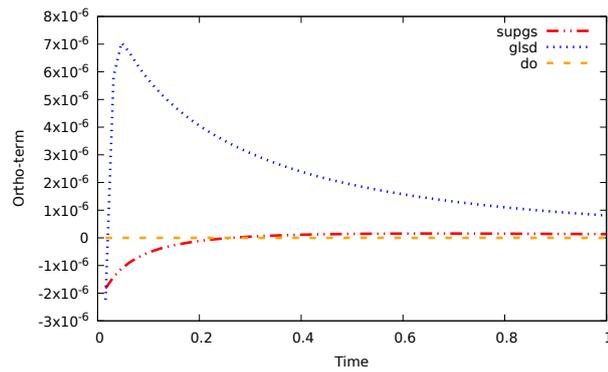


Figure 2.9: Time evolution of the global orthogonality-term  $(\kappa \Delta \phi^h, \phi')_{\Omega}$  for SUPGS, GLSD and DO.

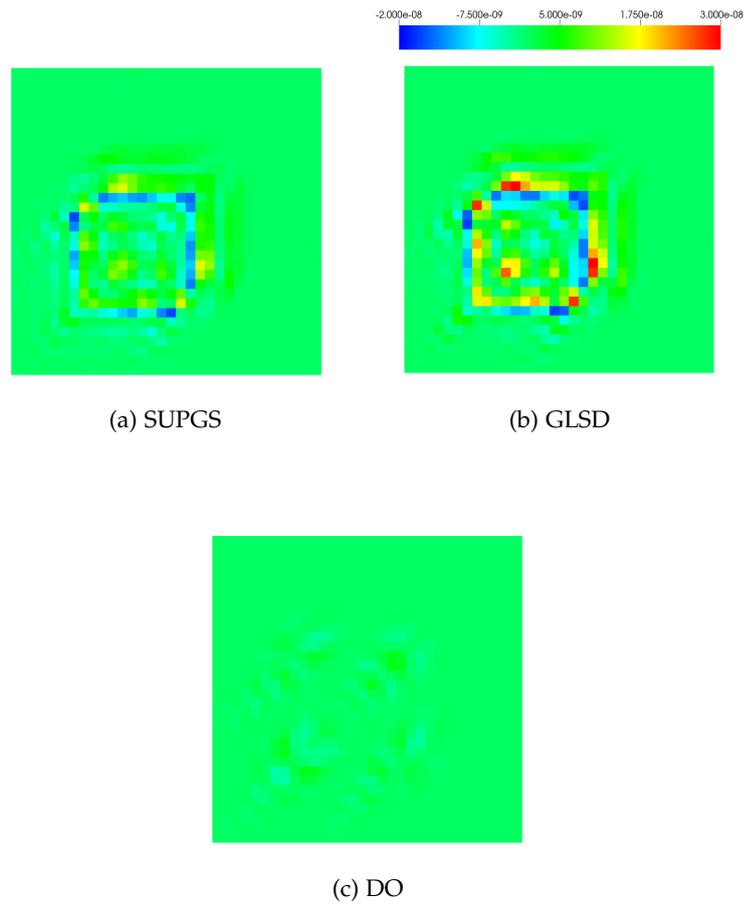


Figure 2.10: Local behavior of the orthogonality-term  $(\kappa\Delta\phi^h, \phi^l)_{\tilde{\omega}}$  (at  $t = 1.0$ ) for SUPGS, GLSD and DO.

## 2.8 CONCLUSIONS

In this work we have proposed an approach to rectify the incorrect-energy behavior of the standard stabilized methods. To this purpose we have employed the concepts of orthogonal small-scales and the dynamic behavior of the small-scales.

This work takes a road through the various standard weak formulations. The standard Galerkin shows correct-energy evolution but suffers from stability issues. Standard stabilized methods display the opposite. Starting from the variational multiscale approach, we have formulated a design condition to step-by-step remedy the incorrect-energy behavior. The first part towards rectification employs dynamic behavior of the small-scales and henceforth leads to a variational multiscale approach with dynamic small-scales. Next, an orthogonality demand of the large- and small-scales, which can be understood as a  $H_0^1$ -projection operator, appears. This leads to several options for the variational formulation. It links the form to, both employing dynamic small-scales, the streamline-upwind Petrov-Galerkin method or the Galerkin/least-squares method of which the latter one, in contrast to the former one, displays the energy behavior aimed at. Explicitly enforcing the orthogonality of the large- and small-scales returns us to the variational multiscale framework with the correct-energy behavior.

The proposed variational formulations which depict correct-energy behavior are:

- the Galerkin/least-squares formulation with dynamic small-scales (GLSD)
- the approach with dynamic orthogonal small-scales (DO)

Numerical results show that the energy convergence of the novel methods displays very similar performance in comparison with the existing stabilized finite element methods. However, the standard methods display both positive and negative small-scale contributions to energy dissipation. The GLSD and the DO method do not suffer from these deficiencies. Furthermore, the numerical results show activity of the unwanted terms in the standard stabilized forms and confirm the enforced orthogonality of the large- and small-scales. The numerical computations have been performed with isogeometric analysis, which is required for a solenoidal velocity field and seems a natural choice when employing orthogonal small-scales.

This work serves as an important first step for generalizations in other contexts. Future work will entail a similar methodology for the incompressible Navier-Stokes equations.

---

## CORRECT ENERGY EVOLUTION OF STABILIZED INCOMPRESSIBLE NAVIER-STOKES

---

This chapter is reproduced from [58]:

M.F.P. ten Eikelder and I. Akkerman, *Correct energy evolution of stabilized formulations: The relation between VMS, SUPG and GLS via dynamic orthogonal small-scales and isogeometric analysis. II: The incompressible Navier-Stokes equations*, *Comput. Meth. Appl. Mech. Engrg.* 340 (2018) 1135-1159.

### ABSTRACT

*This chapter presents the construction of a correct-energy stabilized finite element method for the incompressible Navier-Stokes equations. The framework of the methodology and the correct-energy concept have been developed in the convective–diffusive context in the preceding chapter [M.F.P. ten Eikelder, I. Akkerman, Correct energy evolution of stabilized formulations: The relation between VMS, SUPG and GLS via dynamic orthogonal small-scales and isogeometric analysis. I: The convective–diffusive context, *Comput. Methods Appl. Mech. Engrg.* 331 (2018) 259–280]. The current work extends ideas of the preceding chapter to build a stabilized method within the variational multiscale (VMS) setting which displays correct-energy behavior for the incompressible Navier-Stokes equations. Similar to the convection–diffusion case, a key ingredient is the proper dynamic and orthogonal behavior of the small-scales. This is demanded for correct energy behavior and links the VMS framework to the streamline-upwind Petrov-Galerkin (SUPG) and the Galerkin/least-squares method (GLS).*

*The presented method is a Galerkin/least-squares formulation with dynamic divergence-free small-scales (GLSDD). It is locally mass-conservative for both the large- and small-scales separately. In addition, it locally conserves linear and angular momentum. The computations require and employ NURBS-based isogeometric analysis for the spatial discretization. The resulting formulation numerically shows improved energy behavior for turbulent flows comparing with the original VMS method.*

### 3.1 INTRODUCTION

The creation of artificial energy in numerical methods is undesirable from both a physical and a numerical stability point of view. Therefore methods precluding this deficiency are often sought after. This work continues the construction of the correct-energy displaying stabilized finite element methods. The first episode [57] exposes the developed methodology in the convective–diffusive context. The current study deals with the incompressible Navier–Stokes equations and is the second piece of work within the framework. The setup of this chapter is closely related to that of [57]. In particular, the *correct-energy* demand is the same, thus it represents that the method (i) does not create artificial energy and (ii) closely resembles the energy evolution of the continuous setting. The precise definition is stated in Section 3.4. What sets the Navier–Stokes problem apart from convection–diffusion case is the inclusion of the incompressibility constraint. In this work we use a divergence-conforming basis which allows exact pointwise satisfaction of this constraint. This is considered a beneficial property. Therefore it is added as a design criterion. In a two-phase context this property is essential for correct energy behavior [6].

#### 3.1.1 Contributions of this work

This work derives a novel VMS formulation which exhibits the correct energy behavior and to this purpose combines several ingredients. The final formulation is summarized in Appendix 3.A. The new method is a residual-based approach that employs (i) dynamic behavior of the small-scales, (ii) solenoidal NURBS basis functions and (iii) a Lagrange-multiplier construction to ensure the incompressibility of the small-scale velocities. The formulation is of skew-symmetric type, rather than conservative, which is motivated by both the correct-energy demand and its improved behavior in the single scale setting (i.e. the Galerkin method) [117]. Moreover, the formulation reduces to a Galerkin formulation in case of a vanishing Reynolds number due to a Stokes-projector. The use of dynamic small-scales, firstly proposed in [46], is also driven from an energy point of view. In addition, it leads to global momentum conservation and the numerical results of [48] show improved behavior of the dynamic small-scales with respect to their static counterpart.

#### 3.1.2 Context

This work falls within the variational multiscale framework [100, 103]. The basic idea of this method is to split solution into the large/resolved-scales and small/unresolved-scales. The small-scales are modeled in terms of (the residual of) the large scales and substituted into the equation for the large-scales. This approach was first applied in a residual-based LES context to incompressible turbulence computations in [17]. The VMS methodology has enjoyed a lot of progress since then. For an overview of the development consult the review paper [47].

Our work is not the first to analyze the energy behavior of the VMS method. A spectral analysis of the VMS method can be found in [199]. That paper proves dissipation of the model terms under restrictive conditions. Additional to the

optimality projector, they require  $L_2$ -orthogonality of the large- and small-scales. This condition naturally leads to the use of spectral methods.

Principe et al. [159] provide a precise definition of the numerical dissipation within the variational multiscale context for incompressible flows. Equally important, they numerically show that the concept of dynamic small-scales, which we apply in this work, is able to model turbulence.

Colomés et al. [49] assess the performance of several VMS methods for turbulent flow problems and provide an energy analysis of these methods. They conclude that algebraic subgrid scales (ASGS) and orthogonal subscales (OSS) yield similar results, whereas the latter one is more convenient in terms of numerical performance.

We build onto [49, 159, 199] without requiring  $L_2$ -orthogonality. Therefore we are not restricted to the use of spectral methods, while retaining a strict energy relation.

Other recent related work includes the IGA divergence-conforming VMS method of Opstal et al. in [154]. They also employ an  $H_0^1$ -orthogonality between the velocity large- and small-scales on a local level. Our work deviates from [154] in that we motivate the required orthogonalities with the correct energy demand. Furthermore, our work distinguishes itself by enforcing the divergence-free velocity small-scales with a Lagrange-multiplier construction. We believe that the Stokes orthogonality between the large- and small-scales is a natural path to take, since it reduces the scheme to the Galerkin method in the vanishing Reynolds number limit.

The discretizations throughout this work are based on the isogeometric analysis (IGA) concept, proposed by Hughes et al. in [101]. This idea integrates the historically distinct fields of computer aided design (CAD) and finite element analysis. Isogeometric analysis rapidly became a valuable tool in computational fluid dynamics, in particular in turbulence computations. It provides several advantages over standard finite element analysis, including an exact description of CAD geometries, increased robustness and superior approximation properties [4, 101, 142]. This work requires in particular inf-sup stable discretizations for which we use [68, 69]. Moreover these spaces allow the pointwise satisfaction of the incompressibility constraint. The smooth NURBS basis functions are convenient for the computation of second derivatives.

### 3.1.3 Outline

The organization of this chapter in Section 3.2 and 3.3 is very comparable with that of the convective–diffusive context [57], and at some points mirrors it. The purpose thereof is (i) to indicate the great similarities of the methodologies and (ii) to clarify the approach. The remainder of this work presents the actual construction of a stabilized variational formulation for the incompressible Navier–Stokes equations which displays correct-energy behavior. We summarize it as follows. Section 3.2 states the continuous form of the governing incompressible flow equations, both in the strong formulation and the standard weak formulation. It additionally provides the energy evolution of the continuous equation, in both global and local form. Section 3.3 discusses the energy evolution of the variational multiscale approach with dynamic small-scales. The path toward correct energy behavior actually starts in Section 3.4. This section presents the required orthogonality of the large-scales and small-scales. This converts the residual-based variational multiscale method

into the Galerkin/least-squares method with the correct energy behavior. Section 3.5 presents conservation properties of the method. Section 3.6 provides a computational test case, namely a three-dimensional Taylor–Green vortical flow. In particular it examines the energy behavior and compares the novel method with the standard VMS method with static small-scales [17]. The calculations employ the generalized- $\alpha$  method with favorable energy behavior which is also discussed in [57]. In Section 3.7, we wrap up and present avenues for future research.

## 3.2 THE CONTINUOUS INCOMPRESSIBLE NAVIER–STOKES EQUATIONS

### 3.2.1 Strong formulation

Let  $\Omega \in \mathbb{R}^d$ ,  $d = 2, 3$ , denote the spatial domain and  $\partial\Omega = \Gamma = \Gamma_g \cup \Gamma_h$  its boundary, see Figure 3.1.

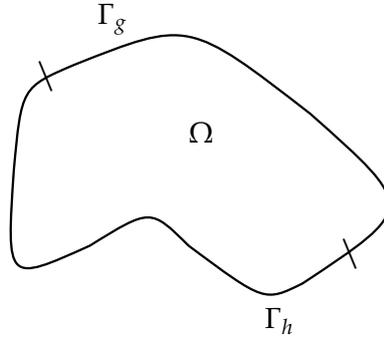


Figure 3.1: Spatial domain  $\Omega$  with its boundaries  $\Gamma = \Gamma_g \cup \Gamma_h$ . This is the same figure as in [57].

The problem consists of solving the incompressible Navier–Stokes equations governing the fluid flow, which read in strong form

$$\partial_t \mathbf{u} + \nabla \cdot (\mathbf{u} \otimes \mathbf{u}) + \nabla p - \nabla \cdot (2\nu \nabla^s \mathbf{u}) = \mathbf{f} \quad \text{in } \Omega \times \mathcal{I}, \quad (3.2.1a)$$

$$\nabla \cdot \mathbf{u} = 0 \quad \text{in } \Omega \times \mathcal{I}, \quad (3.2.1b)$$

$$\mathbf{u} = \mathbf{g} \quad \text{in } \Gamma_g \times \mathcal{I}, \quad (3.2.1c)$$

$$-u_n^- \mathbf{u} - pn + \nu \partial_n \mathbf{u} = \mathbf{h} \quad \text{in } \Gamma_h \times \mathcal{I}, \quad (3.2.1d)$$

$$\mathbf{u}(\mathbf{x}, 0) = \mathbf{u}_0(\mathbf{x}) \quad \text{in } \Omega, \quad (3.2.1e)$$

for the velocity  $\mathbf{u} : \Omega \times \mathcal{I} \rightarrow \mathbb{R}^d$  and the pressure divided by the density  $p : \Omega \times \mathcal{I} \rightarrow \mathbb{R}$ . A constant density is assumed. Eqs. (3.2.1a)–(3.2.1e) describe the balance of linear momentum, the conservation of mass, the inhomogeneous Dirichlet boundary condition, the traction boundary condition and the initial conditions, respectively. The spatial coordinate denotes  $\mathbf{x} \in \Omega$  and the time denotes  $t \in \mathcal{I} = (0, T)$  with end time  $T > 0$ . The given dynamic viscosity is  $\nu : \Omega \rightarrow \mathbb{R}^+$ , the body force is  $\mathbf{f} : \Omega \times \mathcal{I} \rightarrow \mathbb{R}^d$ , the initial velocity is  $\mathbf{u}_0 : \Omega \rightarrow \mathbb{R}^d$  and the boundary data are  $\mathbf{g} : \Gamma_g \times \mathcal{I} \rightarrow \mathbb{R}^d$  and  $\mathbf{h} : \Gamma_h \times \mathcal{I} \rightarrow \mathbb{R}^d$ . We assume a zero-average pressure for all  $t \in \mathcal{I}$  in case of an empty Neumann boundary. The normal velocity denotes

$u_n = \mathbf{u} \cdot \mathbf{n}$  with positive and negative parts  $u_n^\pm = \frac{1}{2}(u_n \pm |u_n|)$ . The various derivative operators are the temporal one  $\partial_t$ , the symmetric gradient  $\nabla^s \cdot = \frac{1}{2}(\nabla \cdot + \nabla^T \cdot)$  and the normal gradient  $\partial_n = \mathbf{n} \cdot \nabla$ , with  $\mathbf{n}$  the outward unit normal.

### 3.2.2 Weak formulation

Let  $\mathcal{W}^0$  denote the trial weighting function space satisfying the homogeneous Dirichlet conditions on  $\mathbf{u}$  and  $\mathcal{W}^8$  the trial solution space with non-homogeneous Dirichlet conditions on  $\mathbf{u}$ . The standard variational formulation writes:

Find  $\{\mathbf{u}, p\} \in \mathcal{W}^8$  such that for all  $\{\mathbf{w}, q\} \in \mathcal{W}^0$ ,

$$B_{\Omega, \Gamma_h}(\{\mathbf{u}, p\}, \{\mathbf{w}, q\}) = L_{\Omega, \Gamma_h}(\{\mathbf{w}, q\}), \quad (3.2.2a)$$

where

$$B_{D, \Gamma_h}(\{\mathbf{u}, p\}, \{\mathbf{w}, q\}) = B_D(\{\mathbf{u}, p\}, \{\mathbf{w}, q\}) + (\mathbf{w}, u_n^+ \mathbf{u})_{\Gamma_h(D)}, \quad (3.2.2b)$$

$$L_{D, \Gamma_h}(\{\mathbf{w}, q\}) = L_D(\{\mathbf{w}, q\}) + (\mathbf{w}, \mathbf{h})_{\Gamma_h(D)}, \quad (3.2.2c)$$

$$\begin{aligned} B_D(\{\mathbf{u}, p\}, \{\mathbf{w}, q\}) &= (\mathbf{w}, \partial_t \mathbf{u})_D - (\nabla \mathbf{w}, \mathbf{u} \otimes \mathbf{u})_D + (\nabla \mathbf{w}, 2\nu \nabla^s \mathbf{u})_D \\ &\quad + (q, \nabla \cdot \mathbf{u})_D - (\nabla \cdot \mathbf{w}, p)_D, \end{aligned} \quad (3.2.2d)$$

$$L_D(\{\mathbf{w}, q\}) = (\mathbf{w}, \mathbf{f})_D. \quad (3.2.2e)$$

Here  $B_D$  is the bilinear form and  $(\cdot, \cdot)_D$  is the  $L^2(D)$  inner product over  $D$ . The Dirichlet and traction boundary of domain  $D$  denote  $\Gamma_g(D) := \Gamma_g \cap \partial D$  and  $\Gamma_h(D) := \Gamma_h \cap \partial D$  respectively. The strong (3.2.1) and the weak formulation (3.2.2) are equivalent for smooth solutions.

**Remark 3.2.1.** *The variational form (3.2.2) is of conservative type: the incompressibility constraint (3.2.1b) is not directly employed in the convective terms. A discretization of the conservative form may lead to spurious oscillations caused by the error in the incompressibility constraint acting as a distribution of sinks and sources. Employing (3.2.1b) can be used to generate a convective form which is sometimes preferred and often adopted in Galerkin computations [117]. Here we write the variational formulation of skew-symmetric type which will be used in Section 3.4:*

Find  $\{\mathbf{u}, p\} \in \mathcal{W}^8$  such that for all  $\{\mathbf{w}, q\} \in \mathcal{W}^0$ ,

$$C_{\Omega, \Gamma_h}(\{\mathbf{u}, p\}, \{\mathbf{w}, q\}) = L_{\Omega, \Gamma_h}(\{\mathbf{w}, q\}), \quad (3.2.3a)$$

where

$$C_{D, \Gamma_h}(\{\mathbf{u}, p\}, \{\mathbf{w}, q\}) = C_D(\{\mathbf{u}, p\}, \{\mathbf{w}, q\}) + \frac{1}{2}(\mathbf{w}, |u_n| \mathbf{u})_{\Gamma_h(D)}, \quad (3.2.3b)$$

$$\begin{aligned} C_D(\{\mathbf{u}, p\}, \{\mathbf{w}, q\}) &= (\mathbf{w}, \partial_t \mathbf{u})_D + \frac{1}{2}(\mathbf{w}, \mathbf{u} \cdot \nabla \mathbf{u})_D - \frac{1}{2}(\mathbf{u} \cdot \nabla \mathbf{w}, \mathbf{u})_D \\ &\quad + (\nabla \mathbf{w}, 2\nu \nabla^s \mathbf{u})_D + (q, \nabla \cdot \mathbf{u})_D - (\nabla \cdot \mathbf{w}, p)_D. \end{aligned} \quad (3.2.3c)$$

Again, this form is equivalent to the strong form (3.2.1). Form (3.2.3) does not possess all conservation properties when discretized in a standard way. However, this can be restored using a multiscale split, see [117] for details. In the following we continue with the conservative form (3.2.2).

**Remark 3.2.2.** Concerning the viscosity term, we apply the incompressibility constraint to remove the non-symmetric part. The divergence-conforming spaces permit this action also in a discrete sense.

To obtain the energy evolution linked to (3.2.1) we want to substitute  $\boldsymbol{w} = \boldsymbol{u}$ . This is not possible in (3.2.2) due to the different boundary conditions of the solution and test function spaces. The enforcement of the Dirichlet boundary conditions in the spaces bypasses when employing a Lagrange multiplier construction. This converts the variational formulation into a *mixed formulation*:

Find  $(\{\boldsymbol{u}, p\}, \boldsymbol{\lambda}_\Omega) \in \mathcal{W} \times \mathcal{V}$  such that for all  $(\{\boldsymbol{w}, q\}, \boldsymbol{\vartheta}) \in \mathcal{W} \times \mathcal{V}$ ,

$$(\boldsymbol{\lambda}_\Omega, \boldsymbol{w})_{\Gamma_g} = B_{\Omega, \Gamma_h}(\{\boldsymbol{u}, p\}, \{\boldsymbol{w}, q\}) - L_{\Omega, \Gamma_h}(\{\boldsymbol{w}, q\}) + (\boldsymbol{\vartheta}, \boldsymbol{u} - \boldsymbol{g})_{\Gamma_g}. \quad (3.2.4)$$

Here  $\mathcal{W}$  is the unrestricted space used for the solution and test functions and  $\mathcal{V}$  is a suitable Lagrange multiplier space. Section 3.2.3 employs formulation (3.2.4) to derive the corresponding global energy statement. The equivalence of this form with the strong form (3.2.1) follows from Green's formula and an appropriate choice of the weighting functions. The expression of the Lagrange multiplier is a by-product of this execution and yields

$$\boldsymbol{\lambda}_\Omega = -\frac{1}{2}u_n \boldsymbol{u} - p \boldsymbol{n} + \nu \partial_n \boldsymbol{u}. \quad (3.2.5)$$

The multiplier can be interpreted as an auxiliary flux with a convective, a pressure and a viscous contribution. Consult [102] for details about auxiliary fluxes in weak formulations.

**Remark 3.2.3.** Note that we get the same expression when employing the skew-symmetric form (3.2.3).

### 3.2.3 Global energy evolution

The evolution of the global energy follows when substituting  $(\{\boldsymbol{w}, q\}, \boldsymbol{\vartheta}) = (\{\boldsymbol{u}, p\}, \boldsymbol{\lambda}_\Omega)$  in (3.2.4). Employing Green's formula and the strong incompressibility constraint (3.2.1b) we see that the convective term only contributes to the energy evolution via a boundary term. The global energy, which is defined as  $\mathcal{E}_\Omega := \frac{1}{2}(\boldsymbol{u}, \boldsymbol{u})_\Omega$ , evolves as

$$\frac{d}{dt} \mathcal{E}_\Omega = -\|v^{1/2} \nabla \boldsymbol{u}\|_\Omega^2 + (\boldsymbol{u}, \boldsymbol{f})_\Omega - (1, \mathcal{F}_\Omega)_\Gamma, \quad (3.2.6)$$

where  $\frac{d}{dt}$  is the time derivative and  $\|\cdot\|_D^2$  defines the standard  $L^2$ -norm over  $D$ . The flux reads:

$$\mathcal{F}_\Omega = \begin{cases} -\boldsymbol{g} \cdot \boldsymbol{\lambda}_\Omega & \text{on } \Gamma_g, \\ |u_n|e - \boldsymbol{u} \cdot \boldsymbol{h} & \text{on } \Gamma_h, \end{cases} \quad (3.2.7)$$

with  $e := \frac{1}{2} \boldsymbol{u} \cdot \boldsymbol{u}$  the pointwise energy. The terms of (3.2.6) represent from left to right: (i) the energy loss due to viscous molecular dissipation, (ii) the power exerted

by the body force and (iii) the energy change due to the boundary conditions. Substitution of the Lagrange multiplier and the boundary conditions leads to the expected expression of the flux

$$\mathcal{F}_\Omega = u_n(e + p) - \nu \partial_n e \quad \text{on } \Gamma. \quad (3.2.8)$$

These terms represent the convective and viscous flux as well as the rate of work due to the pressure. We emphasize that the continuous convective–diffusive equation displays very similar energy behavior (obviously the pressure term is absent there) [57]. This provides an additional indication of the similarity in the discrete setting.

**Remark 3.2.4.** *The transition from expression (3.2.7) to (3.2.8) is only possible in the continuous setting. In the discrete setting no closed-form expression for the Lagrange multiplier exists. This also applies to the localized version in Section 3.2.4.*

### 3.2.4 Local energy evolution

The procedure to find the local energy evolution is very similar to that of the global energy. Let  $\omega \subset \Omega$  be an arbitrary subdomain with boundary  $\partial\omega$ , let  $\Omega - \omega$  denote its complement and let their shared boundary denote  $\chi_\omega = \partial\omega \cap \partial(\Omega - \omega)$ . Figure 3.2 shows the subdomains and their boundaries.

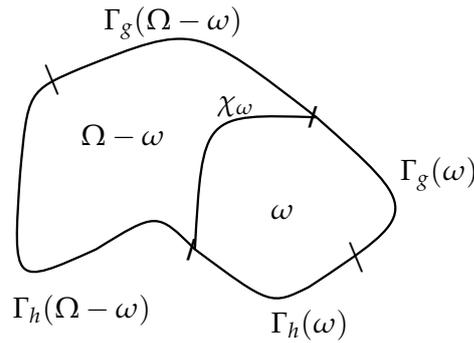


Figure 3.2: Spatial domain  $\Omega$  with a subdomain  $\omega \subset \Omega$ . The shared boundary of  $\omega$  and its complement is  $\chi_\omega$ . The boundaries  $\Gamma_g$  and  $\Gamma_h$  split according to  $\omega$ . This is the same figure as in [57].

The continuity across the interface is enforced with a Lagrange multiplier in the appropriate space  $\mathcal{V}_\omega$ . The discontinuous test function space writes  $\mathcal{W}_\omega$ . The weak statement enforced on  $\omega$  is again a mixed formulation and reads:

$$\begin{aligned} & \text{Find } (\{\mathbf{u}, p\}, \lambda_\omega) \in \mathcal{W} \times \mathcal{V} \text{ such that for all } (\{\mathbf{w}, q\}, \boldsymbol{\vartheta}) \in \mathcal{W} \times \mathcal{V}, \\ & (\mathbf{w}, \lambda_\omega)_{\chi_\omega} + (\mathbf{w}, \lambda_\omega)_{\Gamma_g(\omega)} = B_{\omega, \Gamma_h}(\{\mathbf{u}, p\}, \{\mathbf{w}, q\}) - L_{\omega, \Gamma_h}(\{\mathbf{w}, q\}), \\ & (\boldsymbol{\vartheta}, \llbracket \mathbf{u} \rrbracket)_{\chi_\omega} + (\boldsymbol{\vartheta}, \mathbf{u} - \mathbf{g})_{\Gamma_g(\omega)} = 0. \end{aligned} \quad (3.2.9a)$$

We have here employed the jump term  $\llbracket \mathbf{u} \rrbracket$  given by

$$\llbracket \mathbf{u} \rrbracket := \mathbf{u}|_\omega - \mathbf{u}|_{\Omega - \omega}, \quad (3.2.10)$$

where the terms are defined on  $\omega$  and  $\Omega - \omega$ , respectively. Furthermore,  $\mathbf{n}_\omega$  is the outward normal of domain  $\omega$ ,  $u_{n_\omega}$  is the outward velocity in direction  $\mathbf{n}_\omega$  and  $\partial_{n_\omega}$  the direction derivative outward of  $\omega$ . The equivalence of this form with the strong form (3.2.1) leads to the expression of the Lagrange multiplier:

$$\lambda_\omega = -u_{n_\omega} \mathbf{u} - p \mathbf{n}_\omega + \nu \partial_{n_\omega} \mathbf{u}, \quad (3.2.11)$$

which is clearly the localized version of (3.2.5). A direct consequence is the symmetry of the Lagrange multipliers (these are also called auxiliary fluxes in this setting, see [102]):

$$\lambda_\omega + \lambda_{\Omega-\omega} = \mathbf{0}, \quad (3.2.12)$$

i.e. that what flows out  $\omega$  through  $\chi_\omega$  enters its complement. The energy evolution linked to each of the domains is a natural split of the global energy evolution:

$$\frac{d}{dt} \mathcal{E}_\omega = - \|v^{1/2} \nabla \mathbf{u}\|_\omega^2 + (\mathbf{u}, \mathbf{f})_\omega - (1, \mathcal{F}_\omega)_{\partial\omega}, \quad (3.2.13)$$

with energy fluxes

$$\mathcal{F}_\omega = \begin{cases} -\mathbf{g} \cdot \lambda_\omega & \text{on } \Gamma_g(\omega), \\ |u_{n_\omega}| e - \mathbf{u} \cdot \mathbf{h} & \text{on } \Gamma_h(\omega), \\ -\mathbf{u} \cdot \lambda_\omega & \text{on } \chi_\omega. \end{cases} \quad (3.2.14)$$

The last term of (3.2.14) redistributes energy over the domain. It represents an energy flux across the subdomain interface  $\chi_\omega$  with a convective, a pressure and a viscous component. Similarly as before, substitution of the terms in the energy flux leads to

$$\mathcal{F}_\omega = u_{n_\omega} (e + p) - \nu \partial_{n_\omega} e \quad \text{on } \partial\omega. \quad (3.2.15)$$

This is obviously the localized version of (3.2.8).

**Remark 3.2.5.** *All statements of this section are in the continuous setting. Hence, the standard discretization, i.e. the Galerkin method, displays the same correct energy behavior.*

**Remark 3.2.6.** *The various boundary terms may distract the reader and do not contribute to the goal of this work. Therefore we only consider boundary conditions precluding the energy flux  $\mathcal{F}$  on  $\Gamma$ . The homogeneous Dirichlet and periodic boundary conditions satisfy this purpose. Applying non-homogeneous boundaries is straightforward.*

We continue this work by discretizing the system according to the dynamic variational multiscale method with the target to closely resemble energy evolution (3.2.6) and (3.2.13).

### 3.3 ENERGY EVOLUTION OF THE VARIATIONAL MULTISCALE METHOD WITH DYNAMIC SMALL-SCALES

The convective–diffusive context [57] learns us that the dynamical structure of the small-scales is a requirement for the stabilized formulation to display the correct energy behavior. This allows to skip the static small-scales and to directly apply the dynamic modeling approach. We follow this road.

### 3.3.1 The multiscale split

The variational multiscale split is nowadays a standard execution [100, 103] which we include here for the sake of completeness and notation. Employing the variational multiscale methodology the trial and weighting function spaces split into large- and small-scales as:

$$\mathcal{W} = \mathcal{W}^h \oplus \mathcal{W}', \quad (3.3.1)$$

with  $\mathcal{W}^h$  and  $\mathcal{W}'$  containing the large-scales and small-scales, respectively. The large-scale space is spanned by the finite-dimensional numerical discretization while the fine-scales are its infinite-dimensional complement. The fine-scale space  $\mathcal{W}'$  is also referred to as subgrid-scales since these scales are not reproduced by the grid. This decomposition implies the split of the solution and weighting functions as follows:

$$\mathbf{u} = \mathbf{u}^h + \mathbf{u}', \quad (3.3.2a)$$

$$\mathbf{w} = \mathbf{w}^h + \mathbf{w}', \quad (3.3.2b)$$

where  $\mathbf{u}^h, \mathbf{w}^h \in \mathcal{W}^h$  and  $\mathbf{u}', \mathbf{w}' \in \mathcal{W}'$  with  $\mathbf{u} := \{\mathbf{u}, p\}$ ,  $\mathbf{w} := \{\mathbf{w}, q\}$ . Uniqueness follows when a projector  $\mathcal{P}^h : \mathcal{W} \rightarrow \mathcal{W}^h$  is used for the splitting operation:

$$\mathbf{u}^h = \mathcal{P}^h \mathbf{u}, \quad (3.3.3a)$$

$$\mathbf{u}' = (\mathcal{I} - \mathcal{P}^h) \mathbf{u}, \quad (3.3.3b)$$

where  $\mathcal{I} : \mathcal{W} \rightarrow \mathcal{W}$  is the identity operator. Employing both  $\mathbf{w} = \mathbf{w}^h$  and  $\mathbf{w} = \mathbf{w}'$ , and the solution split (3.3.2a) in (3.2.2) leads to the weak formulation:

Find  $\mathbf{u}^h \in \mathcal{W}^h$ ,  $\mathbf{u}' \in \mathcal{W}'$  for all  $\mathbf{w}^h \in \mathcal{W}^h$ ,  $\mathbf{w}' \in \mathcal{W}'$ ,

$$B_\Omega(\mathbf{u}^h + \mathbf{u}', \mathbf{w}^h) = L_\Omega(\mathbf{w}^h)_\Omega, \quad \text{for all } \mathbf{w}^h \in \mathcal{W}^h, \quad (3.3.4a)$$

$$B_\Omega(\mathbf{u}^h + \mathbf{u}', \mathbf{w}') = L_\Omega(\mathbf{w}')_\Omega, \quad \text{for all } \mathbf{w}' \in \mathcal{W}'. \quad (3.3.4b)$$

Note that this is an infinite-dimensional system with unknowns  $\mathbf{u}^h$  and  $\mathbf{u}'$ . Appropriately parameterizing the small-scales  $\mathbf{u}'$  in terms of  $\mathbf{u}^h$  converts (3.3.4a) into a solvable finite element problem. This conversion can be done with inspiration from (3.3.4b). For the technical details of the parameterization consult [113].

### 3.3.2 Dynamic small-scales

Assume now that the domain  $\Omega$  is partitioned into a set of elements  $\Omega_e$ . The domain of element interiors does not include the interior boundaries and denotes

$$\tilde{\Omega} = \bigcup_e \Omega_e. \quad (3.3.5)$$

We employ the dynamic small-scales, see [46], demanded by the convective-diffusive context for correct energy behavior [57]. The fine-scale model

$$\partial_t \{\hat{\mathbf{u}}', 0\} + \tau^{-1} \{\hat{\mathbf{u}}', \hat{p}'\} + \mathcal{R}(\{\mathbf{u}^h, p^h\}, \hat{\mathbf{u}}') = 0, \quad (3.3.6)$$

is an ordinary differential equation. The hat-sign is used to indicate a small-scale model instead of the actual small-scales. The intrinsic time scale  $\tau$  is a matrix of stabilization parameters, here  $\tau \in \mathbb{R}^{4 \times 4}$ , with contributions for the two equations:

$$\tau = \begin{pmatrix} \tau_M \mathbf{I}_{3 \times 3} & \mathbf{0}_3 \\ \mathbf{0}_3^T & \tau_C \end{pmatrix}. \quad (3.3.7)$$

The local large-scale residual contains a momentum part  $r_M$  and continuity part  $r_C$  linked to the incompressibility constraint, respectively, given by

$$\mathcal{R} \left( \left\{ \mathbf{u}^h, p^h \right\}, \hat{\mathbf{u}}' \right) = \left\{ r_M \left( \left\{ \mathbf{u}^h, p^h \right\}, \hat{\mathbf{u}}' \right), r_C \left( \mathbf{u}^h \right) \right\}^T, \quad (3.3.8a)$$

$$r_M = \partial_t \mathbf{u}^h + \left( \left( \mathbf{u}^h + \hat{\mathbf{u}}' \right) \cdot \nabla \right) \mathbf{u}^h + \nabla p^h - \nu \Delta \mathbf{u}^h - \mathbf{f}, \quad (3.3.8b)$$

$$r_C = \nabla \cdot \mathbf{u}^h. \quad (3.3.8c)$$

In the following we ignore the hat-sign. We employ a dynamic version of the stabilization parameters  $\tau_M, \tau_C$  defined in [17]. The details are provided in Appendix 3.B. The subscripts  $M$  and  $C$  refer to *momentum* and *continuity*, respectively. Mirroring [57], the momentum residual (3.3.8b) uses the full velocity  $\mathbf{u}^h + \mathbf{u}'$ . This creates a nonlinearity in the system. Therefore we apply a standard iterative procedure to determine the small-scales.

The resulting residual-based dynamic VMS weak formulation is

Find  $\mathbf{u}^h \in \mathcal{W}^h$  for all  $\mathbf{w}^h \in \mathcal{W}^h$

$$B_{\Omega}^{\text{VMSD}} \left( \mathbf{u}^h, \mathbf{w}^h \right) = L_{\Omega}(\mathbf{w}^h), \quad (3.3.9a)$$

where

$$\begin{aligned} B_{\Omega}^{\text{VMSD}} \left( \mathbf{u}^h, \mathbf{w}^h \right) &= B_{\Omega} \left( \mathbf{u}^h, \mathbf{w}^h \right) + \left( \mathbf{w}^h, \partial_t \mathbf{u}' \right)_{\hat{\Omega}} - \left( \nu \Delta \mathbf{w}^h, \mathbf{u}' \right)_{\hat{\Omega}} \\ &\quad - \left( \nabla q^h, \mathbf{u}' \right)_{\hat{\Omega}} - \left( \nabla \cdot \mathbf{w}^h, p' \right)_{\hat{\Omega}} \\ &\quad - \left( \nabla \mathbf{w}^h, \mathbf{u}^h \otimes \mathbf{u}' \right)_{\hat{\Omega}} - \left( \nabla \mathbf{w}^h, \mathbf{u}' \otimes \mathbf{u}^h \right)_{\hat{\Omega}} \\ &\quad - \left( \nabla \mathbf{w}^h, \mathbf{u}' \otimes \mathbf{u}' \right)_{\hat{\Omega}}, \end{aligned} \quad (3.3.9b)$$

$$\partial_t \{ \mathbf{u}', 0 \} + \tau^{-1} \{ \mathbf{u}', p' \} + \mathcal{R} \left( \left\{ \mathbf{u}^h, p^h \right\}, \mathbf{u}' \right) = 0, \quad (3.3.9c)$$

and where the additional D stands for *dynamic*. When examining the last line of (3.3.9b), we recognize the following contributions. The first term is the SUPG contribution. The first two terms model the *cross stress*, while the last term models the *Reynolds stress*. Note that no spatial derivatives act on the small-scales. Furthermore, in contrast to static small-scales, the dynamic small-scale model (3.3.9c) is a separate equation and cannot directly be substituted into the large-scale equation (3.3.9b).

### 3.3.3 Local energy evolution of the VMSD form

To arrive at the local energy evolution of (3.3.9), we extend the weak formulation to a Lagrange multiplier setting to allow discontinuous functions across subdomains,

similar as (3.2.9). The weak statement, here stated for domain  $\omega \subset \Omega$ , reads

Find  $(\mathbf{u}^h, \boldsymbol{\lambda}_\omega^h) \in \mathcal{W} \times \mathcal{V}$  such that for all  $(\mathbf{w}^h, \boldsymbol{\vartheta}^h) \in \mathcal{W} \times \mathcal{V}$ ,

$$\begin{aligned} (\mathbf{w}^h, \boldsymbol{\lambda}_\omega^h)_{\chi_\omega} &= B_\omega^{\text{VMSD}}(\mathbf{u}^h, \mathbf{w}^h) \\ &\quad - L_\omega(\mathbf{w}^h), \end{aligned} \quad (3.3.10a)$$

$$(\boldsymbol{\vartheta}^h, \llbracket \mathbf{u}^h \rrbracket)_{\chi_\omega} = 0, \quad (3.3.10b)$$

$$\partial_t \{\mathbf{u}', 0\} + \boldsymbol{\tau}^{-1} \{\mathbf{u}', p'\} + \mathcal{R}(\{\mathbf{u}^h, p^h\}, \mathbf{u}') = 0. \quad (3.3.10c)$$

To obtain the evolution of the *local total energy*  $\mathcal{E}_\omega = \frac{1}{2}(\mathbf{u}^h + \mathbf{u}', \mathbf{u}^h + \mathbf{u}')_{\tilde{\omega}}$  linked to the variational formulation (3.3.9), we employ  $\mathbf{w}^h = \mathbf{u}^h, q^h = p^h$  and  $\boldsymbol{\vartheta}^h = \boldsymbol{\lambda}_\omega^h$  in (3.3.10). Adding  $\mathbf{u}'$  times the momentum component of (3.3.10c) integrated over  $\tilde{\omega}$  eventually leads to

$$\begin{aligned} \frac{d}{dt} \mathcal{E}_\omega &= - \|\nu^{1/2} \nabla \mathbf{u}^h\|_\omega^2 + (\mathbf{u}^h, \mathbf{f})_\omega - (1, \mathcal{F}_\omega^h)_{\chi_\omega} \\ &\quad - \|\tau_M^{-1/2} \mathbf{u}'\|_{\tilde{\omega}}^2 + (\mathbf{u}', \mathbf{f})_{\tilde{\omega}} + 2(\nu \Delta \mathbf{u}^h, \mathbf{u}')_{\tilde{\omega}} + (\nabla \cdot \mathbf{u}^h, p')_{\tilde{\omega}} \\ &\quad + \left( \nabla \mathbf{u}^h, (\mathbf{u}^h + \mathbf{u}') \otimes (\mathbf{u}^h + \mathbf{u}') \right)_{\tilde{\omega}} - \left( \mathbf{u}', (\mathbf{u}^h + \mathbf{u}') \cdot \nabla \mathbf{u}^h \right)_{\tilde{\omega}}, \end{aligned} \quad (3.3.11)$$

where

$$\mathcal{F}_\omega^h = -\boldsymbol{\lambda}_\omega^h \cdot \mathbf{u}^h. \quad (3.3.12)$$

The first line closely resembles the continuous energy evolution relation. Each one of the other terms appears as a result of the VMS stabilization. The first term of the second line represents the numerical dissipation due to the missing small-scales. This contributes to a decay of the energy, which is favorable from a stability argument. The second term is the power exerted by the body force on the small-scales, this term closely resembles its large-scale counterpart. The remaining terms have no continuous counterpart. With the current small-scale model, the small-scale pressure term dissipates energy<sup>1</sup>. The signs of the other terms are undetermined and therefore these can create energy artificially. The term  $2(\nu \Delta \mathbf{u}^h, \mathbf{u}')_{\tilde{\omega}}$  can be bounded by both the physical dissipation  $\|\nu^{1/2} \nabla \mathbf{u}^h\|_\omega^2$  and numerical dissipation  $\|\tau_M^{-1/2} \mathbf{u}'\|_{\tilde{\omega}}^2$  using a standard argument. However, this results in an overall dissipation that can be smaller than the physical one. This is deemed undesirable. Note that it is comparable with that of the dynamic VMS stabilized form in the convective–diffusive context. The contrast occurs in the last line which is linked to the incompressibility constraint (3.2.1b) and the small-scale pressure. Inspired by the convective–diffusive context, the next section rectifies the method to closely resemble the energy behavior of the continuous setting.

<sup>1</sup> The small-scale pressure expression can be substituted into this term to arrive at  $(\nabla \cdot \mathbf{u}^h, p')_{\tilde{\omega}} = -\|\tau_C^{-1/2} p'\|_{\tilde{\omega}}^2$ . Note that it vanishes when employing a divergence-conforming discrete velocity space.

**Remark 3.3.1.** Employing  $\omega = \Omega$ , and hence  $\tilde{\omega} = \tilde{\Omega}$ , provides the global energy evolution of (3.3.9):

$$\begin{aligned} \frac{d}{dt} \mathcal{E}_\Omega &= - \|v^{1/2} \nabla \mathbf{u}^h\|_\Omega^2 + (\mathbf{u}^h, \mathbf{f})_\Omega \\ &\quad - \|\tau_M^{-1/2} \mathbf{u}'\|_{\tilde{\Omega}}^2 + (\mathbf{u}', \mathbf{f})_{\tilde{\Omega}} + 2(v \Delta \mathbf{u}^h, \mathbf{u}')_{\tilde{\Omega}} + (\nabla \cdot \mathbf{u}^h, p')_{\tilde{\Omega}} \\ &\quad + \left( \nabla \mathbf{u}^h, (\mathbf{u}^h + \mathbf{u}') \otimes (\mathbf{u}^h + \mathbf{u}') \right)_{\tilde{\Omega}} - \left( \mathbf{u}', (\mathbf{u}^h + \mathbf{u}') \cdot \nabla \mathbf{u}^h \right)_{\tilde{\Omega}}. \end{aligned} \quad (3.3.13)$$

### 3.4 TOWARD A STABILIZED FORMULATION WITH CORRECT ENERGY BEHAVIOR

This section presents the procedure to remedy the incorrect energy behavior (3.3.11) of the dynamic VMS formulation (3.3.9). The first ingredient is the switch from the conservative form to a skew-symmetric form with the help of the divergence-free velocity field constraint. Next, we employ the natural choice of a Stokes-projector and demand divergence-free small-scales. In view of the convective–diffusive context, we use  $H_0^1$  small-scales to treat the small-scale viscous term.

#### 3.4.1 Design condition

We present a design condition which clarifies the desirable energy behavior of the formulation. The variational weak formulation corresponding to (3.2.1) is demanded to satisfy the local energy behavior:

$$\begin{aligned} \frac{d}{dt} \mathcal{E}_\omega &= - \|v^{1/2} \nabla \mathbf{u}^h\|_\omega^2 + (\mathbf{u}^h, \mathbf{f})_\omega - (1, \mathcal{F}_\omega^h)_{\chi_\omega} \\ &\quad - \|\tau_M^{-1/2} \mathbf{u}'\|_{\tilde{\omega}}^2 + (\mathbf{u}', \mathbf{f})_{\tilde{\omega}}, \end{aligned} \quad (3.4.1)$$

with *exact* divergence-free velocity fields. Note that this requirement is very similar to that of the convective–diffusive context [57] where the convective velocity is assumed solenoidal.

**Remark 3.4.1.** In the following we use the ingredients mentioned above to convert the VMS formulation (3.3.10) into a method that satisfies the design condition. It is important to realize that the small-scales employed in the formulation are determined by a model equation. This implies that these properties are not necessarily valid for the model small-scales. In contrast, the exact small-scales do satisfy these properties. The model small-scales approximate its exact counterpart which justifies the judicious use of these properties to construct a method that satisfies the design condition.

#### 3.4.2 Skew-symmetric form

We employ a multiscale form of the skew-symmetric formulation (see (3.2.3)) to eliminate the convective contributions in (3.3.11). Considering the convective terms in isolation, we cast them into the following form:

$$\begin{aligned}
& - (\nabla \mathbf{w}^h, (\mathbf{u}^h + \mathbf{u}') \otimes (\mathbf{u}^h + \mathbf{u}'))_{\bar{\Omega}} \\
& = - \left( (\mathbf{u}^h + \mathbf{u}') \cdot \nabla \mathbf{w}^h, \mathbf{u}^h \right)_{\bar{\Omega}} - \left( (\mathbf{u}^h + \mathbf{u}') \cdot \nabla \mathbf{w}^h, \mathbf{u}' \right)_{\bar{\Omega}} \\
& = \frac{1}{2} \left( \mathbf{w}^h, (\mathbf{u}^h + \mathbf{u}') \cdot \nabla \mathbf{u}^h \right)_{\bar{\Omega}} - \frac{1}{2} \left( (\mathbf{u}^h + \mathbf{u}') \cdot \nabla \mathbf{w}^h, \mathbf{u}^h \right)_{\bar{\Omega}} \\
& \quad + \frac{1}{2} \left( \mathbf{u}^h, \mathbf{w}^h \nabla \cdot (\mathbf{u}^h + \mathbf{u}') \right)_{\bar{\Omega}} - \left( (\mathbf{u}^h + \mathbf{u}') \cdot \nabla \mathbf{w}^h, \mathbf{u}' \right)_{\bar{\Omega}} \\
& = \frac{1}{2} \left( \mathbf{w}^h, (\mathbf{u}^h + \mathbf{u}') \cdot \nabla \mathbf{u}^h \right)_{\bar{\Omega}} - \frac{1}{2} \left( (\mathbf{u}^h + \mathbf{u}') \cdot \nabla \mathbf{w}^h, \mathbf{u}^h \right)_{\bar{\Omega}} \\
& \quad - \left( (\mathbf{u}^h + \mathbf{u}') \cdot \nabla \mathbf{w}^h, \mathbf{u}' \right)_{\bar{\Omega}}, \tag{3.4.2}
\end{aligned}$$

where we have employed the multiscale incompressibility constraint  $\nabla \cdot \mathbf{u} = \nabla \cdot (\mathbf{u}^h + \mathbf{u}') = 0$  in the last equality. The last expression is incorporated into the formulation. The resulting residual-based skew-symmetric VMS weak formulation is

Find  $\mathbf{u}^h \in \mathcal{W}^h$  such that for all  $\mathbf{w}^h \in \mathcal{W}^h$ ,

$$C_{\Omega}^{\text{VMSD}}(\mathbf{u}^h, \mathbf{w}^h) = L_{\Omega}(\mathbf{w}^h), \tag{3.4.3a}$$

where

$$\begin{aligned}
C_{\Omega}^{\text{VMSD}}(\mathbf{u}^h, \mathbf{w}^h) & = C_{\Omega}(\mathbf{u}^h, \mathbf{w}^h) + (\mathbf{w}^h, \partial_t \mathbf{u}')_{\bar{\Omega}} - (\nu \Delta \mathbf{w}^h, \mathbf{u}')_{\bar{\Omega}} \\
& \quad - (\nabla q^h, \mathbf{u}')_{\bar{\Omega}} - (\nabla \cdot \mathbf{w}^h, p')_{\bar{\Omega}} \\
& \quad + \frac{1}{2} (\mathbf{w}^h, \mathbf{u}' \cdot \nabla \mathbf{u}^h)_{\bar{\Omega}} - \frac{1}{2} (\mathbf{u}' \cdot \nabla \mathbf{w}^h, \mathbf{u}^h)_{\bar{\Omega}} \\
& \quad - \left( (\mathbf{u}^h + \mathbf{u}') \cdot \nabla \mathbf{w}^h, \mathbf{u}' \right)_{\bar{\Omega}}, \tag{3.4.3b}
\end{aligned}$$

$$\partial_t \{\mathbf{u}', 0\} + \tau^{-1} \{\mathbf{u}', p'\} + \mathcal{R}(\{\mathbf{u}^h, p^h\}, \mathbf{u}') = 0. \tag{3.4.3c}$$

This eliminates the convective contributions from the local energy evolution equation:

$$\begin{aligned}
\frac{d}{dt} \mathcal{E}_{\omega} & = - \|\nu^{1/2} \nabla \mathbf{u}^h\|_{\omega}^2 + (\mathbf{u}^h, \mathbf{f})_{\omega} - (1, \mathcal{F}_{\omega}^h)_{\chi_{\omega}} \\
& \quad - \|\tau_M^{-1/2} \mathbf{u}'\|_{\bar{\omega}}^2 + (\mathbf{u}', \mathbf{f})_{\bar{\omega}} + 2(\nu \Delta \mathbf{u}^h, \mathbf{u}')_{\bar{\omega}} + (\nabla \cdot \mathbf{u}^h, p')_{\bar{\omega}}. \tag{3.4.4}
\end{aligned}$$

### 3.4.3 Stokes projector

In the convective–diffusive context a  $H_0^1$ -orthogonality of the small-scale viscous term is required for correct energy behavior. This is the distinguished limit of  $Pe \rightarrow 0$  of the steady convection–diffusion equations, where  $Pe$  is the Péclet number. Its Navier–Stokes counterpart is to apply a Stokes-projector which is based on the distinguished limit  $Re \rightarrow 0$  of the steady incompressible Navier–Stokes equations. Here  $Re$  is the Reynolds number. Thus, applying a Stokes projection on the large-scale equation seems a natural choice. Moreover, it reduces the variational form in the limit  $Re \rightarrow 0$  to the standard Galerkin method. This is a valid and established

method in that regime, provided compatible discretizations for the velocity and pressure spaces are used.

For the scale separation (3.3.3) we select the Stokes projector given by

$\mathcal{P}_{\text{Stokes}}^h : \mathbf{u} \in \mathcal{W} \rightarrow \mathbf{u}^h \in \mathcal{W}^h$ : Find  $\mathbf{u}^h \in \mathcal{W}^h$  such that for all  $\mathbf{w}^h \in \mathcal{W}^h$ ,

$$\left( \nu \Delta \mathbf{w}^h, \mathbf{u}^h \right)_{\Omega} + \left( \nabla \cdot \mathbf{w}^h, p^h \right)_{\Omega} = \left( \nu \Delta \mathbf{w}^h, \mathbf{u} \right)_{\Omega} + \left( \nabla \cdot \mathbf{w}^h, p \right)_{\Omega}, \quad (3.4.5a)$$

$$\left( \nabla q^h, \mathbf{u}^h \right)_{\Omega} = \left( \nabla q^h, \mathbf{u} \right)_{\Omega}, \quad (3.4.5b)$$

in the bilinear form (3.4.3b). Note that this projector only makes sense if the elements of  $\mathcal{W}^h$  are inf-sup stable and the velocities are at least  $C^1$ -continuous. The numerical results presented in Section 3.6 fulfill this requirement: quadratic NURBS basis functions are employed. However, note that the final form, given in Appendix 3.A, does not have the smoothness restriction.

As a consequence we assume the modeled small-scales to satisfy the orthogonality induced by the Stokes operator:

$$\left( \nu \Delta \mathbf{w}^h, \mathbf{u}' \right)_{\tilde{\Omega}} + \left( \nabla \cdot \mathbf{w}^h, p' \right)_{\tilde{\Omega}} = 0, \quad (3.4.6a)$$

$$\left( \nabla q^h, \mathbf{u}' \right)_{\tilde{\Omega}} = 0, \quad (3.4.6b)$$

for all  $\mathbf{w}^h \in \mathcal{W}^h$ . This converts (3.4.3) into the simplified formulation:

Find  $\mathbf{u}^h \in \mathcal{W}^h$  such that for all  $\mathbf{w}^h \in \mathcal{W}^h$

$$S_{\Omega} \left( \mathbf{u}^h, \mathbf{w}^h \right) = L_{\Omega}(\mathbf{w}^h), \quad (3.4.7a)$$

where

$$\begin{aligned} S_{\Omega} \left( \mathbf{u}^h, \mathbf{w}^h \right) &= C_{\Omega} \left( \mathbf{u}^h, \mathbf{w}^h \right) + \left( \mathbf{w}^h, \partial_t \mathbf{u}' \right)_{\tilde{\Omega}} \\ &\quad + \frac{1}{2} \left( \mathbf{w}^h, \mathbf{u}' \cdot \nabla \mathbf{u}^h \right)_{\tilde{\Omega}} - \frac{1}{2} \left( \mathbf{u}' \cdot \nabla \mathbf{w}^h, \mathbf{u}^h \right)_{\tilde{\Omega}} \\ &\quad - \left( \left( \mathbf{u}^h + \mathbf{u}' \right) \cdot \nabla \mathbf{w}^h, \mathbf{u}' \right)_{\tilde{\Omega}}, \end{aligned} \quad (3.4.7b)$$

$$\partial_t \mathbf{u}' + \tau_M^{-1} \mathbf{u}' + \mathbf{r}_M = 0, \quad (3.4.7c)$$

where the  $S$  abbreviates *Stokes*. Note that the small-scale pressure terms have vanished from the formulation. The energy linked to this formulation is

$$\begin{aligned} \frac{d}{dt} \mathcal{E}_{\omega} &= - \|\nu^{1/2} \nabla \mathbf{u}^h\|_{\omega}^2 + (\mathbf{u}^h, \mathbf{f})_{\omega} - (1, \mathcal{F}_{\omega}^h)_{\chi_{\omega}} \\ &\quad - \|\tau_M^{-1/2} \mathbf{u}'\|_{\tilde{\omega}}^2 + (\mathbf{u}', \mathbf{f})_{\tilde{\omega}} + (\nu \Delta \mathbf{u}^h, \mathbf{u}')_{\tilde{\omega}} - (\nabla p^h, \mathbf{u}')_{\tilde{\omega}}. \end{aligned} \quad (3.4.8)$$

To fulfill the design condition (3.4.1), the last two terms of (3.4.8) need to be eliminated, i.e.

$$(\nu \Delta \mathbf{u}^h, \mathbf{u}')_{\tilde{\Omega}} - (\nabla p^h, \mathbf{u}')_{\tilde{\Omega}} = 0. \quad (3.4.9)$$

There are various options available to accomplish this. Before sketching some of these options we first like to note the following. Augmenting the undesirable terms of (3.4.8) with  $(\nabla \cdot \mathbf{u}^h, p')$  results in the requirement

$$(\nu \Delta \mathbf{u}^h, \mathbf{u}')_{\tilde{\omega}} - (\nabla p^h, \mathbf{u}')_{\tilde{\omega}} + (\nabla \cdot \mathbf{u}^h, p')_{\tilde{\omega}} = 0. \quad (3.4.10)$$

This is a well-defined orthogonality induced by the Stokes operator, given in (3.4.6). The augmented term would appear if  $\nabla p'$  in the small-scale momentum equation is not neglected<sup>2</sup>. Note that this term is not (easily) computable and therefore usually omitted in the formulation.

The required orthogonality (3.4.9) can be either *assumed* or *enforced* [57]. We discuss four options here.

- First we could assume the orthogonality in the small-scale equation (3.4.7c). This orthogonality has previously been assumed to modify the large-scale equation (3.4.7a). Assuming it in the small-scale equation results in a stable method with the desired energy property. However the small-scale model is not residual-based anymore. This results in an inconsistent method. We do not further consider this option.
- Alternatively, we could assume the orthogonality in the large-scale equation (3.4.7a) again. This converts the formulation into a GLS method. This method includes a PSPG term,  $-(\nabla q^h, \mathbf{u}')_{\tilde{\Omega}}$ , and therefore pointwise divergence-free solutions cannot be guaranteed. The formulation harms the design condition of Section 3.4.1 and is therefore omitted.
- Another option is to enforce the required orthogonality using Lagrange-multipliers. This is not straightforward and is deemed unnecessarily expensive.
- The path we propose is to cure the unwanted terms separately by combining the second and third options. The approach is to (i) enforce divergence-free small-scales to eliminate the second term of (3.4.9) and (ii) assume an  $H_0^1$ -orthogonality to erase the first term of (3.4.9). Sections 3.4.4 and 3.4.5 respectively describe these steps.

#### 3.4.4 Divergence-free small-scales

The last term of (3.4.8) disappears when enforcing divergence-free small-scales. We handle this with a projection operator on the small-scales:

$\mathcal{P}_{\text{div}}^h : \mathbf{u} \in \mathcal{W} \rightarrow \mathbf{u}^h \in \mathcal{W}^h$ : Find  $\mathbf{u}^h \in \mathcal{W}^h$  such that for all  $\mathbf{w}^h \in \mathcal{W}^h$ ,

$$\left( \nabla q^h, \mathbf{u}^h \right)_{\Omega} = \left( \nabla q^h, \mathbf{u} \right)_{\Omega}, \quad (3.4.11)$$

with corresponding orthogonality:

$$\left( \nabla q^h, \mathbf{u}' \right)_{\tilde{\Omega}} = 0, \quad \text{for all } \mathbf{w}^h \in \mathcal{W}^h. \quad (3.4.12)$$

<sup>2</sup> Including the small-scale pressure in the residual augments the right-hand side of (3.4.8) with the term  $(\nabla p', \mathbf{u}')$ . Next, by using the strong form continuity equation weighted with the small-scale pressure, i.e.  $(p', \nabla \cdot (\mathbf{u}^h + \mathbf{u}')) = 0$ , this term converts into  $(\nabla \cdot \mathbf{u}^h, p')$ .

This orthogonality defines the fine-scale space  $\mathcal{W}'$  which represents the orthogonal component of  $\mathcal{W}^h$  in terms of the projection (3.4.12) as

$$\mathcal{W}' = \mathcal{W}'_{\text{div}} := \left\{ \{\mathbf{u}, p\} \in \mathcal{W}; \quad (\nabla \theta^h, \mathbf{u})_{\Omega} = 0, \quad \text{for all } \theta^h \in \mathcal{P}^h \right\}, \quad (3.4.13)$$

where the space  $\mathcal{P}^h$  is the pressure part of  $\mathcal{W}^h = \mathcal{U}^h \times \mathcal{P}^h$ . Directly employing this divergence-free space indeed eliminates the last term of (3.4.8). However the small-scale solution space is infinite dimensional, and therefore not applicable in the numerical method. As before, we avoid dealing with this space by using a Lagrange-multiplier construction yielding a mixed formulation. Opening the solution space leads to the formulation:

Find  $(\mathbf{u}^h, \zeta^h) \in \mathcal{W}^h \times \mathcal{P}^h$  such that for all  $(\mathbf{w}^h, \theta^h) \in \mathcal{W}^h \times \mathcal{P}^h$ ,

$$S_{\Omega}^{\text{div}} \left( (\mathbf{u}^h, \zeta^h), (\mathbf{w}^h, \theta^h) \right) = L_{\Omega}(\mathbf{w}^h)_{\Omega}, \quad (3.4.14a)$$

where

$$S_{\Omega}^{\text{div}} \left( (\mathbf{u}^h, \zeta^h), (\mathbf{w}^h, \theta^h) \right) = S_{\Omega}(\mathbf{u}^h, \mathbf{w}^h) + (\nabla \theta^h, \mathbf{u}')_{\tilde{\Omega}}, \quad (3.4.14b)$$

$$\partial_t \mathbf{u}' + \tau_M^{-1} \mathbf{u}' + \nabla \zeta^h + \mathbf{r}_M = 0. \quad (3.4.14c)$$

Obviously, this form follows the energy evolution

$$\begin{aligned} \frac{d}{dt} \mathcal{E}_{\omega} &= - \|\nu^{1/2} \nabla \mathbf{u}^h\|_{\omega}^2 + (\mathbf{u}^h, \mathbf{f})_{\omega} - (1, \mathcal{F}_{\omega}^h)_{\chi_{\omega}} \\ &\quad - \|\tau_M^{-1/2} \mathbf{u}'\|_{\tilde{\omega}}^2 + (\mathbf{u}', \mathbf{f})_{\tilde{\omega}} + (\nu \Delta \mathbf{u}^h, \mathbf{u}')_{\tilde{\omega}}. \end{aligned} \quad (3.4.15)$$

**Remark 3.4.2.** Note that enforcing divergence-free small-scales has introduced an additional equation in the system. The new method has 5 global variables instead of 4 leading to a commensurate increase in computational time. The added block diagonal term is a diffusion matrix which does not further complicate the saddle point structure of the problem.

### 3.4.5 $H_0^1$ -orthogonal small-scales

In the energy evolution (3.4.15) unwanted artificial energy can only be created by the term  $(\nu \Delta \mathbf{u}^h, \mathbf{u}')_{\tilde{\omega}}$ . Employing the orthogonality induced by the  $H_0^1$ -seminorm,

$$(\nu \Delta \mathbf{w}^h, \mathbf{u}')_{\tilde{\Omega}} = 0 \quad \text{for all } \mathbf{w}^h \in \mathcal{W}^h, \quad (3.4.16)$$

obviously cancels this term. To avoid dealing with a larger system of equations, we do not enforce the orthogonality but we assume it in the large-scale equation (3.4.14a). This leads to a consistent GLS method. The resulting GLSDD-formulation reads:

Find  $(\mathbf{u}^h, \zeta^h) \in \mathcal{W}^h \times \mathcal{P}^h$  such that for all  $(\mathbf{w}^h, \theta^h) \in \mathcal{W}^h \times \mathcal{P}^h$ ,

$$S_{\Omega}^{\text{GLSDD}} \left( (\mathbf{u}^h, \zeta^h), (\mathbf{w}^h, \theta^h) \right) = L_{\Omega}(\mathbf{w}^h), \quad (3.4.17a)$$

where

$$\begin{aligned} S_{\Omega}^{\text{GLSDD}} \left( \left( \mathbf{u}^h, \zeta^h \right), \left( \mathbf{w}^h, \theta^h \right) \right) &= S_{\Omega}^{\text{div}} \left( \left( \mathbf{u}^h, \zeta^h \right), \left( \mathbf{w}^h, \theta^h \right) \right) + \left( \nu \Delta \mathbf{w}^h, \mathbf{u}' \right)_{\tilde{\Omega}}, \\ \partial_t \mathbf{u}' + \tau_M^{-1} \mathbf{u}' + \nabla \zeta^h + \mathbf{r}_M &= 0. \end{aligned} \quad (3.4.17b)$$

In the abbreviation GLSDD we follow the same structure as before where the last two D's stand for *dynamic, divergence-free small-scale velocities*<sup>3</sup>. This method displays the correct-energy behavior:

$$\begin{aligned} \frac{d}{dt} \mathcal{E}_{\omega} &= -\|v^{1/2} \nabla \mathbf{u}^h\|_{\omega}^2 + (\mathbf{u}^h, \mathbf{f})_{\omega} - (1, \mathcal{F}_{\omega}^h)_{\chi_{\omega}} \\ &\quad - \|\tau_M^{-1/2} \mathbf{u}'\|_{\tilde{\omega}}^2 + (\mathbf{u}', \mathbf{f})_{\tilde{\omega}}. \end{aligned} \quad (3.4.18)$$

The full expansion of this novel formulation is included in Appendix 3.A for clarity.

### 3.4.6 Local energy backscatter

The separate energy evolution of the large- and small-scales deduces in a similar fashion as above. The large-scale energy  $\mathcal{E}_{\omega}^h = \frac{1}{2}(\mathbf{u}^h, \mathbf{u}^h)_{\omega}$  and the small-scale energy  $\mathcal{E}'_{\omega} = \frac{1}{2}(\mathbf{u}', \mathbf{u}')_{\tilde{\omega}}$  do not add up to the total energy  $\mathcal{E}_{\omega}$  because of the missing cross terms. This energy is stored in an intermediate (buffer) regime which we denote with  $\mathcal{E}_{\omega}^{h'} = (\mathbf{u}^h, \mathbf{u}')_{\tilde{\omega}}$ . The energy evolution takes the form:

$$\begin{aligned} \frac{d}{dt} \mathcal{E}_{\omega}^h &= -\|v^{1/2} \nabla \mathbf{u}^h\|_{\omega}^2 + (\mathbf{u}^h, \mathbf{f})_{\omega} - (1, \mathcal{F}_{\omega}^h)_{\chi_{\omega}} \\ &\quad + \left( (\mathbf{u}^h + \mathbf{u}') \cdot \nabla \mathbf{u}^h, \mathbf{u}' \right)_{\tilde{\omega}} - (\mathbf{u}^h, \partial_t \mathbf{u}')_{\tilde{\omega}}, \end{aligned} \quad (3.4.19a)$$

$$\frac{d}{dt} \mathcal{E}'_{\omega} = (\mathbf{u}^h, \partial_t \mathbf{u}')_{\tilde{\omega}} + (\mathbf{u}', \partial_t \mathbf{u}^h)_{\tilde{\omega}}, \quad (3.4.19b)$$

$$\begin{aligned} \frac{d}{dt} \mathcal{E}_{\omega}^{h'} &= -\|\tau_M^{-1/2} \mathbf{u}'\|_{\tilde{\omega}}^2 + (\mathbf{u}', \mathbf{f})_{\tilde{\omega}} \\ &\quad - \left( (\mathbf{u}^h + \mathbf{u}') \cdot \nabla \mathbf{u}^h, \mathbf{u}' \right)_{\tilde{\omega}} - (\mathbf{u}', \partial_t \mathbf{u}^h)_{\tilde{\omega}}. \end{aligned} \quad (3.4.19c)$$

The result mirrors to the convective–diffusive context with as convective velocity now the total velocity  $\mathbf{u}^h + \mathbf{u}'$ . There is a direct exchange of convective energy between the large-scale and small-scales. Obviously, the superposition of (3.4.19) yields (3.4.18).

### 3.4.7 Time-discrete energy behavior

The generalized- $\alpha$  method serves as time-integrator. Mirroring the convective–diffusive context [57], and using the same notation, we eventually obtain for  $\alpha_m = \gamma$ :

$$\begin{aligned} \mathcal{E}_{n+1} &= \mathcal{E}_n - \Delta t^2 \left( \alpha_f - \frac{1}{2} \right) \|\dot{\mathbf{u}}_{n+\alpha_m}\|_{\Omega}^2 - \Delta t \|v^{1/2} \nabla \mathbf{u}_{n+\alpha_f}^h\|_{\Omega}^2 - \Delta t \|\tau_{\text{dyn}}^{-1/2} \mathbf{u}'_{n+\alpha_f}\|_{\tilde{\Omega}}^2 \\ &\quad + \Delta t (\mathbf{u}_{n+\alpha_f}^h, \mathbf{f})_{\Omega} + \Delta t (\mathbf{u}'_{n+\alpha_f}, \mathbf{f})_{\tilde{\Omega}}. \end{aligned} \quad (3.4.20)$$

Hence, we have a decay of the discretized energy when, in absence of forcing,  $\alpha_f \geq \frac{1}{2}$ . In the numerical implementation we use  $\alpha_f = \alpha_m = \gamma = \frac{1}{2}$  for the stability and second-order accuracy properties [44].

<sup>3</sup> The name GLS refers to the convection–diffusion part of the problem.

### 3.5 CONSERVATION PROPERTIES

Conservation of physical quantities in the numerical formulation is an often sought-after property. In this section we derive the various conservation properties (continuity, linear momentum, angular momentum) of the proposed formulation (3.4.17). We prove these by selecting the appropriate weighting functions. The conservation properties hold on both a global and a local scale. Therefore we omit the domain subscript in the following.

#### 3.5.1 Continuity

Employing the weighting function  $\mathbf{w}^h = \mathbf{0}$ ,  $\theta^h = 0$  in (3.4.17) yields

$$(q^h, \nabla \cdot \mathbf{u}^h) = 0. \quad (3.5.1)$$

The choice  $q^h = \nabla \cdot \mathbf{u}^h$  proves the *pointwise* satisfaction of incompressibility constraint<sup>4</sup>

$$\|\nabla \cdot \mathbf{u}^h\|^2 = 0 \quad \Rightarrow \quad \nabla \cdot \mathbf{u}^h = 0 \quad \text{for all } \mathbf{x} \in \Omega. \quad (3.5.2)$$

Furthermore, the choice of weighting functions  $\mathbf{w}^h = \mathbf{0}$ ,  $q^h = 0$  leads to divergence-free small-scale velocities in the following sense:

$$(\nabla \theta^h, \mathbf{u}') = 0. \quad (3.5.3)$$

#### 3.5.2 Linear momentum

We substitute the weighting functions  $(\mathbf{w}^h, q^h, \theta^h) = (\mathbf{e}_i, 0, -\frac{1}{2}\mathbf{e}_i \cdot \mathbf{u}^h)$  in (3.4.17), where  $\mathbf{e}_i$  is the  $i$ th Cartesian basis vector. Using  $\nabla \mathbf{e}_i = \mathbf{0}$  and the pointwise divergence-free velocity (3.5.2), all diffusive and pressure terms drop out and we are left with:

$$\begin{aligned} & \left( \mathbf{e}_i, \partial_t \mathbf{u}^h + \partial_t \mathbf{u}' \right) + \frac{1}{2} \left( \mathbf{e}_i, \left( (\mathbf{u}^h + \mathbf{u}') \cdot \nabla \right) \mathbf{u}^h \right) \\ & \quad + \left( \nabla \left( -\frac{1}{2} \mathbf{e}_i \cdot \mathbf{u}^h \right), \mathbf{u}' \right) = (\mathbf{e}_i, f). \end{aligned} \quad (3.5.4)$$

Consider the convective term in isolation and write

$$\begin{aligned} \frac{1}{2} \left( \mathbf{e}_i, \left( (\mathbf{u}^h + \mathbf{u}') \cdot \nabla \right) \mathbf{u}^h \right) &= \frac{1}{2} \left( \mathbf{e}_i, \nabla \cdot \left( (\mathbf{u}^h + \mathbf{u}') \otimes \mathbf{u}^h \right) \right) \\ & \quad - \frac{1}{2} \left( \mathbf{e}_i, \left( \nabla \cdot (\mathbf{u}^h + \mathbf{u}') \right) \mathbf{u}^h \right) \\ &= -\frac{1}{2} \left( \nabla \mathbf{e}_i, \left( \mathbf{u}^h + \mathbf{u}' \right) \otimes \mathbf{u}^h \right) \\ & \quad - \frac{1}{2} \left( \mathbf{e}_i, \left( \nabla \cdot (\mathbf{u}^h + \mathbf{u}') \right) \mathbf{u}^h \right) \\ &= -\frac{1}{2} \left( \mathbf{e}_i \cdot \mathbf{u}^h, \nabla \cdot \mathbf{u}' \right) \\ &= \left( \nabla \left( \frac{1}{2} \mathbf{e}_i \cdot \mathbf{u}^h \right), \mathbf{u}' \right). \end{aligned} \quad (3.5.5)$$

<sup>4</sup> Note that in general this weighting function choice is not allowed. We employ the IGA spaces with stable velocity and pressure pairs that do allow this choice.

Combining (3.5.4) and (3.5.5) leads to the balance

$$(\mathbf{e}_i, \partial_t \mathbf{u}^h + \partial_t \mathbf{u}') = (\mathbf{e}_i, \mathbf{f}). \quad (3.5.6)$$

Linear momentum is thus conserved in terms of the total solution.

### 3.5.3 Angular momentum

Conservation of global angular momentum is a desirable property, certainly in rotating flows. It has been analyzed by Bazilevs et al. [16] and Evans et al. [69]. When using the appropriate weighting function spaces the formulation conserves angular momentum. The numerical results of Section 3.6 are however not computed with these weighting function spaces. The demonstration of conservation of angular momentum follows the same ideas as [16]. We set the weighting functions  $(\mathbf{w}^h, q^h, \theta^h) = (\mathbf{x} \times \mathbf{e}_j, 0, -\frac{1}{2} (\mathbf{x} \times \mathbf{e}_j) \cdot \mathbf{u}^h)$ . By construction the gradient of the weighting function leads to a skew-symmetric tensor [16]. As a result the gradient tensor is orthogonal to any symmetric tensor. Consequently the divergence, which is the trace of the gradient, is zero.

Employing these weighting functions in the weak form we arrive at

$$\begin{aligned} & (\mathbf{x} \times \mathbf{e}_j, \partial_t \mathbf{u}^h + \partial_t \mathbf{u}') + \frac{1}{2} (\mathbf{x} \times \mathbf{e}_j, ((\mathbf{u}^h + \mathbf{u}') \cdot \nabla) \mathbf{u}^h)_{\Omega} \\ & - \frac{1}{2} \left( ((\mathbf{u}^h + \mathbf{u}') \cdot \nabla) (\mathbf{x} \times \mathbf{e}_j), \mathbf{u}^h \right) - \left( ((\mathbf{u}^h + \mathbf{u}') \cdot \nabla) (\mathbf{x} \times \mathbf{e}_j), \mathbf{u}' \right) \\ & - \frac{1}{2} \left( \nabla \left( (\mathbf{x} \times \mathbf{e}_j) \cdot \mathbf{u}^h \right), \mathbf{u}' \right) = ((\mathbf{x} \times \mathbf{e}_j), \mathbf{f}). \end{aligned} \quad (3.5.7)$$

Consider again the convective terms in isolation. Switching back to a conservative form, see (3.4.2), yields an incompressibility term:

$$\begin{aligned} & \frac{1}{2} (\mathbf{x} \times \mathbf{e}_j, ((\mathbf{u}^h + \mathbf{u}') \cdot \nabla) \mathbf{u}^h) - \frac{1}{2} \left( ((\mathbf{u}^h + \mathbf{u}') \cdot \nabla) (\mathbf{x} \times \mathbf{e}_j), \mathbf{u}^h \right) \\ & - \left( ((\mathbf{u}^h + \mathbf{u}') \cdot \nabla) (\mathbf{x} \times \mathbf{e}_j), \mathbf{u}' \right) \\ & = - (\nabla (\mathbf{x} \times \mathbf{e}_j), (\mathbf{u}^h + \mathbf{u}') \otimes (\mathbf{u}^h + \mathbf{u}')) - \frac{1}{2} \left( \mathbf{u}^h, (\mathbf{x} \times \mathbf{e}_j) \nabla \cdot (\mathbf{u}^h + \mathbf{u}') \right) \\ & = - (\nabla (\mathbf{x} \times \mathbf{e}_j), (\mathbf{u}^h + \mathbf{u}') \otimes (\mathbf{u}^h + \mathbf{u}')) + \frac{1}{2} \left( \nabla \left( (\mathbf{x} \times \mathbf{e}_j) \cdot \mathbf{u}^h \right), \mathbf{u}' \right). \end{aligned} \quad (3.5.8)$$

The antisymmetric tensor and the symmetric tensor in the first and second argument, respectively, cause the first term to vanish. The incompressibility term cancels with the choice of  $\theta^h$  and the conservation of angular momentum is what remains:

$$(\mathbf{x} \times \mathbf{e}_j, \partial_t \mathbf{u}^h + \partial_t \mathbf{u}') = (\mathbf{x} \times \mathbf{e}_j, \mathbf{f}). \quad (3.5.9)$$

## 3.6 NUMERICAL TEST CASE

In this section we test the GLSDD method (3.4.17) on a three-dimensional Taylor–Green vortex flow at Reynolds number  $Re = 1600$ . This test case is challenging and it is often employed to examine the performance of numerical algorithms for turbulence computations. It serves our purpose because (i) the energy behavior of a fully turbulent flow can be studied, (ii) reference data is available and (iii) the

domain is periodic. Other boundary conditions than periodic ones are beyond the scope of this work.

The flow is initially of laminar type. As the time evolves, the vortices begin to evolve and roll-up. The vortical structures undergo changes and subsequently their structures breakdown and form distorted vorticity patches. The flow transitions to one with a turbulence character; the vortex stretching causes the creation of small-scales. The Taylor–Green vortex initial conditions are specified as follows:

$$u(\mathbf{x}, 0) = \sin(x) \cos(y) \cos(z), \quad (3.6.1a)$$

$$v(\mathbf{x}, 0) = -\cos(x) \sin(y) \cos(z), \quad (3.6.1b)$$

$$w(\mathbf{x}, 0) = 0, \quad (3.6.1c)$$

$$p(\mathbf{x}, 0) = \frac{1}{16} (\cos(2x) + \cos(2y)) (\cos(2z) + 2). \quad (3.6.1d)$$

The physical domain is the cube  $\Omega = [0, 2\pi]^3$  with periodic boundary conditions. For this test case the viscosity is given by  $\nu = \frac{1}{Re}$ . Here we consider the transition phase for times  $t \leq 10$  s. Figure 3.3 shows the iso-surfaces of the  $z$ -vorticity of the initial condition (laminar flow) and the final configuration (fully turbulent flow).

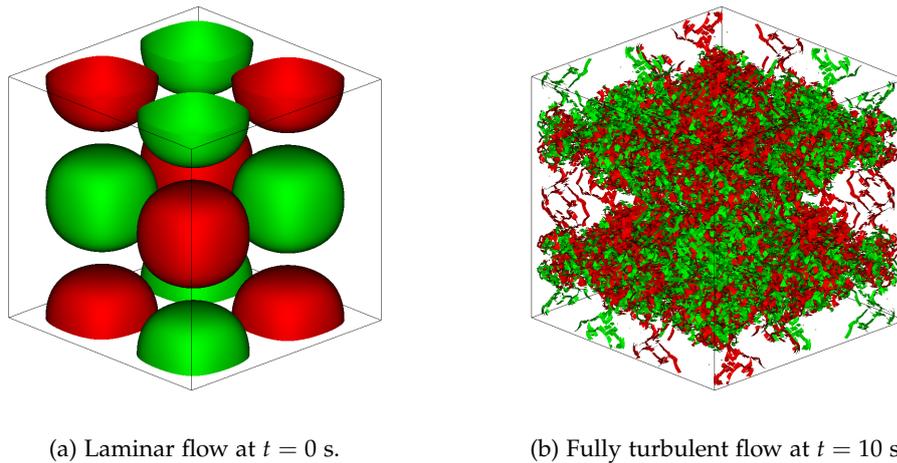


Figure 3.3: Taylor–Green vortex flow at  $Re = 1600$ . Iso-surfaces of  $z$ -vorticity.

Due to the symmetric behavior of the flow, we are allowed to simulate only an eighth part of the domain. Hence, we take as computational domain  $\Omega^h = [0, \pi]^3$  and apply no-penetration boundary conditions. All the implementations employ NURBS basis functions that are mostly  $C^1$ -quadratic, however every velocity space is enriched to be cubic  $C^2$  in the associated direction [34, 35, 68, 69]. Note that conservation of angular momentum cannot be guaranteed, since the choice of the weighting function  $\theta^h$  in Section 3.5.3 is not valid. We apply a standard  $L_2$ -projection to set the initial condition on the mesh. For the time-integration we employ the generalized- $\alpha$  method with the parameter choices of [57] which yield correct energy evolution. This method is stable and shows second-order temporal accuracy. The resulting system of equations is solved with the standard flexible GMRES method with additive Schwartz preconditioning provided by Petsc [11, 12].

We perform simulations with three different methods: (i) the classical Galerkin method, (ii) the VMS method with static small-scales (VMSS), comparable with [17] and (iii) the novel Galerkin/least-squares formulation with dynamic and divergence-free small-scales (GLSDD), i.e. form (3.4.17). The DNS results of Brachet et al. [30] obtained with a spectral method on a fine  $512^3$ -mesh serve as reference data (ref).

First, we perform a brief mesh refinement study for the novel method. Figure 3.4 shows mesh refined results for the novel GLSDD method (3.4.17). For this purpose meshes with  $16^3$ ,  $24^3$ ,  $32^3$ ,  $48^3$  elements have been employed. Clearly, the energy behavior on the coarsest two meshes is quite off. The finer meshes are able to closely capture the turbulence character of the flow. In the following we therefore use meshes of  $32^3$  or  $48^3$  elements.

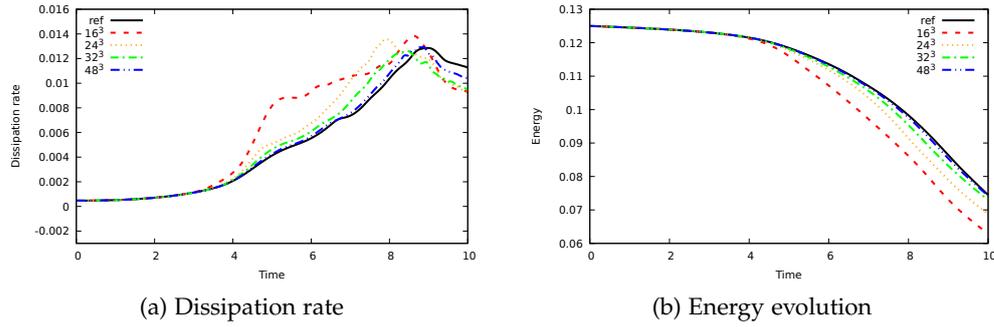


Figure 3.4: Taylor–Green vortex flow at  $Re = 1600$  mesh convergence. The GLS method with dynamic divergence-free small-scales.

We compare the results of the novel GLSDD method with the VMSS and the Galerkin approach. The simulations are carried out on a mesh of  $32^3$  elements, i.e. the mesh size is  $h = \frac{\pi}{32}$ , and on a slightly finer mesh of  $48^3$  elements. The time-step is taken as  $\Delta t = \frac{4h}{5\pi}$ , i.e. the initial CFL-number is roughly 0.25. In the Figures 3.5-3.6 we visualize the time history of the kinetic energy and kinetic energy dissipation rate for each of the three methods and the reference data.

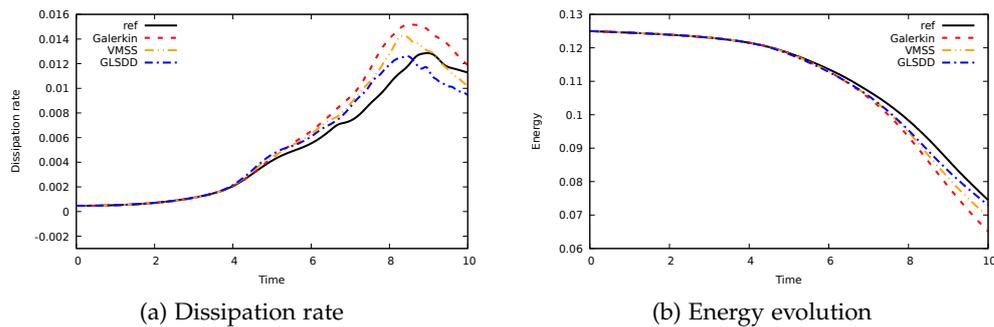


Figure 3.5: Taylor–Green vortex flow at  $Re = 1600$  on  $32^3$ -mesh for various methods: the Galerkin method, the VMS method with static small-scales and the GLS method with dynamic divergence-free small-scales.

The Figure 3.5 shows that each of the methods is able to roughly capture the energy behavior on the coarse mesh. The dissipation peak appears too early in time

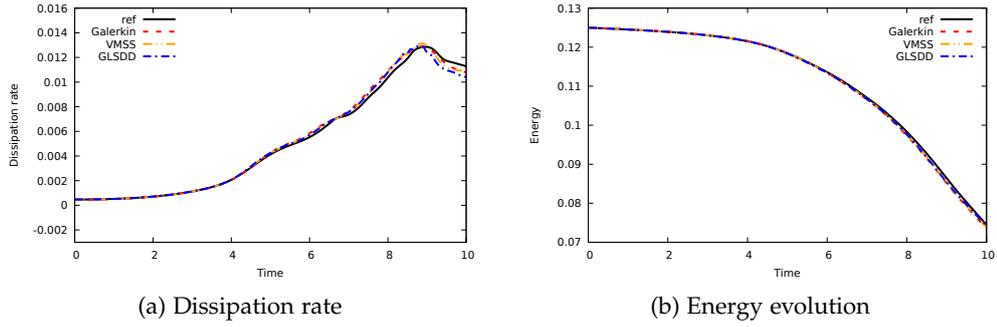


Figure 3.6: Taylor–Green vortex flow at  $Re = 1600$  on  $48^3$ -mesh for various methods: the Galerkin method, the VMS method with static small-scales and the GLS method with dynamic divergence-free small-scales.

for each of the simulations. The Galerkin method displays the least accurate results, it overpredicts the dissipation rate. The VMSS method performs a bit better at all times. The novel GLSDD approach demonstrates an even closer agreement with the reference results. The results on the finer mesh, in Figure 3.6, reveal almost no difference with the reference data.

In the following we further analyze the contributions of the dissipation rate (on the course mesh). The dissipation rate of the Galerkin method only consists of the large-scale/physical dissipation  $\|v^{1/2}\nabla\mathbf{u}^h\|_{\Omega}^2$ . In contrast, the dissipation of the GLSDD method is composed of a large-scale and a small-scale contribution:

$$\frac{d}{dt}\mathcal{E}_{\Omega}^{\text{GLSDD}} = -\|v^{1/2}\nabla\mathbf{u}^h\|_{\Omega}^2 - \|\tau_M^{-1/2}\mathbf{u}'\|_{\Omega}^2. \quad (3.6.2)$$

In Figure 3.7 we display the temporal evolution of both parts and the small-scale dissipation fraction  $(\|\tau_M^{-1/2}\mathbf{u}'\|_{\Omega}^2)/(\|v^{1/2}\nabla\mathbf{u}^h\|_{\Omega}^2 + \|\tau_M^{-1/2}\mathbf{u}'\|_{\Omega}^2)$ . In the laminar regime ( $t < 3$ ) the small-scale contribution is negligible. When the flow has a more turbulent character the contribution of the small-scales is substantial: the maximum of the dissipation fraction exceeds 0.35.

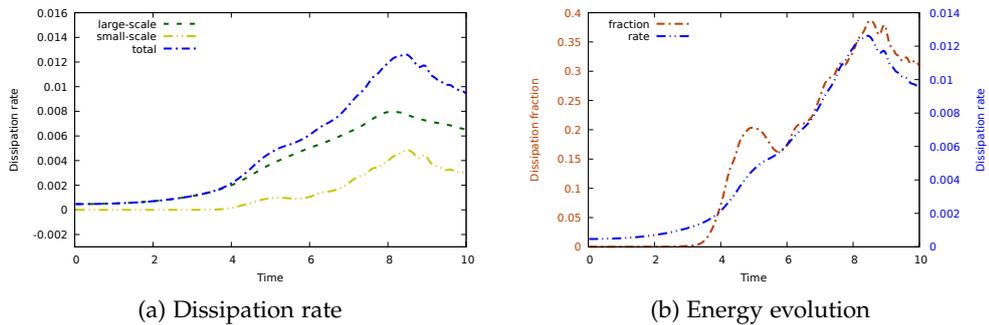


Figure 3.7: Taylor–Green vortex flow at  $Re = 1600$  on  $32^3$ -mesh with the GLSDD method: (a) large-scale and small-scale parts of the dissipation rate and (b) their fraction.

Lastly, we focus on the energy dissipation of the VMSS formulation. The derivation follows the same steps used throughout this work. One might argue that the energy

could also be solely based on the large-scales. This is what we do here. Its evolution reads:

$$\begin{aligned} \frac{d}{dt} \mathcal{E}_\Omega^{h, \text{VMSS}} = & - \|\nu^{1/2} \nabla \mathbf{u}^h\|_\Omega^2 - \|\tau_M^{-1/2} \mathbf{u}'\|_\Omega^2 + (\nu \Delta \mathbf{u}^h, \mathbf{u}')_{\bar{\Omega}} \\ & - (\mathbf{u}', \partial_t \mathbf{u}^h)_{\bar{\Omega}} + (\nabla \cdot \mathbf{u}^h, p')_{\bar{\Omega}} \\ & + \left( \nabla \mathbf{u}^h, (\mathbf{u}^h + \mathbf{u}') \otimes (\mathbf{u}^h + \mathbf{u}') \right)_{\bar{\Omega}} - \left( \mathbf{u}', (\mathbf{u}^h + \mathbf{u}') \cdot \nabla \mathbf{u}^h \right)_{\bar{\Omega}}. \end{aligned} \quad (3.6.3)$$

Figure 3.8 shows the contribution of the separate terms. The two desired dissipation terms are clearly dominant. The small-scale dissipation is smaller than the large-scale dissipation, however it has a significant contribution. Although the contributions are small, the unwanted terms can create artificial energy.

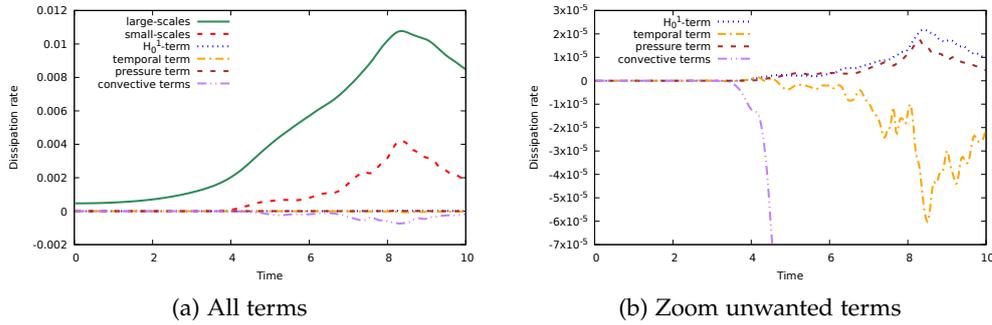


Figure 3.8: Taylor–Green vortex flow at  $Re = 1600$  on  $32^3$ -mesh with the VMSS method: energy dissipation of separate terms.

### 3.7 CONCLUSIONS

We continued the study initiated in [57] concerning the construction of methods displaying correct-energy behavior. In this work we have applied the developed methodology to the incompressible Navier–Stokes equations. It clearly shows that the link between the methods VMS, SUPG and GLS, established in [57], is also valid for the incompressible Navier–Stokes equations.

The novel GLSDD methodology employs divergence-conforming NURBS basis functions and uses a Lagrange multiplier setting to ensure divergence-free small-scales. Furthermore, it enjoys the favorable behavior of the dynamic small-scales and reduces to the Galerkin method in the Stokes regime. These properties all emerge from the correct-energy design condition. A pleasant byproduct of the method is the conservation of linear momentum. The conservation of angular momentum can be achieved when employing the appropriate weighting function spaces. The numerical results support the theoretical framework in that the energy behavior improves upon the VMS method with static small-scales. The variational multiscale method with static small-scales has unwanted small-scale contributions which create artificial energy.

The novel formulation requires a bit more effort to implement compared to the variational multiscale method with static small-scales. One has to include an additional variable to ensure the divergence-free behavior of the small-scales. In addition

the formulation needs to be equipped with the dynamic small-scale model. However, the resulting system of equations does not demand a sophisticated preconditioner; we have employed the standard ASM (Additive Schwarz Method) technique. In our opinion, the accuracy gain outweighs the little extra implementation effort and calculation cost.

A possible directions for future work concerns the development of an energy-dissipative method for the computation of free surface flow problems. This is an important step, since artificial energy creation can yield highly unstable behavior, as demonstrated in [3].

## 3.A GLS FORMULATION WITH DYNAMIC DIVERGENCE-FREE SMALL-SCALES

We repeat the Galerkin/least-squares formulation with dynamic divergence-free small-scales (GLSDD), i.e. form (3.4.17), to provide an overview of the separate terms. The formulation is of skew-symmetric type, applies GLS stabilization and uses divergence-free dynamic small-scales. The method requires a stable velocity–pressure pair and reads:

Find  $(\mathbf{u}^h, p^h, \zeta^h) \in \mathcal{W}^h \times \mathcal{P}^h$  such that for all  $(\mathbf{w}^h, q^h, \theta^h) \in \mathcal{W}^h \times \mathcal{P}^h$ ,

$$\begin{aligned}
 & \left( \mathbf{w}^h, \partial_t \mathbf{u}^h \right)_{\Omega} + \left( \mathbf{w}^h, \partial_t \mathbf{u}' \right)_{\tilde{\Omega}} \\
 & \quad + \frac{1}{2} (\mathbf{w}^h, (\mathbf{u}^h + \mathbf{u}') \cdot \nabla \mathbf{u}^h)_{\Omega} - \frac{1}{2} ((\mathbf{u}^h + \mathbf{u}') \cdot \nabla \mathbf{w}^h, \mathbf{u}^h)_{\Omega} \\
 & \quad - \left( (\mathbf{u}^h + \mathbf{u}') \cdot \nabla \mathbf{w}^h, \mathbf{u}' \right)_{\tilde{\Omega}} \\
 & \quad + (\nabla \mathbf{w}^h, 2\nu \nabla^s \mathbf{u}^h)_{\Omega} + \left( \nu \Delta \mathbf{w}^h, \mathbf{u}' \right)_{\tilde{\Omega}} \\
 & \quad + (q^h, \nabla \cdot \mathbf{u}^h)_{\Omega} - (\nabla \cdot \mathbf{w}^h, p^h)_{\Omega} \\
 & \quad + \left( \nabla \theta^h, \mathbf{u}' \right)_{\tilde{\Omega}} = (\mathbf{w}, \mathbf{f})_{\Omega},
 \end{aligned} \tag{3.A.1a}$$

$$\partial_t \mathbf{u}' + \tau_M^{-1} \mathbf{u}' + \nabla \zeta^h + \mathbf{r}_M = 0, \tag{3.A.1b}$$

where momentum residual is

$$\mathbf{r}_M = \partial_t \mathbf{u}^h + \left( (\mathbf{u}^h + \mathbf{u}') \cdot \nabla \right) \mathbf{u}^h + \nabla p^h - \nu \Delta \mathbf{u}^h - \mathbf{f}. \tag{3.A.2}$$

The separate terms of (3.A.1a) are from left to right: the temporal terms, the convective contributions, the viscous contributions, the incompressibility constraint, the pressure term, the divergence-free small-scale velocity constraint and the forcing term. This form follows the correct-energy evolution (on a local scale):

$$\begin{aligned}
 \frac{d}{dt} \mathcal{E}_{\omega} &= - \| \nu^{1/2} \nabla \mathbf{u}^h \|_{\omega}^2 + (\mathbf{u}^h, \mathbf{f})_{\omega} - (1, \mathcal{F}_{\omega}^h)_{\chi_{\omega}} \\
 & \quad - \| \tau_M^{-1/2} \mathbf{u}' \|_{\tilde{\omega}}^2 + (\mathbf{u}', \mathbf{f})_{\tilde{\omega}},
 \end{aligned} \tag{3.A.3}$$

and possesses the conservation properties of Section 3.5.

## 3.B DEFINITION DYNAMIC STABILIZATION PARAMETER

The dynamic stabilization parameter  $\tau_M$  is the discrete approximation of the inverse of the convective and viscous parts of momentum Navier–Stokes operator. It mirrors the dynamic stabilization parameter of convection–diffusion equation (see [57]). The continuity stabilization parameter  $\tau_C$  is on its turn the discrete approximation of the inverse of the divergence operator, here we use the objective definition introduced in [16]. The parameters take the form:

$$\tau_M = \left( \tau_{\text{conv}}^{-2} + \tau_{\text{visc}}^{-2} \right)^{-1/2}, \tag{3.B.1a}$$

$$\tau_C = \left( \tau_M \sqrt{\mathbf{g} : \mathbf{g}} \right)^{-1}, \tag{3.B.1b}$$

where the convective and viscous contributions of  $\tau_M$  are

$$\tau_{\text{conv}}^{-2} = 4\mathbf{u} \cdot \mathbf{g}\mathbf{u}, \quad (3.B.2a)$$

$$\tau_{\text{visc}}^{-2} = C_I \nu^2 \mathbf{g} : \mathbf{g}. \quad (3.B.2b)$$

Here the following definition is employed:

$$\mathbf{g} = \frac{\partial \xi}{\partial x} \frac{\partial \xi}{\partial x'} \quad (3.B.3a)$$

$$\mathbf{g} : \mathbf{g} = \sum_{i,j=1}^3 G_{ij} G_{ij}, \quad (3.B.3b)$$

where  $\partial \xi / \partial x$  is the inverse Jacobian of the map between the elements in the reference and physical domain. The positive constant  $C_I$  is determined by an inverse estimate.

## Part II

### A FRAMEWORK FOR DISCONTINUITY CAPTURING METHODS

This part presents a variational multiscale analysis framework for discontinuity capturing methods. The methodology relies on proposed variation entropy theory. First we describe variation entropy theory and then use this to derive discontinuity capturing methods within the multiscale framework.



# 4

---

## VARIATION ENTROPY: A CONTINUOUS LOCAL GENERALIZATION OF THE TVD PROPERTY USING ENTROPY PRINCIPLES.

---

*“Sometimes what you want is right in front of you. All you have to do is open your eyes and see it.” – Meg Cabot (writer)*

This chapter is reproduced from [59]:

M.F.P. ten Eikelder and I. Akkerman, *Variation entropy: a continuous local generalization of the TVD property using entropy principles.*, *Comput. Meth. Appl. Mech. Engrg.* 355 (2019) 261-283.

### ABSTRACT

*This work presents the notion of a variation entropy. This concept is an entropy framework for the gradient of the solution of a conservation law instead of on the solution itself. It appears that all semi-norms are admissible variation entropies. This provides insight into the total variation diminishing property and justifies it from entropy principles. The framework allows to derive new local variation diminishing properties in the continuous form. This can facilitate the design of new numerical methods for problems that contain discontinuities.*

## 4.1 INTRODUCTION

Violent disturbances emerging from sudden changes in velocity, pressure and temperature, known as shock waves appear everywhere in nature, science and industrial applications. Examples are water–air flows, supersonic flights, the water hammer phenomena, shock–bubble interaction, material impact and sudden changes in crowd dynamics. The behavior of these phenomena is usually governed by nonlinear conservation laws. The development of numerical techniques for the solution procedure of conservation laws is challenging because higher-order methods produce oscillations near shocks. There exists a large class of numerical methods which aim to tackle these oscillations via reducing to first-order spatial accuracy at the shock wave. These techniques augment the numerical method in one way or another with artificial diffusion or viscosity in the shock wave region.

Most numerical methods developed for problems involving shock waves use finite-difference or finite-volume approaches. These methods are often well-established and show good performance in numerical computations. The employed mechanisms can often be linked to one of the following. The concept of flux limiters (MUSCL), see e.g., [132, 182, 192, 195, 196], reduces the scheme at the shock to first-order by adding diffusion. The monotonicity property introduced by Harten in 1983 [88] precludes the creation of local extrema and ensures that local minima (maxima) are non-decreasing (non-increasing). Perhaps the most relevant in some numerical simulations is the maximum principle, see [88] or the more recent work [148, 204]. This principle states that the solution values remain between the minimum and maximum of the initial condition. This is in particular important in simulations of physical quantities that should remain non-negative, e.g., densities and also, in the case of two-fluid problems, volume-fractions. A negative density or a volume-fraction exceeding the zero-one range can directly lead to a blow-up of the simulations. Therefore numerical methods that preclude this by design are often sought after, see e.g., for compressible two-fluid flow simulations [53, 63, 172].

Possibly the most famous property is the total variation diminishing (TVD) property [88, 90, 175]. The total variation diminishing schemes preclude the growth of the total variation of the solution. These methods ensure that the numerical solution  $\phi$  of a PDE satisfies

$$\text{TV}(\phi^{n+1}) \leq \text{TV}(\phi^n),$$

where  $n$  denotes the time-level. Here the total variation  $\text{TV}$  is defined in one and two dimensions respectively as:

$$\begin{aligned} \text{TV}(\phi) &= \sum_j |\phi_{j+1} - \phi_j|, \\ \text{TV}(\phi) &= \sum_{j,k} (\Delta y |\phi_{j+1,k} - \phi_{j,k}| + \Delta x |\phi_{j,k+1} - \phi_{j,k}|), \end{aligned}$$

with the subscript the spatial index and with  $|\cdot|$  the absolute value. This definition is based on the discrete approximation of the  $L_1$ -norm of the gradient. It is important to emphasize that it represents a global property that assumes an underlying Cartesian grid that is possibly non-uniform and does not have a straightforward equivalent for unstructured grids. Furthermore, the TVD property does not readily extend to

multiple dimensions. This has been the motivation for the development of local extremum diminishing (LED) schemes [118, 119]. In the two-dimensional case, the definition of the total variation depends on the orientation of the grid, i.e. it would change if one would rotate the grid. This is undesirable since (i) this total variation does not have a continuous counterpart and (ii) one generally does not want the numerical method to depend on the coordinate system (objectivity). Note that all of the above mentioned features and properties are satisfied by the exact solution and are in the discrete setting closely linked. In particular cases (e.g., one-dimensional, scalar) one implies another [88].

In the framework of finite element methods several stabilized methods have been proposed to deal with spurious wiggles in the solution profiles of convection-dominated problems. The well-known methods are the Streamline upwind-Petrov Galerkin (SUPG) method [33], the Galerkin/least-squares method [106] and the variational multiscale method [100, 103, 115]. The latter method offers a rich prospect for design new stabilized methods and has gained a lot of attention recently [15, 49, 57, 58, 202]. In the direction of TVD schemes and maximum principles, several VMS methods have been proposed. For example a total variation bounding constraint [67] and the maximum principle [70] has been enforced in the VMS context. When shocks waves form, plain stabilized methods are not sufficient and additional dissipation mechanisms are necessary. These mechanisms are called discontinuity capturing (DC) operators [110, 139], and are sometimes residual-based [5, 19, 185] or entropy-residual based [85]. We refer for an overview of DC to the review papers [114, 121, 122].

To single out the physically relevant solutions the concept of entropy solutions has been proposed by Krupnikov in his seminal 1970 paper [135]. The entropy solution is a limiting case of a generalized solution which perturbs the conservation law with a diffusion term and can be used to prove existence, uniqueness and stability theorems. In the case of systems of conservation laws, Friedrichs, Kurt and Lax show in 1971 that if an additional conserved quantity is a convex function of the solution then the system of equations can be symmetrized and provides a corresponding entropy inequality [76]. Harten continues the research on symmetrizability of systems of conservation laws which possess entropy functions [89]. Additionally, he provides symmetric formulations in conservative variables for the Euler equations of gas dynamics. Tadmor shows a year later, in 1984, that the concepts of symmetrizability, having an entropy function and having a so-called skew-selfadjoint form are equivalent [184]. Furthermore, Tadmor identifies in [183] that any symmetric system of conservation laws is equipped with a one-parameter family of entropy functions. The work of Harten and Tadmor has been generalized by Hughes et al. [107] to the compressible Navier-Stokes equations with heat conduction effects. The corresponding finite element schemes satisfy by design the second law of thermodynamics, see also [40].

Although total variation diminishing schemes have proven their power and relevance, their use seems to be restricted to finite-difference/finite-volume discretizations and a generic finite element variant seems to be missing. Moreover, the different concepts of total variation diminishing schemes and entropy solutions/entropy variables both target to improve the solution quality at shock waves. Despite

that they serve the same goal, a clear connection (on the continuous level) is missing. These observations led to ponder the following two questions:

- *How can we construct a local continuous generalization of the TVD stability condition?*
- *Is there a connection between entropy solutions and the TVD property?*

The current chapter aims to answer these questions. To that purpose, we introduce a new stability concept which we call *variation entropy*. Similar to the well-known entropy concept, *variation entropy solutions* are those solutions for which an additional quantity is conserved or dissipated. The fundamental difference is that a variation entropy is a function of the *gradient of the solution* rather than the solution itself. Variation entropy solutions are presented in the continuous setting and are as such not restricted to a particular discretization. Therefore, in contrast to the TVD property, numerical methods employing variation entropy concept may be based on a variational setting (e.g., finite element methods). An important observation is that the TVD property may be derived from a specific variation entropy solution.

We summarize the main definitions and results formally (i.e. up to regularization). Consider a conservation law for the unknown  $\phi$  and flux  $f = f(\phi)$ :

$$\partial_t \phi + \nabla \cdot f = 0.$$

The *variation entropy* and the *variation entropy solution* are defined as follows.

**Definition 4.1.1.** *The pair  $(\eta, q)$  with  $\eta = \eta(\nabla \phi)$  is termed a variation entropy–variation entropy flux pair if*

- *$\eta$  is convex*
- *$\eta$  satisfies the homogeneity property:  $\nabla \phi \cdot \frac{\partial \eta}{\partial \nabla \phi} = \eta$*
- *$q$  is given by  $q = \eta \frac{\partial f}{\partial \phi}$ .*

**Definition 4.1.2.** *We call the function  $\phi$  a variation entropy solution if it is an integral solution of the conservation law and formally satisfies for each variation entropy–variation entropy flux pair:*

$$\partial_t \eta + \nabla \cdot q \leq 0.$$

An important class of variation entropies is formed by the semi-norms.

**Theorem 4.1.3.** *If a function is a semi-norm then it is a variation entropy.*

The famous TVD property is special case of the following corollary.

**Corollary 4.1.4.** *In case of periodic or no-inflow boundaries the variation entropy decays in time via:*

$$\int_{\Omega} \eta(\nabla \phi(x, t)) \, d\Omega \leq \int_{\Omega} \eta(\nabla \phi_0(x)) \, d\Omega, \quad \text{for all } t > 0.$$

The remainder of the chapter can be summarized as follows. Section 4.2 provides a brief summary of the entropy solutions in the classical sense. In Section 4.3 the concept of the variation entropy solutions is presented. This section identifies the class of possible variation entropies. Section 4.4 discusses the selection of variation entropies. In particular, the well-known TVD property is here presented in an entropy context. Finally, Section 4.5 draws the conclusions and outline avenues for future research.

4.2 ENTROPY SOLUTIONS IN THE CLASSICAL SENSE

Let  $\Omega \subset \mathbb{R}^d$  be an open connected domain. Let us consider the scalar conservation problem: find  $\phi : \Omega \times \mathcal{I} \rightarrow \mathbb{R}$  such that

$$\partial_t \phi + \nabla \cdot \mathbf{f} = 0, \quad (\mathbf{x}, t) \in \Omega \times \mathcal{I}, \tag{4.2.1}$$

subject to the initial condition  $\phi(\mathbf{x}, 0) = \phi_0(\mathbf{x}) \in L^\infty(\Omega)$ . Here  $\mathbf{f} = \mathbf{f}(\phi) \in \mathcal{C}(\Omega, \mathbb{R})$  is the (nonlinear) flux, the spatial coordinate denotes  $\mathbf{x} \in \Omega$ , the time is  $t \in \mathcal{I} = (0, T)$  with  $T > 0$ . Solutions of (4.2.1) can contain discontinuities (shocks, rarefaction waves) which motivates the search for weak solutions. A weak solution  $\phi \in L^\infty(\Omega, \mathbb{R}_+)$  of (4.2.1) satisfies

$$\int_{\mathbb{R}_+} \int_{\Omega} \phi \partial_t \psi + \mathbf{f}(\phi) \nabla \psi \, d\Omega \, dt + \int_{\Omega} \phi_0 \psi_0 \, d\Omega = 0 \tag{4.2.2}$$

for all test functions  $\psi \in \mathcal{C}_c^1(\Omega, \mathbb{R}^+)$  with  $\psi_0(\mathbf{x}) = \psi(\mathbf{x}, 0)$ . Weak solutions are generally not unique.

Let us first consider the case of smooth solutions. Let  $\eta = \eta(\phi) \in \mathcal{C}^1(\mathbb{R})$  be a convex function. Multiplying (4.2.1) with  $\partial \eta / \partial \phi$  shows that  $\eta$  satisfies a conservation law

$$\partial_t \eta + \nabla \cdot \mathbf{q} = 0, \tag{4.2.3}$$

when the flux  $\mathbf{q}$  satisfies the *compatibility condition*:

$$\frac{\partial \mathbf{q}}{\partial \phi} = \frac{\partial \eta}{\partial \phi} \frac{\partial \mathbf{f}}{\partial \phi}. \tag{4.2.4}$$

**Definition 4.2.1.** (*Entropy/entropy-flux pair*) The pair  $(\eta, \mathbf{q})$  is called an *entropy/entropy-flux pair* when  $\eta$  is a convex function and  $\mathbf{q}$  fulfills the compatibility condition (4.2.4). The function  $\eta$  is referred to as the *entropy function* and  $\mathbf{q}$  as the *entropy flux*.

When discontinuities appear the chain rule cannot be applied anymore and (4.2.3) is replaced by an inequality:

$$\partial_t \eta + \nabla \cdot \mathbf{q} \leq 0. \tag{4.2.5}$$

The entropy condition (4.2.5) tells us that the entropy  $\eta$  dissipates at shock waves. This inequality should be understood as

$$\int_0^\infty \int_{\Omega} \eta(\phi) \partial_t w + \mathbf{q}(\phi) \cdot \nabla w \, d\Omega \, dt \geq 0, \tag{4.2.6}$$

for all  $w \in \mathcal{C}_c^\infty(\Omega \times (0, \infty))$ ,  $w \geq 0$ .

**Definition 4.2.2.** (*Entropy solution*) The function  $\phi$  is called an entropy solution or entropic if it is an integral solution and additionally satisfies (4.2.6) for each entropy/entropy-flux pair.

Consider solutions  $\phi^\epsilon : \Omega \times \mathcal{I} \rightarrow \mathbb{R}$  of the approximate viscous problem:

$$\partial_t \phi^\epsilon + \nabla \cdot \mathbf{f}^\epsilon = \epsilon \Delta \phi^\epsilon, \quad (\mathbf{x}, t) \in \Omega \times \mathcal{I}, \quad (4.2.7)$$

with  $\mathbf{f}^\epsilon = \mathbf{f}(\phi^\epsilon)$ . The vanishing viscosity solution  $\phi$  is now defined as:  $\phi^\epsilon \rightarrow \phi$  a.e. for  $\epsilon \rightarrow 0$ .

**Theorem 4.2.3.** *Vanishing viscosity solutions are entropy solutions.*

*Proof.* This is a direct consequence of the convexity of  $\eta$  and the compatibility condition (4.2.4). For details see Evans [72].  $\square$

**Theorem 4.2.4.** *Entropy solutions of scalar conservation laws are unique.*

*Proof.* See Evans [72].  $\square$

In case  $\Omega$  is a periodic domain or has no-inflow boundaries (the inflow is characterized by  $\partial \mathbf{f} / \partial \phi \cdot \mathbf{n} \leq 0$  where  $\mathbf{n}$  is the outward normal), integration of (4.2.5) over  $\Omega$  leads to a decay of the overall entropy:

$$\frac{d}{dt} \int_{\Omega} \eta(\phi(\mathbf{x}, t)) \, d\Omega \leq 0, \quad (4.2.8)$$

which implies:

$$\int_{\Omega} \eta(\phi(\mathbf{x}, t)) \, d\Omega \leq \int_{\Omega} \eta(\phi_0(\mathbf{x})) \, d\Omega, \quad \text{for all } t \geq 0. \quad (4.2.9)$$

Note that taking  $\eta(\phi) = \phi^2/2$  leads to the usual  $L^2$ -stability from linear theory for hyperbolic equations. We refer to [140] for more details.

**Remark 4.2.5.** *We can also consider flux functions of the form  $\mathbf{f} = \mathbf{f}(\phi, \nabla \phi)$ . For the sake of simplicity we restrict ourselves to the case where the matrix  $\partial \mathbf{f} / \partial \nabla \phi$  is of the form  $-k\mathbf{I}$  with the scalar  $k \geq 0$  and the identity matrix  $\mathbf{I}$ . The entropy condition (4.2.5) now takes the form*

$$\partial_t \eta + \nabla \cdot \mathbf{q} - k \Delta \eta \leq 0. \quad (4.2.10)$$

### 4.3 VARIATION ENTROPY SOLUTIONS

In this section we introduce the notion of variation entropy solutions.

#### 4.3.1 The concept

We present the variation entropy concept for scalar conservation laws. The extension to systems of conservation laws may be the topic of another work. We consider the nonlinear conservation law

$$\partial_t \phi + \nabla \cdot \mathbf{f} = 0, \quad (\mathbf{x}, t) \in \Omega \times \mathcal{I}, \quad (4.3.1)$$

with flux  $f = f(\phi)$ . Let us first consider smooth solutions. The main idea is to look at the associated entropy relation of the spatial gradient (or variation) of the conservation law instead of that of the plain conservation law, i.e. consider the entropy relation of the system of equations:

$$\nabla(\partial_t \phi) + \nabla(\nabla \cdot f) = 0. \quad (4.3.2)$$

The motivation of the approach stems from the observation that sharp layers in solution profiles are characterized by large gradients. By considering a convex function of the solution gradient these sharp layers can be identified. In Figure 4.1 we sketch the concept.

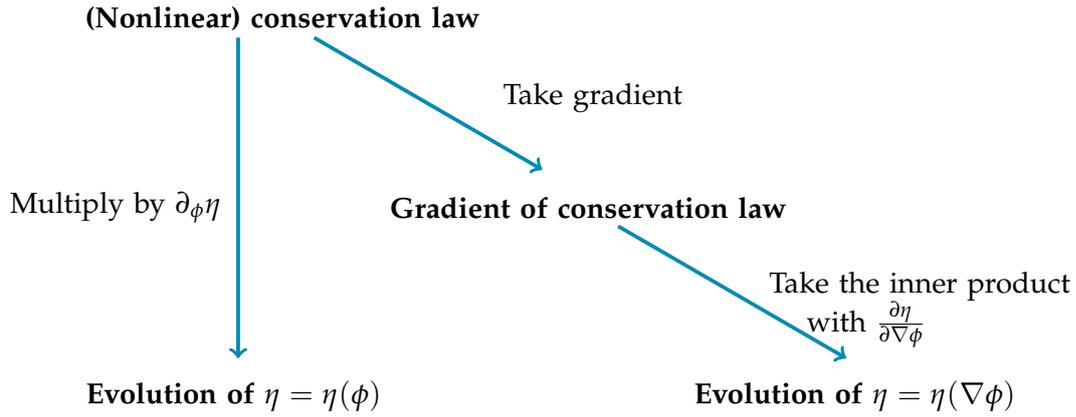


Figure 4.1: The concept of the variation entropy for a smooth solution. In the classical approach one considers the entropy of the conservation law. The idea of the variation entropy approach is to first take the gradient of the conservation law and subsequently introducing the entropy concept.

**Lemma 4.3.1.** (Evolution equation) Let  $\phi : \Omega \times \mathcal{I} \rightarrow \mathbb{R}$  be a smooth solution of the conservation law (4.3.1) and let  $\eta : \mathbb{R}^d \rightarrow \mathbb{R}$  be a twice differentiable convex function of the gradient of  $\phi$ , i.e.  $\eta = \eta(\nabla \phi)$ . The temporal evolution of  $\eta$  reads:

$$\partial_t \eta + \nabla \cdot \mathbf{q} = \mathcal{A}, \quad (4.3.3)$$

where the flux  $\mathbf{q}$  and the non-conservative term  $\mathcal{A}$  are respectively given by:

$$\mathbf{q} = \left( \frac{\partial \eta}{\partial \nabla \phi} \cdot \nabla \phi \right) \frac{\partial f}{\partial \phi}, \quad (4.3.4a)$$

$$\mathcal{A} = (\mathbf{H}_{\nabla \phi} \eta \nabla \phi) \cdot \left( \mathbf{H}_x \phi \frac{\partial f}{\partial \phi} \right). \quad (4.3.4b)$$

The contraction is defined as  $\mathbf{A} : \mathbf{B} = \text{Tr}(\mathbf{A}\mathbf{B}^T)$ ,  $(\mathbf{H}_x \phi)_{mn} = \partial^2 \phi / \partial x_m \partial x_n$  is the (symmetric) Hessian of  $\phi$  and  $(\mathbf{H}_{\nabla \phi} \eta)_{mn} = \partial^2 \eta / \partial \nabla_m \phi \partial \nabla_n \phi$  is the (symmetric) Hessian of  $\eta$ .

Under certain conditions presented in following subsections, the function  $\eta$  plays the role of an entropy. To distinguish from the classical entropy setting, we will use the term *variation entropy* which will be defined later.

*Proof.* (Lemma 4.3.1) Changing the order of differentiation and subsequently taking the inner product of (4.3.2) with the *variation entropy variables*  $\partial\eta/\partial\nabla\phi$  (vector-valued) we find:

$$\partial_t\eta + \frac{\partial\eta}{\partial\nabla\phi} \cdot \nabla(\nabla \cdot f) = 0. \quad (4.3.5)$$

Consider the second term of (4.3.5) in isolation. We interchange the gradient and divergence operators and use the product rule to arrive at:

$$\begin{aligned} \frac{\partial\eta}{\partial\nabla\phi} \cdot (\nabla(\nabla \cdot f)) &= \frac{\partial\eta}{\partial\nabla\phi} \cdot (\nabla \cdot (\nabla f)^T) \\ &= \nabla \cdot \left( \nabla f \frac{\partial\eta}{\partial\nabla\phi} \right) - \left( \nabla \left( \frac{\partial\eta}{\partial\nabla\phi} \right) \right)^T : \nabla f. \end{aligned} \quad (4.3.6)$$

Here we use the notation  $\nabla \cdot \mathbf{T} = \partial_{x_j} T_{ij}$  for the divergence of a tensor. The first term of (4.3.6) yields the flux term on the left-hand side of (4.3.3). Using the chain rule we get straightforwardly

$$\nabla f \frac{\partial\eta}{\partial\nabla\phi} = \left( \frac{\partial\eta}{\partial\nabla\phi} \cdot \nabla\phi \right) \frac{\partial f}{\partial\phi}. \quad (4.3.7)$$

Next, by using the identity

$$\left( \nabla \left( \frac{\partial\eta}{\partial\nabla\phi} \right) \right)^T = \mathbf{H}_x\phi \mathbf{H}_{\nabla\phi}\eta, \quad (4.3.8)$$

the second term of (4.3.6) can be written as

$$\left( \nabla \left( \frac{\partial\eta}{\partial\nabla\phi} \right) \right)^T : \nabla f = (\mathbf{H}_{\nabla\phi}\eta \nabla\phi) \cdot \left( \mathbf{H}_x\phi \frac{\partial f}{\partial\phi} \right). \quad (4.3.9)$$

Combining (4.3.5), (4.3.6), (4.3.7) and (4.3.9) leads to the claim.  $\square$

The flux term on the left-hand side of (4.3.3), composed of a convective component, redistributes  $\eta$  over the domain. In absence of the non-conservative term  $\mathcal{A}$  the temporal evolution would mirror the classical entropy case: it would satisfy a conservation law. We proceed with identifying the class of functions  $\eta$  that closely resembles the classical entropy case.

#### 4.3.2 Characterization of variation entropies

To closely match the classical entropy concept, the variation entropy should equip solutions with a dissipative condition. As such the influence of the advection term  $\mathcal{A}$  on the right-hand side of (4.3.3) is unwanted. Thus it cannot be part of the variation entropy concept. As a second property, we copy the convexity demand of  $\eta$  from the classical case. Furthermore, since the concept serves to identify sharp layers, we demand that  $\eta$  vanishes in absence of spatial gradients. We propose the following design condition.

*Design condition:*

We seek a function  $\eta$  such that:

- $\mathcal{A} = 0$  (dissipative condition),
- $\eta$  is convex,
- $\eta(\mathbf{0}) = 0$ .

The following lemma identifies the class of functions  $\eta$  that satisfy the design condition.

**Lemma 4.3.2.** (*Variation entropy design condition*) *A convex function  $\eta$  satisfies the design condition for a general conservation law if and only if it fulfills the homogeneity property:*

$$\nabla\phi \cdot \frac{\partial\eta}{\partial\nabla\phi} = \eta. \quad (4.3.10)$$

*Proof.* The advection term  $\mathcal{A}$  vanishes for a general conservation law if and only if the vectors  $\mathbf{H}_{\nabla\phi}\eta\nabla\phi$  and  $\mathbf{H}_x\phi\partial f/\partial\phi$  are orthogonal. Since the sign of each of the components of the advective speed  $\partial f/\partial\phi$  is undetermined, the entries of  $\mathbf{H}_x\phi\partial f/\partial\phi$  can be positive or negative. Thus the only problem-independent vector orthogonal to  $\mathbf{H}_x\phi\partial f/\partial\phi$  is the zero vector. Hence seeking a function  $\eta$  for which  $\mathcal{A}$  vanishes for general conservation law is equivalent to searching a function  $\eta$  for which

$$\mathbf{H}_{\nabla\phi}\eta\nabla\phi = 0. \quad (4.3.11)$$

Observe that this system of equations can be cast into the form:

$$\frac{\partial}{\partial\nabla\phi} \left( \nabla\phi \cdot \frac{\partial\eta}{\partial\nabla\phi} \right) = \frac{\partial\eta}{\partial\nabla\phi}. \quad (4.3.12)$$

Integration of the  $i$ -th equation with respect to  $\nabla_i\phi$  provides

$$\nabla\phi \cdot \frac{\partial\eta}{\partial\nabla\phi} = \eta + c, \quad (4.3.13)$$

where  $c$  is a constant. The condition  $\eta(\mathbf{0}) = 0$  leads to  $c = 0$ , and thus the homogeneity property (4.3.10) follows.  $\square$

We call a function that satisfies the design condition a *variation entropy*, as stated in the following definition.

**Definition 4.3.3.** (*Variation entropy*) *A function  $\eta = \eta(\nabla\phi)$  is termed a variation entropy if*

- $\eta$  is convex
- $\eta$  satisfies the homogeneity property (4.3.10).

An alternative form of the homogeneity property is stated in the following lemma.

**Lemma 4.3.4.** (*Homogeneous function*) *A variation entropy is a positive homogeneous function of degree 1:*

$$\eta(\alpha\mathbf{v}) = \alpha\eta(\mathbf{v}), \quad \text{for all } \mathbf{v} \in \mathbb{R}^d, \alpha \geq 0. \quad (4.3.14)$$

*Proof.* This is a direct consequence of Euler's homogeneous function theorem.  $\square$

By employing Lemma 4.3.4, convexity is equivalent to sub-additivity.

**Lemma 4.3.5.** (*Sub-additivity*) A positive homogeneous function of degree 1 is convex if and only if it is sub-additive:

$$\eta(\mathbf{v}_1 + \mathbf{v}_2) \leq \eta(\mathbf{v}_1) + \eta(\mathbf{v}_2), \quad \text{for all } \mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^d. \quad (4.3.15)$$

Combining Lemma 4.3.4 and 4.3.5 leads to the following theorem.

**Theorem 4.3.6.** (*Variation entropy*) A function  $\eta = \eta(\nabla\phi)$  is a variation entropy if and only if

- $\eta$  is positive homogeneous function of degree 1
- $\eta$  is sub-additive.

**Theorem 4.3.7.** (*Semi-norm*) If a function is a semi-norm then it is a variation entropy.

*Proof.* The axioms of a seminorm are absolute homogeneity, i.e.

$$\eta(\alpha\mathbf{v}) = |\alpha|\eta(\mathbf{v}), \quad \text{for all } \mathbf{v} \in \mathbb{R}^d, \alpha \in \mathbb{R}, \quad (4.3.16)$$

and sub-additivity. The absolute homogeneity demand (4.3.16) is a specific case of the positive homogeneity (4.3.14).  $\square$

**Theorem 4.3.8.** (*Linear function*) If a function is linear then it is a variation entropy.

Note that Theorems 4.3.7 and 4.3.8 provide sufficient but not necessary conditions. We remark that a linear combinations of variation entropies form again a variation entropy.

**Corollary 4.3.9.** (*Linear combination*) Let  $\eta_k$  be variation entropies for  $k = 1, \dots, n$  for some integer  $n$ . The linear combination  $\eta := \sum_k \alpha_k \eta_k$  with  $\alpha_k \in \mathbb{R}^+$  is a variation entropy.

**Lemma 4.3.10.** (*Alternative form of the variation entropy*) A convex function  $\eta = \eta(\nabla\phi)$  is a variation entropy if and only if it is of the form

$$\eta = \eta(\nabla\phi) = \hat{\eta}(r, \boldsymbol{\theta}) = F(\boldsymbol{\theta})r, \quad (4.3.17)$$

where  $F = F(\boldsymbol{\theta})$  is a scalar-valued function and the spherical coordinates are the radius  $r$  and the angles  $\boldsymbol{\theta}$  corresponding to the coordinates  $\nabla\phi$ . The convexity demand translates in the two-dimensional case to

$$F(\theta) + F''(\theta) \geq 0. \quad (4.3.18)$$

The restrictions on the function  $F$  in the 3-dimensional case are more involved. We refer the interested reader to Appendix 4.A for the details.

**Remark 4.3.11.** Note that  $\hat{\eta} = rF(\boldsymbol{\theta})$  is not differentiable in the origin. However the origin is an important part of the domain since this is where  $\phi$  attains a local extremum. A possible way to bypass non-differentiability is to regularize  $\eta$  and thus to work with an approximate variation entropy.

*Proof.* (Lemma 4.3.10) We start from the homogeneity property

$$\mathbf{v} \cdot \frac{\partial \eta}{\partial \mathbf{v}} = \eta, \quad \text{for all } \mathbf{v} \in \mathbb{R}^d. \quad (4.3.19)$$

We now switch to spherical coordinates, i.e. the coordinates consist of a radial coordinate  $r$  and  $d - 1$  angular coordinates  $\theta \in [0, 2\pi)$ ,  $\varphi_m \in [0, \pi]$ ,  $m = 1, \dots, d - 2$ . The transformation is given by:

$$\begin{aligned} v_1 &= r \cos \theta \prod_{l=1}^{d-2} \sin \varphi_l, \\ v_2 &= r \sin \theta \prod_{l=1}^{d-2} \sin \varphi_l, \\ v_m &= r \cos \varphi_{m-2} \prod_{l=m-1}^{d-2} \sin \varphi_l \quad \text{for } m = 3, \dots, d - 1, \\ v_d &= r \cos \varphi_{d-2}. \end{aligned} \quad (4.3.20)$$

For  $d = 3$  the third line drops out, and for  $d = 2$  both the third and last lines vanish. Both cases reduce to the well-known transformations. Consult [27] for a derivation of a similar form. The direct consequence  $r \partial \mathbf{v} / \partial r = \mathbf{v}$  provides

$$r \frac{\partial \hat{\eta}}{\partial r} = r \frac{\partial \eta}{\partial \mathbf{v}} \cdot \frac{\partial \mathbf{v}}{\partial r} = \mathbf{v} \cdot \frac{\partial \eta}{\partial \mathbf{v}}. \quad (4.3.21)$$

This allows us to cast (4.3.13) into the differential equation

$$r \frac{\partial \hat{\eta}}{\partial r} = \hat{\eta}. \quad (4.3.22)$$

The corresponding solution follows straightforwardly

$$\hat{\eta} = \hat{\eta}(r, \varphi_1, \dots, \varphi_{d-2}, \theta) = \hat{\eta}(r, \boldsymbol{\theta}) = F(\boldsymbol{\theta})r, \quad (4.3.23)$$

with  $F$  a scalar-valued function.

We characterize the convexity of  $\eta$  by the positivity of the eigenvalues of the Hessian. In the two-dimensional case we have  $F = F(\theta)$ . The Hessian in polar coordinates takes the form:

$$\mathbf{H}_{\nabla \phi \eta} = \frac{F(\theta) + F''(\theta)}{r} \begin{pmatrix} \sin^2 \theta & -\cos \theta \sin \theta \\ -\cos \theta \sin \theta & \cos^2 \theta \end{pmatrix}. \quad (4.3.24)$$

Note that this is in line with the first demand. The eigenvalues  $\lambda_1, \lambda_2$  of  $\mathbf{H}_{\nabla \phi \eta}$  are

$$\lambda_1 = 0, \quad (4.3.25a)$$

$$\lambda_2 = \frac{F(\theta) + F''(\theta)}{r}, \quad (4.3.25b)$$

and convexity of  $\eta$  follows when  $F(\theta) + F''(\theta) \geq 0$ .

□

## 4.3.3 Variation entropy-variation entropy flux pairs

Employing the homogeneity property (4.3.10) the evolution equation of a variation entropy takes the following form.

**Lemma 4.3.12.** (*Evolution equation variation entropy*) Let  $\phi : \Omega \times \mathcal{I} \rightarrow \mathbb{R}$  be a smooth solution of the conservation law (4.3.1) and let  $\eta : \mathbb{R}^d \rightarrow \mathbb{R}$  be a twice differentiable convex function of the gradient of  $\phi$ , i.e.  $\eta = \eta(\nabla\phi)$ . The temporal evolution of a variation entropy  $\eta$  reads:

$$\partial_t \eta + \nabla \cdot \mathbf{q} = 0, \quad (4.3.26)$$

where the flux  $\mathbf{q}$  is given by:

$$\mathbf{q} = \eta \frac{\partial \mathbf{f}}{\partial \phi}. \quad (4.3.27)$$

**Definition 4.3.13.** (*Variation entropy-variation entropy flux pair*) The pair  $(\eta, \mathbf{q})$  is called a variation entropy-variation entropy flux pair provided

- $\eta$  is a variation entropy
- $\mathbf{q}$  satisfies the compatibility condition  $\mathbf{q} = \eta \frac{\partial \mathbf{f}}{\partial \phi}$ .

**Remark 4.3.14.** Taking the derivative of the compatibility condition with respect to  $\nabla\phi$  yields for  $\eta = \eta(\nabla\phi)$ :

$$\frac{\partial \mathbf{q}}{\partial \nabla \phi} = \frac{\partial \eta}{\partial \nabla \phi} \frac{\partial \mathbf{f}}{\partial \phi}. \quad (4.3.28)$$

This form highlights the relation with the compatibility condition of the classical case (4.2.4) which states for  $\eta = \eta(\phi)$ :

$$\frac{\partial \mathbf{q}}{\partial \phi} = \frac{\partial \eta}{\partial \phi} \frac{\partial \mathbf{f}}{\partial \phi}. \quad (4.3.29)$$

We proceed with presenting variation entropy solutions. In practice hyperbolic problems with discontinuities are an interesting application. When discontinuities appear the chain rule cannot be applied anymore and thus the foregoing derivation is not valid in the current form. In the classical entropy case at this point the conservation law of the entropy is replaced by an entropy inequality. This is well-defined at discontinuities. For the variation entropy concept, this is not possible at a discontinuity due to the gradient in  $\eta(\nabla\phi)$ . As such the concept of a variation entropy solution can be applied to problems with sharp layers but is in the current form not defined in case of discontinuities. We wish to emphasize that variation entropy solutions may be very suitable in numerics though. A numerical method approximates gradients at discontinuities (e.g., continuous finite-elements) and as such ill-defined gradients do not appear. The variation entropy solution provides a local stability condition that may be used to eliminate the spurious oscillations near sharp layers or shocks.

In the following we proceed with a smooth solution to bypass this singularity problem. We aim to closely resemble the classical entropy case in the following and

accordingly replace the variation entropy evolution equation by the *variation entropy condition*:

$$\partial_t \eta + \nabla \cdot \mathbf{q} \leq 0. \quad (4.3.30)$$

This means that the variation entropy evolves according to the flux  $\mathbf{q}$  but may experience sudden increases. The rigorous counterpart of (4.3.30) is:

$$\int_0^\infty \int_\Omega \eta \left( \partial_t w + \frac{\partial f}{\partial \phi} \cdot \nabla w \right) dx dt \geq 0. \quad (4.3.31)$$

In a similar fashion as for entropy solutions in the classical sense we define *variation entropy solutions*.

**Definition 4.3.15.** (*Variation entropy solution*) We call the smooth function  $\phi$  a *variation entropy solution* if it is an integral solution and additionally satisfies (4.3.31) for each variation entropy–variation entropy flux pair.

Let us now consider the solutions  $\phi^\epsilon : \Omega \times \mathcal{I} \rightarrow \mathbb{R}$  of the approximate viscous problem:

$$\partial_t \phi^\epsilon + \nabla \cdot \mathbf{f}^\epsilon = \epsilon \Delta \phi^\epsilon, \quad (\mathbf{x}, t) \in \Omega \times \mathcal{I}, \quad (4.3.32)$$

with viscosity parameter  $\epsilon > 0$ , flux  $\mathbf{f}^\epsilon \equiv \mathbf{f}(\phi^\epsilon)$ . Let the vanishing viscosity solution denote  $\phi = \lim_{\epsilon \rightarrow 0} \phi^\epsilon$ .

**Theorem 4.3.16.** (*Vanishing viscosity*) The smooth function  $\phi$  is a *variation entropy solution*.

*Proof.* The proof basically is the variation entropy counterpart of the classical entropy case presented in Evans [72]. The vanishing viscosity solution is an integral solution just as used for the classical entropy case. What remains is to show that it also satisfies the variation entropy condition. Taking the gradient of (4.3.32) and subsequently the inner product with  $\partial \eta / \partial \nabla \phi^\epsilon$  provides

$$\partial_t \eta^\epsilon + \nabla \cdot \mathbf{q}^\epsilon = \epsilon \frac{\partial \eta^\epsilon}{\partial \nabla \phi^\epsilon} \cdot \nabla (\Delta \phi^\epsilon), \quad (4.3.33)$$

where the superscript  $\epsilon$  indicates dependence on  $\nabla \phi^\epsilon$ . Next, by applying the chain rule we can evaluate the expression as:

$$\begin{aligned} \Delta \eta^\epsilon &= \nabla \cdot (\nabla \eta^\epsilon) \\ &= \nabla \cdot \left( \mathbf{H}_x \phi^\epsilon \frac{\partial \eta^\epsilon}{\partial \nabla \phi^\epsilon} \right) \\ &= (\mathbf{H}_x \phi^\epsilon \mathbf{H}_{\nabla \phi^\epsilon} \eta^\epsilon) : \mathbf{H}_x \phi^\epsilon + \frac{\partial \eta^\epsilon}{\partial \nabla \phi^\epsilon} \cdot \nabla (\Delta \phi^\epsilon). \end{aligned} \quad (4.3.34)$$

Substitution yields:

$$\partial_t \eta^\epsilon + \nabla \cdot \mathbf{q}^\epsilon = \epsilon \Delta \eta^\epsilon - \epsilon \mathbf{H}_x \phi^\epsilon \mathbf{H}_{\nabla \phi^\epsilon} \eta^\epsilon \mathbf{H}_x \phi^\epsilon, \quad (4.3.35)$$

where the flux  $\mathbf{q}^\epsilon$  satisfies the compatibility condition (4.3.27) and where the superscript  $\epsilon$  denotes dependence on  $\phi^\epsilon$  instead of  $\phi$ . By the convexity of  $\eta^\epsilon$  the second term on the right-hand side<sup>1</sup> has the sign:

$$(\mathbf{H}_x \phi^\epsilon \mathbf{H}_{\nabla \phi^\epsilon} \eta^\epsilon) : \mathbf{H}_x \phi^\epsilon \geq 0, \quad (4.3.36)$$

which implies

$$\partial_t \eta^\epsilon + \nabla \cdot \mathbf{q}^\epsilon \leq \epsilon \Delta \eta^\epsilon. \quad (4.3.37)$$

We proceed by multiplying (4.3.35) by  $w \in C_c^\infty(\Omega \times (0, \infty))$ ,  $w \geq 0$ , integrate and use (4.3.36):

$$\begin{aligned} \int_0^\infty \int_\Omega \eta \left( \partial_t w + \frac{\partial f}{\partial \phi} \cdot \nabla w \right) dx dt &= \int_0^\infty \int_\Omega \epsilon (\mathbf{H}_x \phi \mathbf{H}_{\nabla \phi} \eta) : \mathbf{H}_x \phi w - \epsilon \eta \Delta w dx dt \\ &\geq - \int_0^\infty \int_\Omega \epsilon \eta \Delta w dx dt. \end{aligned} \quad (4.3.38)$$

Applying the dominated convergence theorem we get

$$\int_0^\infty \int_\Omega \eta \left( \partial_t w + \frac{\partial f}{\partial \phi} \cdot \nabla w \right) dx dt \geq 0. \quad (4.3.39)$$

□

**Corollary 4.3.17.** (*Decay of variation entropy*) Let  $\Omega$  be a periodic domain or have no-inflow boundaries and let  $\phi$  be a smooth function. The variation entropy of a vanishing viscosity solution decays in time:

$$\int_\Omega \eta(\nabla \phi(x, t)) d\Omega \leq \int_\Omega \eta(\nabla \phi_0(x)) d\Omega, \quad \text{for all } t > 0. \quad (4.3.40)$$

*Proof.* Integration of (4.3.30) yields:

$$\begin{aligned} \frac{d}{dt} \int_\Omega \eta(\nabla \phi(x, t)) d\Omega &\leq 0 \quad \Rightarrow \\ \int_\Omega \eta(\nabla \phi(x, t)) d\Omega &\leq \int_\Omega \eta(\nabla \phi_0(x)) d\Omega, \quad \text{for all } t > 0. \end{aligned} \quad (4.3.41)$$

□

#### 4.3.4 Augmented conservation laws

We consider ‘augmented conservation laws’, i.e. PDEs with a convective, diffusive and source component of the form:

$$\partial_t \phi + \nabla \cdot \mathbf{f} = s, \quad (\mathbf{x}, t) \in \Omega \times \mathcal{I}, \quad (4.3.42)$$

with flux  $\mathbf{f} = \mathbf{f}(\phi, \nabla \phi)$  and source term  $s = s(\phi)$ . Here we assume that the matrix  $\partial \mathbf{f} / \partial \nabla \phi$  is negative semi-definite. Let  $(\eta, \mathbf{q})$  be a variation entropy/variation

<sup>1</sup> Note that its classical entropy counterpart is  $\nabla \phi^\epsilon \cdot \nabla \phi^\epsilon \partial^2 \eta^\epsilon / \partial (\phi^\epsilon)^2$ .

entropy flux pair. A straightforward computation shows that using the homogeneity property (4.3.10), the evolution equation takes the form:

$$\partial_t \eta + \nabla \cdot \mathbf{q} = \mathcal{D} + \mathcal{S}, \quad (4.3.43)$$

where the flux  $\mathbf{q}$  and the non-conservative terms  $\mathcal{D}$  and  $\mathcal{S}$  are respectively given by:

$$\mathbf{q} = \eta \frac{\partial \mathbf{f}}{\partial \phi} + \frac{\partial \mathbf{f}}{\partial \nabla \phi} \nabla \eta, \quad (4.3.44a)$$

$$\mathcal{D} = (\mathbf{H}_x \phi \mathbf{H}_{\nabla \phi} \eta) : \left( \frac{\partial \mathbf{f}}{\partial \nabla \phi} \mathbf{H}_x \phi \right), \quad (4.3.44b)$$

$$\mathcal{S} = \frac{\partial s}{\partial \phi} \eta. \quad (4.3.44c)$$

We emphasize that this form closely resembles an augmented conservation law with convection, diffusion and reaction components. The reaction term is the only term that can create variation entropy (remark that it vanishes for a constant source  $s$ ). Note that negative eigenvalues of the diffusion matrix are an essential requirement for well-posedness. Positive eigenvalues create variation entropy which leads to a blow-up of the solutions. In the next lemma we show that the term  $\mathcal{D}$  on the right-hand side of (4.3.43) destroys variation entropy.

**Lemma 4.3.18.** *(Negativity of  $\mathcal{D}$ ) The term  $\mathcal{D}$  on the right-hand side of (4.3.43) contributes to dissipation of the variation entropy, i.e. it takes negative values only.*

*Proof.* The convexity of the function  $\eta = \eta(\nabla \phi)$  implies that there exists a real nonsingular matrix  $\mathbf{M}$  such that

$$\mathbf{H}_{\nabla \phi} \eta = \mathbf{M}^T \mathbf{M}. \quad (4.3.45)$$

Substitution into the expression for  $\mathcal{D}$  gives

$$\mathcal{D} = (\mathbf{H}_x \phi \mathbf{H}_{\nabla \phi} \eta) : \left( \frac{\partial \mathbf{f}}{\partial \nabla \phi} \mathbf{H}_x \phi \right) = \left( \mathbf{M} \mathbf{H}_x \phi \frac{\partial \mathbf{f}}{\partial \nabla \phi} \right) : (\mathbf{M} \mathbf{H}_x \phi) \leq 0, \quad (4.3.46)$$

where the inequality follows from the fact that the diffusivity matrix is negative semi-definite.  $\square$

#### 4.4 SELECTION OF THE VARIATION ENTROPY

Here we provide some examples of variation entropies and discuss objectivity of variation entropies. We present regularization of the 2-norm in detail. For the sake of simplicity we assume smooth solutions in this section.

##### 4.4.1 Some examples

Some examples of variation entropies are:

- The linear function  $\eta(\nabla \phi) = \mathbf{a} \cdot \nabla \phi$ , with  $\mathbf{a} \in \mathbb{R}^d$ .

- The function  $\eta(\nabla\phi) = \|\nabla\phi\|_{\mathbf{A}}$  defined as  $\|\nabla\phi\|_{\mathbf{A}}^2 := \nabla\phi^T \mathbf{A} \nabla\phi$ , with  $\mathbf{A} \in \mathbb{R}^{d \times d}$  a symmetric positive-semidefinite matrix.
- The standard  $p$ -norm, i.e.  $\eta(\nabla\phi) = \|\nabla\phi\|_p$  with  $p \geq 1$ .

By Corollary 4.3.9, any superposition of the previous variation entropies is also a variation entropy.

**Proposition 4.4.1.** *The alternative form of the variation entropy is*

$$\hat{\eta}(r, \theta) = rF(\theta) \quad (4.4.1)$$

where  $F = F(\theta)$  is given by

for  $d = 2$ :

$$F(\theta) = \begin{cases} a_1 \cos \theta + a_2 \sin \theta & \text{for } \eta(\nabla\phi) = \mathbf{a} \cdot \nabla\phi \\ (a_{11} \cos^2 \theta + 2a_{21} \cos \theta \sin \theta + a_{22} \sin^2 \theta)^{1/2} & \text{for } \eta(\nabla\phi) = \|\nabla\phi\|_{\mathbf{A}} \\ (|\cos \theta|^p + |\sin \theta|^p)^{1/p} & \text{for } \eta(\nabla\phi) = \|\nabla\phi\|_p \end{cases} \quad (4.4.2)$$

for  $d = 3$ :

$$F(\theta, \varphi) = \begin{cases} a_1 \cos \theta \sin \varphi + a_2 \sin \theta \sin \varphi + a_3 \cos \varphi & \text{for } \eta(\nabla\phi) = \mathbf{a} \cdot \nabla\phi \\ (a_{11} \cos^2 \theta \sin^2 \varphi + a_{22} \sin^2 \theta \sin^2 \varphi + a_{33} \cos^2 \varphi \\ + 2a_{21} \cos \theta \sin \theta \sin^2 \varphi + 2a_{31} \cos \theta \sin \varphi \cos \varphi \\ + 2a_{32} \sin \theta \sin \varphi \cos \varphi)^{1/2} & \text{for } \eta(\nabla\phi) = \|\nabla\phi\|_{\mathbf{A}} \\ (|\cos \theta \sin \varphi|^p + |\sin \theta \sin \varphi|^p + |\cos \varphi|^p)^{1/p} & \text{for } \eta(\nabla\phi) = \|\nabla\phi\|_p. \end{cases}$$

*Proof.* We present the proof only for  $\eta(\nabla\phi) = \|\nabla\phi\|_{\mathbf{A}}$  with  $d = 2$ , the other forms follow similarly. A direct calculation provides

$$\nabla\phi^T \mathbf{A} \nabla\phi = a_{11} \partial_x \phi^2 + 2a_{21} \partial_x \phi \partial_y \phi + a_{22} \partial_y \phi^2. \quad (4.4.3)$$

Next, we can trivially write

$$\|\nabla\phi\|_{\mathbf{A}} = \|\nabla\phi\|_2 \left( a_{11} \frac{\partial_x \phi^2}{\|\nabla\phi\|_2^2} + 2a_{21} \frac{\partial_x \phi \partial_y \phi}{\|\nabla\phi\|_2^2} + a_{22} \frac{\partial_y \phi^2}{\|\nabla\phi\|_2^2} \right)^{1/2}. \quad (4.4.4)$$

Switching to polar coordinates with radial distance  $r$  and angle  $\theta$ , the right-hand side of (4.4.4) takes the form

$$rF(\theta) = r (a_{11} \cos^2 \theta + 2a_{21} \cos \theta \sin \theta + a_{22} \sin^2 \theta)^{1/2}. \quad (4.4.5)$$

Note that fulfilling the convexity condition requires the positive semi-definiteness of the matrix  $\mathbf{A}$ :

$$F(\theta) + F''(\theta) = \frac{\det \mathbf{A}}{F(\theta)^3} \geq 0. \quad (4.4.6)$$

Here  $\det \mathbf{A}$  denotes the determinant of matrix  $\mathbf{A}$ . □

The linear variation entropy fulfills the convexity condition in spherical coordinates with equality, for  $d = 2$  this reads

$$F(\theta) + F''(\theta) = 0. \quad (4.4.7)$$

Thus there is no space for the variation entropy to decrease. It is known that the entropy does not remain constant at shocks. This makes the linear variation entropy a non-suitable mechanism to deal with shock waves in a numerical simulation.

For the 1-norm variation entropy, i.e.  $\eta = \|\nabla\phi\|_1$ , we have

$$F(\theta) + F''(\theta) = 0, \quad \theta \neq m\pi/2 \quad m \in \mathbb{Z}. \quad (4.4.8)$$

We remark that the 1-norm is not differentiable along the axis. Thus also for the 1-norm there is also no space for the variation entropy to decrease. For  $p > 1$  the convexity condition is not an equality which makes it suitable for shocks.

Next we consider the quadratic form. Proposition 4.4.1 reveals that the matrix  $\mathbf{A} \in \mathbb{R}^{d \times d}$  in the quadratic form may depend on  $\theta$  (the spherical coordinate angles). We explicitly state the evolution equation of  $\eta = \|\|\nabla\phi\|\|_{\mathbf{A}}$  for an augmented conservation law. Note that the derivatives take the form

$$\frac{\partial\eta}{\partial\nabla\phi} = \frac{1}{\eta} \mathbf{A} \nabla\phi, \quad (4.4.9a)$$

$$(\mathbf{H}_{\nabla\phi}\eta) = \frac{1}{\eta} \mathbf{A} - \frac{1}{\eta^3} \mathbf{A} \nabla\phi (\mathbf{A} \nabla\phi)^T. \quad (4.4.9b)$$

Substitution of (4.4.9) into (4.3.43)-(4.3.44) reveals that the variation entropy evolves as:

$$\partial_t \eta + \nabla \cdot \mathbf{q} = \mathcal{D} + \mathcal{S}, \quad (4.4.10)$$

where the flux  $\mathbf{q}$  and the non-conservative terms  $\mathcal{D}$  and  $\mathcal{S}$  are respectively given by:

$$\mathbf{q} = \eta \frac{\partial f}{\partial \phi} + \left( \frac{\partial f}{\partial \nabla \phi} \mathbf{H}_x \phi \right) \frac{\nabla \phi^T \mathbf{A}}{\eta}, \quad (4.4.11a)$$

$$\mathcal{D} = \left( \mathbf{H}_x \phi \left( \frac{\mathbf{A}}{\eta} - \frac{1}{\eta^3} \mathbf{A} \nabla \phi : \mathbf{A} \nabla \phi \right) \right) : \left( \frac{\partial f}{\partial \nabla \phi} \mathbf{H}_x \phi \right), \quad (4.4.11b)$$

$$\mathcal{S} = \frac{\partial s}{\partial \phi} \eta. \quad (4.4.11c)$$

#### 4.4.2 Objectivity

A reasonable demand on the continuous level is to ask for objectivity (frame-invariance) of the variation entropy. Here we solely focus on rotation invariance, as invariance by translation is immediate. Thus we wish to identify those variation entropies which are not affected by a rotation of the coordinate system. We denote with  $\mathbf{x}$  the original spatial coordinates and  $\tilde{\mathbf{x}} = \mathbf{R}\mathbf{x}$  the rotated coordinate system with rotation matrix  $\mathbf{R}$  (i.e.  $\mathbf{R}^T = \mathbf{R}^{-1}$ ).

Let us first consider variation entropies which depend on spatial coordinates solely via  $\nabla\phi$ . This means that possible coefficients, like in  $\mathbf{a}$  and  $\mathbf{A}$  appearing in

$\eta = \mathbf{a} \cdot \nabla \phi$  and  $\eta = \|\|\nabla \phi\|\|_{\mathbf{A}}$  respectively, are constant with respect to the spatial coordinates. In this case rotation invariance may be written as:

$$\eta(\nabla_{\tilde{\mathbf{x}}}\phi) = \eta(\nabla_{\mathbf{x}}\phi), \quad (4.4.12)$$

or equivalently:

$$\eta(\mathbf{R}\nabla_{\tilde{\mathbf{x}}}\phi) = \eta(\nabla_{\tilde{\mathbf{x}}}\phi), \quad \eta(\mathbf{R}^T\nabla_{\mathbf{x}}\phi) = \eta(\nabla_{\mathbf{x}}\phi). \quad (4.4.13)$$

The subscript refers to the corresponding coordinate system.

**Theorem 4.4.2.** (*Objectivity*) *The only objective variation entropy with constant coefficients is the total variation measured in the 2-norm,  $\eta = \|\nabla \phi\|_2$  (up to multiplication by a constant).*

*Proof.* We use the alternative form of Lemma 4.3.10. Let the angles of the rotation matrix  $\mathbf{R}$  be  $\boldsymbol{\varrho}$  and let polar angles of  $\nabla \phi$  denote  $\boldsymbol{\theta}$ . A direct computation results in

$$\eta(\mathbf{R}\nabla \phi) = rF(\boldsymbol{\theta} + \boldsymbol{\varrho}) = \hat{\eta}(r, \boldsymbol{\theta} + \boldsymbol{\varrho}). \quad (4.4.14)$$

Demanding  $\eta(\mathbf{R}\nabla \phi) = \eta(\nabla \phi)$  provides that  $F(\boldsymbol{\theta} + \boldsymbol{\varrho}) = F(\boldsymbol{\theta})$  for all angles  $\boldsymbol{\varrho}$ , i.e.  $F$  is a constant.  $\square$

Thus the only objective  $p$ -norm with constant coefficients is the norm with  $p = 2$ . In particular, we wish to emphasize that *the 1-norm is not objective*. This makes it unsuitable for usage in non-Cartesian grid computations.

Next we proceed with the case in which the coefficients may depend on spatial coordinates. We first consider the linear variation entropy. Objectivity results when

$$\mathbf{a}(\mathbf{x}) \cdot \nabla_{\mathbf{x}}\phi = \mathbf{a}(\tilde{\mathbf{x}}) \cdot \nabla_{\tilde{\mathbf{x}}}\phi. \quad (4.4.15)$$

By applying the chain rule we get

$$\mathbf{a}(\mathbf{x}) \cdot \nabla_{\mathbf{x}}\phi = \mathbf{a}(\mathbf{R}\mathbf{x}) \cdot \mathbf{R}^T \nabla_{\mathbf{x}}\phi. \quad (4.4.16)$$

Thus objectivity follows when the coefficient vector satisfies:

$$\mathbf{a}(\mathbf{R}\mathbf{x}) = \mathbf{R}\mathbf{a}(\mathbf{x}). \quad (4.4.17)$$

Note that when  $\mathbf{a}$  represents a convection velocity, a rotation of coordinate system naturally implies the same rotation of the convective velocity.

Consider now the variation entropy  $\eta = \|\|\nabla \phi\|\|_{\mathbf{A}}$  with  $\mathbf{A}$  a symmetric positive semi-definite matrix. This variation entropy is objective when

$$\|\|\nabla_{\mathbf{x}}\phi\|\|_{\mathbf{A}(\mathbf{x})} = \|\|\nabla_{\tilde{\mathbf{x}}}\phi\|\|_{\mathbf{A}(\tilde{\mathbf{x}})}. \quad (4.4.18)$$

A similar argument leads to the constraint:

$$\mathbf{A}(\mathbf{R}\mathbf{x}) = \mathbf{R}\mathbf{A}(\mathbf{x})\mathbf{R}^T. \quad (4.4.19)$$

Remark that  $\mathbf{A} = \mathbf{I}$  is included in this case.

An overview of the variation entropy results is presented in Figure 4.2.

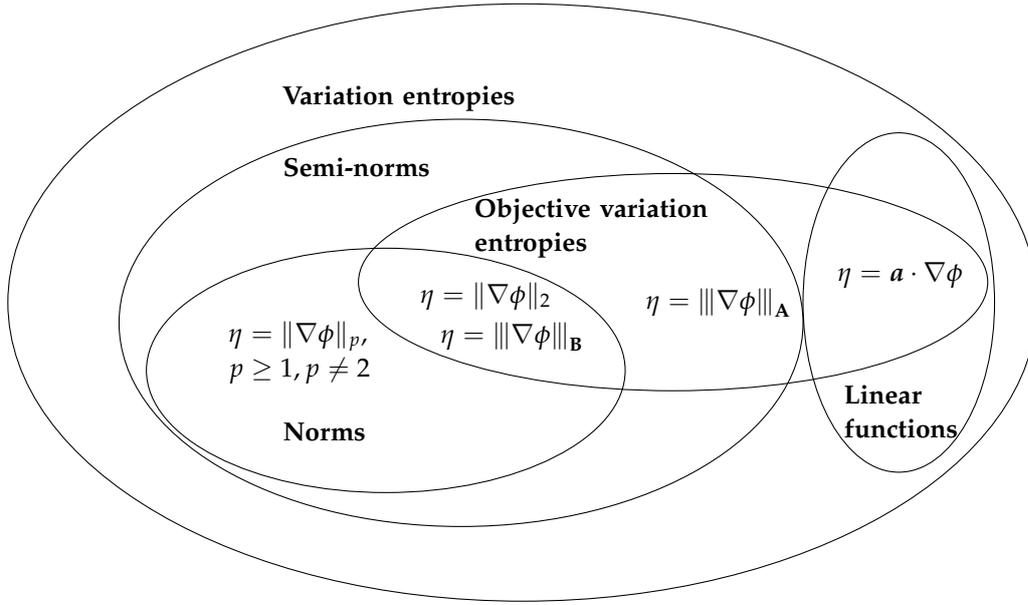


Figure 4.2: Overview of objective variation entropy results. The variation entropy  $\|\nabla\phi\|_p$  is not objective unless  $p = 2$ . The matrices  $\mathbf{A}$  and  $\mathbf{B}$  are positive semi-definite and positive definite respectively. The vector and matrices in  $\eta = \mathbf{a} \cdot \nabla\phi$ ,  $\eta = \|\nabla\phi\|_{\mathbf{A}}$  and  $\eta = \|\nabla\phi\|_{\mathbf{B}}$  are chosen according to (4.4.17) and (4.4.19), and as such we classify these variation entropies as objective.

#### 4.4.3 Regularization of 2-norm variation entropy

Due to its importance, we discuss the 2-norm variation entropy here separately. In particular we consider a regularized version in order to allow evaluation everywhere. Thus we study the case where the variation entropy function is the regularized absolute value operator  $\|\cdot\|_{\varepsilon,2} : \mathbb{R}^d \rightarrow \mathbb{R}_+$  which is defined for  $\mathbf{b} \in \mathbb{R}^d, \varepsilon > 0$  as:

$$\|\mathbf{b}\|_{\varepsilon,2}^2 := \mathbf{b} \cdot \mathbf{b} + \varepsilon^2. \quad (4.4.20)$$

Notice that

$$\|\mathbf{b}\|_2 \leq \|\mathbf{b}\|_{\varepsilon,2} \leq \|\mathbf{b}\|_2 + \varepsilon, \quad (4.4.21)$$

as displayed in Figure 4.3.

The regularized absolute value has the derivatives:

$$\partial_{\mathbf{b}} \|\mathbf{b}\|_{\varepsilon,2} = \frac{\mathbf{b}}{\|\mathbf{b}\|_{\varepsilon,2}}, \quad (4.4.22a)$$

$$\partial_{\mathbf{b}}^2 \|\mathbf{b}\|_{\varepsilon,2} = \left( \mathbf{I} - \frac{\mathbf{b}\mathbf{b}^T}{\|\mathbf{b}\|_{\varepsilon,2}^2} \right) \frac{1}{\|\mathbf{b}\|_{\varepsilon,2}}, \quad (4.4.22b)$$

which exist everywhere. The homogeneity constraint (4.3.10) is violated by a term that scales with  $\varepsilon^2$ :

$$\mathbf{b} \cdot \partial_{\mathbf{b}} \|\mathbf{b}\|_{\varepsilon,2} - \|\mathbf{b}\|_{\varepsilon,2} = \frac{\varepsilon^2}{\|\mathbf{b}\|_{\varepsilon,2}}. \quad (4.4.23)$$

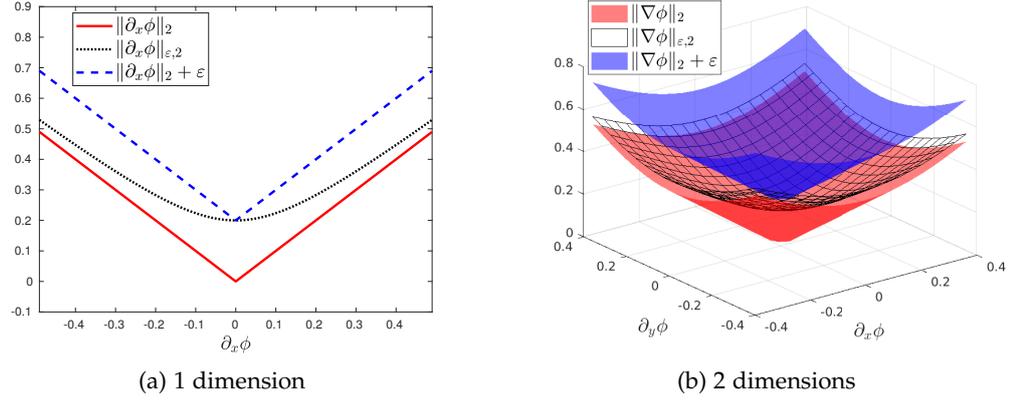


Figure 4.3: Plot of the regularized norm in (a) 1-dimension and (b) 2-dimensions. Here  $\varepsilon = 0.2$ .

Also the term that appears in  $\mathcal{A}$  scales with  $\varepsilon^2$ :

$$\partial_{\mathbf{b}}^2 \|\mathbf{b}\|_{\varepsilon,2} \mathbf{b} = \varepsilon^2 \frac{\mathbf{b}}{\|\mathbf{b}\|_{\varepsilon,2}^3}. \quad (4.4.24)$$

**Corollary 4.4.3.** (Evolution equation of a regularized variation entropy) The regularized variation  $\eta = \eta_\varepsilon = \|\nabla\phi\|_{\varepsilon,2}$  satisfies the evolution equation:

$$\partial_t \eta_\varepsilon + \nabla \cdot \mathbf{q}_\varepsilon = \mathcal{D}_\varepsilon + \mathcal{S}_\varepsilon + \mathcal{R}_\varepsilon, \quad (4.4.25)$$

where the flux  $\mathbf{q}_\varepsilon$ , the non-conservative terms  $\mathcal{A}_\varepsilon$  and  $\mathcal{D}_\varepsilon$  and the source term  $\mathcal{S}_\varepsilon$  are respectively defined as:

$$\mathbf{q}_\varepsilon = \frac{\partial f}{\partial \phi} \eta_\varepsilon + \frac{\partial f}{\partial \nabla \phi} \nabla \eta_\varepsilon, \quad (4.4.26a)$$

$$\mathcal{D}_\varepsilon = \frac{1}{\eta_\varepsilon^3} \left( (\|\nabla\phi\|_2^2 \mathbf{I} - \nabla\phi \nabla\phi^T) \mathbf{H}\mathbf{x}\phi \right) : \left( \frac{\partial f}{\partial \nabla \phi} \mathbf{H}\mathbf{x}\phi \right), \quad (4.4.26b)$$

$$\mathcal{S}_\varepsilon = \frac{\partial s}{\partial \phi} \eta_\varepsilon, \quad (4.4.26c)$$

$$\mathcal{R}_\varepsilon = \frac{\varepsilon^2}{\eta_\varepsilon} \left( \nabla \cdot \left( \frac{\partial f}{\partial \phi} \right) + \frac{1}{\eta_\varepsilon^2} \mathbf{H}\mathbf{x}\phi : \left( \frac{\partial f}{\partial \nabla \phi} \mathbf{H}\mathbf{x}\phi \right) - \frac{\partial s}{\partial \phi} \right). \quad (4.4.26d)$$

*Proof.* A direct substitution of  $\eta = \eta_\varepsilon$  into (4.3.43)-(4.3.44) using (4.4.20)-(4.4.24) yields

$$\partial_t \eta_\varepsilon + \nabla \cdot \mathbf{q} = \mathcal{A} + \mathcal{D} + \mathcal{S}, \quad (4.4.27)$$

where the flux  $\mathbf{q}$ , the non-conservative terms  $\mathcal{A}$  and  $\mathcal{D}$  and the source term  $\mathcal{S}$  are respectively given by:

$$\mathbf{q} = \frac{\partial f}{\partial \phi} \left( \eta_\varepsilon - \frac{\varepsilon^2}{\eta_\varepsilon} \right) + \frac{\partial f}{\partial \nabla \phi} \nabla \eta, \quad (4.4.28a)$$

$$\mathcal{A} = \frac{\varepsilon^2}{\eta_\varepsilon^2} \frac{\nabla \phi}{\eta_\varepsilon} \cdot \left( \mathbf{H}_x \phi \frac{\partial f}{\partial \phi} \right), \quad (4.4.28b)$$

$$\mathcal{D} = \frac{1}{\eta_\varepsilon} \left( \left( \mathbf{I} - \frac{\nabla \phi \nabla \phi^T}{\eta_\varepsilon^2} \right) \mathbf{H}_x \phi \right) : \left( \frac{\partial f}{\partial \nabla \phi} \mathbf{H}_x \phi \right), \quad (4.4.28c)$$

$$\mathcal{S} = \frac{\partial s}{\partial \phi} \left( \eta_\varepsilon - \frac{\varepsilon^2}{\eta_\varepsilon} \right). \quad (4.4.28d)$$

The divergence of the flux writes as

$$\begin{aligned} \nabla \cdot \mathbf{q} &= \nabla \cdot \left( \frac{\partial f}{\partial \phi} \left( \eta_\varepsilon - \frac{\varepsilon^2}{\eta_\varepsilon} \right) + \frac{\partial f}{\partial \nabla \phi} \nabla \eta \right), \\ &= \nabla \cdot \left( \frac{\partial f}{\partial \phi} \eta_\varepsilon + \frac{\partial f}{\partial \nabla \phi} \nabla \eta \right) + \frac{\varepsilon^2}{\eta_\varepsilon^3} \nabla \phi \cdot \left( \mathbf{H}_x \phi \frac{\partial f}{\partial \phi} \right) - \frac{\varepsilon^2}{\eta_\varepsilon} \nabla \cdot \left( \frac{\partial f}{\partial \phi} \right) \\ &= \nabla \cdot \mathbf{q}_\varepsilon + \mathcal{A} - \frac{\varepsilon^2}{\eta_\varepsilon} \nabla \cdot \left( \frac{\partial f}{\partial \phi} \right). \end{aligned} \quad (4.4.29)$$

The diffusion term  $\mathcal{D}$  can be written as

$$\begin{aligned} \mathcal{D} &= \left( \frac{1}{\eta_\varepsilon} \left( \mathbf{I} - \frac{\nabla \phi \nabla \phi^T}{\eta_\varepsilon^2} \right) \mathbf{H}_x \phi \right) : \left( \frac{\partial f}{\partial \nabla \phi} \mathbf{H}_x \phi \right), \\ &= \frac{1}{\eta_\varepsilon^3} \left( \left( \|\nabla \phi\|_2^2 \mathbf{I} - \nabla \phi \nabla \phi^T \right) \mathbf{H}_x \phi \right) : \left( \frac{\partial f}{\partial \nabla \phi} \mathbf{H}_x \phi \right) + \frac{\varepsilon^2}{\eta_\varepsilon^3} \mathbf{H}_x \phi : \left( \frac{\partial f}{\partial \nabla \phi} \mathbf{H}_x \phi \right) \\ &= \mathcal{D}_\varepsilon + \frac{\varepsilon^2}{\eta_\varepsilon^3} \mathbf{H}_x \phi : \left( \frac{\partial f}{\partial \nabla \phi} \mathbf{H}_x \phi \right). \end{aligned} \quad (4.4.30)$$

Substitution of (4.4.29)-(4.4.30) into (4.4.27)-(4.4.28) proves the claim.  $\square$

Consider the one-dimensional case of (4.4.25)-(4.4.26), i.e.

$$\frac{\partial}{\partial t} \|\partial_x \phi\|_{\varepsilon,2} + \partial_x \left( \frac{\partial f}{\partial \phi} \|\partial_x \phi\|_{\varepsilon,2} + \frac{\partial f}{\partial (\partial_x \phi)} \partial_x (\|\partial_x \phi\|_{\varepsilon,2}) \right) = \frac{\partial s}{\partial \phi} \|\partial_x \phi\|_{\varepsilon,2} + \mathcal{R}_\varepsilon. \quad (4.4.31)$$

We focus on the last term on the right-hand side. It can be written as

$$\mathcal{R}_\varepsilon = \left( \partial_x \left( \frac{\partial f}{\partial \phi} \right) - \frac{\partial s}{\partial \phi} \right) g_1^\varepsilon(\partial_x \phi) + \left( (\partial_{xx} \phi)^2 \frac{\partial f}{\partial \partial_x \phi} \right) g_2^\varepsilon(\partial_x \phi), \quad (4.4.32)$$

where the functions  $g_1^\varepsilon$  and  $g_2^\varepsilon$  are defined as

$$g_1^\varepsilon(\partial_x \phi) = \frac{\varepsilon^2}{\|\partial_x \phi\|_{\varepsilon,2}}, \quad (4.4.33a)$$

$$g_2^\varepsilon(\partial_x \phi) = \frac{\varepsilon^2}{\|\partial_x \phi\|_{\varepsilon,2}^3}. \quad (4.4.33b)$$

In Figure 4.4 we plot  $g_1^\varepsilon$  and  $g_2^\varepsilon$  for several values of  $\varepsilon$ .

The function  $g_1^\varepsilon$  vanishes in the limit  $\varepsilon \downarrow 0$  and its value at the origin is  $g_1^\varepsilon(0) = \varepsilon$ .

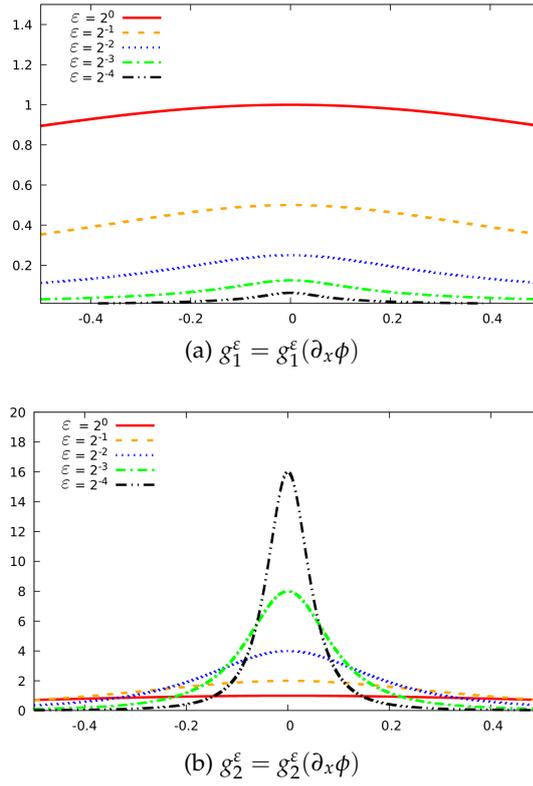


Figure 4.4: Plot of the functions  $g_1^\varepsilon$  (a) and  $g_2^\varepsilon$  (b). On the horizontal axis  $\partial_x \phi$  varies.

On the other hand,  $g_2^\varepsilon$  behaves as a (scaled) delta distribution centered at the origin. The value at the origin is  $g_2^\varepsilon(0) = \varepsilon^{-1}$  and the area under the profile  $g_2^\varepsilon$  is 2 (which is independent of  $\varepsilon$ ). Thus the regularization focuses the diffusion contribution at points where  $\partial_x \phi$  approaches zero, i.e. the extrema of  $\phi$ . For  $\varepsilon \downarrow 0$  we conclude that in one dimension variation entropy can either be produced by diffusion at local extrema or by the source term.

In the multi-dimensional case ( $d > 1$ ) the variation entropy diffusion  $\mathcal{D}$  does not vanish. This is a clear separation of the 1-dimensional case and the multi-dimensional case. We explicitly state the temporal evolution of the non-regularized variation. The limit of  $\varepsilon \downarrow 0$  in (4.4.25)-(4.4.26) yields:

$$\partial_t \|\nabla \phi\|_2 + \nabla \cdot \mathbf{q} = \mathcal{D} + \mathcal{S}, \quad (4.4.34)$$

where the flux  $\mathbf{q}$ , the non-conservative terms  $\mathcal{A}$  and  $\mathcal{D}$  and the source term  $\mathcal{S}$  are respectively defined as:

$$\mathbf{q} = \frac{\partial \mathbf{f}}{\partial \phi} \|\nabla \phi\|_2 + \frac{\partial \mathbf{f}}{\partial \nabla \phi} \nabla \|\nabla \phi\|_2, \quad (4.4.35a)$$

$$\mathcal{D} = \frac{1}{\|\nabla \phi\|_2} \left( \left( \mathbf{I} - \frac{\nabla \phi \nabla \phi^T}{\|\nabla \phi\|_2^2} \right) \mathbf{H}_x \phi \right) : \left( \frac{\partial \mathbf{f}}{\partial \nabla \phi} \mathbf{H}_x \phi \right), \quad (4.4.35b)$$

$$\mathcal{S} = \frac{\partial s}{\partial \phi} \|\nabla \phi\|_2, \quad (4.4.35c)$$

which is not defined for  $\nabla \phi = \mathbf{0}$ .

**Remark 4.4.4.** We emphasize that the well-known total variation diminishing (TVD) constraint:

$$\frac{d}{dt} \int_{\Omega} \|\nabla \phi(x, t)\|_2 d\Omega \leq 0 \Rightarrow \int_{\Omega} \|\nabla \phi(x, t)\|_2 d\Omega \leq \int_{\Omega} \|\nabla \phi_0(x)\|_2 d\Omega, \quad \text{for } t > 0 \quad (4.4.36)$$

is special case of decay of variation entropy (substitute  $\eta = \|\nabla \phi\|_2$  into Corollary 4.3.17).

We wish to indicate the effect of regularization on the total variation. To that purpose we compute the total variation and its regularized counterpart for two functions: (i) a linear approximation of the Heaviside function

$$\phi_{L,\mathcal{E}}(x) = \begin{cases} 0 & \text{if } x \leq -\mathcal{E} \\ \frac{1}{2} \left(1 + \frac{x}{\mathcal{E}}\right) & \text{if } x < |\mathcal{E}|, \\ 1 & \text{if } x \geq \mathcal{E} \end{cases} \quad (4.4.37)$$

and (ii) a smoothed Heaviside function:

$$\phi_{S,\mathcal{E}}(x) = \begin{cases} 0 & \text{if } x \leq -\mathcal{E} \\ \frac{1}{2} \left(1 + \frac{x}{\mathcal{E}} + \frac{1}{\pi} \sin\left(\frac{x\pi}{\mathcal{E}}\right)\right) & \text{if } x < |\mathcal{E}|. \\ 1 & \text{if } x \geq \mathcal{E} \end{cases} \quad (4.4.38)$$

This smoothed Heaviside is often used for levelset computations, see e.g., [3, 5]. We plot the approximated Heaviside functions and the (regularized) 2-norms of their derivatives in Figure 4.5.

The total variation and its regularizations are defined as:

$$TV(\phi_{\mathcal{E}}) = \int_{-\mathcal{E}}^{\mathcal{E}} \|\partial_x \phi_{\mathcal{E}}(x)\|_2 dx, \quad (4.4.39a)$$

$$TV_{\varepsilon}(\phi_{\mathcal{E}}) = \int_{-\mathcal{E}}^{\mathcal{E}} \|\partial_x \phi_{\mathcal{E}}(x)\|_{\varepsilon,2} dx, \quad (4.4.39b)$$

$$\overline{TV}_{\varepsilon}(\phi_{\mathcal{E}}) = \int_{-\mathcal{E}}^{\mathcal{E}} \|\partial_x \phi_{\mathcal{E}}(x)\|_2 + \varepsilon dx. \quad (4.4.39c)$$

The total variation (4.4.39a) is independent of the regularization parameter  $\mathcal{E}$  since the functions (4.4.37)-(4.4.38) monotonically increase from 0 to 1 and thus we have:

$$TV(\phi_{L,\mathcal{E}}) = TV(\phi_{S,\mathcal{E}}) = 1. \quad (4.4.40)$$

The regularized total variation (4.4.39b) for the linear approximation (4.4.37) equals:

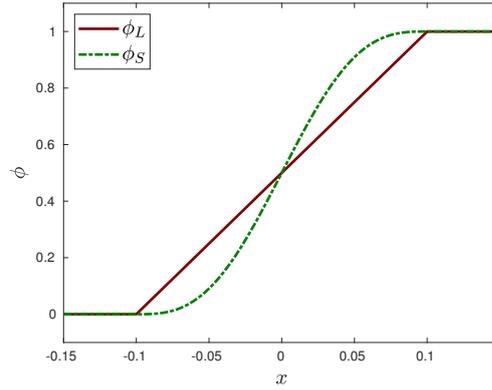
$$TV_{\varepsilon}(\phi_{L,\mathcal{E}}) = \sqrt{1 + 4\mathcal{E}^2\varepsilon^2}. \quad (4.4.41)$$

Generally a small value is chosen for the parameters  $\mathcal{E}$  and  $\varepsilon$  which indicates that for the linear approximation the regularized total variation is nearly indistinguishable from its exact counterpart. In Figure 4.6 we show the (regularized) total variation for the both approximations.

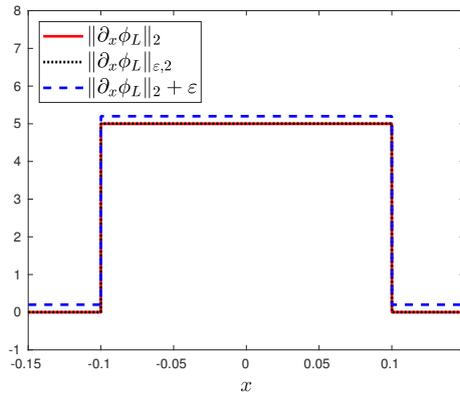
We see that also for the smooth Heaviside small regularization parameters  $\mathcal{E}$  and  $\varepsilon$  indicate a close resemblance with the exact value. Note that the estimate

$$TV(\phi) \leq TV_{\varepsilon}(\phi) \leq \overline{TV}_{\varepsilon}(\phi), \quad (4.4.42)$$

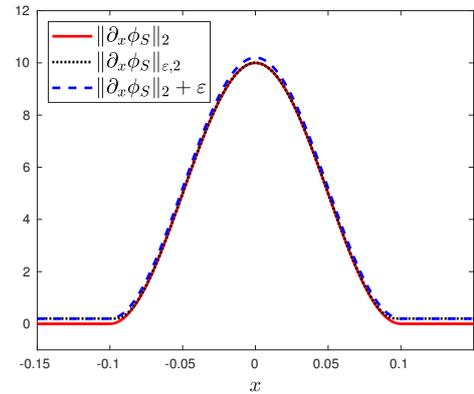
which is a direct consequence of (4.4.21), is confirmed by Figure 4.6.



(a)  $\phi_\varepsilon = \phi_\varepsilon(x)$



(b)  $\partial_x \phi_{L,\varepsilon} = \partial_x \phi_{L,\varepsilon}(x)$



(c)  $\partial_x \phi_{S,\varepsilon} = \partial_x \phi_{S,\varepsilon}(x)$

Figure 4.5: Plot of (a) the smooth Heaviside and (b) the (regularized) 2-norms of its derivative. Here  $\varepsilon = 0.2$  and  $\mathcal{E} = 0.1$ .

#### 4.5 CONCLUSION AND DISCUSSION

The purpose of this chapter is to answer the two questions:

- *How can we construct a local continuous generalization of the TVD stability condition?*
- *Is there a connection between entropy solutions and the TVD property?*

To accomplish this we have developed the new stability concept *variation entropy* for nonlinear conservation laws. The core idea is to develop an entropy concept based on the gradient of the solution of a conservation law instead of on the solution itself. This may be a more natural and suitable approach when dealing with shock waves, which are characterized by their large gradients.

Variation entropy solutions are formulated in the continuous setting and as such employing this concept eliminates the restrictions of the TVD property. Variation entropies are homogeneous convex functions. As a result, all semi-norms are suitable variation entropies. A particular choice is the standard 2-norm variation entropy which can be viewed as local continuous evolution equation of the TVD property. This sheds light on the TVD property from the perspective of entropy solutions. In

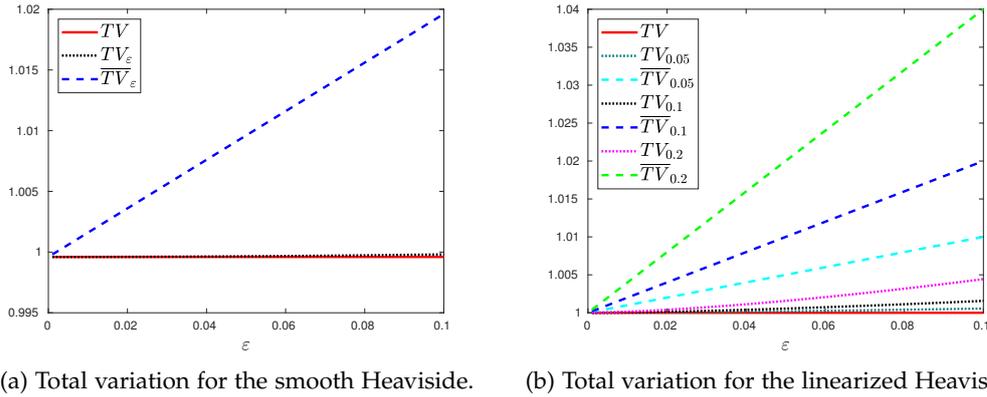


Figure 4.6: Plot of total variation and its regularization for (a) the linearized Heaviside ( $\mathcal{E} = 0.1$ ) and (b) the smooth Heaviside for  $\mathcal{E} = 0.05, 0.1, 0.02$ .

other words, *variation entropy solutions are the link between classical entropy solutions and the TVD property.*

This work opens several doors for future research. The class of variation entropy solutions can be used to design new numerical methods. Numerical solutions in this class cannot create variation entropy and thus satisfy a local TVD stability property. This precludes spurious oscillations which is crucial near shock waves.

A particular open problem is the construction of discontinuities capturing mechanisms in finite element methods. A discontinuity capturing operator could be directly based on the variation entropy condition. It could add diffusion where the variation entropy condition is harmed and render inoperative elsewhere. We believe that a natural way to arrive at a discontinuity capturing method is via the variational multiscale (VMS) method. In Chapter 5 we explore this approach and use the multiscale split projector to demand the variation entropy condition. This naturally results in a consistent VMS method equipped with a penalty term that adds diffusion when the variation entropy condition is violated.



## 4.A THE 3-DIMENSIONAL VERSION OF THE CONVEXITY CONDITION

**Theorem 4.A.1.** *Let the dimension  $d = 3$ . Assume a zero source term, i.e.  $s = 0$ . The variation entropy  $\eta = \eta(\nabla\phi)$  diminishes in time, i.e.*

$$\int_{\Omega} \eta(\nabla\phi(x, t)) d\Omega \leq \int_{\Omega} \eta(\nabla\phi_0(x)) d\Omega, \quad (4.A.1)$$

for all  $t > 0$ , if and only if the variation entropy is given by

$$\eta = \eta(\nabla\phi) = \eta(r, \varphi) = F(\theta, \varphi)r, \quad (4.A.2)$$

where  $r$  and the  $\varphi$  are the polar coordinates corresponding to  $\nabla\phi$ . The scalar-valued function  $F = F(\varphi, \theta)$  satisfies

$$A \geq 0, \quad (4.A.3a)$$

$$B \leq A. \quad (4.A.3b)$$

with

$$\begin{aligned} A &= 4 \left( 2F + \frac{\partial F}{\partial \varphi} \cot \varphi + \frac{\partial^2 F}{\partial \theta^2} \csc^2 \varphi \right), \\ B &= \sqrt{2} \csc^2 \varphi \left( \frac{\partial F^2}{\partial \varphi} (1 - \cos 4\varphi) + 32 \left( \frac{\partial F}{\partial \theta} \right)^2 + 8 \left( \frac{\partial^2 F}{\partial \theta^2} \right)^2 \right. \\ &\quad \left. + 32 \left( \left( \frac{\partial^2 F}{\partial \varphi \partial \theta} \right)^2 - \left( \frac{\partial F}{\partial \theta} \right)^2 \right) \sin^2 \varphi \right. \\ &\quad \left. + 8 \left( \frac{\partial F}{\partial \varphi} \frac{\partial^2 F}{\partial \theta^2} - 4 \frac{\partial^2 F}{\partial \varphi \partial \theta} \frac{\partial F}{\partial \theta} \right) \sin(2\varphi) \right)^{1/2}. \end{aligned} \quad (4.A.4a)$$

*Proof.* We provide the details of the restriction (4.A.3a)-(4.A.3b) on  $F$ . We follow the same procedure as in the 2-dimensional case and thus we show that the eigenvalues of the Hessian are positive. Therefore we employ spherical coordinates:

$$v_1 = r \cos \theta \sin \varphi \quad (4.A.5a)$$

$$v_2 = r \sin \theta \sin \varphi \quad (4.A.5b)$$

$$v_3 = r \cos \varphi \quad (4.A.5c)$$

The first derivatives can be written in spherical coordinates as:

$$\frac{\partial}{\partial v_1} = \cos \theta \sin \varphi \frac{\partial}{\partial r} - \frac{\sin \theta}{r \sin \varphi} \frac{\partial}{\partial \theta} + \frac{\cos \theta \cos \varphi}{r} \frac{\partial}{\partial \varphi} \quad (4.A.6a)$$

$$\frac{\partial}{\partial v_2} = \sin \theta \sin \varphi \frac{\partial}{\partial r} + \frac{\cos \theta}{r \sin \varphi} \frac{\partial}{\partial \theta} + \frac{\sin \theta \cos \varphi}{r} \frac{\partial}{\partial \varphi} \quad (4.A.6b)$$

$$\frac{\partial}{\partial v_3} = \cos \varphi \frac{\partial}{\partial r} - \frac{\sin \varphi}{r} \frac{\partial}{\partial \varphi} \quad (4.A.6c)$$

The computation of the second derivatives is straightforward but at the same time quite involved. Here we only provide the resulting components of the Hessian, which are

$$\begin{aligned} \frac{\partial^2}{\partial v_1^2} \eta &= \frac{\sin^2 \theta}{r} F - \frac{\sin \theta \cos \theta}{r} \frac{\partial F}{\partial \theta} + \frac{\sin \theta \cos \theta}{r \sin^2 \varphi} \frac{\partial F}{\partial \theta} + \frac{\sin^2 \theta}{r \sin^2 \varphi} \frac{\partial^2 F}{\partial \theta^2} + \frac{\sin^2 \theta \cos \varphi}{r \sin \varphi} \frac{\partial F}{\partial \varphi} \\ &\quad - \frac{\sin \theta \cos \theta \cos \varphi}{r \sin \varphi} \frac{\partial^2 F}{\partial \varphi \partial \theta} + \frac{\cos^2 \theta \cos^2 \varphi}{r} F + \frac{\cos \theta \cos^2 \varphi \sin \theta}{r \sin^2 \varphi} \frac{\partial F}{\partial \theta} \\ &\quad - \frac{\cos \theta \sin \theta \cos \varphi}{r \sin \varphi} \frac{\partial^2 F}{\partial \varphi \partial \theta} + \frac{\cos^2 \theta \cos^2 \varphi}{r} \frac{\partial^2 F}{\partial \varphi^2}, \end{aligned} \quad (4.A.7a)$$

$$\begin{aligned} \frac{\partial^2}{\partial v_2^2} \eta &= \frac{\cos^2 \theta}{r} F + \frac{\sin \theta \cos \theta}{r} \frac{\partial F}{\partial \theta} - \frac{\sin \theta \cos \theta}{r \sin^2 \varphi} \frac{\partial F}{\partial \theta} + \frac{\cos^2 \theta}{r \sin^2 \varphi} \frac{\partial^2 F}{\partial \theta^2} + \frac{\cos^2 \theta \cos \varphi}{r \sin \varphi} \frac{\partial F}{\partial \varphi} \\ &\quad + \frac{\sin \theta \cos \theta \cos \varphi}{r \sin \varphi} \frac{\partial^2 F}{\partial \varphi \partial \theta} + \frac{\sin^2 \theta \cos^2 \varphi}{r} F - \frac{\cos \theta \cos^2 \varphi \sin \theta}{r \sin^2 \varphi} \frac{\partial F}{\partial \theta} \\ &\quad + \frac{\cos \theta \sin \theta \cos \varphi}{r \sin \varphi} \frac{\partial^2 F}{\partial \varphi \partial \theta} + \frac{\sin^2 \theta \cos^2 \varphi}{r} \frac{\partial^2 F}{\partial \varphi^2}, \end{aligned} \quad (4.A.7b)$$

$$\frac{\partial^2}{\partial v_3^2} \eta = \left( F + \frac{\partial^2 F}{\partial \varphi^2} \right) \frac{\sin^2 \varphi}{r}, \quad (4.A.7c)$$

$$\begin{aligned} \frac{\partial^2 \eta}{\partial v_1 \partial v_2} &= \frac{\partial^2 \eta}{\partial v_2 \partial v_1} = \cos^2 \varphi \frac{\sin^2 \theta - \cos^2 \theta}{r \sin^2 \varphi} \frac{\partial F}{\partial \theta} - \frac{\sin \theta \cos \theta}{r \sin^2 \varphi} \frac{\partial^2 F}{\partial \theta^2} \\ &\quad - \frac{\sin \theta \cos \theta \cos \varphi}{r \sin \varphi} \frac{\partial F}{\partial \varphi} - \frac{\cos \theta \sin^2 \varphi \sin \theta}{r} F \\ &\quad + \frac{(\cos^2 \theta - \sin^2 \theta) \cos \varphi}{r \sin \varphi} \frac{\partial^2 F}{\partial \varphi \partial \theta} + \frac{\sin \theta \cos \theta \cos^2 \varphi}{r} \frac{\partial^2 F}{\partial \varphi^2}, \end{aligned} \quad (4.A.7d)$$

$$\begin{aligned} \frac{\partial^2}{\partial v_1 \partial v_3} \eta &= \frac{\partial^2}{\partial v_3 \partial v_1} \eta = -\frac{\sin \varphi \cos \theta \cos \varphi}{r} \left( F + \frac{\partial^2 F}{\partial \varphi^2} \right) - \frac{\sin \theta \cos \varphi}{r \sin \varphi} \frac{\partial F}{\partial \theta} \\ &\quad + \frac{\sin \theta}{r} \frac{\partial^2 F}{\partial \varphi \partial \theta}, \end{aligned} \quad (4.A.7e)$$

$$\begin{aligned} \frac{\partial^2}{\partial v_2 \partial v_3} \eta &= \frac{\partial^2}{\partial v_3 \partial v_2} \eta = -\frac{\sin \varphi \sin \theta \cos \varphi}{r} \left( F + \frac{\partial^2 F}{\partial \varphi^2} \right) + \frac{\cos \theta \cos \varphi}{r \sin \varphi} \frac{\partial F}{\partial \theta} \\ &\quad - \frac{\cos \theta}{r} \frac{\partial^2 F}{\partial \varphi \partial \theta}. \end{aligned} \quad (4.A.7f)$$

The eigenvalues of the Hessian can be computed to be:

$$\lambda_1 = 0, \tag{4.A.8a}$$

$$\begin{aligned} \lambda_2 = \frac{1}{8r} & \left[ 8F + 4 \frac{\partial F}{\partial \varphi} \cot \varphi + 4 \frac{\partial^2 F}{\partial \theta^2} \csc^2 \varphi \right. \\ & + \sqrt{2} \csc^2 \varphi \left( \frac{\partial F^2}{\partial \varphi} (1 - \cos 4\varphi) + 32 \frac{\partial F^2}{\partial \theta} + 8 \frac{\partial^2 F^2}{\partial \theta^2} \right. \\ & + 32 \left( \left( \frac{\partial^2 F}{\partial \varphi \partial \theta} \right)^2 - \left( \frac{\partial F}{\partial \theta} \right)^2 \right) \sin^2 \varphi \\ & \left. \left. + 8 \left( \frac{\partial F}{\partial \varphi} \frac{\partial^2 F}{\partial \theta^2} - 4 \frac{\partial^2 F}{\partial \varphi \partial \theta} \frac{\partial F}{\partial \theta} \right) \sin(2\varphi) \right)^{1/2} \right], \tag{4.A.8b} \end{aligned}$$

$$\begin{aligned} \lambda_3 = \frac{1}{8r} & \left[ 8F + 4 \frac{\partial F}{\partial \varphi} \cot \varphi + 4 \frac{\partial^2 F}{\partial \theta^2} \csc^2 \varphi \right. \\ & - \sqrt{2} \csc^2 \varphi \left( \frac{\partial F^2}{\partial \varphi} (1 - \cos 4\varphi) + 32 \frac{\partial F^2}{\partial \theta} + 8 \frac{\partial^2 F^2}{\partial \theta^2} \right. \\ & + 32 \left( \left( \frac{\partial^2 F}{\partial \varphi \partial \theta} \right)^2 - \left( \frac{\partial F}{\partial \theta} \right)^2 \right) \sin^2 \varphi \\ & \left. \left. + 8 \left( \frac{\partial F}{\partial \varphi} \frac{\partial^2 F}{\partial \theta^2} - 4 \frac{\partial^2 F}{\partial \varphi \partial \theta} \frac{\partial F}{\partial \theta} \right) \sin(2\varphi) \right)^{1/2} \right]. \tag{4.A.8c} \end{aligned}$$

Positivity of the eigenvalues leads to the restrictions on  $F$ . □



---

## A THEORETICAL FRAMEWORK FOR DISCONTINUITY CAPTURING

---

*“We believe that the multiscale framework with a proper set of optimality conditions is the right underlying theoretical structure that may more naturally lead to discontinuity capturing formulations. — Bazilevs et al. [19]*

This chapter is reproduced from [61]:

M.F.P. ten Eikelder, Y. Bazilevs and I. Akkerman, *A theoretical framework for discontinuity capturing: Joining variational multiscale analysis and variation entropy theory*, *Comput. Methods. Appl. Mech. Engrg.* (2020) 112664

### ABSTRACT

*In this chapter we show that the variational multiscale method together with the variation entropy concept form the underlying theoretical framework of discontinuity capturing. Variation entropy theory (Chapter 4) is the recently introduced concept that equips total variation diminishing solutions with an entropy foundation. This is the missing ingredient in order to show that the variational multiscale method can capture sharp layers. The novel framework naturally equips the variational multiscale method with a class of discontinuity capturing operators. This class includes the popular  $YZ\beta$  method and methods based on the residual of the variation-entropy. The discontinuity capturing mechanisms do not contain ad hoc devices and appropriate length scales are derived. Numerical results obtained with quadratic NURBS are virtually oscillation-free and show sharp layers, which confirms the viability of the methodology.*

## 5.1 INTRODUCTION

Discontinuities in physical quantities such as densities, pressures and velocities often occur in scientific and industrial problems. Common examples include explosions, cavitation events, two-fluid flows and traffic congestion. These phenomena are generally modeled by (nonlinear) conservation laws. Numerical methods that aim to solve these conservation laws encounter difficulties at the shock waves. Straightforward discretizations pollute the discrete solution by spurious oscillations. To overcome this, the numerical method typically introduces additional diffusion/viscosity near the shock. There exist many possibilities, depending on the underlying numerical method, on how to determine this diffusion term.

In the finite-difference and finite-volume world, additional diffusion is often the result of one of the following approaches. Perhaps the simplest technique to introduce diffusion is to use a standard upwind method. This removes the spurious oscillations, but the price one has to pay is a significant decrease of accuracy. An alternative is to use a monotonic upwind scheme for conservation laws (MUSCL) [182, 192, 195, 196], also known as a limiter scheme, which reduces the numerical flux to first-order near the shock. Several other approaches equip the numerical method with discrete features. Examples include schemes with the monotonicity property [88], the total variation diminishing schemes [88, 90, 175] or methods that ensure the maximum principle [88, 204], e.g., in two-fluid flow simulations [8, 53, 63].

In the context of finite element methods, the spurious wiggles were first addressed with the Streamline upwind-Petrov Galerkin (SUPG) method for incompressible flow problems in the well-known 1982 paper [33] and for compressible flow problems in [116, 187]. The compressible flow case required a quasi-linear form which leads to the concept of generalized advection operators. In both cases the SUPG method adds diffusion only in the direction of the flow and is not subject to artificial diffusion criticism. The SUPG method provides accurate solutions without oscillations when strong shocks are absent. In regions near sharp layers a more robust formulation was needed. To this purpose, several shock capturing operator mechanisms are introduced. One of the first of these techniques has been proposed in [112]. This method provides control of gradient of the solution. The sharp layers in compressible flows were addressed by Hughes et al. who proposed to use entropy variables [107] and entropy variables in combination with the SUPG operator [91, 110]. Le Beau and Tezduyar equipped the original SUPG method with a shock capturing operator in conservative variables [138, 139]. The numerical computations reveal that results of using entropy variables without shock-capturing are nearly indistinguishable from using conservative variables with shock-capturing [138]. This might indicate a similarity or relation between shock capturing methods and entropy variables. Another important discontinuity capturing method is  $\gamma Z\beta$  shock-capturing [19, 185, 189, 190]. This shock-capturing method is based on scaled residuals and contains user-defined parameters which can be chosen depending on the smoothness of the layer. Other work on discontinuity capturing includes the CAU method [77] in which the flow velocity in the SUPG term is replaced by an approximate upwind direction and the work of Sampaio and Coutinho [54] in which an effective transport velocity is used. For a more complete overview of stabilized methods and shock

capturing techniques for compressible flows we refer to the review papers [114, 121, 122].

A particular class of stabilized finite element methods is that of the variational multiscale methods [100, 103, 115]. The idea is to incorporate the effect of the small-scales via a model equation in the resolved part of the solution. This improves the stability of the finite element scheme. This framework provides a theoretical foundation of stabilized methods. It has been widely applied for the computation of incompressible turbulence [4, 16, 17, 49, 58, 94, 131, 163]. The corresponding turbulence model is residual-based and does not contain ad-hoc mechanisms. The technique finds also applications in free surface flow and FSI computations [3, 5, 133, 141]. The multiscale formulation is often augmented with an artificial discontinuity capturing term when sharp layers may occur. The VMS method offers rich possibilities to design new methods. It can be used to enforce a particular property in the numerical method, such as a total variation bounding constraint [67] and the maximum principle [70]. Other recent work includes a VMS method that employs particular fine-scale models to arrive at a discontinuity capturing term [147].

Recently a popular discontinuity capturing method known by the name *entropy viscosity method* has been introduced [85]. This method bases the added nonlinear viscosity on the entropy residual. The motivation originates from the fact that the entropy satisfies a conservation equation in smooth regions and an inequality at shock waves. Basing the viscosity on the entropy production does not affect the smooth regions while in shock regions numerical dissipation is added. The entropy viscosity method has been further developed in the framework of discontinuous Galerkin methods in [206]. Furthermore, the stability of explicit entropy viscosity methods has been analyzed [28]. The method is a promising technique and has shown quite well behavior on many benchmark problems. It is however an heuristic approach for which, to the best knowledge of the authors, the theoretical justification is still missing. We cite

Guermond et al. [85]: *‘the amount of theory to justify the approach is almost non-existent. The justification of the method is mainly heuristic for the time being.’*

The idea of using an entropy concept to locate sharp layers is interesting. We have recently proposed the variation entropy theory [59] which provides entropy solutions with a new perspective and can be viewed an extension of total variation diminishing solutions. In order to identify sharp layers in solution profiles, the idea is to look at the gradient of the solution instead of at the solution itself. The variation entropy concept provides an entropy framework to analyze the behavior of the gradient of the solution using the so-called variation entropy condition. In a numerical setting this can be a tool to locate Gibbs oscillations.

All the previously mentioned techniques that add numerical diffusion in the region of the sharp layer are in some way the result of equipping the method with a favorable numerical property. The methods are *ad hoc* technologies that are not derived from the continuous partial differential equation. We note that Bazilevs et al. [19] conjecture that the variational multiscale method is the correct theoretical groundwork for discontinuity capturing methods:

Bazilevs et al. [19]: ‘While stabilized methods may be derived on the basis of the variational multiscale methodology, discontinuity capturing is an *ad hoc* technique. Nevertheless, it is a widely used technology that enables a practitioner to successfully tackle real-world applications. We believe that the multiscale framework with a proper set of optimality conditions is the right underlying theoretical structure that may more naturally lead to discontinuity capturing formulations. This conjecture is intriguing and warrants further investigation.’

In the current chapter we prove that this conjecture is valid. To establish this, we unify previous ideas and concepts into a variational multiscale-variation entropy framework. We believe that the variation entropy theory was the missing element in order to be able to demonstrate the correctness of the conjecture in [19]. The variation entropy idea tells us the location of the viscosity whereas the multiscale concept provides a way to model the viscosity via the missing scales. Merging the variational multiscale method with the variation entropy framework *naturally* augments the VMS method with a discontinuity capturing operator. We sketch this in Figure 5.1. We propose a discontinuity capturing viscosity that is variation-entropy residual-based. In some sense this is similar to the entropy viscosity method [85] where the residual is based on the entropy. In contrast, our discontinuity capturing term comes with theoretical foundations. We emphasize that the proposed framework does not contain *ad hoc* devices. The approximate small-scale models are physics-based by means of Green’s functions and residuals.

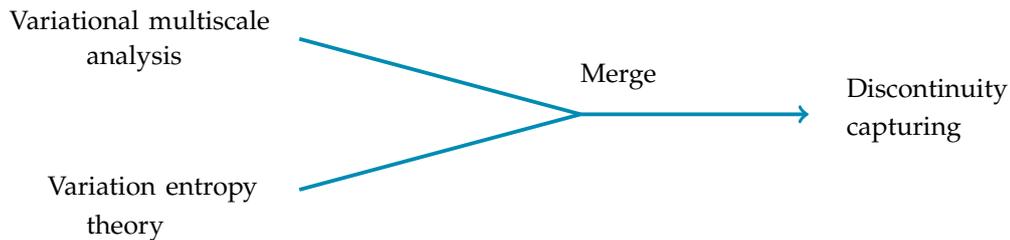


Figure 5.1: Merging the variational multiscale method and the variation entropy concept leads to a discontinuity capturing term

The remainder of the work can be summarized as follows. The Section 5.2 briefly introduces the notion of entropy and variation entropy solutions. In Section 5.3 we present the discontinuity capturing framework based on the variational multiscale analysis and variation entropy theory. Section 5.4 presents numerical results and in Section 5.5 we draw the conclusions and outline avenues for future research.

## 5.2 ENTROPY SOLUTIONS

### 5.2.1 The classical entropy

Let  $\Omega \subset \mathbb{R}^d$  be an open and connected domain. Consider the scalar-valued conservation problem:

find  $\phi : \Omega \times \mathcal{I} \rightarrow \mathbb{R}$  such that

$$\partial_t \phi + \nabla \cdot \mathbf{f} = 0, \quad (\mathbf{x}, t) \in \Omega \times \mathcal{I}, \quad (5.2.1a)$$

$$\phi(\mathbf{x}, 0) = \phi_0(\mathbf{x}). \quad (5.2.1b)$$

The problem is equipped with appropriate boundary conditions. We assume that the initial condition  $\phi_0 \in L^\infty(\Omega)$  has compact support in  $\Omega$ . The smooth (nonlinear) flux denotes  $\mathbf{f} = \mathbf{f}(\phi)$ , the spatial coordinate is  $\mathbf{x} \in \Omega$  and time is  $t \in \mathcal{I} = (0, t_e)$  with  $t_e > 0$ . The problem (5.2.1) can produce discontinuities and shocks which motivates the usage of weak solutions. A weak solution  $\phi$  is a bounded function that satisfies

$$\int_0^\infty \int_\Omega (\phi \partial_t v + \mathbf{f} \cdot \nabla v) \, dx dt + \int_\Omega \phi_0(\mathbf{x}) v(\mathbf{x}, 0) \, dx = 0, \quad (5.2.2)$$

for all test functions  $v \in C_c^\infty(\Omega \times (0, \infty))$  (i.e.  $v$  is smooth and has compact support). An important observation is that physically and mathematically correct solutions are vanishing viscosity solutions. This is a key ingredient in the concept of *entropy solutions* which are weak solutions that satisfy an additional inequality, denoted as the *entropy condition*.

**Definition 5.2.1.** A solution of (5.2.1) is called an *entropy solution* if it satisfies, in the distributional sense, the *entropy condition*:

$$\partial_t \eta(\phi) + \nabla \cdot \mathbf{q}(\phi) \leq 0, \quad (5.2.3)$$

for all convex entropy functions  $\eta$ . Condition (5.2.3) is rigorously understood as

$$\int_0^\infty \int_\Omega (\eta(\phi) \partial_t v + \mathbf{q}(\phi) \cdot \nabla v) \, dx dt \geq 0, \quad (5.2.4)$$

for all test functions  $v \in C_c^\infty(\Omega \times (0, \infty))$ ,  $v \geq 0$ .

**Definition 5.2.2.** A pair of functions  $(\eta, \mathbf{q}) = (\eta(\phi), \mathbf{q}(\phi))$  is called an *entropy-entropy flux pair* for the conservation law (5.2.1) if

- $\eta$  is convex
- the compatibility condition is satisfied:

$$\frac{\partial \mathbf{q}}{\partial \phi} = \frac{\partial \eta}{\partial \phi} \frac{\partial \mathbf{f}}{\partial \phi}. \quad (5.2.5)$$

For smooth solutions the entropy condition is satisfied with equality, while the entropy dissipates at shock waves. In absence of boundary conditions, integration of (5.2.3) over  $\Omega$  leads to dissipation of the overall entropy:

$$\frac{d}{dt} \int_\Omega \eta(\phi) \, d\Omega \leq 0 \Rightarrow \int_\Omega \eta(\phi(\mathbf{x}, t)) \, d\Omega \leq \int_\Omega \eta(\phi_0(\mathbf{x})) \, d\Omega, \quad \text{for all } t > 0. \quad (5.2.6)$$

This *a-priori* estimate is the so-called entropy stability property and can be understood as a nonlinear  $L^2$ -stability for conservation laws when taking  $\eta(\phi) = \phi^2/2$ .

**Theorem 5.2.3.** *Entropy solutions are unique (in the scalar case).*

*Proof.* See [72, 135]. □

For more details about entropy solutions one can consult e.g., [14, 72, 135, 183].

### 5.2.2 The variation entropy

The idea of the variation entropy (VE), developed in [59], is to consider the entropy of the conservation law of  $\nabla\phi$  instead of  $\phi$ . We start off with the same conservation law:

find  $\phi : \Omega \times \mathcal{I} \rightarrow \mathbb{R}$  such that

$$\partial_t \phi + \nabla \cdot \mathbf{f} = 0, \quad (x, t) \in \Omega \times \mathcal{I}, \quad (5.2.7a)$$

$$\phi(x, 0) = \phi_0(x). \quad (5.2.7b)$$

in which the initial condition  $\phi(x, 0) = \phi_0(x) \in L^\infty(\Omega)$  is assumed to have compact support. The flux  $\mathbf{f} = \mathbf{f}(\phi) \in \mathcal{C}(\Omega, \mathbb{R})$  is possibly nonlinear.

**Remark 5.2.4.** In this work we restrict ourselves to the hyperbolic case, i.e.  $\mathbf{f} = \mathbf{f}(\phi)$ , unless explicitly indicated. It is also possible to consider the parabolic case in which the flux depends on  $\nabla\phi$ , i.e.  $\mathbf{f} = \mathbf{f}(\phi, \nabla\phi)$ . In that case one needs to assume that the matrix  $\partial\mathbf{f}/\partial\nabla\phi$  is symmetric negative definite.

**Definition 5.2.5.** The convex function  $\eta = \eta(\nabla\phi)$  is said to be a variation entropy if  $\eta(\mathbf{0}) = 0$  and it satisfies the variation entropy condition

$$\partial_t \eta + \nabla \cdot \mathbf{q} \leq 0, \quad (5.2.8)$$

in weak sense where the flux  $\mathbf{q}$  satisfies the compatibility condition

$$\mathbf{q} = \frac{\partial\mathbf{f}}{\partial\phi} \nabla\phi \cdot \frac{\partial\eta}{\partial\nabla\phi}. \quad (5.2.9)$$

**Remark 5.2.6.** For parabolic problems the variation entropy condition reads

$$\partial_t \eta + \nabla \cdot \mathbf{q} - \mathcal{D} \leq 0, \quad (5.2.10)$$

in weak sense where the flux  $\mathbf{q}$  and the non-conservative term  $\mathcal{D}$  are respectively given by:

$$\mathbf{q} = \frac{\partial\mathbf{f}}{\partial\phi} \nabla\phi \cdot \frac{\partial\eta}{\partial\nabla\phi} + \frac{\partial\mathbf{f}}{\partial\nabla\phi} \nabla\eta, \quad (5.2.11a)$$

$$\mathcal{D} = (\mathbf{H}_x\phi \mathbf{H}_{\nabla\phi}\eta) : \left( \frac{\partial\mathbf{f}}{\partial\nabla\phi} \mathbf{H}_x\phi \right). \quad (5.2.11b)$$

Here  $\mathbf{H}_x\phi$  and  $\mathbf{H}_{\nabla\phi}\eta$  are the Hessians of  $\phi$  and  $\eta$ .

**Proposition 5.2.7.** A variation entropy satisfies the homogeneity property:

$$\mathbf{v} \cdot \frac{\partial\eta}{\partial\mathbf{v}} = \eta, \quad \text{for all } \mathbf{v} \in \mathbb{R}^d. \quad (5.2.12)$$

**Theorem 5.2.8.** A function  $\eta$  is a variation entropy if and only if

- $\eta$  is positive homogeneous function of degree 1:

$$\eta(\gamma\mathbf{v}) = \gamma\eta(\mathbf{v}) \quad \text{for all } \gamma \geq 0, \mathbf{v} \in \mathbb{R}^d. \quad (5.2.13)$$

- $\eta$  is sub-additive:

$$\eta(\mathbf{v}_1 + \mathbf{v}_2) \leq \eta(\mathbf{v}_1) + \eta(\mathbf{v}_2) \quad \text{for all } \mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^d. \quad (5.2.14)$$

**Proposition 5.2.9.** *A convex function  $\eta$  is a variation entropy if and only if it is given by*

$$\eta = \eta(\nabla\phi) = \hat{\eta}(r, \boldsymbol{\theta}) = F(\boldsymbol{\theta})r, \quad (5.2.15)$$

where  $F = F(\boldsymbol{\theta})$  is a scalar-valued function and  $r$  and  $\boldsymbol{\theta}$  are the spherical polar coordinates of  $\nabla\phi$ . The convexity condition is in the 2-dimensional case:

$$F(\theta) + F''(\theta) \geq 0. \quad (5.2.16)$$

**Remark 5.2.10.** *The convexity demand in three dimensions is more involved. We refer to [59] for details.*

**Proposition 5.2.11.** *All semi-norms of  $\nabla\phi$  are variation entropies.*

**Corollary 5.2.12.** *A direct consequence of the homogeneity property (5.2.12) in Theorem 5.2.8 is that the variation entropy flux is given by*

$$\mathbf{q} = \frac{\partial \mathbf{f}}{\partial \phi} \eta, \quad (5.2.17)$$

and that the variation entropy condition is thus rigorously understood as

$$\int_0^\infty \int_{\mathbb{R}^d} \eta \left( \partial_t v + \frac{\partial \mathbf{f}}{\partial \phi} \cdot \nabla v \right) dx dt \geq 0, \quad (5.2.18)$$

for all test functions  $v \in C_c^\infty(\Omega \times (0, \infty))$ ,  $v \geq 0$ .

In the case of parabolic conservation laws we can write

$$\mathbf{q} = \frac{\partial \mathbf{f}}{\partial \phi} \eta + \frac{\partial \mathbf{f}}{\partial \nabla \phi} \nabla \eta. \quad (5.2.19)$$

Note that (5.2.17) is similar to the form of the compatibility condition of the classical entropy case when taking the derivative with respect to  $\nabla\phi$ :

$$\frac{\partial \mathbf{q}}{\partial \nabla \phi} = \frac{\partial \mathbf{f}}{\partial \phi} \otimes \frac{\partial \eta}{\partial \nabla \phi}. \quad (5.2.20)$$

Examples of variation entropies are

$$\eta = \eta(\nabla\phi) = \|\nabla\phi\|_2, \quad (5.2.21a)$$

$$\eta = \eta(\nabla\phi) = \mathbf{c} \cdot \nabla\phi, \text{ for } \mathbf{c} \in \mathbb{R}^d, \quad (5.2.21b)$$

$$\eta = \eta(\nabla\phi) = \|\|\nabla\phi\|_{\mathbf{A}} := \left( \nabla\phi^T \mathbf{A} \nabla\phi \right)^{1/2}, \quad (5.2.21c)$$

for positive semi-definite matrix  $\mathbf{A}$ .

where  $\|\cdot\|_2$  is the standard 2-norm. Variation entropy (5.2.21b) is the only linear variation entropy.

**Definition 5.2.13.** *A pair of functions  $(\eta, \mathbf{q})$  is called a variation entropy-variation entropy flux pair for the conservation law (5.2.7) if*

- $\eta$  is a variation entropy,
- the flux  $\mathbf{q}$  is given by (5.2.17).

**Definition 5.2.14.** We call  $\phi = \phi(\mathbf{x}, t)$  a variation entropy solution of (5.2.7) if  $\phi$  is a weak solution and  $\phi$  satisfies (5.2.8) in a weak sense for each variation entropy-variation entropy flux pair  $(\eta, \mathbf{q})$ .

Physically relevant solutions are vanishing viscous solutions  $\phi^\epsilon$  satisfying:

$$\partial_t \phi^\epsilon + \nabla \cdot \mathbf{f}(\phi^\epsilon) = \epsilon \Delta \phi^\epsilon. \quad (5.2.22)$$

Suppose  $\phi^\epsilon$  is uniformly bounded in  $L^\infty(\Omega)$  and

$$\phi^\epsilon \rightarrow \phi \quad \text{a.e. as } \epsilon \rightarrow 0, \quad (5.2.23)$$

then we say that  $\phi$  is a physically relevant solution. In the following theorem we state that physically relevant solutions are, apart from classical entropy solutions, also variation entropy solutions.

**Theorem 5.2.15.** The limit solution  $\phi = \lim_{\epsilon \rightarrow 0} \phi^\epsilon$  is a variation entropy solution.

*Proof.* See [59]. □

Analogously to the classical entropy case, in absence of boundary conditions we can integrate over the domain  $\Omega$  to get a dissipation of the overall variation entropy:

$$\frac{d}{dt} \int_{\Omega} \eta(\nabla \phi) d\Omega \leq 0 \Rightarrow \int_{\Omega} \eta(\nabla \phi(\mathbf{x}, t)) d\Omega \leq \int_{\Omega} \eta(\nabla \phi_0(\mathbf{x})) d\Omega \quad \text{for all } t > 0. \quad (5.2.24)$$

### 5.3 THE VMS-VARIATION ENTROPY FRAMEWORK FOR DISCONTINUITY CAPTURING METHODS

In this section we employ the variation entropy concepts within the variational multiscale framework to derive a class of discontinuity capturing methods.

#### 5.3.1 Starting point

We take the point of view that a good numerical method solves the conservation law problem:

find  $\phi : \Omega \times \mathcal{I} \rightarrow \mathbb{R}$  such that

$$\partial_t \phi + \nabla \cdot \mathbf{f} = 0, \quad (\mathbf{x}, t) \in \Omega \times \mathcal{I}, \quad (5.3.1a)$$

$$\phi(\mathbf{x}, 0) = \phi_0(\mathbf{x}), \quad \mathbf{x} \in \Omega, \quad (5.3.1b)$$

with smooth flux  $\mathbf{f} = \mathbf{f}(\phi)$  and at the same time does not harm the variation entropy condition

$$\partial_t \eta + \nabla \cdot \mathbf{q} \leq 0, \quad (5.3.2)$$

for variation entropy  $\eta$ . In the remainder of this section we derive a multiscale framework which endeavors this.

**Remark 5.3.1.** *As stated in Remark 5.2.4 we focus here on hyperbolic conservation laws. We want to emphasize that changing to the parabolic case is a trivial execution. Furthermore, one can augment the conservation law with a non-zero source term.*

We start off with the regularized conservation law:

find  $\phi^\epsilon : \Omega \times \mathcal{I} \rightarrow \mathbb{R}$  such that

$$\partial_t \phi^\epsilon + \nabla \cdot \mathbf{f}(\phi^\epsilon) = \epsilon \Delta \phi^\epsilon, \quad (x, t) \in \Omega \times \mathcal{I}, \quad (5.3.3a)$$

$$\phi^\epsilon(x, 0) = \phi_0(x), \quad x \in \Omega, \quad (5.3.3b)$$

with regularization parameter  $\epsilon \geq 0$ . The initial condition  $\phi(x, 0) = \phi_0(x) \in L^\infty(\Omega)$  is assumed to have compact support. Note that the limit solution is a variation entropy solution (Theorem 5.2.15). The weak form of this problem is:

find  $\phi^\epsilon \in \mathcal{W}$  such that for all  $w \in \mathcal{W}$

$$(w, \partial_t \phi^\epsilon)_{L^2(\Omega)} - (\mathbf{f}(\phi^\epsilon), \nabla w)_{L^2(\Omega)} = -(\phi^\epsilon, w)_{\mathcal{W}}. \quad (5.3.4)$$

Here  $(\cdot, \cdot)_{L^2(\Omega)}$  is the standard  $L^2$ -innerproduct and we have used the self-adjoint positive-definite linear viscosity operator to define an inner product:

$$(u, v)_{\mathcal{W}} := (\epsilon \nabla u, \nabla v)_{L^2(\Omega)}. \quad (5.3.5)$$

A natural norm is the *energy norm*:

$$\|v\|_{\mathcal{W}}^2 := \left\| \epsilon^{1/2} \nabla v \right\|_{L^2(\Omega)}^2. \quad (5.3.6)$$

For more details about the construction of an energy norm we refer to [165].

### 5.3.2 Mesh representation and geometrical mapping

Let the parametric domain be  $\hat{\Omega} := (-1, 1)^d \subset \mathbb{R}^d$  and let us denote the mesh in the parametric domain with  $\mathcal{M}$ . The elements  $Q$  of  $\mathcal{M}$  have element size  $h_Q = \text{diag}(Q)$  (diagonal length). We denote the physical domain by  $\Omega \subset \mathbb{R}^d$  and the continuously differentiable geometrical map (with continuously differentiable inverse) by  $\mathbf{F} : \hat{\Omega} \rightarrow \Omega$ . Each parametric element  $Q \in \mathcal{M}$  maps into a physical element

$$\Omega_K = \mathbf{F}(Q), \quad (5.3.7)$$

which induces a physical mesh:

$$\mathcal{K} = \mathbf{F}(\mathcal{M}) := \{\Omega_K : \Omega_K = \mathbf{F}(Q), Q \in \mathcal{M}\}. \quad (5.3.8)$$

We denote the corresponding Jacobian by  $\mathbf{J} = D\mathbf{F} = \partial x / \partial \xi$ , or in index notation  $J_{ij} = \partial x_i / \partial \xi_j$ . We define the second-rank element metric tensor as

$$\mathbf{G} = \frac{\partial \xi^T}{\partial x} \frac{\partial \xi}{\partial x} = \mathbf{J}^{-T} \mathbf{J}^{-1}, \quad \text{or in index notation} \quad G_{ij} = \frac{\partial \xi_k}{\partial x_i} \frac{\partial \xi_k}{\partial x_j}. \quad (5.3.9)$$

The inverse is given by

$$\mathbf{G}^{-1} = \frac{\partial \mathbf{x}}{\partial \boldsymbol{\xi}} \frac{\partial \mathbf{x}^T}{\partial \boldsymbol{\xi}} = \mathbf{J}\mathbf{J}^T, \quad \text{with the index notation} \quad G_{ij}^{-1} = \frac{\partial x_i}{\partial \xi_k} \frac{\partial x_j}{\partial \xi_k}. \quad (5.3.10)$$

Furthermore we define the physical mesh size  $h_K$  as

$$h_K^2 = \frac{h_Q^2}{d} \|\mathbf{J}\|_F^2, \quad (5.3.11)$$

where the subscript  $F$  refers to the Frobenius norm given by

$$\|\mathbf{J}\|_F^2 = \sum_{i,j=1}^d \left( \frac{\partial x_i}{\partial \xi_j} \right)^2 = \text{Tr}(\mathbf{J}\mathbf{J}^T) = \text{Tr}(\mathbf{G}^{-1}), \quad (5.3.12)$$

with  $\text{Tr}$  the trace operator. The Frobenius norm is a natural choice for mesh metrics since it is rotation-invariant and appears in several well-known mesh quality measures. Another benefit is its lower computational costs compared to the standard  $p$ -norm [130]. Furthermore, on a Cartesian mesh it reduces to the length of the diagonal of an element. We use the notation  $\nabla_{\boldsymbol{\xi}}$  to distinguish differentiation with respect to the reference coordinates  $\boldsymbol{\xi}$  from the gradient in physical coordinates  $\nabla$ .

On uniform Cartesian meshes we use the notation  $\partial x / \partial \xi := \partial x_1 / \partial \xi_1 = \partial x_2 / \partial \xi_2 = \partial x_3 / \partial \xi_3 > 0$ .

### 5.3.3 The multiscale split

The residual-based variational multiscale approach splits the solution into a large-scale and a small-scale component. The large-scale component is solved numerically, whereas the small-scale contribution is treated in an approximate sense. Assume that there exists an idempotent (and possibly nonlinear) projector  $\mathcal{P}^h : \mathcal{W} \rightarrow \mathcal{W}^h$ . The trial solution and weighting function spaces split as

$$\mathcal{W} = \mathcal{P}^h \mathcal{W} \oplus (\mathcal{I} - \mathcal{P}^h) \mathcal{W} = \mathcal{W}^h \oplus \mathcal{W}', \quad (5.3.13)$$

where  $\mathcal{W}^h$  is the coarse-scale linear subspace and  $\mathcal{W}'$  is its infinite-dimensional complement. This allows us to decompose  $\phi^\epsilon \in \mathcal{W}$  and  $w \in \mathcal{W}$  as:

$$\phi^\epsilon = (\phi^\epsilon)^h + (\phi^\epsilon)' \quad \text{and} \quad w = w^h + w', \quad (5.3.14)$$

where  $(\phi^\epsilon)^h = \mathcal{P}^h \phi^\epsilon$  and  $w^h = \mathcal{P}^h w$ . Because  $\mathcal{W}^h$  is a subset of  $\mathcal{W}$ , (5.3.4) implies that for all  $w^h \in \mathcal{W}^h$

$$\begin{aligned} & (w^h, \partial_t((\phi^\epsilon)^h + (\phi^\epsilon)'))_{L^2(\Omega)} - (\mathbf{f}((\phi^\epsilon)^h + (\phi^\epsilon)'), \nabla w^h)_{L^2(\Omega)} \\ &= - \left( (\phi^\epsilon)^h + (\phi^\epsilon)', w^h \right)_{\mathcal{W}'}, \end{aligned} \quad (5.3.15)$$

regardless of the possible nonlinearity of  $\mathcal{P}^h$ . Take for the coarse-scale space  $\mathcal{W}^h \subset H^1(\Omega)$ . Sending  $\epsilon \rightarrow 0$  in (5.3.15) and noting that due to

$$\left| \left( (\phi^\epsilon)^h, w^h \right)_{\mathcal{W}'} \right| \leq \|(\phi^\epsilon)^h\|_{\mathcal{W}} \|w^h\|_{\mathcal{W}} \rightarrow 0 \quad \text{as} \quad \epsilon \rightarrow 0, \quad (5.3.16)$$

we arrive at

$$(w^h, \partial_t(\phi^h + \phi'))_{L^2(\Omega)} - (\mathbf{f}(\phi^h + \phi'), \nabla w^h)_{L^2(\Omega)} = - (\phi', w^h)_{\mathcal{W}'}, \quad (5.3.17)$$

for all  $w^h \in \mathcal{W}^h$ , with  $\phi^h := \lim_{\epsilon \rightarrow 0} (\phi^\epsilon)^h$  and  $\phi' := \lim_{\epsilon \rightarrow 0} (\phi^\epsilon)'$ . Here we use an abuse of notation for the term on the right-hand side which is the limit of the small-scale regularization term. In contrast to the large-scale component, the small-scale term does not vanish in general due to the (possibly) unbounded gradient  $\nabla \phi'$ . Note that this weak formulation is still exact. However, the infinite-dimensionality of  $\mathcal{W}'$  does not allow for a discrete implementation.

Let  $\eta$  be a positive-valued variation entropy function  $\eta = \eta(\nabla \phi) : \mathbb{R}^d \rightarrow \mathbb{R}_+$  (and thus nonlinear, eliminating the linear variation entropy (5.2.21b)). We assume that  $\eta(\nabla w) \in L^2(\Omega) \forall w \in \mathcal{W}$ . The large-scale solution space associated with  $\eta$  is defined as

$$\mathcal{V}^h := \eta(\nabla \mathcal{W}^h), \quad (5.3.18)$$

with the elements

$$\eta^h := \eta(\nabla \phi^h) \in \mathcal{V}^h. \quad (5.3.19)$$

We define the small-scale variation entropy as  $\eta' := \eta(\nabla \phi) - \eta^h$ . Let us denote the residual of the conservation law and that of the variation entropy condition as:

$$\mathcal{R}_{\text{CL}}\phi := \partial_t \phi + \nabla \cdot \mathbf{f}, \quad (5.3.20a)$$

$$\begin{aligned} \mathcal{R}_{\text{VE}}\eta &:= \partial_t \eta + \nabla \cdot \mathbf{q} \\ &= \frac{\partial \eta}{\partial \nabla \phi} \cdot \nabla (\mathcal{R}_{\text{CL}}\phi). \end{aligned} \quad (5.3.20b)$$

#### 5.3.4 A standard optimality projector

To establish scale separation the projector  $\mathcal{P}^h$  needs to be selected. We construct the projector via the minimization of a functional. The standard approach is the following. Consider the minimization problem:

find  $\phi^h \in \mathcal{W}^h$  such that:

$$\mathcal{L}(\phi - \phi^h) = \inf_{\theta^h \in \mathcal{W}^h} \mathcal{L}(\phi - \theta^h) \quad (5.3.21)$$

where the quadratic functional is given by

$$\mathcal{L}(\phi) = \frac{1}{2} \|\phi\|_{\mathcal{W}}^2. \quad (5.3.22)$$

**Lemma 5.3.2.** The functional  $\mathcal{M} : \mathcal{W}^h \rightarrow \mathbb{R}$  given by

$$\mathcal{M}(w^h) := \frac{1}{2} \|\phi - w^h\|_{\mathcal{W}}^2 \quad (5.3.23)$$

is strictly convex.

*Proof.* This follows from the positive definiteness of the second derivative which equals

$$d^2 \mathcal{M}(w^h)(u^h, v^h) = (u^h, v^h)_{\mathcal{W}}, \quad (5.3.24)$$

for  $u^h, v^h \in \mathcal{W}^h$ .  $\square$

**Theorem 5.3.3.** *Assuming  $\mathcal{W}^h$  is closed, problem (5.3.21)-(5.3.22) has a unique solution.*

*Proof.* This is a consequence of Lemma 5.3.2. See also [70].  $\square$

The multiscale split projector (5.3.13) is now defined as:

$$\mathcal{P}^h \phi = \operatorname{argmin}_{\phi^h \in \mathcal{W}^h} \mathcal{L}(\phi - \phi^h). \quad (5.3.25)$$

We obtain a stationary point when the Gateaux derivative of the functional  $\mathcal{L}$  in direction  $w^h$  vanishes:

$\mathcal{P}^h : \phi \in \mathcal{W} \rightarrow \phi^h \in \mathcal{W}^h$ : find  $\phi^h \in \mathcal{W}^h$  such that for all  $w^h \in \mathcal{W}^h$ :

$$d\mathcal{L}(\phi - \phi^h)(w^h) = 0. \quad (5.3.26)$$

Evaluating (5.3.26), the multiscale split projector takes the form:

$\mathcal{P}^h : \phi \in \mathcal{W} \rightarrow \phi^h \in \mathcal{W}^h$ : find  $\phi^h \in \mathcal{W}^h$  such that for all  $w^h \in \mathcal{W}^h$ :

$$(w^h, \phi^h - \phi)_{\mathcal{W}} = 0. \quad (5.3.27)$$

Employing this relation in the VMS weak formulation, via the multiscale split (5.3.14), cancels the symmetric contributions on the small-scales:

$$(w^h, \phi')_{\mathcal{W}} = 0. \quad (5.3.28)$$

As a result, the small-scale solution space is given by

$$\mathcal{W}' = \left\{ \phi' \in \mathcal{W} : (w^h, \phi')_{\mathcal{W}} = 0 \quad \text{for all } w^h \in \mathcal{W}^h \right\}. \quad (5.3.29)$$

**Remark 5.3.4.** *The orthogonality (5.3.28) is linked to correct energy behavior for the convection-diffusion and the incompressible Navier-Stokes equations. For details we refer to [57, 58].*

In the standard VMS framework the small-scales of the governing equations need to be modeled to arrive at a numerical method. For general details about small-scale modeling we refer to [17, 57, 113]. We employ the standard small-scale model for  $\phi'$ :

$$\hat{\phi}' = -\tau_{\text{CL}} \mathcal{R}_{\text{CL}} \phi^h, \quad (5.3.30a)$$

$$\partial_t \hat{\phi}' = 0, \quad (5.3.30b)$$

The scalar stabilization parameter  $\tau_{\text{CL}}$  is a mesh-dependent approximation (based on inverse estimates, see e.g [86]) of the inverse of the differential operator. We use the hat-sign to indicate that (5.3.30) is a small-scale model instead of it being

the actual small-scales. In the following we use this approximation and ignore the hat-sign. The current approach is known as the concept of static small-scales, due to assumption (5.3.30b). We note that, as an alternative, one can employ dynamic small-scales. In that case a dynamic version of (5.3.30a) is used and the second modeling assumption, relation (5.3.30b), is not made. This dynamic approach has some advantages [48, 57, 58]. Furthermore, we remark that nonlinear contributions of the small-scales can be incorporated in the residual, see e.g., [17].

By employing the orthogonality (5.3.28) and the small-scale model (5.3.30) in an SUPG fashion in (5.3.15) we arrive at:

$$\begin{aligned} \text{find } \phi^h \in \mathcal{W}^h \text{ such that for all } w^h \in \mathcal{W}^h \\ (w^h, \partial_t \phi^h)_{L^2(\Omega)} - (\nabla \mathbf{w}^h, \mathbf{f}(\phi^h))_{L^2(\Omega)} \\ + \sum_K \left( (\tau_{\text{CL}})_K \frac{\partial \mathbf{f}}{\partial \phi^h} \cdot \nabla w^h, \mathcal{R}_{\text{CL}}(\phi^h) \right)_{L^2(\Omega_K)} = 0. \end{aligned} \quad (5.3.31)$$

The consequence is thus that both the large and small-scale components stemming from the regularized term vanish. However, when incorporating the variation entropy condition in the projector the small-scale contribution does not vanish. We present this approach in the next subsection.

### 5.3.5 A variation entropy optimality projector

Here we present a new optimality projector that uses the variation entropy condition. This naturally leads to a discontinuity capturing term.

**Remark 5.3.5.** *Here we choose to enforce the variation entropy condition in an indirect manner. As an alternative one could use a more direct approach. We present the corresponding steps in Appendix 5.A. This alternative approach does not provide a convex problem and as such uniqueness of the minimization problem can not be guaranteed.*

Consider the minimization problem:

find  $\phi^h \in \mathcal{W}^h$  such that:

$$\mathcal{L}(\phi - \phi^h) = \inf_{\theta^h \in \mathcal{K}^h} \mathcal{L}(\phi - \theta^h), \quad (5.3.32)$$

where the constraint set reads:

$$\mathcal{K}^h := \left\{ \phi^h \in \mathcal{W}^h : (v^h, \eta(\nabla \phi^h) - \eta(\nabla \phi))_{L^2(\Omega)} \leq 0 \quad \text{for all } v^h \in \mathcal{V}^h \right\}. \quad (5.3.33)$$

**Lemma 5.3.6.** *The solution space  $\mathcal{K}^h$  is convex.*

*Proof.* This is direct consequence of the sub-additivity and the homogeneity of the variation entropy. Indeed let  $0 \leq \zeta \leq 1$  and let  $\phi_1^h, \phi_2^h \in \mathcal{K}^h$  then

$$\begin{aligned} & (v^h, \eta(\zeta \nabla \phi_1^h + (1 - \zeta) \nabla \phi_2^h))_{L^2(\Omega)} \\ & \leq (v^h, \eta(\zeta \nabla \phi_1^h) + \eta((1 - \zeta) \nabla \phi_2^h))_{L^2(\Omega)} \quad (\text{sub-additivity}) \\ & \leq (v^h, \zeta \eta(\nabla \phi_1^h) + (1 - \zeta) \eta(\nabla \phi_2^h))_{L^2(\Omega)} \quad (\text{homogeneity}) \\ & \leq (v^h, \zeta \eta(\nabla \phi) + (1 - \zeta) \eta(\nabla \phi))_{L^2(\Omega)} \quad (\phi_1^h, \phi_2^h \in \mathcal{K}^h) \\ & = (v^h, \eta(\nabla \phi))_{L^2(\Omega)}, \end{aligned} \quad (5.3.34)$$

for all  $v^h \in \mathcal{V}^h$ . This implies  $\zeta\phi_1^h + (1 - \zeta)\phi_2^h \in \mathcal{K}^h$ .  $\square$

**Theorem 5.3.7.** *Problem (5.3.32)-(5.3.33) has a unique solution.*

*Proof.* The constraint set  $\mathcal{K}^h$  is convex. Uniqueness follows from Lemma 5.3.2 in a similar fashion as in Theorem 5.3.3. Details can be found in [70].  $\square$

We proceed by opening the solution space. We penalize violating the constraint defined in (5.3.33). The constraint problem (5.3.32)-(5.3.33) converts into the unconstrained minimization problem:

find  $\phi^h \in \mathcal{W}^h$  such that:

$$\mathcal{J}(\phi - \phi^h, \phi, \phi^h) = \inf_{\theta^h \in \mathcal{W}^h} \mathcal{J}(\phi - \theta^h, \phi, \theta^h), \quad (5.3.35a)$$

where we have defined the functional  $\mathcal{J} : \mathcal{W}' \times \mathcal{W} \times \mathcal{W}^h \rightarrow \mathbb{R}$  as

$$\mathcal{J}(w_1, w_2, w_3) = \mathcal{L}(w_1) + \frac{1}{2} \|\sqrt{\mu} \{\eta(\nabla w_2) - \eta(\nabla w_3)\}_-\|_{L^2(\Omega)}^2. \quad (5.3.35b)$$

where  $\{\cdot\}_-$ , defined as  $\{a\}_- = (a - |a|)/2$ , isolates the negative part of its argument. Here  $\mu = \mu(\Omega) \geq 0$  is a parameter penalizing excess variation entropy of the coarse scale solution. The unit of  $\mu$  is  $[\mu] = [\phi]^2 / ([\eta]^2 T)^1$ . In the case that the unit of  $\eta$  is that of the solution over length, i.e.  $[\eta] = [\phi]/L$ ,  $\mu$  has the unit of a viscosity:  $[\mu] = L^2/T$ .

**Proposition 5.3.8.** *The functional  $\mathcal{J} = \mathcal{J}(\phi - \phi^h, \phi, \phi^h)$  is bounded:*

$$\mathcal{L}(\phi - \phi^h) \leq \mathcal{J}(\phi - \phi^h, \phi, \phi^h) \leq \mathcal{J}_{up}(\phi - \phi^h) \quad (5.3.36a)$$

where the upper bound is given by

$$\mathcal{J}_{up}(\phi - \phi^h) = \mathcal{L}(\phi - \phi^h) + \frac{1}{2} \|\sqrt{\mu} \eta(\nabla(\phi - \phi^h))\|_{L^2(\Omega)}^2. \quad (5.3.36b)$$

*Proof.* This follows from the sub-additivity of  $\eta$  (5.2.14):

$$\begin{aligned} \eta(\nabla\phi) - \eta^h &= \eta(\nabla\phi) - \eta(\nabla\phi + (\nabla\phi^h - \nabla\phi)) \\ &\geq \eta(\nabla\phi) - \eta(\nabla\phi) - \eta(\nabla\phi^h - \nabla\phi) \\ &= -\eta(\nabla\phi^h - \nabla\phi). \end{aligned} \quad (5.3.37)$$

$\square$

**Remark 5.3.9.** *In the case that the variation entropy equals  $\eta = \|\nabla\phi\|_2$ , the upper bound converts into*

$$\begin{aligned} \mathcal{J}_{up}(\phi - \phi^h) &= \frac{1}{2} \|\phi - \phi^h\|_{\mathcal{W}}^2 + \frac{1}{2} \|\mu^{1/2} \nabla(\phi - \phi^h)\|_{L^2(\Omega)}^2 \\ &= \frac{1}{2} \|(\epsilon + \mu)^{1/2} \nabla(\phi - \phi^h)\|_{L^2(\Omega)}^2, \end{aligned} \quad (5.3.38)$$

which is the energy norm with viscosity  $\epsilon + \mu$ .

<sup>1</sup> In this work we use the notation  $[a]$  to denote the unit of quantity  $a$ . Furthermore,  $L$  denotes and length unit and  $T$  a time unit.

We penalize violating the constraint defined in (5.3.33) by defining the projector  $\mathcal{P}^h \phi : \mathcal{W} \rightarrow \mathcal{W}^h$  as:

$$\mathcal{P}^h \phi = \operatorname{argmin}_{\phi^h \in \mathcal{W}^h} \mathcal{J}(\phi - \phi^h, \phi, \phi^h). \quad (5.3.39)$$

**Remark 5.3.10.** This can be viewed as a nonlinear Tikhonov-like regularization of orthogonal projection in  $\mathcal{W}$ . Alternatively, it can be seen as a penalty regularization of the inequality-constrained projection

$$\mathcal{P}_c^h \phi = \operatorname{arg} \left\{ \begin{array}{l} \min_{\phi^h \in \mathcal{W}^h} \mathcal{L}(\phi - \phi^h) \\ \text{subject to } \eta(\nabla \phi^h) \leq \eta(\nabla \phi) \text{ a.e. in } \Omega \end{array} \right\}. \quad (5.3.40)$$

The first-order optimality conditions for  $\mathcal{P}^h$  follow from equating the Gateaux derivative in direction  $w^h$  to zero. Noting that

$$d \eta(\nabla \phi^h)(\nabla w^h) = \frac{\partial \eta}{\partial \nabla \phi^h} \cdot \nabla w^h, \quad (5.3.41)$$

we obtain the problem:

$\mathcal{P}^h : \phi \in \mathcal{W} \rightarrow \phi^h \in \mathcal{W}^h$ : find  $\phi^h \in \mathcal{W}^h$  such that for all  $w^h \in \mathcal{W}^h$ :

$$\left( \phi', w^h \right)_{\mathcal{W}} = - \left( \mu \{ \eta' \}_- - \frac{\partial \eta^h}{\partial \nabla \phi^h}, \nabla w^h \right)_{L^2(\Omega)}. \quad (5.3.42)$$

Using the homogeneity of  $\eta^h$ , i.e. relation (5.2.12), we can write:

$$\frac{\partial \eta}{\partial \nabla \phi^h} = \frac{1}{\eta^h} \left( \frac{\partial \eta^h}{\partial \nabla \phi^h} \otimes \frac{\partial \eta^h}{\partial \nabla \phi^h} \right) \nabla \phi^h. \quad (5.3.43)$$

Thus we arrive at

$$\left( \phi', w^h \right)_{\mathcal{W}} = \left( \mathbf{K} \nabla \phi^h, \nabla w^h \right)_{L^2(\Omega)}, \quad (5.3.44)$$

in which the matrix  $\mathbf{K}$  is given by:

$$\mathbf{K} = \nu \left( \frac{\partial \eta^h}{\partial \nabla \phi^h} \otimes \frac{\partial \eta^h}{\partial \nabla \phi^h} \right), \quad (5.3.45a)$$

$$\nu = - \mu \frac{\{ \eta' \}_-}{\eta^h}. \quad (5.3.45b)$$

The parameter  $\nu \geq 0$ , referred to as *variation entropy viscosity*, scales with the relative error of the variation entropy  $\eta^h$  and has unit  $[\nu] = [\mu] = [\phi]^2 [\eta]^{-2} T^{-1}$ . Note that  $\mathbf{K}$  has the unit of a viscosity:

$$[\mathbf{K}] = [\nu] \frac{[\eta]^2}{[\nabla \phi]^2} = \frac{L^2}{T}. \quad (5.3.46)$$

The matrix (5.3.45a) acts as *diffusion based on the variation entropy residual*. Note that the diffusion operator of the streamline upwind method [33] acts in the direction of the flow. For discontinuity capturing control of gradients in the direction  $\nabla \phi^h$  is relevant [112]. The diffusion operator  $\mathbf{K}$  acts in the direction  $\partial \eta / \partial \nabla \phi^h$ . This is the direction in which the variation entropy changes and is thus a natural direction to add diffusion.

**Lemma 5.3.11.** *The matrix  $\mathbf{K}$  is symmetric positive semi-definite.*

*Proof.* Symmetry is trivial and the positive semi-definiteness is a direct consequence of  $v$  being positive.  $\square$

**Remark 5.3.12.** *One could alternatively start from the constrained projector  $\mathcal{P}_c^h$  and approximate the Lagrange multiplier associated with the constraint by penalizing  $-\mu\{\eta'\}_-$  to obtain the same result.*

At this point, no approximation has been made. We may substitute (5.3.44)-(5.3.45) into (5.3.17), illustrating how unresolved viscous dissipation in the fine-scale solution is expressed in terms of variation entropy in the coarse-scale problem when the coarse scales are defined by the nonlinear projector  $\mathcal{P}^h$ . Notice that, when taking the limit  $\epsilon \rightarrow 0$  the right-hand side of (5.3.45) does not vanish in this limit. This is consistent with the necessity of shock-capturing operators in the inviscid limit. Further, the right-hand side of (5.3.45) is independent of the precise choice of viscous operator, as one would hope in the case that an arbitrary regularization introduced for analysis purposes.

### 5.3.6 Small-scale model variation entropy

The current VMS-VE framework requires a model for the negative part of the small-scale variation entropy  $\{\eta'\}_-$ . This opens the door to explore several small-scale models leading to different numerical methods. Note that the small-scale variation entropy  $\eta' = \eta(\nabla\phi) - \eta^h$  is linked to the small-scale solution  $\phi'$  via:

$$\begin{aligned} \{\eta'\}_- &= \{\eta(\nabla\phi) - \eta^h\}_- \\ &= \{\eta(\nabla\phi^h + \nabla\phi') - \eta^h\}_- \\ &\leq \{\eta(\nabla\phi')\}_- \\ &= 0. \end{aligned} \tag{5.3.47}$$

Thus employing the model  $\{\eta'\}_- = \{\eta(\nabla\phi')\}_-$  where the small-scale solution  $\phi'$  is determined by the standard static model (5.3.30) causes the discontinuity capturing operator to vanish.

We propose a model for  $\{\eta'\}_-$  inspired by the variation entropy condition. Other possibilities may lead to an improved method with practical benefits and/or theoretical advantages. Remark that in the case of smooth solutions the variation entropy condition converts into an equality. Here we use the standard VMS method and write down the Euler-Lagrange equations of small-scale equation. Following this reasoning we propose the model for  $\{\eta'\}_-$ :

$$\widehat{\{\eta'\}}_{-\text{VE}} = -\tau_{\text{VE}} \left\{ \mathcal{R}_{\text{VE}} \left( \eta^h \right) \right\}_+, \tag{5.3.48}$$

where the hat-symbol indicates the modeling step. Here  $\tau_{\text{VE}} \geq 0$  represents a time-scale associated with the variation entropy. We note that  $\tau_{\text{VE}}$  is an element-wise parameter:  $\tau_{\text{VE}} = (\tau_{\text{VE}})_K$ . We wish to emphasize that model (5.3.48) is residual-based, both directly with residual  $\mathcal{R}_{\text{VE}}$  and with residual  $\mathcal{R}_{\text{CL}}$ , see (5.3.20b).

5.3.7 Variation entropy viscosity

The variation entropy viscosity corresponding to the model (5.3.48) is:

$$\nu_{VE} = \mu \tau_{VE} \frac{\{\mathcal{R}_{VE}(\eta^h)\}_+}{\eta^h}, \tag{5.3.49}$$

where the subscript *VE* refers to the variation-entropy residual. Despite the fact that the variation entropy viscosities are element-wise parameters, for the ease of notation we do not explicitly write a subscript *K* referring to element *K* in this subsection. Note that the variation entropy viscosity (5.3.49) vanishes when the variation entropy condition is satisfied.

The product  $\mu \tau_{VE}$  needs to be modeled. A natural approach would be to model the terms separately. In this case, a standard VMS approach could be used to model the intrinsic time-scale associated with the variation entropy,  $\tau_{VE}$ , i.e. one could use a discrete approximation of the inverse of the corresponding differential operator<sup>2</sup>. The term  $\mu$  is a penalty parameter that links the variation entropy to the regularization term. Recall that the unit of  $\mu$  is  $[\phi]^2[\eta]^{-2}T^{-1}$ . Since  $\mu$  is associated with the variation entropy, a natural choice for the time-scale in  $\mu$  is  $\tau_{VE}^{-1}$ . Following this approach, the model for  $\tau_{VE}$  would cancel in the product  $\mu \tau_{VE}$ . This suggests the alternative to model the product  $\mu \tau_{VE}$  as one quantity instead of modeling the separate terms. This is how we proceed. As  $\mu$  connects the variation entropy to the regularization term associated with the conservation law and this connection is governed by the operator  $\partial\eta/\partial\nabla\phi \cdot \nabla$  (see (5.3.20b)), we employ this operator to determine the product  $\mu \tau_{VE}$ . By applying the chain rule this operator may be written as:

$$\begin{aligned} \frac{\partial\eta^h}{\partial\nabla\phi^h} \cdot \nabla &= \left( \frac{\partial\xi}{\partial x} \frac{\partial\eta^h}{\partial\nabla\phi^h} \right) \frac{\partial}{\partial\xi} \\ &= \left( \mathbf{J}^{-1} \frac{\partial\eta^h}{\partial\nabla\phi^h} \right) \frac{\partial}{\partial\xi}. \end{aligned} \tag{5.3.50}$$

Note that the unit of the product  $\mu \tau_{VE}$  is the inverse square of that of the operator  $\partial\eta/\partial\nabla\phi \cdot \nabla$  (the units are  $[\phi]^2[\eta]^{-2}$  and  $[\phi]^{-1}[\eta]$  respectively). We propose to use the discrete approximation of the inverse square of the operator  $\partial\eta/\partial\nabla\phi \cdot \nabla$  as a model for  $\mu \tau_{VE}$ . We take:

$$\mu \tau_{VE} = C h_Q^2 \left\| \left\| \frac{\partial\eta^h}{\partial\nabla\phi^h} \right\| \right\|_G^{-2}, \tag{5.3.51}$$

where *C* is some unitless constant. Remark that a similar approximation technique also employing reference coordinates has been used to derive the stabilization parameter  $\tau_{CL}$ , see e.g., [17]. The variation entropy viscosity thus converts into

$$\nu_{VE} = C h_Q^2 \left\| \left\| \frac{\partial\eta^h}{\partial\nabla\phi^h} \right\| \right\|_G^{-2} \frac{\{\mathcal{R}_{VE}(\eta^h)\}_+}{\eta^h}. \tag{5.3.52}$$

<sup>2</sup> In the linear convection-diffusion case one could use  $\tau_{VE} = \tau_{CL}$  as the differential operators equal (see (5.3.78) in Section 5.3.10).

**Remark 5.3.13.** Viscosity coefficients are usually determined via introducing a shock-capturing quantity and a length-scale. In the small-scale model (5.3.48) the quantity  $\{\mathcal{R}_{\text{VE}}(\eta^h)\}_+ / \eta^h$  serves as shock-capturing quantity (and has unit  $T^{-1}$ ).

The next step is to select a variation entropy. We propose two options. The simplest choice for the variation entropy is to take  $\eta^h = \|\nabla\phi^h\|_2$ . Remark that this variation entropy is objective, see [59]. The corresponding variation entropy viscosity takes the form:

$$\eta^h = \|\nabla\phi^h\|_2 \quad \Rightarrow \quad \nu_{\text{VE}} = C h_Q^2 \left( \frac{\|\nabla\phi^h\|_2}{\|\|\nabla\phi^h\|\|_{\mathbf{G}}} \right) \frac{\{\mathcal{R}_{\text{VE}}(\|\nabla\phi^h\|_2)\}_+}{\|\|\nabla\phi^h\|\|_{\mathbf{G}}}. \quad (5.3.53)$$

Another option is to select  $\eta = \|\|\nabla\phi^h\|\|_{\mathbf{A}}$  which is defined for a positive semi-definite symmetric matrix as  $\|\|\nabla\phi\|\|_{\mathbf{A}}^2 := \nabla\phi^T \mathbf{A} \nabla\phi$ , see (5.2.21c). This is indeed a variation entropy and is rotation invariant whenever

$$\mathbf{A}(\mathbf{R}\mathbf{x}) = \mathbf{R}\mathbf{A}(\mathbf{x})\mathbf{R}^T \quad (5.3.54)$$

for rotation matrix  $\mathbf{R}$ , see [59]. We suggest to take  $\mathbf{A} = \mathbf{G}^{-1}$ , i.e.  $\eta^h = \|\|\nabla\phi^h\|\|_{\mathbf{G}^{-1}} = \|\nabla_{\xi}\phi^h\|_2$ . Trivially  $\mathbf{G}^{-1} = \mathbf{G}^{-1}(x)$  satisfies (5.3.54). The variation entropy viscosity corresponding to this choice is:

$$\eta^h = \|\nabla_{\xi}\phi^h\|_2 \quad \Rightarrow \quad \nu_{\text{VE}} = C h_Q^2 \frac{\{\mathcal{R}_{\text{VE}}(\|\nabla_{\xi}\phi^h\|_2)\}_+}{\|\nabla_{\xi}\phi^h\|_2}. \quad (5.3.55)$$

**Proposition 5.3.14.** On an uniform Cartesian mesh there holds on element  $K$ :

$$\mathbf{G}_K^{-1} = \mathbf{J}_K \mathbf{J}_K^T = \left( \frac{\partial x}{\partial \xi} \right)^2 \mathbf{I}_K, \quad (5.3.56a)$$

$$h_K^2 = \frac{h_Q^2}{d} \|\mathbf{J}_K\|_F^2 = h_Q^2 \left( \frac{\partial x}{\partial \xi} \right)^2, \quad (5.3.56b)$$

$$\|\nabla_{\xi}\phi^h\|_2 = \|\|\nabla\phi^h\|\|_{\mathbf{G}_K^{-1}} = \|\nabla\phi^h\|_2 \frac{\partial x}{\partial \xi}, \quad (5.3.56c)$$

$$\|\|\nabla\phi^h\|\|_{\mathbf{G}} = \|\nabla\phi^h\|_2 \frac{\partial \xi}{\partial x}. \quad (5.3.56d)$$

**Lemma 5.3.15.** On uniform Cartesian quadratic/cubic meshes we have

$$\nu_{\text{VE}}(\|\nabla\phi^h\|_2) = C h_K^2 \frac{\{\mathcal{R}_{\text{VE}}(\|\nabla\phi^h\|_2)\}_+}{\|\nabla\phi^h\|_2}, \quad (5.3.57a)$$

$$\nu_{\text{VE}}(\|\nabla_{\xi}\phi^h\|_2) = C h_Q^2 \frac{\{\mathcal{R}_{\text{VE}}(\|\nabla\phi^h\|_2)\}_+}{\|\nabla\phi^h\|_2}. \quad (5.3.57b)$$

*Proof.* Using (5.3.56b) and (5.3.56d) the first identity is obtained:

$$\begin{aligned} \nu_{\text{VE}}(\|\nabla\phi^h\|_2) &= C h_Q^2 \left( \frac{\|\nabla\phi^h\|_2}{\|\|\nabla\phi^h\|\|_{\mathbf{G}}} \right)^2 \frac{\{\mathcal{R}_{\text{VE}}(\|\nabla\phi^h\|_2)\}_+}{\|\nabla\phi^h\|_2} \\ &= C h_Q^2 \left( \frac{\partial x}{\partial \xi} \right)^2 \frac{\{\mathcal{R}_{\text{VE}}(\|\nabla\phi^h\|_2)\}_+}{\|\nabla\phi^h\|_2} \\ &= C h_K^2 \frac{\{\mathcal{R}_{\text{VE}}(\|\nabla\phi^h\|_2)\}_+}{\|\nabla\phi^h\|_2}. \end{aligned} \quad (5.3.58)$$

The second expression follows via (5.3.20b), (5.3.56a) and (5.3.56c):

$$\begin{aligned}
 \nu_{\text{VE}}(\|\nabla_{\xi}\phi^h\|_2) &= C h_Q^2 \frac{\{\mathcal{R}_{\text{VE}}(\|\nabla_{\xi}\phi^h\|_2)\}_+}{\|\nabla_{\xi}\phi^h\|_2} \\
 &= C h_Q^2 \frac{\left\{ \frac{\mathbf{G}^{-1}\nabla\phi^h}{\|\nabla_{\xi}\phi^h\|_2} \cdot \nabla(\mathcal{R}_{\text{CL}}(\phi^h)) \right\}_+}{\|\nabla\phi^h\|_2 \partial x / \partial \xi} \\
 &= C h_Q^2 \frac{\left\{ \frac{(\partial x / \partial \xi)^2 \nabla\phi^h}{\|\nabla\phi^h\|_2 \partial x / \partial \xi} \cdot \nabla(\mathcal{R}_{\text{CL}}(\phi^h)) \right\}_+}{\|\nabla\phi^h\|_2 \partial x / \partial \xi} \\
 &= C h_Q^2 \frac{\{\mathcal{R}_{\text{VE}}(\|\nabla\phi^h\|_2)\}_+}{\|\nabla\phi^h\|_2}. \tag{5.3.59}
 \end{aligned}$$

□

**Corollary 5.3.16.** *On uniform Cartesian quadratic/cubic meshes we have the identity:*

$$\nu_{\text{VE}}(\|\nabla\phi^h\|_2) = \left(\frac{\partial x}{\partial \xi}\right)^2 \nu_{\text{VE}}(\|\nabla_{\xi}\phi^h\|_2). \tag{5.3.60}$$

To avoid singularities we introduce a regularized variation entropy  $\eta_{\varepsilon}^h$ , see also [59]. Let us define the regularization of variation entropy  $\|\|\nabla\phi^h\|\|_{\mathbf{A}}$  for regularization parameter  $0 < \varepsilon \ll 1$  via:

$$(\eta_{\varepsilon}^h)^2 = \|\|\nabla\phi^h\|\|_{\varepsilon, \mathbf{A}}^2 := \|\|\nabla\phi^h\|\|_{\mathbf{A}}^2 + \varepsilon^2 \frac{\text{Tr}(\mathbf{A})}{d}. \tag{5.3.61}$$

Furthermore we also define:

$$\|\nabla\phi\|_{\varepsilon, 2}^2 := \|\nabla\phi\|_2^2 + \varepsilon^2. \tag{5.3.62}$$

The resulting expressions for the variation entropies chosen above are

$$\eta_{\varepsilon}^h = \|\nabla\phi^h\|_2 \quad \Rightarrow \quad (\eta_{\varepsilon}^h)^2 = \|\nabla\phi^h\|_{\varepsilon, 2}^2, \tag{5.3.63a}$$

$$\begin{aligned}
 \eta_{\varepsilon}^h = \|\nabla_{\xi}\phi^h\|_2 \quad \Rightarrow \quad (\eta_{\varepsilon}^h)^2 &= \|\|\nabla\phi^h\|\|_{\varepsilon, \mathbf{G}^{-1}}^2 \\
 &= \|\nabla_{\xi}\phi^h\|_2^2 + \varepsilon^2 \|\mathbf{J}\|_F^2 / d \\
 &=: \|\nabla_{\xi}\phi^h\|_{\varepsilon_{\xi}, 2}^2, \tag{5.3.63b}
 \end{aligned}$$

where the regularization parameters are related as

$$\varepsilon_{\xi}^2 = \varepsilon^2 \|\mathbf{J}\|_F^2 / d. \tag{5.3.64}$$

We apply the regularization both to the shock-capturing quantities and the prefactors yielding the following regularized variation entropy viscosities:

$$\eta_{\varepsilon}^h = \|\nabla\phi^h\|_{\varepsilon, 2} \quad \Rightarrow \quad \nu_{\text{VE}} = C h_Q^2 \frac{\left( \frac{\|\nabla\phi^h\|_{\varepsilon, 2}}{\|\|\nabla\phi^h\|\|_{\varepsilon, \mathbf{G}}} \right) \{\mathcal{R}_{\text{VE}}(\|\nabla\phi^h\|_{\varepsilon, 2})\}_+}{\|\|\nabla\phi^h\|\|_{\varepsilon, \mathbf{G}}}, \tag{5.3.65a}$$

$$\eta_{\varepsilon}^h = \|\nabla_{\xi}\phi^h\|_{\varepsilon_{\xi}, 2} \quad \Rightarrow \quad \nu_{\text{VE}} = C h_Q^2 \frac{\{\mathcal{R}_{\text{VE}}(\|\nabla_{\xi}\phi^h\|_{\varepsilon_{\xi}, 2})\}_+}{\|\nabla_{\xi}\phi^h\|_{\varepsilon_{\xi}, 2}}. \tag{5.3.65b}$$

**Remark 5.3.17.** Applying solely regularization in (5.3.52) to derive regularized versions of the expressions (5.3.53) and (5.3.55) does not exclude singularities.

The specific regularization choice ensures that Corollary 5.3.16 also holds in the regularized case.

**Corollary 5.3.18.** On uniform Cartesian quadratic/cubic meshes we have the identity:

$$\nu_{VE} (\|\nabla\phi^h\|_{\varepsilon,2}) = \left(\frac{\partial x}{\partial \xi}\right)^2 \nu_{VE} (\|\nabla_{\xi}\phi^h\|_{\varepsilon_{\xi},2}), \quad (5.3.66)$$

where the terms in the brackets after  $\nu_{VE}$  refer to (5.3.65a) and (5.3.65b) respectively.

### 5.3.8 Diffusion matrices

Let us first consider the case  $\eta_{\varepsilon}^h = \|\nabla\phi^h\|_{\varepsilon,2}$ . We use the non-regularized  $\eta^h = \|\nabla\phi^h\|_2$  to derive the diffusion matrix and find via (5.3.45):

$$\mathbf{K}_K = (\nu_{VE})_K \frac{\nabla\phi^h}{\|\nabla\phi^h\|_2} \otimes \frac{\nabla\phi^h}{\|\nabla\phi^h\|_2} \quad (5.3.67)$$

with variation entropy  $(\nu_{VE})_K$  given in (5.3.65a). Using the relation (5.3.44) we see that in this case the matrix  $\mathbf{K}_K$  results in isotropic diffusion:

$$\left(\nabla w^h, \mathbf{K}_K \nabla\phi^h\right)_{L^2(\Omega_K)} = \left(\nabla w^h, (\nu_{VE})_K \nabla\phi^h\right)_{L^2(\Omega_K)}. \quad (5.3.68)$$

In the other situation, i.e.  $\eta_{\varepsilon}^h = \|\nabla_{\xi}\phi^h\|_{\varepsilon_{\xi},2}$ , using also the corresponding non-regularized variation entropy yields for the diffusion matrix:

$$\mathbf{K}_K = (\nu_{VE})_K \frac{\mathbf{G}^{-1}\nabla\phi^h}{\|\nabla_{\xi}\phi^h\|_2} \otimes \frac{\mathbf{G}^{-1}\nabla\phi^h}{\|\nabla_{\xi}\phi^h\|_2}. \quad (5.3.69)$$

with variation entropy  $(\nu_{VE})_K$  given in (5.3.65b). The resulting diffusion contribution in the weak form is based on local gradients:

$$\begin{aligned} \left(\nabla w^h, \mathbf{K}_K \nabla\phi^h\right)_{L^2(\Omega_K)} &= \left(\nabla w^h, (\nu_{VE})_K \frac{\mathbf{J}\nabla_{\xi}\phi^h}{\|\nabla_{\xi}\phi^h\|_2} \|\nabla_{\xi}\phi^h\|_2\right)_{L^2(\Omega_K)} \\ &= \left(\nabla_{\xi} w^h, (\nu_{VE})_K \nabla_{\xi}\phi^h\right)_{L^2(\Omega_K)}. \end{aligned} \quad (5.3.70)$$

**Remark 5.3.19.** We note that Guermond and Nazarov [84] use reference coordinates to enforce the maximum principle. They observe that local reference coordinates can provide more control over the gradients.

**Theorem 5.3.20.** On uniform Cartesian quadratic/cubic meshes the choices  $\eta = \|\nabla\phi\|_{\varepsilon,2}$  and  $\eta = \|\nabla_{\xi}\phi\|_{\varepsilon_{\xi},2}$  coincide.

*Proof.* This is a direct consequence of Corollary 5.3.18:

$$\begin{aligned} \left(\nabla_{\xi} w^h, (\nu_{VE})_K (\|\nabla_{\xi}\phi^h\|_{\varepsilon_{\xi},2}) \nabla_{\xi}\phi^h\right)_{L^2(\Omega_K)} &= \\ \left(\nabla w^h, (\nu_{VE})_K (\|\nabla_{\xi}\phi^h\|_{\varepsilon_{\xi},2}) \left(\frac{\partial x}{\partial \xi}\right)^2 \nabla\phi^h\right)_{L^2(\Omega_K)} &= \\ \left(\nabla w^h, (\nu_{VE})_K (\|\nabla\phi^h\|_{\varepsilon,2}) \nabla\phi^h\right)_{L^2(\Omega_K)}. \end{aligned} \quad (5.3.71)$$

□

### 5.3.9 Complete semi-discrete formulations

By substituting the diffusion terms (5.3.68) and (5.3.70) with corresponding variation entropy viscosities (5.3.65) into (5.3.44) and using the SUPG model of (5.3.31) in (5.3.15), we arrive at the following variational formulation:

find  $\phi^h \in \mathcal{W}^h$  such that for all  $w^h \in \mathcal{W}^h$ :

$$\begin{aligned}
 & \underbrace{\left( w^h, \partial_t \phi^h \right)_{L^2(\Omega)} - \left( \nabla w^h, \mathbf{f}(\phi^h) \right)_{L^2(\Omega)}}_{\text{Galerkin}} + \underbrace{\sum_K \left( (\tau_{\text{CL}})_K \frac{\partial \mathbf{f}}{\partial \phi^h} \cdot \nabla w^h, \mathcal{R}_{\text{CL}}(\phi^h) \right)_{L^2(\Omega_K)}}_{\text{Stabilization}} \\
 & + \left\{ \begin{array}{l} \underbrace{\sum_K (\nabla w^h, (\nu_{\text{VE}})_K \nabla \phi^h)_{L^2(\Omega_K)}}_{\substack{\text{Discontinuity capturing} \\ \text{in physical coordinates}}} \quad \text{if } \eta_\varepsilon^h = \|\nabla \phi^h\|_{\varepsilon,2} \\ \underbrace{\sum_K (\nabla_\xi w^h, (\nu_{\text{VE}})_K \nabla_\xi \phi^h)_{L^2(\Omega_K)}}_{\substack{\text{Discontinuity capturing} \\ \text{in reference coordinates}}} \quad \text{if } \eta_\varepsilon^h = \|\nabla_\xi \phi^h\|_{\varepsilon_\xi,2} \end{array} \right\} = 0 \quad (5.3.72a)
 \end{aligned}$$

where the variation entropy viscosity is:

$$\nu_{\text{VE}} = \begin{cases} C h_Q^2 \left( \frac{\|\nabla \phi^h\|_{\varepsilon,2}}{\|\nabla \phi^h\|_{\varepsilon,\mathbf{G}}} \right) \frac{\{\mathcal{R}_{\text{VE}}(\|\nabla \phi^h\|_{\varepsilon,2})\}_+}{\|\nabla \phi^h\|_{\varepsilon,\mathbf{G}}} & \text{if } \eta_\varepsilon^h = \|\nabla \phi^h\|_{\varepsilon,2} \\ C h_Q^2 \frac{\{\mathcal{R}_{\text{VE}}(\|\nabla_\xi \phi^h\|_{\varepsilon_\xi,2})\}_+}{\|\nabla_\xi \phi^h\|_{\varepsilon_\xi,2}} & \text{if } \eta_\varepsilon^h = \|\nabla_\xi \phi^h\|_{\varepsilon_\xi,2}. \end{cases} \quad (5.3.72b)$$

We conclude that the variation entropy optimality projector with the proper modeling choices *naturally* augments the VMS method with a discontinuity capturing term:

$$\boxed{\text{VMS} + \text{VE} \rightsquigarrow \text{DC}}$$

This proves the conjecture of Bazilevs et al. [19]:

*‘the multiscale frame-work with a proper set of optimality conditions is the right underlying theoretical structure that may more naturally lead to discontinuity capturing formulations.’*

**Remark 5.3.21.** It is possible to set a maximum to the introduced viscosity, see e.g., [85]. Based on first-order upwind techniques (which yields in some cases a monotone method) a natural choice would be to take the maximum viscosity as:

$$\nu_{\text{max}} = C_{\text{max}} h_K \left\| \frac{\partial \mathbf{f}}{\partial \phi} \right\|_2, \quad (5.3.73)$$

where  $C_{\text{max}}$  is some constant.

5.3.10 *The convection-diffusion problem*

In the preceding part of this section we have solely focused on the hyperbolic case. As claimed, the parabolic case is a straightforward extension, which we demonstrate here using the convection-diffusion model problem.

Let  $\phi_0 = \phi_0(\mathbf{x})$ , divergence-free velocity field  $\mathbf{a} = \mathbf{a}(\phi)$  and diffusivity  $\kappa \geq 0$  be given. The problem reads:

*find*  $\phi = \phi(\mathbf{x}, t) : \Omega \times \mathbb{R}^+ \rightarrow \mathbb{R}$  *such that:*

$$\partial_t \phi + \mathbf{a} \cdot \nabla \phi - \kappa \Delta \phi = 0 \quad \text{in } \Omega \times \mathcal{I}, \quad (5.3.74a)$$

$$\phi = g \quad \text{on } \partial\Omega, \quad (5.3.74b)$$

$$\phi(\mathbf{x}, 0) = \phi_0(\mathbf{x}) \quad \text{in } \Omega. \quad (5.3.74c)$$

The standard weak formulation is:

*find*  $\phi \in H_g^1(\Omega)$  *such that for all*  $w \in H_0^1(\Omega)$ :

$$(w, \partial_t \phi + \mathbf{a} \cdot \nabla \phi)_{L^2(\Omega)} + (\nabla w, \kappa \nabla \phi)_{L^2(\Omega)} = 0 \quad \text{in } \Omega \times \mathcal{I}, \quad (5.3.75a)$$

$$\phi(\mathbf{x}, 0) = \phi_0(\mathbf{x}) \quad \text{in } \Omega, \quad (5.3.75b)$$

where we use the standard notation for the function spaces with  $H_g^1(\Omega) := \{w \in H^1(\Omega) : w = g \text{ on } \partial\Omega\}$ . We note that this problem suits the abstract framework with function spaces  $\mathcal{W} = H_g^1(\Omega)$ ,  $\mathcal{W}^* = H^{-1}(\Omega)$ . The linear operator  $\mathcal{L} : H_g^1(\Omega) \rightarrow H^{-1}(\Omega)$  is defined as

$${}_{H^{-1}(\Omega)} \langle \mathcal{L}\phi, w \rangle_{H_0^1(\Omega)} = (w, \partial_t \phi)_{L^2(\Omega)} - (\nabla w, \mathbf{a}\phi - \kappa \nabla \phi)_{L^2(\Omega)}. \quad (5.3.76)$$

Applying the methodology results in the following method:

*find*  $\phi^h \in \mathcal{W}_g^h$  *such that for all*  $w^h \in \mathcal{W}_0^h$ :

$$\begin{aligned} & \underbrace{(w^h, \partial_t \phi^h + \mathbf{a} \cdot \nabla \phi^h)_{L^2(\Omega)} + (\nabla w^h, \kappa \nabla \phi^h)_{L^2(\Omega)}}_{\text{Galerkin contribution}} \\ & + \underbrace{\sum_K ((\tau_{\text{CL}})_K (\mathbf{a} \cdot \nabla w^h + \kappa \Delta w^h), \mathcal{R}_{\text{CL}} \phi^h)_{L^2(\Omega_K)}}_{\text{VMS stabilization}} \\ & + \left\{ \begin{array}{l} \underbrace{\sum_K (\nabla w^h, \nu_K \nabla \phi^h)_{L^2(\Omega_K)}}_{\text{Discontinuity capturing}} \quad \text{if } \eta_\varepsilon^h = \|\nabla \phi^h\|_{\varepsilon, 2} \\ \text{in physical coordinates} \\ \underbrace{\sum_K (\nabla_\xi w^h, \nu_K \nabla_\xi \phi^h)_{L^2(\Omega_K)}}_{\text{Discontinuity capturing}} \quad \text{if } \eta_\varepsilon^h = \|\nabla_\xi \phi^h\|_{\varepsilon_\xi, 2} \\ \text{in reference coordinates} \end{array} \right\} = 0 \quad (5.3.77a) \end{aligned}$$

where the variation entropy viscosity is:

$$\nu_{\text{VE}} = \begin{cases} C h_Q^2 \left( \frac{\|\nabla \phi^h\|_{\varepsilon,2}}{\|\nabla \phi^h\|_{\varepsilon,\mathbf{G}}} \right) \frac{\{\mathcal{L}(\|\nabla \phi^h\|_{\varepsilon,2})\}_+}{\|\nabla \phi^h\|_{\varepsilon,\mathbf{G}}} & \text{if } \eta_\varepsilon^h = \|\nabla \phi^h\|_{\varepsilon,2} \\ C h_Q^2 \frac{\{\mathcal{L}(\|\nabla_\xi \phi^h\|_{\varepsilon_\xi,2})\}_+}{\|\nabla_\xi \phi^h\|_{\varepsilon_\xi,2}} & \text{if } \eta_\varepsilon^h = \|\nabla_\xi \phi^h\|_{\varepsilon_\xi,2}. \end{cases} \quad (5.3.77b)$$

We wish to emphasize that convection-diffusion and variation entropy operators coincide:

$$\mathcal{R}_{\text{CL}} \phi^h = \mathcal{L} \phi^h, \quad (5.3.78a)$$

$$\mathcal{R}_{\text{VE}} \eta^h = \mathcal{L} \eta^h. \quad (5.3.78b)$$

The element-wise stabilization parameter  $(\tau_{\text{CL}})_K$  is defined as in [57].

### 5.3.11 Connection to the YZ $\beta$ method

In order to establish the connection to the YZ $\beta$  method [19] we present an alternative small-scale model. Instead of using the model (5.3.48), one can use an approximation. Using the definition (5.3.20b) we may write:

$$\widehat{\{\eta'\}}_{-\text{VE}} = -\tau_{\text{VE}} \left\{ \frac{\partial \eta^h}{\partial \nabla \phi^h} \cdot \nabla (\mathcal{R}_{\text{CL}}(\phi^h)) \right\}_+. \quad (5.3.79)$$

Again using (5.3.50) we now approximate (5.3.79) as a residual-based model via:

$$\begin{aligned} \widehat{\{\eta'\}}_{-\text{CL}} &= -\tau_{\text{VE}} h_Q^{-1} \left\| \mathbf{J}^{-1} \frac{\partial \eta^h}{\partial \nabla \phi^h} \right\|_2 |\mathcal{R}_{\text{CL}}(\phi^h)| \\ &= -\tau_{\text{VE}} h_Q^{-1} \left\| \frac{\partial \eta^h}{\partial \nabla \phi^h} \right\|_{\mathbf{G}} |\mathcal{R}_{\text{CL}}(\phi^h)|, \end{aligned} \quad (5.3.80)$$

Using the model (5.3.51) for  $\mu \tau_{\text{VE}}$ , the corresponding variation entropy viscosity takes the form:

$$\begin{aligned} \nu_{\text{CL}} &= \mu \tau_{\text{VE}} h_Q^{-1} \left\| \frac{\partial \eta^h}{\partial \nabla \phi^h} \right\|_{\mathbf{G}} \frac{|\mathcal{R}_{\text{CL}}(\phi^h)|}{\eta^h} \\ &= C h_Q \left\| \frac{\partial \eta^h}{\partial \nabla \phi^h} \right\|_{\mathbf{G}}^{-1} \frac{|\mathcal{R}_{\text{CL}}(\phi^h)|}{\eta^h}, \end{aligned} \quad (5.3.81)$$

where the subscript refers to the conservation law residual. In this case the variation entropy viscosity (5.3.81) scales with the residual of the conservation law but generally does not vanish when the variation entropy condition is satisfied.

Using the same large-scale variation entropies, i.e.  $\eta^h = \|\nabla \phi^h\|_2$  and  $\eta^h = \|\nabla_\xi \phi^h\|_2$ , we get the expressions:

$$\eta^h = \|\nabla \phi^h\|_2 \quad \Rightarrow \quad \nu_{\text{CL}} = C h_Q \frac{|\mathcal{R}_{\text{CL}}(\phi^h)|}{\|\nabla \phi^h\|_{\mathbf{G}}}, \quad (5.3.82a)$$

$$\eta^h = \|\nabla_\xi \phi^h\|_2 \quad \Rightarrow \quad \nu_{\text{CL}} = C h_Q \frac{|\mathcal{R}_{\text{CL}}(\phi^h)|}{\|\nabla_\xi \phi^h\|_2}. \quad (5.3.82b)$$

**Lemma 5.3.22.** *On uniform Cartesian quadratic/cubic meshes we have*

$$\nu_{\text{CL}} (\|\nabla\phi^h\|_2) = C h_K \frac{|\mathcal{R}_{\text{CL}}(\phi^h)|}{\|\nabla\phi^h\|_2}, \quad (5.3.83a)$$

$$\nu_{\text{CL}} (\|\nabla_{\xi}\phi^h\|_2) = C h_Q \left(\frac{\partial x}{\partial \xi}\right)^{-1} \frac{|\mathcal{R}_{\text{CL}}(\phi^h)|}{\|\nabla\phi^h\|_2}. \quad (5.3.83b)$$

*Proof.* The proof is similar to that of Lemma 5.3.15 and uses (5.3.56b)-(5.3.56d).  $\square$

**Corollary 5.3.23.** *On uniform Cartesian quadratic/cubic meshes we have the identity:*

$$\nu_{\text{CL}} (\|\nabla\phi^h\|_2) = \left(\frac{\partial x}{\partial \xi}\right)^2 \nu_{\text{CL}} (\|\nabla_{\xi}\phi^h\|_2). \quad (5.3.84)$$

The regularized versions of the variation entropy viscosities are:

$$\eta_{\varepsilon}^h = \|\nabla\phi^h\|_{\varepsilon,2} \Rightarrow \nu_{\text{CL}} = C h_Q \frac{|\mathcal{R}_{\text{CL}}(\phi^h)|}{\|\nabla\phi^h\|_{\varepsilon,\mathbf{G}}}, \quad (5.3.85a)$$

$$\eta_{\varepsilon}^h = \|\nabla_{\xi}\phi^h\|_{\varepsilon_{\xi},2}^2 \Rightarrow \nu_{\text{CL}} = C h_Q \frac{|\mathcal{R}_{\text{CL}}(\phi^h)|}{\|\nabla_{\xi}\phi^h\|_{\varepsilon_{\xi},2}}. \quad (5.3.85b)$$

**Corollary 5.3.24.** *On uniform Cartesian quadratic/cubic meshes we have the identity:*

$$\nu_{\text{CL}} (\|\nabla\phi^h\|_{\varepsilon,2}) = \left(\frac{\partial x}{\partial \xi}\right)^2 \nu_{\text{CL}} (\|\nabla_{\xi}\phi^h\|_{\varepsilon_{\xi},2}). \quad (5.3.86)$$

**Theorem 5.3.25.** *On uniform Cartesian quadratic/cubic meshes the choices  $\eta = \|\nabla\phi\|_{\varepsilon,2}$  and  $\eta = \|\nabla_{\xi}\phi\|_{\varepsilon_{\xi},2}$  coincide.*

Substitution yields weak formulation (5.3.72) where the variation entropy viscosity are now given by (5.3.85). As variation entropy we take  $\eta^h = \|\nabla\phi^h\|_2$  which yields:

$$\left(\nabla w^h, \mathbf{K}_K \nabla \phi^h\right)_{L^2(\Omega_K)} = \left(\nabla w^h, C h_Q \frac{|\mathcal{R}_{\text{CL}}(\phi^h)|}{\|\nabla\phi^h\|_{\varepsilon,\mathbf{G}}} \nabla \phi^h\right)_{L^2(\Omega_K)}. \quad (5.3.87)$$

This is the (regularized version of the) discontinuity capturing term used by Akkerman et al. [3] for the level-set convection equation. On Cartesian uniform meshes it reduces to

$$\left(\nabla w^h, \mathbf{K}_K \nabla \phi^h\right)_{L^2(\Omega_K)} = \left(\nabla w^h, C h_K \frac{|\mathcal{R}_{\text{CL}}(\phi^h)|}{\|\nabla\phi^h\|_{\varepsilon,2}} \nabla \phi^h\right)_{L^2(\Omega_K)}. \quad (5.3.88)$$

For convection-diffusion problems this coincides with the  $YZ\beta$  discontinuity capturing operator [19] with parameter  $\beta = 1$ . This term is used for non-uniform meshes as well.

**Remark 5.3.26.** *Remark that  $YZ\beta$  discontinuity capturing and the beyond SUPG discontinuity capturing are nearly identical in a one-dimensional pure convection case with stabilization parameter  $h/a$ .*

**Remark 5.3.27.** *The discontinuity capturing operator  $YZ\beta$  with parameter  $\beta = 2$ , in contrast to  $\beta = 1$ , does not fit in the presented framework. The fact that the choice  $\beta = 1$  is preferred over  $\beta = 2$ , see [19], confirms the viability of the presented theory.*

## 5.4 NUMERICAL COMPARISON

In this section we evaluate the numerical methods on benchmark problems. All the computations are performed with `TIGAR` [124]. We employ  $C^1$ -continuous quadratic NURBS and use the generalized- $\alpha$  time-integrator with the parameter  $\rho_\infty = 1.0$ . Note that this is the only time-integrator within the generalized- $\alpha$  family linked to correct energy behavior, see e.g., [57]. The regularization parameter is taken as  $\varepsilon^2 = 10^{-2}$ .

We show the results of using

1. the well-known SUPG method.
2. the  $YZ\beta$  method with  $\beta = 1$ . The connection of this method with the developed framework is presented in Section 5.3.11.
3. the new method which is summarized in Section 5.3.9.

All the computations are performed on Cartesian meshes. Here the choices  $\eta = \|\nabla\phi\|_2$  and  $\eta = \|\nabla_{\xi}\phi\|_2$  coincide (Theorem 5.3.20). Non-Cartesian computations may be subject of another work.

First we evaluate the convergence behavior of the new methods on a smooth pure advection problem. Then we evaluate the methods on two nonlinear benchmark problems: (i) the Buckley-Leverett equation with gravity and (ii) the KPP rotating wave problem. Both tests involve non-convex fluxes and are challenging since the corresponding solutions have a two-dimensional composite wave-structure. These problems have been employed in other works concerning discontinuity capturing mechanisms, see e.g., [19, 43, 85, 99, 137]. We refer the reader for a comparison of the results to those works.

## 5.4.1 Convergence for smooth solutions

In this first numerical experiment we consider a smooth profile to test the convergence of the methods. The problem reads:

$$\partial_t\phi + \nabla \cdot \mathbf{f} = 0, \quad (5.4.1a)$$

$$\phi(\mathbf{x}, 0) = \begin{cases} \exp\left(-\frac{1}{1-r^2}\right) & \text{if } r < 1.0, \\ 0.0 & \text{otherwise,} \end{cases} \quad (5.4.1b)$$

$$\mathbf{f}(\phi) = \mathbf{a}\phi. \quad (5.4.1c)$$

with radius  $r = \sqrt{x^2 + y^2}$ .

The convection velocity field is constant and has value  $\mathbf{a} = (0.1, 0.15)$ . The time-step size is chosen as  $\Delta t = 4h_K$  and we take  $C = 0.5$ . Figure 5.3 shows second-order/third-order convergence in the  $L^2$ -norm for each of the three methods which for finer meshes yields second-order convergence due to the choice of the time-integrator.

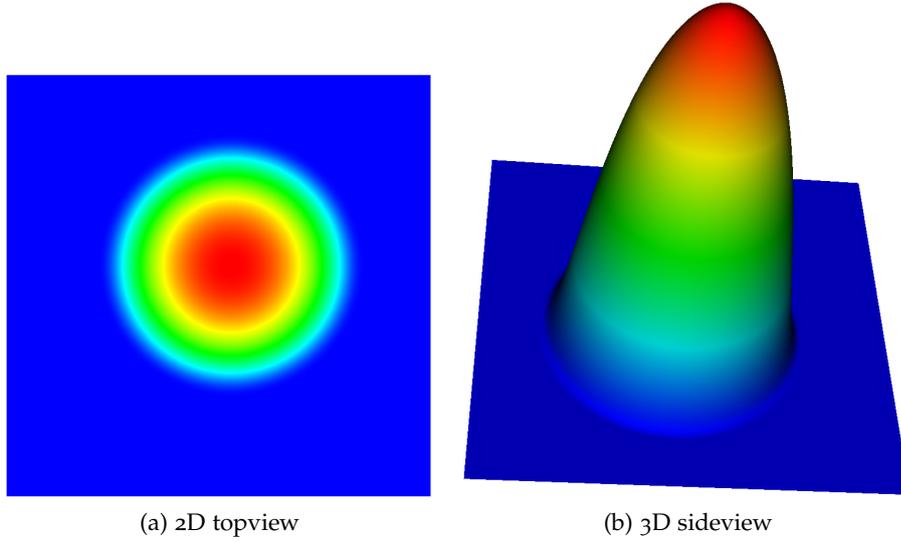


Figure 5.2: Smooth solution problem with quadratic NURBS. The final solution at  $t = 1.0$ .

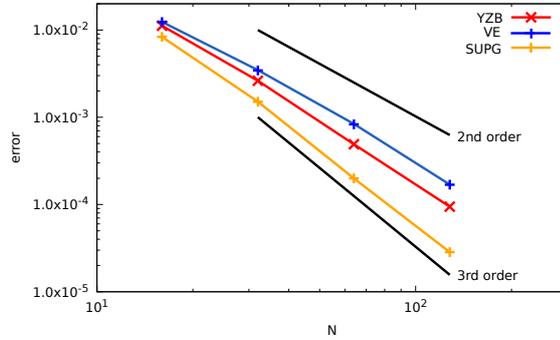


Figure 5.3:  $L^2$  convergence of the smooth solution problem with quadratic NURBS.

#### 5.4.2 Buckley-Leverett with gravity

The gravitational Buckley-Leverett problem with a Riemann initial configuration reads:

find  $\phi = \phi(x, t) : \Omega \times \mathcal{I} \rightarrow \mathbb{R}$  such that:

$$\partial_t \phi + \nabla \cdot \mathbf{f} = 0, \quad (5.4.2a)$$

$$\phi(x, 0) = \begin{cases} 1.0 & \text{if } x^2 + y^2 \leq 0.5, \\ 0.0 & \text{otherwise,} \end{cases} \quad (5.4.2b)$$

$$\mathbf{f}(\phi) = \left( \frac{\phi^2}{\phi^2 + (1 - \phi)^2}, \frac{\phi^2(1 - 5(1 - \phi)^2)}{\phi^2 + (1 - \phi)^2} \right). \quad (5.4.2c)$$

The Buckley-Leverett problem emerges from a two-phase immiscible incompressible fluid problem. It represents a saturation equation in which gravitational effects are incorporated. This results in different fluxes in both spatial directions. The problem

has also been considered in [43, 85, 127]. The solution is advanced in time until  $t = 0.5$ .

All computations are performed on a  $100 \times 100$  mesh with time-step size  $\Delta t = 0.01$ . We show in Figures 5.4-5.6 the solution profiles at final time  $t = 0.5$ . The gray scale of the viscosity magnitude is per Figure chosen such that the location of the diffusion becomes most apparent. The results of the SUPG method contain excessive oscillations. The discontinuity capturing viscosity based on the variation entropy condition focuses on the sharp layer, whereas basing it on the residual of the conservation law spreads it out.

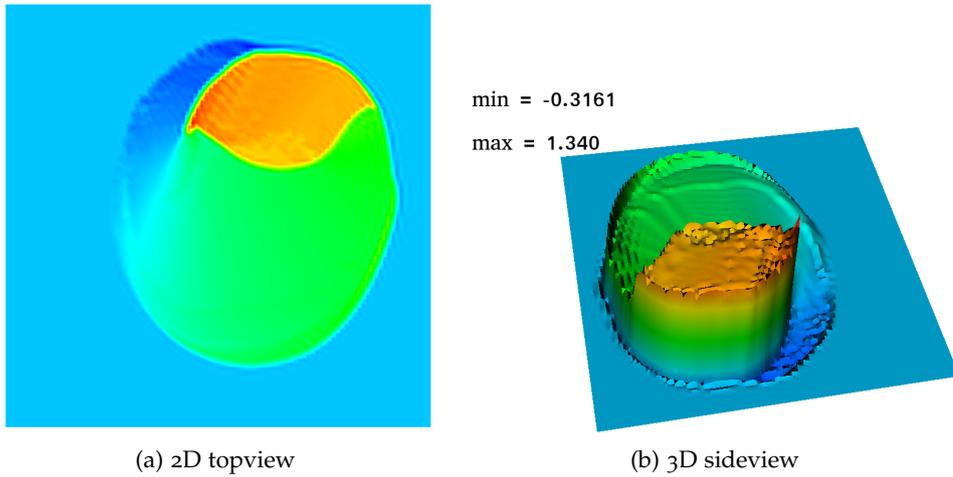


Figure 5.4: Buckley-Leverett problem, the solution at final time  $t = 0.5$  using the SUPG method.

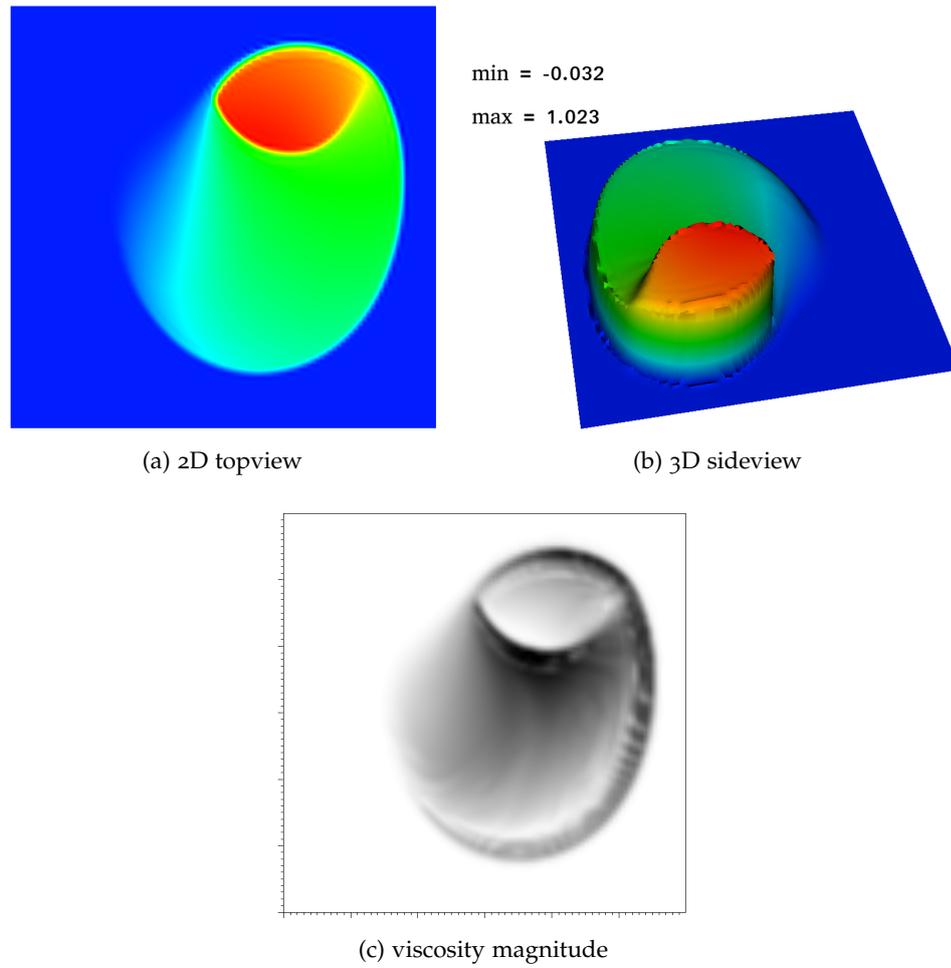


Figure 5.5: Buckley-Leverett problem, the solution at final time  $t = 0.5$  using the  $YZ\beta$  method with constant  $C = 0.25$ .

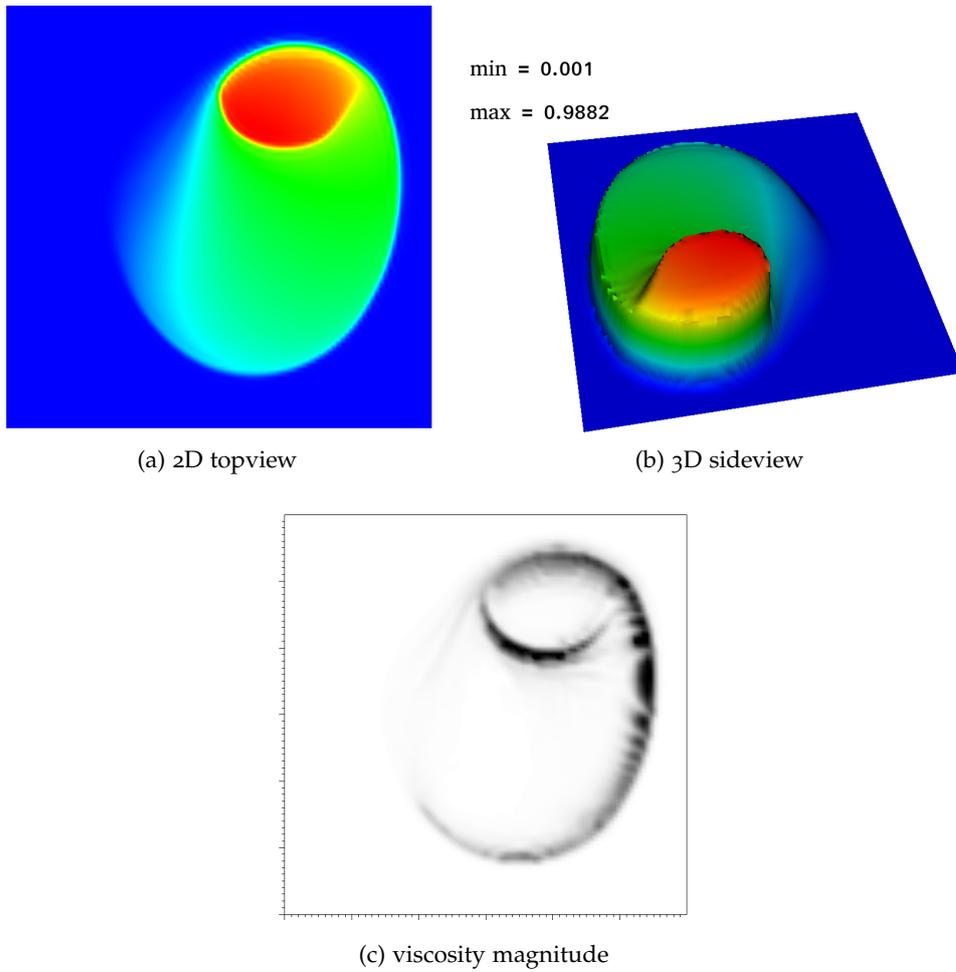


Figure 5.6: Buckley-Leverett problem, the solution at final time  $t = 0.5$  using the VE method with constant  $C = 0.25$ .

## 5.4.3 KPP rotating wave

The KPP rotating wave problem is:

find  $\phi = \phi(\mathbf{x}, t) : \Omega \times \mathcal{I} \rightarrow \mathbb{R}$  such that:

$$\partial_t \phi + \nabla \cdot \mathbf{f} = 0, \quad (5.4.3a)$$

$$\phi(\mathbf{x}, 0) = \begin{cases} 3.5\pi & \text{if } x^2 + y^2 \leq 1, \\ 0.25\pi & \text{otherwise,} \end{cases} \quad (5.4.3b)$$

$$\mathbf{f}(\phi) = (\sin \phi, \cos \phi). \quad (5.4.3c)$$

The test case was proposed in [137] and is named after the authors Kurganov, Petrova, and Popov. Several reconstruction schemes, e.g., central-upwind schemes as WENO5, Minmod 2 and SuperBee, are not successful for this test.

All computations are performed on a 100x100 mesh with time-step size  $\Delta t = 0.01$ . Figures 5.7-5.9 show the solution profiles of the various methods at final time  $t = 1$ .

The results of the SUPG method display sharp layers with excessive oscillations. The solution quality improves greatly when using any of the other methods. Again, the discontinuity capturing viscosity is more localized near the sharp layers when it is based on the variation entropy condition (displayed in Figure 5.9) than on the residual of the conservation law (see Figure 5.8). We see at some locations a viscosity value that is higher than required. In Figure 5.10 we show the results of using a maximum for the viscosity via equation (5.3.73) with  $C_{\max} = 1.0$ . The gray scale of the viscosity magnitude of Figures 5.9 and 5.10 is the same to highlight the effect of using a maximum viscosity. The overly diffusive regions are now removed and the resulting profile has minimal smearing and the spurious oscillations are virtually absent.

**Remark 5.4.1.** *In this testcase it is apparent that the viscosity of the new method is active in regions where variation entropy is created. Gibbs oscillations appear right next to the discontinuity and this is where the viscosity acts. Note that the viscosity is absent at the location of the shockwave itself. This is in contrast to the entropy viscosity method [85] in which the viscosity focuses on the shockwave itself rather than on the oscillations next to it.*

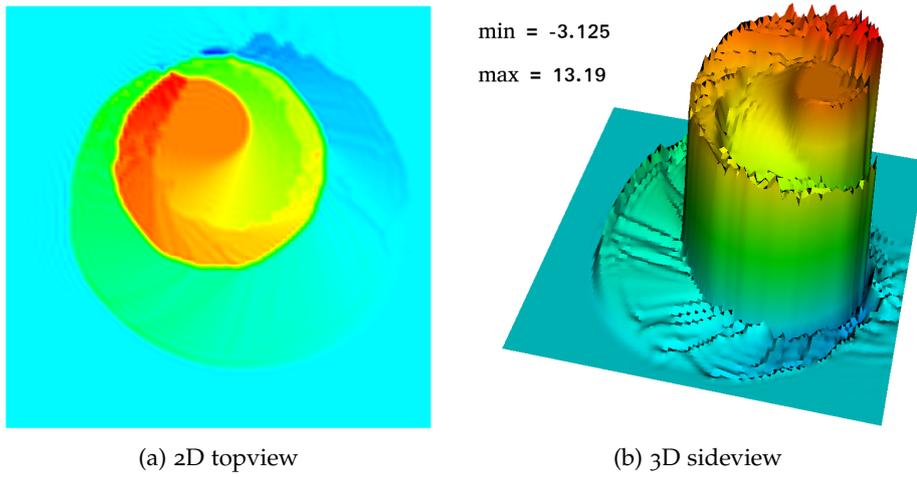
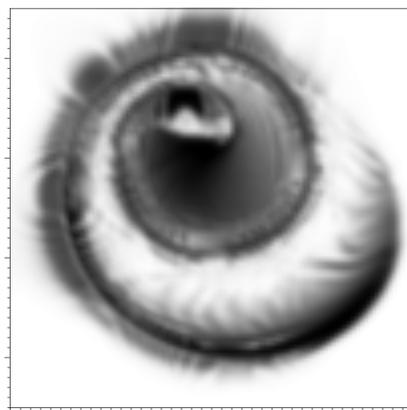
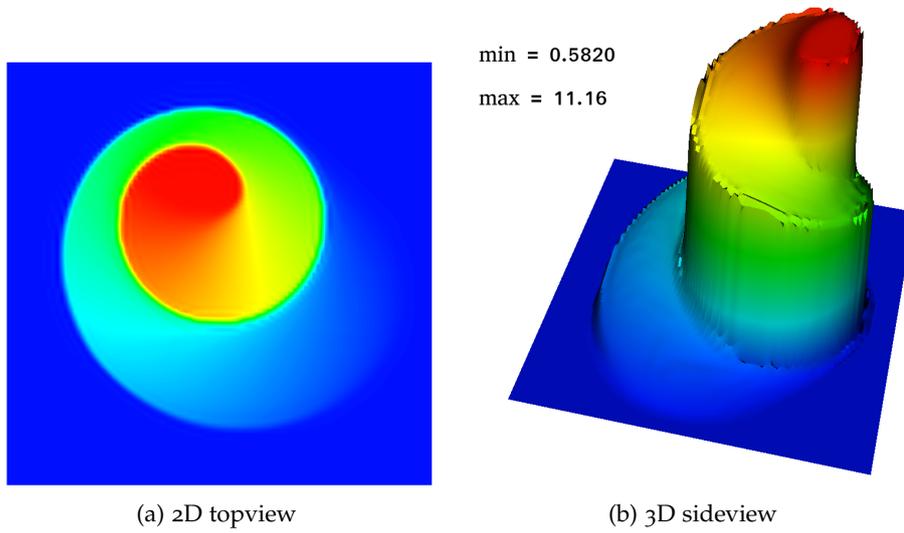


Figure 5.7: KPP rotating wave problem, the solution at final time  $t = 1.0$  using the SUPG method.



(c) viscosity magnitude

Figure 5.8: KPP rotating wave problem, the solution at final time  $t = 1.0$  using the  $YZ\beta$  method with constant  $C = 0.25$ .

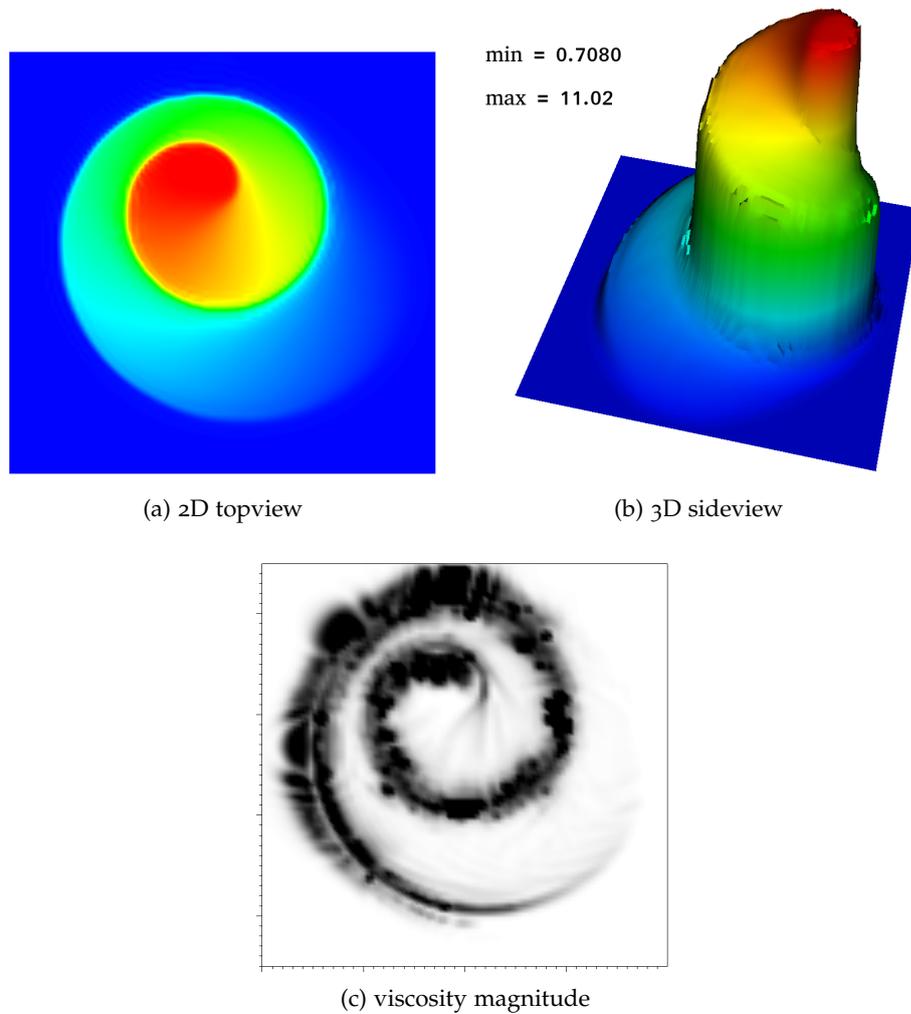


Figure 5.9: KPP rotating wave problem, the solution at final time  $t = 1.0$  using the VE method with constant  $C = 0.25$ .

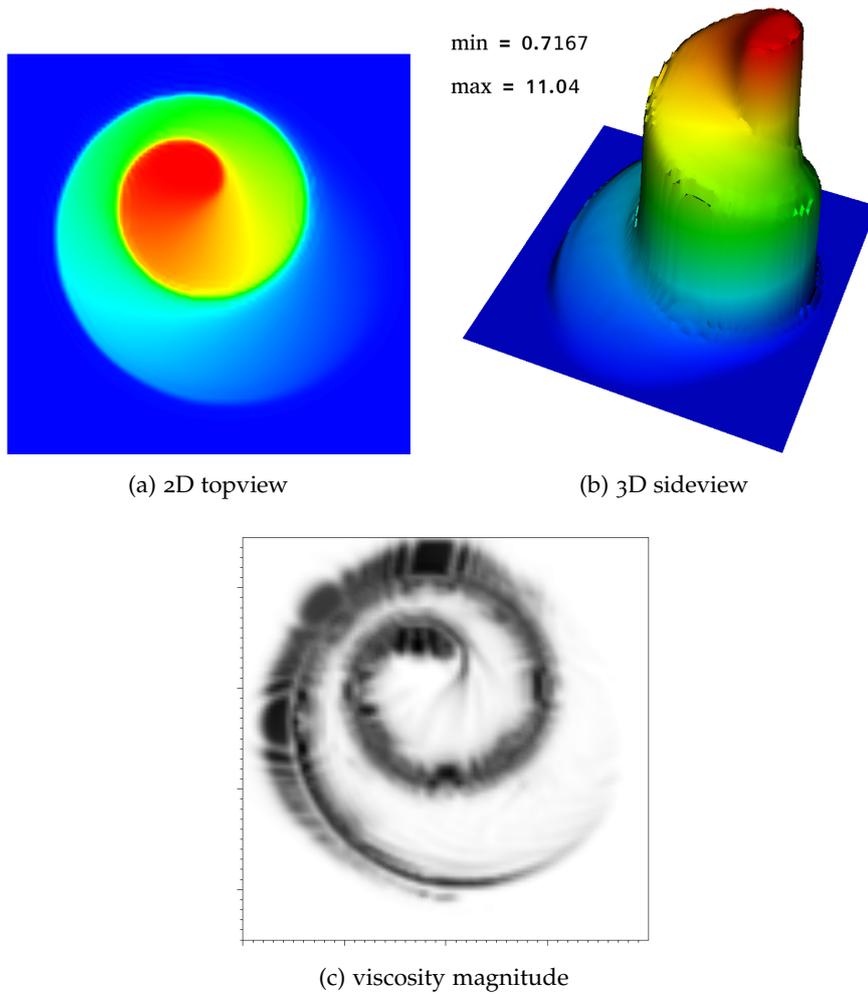


Figure 5.10: KPP rotating wave problem, the solution at final time  $t = 1.0$  using the VE method with constant  $C = 0.25$  and using the maximum viscosity (5.3.73).

## 5.5 CONCLUSIONS

In this chapter we have presented a general framework for discontinuity capturing mechanisms. The framework does not employ *ad hoc* devices which is, to the best knowledge of the authors, in contrast to previous discontinuity capturing methods. The developed theory contains two key ingredients, namely variation entropy theory and variational multiscale analysis. Variation entropy provides us the location of the viscosity and VMS models this viscosity via the missing scales. Merging the variation entropy concept into the variational multiscale method naturally equips the variational multiscale method with a discontinuity capturing term.

The discontinuity capturing viscosity is based on the variation entropy condition. In smooth regions the variation entropy relation is governed with an equality, and this is where the discontinuity capturing term vanishes. Near sharp layers, however, the variation entropy relation becomes an inequality. Here the discontinuity capturing term switches on; dissipation based on the variation entropy production is added to the formulation.

Many spurious oscillation diminishing methods are isotropic of nature or add diffusion in the crosswind direction. The discontinuity capturing viscosity acts in the direction identified by the change of the variation entropy. We believe that this is a natural direction, since this is where sharp layers are expected. In particular cases the viscosity reduces to an isotropic one.

The steps of the framework to arrive at a discontinuity capturing term can be summarized as follows:

1. Regularize the conservation law.
2. Perform a multiscale split and subsequently take the limit of regularization parameter to zero.
3. Select a projector based on the variation entropy condition.
4. Select a small-scale variation entropy model.
5. Compute the variation entropy viscosity.
6. Select a large-scale variation entropy.

We have tested the new discontinuity capturing method on nonlinear benchmark problems. The computations are performed with quadratic NURBS. The numerical results are virtually oscillation-free and have minimal smearing. Compared to the well-known  $YZ\beta$  method [19], the diffusion of the new discontinuity method is more localized *near* sharp layers. These are the locations where variation entropy can be created. We emphasize that the diffusion should not be added at the shock but right next to it.

This chapter sheds light on the different concepts of entropy solutions, the total variation diminishing property and their relation to discontinuity capturing mechanisms. In particular, it establishes a connection between total variation/variation entropy and discontinuity capturing.

The current framework provides some insight into discontinuity capturing techniques, however we certainly do not claim that it is sufficient in this context. There are several openings.

- The first concerns the choice of the variation entropy. Numerical results indicate that taking the 2-norm of the gradient leads to good behavior. Improvement might be achieved with another choice of the variation entropy.
- Another point that deserves interest is the small-scale variation entropy model. We have taken the simplest options and perhaps at this point progress can be made.
- Furthermore, a numerical investigation of the performance of the new method on curved/non-Cartesian meshes could be investigated.

Summarizing, this work proposes a novel paradigm for the construction of discontinuity capturing operators. We think that the framework has a more fundamental mathematical foundations than previously proposed methods. The reason is that it naturally emerges from the conservation law and does not contain ad hoc devices. This, together with the good numerical results illustrate the viability of the framework. The basic questions of discontinuity capturing operators, i.e. (i) where to add diffusion?, and (ii) how much diffusion should be added? are answered. The results of this work indicate that diffusion should be added there where variation entropy is being produced with an amount that scales with the variation entropy production.

We close this chapter with the following note. The variational multiscale method has proven to be a powerful tool for the simulation of turbulent flows, as displayed in the seminal work [17]. In the current work we have demonstrated that, in addition to this,

*the variational multiscale method is suitable to deal with sharp layers/discontinuities.*

The reason for this is simple: both turbulence and shock wave problems contain features that do not 'fit' on a coarse mesh; the variational multiscale framework incorporates these features into the numerical method.



## 5.A AN ALTERNATIVE OPTIMALITY PROJECTOR

Here we present an alternative projector that directly penalizes violation of the variation entropy condition. Consider the minimization problem:

find  $\phi^h \in \mathcal{W}^h$  such that:

$$\mathcal{L}(\phi - \phi^h) = \inf_{\theta^h \in \mathcal{K}^h} \mathcal{L}(\phi - \theta^h), \quad (5.A.1)$$

where the constraint set reads:

$$\mathcal{K}^h := \left\{ \phi^h \in \mathcal{W}^h : (v^h, \mathcal{R}_{\text{VE}} \eta^h)_{L^2(\Omega)} \leq 0 \quad \text{for all } v^h \in \mathcal{V}^h \right\}. \quad (5.A.2)$$

We proceed by opening the solution space with a penalty approach. We define the projector by

$$\mathcal{P}^h \phi = \operatorname{argmin}_{\phi^h \in \mathcal{W}^h} \left\{ \frac{1}{2} \|\phi - \phi^h\|_{\mathcal{W}}^2 + \frac{1}{2} \left\| \sqrt{\mu} \tau_{\text{VE}} \{ \mathcal{R}_{\text{VE}} \eta^h \}_+ \right\|_{L^2(\Omega)}^2 \right\}, \quad (5.A.3)$$

where  $\mu$  and  $\tau_{\text{VE}}$  play the same role as before. Just like for the projector of Section 5.3.5, when the variation entropy condition is not harmed the first-order optimality conditions reduce to an  $H^1(\Omega)$  orthogonality. This optimality projector  $\mathcal{P}^h$  implies:

find  $\phi^h \in \mathcal{W}^h$  such that, for all  $w^h \in \mathcal{W}^h$

$$\left( \phi', w^h \right)_{\mathcal{W}} = \left( \mu \tau_{\text{VE}}^2 \{ \mathcal{R}_{\text{VE}} \eta^h \}_+, \operatorname{d} \mathcal{R}_{\text{VE}} \eta^h (\nabla \phi^h) (\nabla w^h) \right)_{L^2(\Omega)}. \quad (5.A.4)$$

Employing the definition of  $\mathcal{R}_{\text{VE}}$  and the chain rule we arrive at:

$$\begin{aligned} \left( \phi', w^h \right)_{\mathcal{W}} &= \left( \mu \tau_{\text{VE}}^2 \{ \mathcal{R}_{\text{VE}} \eta^h \}_+, \frac{\partial \eta^h}{\partial \nabla \phi^h}, \operatorname{d} (\nabla \mathcal{R}_{\text{CL}}) (\phi^h) (w^h) \right)_{L^2(\Omega)} \\ &\quad + \left( \mu \tau_{\text{VE}}^2 \{ \mathcal{R}_{\text{VE}} \eta^h \}_+, \mathbf{H}_{\nabla \phi^h} \eta^h \nabla w^h, \nabla \mathcal{R}_{\text{CL}} \phi^h \right)_{L^2(\Omega)}. \end{aligned} \quad (5.A.5)$$

Using the homogeneity property and interchanging differential operators we may write:

$$\begin{aligned} \left( \phi', w^h \right)_{\mathcal{W}} &= \left( \mathbf{K} \nabla \phi^h, \tau_{\text{VE}} \nabla \left( \operatorname{d} \mathcal{R}_{\text{CL}} (\phi^h) (w^h) \right) \right)_{L^2(\Omega)} \\ &\quad + \left( \bar{\mathbf{K}} \nabla w^h, \tau_{\text{VE}} \nabla \mathcal{R}_{\text{CL}} \phi^h \right)_{L^2(\Omega)}. \end{aligned} \quad (5.A.6)$$

where  $\tau_{\text{VE}}$  denotes the time-scale linked to the variation entropy and where the matrices are given by:

$$\mathbf{K} = \nu_{\text{VE}} \frac{\partial \eta^h}{\partial \nabla \phi^h} \otimes \frac{\partial \eta^h}{\partial \nabla \phi^h}, \quad (5.A.7a)$$

$$\bar{\mathbf{K}} = \nu_{\text{VE}} \eta^h \mathbf{H}_{\nabla \phi^h} \eta^h, \quad (5.A.7b)$$

$$\nu_{\text{VE}} = \mu \tau_{\text{VE}} \frac{\{ \mathcal{R}_{\text{VE}} \eta^h \}_+}{\eta^h}. \quad (5.A.7c)$$

We arrive at the same expression for  $\nu_{\text{VE}}$ , we may employ the model (5.3.51). At this point it is unclear how to arrive at a numerical method from the small-scale model. There are several options however these include unwanted approximations and/or require neglecting some terms. We do not proceed with this projector however we present some discussion on the diffusion matrices below.

**Proposition 5.A.1.** *The matrices  $\mathbf{K}$  and  $\bar{\mathbf{K}}$  are symmetric positive semi-definite.*

*Proof.* Symmetry is trivial and the positive semi-definiteness of is a direct consequence of  $\nu_{\text{VE}}$  being positive and the convexity of  $\eta^h$  (for  $\bar{\mathbf{K}}$ ).  $\square$

Note that both  $\mathbf{K}$  and  $\bar{\mathbf{K}}$  have the unit of a viscosity and both are based on the variation entropy residual. The matrix  $\mathbf{K}$  is the same as found before and acts in the direction is represented by  $\partial\eta/\partial\nabla\phi^h$ . Below we analyze the matrix  $\bar{\mathbf{K}}$ .

Let  $\mathbf{u}_{\parallel}$  denote the projection of  $\mathbf{u}$  onto  $\partial\eta/\partial\nabla\phi^h$ :

$$\mathbf{u}_{\parallel} := \frac{\partial\hat{\eta}^h}{\partial\nabla\phi^h} \otimes \frac{\partial\hat{\eta}^h}{\partial\nabla\phi^h} \mathbf{u} \quad (5.A.8)$$

where the hat-symbol indicates scaling to unit size:  $\hat{v} = v/\|v\|_2$ . Note that we have the identity

$$\mathbf{u}_{\parallel} \cdot \frac{\partial\eta^h}{\partial\nabla\phi^h} = \mathbf{u} \cdot \frac{\partial\eta^h}{\partial\nabla\phi^h}, \quad (5.A.9)$$

and that the vector

$$\mathbf{u}_{\perp} := \mathbf{u} - \mathbf{u}_{\parallel} = \left( \mathbf{I} - \frac{\partial\hat{\eta}^h}{\partial\nabla\phi^h} \otimes \frac{\partial\hat{\eta}^h}{\partial\nabla\phi^h} \right) \mathbf{u} \quad (5.A.10)$$

is perpendicular to  $\mathbf{u}$ . Whereas the matrix  $\mathbf{K}$  provides control over gradient in the direction  $\partial\eta/\partial\nabla\phi^h$  (represented by  $\mathbf{u}_{\parallel}$ ), the matrix  $\bar{\mathbf{K}}$  can provide control of gradients in the direction orthogonal to that (represented by  $\mathbf{u}_{\perp}$ ). In this case  $\bar{\mathbf{K}}\mathbf{u}$  should be proportional to  $\mathbf{u}_{\perp}$ . This is only the case if  $\eta = \|\nabla\phi^h\|_2$ , as stated in the next proposition.

**Proposition 5.A.2.** *The matrix  $\bar{\mathbf{K}}$  acts in the direction orthogonal to  $\partial\eta/\partial\nabla\phi^h$  if and only if  $\eta = \|\nabla\phi^h\|_2$  (up to multiplication with a constant).*

*Proof.* Up to scaling by a constant, we need to find  $\eta^h$  such that:

$$\bar{\mathbf{K}}\mathbf{u} = \nu_{\text{VE}}\mathbf{u}_{\perp} \quad (5.A.11)$$

for all vectors  $\mathbf{u}$ . Substitution of (5.A.7b) and (5.A.10) gives:

$$\nu_{\text{VE}}\eta^h\mathbf{H}_{\nabla\phi^h}\eta\mathbf{u} = \nu_{\text{VE}} \left( \mathbf{I} - \frac{\partial\hat{\eta}^h}{\partial\nabla\phi^h} \otimes \frac{\partial\hat{\eta}^h}{\partial\nabla\phi^h} \right) \mathbf{u}. \quad (5.A.12)$$

Taking  $\mathbf{u} = \nabla\phi^h$  provides

$$\eta^h\mathbf{H}_{\nabla\phi^h}\eta\nabla\phi^h = \nabla\phi^h - \eta^h \frac{\partial\hat{\eta}^h}{\partial\nabla\phi^h} \left\| \frac{\partial\eta}{\partial\nabla\phi^h} \right\|_2^{-1}. \quad (5.A.13)$$

Next, we use the homogeneity property (5.2.12) to find:

$$\nabla\phi^h - \eta^h \frac{\partial\hat{\eta}^h}{\partial\nabla\phi^h} \left\| \frac{\partial\eta}{\partial\nabla\phi^h} \right\|_2^{-1} = 0. \quad (5.A.14)$$

Rearranging gives

$$\left\| \frac{\partial\eta^h}{\partial\nabla\phi^h} \right\|_2 \nabla\phi^h = \frac{\partial\hat{\eta}^h}{\partial\nabla\phi^h} \eta^h. \quad (5.A.15)$$

Taking the norm leads to:

$$\left\| \frac{\partial\eta^h}{\partial\nabla\phi^h} \right\|_2 \left\| \nabla\phi^h \right\|_2 = \eta^h. \quad (5.A.16)$$

By again using homogeneity we arrive at

$$\left\| \frac{\partial\eta^h}{\partial\nabla\phi^h} \right\|_2 \left\| \nabla\phi^h \right\|_2 = \frac{\partial\eta^h}{\partial\nabla\phi^h} \cdot \nabla\phi^h. \quad (5.A.17)$$

This means that the vectors  $\partial\eta/\partial\nabla\phi^h$  and  $\nabla\phi^h$  point in the same direction. We conclude  $\eta = \|\nabla\phi^h\|_2$ , up to multiplication with a constant.  $\square$



## Part III

### AN ENERGY-DISSIPATIVE METHOD FOR FREE-SURFACE FLOW

In this part we present an energy-dissipative level-set method for the incompressible Navier-Stokes equations with surface tension. The construction is based on so-called functional entropy variables.



---

## AN ENERGY-DISSIPATIVE LEVEL-SET METHOD FOR THE INCOMPRESSIBLE TWO-PHASE NAVIER-STOKES EQUATIONS WITH SURFACE TENSION

---

This chapter is reproduced from [60]:

M.F.P. ten Eikelder and I. Akkerman, *An energy-dissipative level-set method for the incompressible two-phase Navier-Stokes equations with surface tension using functional entropy variables*, under review (2020).

### ABSTRACT

*This chapter presents the first energy-dissipative level-set method for the incompressible Navier-Stokes equations with surface tension. The methodology relies on the recently proposed concept of functional entropy variables. Discretization in space is performed with isogeometric analysis. A new perturbed midpoint scheme is proposed for the temporal-integration. The fully-discrete scheme is unconditionally energy-dissipative, pointwise divergence-free and satisfies the maximum principle for the density. Numerical examples in two and three dimensions verify the energetic stability of the methodology.*

## 6.1 INTRODUCTION

This chapter proposes a novel energy-dissipative numerical method for the computation of the incompressible Navier-Stokes equations with surface tension. Our discretization employs the level-set method to capture the fluid interface and hinges on so-called functional entropy variables. The method unconditionally dissipates the total energy of the system, is pointwise divergence-free and satisfies the maximum principle for the density. The energetic stability improves robustness features and as such the proposed approach is suitable choice for the simulation of immiscible fluids.

### 6.1.1 *Free-surface flow modeling*

Incompressible free-surface flows with surface tension appear in a large class of applications ranging from marine and offshore engineering, e.g., sloshing of LNG in tanks or wave impacts, to bubble dynamics. Applications typically involve violent free-surface flows. As a result topological changes (e.g., break-up or coalescence) occur. Numerical methods for two-fluid flow problems typically follow the free-surface motion with either mesh-motion or with an extra variable that captures the topological changes. The first class of methods is known as interface-tracking methods whereas the second are the interface-capturing methods. When there is a large amount of topological changes interface-tracking methods are an unfortunate choice. On the other hand, interface capturing methods [109, 186, 194] naturally deal with the interface and seem in this case to be the more suitable choice.

Interface capturing methods can roughly be divided into phase-field methods, volume-of-fluid methods and level-set methods, see [64] for a discussion. The phase field models [80, 82, 143, 144] are known for their rigorous thermodynamical structure. The main issue is that numerical methods for phase field models do not provably satisfy the maximum principle for the density [174]. Volume-of-fluid methods [93, 157, 167] are popular methods, also for compressible flows modeling [10, 126], but suffer from the same discrepancy. The maximum principle is generally only guaranteed if a CFL-like condition is fulfilled, see e.g., [63]. When simulating air-water flows the maximum principle is crucial due to the large density jump. Therefore we employ in this chapter the level-set method [2, 168, 169, 181] which by construction satisfies the maximum principle for the density. The level set method does not limit the complexity of the free-surface flow nor the flow regime. It has proven to be suitable tool for free-surface flows in marine applications, e.g., [3, 5, 6, 151].

### 6.1.2 *Surface tension*

Apart from the ability to capture the interface location, the extra variable in interface capturing methods may be used to evaluate the surface tension contribution. In volume-of-fluid and level-set methods the interface normal and curvature may be computed similarly. It is well-known, see e.g., [1, 158], that surface tension effects are better represented when using the level-set approach as compared with the

volume-of-fluid approach. We refer to [83] for error analysis of the surface tension force in the level-set method. The standard and most popular approach is to use the continuum model of Brackbill et al. [31]. In the discrete approximation the evaluation of the curvature often employs a projection step for lower-order methods which leads to inaccuracies. In a recently paper [203] the authors show that the accuracy of the curvature improves significantly when using a smooth higher-order NURBS-based isogeometric discretization [101].

### 6.1.3 Energetic stability

Level-set methods are, to the best knowledge of the authors, never equipped with a thermodynamically stable algorithm. However the notion of energetic stability<sup>1</sup> is of practical importance. In [3] it is shown that for a viscous air-water level-set simulation in certain situations artificial energy may be created. This leads to a nonphysical prediction of the fluid behavior. The approach of proving an energetic stability result in a Galerkin-type formulation includes the selection of appropriate weights. Unfortunately, the suitable test functions are not available in typical finite element methods. This applies to the spatial and temporal discretization independently.

### 6.1.4 This work

In this chapter we address one of the main discrepancies of regularized-interface level-set methods, namely the above mentioned absence of an energetic stability property. We circumvent the limitation caused by the function spaces by introducing the unavailable weighting function as a new variable via so-called functional entropy variables. This concept is the natural alternative to entropy variables when the mathematical entropy associated with the system of equations is a functional (instead of a function) of the conservation variables. We naturally integrate this new variable into the level-set model via the surface tension term. This creates the required extra freedom and as a result the associated weak form is equipped with energetic stability for standard divergence-conforming function spaces. The formulation does not require the evaluation of the curvature and is as such not limited to higher-order discretizations. To inherit energetic stability in a semi-discrete sense we employ a NURBS-based isogeometric analysis Galerkin-type discretization. Furthermore, we introduce a SUPG stabilization mechanism that does not upset the energy-dissipative property of the method. Additionally, we augment the momentum equation with a residual-based discontinuity capturing term. For the temporal discretization we propose a new time-stepping scheme which can be understood as a perturbation of the midpoint rule. The result is a consistent fully-discrete energy-dissipative scheme that is pointwise divergence-free and satisfies the maximum principle for the density.

---

<sup>1</sup> Note that thermodynamically stable resembles energetically stable in the isothermal case as Clausius-Duhem inequality reduces to an energy-dissipative inequality.

### 6.1.5 Structure of this chapter

The remainder of this chapter is organized as follows. Section 6.2 presents and analyzes the energy behavior of the sharp-interface incompressible Navier-Stokes equations with surface tension. In Section 6.3 we use the sharp-interface model as a starting point to derive the regularized level-set model and provide a detailed analysis in terms of energy behavior. Additionally, we extensively discuss the level-set form of the surface tension contribution. In Section 6.4 we employ the functional entropy variables to obtain a modified energy-dissipative formulation. Then, in Section 6.5 we present the semi-discrete energetically stable formulation. Next, in Section 6.6 we present the fully-discrete energy-dissipative method. Section 6.7 shows the numerical experiments in two and three dimensions which verify the energy-dissipative property of the scheme. We draw conclusions in Section 6.8.

**Remark 6.1.1.** *To keep the work comprehensible we have intentionally not included multi-scale stabilization mechanisms and redistancing procedures. Incorporating these additional techniques in the currently proposed algorithm would allow the simulation of violent flows. These developments lie beyond the scope of this paper.*

## 6.2 SHARP-INTERFACE FORMULATION

### 6.2.1 Governing equations

Let  $\Omega \subset \mathbb{R}^d$ ,  $d = 2, 3$ , denote the spatial domain with boundary  $\partial\Omega$ . We consider two immiscible incompressible fluids that occupy subdomains  $\Omega_i \subset \Omega$ ,  $i = 1, 2$ , in the sense  $\bar{\Omega} = \bar{\Omega}_1 \cup \bar{\Omega}_2$  and  $\Omega_1 \cap \Omega_2 = \emptyset$ . A time-dependent smooth interface  $\Gamma = \partial\Omega_1 \cap \partial\Omega_2$  separates the fluids. The problem under consideration consists of solving the incompressible Navier-Stokes equations with surface tension dictating the two-fluid flow:

$$\rho_i (\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u}) - \mu_i \Delta \mathbf{u} + \nabla p = \rho_i \mathbf{g}, \quad \text{in } \Omega_i(t) \quad (6.2.1a)$$

$$\nabla \cdot \mathbf{u} = 0 \quad \text{in } \Omega_i(t), \quad (6.2.1b)$$

$$\llbracket \mathbf{u} \rrbracket = 0 \quad \text{on } \Gamma(t), \quad (6.2.1c)$$

$$\llbracket \mathbf{S}(\mathbf{u}, p) \mathbf{v} \rrbracket = \sigma \kappa \mathbf{v} \quad \text{on } \Gamma(t), \quad (6.2.1d)$$

$$V = \mathbf{u} \cdot \mathbf{v} \quad \text{on } \Gamma(t), \quad (6.2.1e)$$

with  $\mathbf{u}(\mathbf{x}, 0) = \mathbf{u}_0(\mathbf{x})$  in  $\Omega_i(0)$  and  $\Gamma(0) = \Gamma_0$  for the fluid velocity  $\mathbf{u} : \Omega \rightarrow \mathbb{R}^d$  and the pressure  $p : \Omega \rightarrow \mathbb{R}$ . The stress tensor is given by:

$$\mathbf{S}(\mathbf{u}, p) = \boldsymbol{\tau}(\mathbf{u}) - p\mathbf{I} \quad \text{in } \Omega_i(t) \quad (6.2.2)$$

with viscous stress tensor:

$$\boldsymbol{\tau}(\mathbf{u}) = 2\mu_i \nabla^s \mathbf{u} \quad \text{in } \Omega_i(t). \quad (6.2.3)$$

The jump of a vector  $\mathbf{v}$  is denoted as

$$\llbracket \mathbf{v} \rrbracket = (\mathbf{v}|_{\Omega_1} - \mathbf{v}|_{\Omega_2})|_{\Gamma}. \quad (6.2.4)$$

The problem is augmented with appropriate boundary conditions. We denote with  $\mathbf{x} \in \Omega$  the spatial parameter and with  $t \in \mathcal{T} = (0, T)$  the time with end time  $T > 0$ . Furthermore, we set  $\mathbf{g} = -g\mathbf{j}$  where  $g$  is the gravitational acceleration and  $\mathbf{j}$  is the vertical unit vector. The initial velocity is  $\mathbf{u}_0 : \Omega \rightarrow \mathbb{R}^d$ . We use the standard convention for the various differential operators, i.e. the temporal derivative reads  $\partial_t$  and the symmetric gradient denotes  $\nabla^s \cdot = \frac{1}{2} (\nabla \cdot + \nabla^T \cdot)$ . The constants  $\mu_i > 0$  and  $\rho_i > 0$  denote the dynamic viscosity and density of fluid  $i$  respectively. The normal speed of  $\Gamma(t)$  is denoted as  $V$ , the normal of  $\Gamma(t)$ , denoted  $\boldsymbol{\nu}$ , is pointing from  $\Omega_2(t)$  into  $\Omega_1(t)$  and the tangential vector is  $\mathbf{t}$ . The curvature is  $\kappa = \nabla \cdot \boldsymbol{\nu}$ , i.e.  $\kappa(\mathbf{x}, t)$  is negative when  $\Omega_1(t)$  is convex in a neighborhood of  $\mathbf{x} \in \Gamma(t)$ . Furthermore, the outward-pointing normal of  $\partial\Omega$  denotes  $\mathbf{n}$ . We defined  $u_n = \mathbf{u} \cdot \mathbf{n}$  and  $u_\nu = \mathbf{u} \cdot \boldsymbol{\nu}$  as the normal velocity of  $\partial\Omega$  and  $\Gamma(t)$ , respectively. The equation (6.2.1a) represents the the balance of momentum while (6.2.1b) is the continuity equation. Next, (6.2.1c) states that the velocities are continuous across the separating interface. The fourth equation, (6.2.1d), stipulates that the discontinuity of the stresses at the interface is governed by surface tension. In absence of surface tension it reduces to an equilibrium of the stresses. Note that a direct consequence of (6.2.1d) is the continuity of tangential stress at the interface:

$$\llbracket 2\mu_i(\nabla^s \mathbf{u})\boldsymbol{\nu} \rrbracket \cdot \mathbf{t} = 0 \quad \text{on } \Gamma(t). \quad (6.2.5)$$

We assume that the surface tension coefficient  $\sigma \geq 0$  is constant, i.e. Maragoni effects are precluded. Furthermore, we assume that line force terms vanish as a result of boundary conditions or additional conditions (see also [173]). We refer to [160] for some well-posed properties of the problem.

We introduce the notation

$$\rho = \rho_1\chi_{\Omega_1(t)} + \rho_2\chi_{\Omega_2(t)}, \quad (6.2.6a)$$

$$\mu = \mu_1\chi_{\Omega_1(t)} + \mu_2\chi_{\Omega_2(t)}, \quad (6.2.6b)$$

with indicator  $\chi_D$  of domain  $D$ . System (6.2.1) may now be written as:

$$\rho(\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u}) - \mu \Delta \mathbf{u} + \nabla p = \rho \mathbf{g} \quad \text{in } \Omega, \quad (6.2.7a)$$

$$\nabla \cdot \mathbf{u} = 0 \quad \text{in } \Omega, \quad (6.2.7b)$$

$$\llbracket \mathbf{u} \rrbracket = 0 \quad \text{on } \Gamma(t), \quad (6.2.7c)$$

$$\llbracket \mathbf{S}(\mathbf{u}, p)\boldsymbol{\nu} \rrbracket = \sigma \kappa \boldsymbol{\nu} \quad \text{on } \Gamma(t), \quad (6.2.7d)$$

$$V = \mathbf{u} \cdot \boldsymbol{\nu} \quad \text{on } \Gamma(t), \quad (6.2.7e)$$

where  $\boldsymbol{\tau}(\mathbf{u}) \equiv 2\mu \nabla^s \mathbf{u}$  and  $\mathbf{u}(\mathbf{x}, 0) = \mathbf{u}_0(\mathbf{x})$  in  $\Omega_i(0)$  and  $\Gamma(0) = \Gamma_0$ .

As we aim to develop an energy-dissipative level-set method, we first study the energy behavior of the sharp-interface model associated with system (6.2.7). This is the purpose of the remainder of Section 6.2. After the energy analysis in Section 6.2.2 we present a standard weak formulation of (6.2.7) in Section 6.2.3.

## 6.2.2 Energy evolution

We consider the dissipation of the energy of the problem (6.2.7). The total energy consists of three contributions, namely kinetic ( $K$ ), gravitational ( $G$ ) and surface energy ( $S$ ):

$$\mathcal{E}_s(\mathbf{u}) = \mathcal{E}_s^K(\mathbf{u}) + \mathcal{E}_s^G + \mathcal{E}_s^S, \quad (6.2.8a)$$

$$\mathcal{E}_s^K(\mathbf{u}) := \int_{\Omega} \frac{1}{2} \rho \|\mathbf{u}\|_2^2 \, d\Omega, \quad (6.2.8b)$$

$$\mathcal{E}_s^G := \int_{\Omega} \rho g y \, d\Omega, \quad (6.2.8c)$$

$$\mathcal{E}_s^S := \int_{\Gamma(t)} \sigma \, d\Gamma, \quad (6.2.8d)$$

with  $y = \mathbf{x} \cdot \mathbf{j}$  the vertical coordinate. The subscript  $s$  refers to the *sharp-interface* model.

**Theorem 6.2.1.** *Let  $\mathbf{u}$  and  $p$  be smooth solutions of the incompressible Navier-Stokes equations with surface tension (6.2.7). The total energy  $\mathcal{E}_s$ , given in (6.2.8), satisfies the dissipation inequality:*

$$\frac{d}{dt} \mathcal{E}_s(\mathbf{u}) = - \int_{\Omega} \boldsymbol{\tau}(\mathbf{u}) : \nabla \mathbf{u} \, d\Omega + \mathcal{B}_s \leq 0 + \mathcal{B}_s, \quad (6.2.9)$$

where  $\mathcal{B}_s$  contains the boundary contributions:

$$\mathcal{B}_s = \int_{\partial\Omega} \mathbf{n}^T (\mathbf{S}(\mathbf{u}, p) - (\frac{1}{2} \rho \|\mathbf{u}\|^2 + \rho g y) \mathbf{I}) \mathbf{u} \, dS. \quad (6.2.10)$$

*Proof.* To establish the dissipative property (6.2.9) we will first consider the evolution of each of the energy contributions (6.2.8) separately and subsequently substitute these in the strong form (6.2.7).

We start off with the kinetic energy evolution. The following sequence of identities holds:

$$\begin{aligned} \frac{d}{dt} \mathcal{E}_s^K &= \int_{\Omega_1(t)} \rho \mathbf{u} \cdot \partial_t \mathbf{u} \, d\Omega + \int_{\Omega_2(t)} \rho \mathbf{u} \cdot \partial_t \mathbf{u} \, d\Omega \\ &\quad + \int_{\partial\Omega_1(t) \cap \Gamma(t)} \frac{1}{2} \rho \|\mathbf{u}\|^2 \mathbf{u} \cdot \mathbf{n}_1 \, dS + \int_{\partial\Omega_2(t) \cap \Gamma(t)} \frac{1}{2} \rho \|\mathbf{u}\|^2 \mathbf{u} \cdot \mathbf{n}_2 \, dS \\ &= \int_{\Omega} \rho \mathbf{u} \cdot (\partial_t \mathbf{u} + (\mathbf{u} \cdot \nabla) \mathbf{u}) \, d\Omega + \int_{\Omega} \frac{1}{2} \rho \|\mathbf{u}\|^2 \nabla \cdot \mathbf{u} \, d\Omega \\ &\quad - \int_{\partial\Omega} \frac{1}{2} \rho \|\mathbf{u}\|^2 u_n \, dS, \end{aligned} \quad (6.2.11)$$

where  $\mathbf{n}_1$  and  $\mathbf{n}_2$  denote the outward unit normal of  $\Omega_1(t)$  and  $\Omega_2(t)$ , respectively. The first identity results from the Leibniz-Reynolds transport theorem. To obtain the second equality one adds a suitable partition of zero, subsequently applies the divergence theorem on both  $\Omega_1(t)$  and  $\Omega_2(t)$ , and lastly uses the chain rule.

In a similar fashion we have the identities for the gravitational energy evolution:

$$\begin{aligned}
\frac{d}{dt} \mathcal{E}_s^G &= \int_{\partial\Omega_1(t) \cap \Gamma(t)} \rho g y \mathbf{u} \cdot \mathbf{n}_1 \, dS + \int_{\partial\Omega_2(t) \cap \Gamma(t)} \rho g y \mathbf{u} \cdot \mathbf{n}_2 \, dS \\
&= \int_{\Omega_1(t)} \rho g \mathbf{j} \cdot \mathbf{u} \, d\Omega + \int_{\Omega_1(t)} \rho g y \nabla \cdot \mathbf{u} \, d\Omega \\
&\quad + \int_{\Omega_2(t)} \rho g \mathbf{j} \cdot \mathbf{u} \, d\Omega + \int_{\Omega_2(t)} \rho g y \nabla \cdot \mathbf{u} \, d\Omega - \int_{\partial\Omega} \rho g y \mathbf{u} \cdot \mathbf{n} \, dS \\
&= \int_{\Omega} \rho g \mathbf{u} \cdot \mathbf{j} \, d\Omega + \int_{\Omega} \rho g y \nabla \cdot \mathbf{u} \, d\Omega - \int_{\partial\Omega} \rho g y u_n \, dS. \tag{6.2.12}
\end{aligned}$$

The first identity emanates from the Leibniz-Reynolds transport theorem and the second is a direct consequence of the divergence theorem.

Finally, we consider the energetic contribution due to surface tension. We have from the Reynolds transport theorem in tangential calculus, see e.g., [176], the identity:

$$\frac{d}{dt} \mathcal{E}_s^S = \int_{\Gamma(t)} \sigma \kappa u_\nu \, d\Gamma - \int_{\partial\Gamma(t)} \sigma \mathbf{u} \cdot \boldsymbol{\nu}_\partial \, d(\partial\Gamma), \tag{6.2.13}$$

where we recall that we do not account for Maragoni forces ( $\sigma$  is constant). Here  $\boldsymbol{\nu}_\partial$  is the unit-normal vector to  $\partial\Gamma(t)$ , tangent to  $\Gamma(t)$ . We refer to [37, 177] for alternative insightful derivations of (6.2.13). We discard the last member of the right-hand side of (6.2.13) as it represents a line force.

We multiply the momentum equation by  $\mathbf{u}$  and subsequently integrate over the domain:

$$\int_{\Omega} \mathbf{u}^T \rho (\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u}) \, d\Omega + \int_{\Omega} \mathbf{u}^T (\nabla p - \mu \Delta \mathbf{u}) \, d\Omega + \int_{\Omega} \rho g \mathbf{u} \cdot \mathbf{j} \, d\Omega = 0. \tag{6.2.14}$$

Considering the second expression in (6.2.14) in isolation we have the two identities:

$$\begin{aligned}
\int_{\Omega} \mathbf{u}^T (\nabla p - \mu \Delta \mathbf{u}) \, d\Omega &= \int_{\Omega_1(t)} \mathbf{u}^T \nabla (p \mathbf{I} - \mu_1 \nabla^s \mathbf{u}) \, d\Omega \\
&\quad + \int_{\Omega_2(t)} \mathbf{u}^T \nabla (p \mathbf{I} - \mu_2 \nabla^s \mathbf{u}) \, d\Omega \\
&\quad + \int_{\Omega_1(t)} \mu_1 \mathbf{u} \cdot \nabla (\nabla \cdot \mathbf{u}) \, d\Omega + \int_{\Omega_2(t)} \mu_2 \mathbf{u} \cdot \nabla (\nabla \cdot \mathbf{u}) \, d\Omega \\
&= \int_{\Omega} \nabla \mathbf{u} : \mathbf{S}(\mathbf{u}, p) \, d\Omega - \int_{\partial\Omega} \mathbf{n}^T \mathbf{S}(\mathbf{u}, p) \mathbf{u} \, d\Omega \\
&\quad + \int_{\Omega} \mu \mathbf{u} \cdot \nabla (\nabla \cdot \mathbf{u}) \, d\Omega + \int_{\Gamma(t)} \sigma \kappa u_\nu \, d\Gamma. \tag{6.2.15}
\end{aligned}$$

The first identity follows from adding a suitable partition of zero. For the second equality we perform integration by parts and make use of the jump (6.2.7d) where we note that on  $\Gamma(t)$  we have  $\mathbf{n}_1 = -\boldsymbol{\nu}$  and  $\mathbf{n}_2 = \boldsymbol{\nu}$ .

We deduce from the continuity equation:

$$-\int_{\Omega} (p + \frac{1}{2} \rho \|\mathbf{u}\|_2^2 + \rho g y) \nabla \cdot \mathbf{u} \, d\Omega + \int_{\Omega} \mu \mathbf{u} \cdot \nabla (\nabla \cdot \mathbf{u}) \, d\Omega = 0. \tag{6.2.16}$$

Next, we collect the identities (6.2.11), (6.2.12), (6.2.13), (6.2.15) and (6.2.16), substitute these into (6.2.14). The first member in (6.2.14) cancels with the first term in the

ultimate expression in (6.2.11). By virtue of (6.2.15) the second term in (6.2.14) drops out. The third member of (6.2.14) disappears due to (6.2.12). Some of the other terms in (6.2.11), (6.2.12) and (6.2.15) vanish due to (6.2.13) and (6.2.16). Gathering the expressions we eventually arrive at:

$$\frac{d}{dt} \mathcal{E}_s = - \int_{\Omega} \boldsymbol{\tau}(\mathbf{u}) : \nabla \mathbf{u} \, d\Omega + \int_{\partial\Omega} \mathbf{n}^T (\mathbf{S}(\mathbf{u}, p) - (\frac{1}{2}\rho \|\mathbf{u}\|^2 + \rho g y) \mathbf{I}) \mathbf{u} \, dS. \quad (6.2.17)$$

This completes the proof with

$$\mathcal{B}_s = \int_{\partial\Omega} \mathbf{n}^T (\mathbf{S}(\mathbf{u}, p) - (\frac{1}{2}\rho \|\mathbf{u}\|^2 + \rho g y) \mathbf{I}) \mathbf{u} \, dS. \quad (6.2.18)$$

□

### 6.2.3 Weak formulation

Recall that we suppress line force contributions as a result of boundary or auxiliary conditions. At this point we also assume homogeneous boundary conditions to increase readability of the remainder of this chapter. Results can be easily extended to non-homogeneous boundary conditions. We define  $(\cdot, \cdot)_{\Omega}$  as the  $L^2(\Omega)$  inner product on the interior and  $(\cdot, \cdot)_{\Gamma}$  as the  $L^2(\Gamma)$  inner product on the boundary. We take zero-average pressures for all  $t \in \mathcal{T}$ . The space-time velocity-pressure function-space satisfying homogeneous boundary condition  $\mathbf{u} = \mathbf{0}$  denotes  $\mathcal{U}_T$  and the corresponding weighting function space denotes  $\mathcal{U}$ . The standard conservative weak formulation corresponding to strong form (6.2.7) reads:

Find  $\{\mathbf{u}, p\} \in \mathcal{U}$  such that for all  $\{\mathbf{w}, q\} \in \mathcal{U}$ :

$$\begin{aligned} (\mathbf{w}, \rho (\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u}))_{\Omega} - (\nabla \cdot \mathbf{w}, p)_{\Omega} + (\nabla \mathbf{w}, \boldsymbol{\tau}(\mathbf{u}))_{\Omega} \\ + (\mathbf{w}, \sigma \kappa \nu)_{\Gamma(t)} - (\mathbf{w}, \rho \mathbf{g})_{\Omega} = 0, \end{aligned} \quad (6.2.19a)$$

$$(q, \nabla \cdot \mathbf{u})_{\Omega} = 0, \quad (6.2.19b)$$

with interface speed  $V = \mathbf{u} \cdot \boldsymbol{\nu}$ . The weak formulation (6.2.19) is equivalent to the strong form (6.2.7) for smooth solutions and the associated energy evolution relation coincides with that of the strong form (6.2.7).

**Remark 6.2.2.** To show the energy evolution for the case of non-homogeneous boundary conditions one may enforce boundary conditions with a Lagrange multiplier construction [57, 58, 102, 117] and subsequently use (6.2.13) to identify the surface energy contribution.

**Remark 6.2.3.** In order to avoid evaluating second-derivatives the alternative form  $+(\nabla \mathbf{w}, \sigma \mathbf{P}_T)_{\Gamma}$  for the surface tension term in (6.2.19) with tangential projection  $\mathbf{P}_T = \mathbf{I} - \boldsymbol{\nu} \otimes \boldsymbol{\nu}$  may be used. In Appendix 6.A.1 we provide the derivation of this alternative form.

## 6.3 REGULARIZED-INTERFACE LEVEL-SET MODEL

In this section we present the regularized-interface level-set model and analyze its energy behavior. To do so, in Section 6.3.1 we provide the level-set formulation of (6.2.7) which we subsequently present in non-dimensional form Section 6.3.2. Then

in Section 6.3.3 we regularize the sharp-interface level-set formulation. We conclude with a detailed study of the energy behavior of this level-set formulation in Section 6.3.4.

### 6.3.1 Sharp-interface level-set formulation

We employ the interface capturing level-set method to reformulate model (6.2.19). To this purpose we introduce the level-set function  $\phi : \Omega(t) \rightarrow \mathbb{R}$  to describe the evolution of the interface  $\Gamma(t)$ . The sub-domains and interface are identified as:

$$\Omega_1(t) \equiv \{\mathbf{x} \in \Omega(t) | \phi(\mathbf{x}, t) > 0\}, \quad (6.3.1a)$$

$$\Omega_2(t) \equiv \{\mathbf{x} \in \Omega(t) | \phi(\mathbf{x}, t) < 0\}, \quad (6.3.1b)$$

$$\Gamma(t) \equiv \{\mathbf{x} \in \Omega(t) | \phi(\mathbf{x}, t) = 0\}. \quad (6.3.1c)$$

The motion of the interface  $\Gamma(t)$  is governed by pure convection:

$$\partial_t \phi + \mathbf{u} \cdot \nabla \phi = \partial_t \phi + V \|\nabla \phi\| = 0. \quad (6.3.2)$$

This results from taking the temporal derivative of the zero-level set. The domain indicator may now be written as:

$$\chi_{\Omega_1} = H(\phi), \quad (6.3.3a)$$

$$\chi_{\Omega_2} = 1 - H(\phi), \quad (6.3.3b)$$

where  $H$  is the Heaviside function with the half-maximum convention:

$$H(\phi) = \begin{cases} 0 & \phi < 0 \\ \frac{1}{2} & \phi = 0 \\ 1 & \phi > 0. \end{cases} \quad (6.3.4)$$

The resulting density and fluid viscosity are:

$$\rho(\phi) = \rho_1 H(\phi) + \rho_2 (1 - H(\phi)), \quad (6.3.5a)$$

$$\mu(\phi) = \mu_1 H(\phi) + \mu_2 (1 - H(\phi)), \quad (6.3.5b)$$

and the viscous stress now depends on  $\mathbf{u}$  and  $\phi$ :

$$\boldsymbol{\tau}(\mathbf{u}, \phi) = 2\mu(\phi) \nabla^s \mathbf{u}. \quad (6.3.6)$$

In order to write the surface term in (6.2.19) in the level-set context we need expressions for the surface normal, the curvature and require to convert the surface integral into a domain integral. This is how we proceed. We first define the regularized 2-norm  $\|\cdot\|_{\epsilon,2} : \mathbb{R} \rightarrow \mathbb{R}_+$  for dimensionless  $\mathbf{b} \in \mathbb{R}^d$  and  $\epsilon \geq 0$  as:

$$\|\mathbf{b}\|_{\epsilon,2}^2 := \mathbf{b} \cdot \mathbf{b} + \epsilon^2. \quad (6.3.7)$$

The surface normal is now continuously extended into the domain via

$$\hat{\mathbf{v}}(\phi) = \frac{\nabla \phi}{\|\nabla \phi\|_{\epsilon,2}}. \quad (6.3.8)$$

The curvature results from taking the divergence of (6.3.8):

$$\hat{\kappa}(\phi) := \nabla \cdot \hat{\nu} = \nabla \cdot \left( \frac{\nabla \phi}{\|\nabla \phi\|_{\epsilon,2}} \right). \quad (6.3.9)$$

We may now convert the surface integral into

$$\int_{\Gamma(t)} \sigma \boldsymbol{\omega} \cdot \boldsymbol{\nu} \kappa \, d\Gamma = \int_{\Omega} \sigma \boldsymbol{\omega} \cdot \hat{\nu}(\phi) \hat{\kappa}(\phi) \delta_{\Gamma}(\phi) \, d\Omega. \quad (6.3.10)$$

Here  $\delta_{\Gamma} = \delta_{\Gamma}(\phi)$  denotes the Dirac delta concentrated on the interface  $\Gamma(t)$ :

$$\delta_{\Gamma}(\phi) := \delta(\phi) \|\nabla \phi\|_{\epsilon,2}. \quad (6.3.11)$$

which extends the integral over boundary  $\Gamma(t)$  to the domain  $\Omega$  [155]. In (6.3.11)  $\delta(\phi)$  represents the Dirac delta distribution. The expression in (6.3.10) is exact for  $\epsilon = 0$  and an approximation otherwise. We refer to Chang et al. [41] for an insightful derivation. For more rigorous details the reader may consult [95]. Note that we have suppressed  $\epsilon$  in (6.3.8)-(6.3.11). The corresponding strong form writes in terms of the variables  $\mathbf{u}, p$  and  $\phi$  as:

$$\rho(\phi) (\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u}) - \nabla \cdot \boldsymbol{\tau}(\mathbf{u}, \phi) + \nabla p + \sigma \delta_{\Gamma}(\phi) \hat{\kappa}(\phi) \hat{\nu}(\phi) - \rho(\phi) \mathbf{g} = 0, \quad (6.3.12a)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (6.3.12b)$$

$$\partial_t \phi + \mathbf{u} \cdot \nabla \phi = 0, \quad (6.3.12c)$$

with  $\mathbf{u}(\mathbf{x}, 0) = \mathbf{u}_0(\mathbf{x})$  and  $\phi(\mathbf{x}, 0) = \phi_0(\mathbf{x})$  in  $\Omega$ . From this point onward we skip the hat symbols for simplicity.

### 6.3.2 Non-dimensionalization

We now perform the non-dimensionalization of the incompressible Navier-Stokes equations with surface tension. Here we re-scale the system (6.3.12) based on physical variables. The dimensionless variables are given by:

$$\begin{aligned} \mathbf{x}^* &= \frac{\mathbf{x}}{L_0}, & \mathbf{u}^* &= \frac{\mathbf{u}}{U_0}, & t^* &= \frac{t U_0}{L_0}, & \rho^* &= \frac{\rho}{\rho_1}, \\ \mu^* &= \frac{\mu}{\mu_1}, & \phi^* &= \frac{\phi}{L_0}, & p^* &= \frac{p}{\rho_1 U_0^2}, \end{aligned} \quad (6.3.13)$$

where  $L_0$  is a characteristic length scale and  $U_0$  is a characteristic velocity. A direct consequence is

$$\kappa^*(\phi^*) := \nabla^* \cdot \left( \frac{\nabla^* \phi^*}{\|\nabla^* \phi^*\|_{\epsilon,2}} \right) = L_0 \kappa(\phi), \quad (6.3.14a)$$

$$\delta_{\Gamma}^*(\phi^*) := \delta(\phi^*) \|\nabla^* \phi^*\|_{\epsilon,2} = L_0 \delta_{\Gamma}(\phi), \quad (6.3.14b)$$

where we have used the scaling property of the Dirac delta:

$$\delta(\alpha \phi) = \frac{1}{|\alpha|} \delta(\phi), \quad \alpha \neq 0. \quad (6.3.15)$$

The dimensionless system reads:

$$\begin{aligned} \rho^*(\phi^*) (\partial_{t^*} \mathbf{u}^* + \mathbf{u}^* \cdot \nabla^* \mathbf{u}^*) - \nabla^* \cdot \boldsymbol{\tau}^*(\mathbf{u}^*, \phi^*) + \nabla^* p^* \\ + \frac{1}{\text{We}} \delta_{\Gamma}^*(\phi^*) \kappa^*(\phi^*) \mathbf{v}^*(\phi^*) + \frac{1}{\text{Fr}^2} \rho^*(\phi^*) \mathbf{j} = 0, \end{aligned} \quad (6.3.16a)$$

$$\nabla^* \cdot \mathbf{u}^* = 0, \quad (6.3.16b)$$

$$\partial_{t^*} \phi^* + \mathbf{u}^* \cdot \nabla^* \phi^* = 0, \quad (6.3.16c)$$

where dimensionless viscous stress is given by:

$$\boldsymbol{\tau}^* = \boldsymbol{\tau}^*(\mathbf{u}^*, \phi^*) = \frac{1}{\text{Re}} \mu^*(\phi^*) \left( \nabla^* \mathbf{u}^* + \nabla^{*T} \mathbf{u}^* \right). \quad (6.3.17)$$

The used dimensionless coefficients are the Reynolds number (Re) which expresses relative strength of inertial forces and viscous forces, the Weber number (We) measuring the ratio of inertia to surface tension and the Froude number (Fr) which quantifies inertia with respect to gravity. The expressions are given by:

$$\text{Re} = \frac{\rho_1 U_0 L_0}{\mu_1}, \quad (6.3.18a)$$

$$\text{We} = \frac{\rho_1 U_0^2 L_0}{\sigma}, \quad (6.3.18b)$$

$$\text{Fr} = \frac{U_0}{\sqrt{g L_0}}. \quad (6.3.18c)$$

We suppress the star symbols in the remainder of this Chapter.

### 6.3.3 Regularization

In the following we smear the interface over an interface-width of  $\varepsilon > 0$  via replacing the (sharp) Heaviside function (6.3.4) by a regularized differentiable Heaviside  $H_\varepsilon(\phi)$ . We postpone the specific form of  $H_\varepsilon(\phi)$  to Section 6.6. The regularized delta function is  $\delta_{\Gamma, \varepsilon}(\phi) = \delta_\varepsilon(\phi) \|\nabla \phi\|_{\varepsilon, 2}$  with one-dimensional continuous regularized delta function  $\delta_\varepsilon(\phi) = H'_\varepsilon(\phi)$ . We refer to [136] for details concerning the approximation of the Dirac delta. The density and fluid viscosity are computed as

$$\rho_\varepsilon \equiv \rho_\varepsilon(\phi) := \rho_1 H_\varepsilon(\phi) + \rho_2 (1 - H_\varepsilon(\phi)), \quad (6.3.19a)$$

$$\mu_\varepsilon \equiv \mu_\varepsilon(\phi) := \mu_1 H_\varepsilon(\phi) + \mu_2 (1 - H_\varepsilon(\phi)). \quad (6.3.19b)$$

Our procedure to arrive at an energy-dissipative formulation, presented in Section 6.4, requires a conservative formulation of the momentum equation. Using the continuity and level-set equations, the associated model follows straightforwardly:

$$\begin{aligned} \partial_t (\rho_\varepsilon(\phi) \mathbf{u}) + \nabla \cdot (\rho_\varepsilon(\phi) \mathbf{u} \otimes \mathbf{u}) - \nabla \cdot \boldsymbol{\tau}_\varepsilon(\mathbf{u}, \phi) + \nabla p \\ + \frac{1}{\text{We}} \delta_{\Gamma, \varepsilon}(\phi) \kappa(\phi) \mathbf{v}(\phi) + \frac{1}{\text{Fr}^2} \rho_\varepsilon(\phi) \mathbf{j} = 0, \end{aligned} \quad (6.3.20a)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (6.3.20b)$$

$$\partial_t \phi + \mathbf{u} \cdot \nabla \phi = 0, \quad (6.3.20c)$$

where  $\boldsymbol{\tau}_\varepsilon(\mathbf{u}, \phi) = 2\mu_\varepsilon(\phi) \nabla^s \mathbf{u}$  and  $\mathbf{u}(\mathbf{x}, 0) = \mathbf{u}_0(\mathbf{x})$  and  $\phi(\mathbf{x}, 0) = \phi_0(\mathbf{x})$  in  $\Omega$ . At this point we have assumed a constant interface width  $\varepsilon$ . In the following we omit the  $\varepsilon$  for the sake of notational simplicity.

**Remark 6.3.1.** In case of a non-constant  $\epsilon$  one requires to augment the right-hand side of (6.3.20a) with  $\partial\rho_\epsilon/\partial\epsilon(\partial_t\epsilon + \mathbf{u} \cdot \nabla\epsilon)$ .

**Remark 6.3.2.** At this point we remark that as an alternative one may also employ a skew-symmetric form for the convective terms. Via a partial integration step,

$$\begin{aligned} (\mathbf{w}, \nabla \cdot (\rho\mathbf{u} \otimes \mathbf{u}))_\Omega &= \frac{1}{2}(\mathbf{w}, \rho\mathbf{u} \cdot \nabla\mathbf{u})_\Omega - \frac{1}{2}(\nabla\mathbf{w}, \rho\mathbf{u} \otimes \mathbf{u})_\Omega + \frac{1}{2}(\mathbf{w}, \mathbf{u}\mathbf{u} \cdot \nabla\rho)_\Omega \\ &\quad + \frac{1}{2}(\mathbf{w}, \rho\mathbf{u}\nabla \cdot \mathbf{u})_\Omega, \end{aligned} \quad (6.3.21)$$

we may replace the convective term in (6.3.20) by the first three terms on the right-hand side of (6.3.21). In the current situation the specific form of the convective terms (conservative or skew-symmetric) is not essential. This changes when the formulation is equipped with multiscale stabilization terms. In the single-fluid case (in absence of surface tension) the well-known multiscale discretization that represents an energy-stable system is the skew-symmetric form, see e.g., [57, 58, 71]. In contrast to the current two-phase model, this property is for the single-fluid case directly inherited by the fully-discrete case when employing the mid-point rule for time integration.

### 6.3.4 Energy evolution

In the following we show the energy balance of the level-set formulation (6.3.20). The kinetic, gravitational and surface energy associated with system (6.3.20) are:

$$\mathcal{E}^K(\mathbf{u}, \phi) := \left(\frac{1}{2}\rho(\phi)\mathbf{u}, \mathbf{u}\right)_\Omega, \quad (6.3.22a)$$

$$\mathcal{E}^G(\phi) := \frac{1}{\mathbb{F}\mathbb{r}^2} (\rho(\phi), y)_\Omega, \quad (6.3.22b)$$

$$\mathcal{E}^S(\phi) := \frac{1}{\mathbb{W}\mathbb{e}} (1, \delta_\Gamma(\phi))_\Omega. \quad (6.3.22c)$$

The total energy is the superposition of the separate energies:

$$\mathcal{E}(\mathbf{u}, \phi) = \mathcal{E}^K(\mathbf{u}, \phi) + \mathcal{E}^G(\phi) + \mathcal{E}^S(\phi). \quad (6.3.23)$$

The local energy is given by:

$$\mathcal{H} = \frac{1}{2}\rho(\phi)\|\mathbf{u}\|_2^2 + \frac{1}{\mathbb{F}\mathbb{r}^2}\rho(\phi)y + \frac{1}{\mathbb{W}\mathbb{e}}\delta_\Gamma(\phi). \quad (6.3.24)$$

We present the local energy balance and subsequently the global balance. To that purpose we first need to introduce some notation and Lemmas associated with the surface energy. Let us define the normal projection operator:

$$\mathbf{P}_N(\phi) := \frac{\nabla\phi}{\|\nabla\phi\|_{\epsilon,2}} \otimes \frac{\nabla\phi}{\|\nabla\phi\|_{\epsilon,2}}, \quad (6.3.25)$$

and the tangential projection operator:

$$\mathbf{P}_T(\phi) := \mathbf{I} - \mathbf{P}_N(\phi). \quad (6.3.26)$$

The associated gradient operators are the gradient along the direction normal to the interface:

$$\nabla_N = \mathbf{P}_N(\phi)\nabla, \quad (6.3.27)$$

and the gradient tangent to the interface:

$$\nabla_\Gamma = \mathbf{P}_T(\phi)\nabla = \nabla - \nabla_N. \quad (6.3.28)$$

**Lemma 6.3.3.** *The term  $\|\nabla\phi\|_{\epsilon,2}$  evolves in time according to:*

$$\partial_t \|\nabla\phi\|_{\epsilon,2} + \nabla \cdot (\|\nabla\phi\|_{\epsilon,2} \mathbf{u}) - \|\nabla\phi\|_{\epsilon,2} \nabla_{\Gamma} \mathbf{u} = 0. \quad (6.3.29)$$

*Proof.* This follows when evaluating the normal derivative of the level-set equation. Taking the gradient of the level-set equation and subsequently evaluating the inner product of the result with  $\nu(\phi)$  yields:

$$\frac{\nabla\phi}{\|\nabla\phi\|_{\epsilon,2}} \cdot \nabla (\partial_t \phi + \mathbf{u} \cdot \nabla\phi) = 0. \quad (6.3.30)$$

Applying the gradient operator to each of the members provides

$$\begin{aligned} \frac{\nabla\phi}{\|\nabla\phi\|_{\epsilon,2}} \cdot \nabla (\partial_t \phi) + \mathbf{u} \cdot \left( \nabla (\nabla\phi) \frac{\nabla\phi}{\|\nabla\phi\|_{\epsilon,2}} \right) \\ + \nabla \mathbf{u} : \left( \frac{\nabla\phi}{\|\nabla\phi\|_{\epsilon,2}} \otimes \nabla\phi \right) = 0. \end{aligned} \quad (6.3.31)$$

The first term in (6.3.31) coincides with the first member in expression (6.3.29). For the second term in (6.3.31) we note that the term in brackets equals the gradient of  $\|\nabla\phi\|_{\epsilon,2}$ . Finally, one recognizes the normal projection operator in the latter term of (6.3.31). This delivers:

$$\partial_t \|\nabla\phi\|_{\epsilon,2} + \mathbf{u} \cdot \nabla \|\nabla\phi\|_{\epsilon,2} + \|\nabla\phi\|_{\epsilon,2} \nabla_N \mathbf{u} = 0. \quad (6.3.32)$$

Adding a suitable partition of zero completes the proof.  $\square$

**Remark 6.3.4.** *The evolution (6.3.29) may be linked to the recently proposed variation entropy theory [59]. Variation entropy is local continuous generalization of the celebrated TVD (total variation diminishing) property derived from entropy principles. It serves as a derivation of discontinuity capturing mechanisms [61]. Using the continuity equation (6.3.20b) we obtain an alternative form of (6.3.29):*

$$\partial_t \eta(\nabla\phi) + \nabla \cdot \left( \eta(\nabla\phi) \frac{\partial \mathbf{f}}{\partial \phi} \right) + \eta(\nabla\phi) \nabla_N \frac{\partial \mathbf{f}}{\partial \phi} = 0, \quad (6.3.33)$$

with  $\eta(\nabla\phi) = \|\nabla\phi\|_{\epsilon,2}$  and  $\mathbf{f}(\phi, \mathbf{u}) = \mathbf{u}\phi$ . In the stationary case, i.e. when the term  $\nabla_N(\partial \mathbf{f} / \partial \phi)$  is absent, relation (6.3.33) represents the evolution of variation entropy  $\eta(\nabla\phi)$ . This occurs when the velocity normal to the interface is constant.

**Lemma 6.3.5.** *The surface Dirac  $\delta_{\Gamma}(\phi)$  evolves in time according to:*

$$\partial_t \delta_{\Gamma}(\phi) + \nabla \cdot (\delta_{\Gamma}(\phi) \mathbf{u}) - \delta_{\Gamma}(\phi) \nabla_{\Gamma} \mathbf{u} = 0. \quad (6.3.34)$$

*Proof.* Multiplying the level-set equation by  $\delta'(\phi)$  provides:

$$\partial_t \delta(\phi) + \mathbf{u} \cdot \nabla \delta(\phi) = 0. \quad (6.3.35)$$

The superposition of (6.3.29) multiplied by  $\delta(\phi)$  and (6.3.35) multiplied by  $\|\nabla\phi\|_{\epsilon,2}$  provides the result. In other words, the operator

$$\delta(\phi) \frac{\nabla\phi}{\|\nabla\phi\|_{\epsilon,2}} \cdot \nabla + \|\nabla\phi\|_{\epsilon,2} \delta'(\phi) \mathcal{I}, \quad (6.3.36)$$

in which  $\mathcal{I}$  denotes the identity operator, applied to the level-set equation delivers the evolution of the surface Dirac (6.3.34).  $\square$

To derive the local energy balance we introduce the following identity.

**Proposition 6.3.6.** *It holds:*

$$-\nabla \cdot (\mathbf{P}_\Gamma(\phi)\delta_\Gamma(\phi)) = \delta_\Gamma(\phi)\mathbf{v}(\phi)\kappa(\phi) - \epsilon^2\delta'(\phi)\mathbf{v}(\phi). \quad (6.3.37)$$

*Proof.* See Appendix 6.A.2. □

We now present the local energy balance.

**Lemma 6.3.7.** *The local energy balance associated with system (6.3.20) takes the form:*

$$\begin{aligned} \partial_t \mathcal{H} + \nabla \cdot (((\mathcal{H} + p)\mathbf{I} - \boldsymbol{\tau}(\mathbf{u}, \phi))\mathbf{u}) - \frac{1}{\mathbb{W}\mathbf{e}} \nabla \cdot (\delta_\Gamma(\phi)\mathbf{P}_\Gamma\mathbf{u}) + \boldsymbol{\tau}(\mathbf{u}, \phi) : \nabla \mathbf{u} \\ + \epsilon^2 \frac{1}{\mathbb{W}\mathbf{e}} \delta'(\phi)u_\nu = 0. \end{aligned} \quad (6.3.38)$$

The divergence terms represent the redistribution of energy over the domain and the second to last term accounts for energy dissipation due to diffusion. The last term that emanates from the regularization is unwanted. We return to this issue in Section 6.4.

*Proof.* First we consider the local kinetic energy of the system (6.3.20). By straightforwardly applying the chain-rule we find:

$$\partial_t \left( \rho \frac{1}{2} \|\mathbf{u}\|_2^2 \right) = \mathbf{u} \cdot \partial_t(\rho\mathbf{u}) - \frac{1}{2} \|\mathbf{u}\|_2^2 \frac{\partial \rho}{\partial \phi} \partial_t \phi. \quad (6.3.39)$$

From the momentum and level-set equations, i.e. (6.3.20a) and (6.3.20c), we deduce:

$$\begin{aligned} \mathbf{u} \cdot \partial_t(\rho\mathbf{u}) - \frac{1}{2} \|\mathbf{u}\|_2^2 \frac{\partial \rho}{\partial \phi} \partial_t \phi = -\mathbf{u} \cdot \nabla \cdot (\rho\mathbf{u} \otimes \mathbf{u}) + \mathbf{u}^T \nabla \cdot \boldsymbol{\tau}(\mathbf{u}, \phi) - \mathbf{u} \cdot \nabla p \\ - \frac{1}{\mathbb{W}\mathbf{e}} \kappa \delta_\Gamma u_\nu - \frac{1}{\mathbb{F}\mathbf{r}^2} \rho \mathbf{u} \cdot \mathbf{j} + \frac{1}{2} \|\mathbf{u}\|_2^2 \frac{\partial \rho}{\partial \phi} \mathbf{u} \cdot \nabla \phi. \end{aligned} \quad (6.3.40)$$

For the energetic contribution due the gravitational force, the chain-rule and the level-set equation (6.3.20c) convey that:

$$\partial_t \left( \frac{1}{\mathbb{F}\mathbf{r}^2} \rho y \right) = \frac{1}{\mathbb{F}\mathbf{r}^2} y \frac{\partial \rho}{\partial \phi} \partial_t \phi = -\frac{1}{\mathbb{F}\mathbf{r}^2} y \frac{\partial \rho}{\partial \phi} \mathbf{u} \cdot \nabla \phi. \quad (6.3.41)$$

And for the local surface energy evolution we invoke Lemma 6.3.5:

$$\begin{aligned} \partial_t \left( \frac{1}{\mathbb{W}\mathbf{e}} \delta_\Gamma(\phi) \right) &= \frac{1}{\mathbb{W}\mathbf{e}} \left( \delta(\phi) \frac{\nabla \phi}{\|\nabla \phi\|_{\epsilon,2}} \cdot \nabla + \|\nabla \phi\|_{\epsilon,2} \delta'(\phi) \mathcal{I} \right) \partial_t \phi \\ &= -\nabla \cdot \left( \frac{1}{\mathbb{W}\mathbf{e}} \delta_\Gamma(\phi) \mathbf{u} \right) + \frac{1}{\mathbb{W}\mathbf{e}} \delta_\Gamma(\phi) \nabla_\Gamma \mathbf{u}. \end{aligned} \quad (6.3.42)$$

Superposition of (6.3.40)-(6.3.42) yields:

$$\begin{aligned} \partial_t \mathcal{H} &= -\mathbf{u} \cdot \nabla \cdot (\rho\mathbf{u} \otimes \mathbf{u}) + \frac{1}{2} \|\mathbf{u}\|_2^2 \frac{\partial \rho}{\partial \phi} \mathbf{u} \cdot \nabla \phi \\ &\quad - \frac{1}{\mathbb{F}\mathbf{r}^2} \rho \mathbf{u} \cdot \mathbf{j} - \frac{1}{\mathbb{F}\mathbf{r}^2} y \frac{\partial \rho}{\partial \phi} \mathbf{u} \cdot \nabla \phi \\ &\quad - \frac{1}{\mathbb{W}\mathbf{e}} \kappa \delta_\Gamma u_\nu - \nabla \cdot \left( \frac{1}{\mathbb{W}\mathbf{e}} \delta_\Gamma(\phi) \mathbf{u} \right) + \frac{1}{\mathbb{W}\mathbf{e}} \delta_\Gamma(\phi) \nabla_\Gamma \mathbf{u} \\ &\quad + \mathbf{u}^T \nabla \cdot \boldsymbol{\tau}(\mathbf{u}, \phi) - \mathbf{u} \cdot \nabla p. \end{aligned} \quad (6.3.43)$$

With the aim of simplifying (6.3.43) we introduce the identities:

$$-\mathbf{u}^T \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u}) + \frac{1}{2} \|\mathbf{u}\|_2^2 \frac{\partial \rho}{\partial \phi} \mathbf{u} \cdot \nabla \phi = -\nabla \cdot \left( \frac{1}{2} \rho \|\mathbf{u}\|_2^2 \mathbf{u} \right) - \frac{1}{2} \rho \|\mathbf{u}\|_2^2 \nabla \cdot \mathbf{u}, \quad (6.3.44a)$$

$$-\frac{1}{\mathbb{F}r^2} \rho \mathbf{u} \cdot \mathbf{j} - \frac{1}{\mathbb{F}r^2} y \frac{\partial \rho}{\partial \phi} \mathbf{u} \cdot \nabla \phi = -\nabla \cdot \left( \frac{1}{\mathbb{F}r^2} \rho y \mathbf{u} \right) + \frac{1}{\mathbb{F}r^2} \rho y \nabla \cdot \mathbf{u}, \quad (6.3.44b)$$

$$\begin{aligned} \frac{1}{\mathbb{W}e} \delta_\Gamma(\phi) \nabla_\Gamma \mathbf{u} &= \nabla \cdot \left( \frac{1}{\mathbb{W}e} \delta_\Gamma(\phi) \mathbf{P}_T \mathbf{u} \right) + \frac{1}{\mathbb{W}e} \delta_\Gamma \kappa u_\nu \\ &\quad - \epsilon^2 \frac{1}{\mathbb{W}e} \delta'(\phi) u_\nu. \end{aligned} \quad (6.3.44c)$$

The first and the second identity follow from expanding the gradient and divergence operators. To obtain the third we note

$$\delta_\Gamma(\phi) \nabla_\Gamma \mathbf{u} = \nabla \cdot (\delta_\Gamma(\phi) \mathbf{P}_T \mathbf{u}) - \mathbf{u} \cdot \nabla (\delta_\Gamma(\phi) \mathbf{P}_T) \quad (6.3.45)$$

and apply Proposition 6.3.6 on the second term. Invoking (6.3.44) into (6.3.43) and adding a suitable partition of zero yields:

$$\begin{aligned} \partial_t \mathcal{H} + \nabla \cdot ((\mathcal{H} + p) \mathbf{I} - \boldsymbol{\tau}(\mathbf{u}, \phi)) \mathbf{u} - \nabla \cdot \left( \frac{1}{\mathbb{W}e} \delta_\Gamma(\phi) \mathbf{P}_T \mathbf{u} \right) \\ = -\boldsymbol{\tau}(\mathbf{u}, \phi) : \nabla \mathbf{u} - \epsilon^2 \frac{1}{\mathbb{W}e} \delta'(\phi) u_\nu \\ + \left( -\frac{1}{2} \rho \|\mathbf{u}\|_2^2 + p + \frac{1}{\mathbb{F}r^2} \rho y \right) \nabla \cdot \mathbf{u}. \end{aligned} \quad (6.3.46)$$

With the aid of the continuity equation (6.3.20b) the latter member on the right-hand side of (6.3.46) vanishes. This completes the proof.  $\square$

**Remark 6.3.8.** *The energy balance of Lemma 6.3.7 may also be written as:*

$$\begin{aligned} \partial_t \mathcal{H} + \nabla \cdot ((\mathcal{H} + p) \mathbf{u}) - \frac{1}{\mathbb{R}e} \nabla \cdot (2\mu(\phi) \nabla \left( \frac{1}{2} \|\mathbf{u}\|_2^2 \right)) - \frac{1}{\mathbb{W}e} \nabla \cdot (\delta_\Gamma(\phi) \mathbf{P}_T \mathbf{u}) \\ + \boldsymbol{\tau}(\mathbf{u}, \phi) : \nabla \mathbf{u} + \epsilon^2 \frac{1}{\mathbb{W}e} \delta'(\phi) u_\nu = 0. \end{aligned} \quad (6.3.47)$$

*In this form we clearly see that the second divergence term represents the diffusion of kinetic energy density.*

We can now present the global energy evolution.

**Theorem 6.3.9.** *Let  $\mathbf{u}, p$  and  $\phi$  be smooth solutions of the strong form (6.3.20). The associated total energy  $\mathcal{E}$ , given in (6.3.23), satisfies the dissipation inequality:*

$$\frac{d}{dt} \mathcal{E}(\mathbf{u}, \phi) = -(\boldsymbol{\tau}(\mathbf{u}, \phi), \nabla \mathbf{u})_\Omega + \mathcal{B} \leq 0 + \mathcal{B}, \quad (6.3.48)$$

where  $\mathcal{B}$  contains the boundary contributions:

$$\mathcal{B} = \int_{\partial\Omega} \mathbf{n}^T \boldsymbol{\tau}(\mathbf{u}, \phi) \mathbf{u} - u_n \left( \rho \frac{1}{2} \|\mathbf{u}\|_2^2 + \frac{1}{\mathbb{F}r^2} \rho y + p \right) dS, \quad (6.3.49)$$

and where we have set  $\epsilon = 0$ .

*Proof.* This follows from integrating the energy balance of Lemma 6.3.7 over  $\Omega$  and using the divergence theorem:

$$\begin{aligned} \int_{\Omega} \partial_t \mathcal{H} \, d\Omega + \int_{\Omega} \boldsymbol{\tau}(\mathbf{u}, \phi) : \nabla \mathbf{u} \, d\Omega + \int_{\partial\Omega} u_n (\mathcal{H} + p) - \mathbf{n}^T \boldsymbol{\tau}(\mathbf{u}, \phi) \mathbf{u} \, d\Omega \\ + \int_{\Omega} \epsilon^2 \frac{1}{\mathbb{W}e} \delta'(\phi) u_v \, d\Omega - \int_{\partial\Omega} \frac{1}{\mathbb{W}e} \delta_{\Gamma}(\phi) \mathbf{n}^T \mathbf{P}_T(\phi) \mathbf{u} \, dS = 0. \end{aligned} \quad (6.3.50)$$

We discard the line force terms on the right-hand side and reorganize to get:

$$\begin{aligned} \frac{d}{dt} \mathcal{E}(\mathbf{u}, \phi) = - \int_{\Omega} \boldsymbol{\tau}(\mathbf{u}, \phi) : \nabla \mathbf{u} \, d\Omega - \int_{\Omega} \epsilon^2 \frac{1}{\mathbb{W}e} \delta'(\phi) u_v \, d\Omega \\ + \int_{\partial\Omega} \mathbf{n}^T \boldsymbol{\tau}(\mathbf{u}, \phi) \mathbf{u} - u_n \left( \rho_{\frac{1}{2}} \|\mathbf{u}\|_2^2 + \frac{1}{\mathbb{F}r^2} \rho y + p \right) \, dS. \end{aligned} \quad (6.3.51)$$

Using the homogeneous boundary condition and setting  $\epsilon = 0$  finalizes the proof.  $\square$

The energy balance associated with the original model (6.2.7) and that of the level-set formulation (6.3.20) comply.

**Corollary 6.3.10.** *The energetic balance associated with regularized model (6.3.20) (Theorem 6.3.9) is consistent that of the original model (6.2.7) (Theorem 6.2.1).*

*Proof.* In the limit  $\epsilon \rightarrow 0$  we may transform (6.3.51) back to get:

$$\begin{aligned} \frac{d}{dt} \mathcal{E}_s(\mathbf{u}) = \int_{\partial\Omega} \mathbf{n}^T \boldsymbol{\tau}(\mathbf{u}) \mathbf{u} - u_n \left( \rho_{\frac{1}{2}} \|\mathbf{u}\|_2^2 + \frac{1}{\mathbb{F}r^2} \rho y + p \right) \, dS \\ - \int_{\Omega} \boldsymbol{\tau}(\mathbf{u}) : \nabla \mathbf{u} \, d\Omega. \end{aligned} \quad (6.3.52)$$

$\square$

To close this section we note that one may avoid evaluating second derivatives appearing in the surface tension term. This holds for the original model (6.2.7) which we have addressed with briefly in Remark 6.2.3. In the following Proposition we note that this alternative form directly converts to the regularized model (6.3.20).

**Proposition 6.3.11.** *We have the identity:*

$$\begin{aligned} \int_{\Omega} \frac{1}{\mathbb{W}e} \delta_{\Gamma}(\phi) \kappa(\phi) \boldsymbol{\nu}(\phi) \cdot \mathbf{w} \, d\Omega = \int_{\Omega} \frac{1}{\mathbb{W}e} \delta_{\Gamma}(\phi) \nabla \mathbf{w} : \mathbf{P}_T(\phi) \, d\Omega \\ + \frac{1}{\mathbb{W}e} \int_{\Omega} \epsilon^2 \delta'(\phi) \boldsymbol{\nu}(\phi) \mathbf{w} \, d\Omega. \end{aligned} \quad (6.3.53)$$

*Proof.* See Appendix 6.A.2.  $\square$

With the aid of Proposition 6.3.11 one can directly evaluate the surface tension term and does not require any additional procedure such as the one from [120].

## 6.4 ENERGY-DISSIPATIVE FORMULATION

We aim to develop an energetically stable Galerkin-type finite element method for the regularized level-set model (6.3.20). In Sections 6.2 and 6.3 we have in great detail depicted the procedure to arrive at the energy-dissipative statement. This procedure involves several steps that are not valid when dealing with standard finite element discretization spaces. For instance the operator (6.3.36) associated with the surface energy is not permissible in a standard discrete setting. Independently, the temporal discretization also gives rise to issues. Standard second-order semi time-discrete formulations of (6.3.20) are also not equipped with an energy-dissipative structure. We demonstrate this in Appendix 6.B. Lastly, we note that the standard regularized-interface model contains an unwanted term stemming from the regularization.

The first two issues arise from the fact that the standard model is too restrictive with regard to the function spaces. Enlarging the standard function spaces introduces many complications and as such we do not further look into this strategy. The alternative is modify the regularized model (6.3.20). This is the road we pursue. We employ the concept of *functional entropy variables* proposed by Liu et al. [144]. Liu and co-workers introduce the concept of functional entropy variables for the isothermal Navier-Stokes-Korteweg equations [144] and for the Navier-Stokes-Korteweg equations including the interstitial working flux term [145]. Here we apply the formalism to the level-set formulation of the incompressible Navier-Stokes equations with surface tension. This creates the extra space to resolve both discrepancies mentioned above. Additionally, the unwanted regularization term also vanishes.

## 6.4.1 Functional entropy variables

Energetic stability for the incompressible Navier-Stokes equations with surface tension coincides with stability with respect to a mathematical entropy function. Thus to construct an energy-dissipative formulation for the incompressible Navier-Stokes equations the natural approach seems to adopt entropy principles. For systems of conservation laws classical entropy variables are defined as the partial derivatives of an entropy with respect to the conservation variables. The Clausius-Duhem inequality plays the role of energetic stability and this results from pre-multiplication of the system of conservation laws by the entropy variables. The standard approach of constructing an entropy stable discretization as in Hughes et al. [107, 171] is not applicable since the mathematical entropy is not an algebraic function of the conservation variables. In the situation of a general mathematical entropy functional the derivatives should be taken in the functional setting. The corresponding Clausius-Duhem inequality is then the result from the action of the entropy variables on the system of conservation laws.

In the current study we wish to inherit the notion of energetic stability for the incompressible model with surface tension. To this purpose we use as mathematical entropy functional the energy density (6.3.24) which we recall here:

$$\mathcal{H} = \frac{1}{2}\rho\|\mathbf{u}\|_2^2 + \frac{1}{\text{Fr}^2}\rho y + \frac{1}{\text{We}}\delta_\Gamma. \quad (6.4.1)$$

Following the approach described above, energetic stability results from the action of the entropy variables on the system of equations. In contrast to [144] and [145] the notion of conservation variables does not exist. Instead, the derivatives of  $\mathcal{H}$  should here be taken with respect to the model variables  $\mathbf{U} = (\phi, \rho \mathbf{u})$ . Remark that (6.4.1) is a functional of the model variables  $\mathbf{U}$ :

$$\mathcal{H} = \mathcal{H}(\mathbf{U}) = \frac{\|\rho \mathbf{u}\|_2^2}{2\rho(\phi)} + \frac{1}{\mathbb{F}r^2} \rho(\phi) y + \frac{1}{\mathbb{W}e} \delta(\phi) \|\nabla \phi\|_{\epsilon,2}. \quad (6.4.2)$$

Note that  $\mathcal{H}$  contains a gradient term  $\|\nabla \phi\|_{\epsilon,2}$  which is non-local and thus the appropriate derivative is the functional derivative. We define the entropy variables as functional derivatives:

$$\mathbf{V} = [V_1; V_2; V_3; V_4]^T := \frac{\delta \mathcal{H}}{\delta \mathbf{U}} = \left[ \frac{\delta \mathcal{H}}{\delta \phi}; \frac{\delta \mathcal{H}}{\delta(\rho u_1)}; \frac{\delta \mathcal{H}}{\delta(\rho u_2)}; \frac{\delta \mathcal{H}}{\delta(\rho u_3)} \right]^T. \quad (6.4.3)$$

The resulting functional derivatives are for test functions  $\delta v = [\delta v_1, \delta v_2, \delta v_3, \delta v_4]^T$ :

$$\begin{aligned} \frac{\delta \mathcal{H}}{\delta \phi} [\delta v_1] &= -\frac{1}{2} \|\mathbf{u}\|_2^2 \rho'(\phi) \delta v_1 + \frac{1}{\mathbb{F}r^2} \rho'(\phi) y \delta v_1 \\ &\quad + \frac{1}{\mathbb{W}e} \delta(\phi) \frac{\nabla \phi}{\|\nabla \phi\|_{\epsilon,2}} \cdot \nabla \delta v_1 + \frac{1}{\mathbb{W}e} \|\nabla \phi\|_{\epsilon,2} \delta'(\phi) \delta v_1, \end{aligned} \quad (6.4.4a)$$

$$\frac{\delta \mathcal{H}}{\delta(\rho u_1)} [\delta v_2] = u_1 \delta v_2, \quad (6.4.4b)$$

$$\frac{\delta \mathcal{H}}{\delta(\rho u_2)} [\delta v_3] = u_2 \delta v_3, \quad (6.4.4c)$$

$$\frac{\delta \mathcal{H}}{\delta(\rho u_3)} [\delta v_4] = u_3 \delta v_4. \quad (6.4.4d)$$

We emphasize that it is essential to use the expression in terms of the model variables (6.4.2) to evaluate (6.4.4). The associated explicit form of (6.4.4) reads:

$$\begin{aligned} \frac{\delta \mathcal{H}}{\delta \phi} &= -\frac{1}{2} \|\mathbf{u}\|_2^2 \rho'(\phi) + \frac{1}{\mathbb{F}r^2} \rho'(\phi) y - \frac{1}{\mathbb{W}e} \delta(\phi) \nabla \cdot \left( \frac{\nabla \phi}{\|\nabla \phi\|_{\epsilon,2}} \right) \\ &\quad + \frac{1}{\mathbb{W}e} \delta'(\phi) \frac{\epsilon^2}{\|\nabla \phi\|_{\epsilon,2}}, \end{aligned} \quad (6.4.5a)$$

$$\frac{\delta \mathcal{H}}{\delta(\rho \mathbf{u})} = \mathbf{u}^T. \quad (6.4.5b)$$

We may use the functional entropy variables to systematically recover the energy balance (6.3.38).

**Theorem 6.4.1.** *Applying the functional entropy variables to the incompressible two-phase Navier-Stokes equations with surface tension recovers the energy balance (6.3.38):*

$$\begin{aligned} \partial_t \mathcal{H} + \nabla \cdot (((\mathcal{H} + p) \mathbf{I} - \boldsymbol{\tau}(\mathbf{u}, \phi)) \mathbf{u}) + \boldsymbol{\tau}(\mathbf{u}, \phi) : \nabla \mathbf{u} - \frac{1}{\mathbb{W}e} \nabla \cdot (\delta_\Gamma(\phi) \mathbf{P}_T \mathbf{u}) \\ + \epsilon^2 \frac{1}{\mathbb{W}e} \delta'(\phi) u_\nu = 0. \end{aligned} \quad (6.4.6)$$

*Proof.* Application of the functional entropy variables on the time-derivatives provides:

$$\mathbf{V} \left[ \frac{\partial \mathbf{U}}{\partial t} \right] = \frac{\delta \mathcal{H}}{\delta \mathbf{U}} \left[ \frac{\partial \mathbf{U}}{\partial t} \right] = \frac{\partial \mathcal{H}}{\partial t}. \quad (6.4.7)$$

Next we apply the entropy variables on the fluxes to get:

$$\begin{aligned} \mathbf{V} \left[ \begin{array}{c} \mathbf{u} \cdot \nabla \phi \\ \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u}) + \nabla p \end{array} \right] &= - \left( \frac{1}{2} \|\mathbf{u}\|_2^2 - \frac{1}{\mathbb{F}r^2} y \right) \frac{\partial \rho}{\partial \phi} \mathbf{u} \cdot \nabla \phi + \mathbf{u}^T \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u}) \\ &\quad + \nabla \cdot (p \mathbf{u}) - p \nabla \cdot \mathbf{u} \\ &\quad + \frac{1}{\mathbb{W}e} \delta(\phi) \frac{\nabla \phi}{\|\nabla \phi\|_{\epsilon, 2}} \cdot \nabla (\mathbf{u} \cdot \nabla \phi) \\ &\quad + \frac{1}{\mathbb{W}e} \|\nabla \phi\|_{\epsilon, 2} \delta'(\phi) (\mathbf{u} \cdot \nabla \phi). \end{aligned} \quad (6.4.8)$$

Testing the entropy variables with the surface tension term gives:

$$\mathbf{V} \left[ \begin{array}{c} 0 \\ \frac{1}{\mathbb{W}e} \delta_\Gamma(\phi) \nu(\phi) \kappa(\phi) \end{array} \right] = \frac{1}{\mathbb{W}e} \delta_\Gamma(\phi) \kappa(\phi) u_\nu(\phi). \quad (6.4.9)$$

Testing the entropy variables with the viscous stress yields:

$$\mathbf{V} \left[ \begin{array}{c} 0 \\ -\nabla \cdot \boldsymbol{\tau}(\mathbf{u}, \phi) \end{array} \right] = -\nabla \cdot (\boldsymbol{\tau}(\mathbf{u}, \phi) \mathbf{u}) + \boldsymbol{\tau}(\mathbf{u}, \phi) : \nabla \mathbf{u}. \quad (6.4.10)$$

And finally testing with the body force yields:

$$\mathbf{V} \left[ \begin{array}{c} 0 \\ \frac{1}{\mathbb{F}r^2} \rho \mathbf{j} \end{array} \right] = \frac{1}{\mathbb{F}r^2} \rho \mathbf{u} \cdot \mathbf{j}. \quad (6.4.11)$$

Addition of (6.4.8), (6.4.9), (6.4.10) and (6.4.11) gives:

$$\begin{aligned} \mathbf{V} \left[ \begin{array}{c} \mathbf{u} \cdot \nabla \phi \\ \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u}) + \nabla p - \nabla \cdot \boldsymbol{\tau} + \frac{1}{\mathbb{F}r^2} \rho \mathbf{j} + \frac{1}{\mathbb{W}e} \delta_\Gamma(\phi) \nu(\phi) \kappa(\phi) \end{array} \right] \\ &= -\frac{1}{2} \|\mathbf{u}\|_2^2 \frac{\partial \rho}{\partial \phi} \mathbf{u} \cdot \nabla \phi + \mathbf{u}^T \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u}) \\ &\quad + \nabla \cdot (p \mathbf{u}) - p \nabla \cdot \mathbf{u} \\ &\quad + \frac{1}{\mathbb{F}r^2} \rho \mathbf{u} \cdot \mathbf{j} + \frac{1}{\mathbb{F}r^2} y \mathbf{u} \cdot \nabla \rho \\ &\quad + \frac{1}{\mathbb{W}e} \delta(\phi) \frac{\nabla \phi}{\|\nabla \phi\|_{\epsilon, 2}} \cdot \nabla (\mathbf{u} \cdot \nabla \phi) + \frac{1}{\mathbb{W}e} \|\nabla \phi\|_{\epsilon, 2} \delta'(\phi) (\mathbf{u} \cdot \nabla \phi) \\ &\quad + \frac{1}{\mathbb{W}e} \nu(\phi) \kappa(\phi) u_\nu(\phi) \\ &\quad - \nabla \cdot (\boldsymbol{\tau}(\mathbf{u}, \phi) \mathbf{u}) + \boldsymbol{\tau}(\mathbf{u}, \phi) : \nabla \mathbf{u}. \end{aligned} \quad (6.4.12)$$

Recognize the operator (6.3.36) on the fourth line of the right-hand side of (6.4.12). We may thus use Lemma 6.3.5 and write:

$$\begin{aligned} & \frac{1}{\mathbb{W}e} \delta(\phi) \frac{\nabla \phi}{\|\nabla \phi\|_{\epsilon,2}} \cdot \nabla(\mathbf{u} \cdot \nabla \phi) + \frac{1}{\mathbb{W}e} \|\nabla \phi\|_{\epsilon,2} \delta'(\phi) (\mathbf{u} \cdot \nabla \phi) \\ & = \nabla \cdot \left( \frac{1}{\mathbb{W}e} \delta_{\Gamma}(\phi) \mathbf{u} \right) - \frac{1}{\mathbb{W}e} \delta_{\Gamma}(\phi) \nabla_{\Gamma} \mathbf{u}. \end{aligned} \quad (6.4.13)$$

Invoking the identities (6.3.44) and (6.4.13) the expression (6.4.12) collapses to

$$\begin{aligned} \mathbf{V} & \left[ \begin{array}{c} \mathbf{u} \cdot \nabla \phi \\ \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u}) + \nabla p - \nabla \cdot \boldsymbol{\tau}(\mathbf{u}, \phi) + \frac{1}{\mathbb{F}r^2} \rho \mathbf{j} + \frac{1}{\mathbb{W}e} \delta_{\Gamma}(\phi) \nu(\phi) \kappa(\phi) \end{array} \right] \\ & = \nabla \cdot \left( \frac{1}{2} \rho \|\mathbf{u}\|^2 \mathbf{u} \right) + \frac{1}{2} \rho \|\mathbf{u}\|^2 \nabla \cdot \mathbf{u} \\ & \quad + \nabla \cdot (p \mathbf{u}) - p \nabla \cdot \mathbf{u} \\ & \quad + \nabla \cdot \left( \frac{1}{\mathbb{F}r^2} \rho y \mathbf{u} \right) - \frac{1}{\mathbb{F}r^2} \rho y \nabla \cdot \mathbf{u} \\ & \quad + \nabla \cdot \left( \frac{1}{\mathbb{W}e} \delta_{\Gamma}(\phi) \mathbf{u} \right) - \nabla \cdot \left( \frac{1}{\mathbb{W}e} \delta_{\Gamma}(\phi) \mathbf{P}_T \mathbf{u} \right) \\ & \quad + \epsilon^2 \frac{1}{\mathbb{W}e} \delta'(\phi) u_{\nu} = 0 \\ & \quad - \nabla \cdot (\boldsymbol{\tau}(\mathbf{u}, \phi) \mathbf{u}) + \boldsymbol{\tau}(\mathbf{u}, \phi) : \nabla \mathbf{u}. \end{aligned} \quad (6.4.14)$$

We merge the terms in (6.4.14) and use the continuity equation (6.3.20b) to cancel the terms containing the divergence of velocity. Taking the superposition of (6.4.7) and (6.4.14) while recognizing  $\mathcal{H}$  on the right-hand side of (6.4.14) completes the proof.  $\square$

#### 6.4.2 Modified formulation

Theorem 6.4.1 implies that an energy-dissipative relation may be recovered when the functional entropy variables are available as test functions. For standard test function spaces we can not select the weight  $V_1$ . We circumvent this issue, similar as in [144], by explicitly adding  $V_1$  as a new unknown  $v$  to the system of equations. Thus we introduce the extra variable:

$$v := -\frac{\rho}{2} \|\mathbf{u}\|_2^2 + \frac{1}{\mathbb{F}r^2} \rho y - \frac{1}{\mathbb{W}e} \delta(\phi) \nabla \cdot \left( \frac{\nabla \phi}{\|\nabla \phi\|_{\epsilon,2}} \right) + \frac{1}{\mathbb{W}e} \delta'(\phi) \frac{\epsilon^2}{\|\nabla \phi\|_{\epsilon,2}}, \quad (6.4.15)$$

where we use the notation  $\varrho = \varrho(\phi) := \rho'(\phi)$ . The question arises how to couple the extra variable (6.4.15) to the regularized-interface model (6.3.20). In this regard, note that a direct consequence of (6.4.15) is:

$$\begin{aligned} - \left( v + \frac{\varrho}{2} \|\mathbf{u}\|_2^2 - \frac{1}{\mathbb{F}\mathbb{R}^2} \varrho y \right) \nabla \phi &= \frac{1}{\mathbb{W}\mathbb{e}} \delta(\phi) \nabla \cdot \left( \frac{\nabla \phi}{\|\nabla \phi\|_{\epsilon,2}} \right) \nabla \phi \\ &\quad - \frac{1}{\mathbb{W}\mathbb{e}} \delta'(\phi) \frac{\epsilon^2}{\|\nabla \phi\|_{\epsilon,2}} \nabla \phi \\ &= \frac{1}{\mathbb{W}\mathbb{e}} \nabla \cdot \left( \frac{\nabla \phi}{\|\nabla \phi\|_{\epsilon,2}} \right) \frac{\nabla \phi}{\|\nabla \phi\|_{\epsilon,2}} \delta_{\Gamma}(\phi) \\ &\quad - \epsilon^2 \frac{1}{\mathbb{W}\mathbb{e}} \delta'(\phi) \nu(\phi). \end{aligned} \quad (6.4.16)$$

Recall that the regularized-interface model (6.3.20) is only associated with an energy-dissipative structure for  $\epsilon = 0$ , see Theorem 6.3.9. This dissipative structure does not change when performing a consistent modification. Thus adding a suitable partition of zero based on (6.4.16) to the momentum equation (6.3.20a) keeps the same energy behavior. Instead, we suggest to replace the surface tension term in (6.3.20), i.e.

$$\frac{1}{\mathbb{W}\mathbb{e}} \nabla \cdot \left( \frac{\nabla \phi}{\|\nabla \phi\|_{\epsilon,2}} \right) \frac{\nabla \phi}{\|\nabla \phi\|_{\epsilon,2}} \delta_{\Gamma}(\phi), \quad (6.4.17)$$

by the left-hand side of (6.4.16), i.e.

$$- \left( v + \frac{\varrho}{2} \|\mathbf{u}\|_2^2 - \frac{1}{\mathbb{F}\mathbb{R}^2} \varrho y \right) \nabla \phi. \quad (6.4.18)$$

In this way we eliminate the unwanted regularization term. The new strong form writes in terms of the variables  $\mathbf{u}$ ,  $p$ ,  $\phi$  and  $v$  as:

$$\begin{aligned} \partial_t(\rho(\phi)\mathbf{u}) + \nabla \cdot (\rho(\phi)\mathbf{u} \otimes \mathbf{u}) - \nabla \cdot \boldsymbol{\tau}(\mathbf{u}, \phi) + \nabla p \\ - \left( v + \frac{\varrho}{2} \|\mathbf{u}\|_2^2 - \frac{1}{\mathbb{F}\mathbb{R}^2} \varrho y \right) \nabla \phi + \frac{1}{\mathbb{F}\mathbb{R}^2} \rho(\phi) \mathbf{j} = 0, \end{aligned} \quad (6.4.19a)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (6.4.19b)$$

$$\partial_t \phi + \mathbf{u} \cdot \nabla \phi = 0, \quad (6.4.19c)$$

$$\begin{aligned} v + \varrho \frac{\|\mathbf{u}\|_2^2}{2} - \frac{1}{\mathbb{F}\mathbb{R}^2} \varrho y + \frac{1}{\mathbb{W}\mathbb{e}} \delta(\phi) \nabla \cdot \left( \frac{\nabla \phi}{\|\nabla \phi\|_{\epsilon,2}} \right) \\ - \frac{1}{\mathbb{W}\mathbb{e}} \delta'(\phi) \frac{\epsilon^2}{\|\nabla \phi\|_{\epsilon,2}} = 0, \end{aligned} \quad (6.4.19d)$$

with  $\mathbf{u}(0) = \mathbf{u}_0$  and  $\phi(0) = \phi_0$  in  $\Omega$ .

**Remark 6.4.2.** *Even in absence of surface tension effects the substitution (6.4.16) is essential to arrive at an energy-dissipative system.*

The corresponding weak formulation reads:

Find  $(\mathbf{u}, p, \phi, v) \in \mathcal{W}_T$  such that for all  $(\mathbf{w}, q, \psi, \zeta) \in \mathcal{W}$ :

$$(\mathbf{w}, \partial_t(\rho \mathbf{u}))_\Omega - (\nabla \mathbf{w}, \rho \mathbf{u} \otimes \mathbf{u})_\Omega - (\nabla \cdot \mathbf{w}, p)_\Omega + (\nabla \mathbf{w}, \boldsymbol{\tau}(\mathbf{u}, \phi))_\Omega + \frac{1}{\mathbb{F}r^2}(\mathbf{w}, \rho \mathbf{j})_\Omega - (\mathbf{w}, v \nabla \phi)_\Omega - \left( \mathbf{w}, \left( \frac{\varrho}{2} \|\mathbf{u}\|_2^2 - \frac{1}{\mathbb{F}r^2} \varrho y \right) \nabla \phi \right)_\Omega = 0, \quad (6.4.20a)$$

$$(q, \nabla \cdot \mathbf{u})_\Omega = 0, \quad (6.4.20b)$$

$$(\psi, \partial_t \phi + \mathbf{u} \cdot \nabla \phi)_\Omega = 0, \quad (6.4.20c)$$

$$\left( \zeta, v + \varrho \frac{\|\mathbf{u}\|_2^2}{2} - \frac{1}{\mathbb{F}r^2} \varrho y \right)_\Omega - \left( \frac{1}{\mathbb{W}e} \delta(\phi) \frac{\nabla \phi}{\|\nabla \phi\|_{\epsilon, 2}}, \nabla \zeta \right)_\Omega - \left( \frac{1}{\mathbb{W}e} \|\nabla \phi\|_{\epsilon, 2} \delta'(\phi), \zeta \right)_\Omega = 0, \quad (6.4.20d)$$

where we recall  $\varrho = \partial \rho / \partial \phi$  and have  $\mathbf{u}(x, 0) = \mathbf{u}_0(x)$  and  $\phi(x, 0) = \phi_0(x)$  in  $\Omega$ . The solution space  $\mathcal{W}_T$  and corresponding test-function space  $\mathcal{W}$  are divergence-compatible. We take  $\mathcal{W}_T := \mathcal{V}_T \times \mathcal{Q}_T^3$  and  $\mathcal{W} := \mathcal{V}_T \times \mathcal{Q}_T^3$  where we refer to [65, 69] for the precise definitions of  $\mathcal{V}_T$  and  $\mathcal{Q}_T$ .

**Theorem 6.4.3.** *Let  $(\mathbf{u}, p, \phi)$  be a smooth solution of the weak form (6.4.20). The formulation (6.4.20) has the properties:*

1. *The formulation satisfies the maximum principle for the density, i.e. without loss of generality we assume that  $\rho_2 \leq \rho_1$  and then have:*

$$\rho_2 \leq \rho(\phi) \leq \rho_1. \quad (6.4.21)$$

2. *The formulation is divergence-free as a distribution:*

$$\nabla \cdot \mathbf{u} \equiv 0. \quad (6.4.22)$$

3. *The formulation satisfies the dissipation inequality:*

$$\frac{d}{dt} \mathcal{E}(\mathbf{u}, \phi) = -(\nabla \mathbf{u}, \boldsymbol{\tau}(\mathbf{u}, \phi))_\Omega \leq 0. \quad (6.4.23)$$

Dissipation inequality (6.4.23) is not equipped with terms supported on the outer boundary  $\partial\Omega$  since these vanish due to assumed boundary conditions.

*Proof.* 1. This is a direct consequence of the definition of  $\rho = \rho(\phi)$ .

2. The divergence-conforming space allows to take  $q = \nabla \cdot \mathbf{u}$  in (6.4.20b) and hence we find:

$$0 = (\nabla \cdot \mathbf{u}, \nabla \cdot \mathbf{u})_\Omega \quad \Rightarrow \quad \nabla \cdot \mathbf{u} \equiv 0. \quad (6.4.24)$$

3. Selection of the weights  $\psi = v$  in (6.4.20c) and  $\zeta = -\partial_t \phi$  in (6.4.20d) yields:

$$(v, \partial_t \phi + \mathbf{u} \cdot \nabla \phi)_\Omega = 0, \quad (6.4.25a)$$

$$-\left( \partial_t \phi, v + \varrho \frac{\|\mathbf{u}\|_2^2}{2} - \frac{1}{\mathbb{F}r^2} \varrho y \right)_\Omega + \left( \frac{1}{\mathbb{W}e} \delta(\phi) \frac{\nabla \phi}{\|\nabla \phi\|_{\epsilon, 2}}, \nabla \partial_t \phi \right)_\Omega + \left( \frac{1}{\mathbb{W}e} \|\nabla \phi\|_{\epsilon, 2} \delta'(\phi), \partial_t \phi \right)_\Omega = 0. \quad (6.4.25b)$$

We add the equations (6.4.25) and find:

$$\begin{aligned} & (v, \mathbf{u} \cdot \nabla \phi)_\Omega - \left( \frac{\varrho}{2} \|\mathbf{u}\|_2^2, \partial_t \phi \right)_\Omega + \frac{1}{\mathbb{F}r^2} (\partial_t \phi, \varrho y)_\Omega \\ & + \left( \frac{1}{\mathbb{W}e} \delta(\phi) \frac{\nabla \phi}{\|\nabla \phi\|_{\epsilon,2}}, \nabla \partial_t \phi \right)_\Omega + \left( \frac{1}{\mathbb{W}e} \|\nabla \phi\|_{\epsilon,2} \delta'(\phi), \partial_t \phi \right)_\Omega = 0. \end{aligned} \quad (6.4.26)$$

Performing integration by parts yields:

$$\begin{aligned} & \left( \partial_t \phi, -\frac{\varrho}{2} \|\mathbf{u}\|_2^2 + \varrho \frac{1}{\mathbb{F}r^2} y - \frac{1}{\mathbb{W}e} \delta(\phi) \nabla \cdot \left( \frac{\nabla \phi}{\|\nabla \phi\|_{\epsilon,2}} \right) \right. \\ & \left. + \frac{1}{\mathbb{W}e} \delta'(\phi) \frac{\epsilon^2}{\|\nabla \phi\|_{\epsilon,2}} \right)_\Omega = - (v, \mathbf{u} \cdot \nabla \phi)_\Omega. \end{aligned} \quad (6.4.27)$$

Recall that the line integral terms vanish due to auxiliary boundary conditions. Noting that  $\frac{\delta \mathcal{H}}{\delta \phi} = -\frac{\varrho}{2} \|\mathbf{u}\|_2^2 + \varrho \frac{1}{\mathbb{F}r^2} y - \frac{1}{\mathbb{W}e} \delta(\phi) \nabla \cdot \left( \frac{\nabla \phi}{\|\nabla \phi\|_{\epsilon,2}} \right) + \frac{1}{\mathbb{W}e} \delta'(\phi) \frac{\epsilon^2}{\|\nabla \phi\|_{\epsilon,2}}$  we arrive at:

$$\frac{\delta \mathcal{E}}{\delta \phi} \left[ \frac{\partial \phi}{\partial t} \right] := \left( \frac{\partial \phi}{\partial t}, \frac{\delta \mathcal{H}}{\delta \phi} \right)_\Omega = - (v, \mathbf{u} \cdot \nabla \phi)_\Omega. \quad (6.4.28)$$

Next we take  $\mathbf{w} = \mathbf{u}$  in (6.4.20a) to get:

$$\begin{aligned} & (\mathbf{u}, \partial_t(\rho \mathbf{u}))_\Omega - (\nabla \mathbf{u}, \rho \mathbf{u} \otimes \mathbf{u})_\Omega - (\mathbf{w}, \frac{1}{2} \|\mathbf{u}\|_2^2 \varrho(\phi) \nabla \phi)_\Omega - (\nabla \cdot \mathbf{u}, p)_\Omega \\ & + (\nabla \mathbf{u}, \boldsymbol{\tau}(\mathbf{u}, \phi))_\Omega - (\mathbf{u}, v \nabla \phi)_\Omega + \frac{1}{\mathbb{F}r^2} (\mathbf{u}, \varrho y \nabla \phi)_\Omega + \frac{1}{\mathbb{F}r^2} (\mathbf{u}, \rho \mathbf{j})_\Omega = 0. \end{aligned} \quad (6.4.29)$$

From the identities (6.3.44), the continuity equation (6.4.22), homogeneous boundary conditions and integration by parts we extract the identities:

$$-(\nabla \mathbf{u}, \rho \mathbf{u} \otimes \mathbf{u})_\Omega - (\mathbf{u}, \frac{1}{2} \|\mathbf{u}\|_2^2 \varrho(\phi) \nabla \phi)_\Omega = 0 \quad (6.4.30a)$$

$$-(\nabla \cdot \mathbf{u}, p)_\Omega = 0, \quad (6.4.30b)$$

$$\frac{1}{\mathbb{F}r^2} (\mathbf{u}, \varrho y \nabla \phi)_\Omega + \frac{1}{\mathbb{F}r^2} (\mathbf{u}, \rho \mathbf{j})_\Omega = 0. \quad (6.4.30c)$$

Noting that  $\frac{\delta \mathcal{H}}{\delta(\rho \mathbf{u})} = \mathbf{u}^T$  and employing (6.4.30) we arrive at:

$$\frac{\delta \mathcal{E}}{\delta(\rho \mathbf{u})} \left[ \frac{\partial(\rho \mathbf{u})}{\partial t} \right] := \left( \frac{\partial(\rho \mathbf{u})}{\partial t}, \frac{\delta \mathcal{H}}{\delta(\rho \mathbf{u})} \right)_\Omega = - (\nabla \mathbf{u}, \boldsymbol{\tau}(\mathbf{u}, \phi))_\Omega + (\mathbf{u}, v \nabla \phi)_\Omega. \quad (6.4.31)$$

Addition of (6.4.28) and (6.4.31) yields:

$$\frac{d}{dt} \mathcal{E} = \frac{\delta \mathcal{E}}{\delta \phi} \left[ \frac{\partial \phi}{\partial t} \right] + \frac{\delta \mathcal{E}}{\delta(\rho \mathbf{u})} \left[ \frac{\partial(\rho \mathbf{u})}{\partial t} \right] = - (\nabla \mathbf{u}, \boldsymbol{\tau}(\mathbf{u}, \phi))_\Omega. \quad (6.4.32)$$

□

## 6.5 ENERGY-DISSIPATIVE SPATIAL DISCRETIZATION

In this section we present the spatial discretization of the modified model (6.4.20). First we introduce some notation, then discuss the stabilization mechanisms and subsequently provide the semi-discrete formulation.

## 6.5.1 Notation

We employ an isogeometric analysis discretization. To provide the appropriate setting, we introduce the parametric domain denoted as  $\hat{\Omega} := (-1, 1)^d \subset \mathbb{R}^d$  with corresponding mesh  $\mathcal{M}$ . The element size  $h_Q = \text{diag}(Q)$  of an element  $Q$  in  $\mathcal{M}$  is its diagonal length. The physical domain  $\Omega \subset \mathbb{R}^d$  follows as usual via the continuously differentiable geometrical map (with continuously differentiable inverse)  $\mathbf{F} : \hat{\Omega} \rightarrow \Omega$  and the corresponding physical mesh reads:

$$\mathcal{K} = \mathbf{F}(\mathcal{M}) := \{\Omega_K : \Omega_K = \mathbf{F}(Q), Q \in \mathcal{M}\}. \quad (6.5.1)$$

The Jacobian mapping is  $\mathbf{J} = \partial \mathbf{x} / \partial \xi$ . The physical mesh size  $h_K$  is given by

$$h_K^2 := \frac{h_Q^2}{d} \|\mathbf{J}\|_F^2, \quad (6.5.2)$$

with the subscript  $F$  referring to the Frobenius norm. Note that on a Cartesian mesh it reduces to the diagonal-length of an element. The element metric tensor reads

$$\mathbf{G} = \frac{\partial \xi^T}{\partial \mathbf{x}} \frac{\partial \xi}{\partial \mathbf{x}} = \mathbf{J}^{-T} \mathbf{J}^{-1}, \quad (6.5.3)$$

with inverse

$$\mathbf{G}^{-1} = \frac{\partial \mathbf{x}}{\partial \xi} \frac{\partial \mathbf{x}^T}{\partial \xi} = \mathbf{J} \mathbf{J}^T. \quad (6.5.4)$$

Using the metric tensor we see that the Frobenius norm is objective:

$$\|\mathbf{J}\|_F^2 = \text{Tr}(\mathbf{G}^{-1}), \quad (6.5.5)$$

where  $\text{Tr}$  denotes the trace operator.

We define approximation spaces  $\mathcal{W}_T^h \subset \mathcal{W}_T$ ,  $\mathcal{W}^h \subset \mathcal{W}$  spanned by finite element or NURBS basis functions. The div-conforming solution space is  $\mathcal{W}_T^h := \mathcal{V}_T^h \times (\mathcal{Q}_T^h)^3$  and the corresponding test-function space is  $\mathcal{W}_{0,h} := \mathcal{V}_{0,h} \times (\mathcal{Q}_{0,h})^3$ . We refer to [68, 69] for the precise definitions. Furthermore, we use the conventional notation superscript  $h$  to indicate the discretized (vector) field of the corresponding quantity.

## 6.5.2 Stabilization

It is well-known that a plain Galerkin discretization is prone to the development of numerical instabilities. This motivates the use of stabilization mechanisms. We employ the standard SUPG stabilization [33] for the level-set convection, i.e. we augment the discrete level-set equation with

$$+ \sum_K \left( \tau_K \mathbf{u}^h \cdot \nabla \psi^h, \mathcal{R}_I \phi^h \right)_{\Omega_K}, \quad (6.5.6)$$

with residual

$$\mathcal{R}_I \phi^h := \partial_t \phi^h + \mathbf{u}^h \cdot \nabla \phi^h. \quad (6.5.7)$$

We use the standard definition for stabilization parameter  $\tau$  as also given in [57]. To ensure that the stabilization term does not upset the energetic stability property we balance it with the term:

$$-\sum_K \left( \tau_K \mathbf{w}^h \cdot \nabla v^h, \mathcal{R}_1 \phi^h \right)_{\Omega_K} \quad (6.5.8)$$

in the momentum equation.

**Remark 6.5.1.** *In the current work we focus on an energy-dissipative method without multiscale stabilization contributions in the momentum equation such as [17]. Standard stabilized methods are not directly associated with an energy-dissipative property and thus specific techniques are required to establish such a property, see e.g., [58, 71, 159]. We note that these methods are developed for the single-fluid case. An extension to the current two-fluid case may be the topic of another work.*

A popular method to stabilize the momentum equation is to use discontinuity capturing devices. We follow this road and augment the momentum equation with the discontinuity capturing term:

$$+\sum_K \left( \nabla \mathbf{w}^h, \theta_K \nabla \mathbf{u}^h \right)_{\Omega_K}. \quad (6.5.9)$$

The discontinuity capturing viscosity is given by:

$$\theta_K = Ch_K \frac{\|\mathcal{R}_M(\rho^h \mathbf{u}^h)\|_{\epsilon,2}}{\|\nabla \mathbf{u}^h\|_{\epsilon,2}}, \quad (6.5.10)$$

with conservative momentum residual

$$\begin{aligned} \mathcal{R}_M(\rho^h \mathbf{u}^h) &:= \partial_t(\rho^h \mathbf{u}^h) + \nabla \cdot (\rho^h \mathbf{u}^h \otimes \mathbf{u}^h) + \nabla \cdot \boldsymbol{\tau}(\mathbf{u}^h, \phi^h) + \nabla p^h \\ &+ \frac{1}{\text{We}} \delta(\phi^h) \kappa \nabla \phi^h + \frac{1}{\text{Fr}^2} \rho^h \mathbf{J}, \end{aligned} \quad (6.5.11)$$

and  $C$  a user-defined constant. The term clearly dissipates energy.

**Remark 6.5.2.** *In order to avoid evaluating second derivatives in the surface tension contribution, one may project the residual onto the mesh and subsequently use Proposition 6.3.6.*

**Remark 6.5.3.** *Even though we present the stabilization and discontinuity capturing terms in an ad hoc fashion, we wish to emphasize that these may be derived with the aid of the multiscale framework. The natural derivation for discontinuity capturing terms can be found in [61].*

## 6.5.3 Semi-discrete formulation

The semi-discrete approximation of (6.4.20) is stated as follows:

Find  $(\mathbf{u}^h, p^h, \phi^h, v^h) \in \mathcal{W}_T^h$  such that for all  $(\mathbf{w}^h, q^h, \psi^h, \zeta^h) \in \mathcal{W}_{0,h}$ :

$$\begin{aligned} & (\mathbf{w}^h, \partial_t(\rho^h \mathbf{u}^h))_\Omega - (\nabla \mathbf{w}^h, \rho^h \mathbf{u}^h \otimes \mathbf{u}^h)_\Omega - (\nabla \cdot \mathbf{w}^h, p^h)_\Omega \\ & \quad + (\nabla \mathbf{w}^h, \boldsymbol{\tau}(\mathbf{u}^h, \phi^h))_\Omega + \frac{1}{\text{Fr}^2} (\mathbf{w}^h, \rho^h \mathbf{j})_\Omega \\ & - (\mathbf{w}^h, v^h \nabla \phi^h)_\Omega - \left( \mathbf{w}^h, \varrho(\phi^h) \left( \frac{\|\mathbf{u}^h\|_2^2}{2} - \frac{1}{\text{Fr}^2} y \right) \nabla \phi^h \right)_\Omega \\ & + \sum_K (\nabla \mathbf{w}^h, \theta_K \nabla \mathbf{u}^h)_{\Omega_K} - \sum_K (\tau_K \mathbf{w}^h \cdot \nabla v^h, \mathcal{R}_1 \phi^h)_{\Omega_K} = 0, \end{aligned} \quad (6.5.12a)$$

$$(q^h, \nabla \cdot \mathbf{u}^h)_\Omega = 0, \quad (6.5.12b)$$

$$(\psi^h, \partial_t \phi^h + \mathbf{u}^h \cdot \nabla \phi^h)_\Omega + \sum_K (\tau_K \mathbf{u}^h \cdot \nabla \psi^h, \mathcal{R}_1 \phi^h)_{\Omega_K} = 0, \quad (6.5.12c)$$

$$\begin{aligned} & \left( \zeta^h, v^h + \varrho(\phi^h) \left( \frac{\|\mathbf{u}^h\|_2^2}{2} - \frac{1}{\text{Fr}^2} y \right) \right)_\Omega \\ & - \left( \frac{1}{\text{We}} \delta(\phi^h) \frac{\nabla \phi^h}{\|\nabla \phi^h\|_{\epsilon,2}}, \nabla \zeta^h \right)_\Omega \\ & - \left( \frac{1}{\text{We}} \|\nabla \phi^h\|_{\epsilon,2} \delta'(\phi^h), \zeta^h \right)_\Omega = 0. \end{aligned} \quad (6.5.12d)$$

where  $\mathbf{u}^h(0) = \mathbf{u}_0^h$  and  $\phi^h(0) = \phi_0^h$  in  $\Omega$  and we recall  $\varrho^h(0) = \varrho(\phi^h(0))$ . The initial fields  $\mathbf{u}_0^h$  and  $\phi_0^h$  are obtained via standard  $L^2$ -projections of respectively  $\mathbf{u}_0(x)$  and  $\phi_0(x)$  onto the mesh. The density and fluid viscosity are computed as

$$\rho^h \equiv \rho(\phi^h), \quad (6.5.13a)$$

$$\mu^h \equiv \mu(\phi^h). \quad (6.5.13b)$$

The discrete counterparts of the kinetic, gravitational and surface energy are:

$$\mathcal{E}^{\text{K},h} \equiv \mathcal{E}^{\text{K}}(\mathbf{u}^h; \phi^h), \quad (6.5.14a)$$

$$\mathcal{E}^{\text{G},h} \equiv \mathcal{E}^{\text{G}}(\phi^h), \quad (6.5.14b)$$

$$\mathcal{E}^{\text{S},h} \equiv \mathcal{E}^{\text{S}}(\phi^h). \quad (6.5.14c)$$

The total energy is the superposition of the separate energies:

$$\mathcal{E}^h := \mathcal{E}^{\text{K},h} + \mathcal{E}^{\text{G},h} + \mathcal{E}^{\text{S},h}. \quad (6.5.15)$$

Similarly, the semi-discrete local energy reads

$$\mathcal{H}^h \equiv \mathcal{H}(\mathbf{U}^h). \quad (6.5.16)$$

The semi-discrete formulation (6.5.12) inherits to a large extent Theorem 6.4.3. The notable difference lies in the usage of stabilization terms.

**Theorem 6.5.4.** Let  $(\mathbf{u}^h, p^h, \phi^h, v^h)$  be a smooth solution of the weak form of incompressible Navier-Stokes equations with surface tension (6.5.12). The formulation (6.5.12) has the properties:

1. The formulation satisfies the maximum principle for the density, i.e. without loss of generality we assume that  $\rho_2 \leq \rho_1$  and then have:

$$\rho_2 \leq \rho(\phi^h) \leq \rho_1. \quad (6.5.17)$$

2. The formulation is divergence-free as a distribution:

$$\nabla \cdot \mathbf{u}^h \equiv 0. \quad (6.5.18)$$

3. The formulation satisfies the dissipation inequality:

$$\frac{d}{dt} \mathcal{E}^h = - \left( \nabla \mathbf{u}^h, \boldsymbol{\tau}(\mathbf{u}^h, \phi^h) \right)_\Omega - \sum_K \left( \nabla \mathbf{u}^h, \theta_K \nabla \mathbf{u}^h \right)_{\Omega_K} \leq 0. \quad (6.5.19)$$

The proof of Theorem 6.5.4 goes along the same lines as that of Theorem 6.4.3.

*Proof.* 1 & 2. The first two properties are directly inherited from the continuous case. Note that the weighting function choice for the second property is in general not permitted. The specific NURBS function spaces proposed by Evans et al. [68, 69] do allow this selection.

3. Selection of the weights  $\psi^h = v^h$  in (6.5.12c) and  $\zeta^h = -\partial_t \phi^h$  in (6.5.12d) gives:

$$\left( v^h, \partial_t \phi^h + \mathbf{u}^h \cdot \nabla \phi^h \right)_\Omega + \sum_K \left( \tau_K \mathbf{u}^h \cdot \nabla v^h, \mathcal{R}_I \phi^h \right)_{\Omega_K} = 0, \quad (6.5.20a)$$

$$\begin{aligned} & - \left( \partial_t \phi^h, v^h + \varrho^h \frac{\|\mathbf{u}^h\|_2^2}{2} - \frac{1}{\text{Fr}^2} \varrho^h y \right)_\Omega \\ & + \left( \frac{1}{\text{We}} \delta(\phi^h) \frac{\nabla \phi^h}{\|\nabla \phi^h\|_{\epsilon,2}}, \nabla \partial_t \phi^h \right)_\Omega \\ & + \left( \frac{1}{\text{We}} \|\nabla \phi^h\|_{\epsilon,2} \delta'(\phi^h), \partial_t \phi^h \right)_\Omega = 0. \end{aligned} \quad (6.5.20b)$$

Addition of the equations (6.5.20) results in:

$$\begin{aligned} & (v^h, \mathbf{u}^h \cdot \nabla \phi^h)_\Omega - \left( \frac{\varrho^h}{2} \|\mathbf{u}^h\|_2^2, \partial_t \phi^h \right)_\Omega + \frac{1}{\text{Fr}^2} \left( \partial_t \phi^h, \varrho^h y \right)_\Omega \\ & + \left( \frac{1}{\text{We}} \delta(\phi^h) \frac{\nabla \phi^h}{\|\nabla \phi^h\|_{\epsilon,2}}, \nabla \partial_t \phi^h \right)_\Omega + \left( \frac{1}{\text{We}} \|\nabla \phi^h\|_{\epsilon,2} \delta'(\phi^h), \partial_t \phi^h \right)_\Omega \\ & - \sum_K \left( \tau_K \mathbf{u}^h \cdot \nabla v^h, \mathcal{R}_I \phi^h \right)_{\Omega_K} = 0. \end{aligned} \quad (6.5.21)$$

By performing integration by parts we obtain:

$$\begin{aligned} & \left( \partial_t \phi^h, -\frac{\varrho^h}{2} \|\mathbf{u}^h\|_2^2 + \varrho^h \frac{1}{\text{Fr}^2} y - \frac{1}{\text{We}} \delta(\phi^h) \nabla \cdot \left( \frac{\nabla \phi^h}{\|\nabla \phi^h\|_{\epsilon,2}} \right) + \frac{1}{\text{We}} \delta'(\phi^h) \frac{\epsilon^2}{\|\nabla \phi^h\|_{\epsilon,2}} \right)_\Omega \\ & = - (v^h, \mathbf{u}^h \cdot \nabla \phi^h)_\Omega - \sum_K \left( \tau_K \mathbf{u}^h \cdot \nabla v^h, \mathcal{R}_I \phi^h \right)_{\Omega_K}. \end{aligned} \quad (6.5.22)$$

Recognize  $\frac{\delta \mathcal{H}^h}{\delta \phi^h}$  on the left-hand side to arrive at:

$$\begin{aligned} \frac{\delta \mathcal{E}^h}{\delta \phi^h} \left[ \frac{\partial \phi^h}{\partial t} \right] &:= \left( \frac{\partial \phi^h}{\partial t}, \frac{\delta \mathcal{H}^h}{\delta \phi^h} \right)_{\Omega} = - (v^h, \mathbf{u}^h \cdot \nabla \phi^h)_{\Omega} \\ &\quad - \sum_K \left( \tau_K \mathbf{u}^h \cdot \nabla v^h, \mathcal{R}_I \phi^h \right)_{\Omega_K}. \end{aligned} \quad (6.5.23)$$

Next we take  $w^h = \mathbf{u}^h$  in (6.4.20a) to get:

$$\begin{aligned} (\mathbf{u}^h, \partial_t(\rho^h \mathbf{u}^h))_{\Omega} - (\nabla \mathbf{u}^h, \rho^h \mathbf{u}^h \otimes \mathbf{u}^h)_{\Omega} - (\mathbf{u}^h, \frac{1}{2} \|\mathbf{u}^h\|_2^2 \varrho^h \nabla \phi^h)_{\Omega} \\ - (\nabla \cdot \mathbf{u}^h, p^h)_{\Omega} + (\nabla \mathbf{u}^h, \boldsymbol{\tau}(\mathbf{u}^h, \phi^h))_{\Omega} \\ - (\mathbf{u}^h, v^h \nabla \phi^h)_{\Omega} + \frac{1}{\mathbb{F}r^2} (\mathbf{u}^h, \varrho^h y \nabla \phi^h)_{\Omega} + \frac{1}{\mathbb{F}r^2} (\mathbf{u}^h, \rho^h \mathbf{j})_{\Omega} \\ + \sum_K (\nabla \mathbf{u}^h, \theta_K \nabla \mathbf{u}^h)_{\Omega_K} - \sum_K (\tau_K \mathbf{u}^h \cdot \nabla v^h, \mathcal{R}_I \phi^h)_{\Omega_K} = 0. \end{aligned} \quad (6.5.24)$$

Similar as in the continuous case, we have the identities:

$$-(\nabla \mathbf{u}^h, \rho^h \mathbf{u}^h \otimes \mathbf{u}^h)_{\Omega} - (\mathbf{u}^h, \frac{1}{2} \|\mathbf{u}^h\|_2^2 \varrho^h \nabla \phi^h)_{\Omega} = 0, \quad (6.5.25a)$$

$$-(\nabla \cdot \mathbf{u}^h, p^h)_{\Omega} = 0, \quad (6.5.25b)$$

$$\frac{1}{\mathbb{F}r^2} (\mathbf{u}^h, \varrho^h y \nabla \phi^h)_{\Omega} + \frac{1}{\mathbb{F}r^2} (\mathbf{u}^h, \rho^h \mathbf{j})_{\Omega} = 0. \quad (6.5.25c)$$

Noting that  $\frac{\delta \mathcal{H}^h}{\delta(\rho^h \mathbf{u}^h)} = (\mathbf{u}^h)^T$  and employing (6.5.25) we arrive at:

$$\begin{aligned} \frac{\delta \mathcal{E}^h}{\delta(\rho^h \mathbf{u}^h)} \left[ \frac{\partial(\rho^h \mathbf{u}^h)}{\partial t} \right] &:= \left( \frac{\partial(\rho^h \mathbf{u}^h)}{\partial t}, \frac{\delta \mathcal{H}^h}{\delta(\rho^h \mathbf{u}^h)} \right)_{\Omega} \\ &= - (\nabla \mathbf{u}^h, \boldsymbol{\tau}(\mathbf{u}^h, \phi^h))_{\Omega} + (\mathbf{u}^h, v^h \nabla \phi^h)_{\Omega} \\ &\quad - \sum_K (\nabla w^h, \theta_K \nabla \mathbf{u}^h)_{\Omega_K} + \sum_K (\tau_K \mathbf{u}^h \cdot \nabla v^h, \mathcal{R}_I \phi^h)_{\Omega_K}. \end{aligned} \quad (6.5.26)$$

The superposition of (6.5.23) and (6.5.26) yields:

$$\begin{aligned} \frac{d}{dt} \mathcal{E}^h &= \frac{\delta \mathcal{E}^h}{\delta \phi^h} \left[ \frac{\partial \phi^h}{\partial t} \right] + \frac{\delta \mathcal{E}^h}{\delta(\rho^h \mathbf{u}^h)} \left[ \frac{\partial(\rho^h \mathbf{u}^h)}{\partial t} \right] = - (\nabla \mathbf{u}^h, \boldsymbol{\tau}(\mathbf{u}^h, \phi^h))_{\Omega} \\ &\quad - \sum_K (\nabla \mathbf{u}^h, \theta_K \nabla \mathbf{u}^h)_{\Omega_K}. \end{aligned} \quad (6.5.27)$$

□

## 6.6 ENERGY-DISSIPATIVE TEMPORAL DISCRETIZATION

In this section we present the energy-stable time-integration methodology. We present a modified version of the mid-point time-discretization method. First

we introduce some required notation in Section 6.6.1 and then explain the time-discretization of the terms that differ from the standard midpoint rule in Section 6.6.2 and 6.6.3. The eventual method is presented in Section 6.6.4.

The simplest fully-discrete algorithm would be to start from the semi-discrete version of (6.5.12) and then discretize in time using the second-order mid-point time-discretization. An important observation is that this approach does not lead to a provable energy-dissipative formulation, see Appendix 6.B. We note that this is in contrast to the single-fluid case (in absence of surface tension effects).

In the following we present our strategy to arrive at a provable energy-dissipative formulation. Our approach is to mirror the semi-discrete case as closely as possible. We first focus on the terms that are directly associated with temporal derivatives of the energies and then treat the remaining terms.

### 6.6.1 Notation

Let us divide the time-interval  $\mathcal{T}$  into sub-intervals  $\mathcal{T}_n = (t_n, t_{n+1})$  (with  $n = 0, 1, \dots, N$ ) and denote the size of interval  $\mathcal{T}_n$  as time-step  $\Delta t_n = t_{n+1} - t_n$ . We use subscripts to indicate the time-level of the unknown quantities, i.e. the unknowns at time-level  $n$  are  $\mathbf{u}_n^h, p_n^h, \phi_n^h$  and  $v_n^h$ . Lastly, we denote the intermediate time-levels and associated time derivatives as:

$$\mathbf{u}_{n+1/2}^h := \frac{1}{2}(\mathbf{u}_n^h + \mathbf{u}_{n+1}^h), \quad \frac{1}{\Delta t_n} \llbracket \mathbf{u}^h \rrbracket_n := \frac{1}{\Delta t_n} (\mathbf{u}_{n+1}^h - \mathbf{u}_n^h), \quad (6.6.1a)$$

$$\phi_{n+1/2}^h := \frac{1}{2}(\phi_n^h + \phi_{n+1}^h), \quad \frac{1}{\Delta t_n} \llbracket \phi^h \rrbracket_n := \frac{1}{\Delta t_n} (\phi_{n+1}^h - \phi_n^h) \quad (6.6.1b)$$

$$\rho_{n+1/2}^h := \rho(\phi_{n+1/2}^h), \quad \frac{1}{\Delta t_n} \llbracket \rho^h \rrbracket_n := \frac{1}{\Delta t_n} (\rho_{n+1}^h - \rho_n^h), \quad (6.6.1c)$$

$$\frac{1}{\Delta t_n} \llbracket \rho^h \mathbf{u}^h \rrbracket_n := \frac{1}{\Delta t_n} (\rho_{n+1}^h \mathbf{u}_{n+1}^h - \rho_n^h \mathbf{u}_n^h), \quad (6.6.1d)$$

$$\mu_{n+1/2}^h := \mu(\phi_{n+1/2}^h), \quad (6.6.1e)$$

where  $\rho_n^h = \rho(\phi_n^h)$ , and defined  $\llbracket \mathbf{a}^h \rrbracket_n := \mathbf{a}_{n+1}^h - \mathbf{a}_n^h$  for the jump of a certain quantity  $\mathbf{a}^h$ .

### 6.6.2 Identification energy evolution terms

In order to identify the energy evolution terms we wish to have the fully discrete version of

$$\frac{d}{dt} \mathcal{E}^{K,h} = (\mathbf{w}^h, \partial_t(\rho^h \mathbf{u}^h))_\Omega + (\zeta^h, \varrho(\phi^h) \frac{1}{2} \|\mathbf{u}^h\|_2^2)_\Omega, \quad (6.6.2a)$$

$$\frac{d}{dt} \mathcal{E}^{G,h} = - \frac{1}{\mathbb{F}r^2} (\zeta^h, \varrho(\phi^h) \mathbf{y})_\Omega, \quad (6.6.2b)$$

$$\frac{d}{dt} \mathcal{E}^{S,h} = - \left( \frac{1}{\text{We}} \delta(\phi^h) \frac{\nabla \phi^h}{\|\nabla \phi^h\|_{\epsilon,2}}, \nabla \zeta^h \right)_\Omega - \left( \frac{1}{\text{We}} \|\nabla \phi^h\|_{\epsilon,2} \delta'(\phi^h), \zeta^h \right)_\Omega. \quad (6.6.2c)$$

Three issues arise: (i) the approximation of the internal energy density  $\frac{1}{2} \|\mathbf{u}^h\|_2^2$  in the additional equation (6.5.12d), (ii) the approximation of the interface density jump term  $\varrho^h$  and (iii) the approximation of the surface tension contribution.

In the following we discuss the considerations for their time-discretization.

(i) The first matter is resolved when taking a shift in the time-levels in the energy density, analogously as in Liu et al. [144], i.e. we take  $\frac{1}{2}\mathbf{u}_n^h \cdot \mathbf{u}_{n+1}^h$  in the additional equation.

(ii) Concerning the second problem, we require a stable time-discretization of  $q^h$  such that the approximation of  $q^h \partial_t \phi^h$  equals that of  $\partial_t \rho^h$ . This suggests to approximate  $q^h$  at the intermediate time level  $t_{n+1/2}$  as

$$q^h(t_{n+1/2}) \approx q_{F,n+1/2}^h := \frac{\rho(\phi_{n+1}^h) - \rho(\phi_n^h)}{\phi_{n+1}^h - \phi_n^h}, \quad (6.6.3)$$

such that

$$\frac{\llbracket \rho^h \rrbracket_n}{\Delta t_n} = q_{F,n+1/2}^h \frac{\llbracket \phi^h \rrbracket_n}{\Delta t_n}. \quad (6.6.4)$$

Unfortunately, the approximation (6.6.3) is not defined when  $\phi_{n+1}^h = \phi_n^h$  and can behave badly when  $\phi_{n+1}^h \approx \phi_n^h$ . If  $q^h$  is a polynomial function of  $\phi^h$  we may use truncated Taylor expansions around  $\phi_{n+1/2}^h$  to find:

$$q_{F,n+1/2}^h = \sum_{j=0}^M \frac{1}{2^{2j}(2j+1)!} q^{(2j)}(\phi_{n+1/2}^h) \llbracket \phi^h \rrbracket_n^{2j}, \quad (6.6.5)$$

where  $M$  chosen such that latter terms in the sum vanish and where we use the notation  $h^{(m)}(x) = d^m h / dx^m$  for the  $m$ -th derivative of a function  $h = h(x)$ . Remark that (6.6.5) is well-defined. This motivates to use a (piecewise) higher-order polynomial for  $q^h$ . We define the regularized Heaviside as

$$H_\varepsilon(\phi_n^h) := H_p(\phi_n^h / \varepsilon) \quad (6.6.6)$$

where  $H_p = H_p(\phi)$  is the piecewise polynomial regularization:

$$H_p = H_p(\phi) = \begin{cases} 0 & \phi < -1, \\ -\frac{3}{4}\phi^5 - \frac{5}{2}\phi^4 - \frac{5}{2}\phi^3 + \frac{5}{4}\phi + \frac{1}{2} & -1 \leq \phi < 0, \\ -\frac{3}{4}\phi^5 + \frac{5}{2}\phi^4 - \frac{5}{2}\phi^3 + \frac{5}{4}\phi + \frac{1}{2} & 0 \leq \phi < 1, \\ 1 & 1 \leq \phi. \end{cases} \quad (6.6.7)$$

This function is  $C^3$ -continuous at  $\phi = 0$  and  $C^3$ -continuous at  $\phi = -1, \phi = 1$ . Furthermore, we base the regularization of Dirac on the Heaviside, i.e. we have  $\delta_\varepsilon(\phi^h) = H_\varepsilon^{(1)}(\phi^h)$ .

**Remark 6.6.1.** The regularized Dirac delta  $\delta_\varepsilon(\phi^h)$  has area 1.

**Remark 6.6.2.** If  $q^h$  is non-polynomial one may use perturbed trapezoidal rules. In case of positive higher-order derivatives this leads to a stable approximation for  $q^h$ .

**Remark 6.6.3.** This regularization closely resembles the popular goniometric regularization:

$$H_g = H_g(\phi) = \begin{cases} 0 & \phi < -1, \\ \frac{1}{2} \left( 1 + \phi + \frac{1}{\pi} \sin(\pi\phi) \right) & -1 \leq \phi < 1, \\ 1 & 1 \leq \phi. \end{cases} \quad (6.6.8)$$

Figure 6.1 shows the polynomial regularization  $H_p = H_p(\phi)$ , the goniometric regularization  $H_g = H_g(\phi)$  and their first two derivatives. At  $\phi = -1$  and  $\phi = 1$  the goniometric regularization is  $C^2$ -continuous where  $H_p = H_p(\phi)$  is  $C^3$ -continuous.

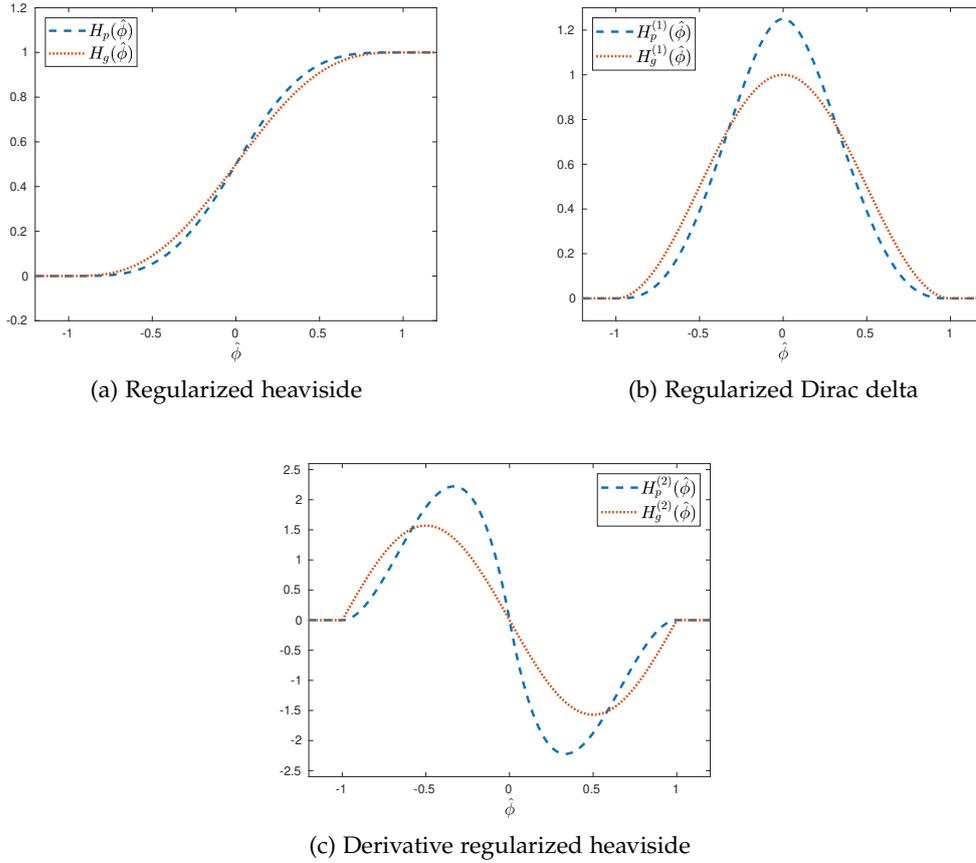


Figure 6.1: Comparison of polynomial and goniometric regularization of the Heaviside.

Since  $\varrho^h(t_{n+1/2})$  is a piecewise polynomial, (6.6.5) only holds if  $\phi_n^h$  and  $\phi_{n+1}^h$  are in the same ‘piece’. In the other case we have  $\phi_n^h \neq \phi_{n+1}^h$  and thus we may use  $\varrho_{F,n+1/2}^h$ . Thus, to define  $\varrho^h(t_{n+1/2})$  in the auxiliary equation we distinguish the cases

1.  $\phi_n^h$  and  $\phi_{n+1}^h$  are in the same ‘piece’ of the polynomial  $H_\epsilon$
2.  $\phi_n^h$  and  $\phi_{n+1}^h$  are in another ‘piece’ of the polynomial  $H_\epsilon$ .

In the first case employ the truncated series (6.6.5) whereas in the second case we directly employ the left-hand side of (6.6.5):

$$q^h(t_{n+1/2}) \approx q_{a,n+1/2}^h := \begin{cases} q_{T,n+1/2}^h & \text{in case 1} \\ q_{F,n+1/2}^h & \text{in case 2,} \end{cases} \quad (6.6.9)$$

with Taylor series representation:

$$q_{T,n+1/2}^h := \llbracket \rho \rrbracket \left( H_\varepsilon^{(1)}(\phi_{n+1/2}^h) + \frac{1}{24} H_\varepsilon^{(3)}(\phi_{n+1/2}^h) \llbracket \phi^h \rrbracket_n^2 + \frac{1}{1920} H_\varepsilon^{(5)}(\phi_{n+1/2}^h) \llbracket \phi^h \rrbracket_n^4 \right). \quad (6.6.10)$$

Definition (6.6.9) satisfies condition (6.6.4):

$$\frac{\llbracket \rho^h \rrbracket_n}{\Delta t_n} = q_{a,n+1/2}^h \frac{\llbracket \phi^h \rrbracket_n}{\Delta t_n}. \quad (6.6.11)$$

**Remark 6.6.4.** The approximation  $q_{F,n+1/2}^h$  (case 2), defined in (6.6.3), is well-behaved when  $\phi_{n+1}^h \approx \phi_n^h$ . To see this, we consider without loss of generality the case where  $\phi_n^h < -\varepsilon$  and  $-\varepsilon < \phi_{n+1}^h < 0$ . A Taylor series representation of  $H_\varepsilon$  around  $\phi_{n+1}^h = -\varepsilon$  reveals:

$$q_{F,n+1/2}^h = \llbracket \rho \rrbracket \frac{\frac{1}{24}(\phi_{n+1}^h + \varepsilon)^4 H_\varepsilon^{(4)}(\xi)}{\phi_{n+1}^h - \phi_n^h}, \quad (6.6.12)$$

for some  $\xi \in (-\varepsilon, \phi_{n+1}^h)$ . It is now convenient to express  $\phi_n^h$  as a perturbation of  $-\varepsilon$  relative to the distance between  $\phi_n^h$  and  $-\varepsilon$ . In other words, we write  $\phi_n^h = -\varepsilon - v(\phi_{n+1}^h + \varepsilon)$  for  $v = -(\phi_n^h + \varepsilon)/(\phi_{n+1}^h + \varepsilon) > 0$ . Substitution into (6.6.12) gives:

$$q_{F,n+1/2}^h = \llbracket \rho \rrbracket \frac{1}{24\varepsilon(1+v)} (\phi_{n+1}^h + \varepsilon)^3 H_\varepsilon^{(4)}(\xi). \quad (6.6.13)$$

Noting that  $|H_p^{(4)}(\phi)| < 60$  we obtain the bound:

$$\begin{aligned} |q_{F,n+1/2}^h| &\leq |\llbracket \rho \rrbracket| \frac{5}{2\varepsilon(1+v)} |\phi_{n+1}^h + \varepsilon|^3 \\ &\leq |\llbracket \rho \rrbracket| \frac{5}{2} \varepsilon^2. \end{aligned} \quad (6.6.14)$$

(iii) We now turn our focus to the surface tension contribution, which writes in semi-discrete form:

$$- \left( \frac{1}{\mathbb{W}e} \delta(\phi^h) \frac{\nabla \phi^h}{\|\nabla \phi^h\|_{\varepsilon,2}}, \nabla \zeta^h \right)_\Omega - \left( \frac{1}{\mathbb{W}e} \|\nabla \phi^h\|_{\varepsilon,2} \delta'(\phi^h), \zeta^h \right)_\Omega. \quad (6.6.15)$$

Recall that in the semi-discrete form the surface energy evolution follows when substituting  $\zeta^h = -\partial_t \phi^h$ :

$$\begin{aligned} &\left( \frac{1}{\mathbb{W}e} \|\nabla \phi^h\|_{\varepsilon,2} \delta'(\phi^h), \partial_t \phi^h \right)_\Omega + \left( \frac{1}{\mathbb{W}e} \delta(\phi^h) \frac{\nabla \phi^h}{\|\nabla \phi^h\|_{\varepsilon,2}}, \nabla \partial_t \phi^h \right)_\Omega = \\ &\frac{1}{\mathbb{W}e} \left( \partial_t \delta(\phi^h), \|\nabla \phi^h\|_{\varepsilon,2} \right)_\Omega + \frac{1}{\mathbb{W}e} \left( \delta(\phi^h), \partial_t \|\nabla \phi^h\|_{\varepsilon,2} \right)_\Omega = \\ &\frac{d}{dt} \left( \delta(\phi^h), \frac{1}{\mathbb{W}e} \|\nabla \phi^h\|_{\varepsilon,2} \right)_\Omega = \\ &\frac{d}{dt} \mathcal{E}^{S,h}. \end{aligned} \quad (6.6.16)$$

Here we have utilized following identities:

- for the first term:

$$(I) \quad \partial_t \phi^h \delta'(\phi^h) = \partial_t \delta(\phi^h), \quad (6.6.17a)$$

- for the second term:

$$(II) \quad \nabla \partial_t \phi^h \cdot \frac{\nabla \phi^h}{\|\nabla \phi^h\|_{\epsilon,2}} = \partial_t \|\nabla \phi^h\|_{\epsilon,2}, \quad (6.6.17b)$$

- and for combining the terms:

$$(III) \quad \|\nabla \phi^h\|_{\epsilon,2} \partial_t \delta(\phi^h) + \delta(\phi^h) \partial_t \|\nabla \phi^h\|_{\epsilon,2} = \partial_t \left( \delta(\phi^h) \|\nabla \phi^h\|_{\epsilon,2} \right). \quad (6.6.17c)$$

We wish to follow the same steps in the fully-discrete setting. However, these identities are not directly guaranteed in a fully discrete sense. In the following we describe the fully-discrete approximation of each of the three terms in (6.6.15), i.e.  $\delta'(\phi^h)$ ,  $\delta(\phi^h)$  and  $\nabla \phi^h / \|\nabla \phi^h\|_{\epsilon,2}$ , that complies with these identities. To that purpose we introduce the mid-point approximation of the time-derivative.

**Proposition 6.6.5.** *The mid-point approximation of the time-derivative satisfies the product-rule in the following sense:*

$$\frac{\llbracket \mathbf{a}^h \cdot \mathbf{b}^h \rrbracket_n}{\Delta t_n} = \mathbf{a}_{n+1/2}^h \cdot \frac{\llbracket \mathbf{b}^h \rrbracket_n}{\Delta t_n} + \frac{\llbracket \mathbf{a}^h \rrbracket_n}{\Delta t_n} \cdot \mathbf{b}_{n+1/2}^h, \quad (6.6.18)$$

where  $\mathbf{a}^h$  and  $\mathbf{b}^h$  are scalar or vector fields.

(III) We start off with the last identity (6.6.17c). The fully-discrete version of the product rule in (6.6.17c) follows from Proposition 6.6.5:

$$\begin{aligned} \frac{\llbracket \delta(\phi^h) \|\nabla \phi^h\|_{\epsilon,2} \rrbracket_n}{\Delta t_n} &= \frac{\llbracket \delta(\phi^h) \rrbracket_n}{\Delta t_n} \left( \|\nabla \phi^h\|_{\epsilon,2} \right)_{n+1/2} \\ &\quad + \left( \delta(\phi^h) \right)_{n+1/2} \frac{\llbracket \|\nabla \phi^h\|_{\epsilon,2} \rrbracket_n}{\Delta t_n}. \end{aligned} \quad (6.6.19)$$

This implies that we require the approximation:

$$\delta(\phi^h)(t_{n+1/2}) \approx (\delta(\phi^h))_{n+1/2}, \quad (6.6.20a)$$

$$\|\nabla \phi^h\|_{\epsilon,2}(t_{n+1/2}) \approx \left( \|\nabla \phi^h\|_{\epsilon,2} \right)_{n+1/2}. \quad (6.6.20b)$$

We now aim to identify the first and the second term on the right-hand side of (6.6.19) with first and second term on the right-hand side of (6.6.16) respectively.

(I) To identify the first term we require, in a similar fashion as for  $\varrho$ , the approximation  $\zeta_{n+1/2}^h \approx \delta'(\phi^h)(t_{n+1/2})$  to satisfy:

$$\frac{\llbracket \delta(\phi^h) \rrbracket_n}{\Delta t_n} = \zeta_{n+1/2}^h \frac{\llbracket \phi^h \rrbracket_n}{\Delta t_n}. \quad (6.6.21)$$

To this purpose we define

$$\zeta_{n+1/2}^h := \begin{cases} \zeta_{T,n+1/2}^h & \text{in case 1} \\ \zeta_{F,n+1/2}^h & \text{in case 2,} \end{cases} \quad (6.6.22)$$

with truncated series:

$$\zeta_{T,n+1/2}^h := \delta_\varepsilon^{(1)}(\phi_{n+1/2}^h) + \frac{[\![\phi^h]\!]_n^2}{24} \delta_\varepsilon^{(3)}(\phi_{n+1/2}^h), \quad (6.6.23)$$

and the fraction:

$$\zeta_{F,n+1/2}^h := \frac{\delta_\varepsilon(\phi_{n+1}^h) - \delta_\varepsilon(\phi_n^h)}{\phi_{n+1}^h - \phi_n^h}. \quad (6.6.24)$$

(II) We take the approximation:

$$\left( \frac{\nabla \phi^h}{\|\nabla \phi^h\|_{\varepsilon,2}} \right) (t_{n+1/2}) \approx \frac{(\nabla \phi^h)_{n+1/2}}{(\|\nabla \phi^h\|_{\varepsilon,2})_{n+1/2}} = \frac{\nabla \phi_{n+1}^h + \nabla \phi_n^h}{\|\nabla \phi_{n+1}^h\|_{\varepsilon,2} + \|\nabla \phi_n^h\|_{\varepsilon,2}}, \quad (6.6.25)$$

such that (6.6.17b) is satisfied in a fully-discrete sense:

$$\nabla \frac{[\![\phi^h]\!]_n}{\Delta t_n} \cdot \frac{(\nabla \phi^h)_{n+1/2}}{(\|\nabla \phi^h\|_{\varepsilon,2})_{n+1/2}} = \frac{\|\nabla \phi_{n+1}^h\|_{\varepsilon,2} - \|\nabla \phi_n^h\|_{\varepsilon,2}}{\Delta t}. \quad (6.6.26)$$

### 6.6.3 Discretization other terms

We discretize the continuity equation using the mid-point rule, i.e.

$$\left( q^h, \nabla \cdot \mathbf{u}_{n+1/2}^h \right)_\Omega = 0, \quad (6.6.27)$$

which implies pointwise divergence-free solutions on a fully-discrete level.

Next, we require the fully-discrete version of the identities:

$$-(\nabla \mathbf{u}^h, \rho^h \mathbf{u}^h \otimes \mathbf{u}^h)_\Omega - (\mathbf{u}^h, \frac{1}{2} \|\mathbf{u}^h\|_2^2 \varrho(\phi^h) \nabla \phi^h)_\Omega = 0, \quad (6.6.28a)$$

$$+ \frac{1}{\mathbb{F}r^2} (\mathbf{u}^h, \rho^h \mathbf{j})_\Omega + \frac{1}{\mathbb{F}r^2} (\mathbf{u}^h, y \varrho(\phi^h) \nabla \phi^h)_\Omega = 0, \quad (6.6.28b)$$

which make use of the pointwise divergence-free property. These identities are fulfilled when we have

$$\nabla \rho(\phi^h) = \varrho(\phi^h) \nabla \phi^h. \quad (6.6.29)$$

Applying the chain-rule implies that we can take as approximation in the momentum equation:

$$q^h(t_{n+1/2}) \approx q_{m,n+1/2}^h := [\![\rho]\!] H'_\varepsilon(\phi_{n+1/2}^h), \quad (6.6.30)$$

where the subscript  $m$  refers to the momentum equation.

**Remark 6.6.6.** Note that we employ two different approximations for  $q^h(t_{n+1/2})$ , namely (6.6.9) in the additional equation (6.5.12d) and (6.6.30) in the momentum equation (6.5.12a).

The remaining terms utilize the standard midpoint discretization.

## 6.6.4 Fully-discrete energy-dissipative method

We are now ready to present the fully-discrete energy-dissipative method:

Given  $\mathbf{u}_n^h, p_n^h, \phi_n^h$  and  $v_n^h$ , find  $\mathbf{u}_{n+1}^h, p_{n+1}^h, \phi_{n+1}^h$  and  $v_{n+1}^h$  such that for all  $(\mathbf{w}^h, q^h, \psi^h, \zeta^h) \in \mathcal{W}_{0,h}$ :

$$\begin{aligned} & (\mathbf{w}^h, \frac{[[\rho^h \mathbf{u}^h]]_n}{\Delta t_n})_\Omega - (\nabla \mathbf{w}^h, \rho_{n+1/2}^h \mathbf{u}_{n+1/2}^h \otimes \mathbf{u}_{n+1/2}^h)_\Omega \\ & - (\nabla \cdot \mathbf{w}^h, p_{n+1}^h)_\Omega + (\nabla \mathbf{w}^h, \boldsymbol{\tau}(\mathbf{u}_{n+1/2}^h, \phi_{n+1/2}^h))_\Omega \\ & + \frac{1}{\mathbb{F}r^2} (\mathbf{w}^h, \rho_{n+1/2}^h \mathbf{J})_\Omega - (\mathbf{w}^h, v_{n+1}^h \nabla \phi_{n+1/2}^h)_\Omega \\ & - \left( \mathbf{w}^h, \varrho_{m,n+1/2}^h \left( \frac{\|\mathbf{u}_{n+1/2}^h\|_2^2}{2} - \frac{1}{\mathbb{F}r^2} y \right) \nabla \phi_{n+1/2}^h \right)_\Omega \\ & - \sum_K (\boldsymbol{\tau}_K \mathbf{w}^h \cdot \nabla v_{n+1}^h, \mathcal{R}_1 \phi_{n+1/2}^h)_{\Omega_K} = 0, \end{aligned} \quad (6.6.31a)$$

$$(\zeta^h, \nabla \cdot \mathbf{u}_{n+1/2}^h)_\Omega = 0, \quad (6.6.31b)$$

$$\begin{aligned} & (\psi^h, \frac{[[\phi^h]]_n}{\Delta t_n} + \mathbf{u}_{n+1/2}^h \cdot \nabla \phi_{n+1/2}^h)_\Omega \\ & + \sum_K (\boldsymbol{\tau}_K \mathbf{u}_{n+1/2}^h \cdot \nabla \psi^h, \mathcal{R}_1 \phi_{n+1/2}^h)_{\Omega_K} = 0, \end{aligned} \quad (6.6.31c)$$

$$\begin{aligned} & \left( \zeta^h, v_{n+1}^h + \varrho_{a,n+1/2}^h \left( \frac{1}{2} \mathbf{u}_{n+1}^h \cdot \mathbf{u}_n^h - \frac{1}{\mathbb{F}r^2} y \right) \right)_\Omega \\ & - \frac{1}{\mathbb{W}e} \left( \zeta^h \varrho_{n+1/2}^h, (\|\nabla \phi^h\|_{\epsilon,2})_{n+1/2} \right)_\Omega \\ & - \frac{1}{\mathbb{W}e} \left( \delta(\phi^h)_{n+1/2} \nabla \zeta^h, \frac{(\nabla \phi^h)_{n+1/2}}{(\|\nabla \phi^h\|_{\epsilon,2})_{n+1/2}} \right)_\Omega = 0. \end{aligned} \quad (6.6.31d)$$

**Remark 6.6.7.** Due to Proposition 6.6.5 the time-derivative in the momentum equation may be implemented as:

$$\frac{[[\rho^h \mathbf{u}^h]]_n}{\Delta t_n} = \rho_{n+1/2}^h \frac{[[\mathbf{u}^h]]_n}{\Delta t_n} + \frac{[[\rho^h]]_n}{\Delta t_n} \mathbf{u}_{n+1/2}^h. \quad (6.6.32)$$

**Theorem 6.6.8.** The algorithm (6.6.31) has the properties:

1. The scheme satisfies the maximum principle for the density, i.e. without loss of generality we assume that  $\rho_2 \leq \rho_1$  and then have:

$$\rho_2 \leq \rho_n^h \leq \rho_1, \quad \text{for all } n = 0, 1, \dots, N. \quad (6.6.33)$$

2. The scheme is divergence-free as a distribution:

$$\nabla \cdot \mathbf{u}_{n+1/2}^h \equiv 0. \quad (6.6.34)$$

3. The scheme satisfies the dissipation inequality:

$$\begin{aligned} \frac{[[\mathcal{E}^h]]_n}{\Delta t_n} &= - \left( \nabla \mathbf{u}_{n+1/2}^h, \boldsymbol{\tau}(\mathbf{u}_{n+1/2}^h, \phi_{n+1/2}^h) \right)_\Omega \\ &\quad - \sum_K \left( \nabla \mathbf{u}_{n+1/2}^h, \theta_K \nabla \mathbf{u}_{n+1/2}^h \right)_{\Omega_K} \\ &\leq 0, \quad \text{for all } n = 0, 1, \dots, N. \end{aligned} \quad (6.6.35)$$

*Proof.* 1 & 2. Analogously to the semi-discrete case.

3. Selection of the weights  $\psi^h = v_{n+1}^h$  in (6.6.31c) and  $\zeta^h = -[[\phi^h]]_n / \Delta t_n$  in (6.6.31d) yields:

$$\begin{aligned} & \left( v_{n+1}^h, \frac{[[\phi^h]]_n}{\Delta t_n} + \mathbf{u}_{n+1/2}^h \cdot \nabla \phi_{n+1/2}^h \right)_\Omega \\ & + \sum_K \left( \tau_K \mathbf{u}_{n+1/2}^h \cdot \nabla v_{n+1}^h, \mathcal{R}_1 \phi_{n+1/2}^h \right)_{\Omega_K} = 0, \end{aligned} \quad (6.6.36a)$$

$$\begin{aligned} & - \left( \frac{[[\phi^h]]_n}{\Delta t_n}, v^h + \varrho_{a,n+1/2}^h \left( \frac{1}{2} \mathbf{u}_{n+1}^h \cdot \mathbf{u}_n^h - \frac{1}{\mathbb{F}r^2} \mathbf{y} \right) \right)_\Omega \\ & + \frac{1}{\text{We}} \left( \frac{[[\phi^h]]_n}{\Delta t_n} \zeta_{n+1/2}^h, \left( \|\nabla \phi^h\|_{\epsilon,2} \right)_{n+1/2} \right)_\Omega \\ & + \frac{1}{\text{We}} \left( \delta(\phi_{n+1/2}^h) \nabla \frac{[[\phi^h]]_n}{\Delta t_n}, \frac{(\nabla \phi^h)_{n+1/2}}{\left( \|\nabla \phi^h\|_{\epsilon,2} \right)_{n+1/2}} \right)_\Omega = 0. \end{aligned} \quad (6.6.36b)$$

We add the equations (6.6.36) and find:

$$\begin{aligned} & \left( v_{n+1}^h, \mathbf{u}_{n+1/2}^h \cdot \nabla \phi_{n+1/2}^h \right)_\Omega - \left( \frac{[[\phi^h]]_n}{\Delta t_n}, \frac{1}{2} \varrho_{a,n+1/2}^h \mathbf{u}_{n+1}^h \cdot \mathbf{u}_n^h \right)_\Omega \\ & + \left( \frac{[[\phi^h]]_n}{\Delta t_n}, \varrho_{a,n+1/2}^h \frac{1}{\mathbb{F}r^2} \mathbf{y} \right)_\Omega \\ & + \sum_K \left( \tau_K \mathbf{u}_{n+1/2}^h \cdot \nabla v_{n+1}^h, \mathcal{R}_1 \phi_{n+1/2}^h \right)_{\Omega_K} \\ & + \left( \frac{[[\phi^h]]_n}{\Delta t_n} \zeta_{n+1/2}^h, \frac{1}{\text{We}} \left( \|\nabla \phi^h\|_{\epsilon,2} \right)_{n+1/2} \right)_\Omega \\ & + \left( \delta(\phi_{n+1/2}^h) \nabla \frac{[[\phi^h]]_n}{\Delta t_n}, \frac{1}{\text{We}} \frac{(\nabla \phi^h)_{n+1/2}}{\left( \|\nabla \phi^h\|_{\epsilon,2} \right)_{n+1/2}} \right)_\Omega = 0. \end{aligned} \quad (6.6.37)$$

Using (6.6.11), (6.6.19), (6.6.21) and (6.6.26) we get

$$\begin{aligned} & \left( \frac{[[\rho^h]]_n}{\Delta t_n}, -\frac{1}{2} \mathbf{u}_{n+1}^h \cdot \mathbf{u}_n^h + \frac{1}{\mathbb{F}r^2} \mathbf{y} \right)_\Omega + \left( \frac{1}{\text{We}}, \frac{[[\delta(\phi^h) \|\nabla \phi^h\|_{\epsilon,2}]]_n}{\Delta t_n} \right)_\Omega \\ & = - \left( v_{n+1}^h, \mathbf{u}_{n+1/2}^h \cdot \nabla \phi_{n+1/2}^h \right)_\Omega \\ & = - \sum_K \left( \tau_K \mathbf{u}_{n+1/2}^h \cdot \nabla v_{n+1}^h, \mathcal{R}_1 \phi_{n+1/2}^h \right)_{\Omega_K}. \end{aligned} \quad (6.6.38)$$

Next we take  $\mathbf{w}^h = \mathbf{u}_{n+1/2}^h$  in (6.6.31a) to get:

$$\begin{aligned}
(\mathbf{u}_{n+1/2}^h, \frac{\llbracket \rho^h \mathbf{u}^h \rrbracket_n}{\Delta t_n})_{\Omega} &= (\nabla \mathbf{u}_{n+1/2}^h, \rho_{n+1/2}^h \mathbf{u}_{n+1/2}^h \otimes \mathbf{u}_{n+1/2}^h)_{\Omega} \\
&+ (\mathbf{u}_{n+1/2}^h, \frac{1}{2} \|\mathbf{u}_{n+1/2}^h\|_2^2 \varrho_{m,n+1/2}^h \nabla \phi_{n+1/2}^h)_{\Omega} \\
&- \frac{1}{\mathbb{F}r^2} (\mathbf{u}_{n+1/2}^h, \rho_{n+1/2}^h \mathbf{j})_{\Omega} \\
&- \frac{1}{\mathbb{F}r^2} (\mathbf{u}_{n+1/2}^h, \varrho_{m,n+1/2}^h \mathbf{y} \nabla \phi_{n+1/2}^h)_{\Omega} \\
&- (\nabla \cdot \mathbf{u}_{n+1/2}^h, p_{n+1}^h)_{\Omega} - (\nabla \mathbf{u}_{n+1/2}^h, \boldsymbol{\tau}(\mathbf{u}_{n+1/2}^h, \phi_{n+1/2}^h))_{\Omega} \\
&+ (\mathbf{u}_{n+1/2}^h, v_{n+1}^h \nabla \phi_{n+1/2}^h)_{\Omega} \\
&- \sum_K (\nabla \mathbf{u}_{n+1/2}^h, \theta_K \nabla \mathbf{u}_{n+1/2}^h)_{\Omega_K} \\
&+ \sum_K (\boldsymbol{\tau}_K \mathbf{u}_{n+1/2}^h \cdot \nabla v_{n+1}^h, \mathcal{R}_I \phi_{n+1/2}^h)_{\Omega_K}. \tag{6.6.39}
\end{aligned}$$

By virtue of (6.6.28) and (6.6.34) we have the identities:

$$\begin{aligned}
&(\nabla \mathbf{u}_{n+1/2}^h, \rho_{n+1/2}^h \mathbf{u}_{n+1/2}^h \otimes \mathbf{u}_{n+1/2}^h)_{\Omega} \\
&+ (\mathbf{u}_{n+1/2}^h, \frac{1}{2} \|\mathbf{u}_{n+1/2}^h\|_2^2 \varrho_{m,n+1/2}^h \nabla \phi_{n+1/2}^h)_{\Omega} = 0, \tag{6.6.40a}
\end{aligned}$$

$$-(\nabla \cdot \mathbf{u}_{n+1/2}^h, p_{n+1}^h)_{\Omega} = 0, \tag{6.6.40b}$$

$$\frac{1}{\mathbb{F}r^2} (\mathbf{u}_{n+1/2}^h, \rho_{n+1/2}^h \mathbf{j})_{\Omega} + \frac{1}{\mathbb{F}r^2} (\mathbf{u}_{n+1/2}^h, \varrho_{m,n+1/2}^h \mathbf{y} \nabla \phi_{n+1/2}^h)_{\Omega} = 0. \tag{6.6.40c}$$

These reduce (6.6.39) to

$$\begin{aligned}
(\mathbf{u}_{n+1/2}^h, \frac{\llbracket \rho^h \mathbf{u}^h \rrbracket_n}{\Delta t_n})_{\Omega} &= -(\nabla \mathbf{u}_{n+1/2}^h, \boldsymbol{\tau}(\mathbf{u}_{n+1/2}^h, \phi_{n+1/2}^h))_{\Omega} \\
&- \sum_K (\nabla \mathbf{u}_{n+1/2}^h, \theta_K \nabla \mathbf{u}_{n+1/2}^h)_{\Omega_K} \\
&+ (\mathbf{u}_{n+1/2}^h, v_{n+1}^h \nabla \phi_{n+1/2}^h)_{\Omega} \\
&+ \sum_K (\boldsymbol{\tau}_K \mathbf{u}_{n+1/2}^h \cdot \nabla v_{n+1}^h, \mathcal{R}_I \phi_{n+1/2}^h)_{\Omega_K}. \tag{6.6.41}
\end{aligned}$$

Addition of (6.6.38) and (6.6.41) by using (6.6.32) gives:

$$\begin{aligned}
&(\mathbf{u}_{n+1/2}^h, \rho_{n+1/2}^h \frac{\llbracket \mathbf{u}^h \rrbracket_n}{\Delta t_n})_{\Omega} + \left( \frac{\llbracket \rho^h \rrbracket_n}{\Delta t_n}, \mathbf{u}_{n+1/2}^h \cdot \mathbf{u}_{n+1/2}^h - \frac{1}{2} \mathbf{u}_{n+1}^h \cdot \mathbf{u}_n^h \right)_{\Omega} \\
&+ \frac{1}{\mathbb{F}r^2} \left( \frac{\llbracket \rho^h \rrbracket_n}{\Delta t_n}, \mathbf{y} \right)_{\Omega} + \frac{1}{\text{We}} \left( 1, \frac{\llbracket \delta(\phi^h) \rrbracket \nabla \phi^h \rrbracket_{\epsilon,2}}{\Delta t_n} \right)_{\Omega} \\
&= -(\nabla \mathbf{u}_{n+1/2}^h, \boldsymbol{\tau}(\mathbf{u}_{n+1/2}^h, \phi_{n+1/2}^h))_{\Omega} \\
&- \sum_K (\nabla \mathbf{u}_{n+1/2}^h, \theta_K \nabla \mathbf{u}_{n+1/2}^h)_{\Omega_K}. \tag{6.6.42}
\end{aligned}$$

Using the identity

$$\|\mathbf{u}_{n+1/2}^h\|^2 - \frac{1}{2} \mathbf{u}_{n+1}^h \cdot \mathbf{u}_n^h = \frac{1}{2} (\|\mathbf{u}^h\|^2)_{n+1/2} \equiv \frac{1}{2} \|\mathbf{u}_{n+1}^h\|^2 + \frac{1}{2} \|\mathbf{u}_n^h\|^2, \tag{6.6.43}$$

we identify the sum of the first two terms on the left-hand side of (6.6.42) as the change of kinetic energy. Next, the third term on the left-hand side of (6.6.42) represents change in gravitational energy. The latter term on the left-hand side of (6.6.42) resembles the surface energy evolution. We are left with:

$$\begin{aligned} \frac{[[\mathcal{E}^h]]_n}{\Delta t_n} &= - \left( \nabla \mathbf{u}_{n+1/2}^h, \boldsymbol{\tau}(\mathbf{u}_{n+1/2}^h, \phi_{n+1/2}^h) \right)_{\Omega} \\ &\quad - \sum_K \left( \nabla \mathbf{u}_{n+1/2}^h, \theta_K \nabla \mathbf{u}_{n+1/2}^h \right)_{\Omega_K}. \end{aligned} \quad (6.6.44)$$

□

**Remark 6.6.9.** Following Brackbill [31] we employ the time-step restriction  $\Delta t_n \leq \Delta t_{\max}$  with

$$\Delta t_{\max} = \left( \frac{\bar{\rho} (\min_Q h_Q)^3 \text{We}}{2\pi} \right)^{1/2}, \quad (6.6.45)$$

where  $\bar{\rho} = (\rho_1 + \rho_2)/2$ .

## 6.7 NUMERICAL EXPERIMENTS

In this section we evaluate the proposed numerical methodology on several numerical examples in two and three dimensions. To test the formulation we use both a static and dynamic equilibrium problem and check the energy-dissipative property of the method. We do not test the method on a ‘violent’ problem in order to avoid the usage of redistancing procedures. All problems are evaluated with NURBS basis functions that are mostly  $C^1$ -quadratic but every velocity space is enriched to cubic  $C^2$  in the associated direction [68, 69].

### 6.7.1 Static spherical droplet

Here we test the surface tension component of the formulation by considering a spherical droplet in equilibrium [31, 75, 201, 203]. Viscous and gravitational forces are absent and hence the surface tension forces are in balance with the pressure difference between the two fluids. The interface balance (6.2.1d) thus reduces to:

$$[[[p]]] = -\sigma\kappa, \quad (6.7.1)$$

which is also referred to as the Young-Laplace equation. The exact curvature is given by:

$$\kappa = -\frac{d-1}{r}, \quad (6.7.2)$$

where we recall  $d = 2, 3$  as the number of spatial dimensions. The spherical droplet of radius  $r = 2$  of fluid 1 with density  $\rho_1 = 1.0$  is immersed in fluid 2 with density  $\rho_2 = 0.1$ . The surface tension coefficient is  $\sigma = 73$ . This implies that surface tension forces dominate since  $\text{We} \approx 1.3 \times 10^{-6}$ , where we have taken  $U_0 = 0.1$  and  $L_0 = 1.0$ . The computational domain is a cubic with a side length of 8 units and the spherical

droplet is positioned in the center of it. On all surfaces a non-penetration boundary condition ( $u_n = 0$ ) is imposed.

We employ three meshes with uniform elements:  $20 \times 20$ ,  $40 \times 40$  and  $80 \times 80$ . We take  $\varepsilon = 2h_K$  for all simulations in this section. The time-step is taken as  $\Delta t_n = 10^{-3}$  which satisfies (6.6.45) for each of the meshes. We exclude the discontinuity capturing mechanisms for this problem, i.e. we set  $\mathcal{C} = 0$ . In Figure 6.2 we display the pressure for the finest mesh.

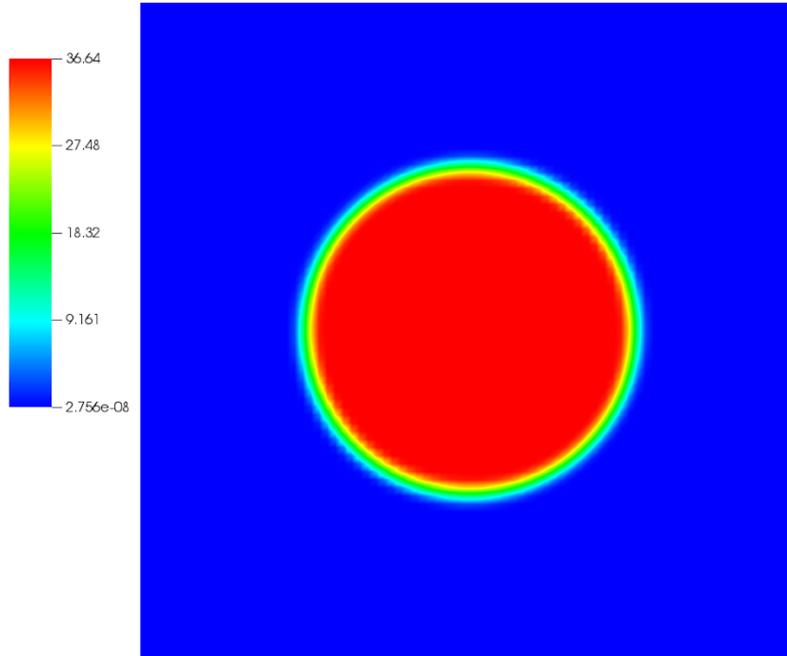


Figure 6.2: Pressure

In Figure 6.3 we display the pressure contours for each of the meshes. The corresponding pressure jump is 37.97, 36.80 and 36.56 for the meshes  $20 \times 20$ ,  $40 \times 40$  and  $80 \times 80$  respectively. This implies second-order convergence.

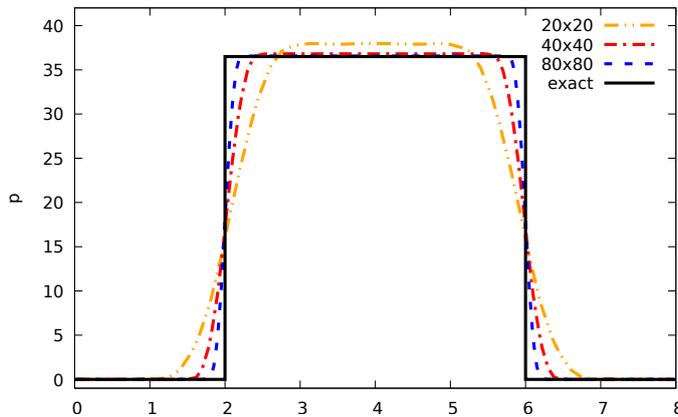


Figure 6.3: Pressure slice at  $y = 4.0$ .

In the Figures 6.4 and 6.5 we depict the energy evolution and dissipation for each of the meshes. The theoretical value of the surface energy is  $2\pi r\sigma \approx 917.34$  which is well represented on the finest mesh. We see that the total and surface energies are (virtually) constant and the kinetic energy grows but has an insignificant contribution to the total energy.

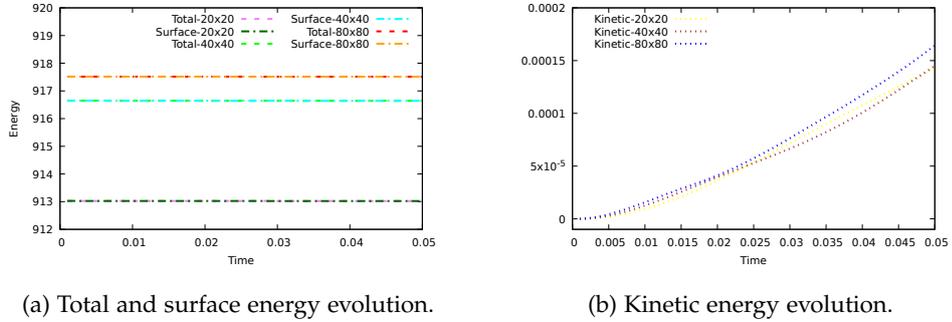


Figure 6.4: Static droplet. Energy evolution for the various meshes.

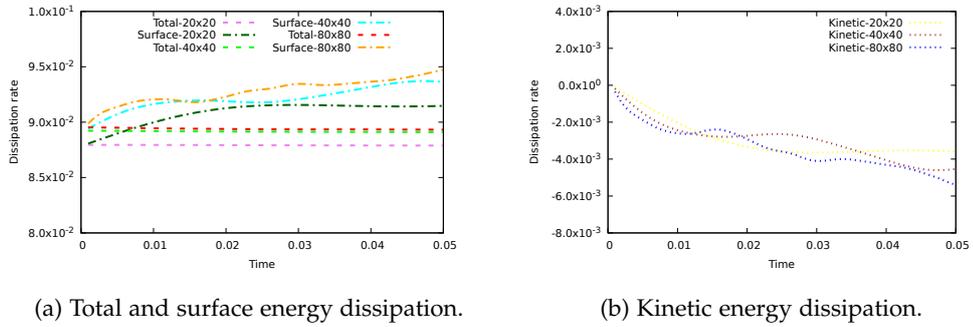


Figure 6.5: Static droplet. Energy dissipation for the various meshes.

Note that this test-case represents a physically stable situation and as such it is desirable that velocities and thus the kinetic energy vanish. Since the system is not in a total energy-stable state we note the occurrence of small parasitic currents. Remark that this is not in conflict with the energy-dissipative property of the numerical discretization; even though kinetic energy increases, the total energy dissipates. We use the typical way to report the magnitude of the parasitic currents (Figure 6.6), i.e. after one time-step and after 50 time-steps. Even though the parasitic currents are very small, they are unfortunately present. This is a well-known problem. One can use several nonphysical ‘tricks’ to reduce parasitic currents. A possibility is to use a so-called balanced-force algorithm [1] which assumes that the curvature is determined analytically.

**Remark 6.7.1.** *The occurrence of parasitic currents is a property of the regularized model, not of the discretization. To see this, note that spurious oscillations would only be absent in a stationary state of the numerical model. A stationary state is the state with the minimum energy, and as such it is a minimizer of*

$$\mathcal{H} = \frac{1}{2}\rho\|\mathbf{u}\|_2^2 + \frac{1}{\text{Fr}^2}\rho y + \frac{1}{\text{We}}\delta_\Gamma, \quad (6.7.3)$$

in which we recall that  $\delta_\Gamma(\phi)$  is regularized:  $\delta_\Gamma(\phi) = \delta_\epsilon(\phi) \|\nabla\phi\|_{\epsilon,2}$ . The minimizers are characterized by

$$\frac{\delta\mathcal{H}}{\delta\phi}[\delta v_1] = -\frac{1}{2}\|\mathbf{u}\|_2^2\rho'(\phi)\delta v_1 + \frac{1}{\mathbb{F}r^2}\rho'(\phi)y\delta v_1 \quad (6.7.4a)$$

$$+ \frac{1}{\mathbb{W}e}\delta_\epsilon(\phi)\frac{\nabla\phi}{\|\nabla\phi\|_{\epsilon,2}} \cdot \nabla\delta v_1 + \frac{1}{\mathbb{W}e}\|\nabla\phi\|_{\epsilon,2}\delta'_\epsilon(\phi)\delta v_1 = 0, \quad (6.7.4b)$$

$$\frac{\delta\mathcal{H}}{\delta(\rho u_1)}[\delta v_2] = u_1\delta v_2 = 0, \quad (6.7.4c)$$

$$\frac{\delta\mathcal{H}}{\delta(\rho u_2)}[\delta v_3] = u_2\delta v_3 = 0, \quad (6.7.4d)$$

$$\frac{\delta\mathcal{H}}{\delta(\rho u_3)}[\delta v_4] = u_3\delta v_4 = 0. \quad (6.7.4e)$$

In absence of gravity we arrive at:

$$-\frac{1}{\mathbb{W}e}\delta_\epsilon(\phi)\nabla \cdot \left( \frac{\nabla\phi}{\|\nabla\phi\|_{\epsilon,2}} \right) + \frac{1}{\mathbb{W}e}\delta'_\epsilon(\phi)\frac{\epsilon^2}{\|\nabla\phi\|_{\epsilon,2}} = 0. \quad (6.7.5)$$

Assuming that near the interface  $\phi$  is not constant, we can discard the second term by setting  $\epsilon = 0$ , which yields the requirement:

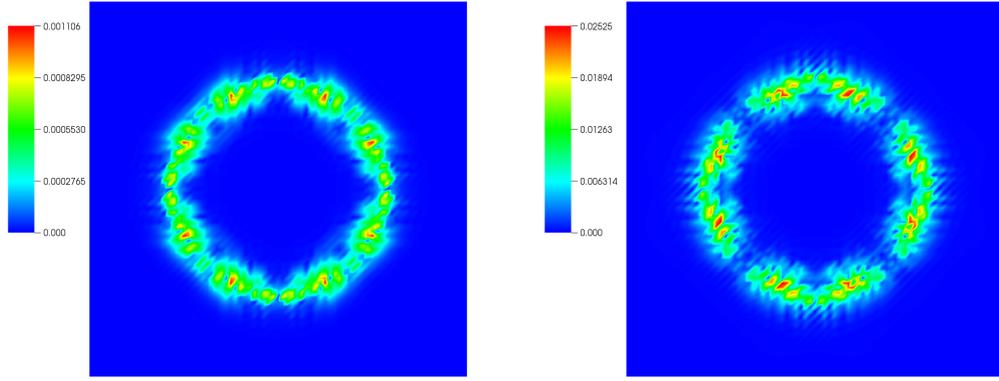
$$-\frac{1}{\mathbb{W}e}\delta_\epsilon(\phi)\kappa(\phi) = 0. \quad (6.7.6)$$

This only holds when

- $\kappa(\phi) = 0$ : a straight interface.
- $\delta_\epsilon(\phi) = 0$ : away from the regularized interface.

In other words, parasitic currents occur near a curved interface. This explains why spurious currents, even though possibly very small, will always be present in this model (except for straight interfaces).

**Remark 6.7.2.** We note that additional dissipation mechanisms for the surface evolution can upset energy-stability of the system. Well-balanced dissipation, introduced for the Navier-Stokes-Korteweg equations [79], is a possible strategy to resolve this.



(a) At time-step 1.

(b) At time-step 50.

Figure 6.6: Magnitude of the velocity, measured in the standard 2-norm.

In Figure 6.7 we plot the variable  $v_{n+1}^h$ . Note that the maximum theoretical value is

$$\begin{aligned} \max_{\mathbf{x} \in \Omega} v &= -\sigma \min_{\mathbf{x} \in \Omega} \left( \delta_\varepsilon(\phi) \nabla \cdot \left( \frac{\nabla \phi}{\|\nabla \phi\|_{\varepsilon,2}} \right) \right) \\ &\approx \frac{\sigma}{2} \max_{\mathbf{x} \in \Omega} \delta_\varepsilon(\phi) \\ &\approx 161.3, \end{aligned} \tag{6.7.7}$$

where the  $\max_{\mathbf{x} \in \Omega} \delta_\varepsilon(\phi) = \max_{\mathbf{x} \in \Omega} \frac{1}{\varepsilon} (H_p)^{(1)}(\frac{\phi}{\varepsilon}) = \frac{1}{2h_K} \max_{\mathbf{x} \in \Omega} H_p^{(1)}(\frac{\phi}{\varepsilon}) = \frac{1}{2 * \frac{8}{80} * \sqrt{2}} \frac{5}{4}$ . We see that the finest mesh is able to accurately represent  $v_{n+1}^h$  whereas on the coarser meshes  $v_{n+1}^h$  is smeared out significantly.

### 6.7.2 Droplet coalescence 2D

In this example, inspired by Gomez et al. [81], we simulate the merging of two droplets into a single one. Gravitational forces are absent. Due to pressure and capillarity forces the single droplet then develops to a circular shape. We take as computational domain the unit box  $\Omega = [0, 1]^d$  and apply no-penetration boundary conditions. The initial configuration consists of two droplet at rest ( $\mathbf{u}_0 = \mathbf{0}$ ) with centers at  $\mathbf{c}_1 = (0.4, 0.5)$  and  $\mathbf{c}_2 = (0.78, 0.5)$  and radii  $r_1 = 0.25$  and  $r_2 = 0.1$  respectively. The diffuse interfaces of the droplets initially overlap on a small part of the domain. If this not the case the droplets remain at their position and thus no merging would occur. In contrast with the Navier-Stokes Korteweg equations, in this situation the interface has a finite width, due to the definition of  $H_\varepsilon(\phi)$ . The Navier-Stokes Korteweg equations have no absolute notion of interface width; its effect is decaying exponentially. The droplets have a larger density ( $\rho_1 = 100$ ) than the surrounding fluid ( $\rho_2 = 1$ ) while the viscosities are equal:  $\mu_1 = \mu_2 = 1$ . We take as surface tension the low value of  $\sigma = 0.1$  which causes a slowly merging process. To initialize the level-set we split the domain into two parts ( $x \leq 0.665$  and

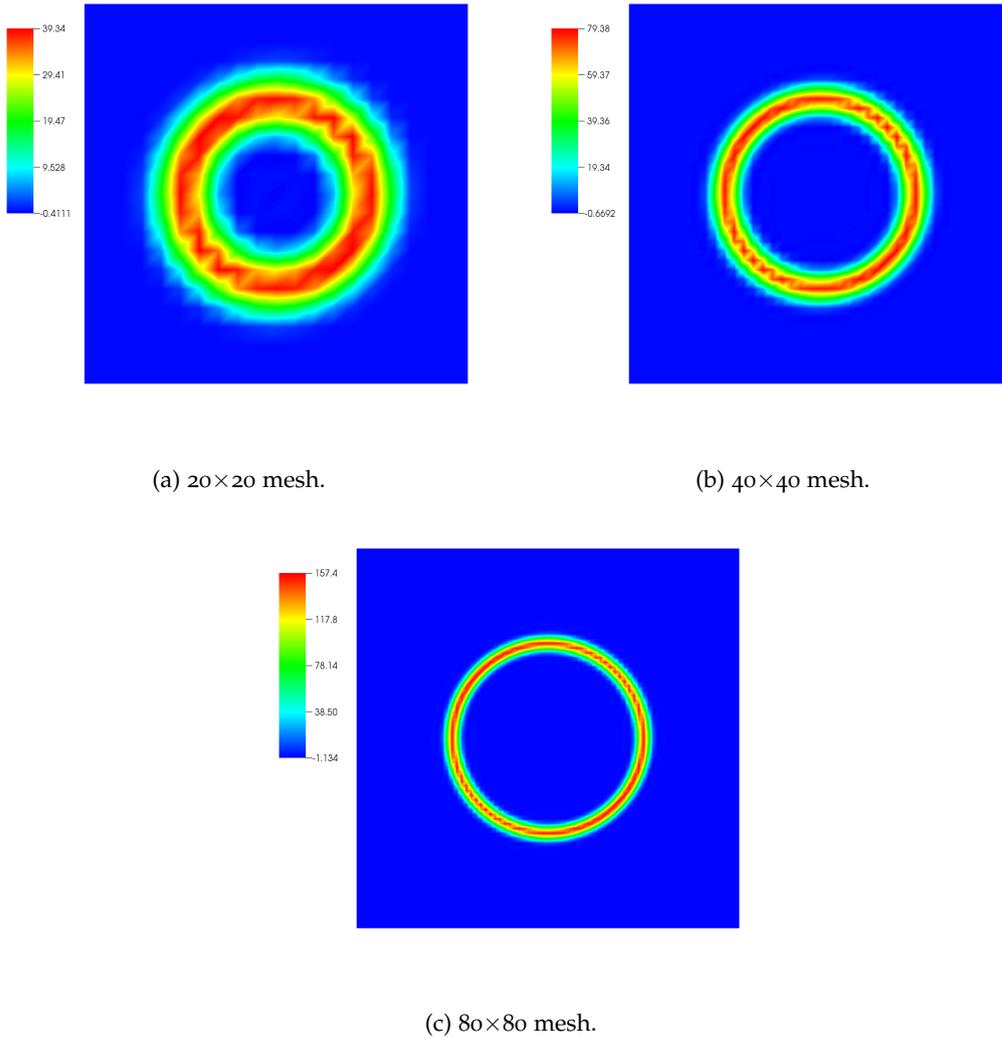


Figure 6.7: The auxiliary variable  $v$  for the various meshes.

$x > 0.665$ ), such that each contains one droplet, and apply the standard distance initialization to each subdomain. We use  $50 \times 50$  elements, set the time-step as  $\Delta t = 0.1$  and take  $C = 0.4$ .

We show in the Figures 6.8-6.13 a detailed view of the merging process. The colors patterns are set per snapshot such that difference are most apparent.

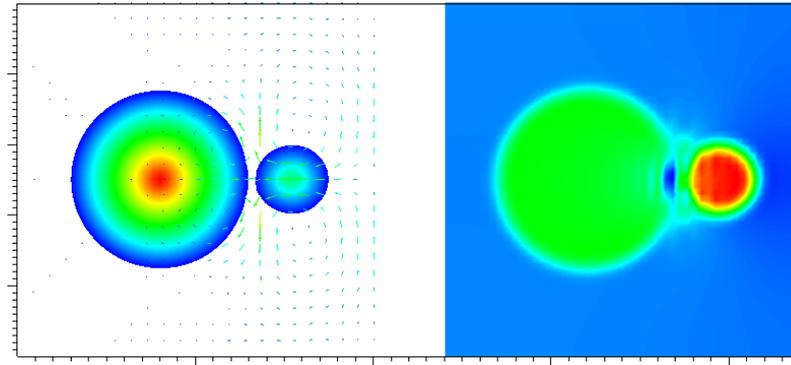


Figure 6.8: Coalescence 2D. Solutions at  $t = 2$ : level-set field and velocity arrows (left) and pressure field (right).

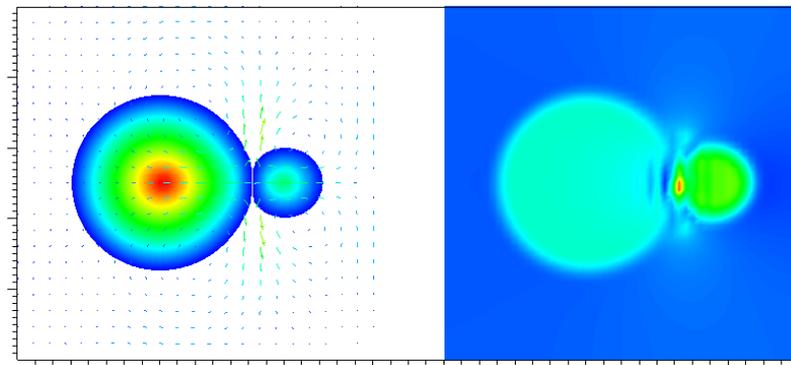


Figure 6.9: Coalescence 2D. Solutions at  $t = 6$ : level-set field and velocity arrows (left) and pressure field (right).

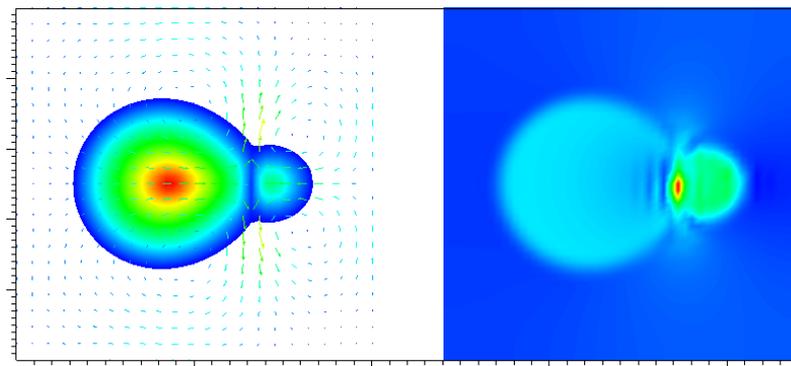


Figure 6.10: Coalescence 2D. Solutions at  $t = 10$ : level-set field and velocity arrows (left) and pressure field (right).

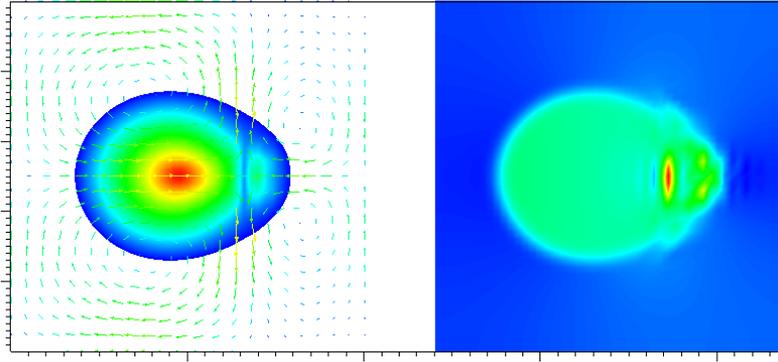


Figure 6.11: Coalescence 2D. Solutions at  $t = 18$ : level-set field and velocity arrows (left) and pressure field (right).

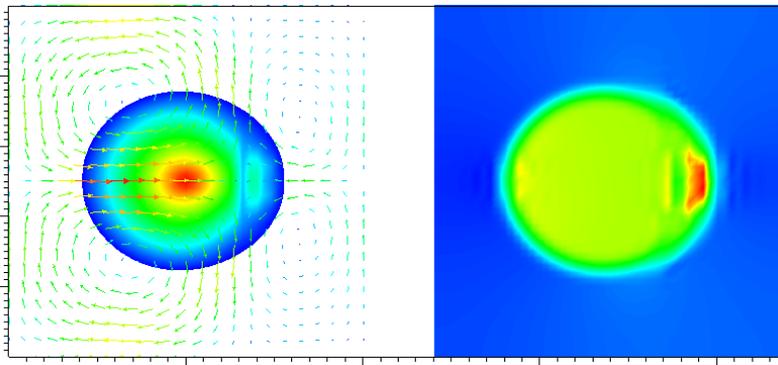


Figure 6.12: Coalescence 2D. Solutions at  $t = 30$ : level-set field and velocity arrows (left) and pressure field (right).

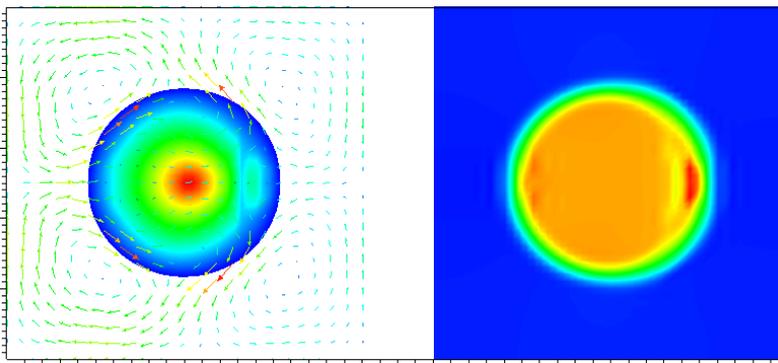


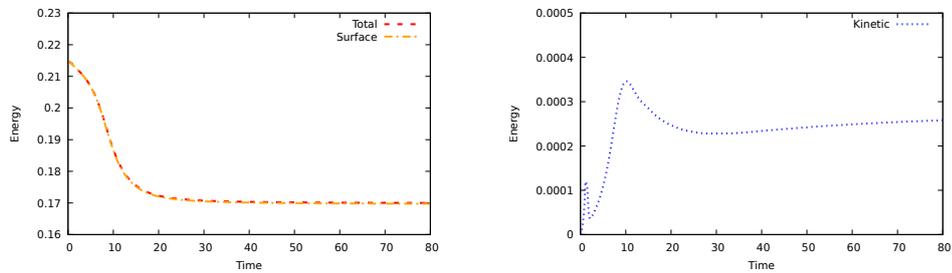
Figure 6.13: Coalescence 2D. Solutions at  $t = 80$ : level-set field and velocity arrows (left) and pressure field (right).

In the Figures 6.14 and 6.15a we show the energy evolution and dissipation. In this case the theoretical value of the initial surface energy is  $2\pi(r_1 + r_2)\sigma \approx 0.2199$ . We observe that the total and surface energies monotonically decrease in time. The kinetic energy increases when the droplets move towards each other ( $t < 10$ ) and decreases during the merging process and subsequently flattens out.

In order to test whether the equilibrium state has been reached we evaluate the circularity of the droplet. The circularity is defined as the fraction of the perimeter evaluated from the droplet volume and the perimeter itself:

$$\gamma = \frac{2 \left( \pi \int_{\{\Omega:\phi>0\}} d\Omega \right)^{1/2}}{\int_{\Omega} \delta_{\epsilon}(\phi) \|\nabla\phi\|_{\epsilon,2} d\Omega}. \tag{6.7.8}$$

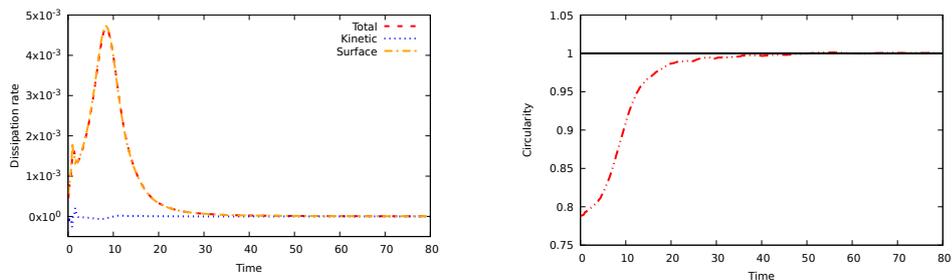
The circularity depicted in Figure 6.15b confirms the equilibrium state as  $\gamma$  tends to 1.



(a) Total and surface energy evolution.

(b) Kinetic energy evolution.

Figure 6.14: Coalescence 2D. Energy evolution.



(a) Energy dissipation rate.

(b) Circularity.

Figure 6.15: Coalescence 2D. Energy dissipation rate and circularity.

## 6.7.3 Droplet coalescence 3D

Here we simulate the merging of two droplets in three dimensions. We use the same physical parameters as in the two-dimensional case. The centers of the droplets are at  $\mathbf{c}_1 = (0.4, 0.5, 0.6)$  and  $\mathbf{c}_2 = (0.75, 0.5, 0.5)$  and the radii remain the same:  $r_1 = 0.25$  and  $r_2 = 0.1$ . Also here the diffuse interfaces of the droplets initially overlap. Again, to initialize the level-set we partition the domain, see Figure 6.16a and apply the standard distance initialization to each subdomain. The initial configuration is depicted in Figure 6.16b. We use  $50 \times 50 \times 50$  elements, set the time-step as  $\Delta t = 0.1$  and take  $\mathcal{C} = 0.1$ .

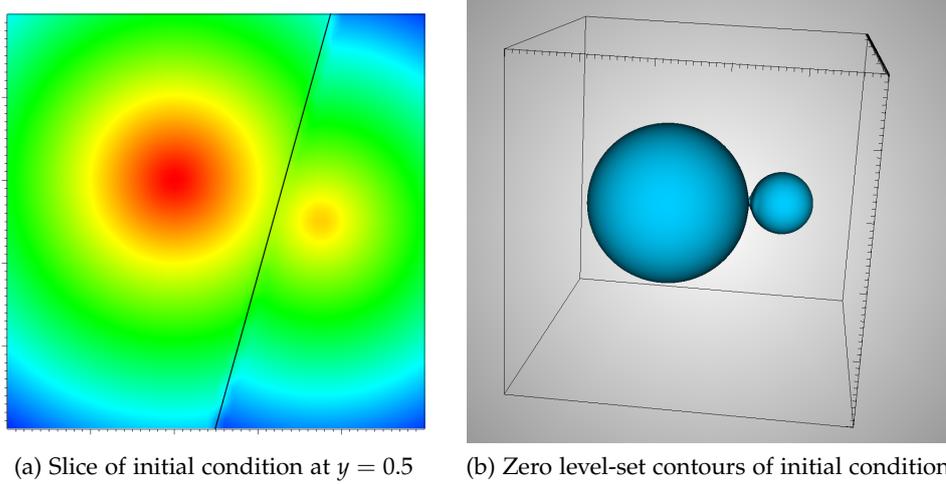


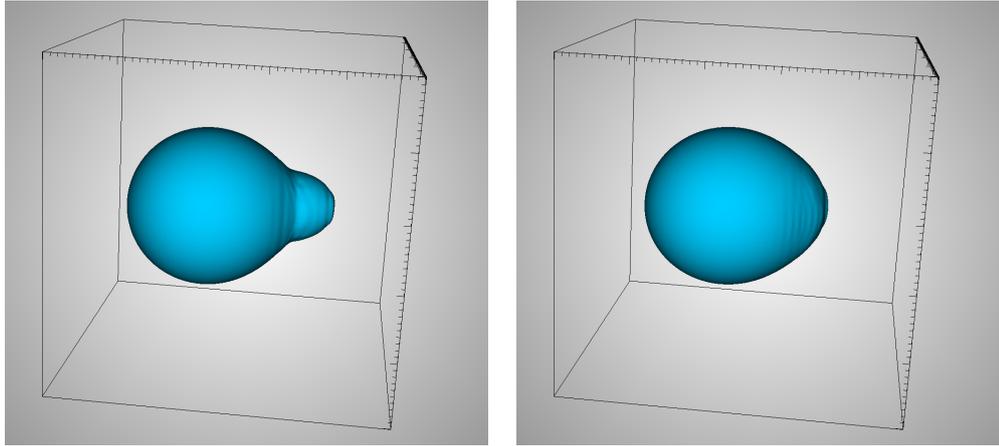
Figure 6.16: Coalescence 3D. Initial condition.

We show in Figure 6.17 snapshots of the merging process. In Figure 6.18 we visualize the energy evolution and dissipation. The theoretical value of the initial surface energy is  $4\pi(r_1^2 + r_2^2)\sigma \approx 0.0911$ . The behavior of the various energies is similar as in the two-dimensional case. Also in this case the energy-dissipative property of the numerical method is confirmed.

## 6.8 CONCLUSION

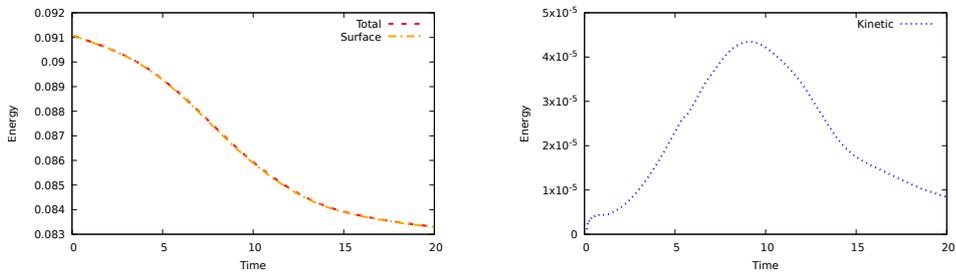
In this work we have proposed a new fully-discrete energy-stable level-set method for the incompressible Navier-Stokes equations with surface tension. To the best knowledge of the authors, this is the first provable energy-dissipative level-set method. Apart from being energetically stable, the method satisfies the maximum principle for the density and is pointwise divergence-free.

We have provided a consistent derivation of our regularized-interface level-set model starting from a sharp-interface model. In addition we have presented a detailed analysis of both models in term of energy behavior. This analysis implies that an energy-dissipative Galerkin-type discretization of the regularized-interface level-set model poses severe restrictions on the functional spaces. Independently, standard second-order temporal discretizations are also not associated with an energy-dissipative structure. Lastly, the standard regularized-interface model con-

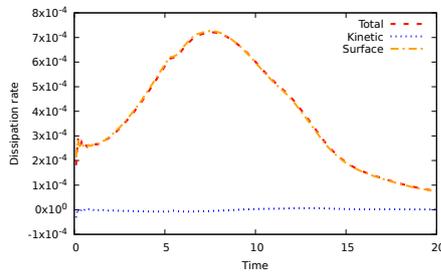


(a) Zero level-set contours at  $t = 10$  (b) Zero level-set contours at  $t = 20$

Figure 6.17: Coalescence 3D. Solutions at  $t = 10$  and  $t = 20$ .



(a) Total and surface energy evolution. (b) Kinetic energy evolution.



(c) Energy dissipation rate.

Figure 6.18: Coalescence 3D. Energy evolution and dissipation rate.

tains an unwanted regularization term. We circumvent each of these problems by creating extra space via the concept of functional entropy variables. This introduces an extra variable to the model which is coupled via the surface tension term. This leads in a natural way to the fully-discrete energy-stable level-set method. The eventual methodology use isogeometric analysis to ensure divergence-free solutions. Furthermore, the method is equipped with an SUPG stabilization mechanism in the level-set equation that is energetically-balanced in the momentum equation. Additionally, we use a residual-based discontinuity capturing term to stabilize the momentum equation. The temporal discretization is performed using a per-

turbed mid-point scheme. We have presented numerical examples in two and three dimensions which confirm the energy-stability of the method.

We see several research directions for further work. A first suggestion is to equip the developed method with multiscale stabilization mechanisms that are energetically stable. Attainable solutions may be inspired by stabilization mechanisms that are energetically stable for single fluid flow [58, 71]. Another possible research direction entails the development of energy-dissipative re-distancing procedures. A first thought here would be to use a two-step numerical scheme in which the redistancing procedure is decoupled from the current algorithm. The two above mentioned extensions would allow to simulate more violent flows, such as a dam-break problem, in an energy-dissipative manner. Lastly we note that another missing feature of the level-set method is local mass conservation. This might perhaps be obtained by using similar techniques as presented in this paper.



## 6.A EQUIVALENCE SURFACE TENSION MODELS

We show equivalence of the surface tension models for the sharp-interface model and the regularized-interface level-set model.

## 6.A.1 Sharp-interface model

In order to avoid directly evaluating the curvature in the surface tension term, one may employ integration by parts as proposed by Bänsch [13]. First we introduce some notation. The normal extensions of the scalar field  $f$  and vector field  $v$  defined on  $\Gamma$  are, see also [37]:

$$\hat{f}(\mathbf{x}) := f(\Pi_\Gamma(\mathbf{x})), \quad (6.A.1a)$$

$$\hat{v}(\mathbf{x}) := v(\Pi_\Gamma(\mathbf{x})), \quad (6.A.1b)$$

where  $\Pi_\Gamma(\mathbf{x})$  is defined as the normal projector of  $\mathbf{x}$  onto the interface  $\Gamma$ . The surface gradients of these fields are now given by

$$\nabla_\Gamma f := \nabla \hat{f}, \quad (6.A.2a)$$

$$\nabla_\Gamma v := \nabla \hat{v}, \quad (6.A.2b)$$

while the tangential divergence of  $v$  is the trace of the surface gradient:

$$\nabla_\Gamma \cdot v := \text{Tr}(\nabla_\Gamma v) = \nabla \cdot \hat{v}. \quad (6.A.3)$$

Note the slight abuse of notation; we use the same notation for the surface gradient as employed for the surface gradient in the regularized level-set model. Alternative expressions for the surface gradients are

$$\nabla_\Gamma f = \mathbf{P}_T \cdot \nabla f, \quad (6.A.4a)$$

$$\nabla_\Gamma v = \nabla v \cdot \mathbf{P}_T, \quad (6.A.4b)$$

where  $\mathbf{P}_T$  denotes the tangential projection tensor:

$$\mathbf{P}_T = \mathbf{I} - \hat{v} \otimes \hat{v}, \quad (6.A.5)$$

where  $\hat{v}$  is continuous extension of the outward unit normal pointing from  $\Omega_1$  into  $\Omega_2$  and  $\mathbf{I}$  is identity matrix. Using the above identities we have

$$\nabla \cdot \hat{w} = \nabla_\Gamma \cdot w = \text{Tr}(\nabla_\Gamma w) = \text{Tr}(\mathbf{P}_T \nabla w) = \mathbf{P}_T : \nabla w. \quad (6.A.6)$$

**Lemma 6.A.1.** *Buscaglia et al. [37]: For any tangentially differentiable vector field  $w$  we have:*

$$\int_{\Gamma(t)} \nabla_\Gamma \cdot w \, d\Gamma = \int_{\Gamma(t)} \kappa \hat{v} \cdot w \, d\Gamma + \int_{\partial\Gamma(t)} \nu_\partial \cdot w \, d(\partial\Gamma). \quad (6.A.7)$$

Using 6.A.6 and Lemma 6.A.1 we may write the surface tension term as

$$\begin{aligned} \frac{1}{\mathbb{W}e} \int_{\Gamma(t)} \kappa v \cdot w \, d\Gamma &= \frac{1}{\mathbb{W}e} \int_{\Gamma(t)} \nabla \cdot \hat{w} \, d\Gamma - \frac{1}{\mathbb{W}e} \int_{\partial\Gamma(t)} \nu_\partial \cdot w \, d(\partial\Gamma) \\ &= \frac{1}{\mathbb{W}e} \int_{\Gamma(t)} \mathbf{P}_T : \nabla w \, d\Gamma - \frac{1}{\mathbb{W}e} \int_{\partial\Gamma(t)} \nu_\partial \cdot w \, d(\partial\Gamma). \end{aligned} \quad (6.A.8)$$

## 6.A.2 Regularized-interface level-set model

In the following we utilize index notation.

**Proposition 6.A.2.** *It holds:*

$$\begin{aligned} \nabla_j ((P_T)_{ij}(\phi)\delta_\Gamma(\phi)) &= -\delta_\Gamma(\phi) \frac{\nabla_i\phi}{\|\nabla\phi\|_{\epsilon,2}} \nabla_j \frac{\nabla_j\phi}{\|\nabla\phi\|_{\epsilon,2}} \\ &\quad + \epsilon^2 \frac{\nabla_i\delta(\phi)}{\|\nabla\phi\|_{\epsilon,2}}. \end{aligned} \quad (6.A.9)$$

*Proof.* We compute

$$\begin{aligned} (P_T)_{ij}(\phi)\nabla_j\delta_\Gamma(\phi) &= \left( I_{ij} - \frac{\nabla_i\phi}{\|\nabla\phi\|_{\epsilon,2}} \frac{\nabla_j\phi}{\|\nabla\phi\|_{\epsilon,2}} \right) \\ &\quad \times \left( \delta(\phi) \frac{\nabla_k\phi}{\|\nabla\phi\|_{\epsilon,2}} \nabla_j \nabla_k\phi + \|\nabla\phi\|_{\epsilon,2} \nabla_j\delta(\phi) \right) \\ &= \delta(\phi) \frac{\nabla_k\phi}{\|\nabla\phi\|_{\epsilon,2}} \nabla_i \nabla_k\phi + \|\nabla\phi\|_{\epsilon,2} \nabla_i\delta(\phi) \\ &\quad - \frac{\nabla_i\phi}{\|\nabla\phi\|_{\epsilon,2}} \frac{\nabla_j\phi}{\|\nabla\phi\|_{\epsilon,2}} \delta(\phi) \frac{\nabla_k\phi}{\|\nabla\phi\|_{\epsilon,2}} \nabla_j \nabla_k\phi \\ &\quad - \frac{\nabla_i\phi}{\|\nabla\phi\|_{\epsilon,2}} \frac{\nabla_j\phi}{\|\nabla\phi\|_{\epsilon,2}} \|\nabla\phi\|_{\epsilon,2} \nabla_j\delta(\phi) \\ &= \delta_\Gamma(\phi) \frac{\nabla_k\phi}{\|\nabla\phi\|_{\epsilon,2}} \left( \frac{\nabla_i \nabla_k\phi}{\|\nabla\phi\|_{\epsilon,2}} \right. \\ &\quad \left. - \frac{\nabla_i\phi}{\|\nabla\phi\|_{\epsilon,2}} \frac{\nabla_j\phi}{\|\nabla\phi\|_{\epsilon,2}} \frac{\nabla_j \nabla_k\phi}{\|\nabla\phi\|_{\epsilon,2}} \right) \\ &\quad + \epsilon^2 \frac{\nabla_i\delta(\phi)}{\|\nabla\phi\|_{\epsilon,2}} \\ &= \delta_\Gamma(\phi) \frac{\nabla_k\phi}{\|\nabla\phi\|_{\epsilon,2}} (P_T)_{ij} \frac{\nabla_j \nabla_k\phi}{\|\nabla\phi\|_{\epsilon,2}} + \epsilon^2 \frac{\nabla_i\delta(\phi)}{\|\nabla\phi\|_{\epsilon,2}}. \end{aligned} \quad (6.A.10)$$

On the other hand we have:

$$\begin{aligned} \delta_\Gamma(\phi)\nabla_j(P_T)_{ij}(\phi) &= -\delta_\Gamma(\phi) \frac{\nabla_i\phi}{\|\nabla\phi\|_{\epsilon,2}} \nabla_j \frac{\nabla_j\phi}{\|\nabla\phi\|_{\epsilon,2}} - \delta_\Gamma(\phi) \nabla_j \frac{\nabla_j\phi}{\|\nabla\phi\|_{\epsilon,2}} \frac{\nabla_i\phi}{\|\nabla\phi\|_{\epsilon,2}} \\ &= -\delta_\Gamma(\phi) \frac{\nabla_i\phi}{\|\nabla\phi\|_{\epsilon,2}} \nabla_j \frac{\nabla_j\phi}{\|\nabla\phi\|_{\epsilon,2}} \\ &\quad - \delta_\Gamma(\phi) \frac{\nabla_j\phi}{\|\nabla\phi\|_{\epsilon,2}} \left( \frac{\nabla_j \nabla_i\phi}{\|\nabla\phi\|_{\epsilon,2}} \right. \\ &\quad \left. - \frac{\nabla_i\phi}{\|\nabla\phi\|_{\epsilon,2}} \frac{\nabla_k\phi}{\|\nabla\phi\|_{\epsilon,2}} \frac{\nabla_j \nabla_k\phi}{\|\nabla\phi\|_{\epsilon,2}} \right) \\ &= -\delta_\Gamma(\phi) \frac{\nabla_i\phi}{\|\nabla\phi\|_{\epsilon,2}} \nabla_j \frac{\nabla_j\phi}{\|\nabla\phi\|_{\epsilon,2}} \\ &\quad - \delta_\Gamma(\phi) \frac{\nabla_j\phi}{\|\nabla\phi\|_{\epsilon,2}} (P_T)_{ik} \frac{\nabla_k \nabla_j\phi}{\|\nabla\phi\|_{\epsilon,2}}. \end{aligned} \quad (6.A.11)$$

Addition of (6.A.10) and (6.A.11) yields:

$$\begin{aligned} \nabla_j ((P_T)_{ij}(\phi)\delta_\Gamma(\phi)) &= (P_T)_{ij}(\phi)\nabla_j\delta_\Gamma(\phi) + \delta_\Gamma(\phi)\nabla_j(P_T)_{ij}(\phi) \\ &= -\delta_\Gamma(\phi)\frac{\nabla_i\phi}{\|\nabla\phi\|_{\epsilon,2}}\nabla_j\frac{\nabla_j\phi}{\|\nabla\phi\|_{\epsilon,2}} \\ &\quad + \epsilon^2\frac{\nabla_i\delta(\phi)}{\|\nabla\phi\|_{\epsilon,2}}. \end{aligned} \quad (6.A.12)$$

□

**Lemma 6.A.3.** *It holds:*

$$\begin{aligned} \frac{1}{\mathbb{W}e}\int_\Omega\delta_\Gamma(\phi)\nabla_jw_i(P_T)_{ij}(\phi)\,d\Omega &= \frac{1}{\mathbb{W}e}\int_\Omega\delta_\Gamma(\phi)\frac{\nabla_i\phi}{\|\nabla\phi\|_{\epsilon,2}}\nabla_j\frac{\nabla_j\phi}{\|\nabla\phi\|_{\epsilon,2}}w_i\,d\Omega \\ &\quad - \frac{1}{\mathbb{W}e}\int_\Omega\epsilon^2\frac{\nabla_i\delta(\phi)}{\|\nabla\phi\|_{\epsilon,2}}w_i\,d\Omega. \end{aligned} \quad (6.A.13)$$

*Proof.* Performing integration by parts we get:

$$\begin{aligned} \frac{1}{\mathbb{W}e}\int_\Omega\delta_\Gamma(\phi)\nabla_jw_i(P_T)_{ij}(\phi)\,d\Omega &= -\frac{1}{\mathbb{W}e}\int_\Omega\nabla_j(\delta_\Gamma(\phi)(P_T)_{ij}(\phi))w_i\,d\Omega \\ &\quad + \frac{1}{\mathbb{W}e}\int_{\partial\Omega}\delta_\Gamma(\phi)n_jw_i(P_T)_{ij}(\phi)\,dS. \end{aligned} \quad (6.A.14)$$

Under the standing assumption we suppress the line force term. Using Proposition 6.A.2 finalizes the proof. □

## 6.B ENERGY EVOLUTION MIDPOINT LEVEL-SET DISCRETIZATION

We provide the energy evolution of a standard time-discrete level-set method using the midpoint rule. We consider the conservative discretization, which reads for time-step  $n$ :

Given  $\mathbf{u}_n, p_n$  and  $\phi_n$ , find  $\mathbf{u}_{n+1}, p_{n+1}$  and  $\phi_{n+1}$  such that:

$$\begin{aligned} \frac{[[\rho\mathbf{u}]]_n}{\Delta t_n} + \nabla \cdot (\rho_{n+1/2}\mathbf{u}_{n+1/2} \otimes \mathbf{u}_{n+1/2}) + \nabla p_{n+1} - \nabla \cdot \boldsymbol{\tau}(\mathbf{u}_{n+1/2}) \\ + \frac{1}{\mathbb{W}e}\kappa(\phi_{n+1/2})\mathbf{v}(\phi_{n+1/2})\delta_\Gamma(\phi_{n+1/2}) + \frac{1}{\mathbb{F}r^2}\rho_{n+1/2}\mathcal{J} = 0, \end{aligned} \quad (6.B.1a)$$

$$\nabla \cdot \mathbf{u}_{n+1/2} = 0, \quad (6.B.1b)$$

$$\frac{[[\phi]]_n}{\Delta t_n} + \mathbf{u}_{n+1/2} \cdot \nabla \phi_{n+1/2} = 0, \quad (6.B.1c)$$

where  $\rho \equiv \rho(\phi)$  on the indicated time-level.

**Theorem 6.B.1.** *The time-discrete formulation (6.B.1) satisfies the energy evolution property:*

$$\frac{[[\mathcal{E}(\mathbf{u}, \phi)]]_n}{\Delta t_n} = - \int_{\Omega} \nabla \mathbf{u}_{n+1/2} : \boldsymbol{\tau}(\mathbf{u}_{n+1/2}) \, d\Omega + \text{error} \quad (6.B.2a)$$

$$\begin{aligned} \text{error} &= \Delta t_n^2 \int_{\Omega} \frac{1}{8} \left\| \frac{[[\mathbf{u}]]_n}{\Delta t_n} \right\|^2 \frac{[[\rho]]_n}{\Delta t_n} \, d\Omega \\ &\quad - \frac{1}{\text{We}\Delta t_n} \int_{\Omega} [[\delta(\phi)]]_n (\|\nabla \phi_{n+1/2}\|_{\epsilon,2} - (\|\nabla \phi\|_{\epsilon,2})_{n+1/2}) \, d\Omega \\ &\quad - \frac{1}{\text{We}\Delta t_n} \int_{\Omega} [[\|\nabla \phi\|_{\epsilon,2}]]_n \\ &\quad \quad \times \left( \delta(\phi_{n+1/2}) \frac{\|\nabla \phi\|_{n+1/2}}{\|\nabla \phi_{n+1/2}\|_{\epsilon,2}} - \delta(\phi)_{n+1/2} \right) \, d\Omega \\ &\quad + \frac{1}{\text{We}\Delta t_n} \int_{\Omega} [[\phi]]_n^3 \times \\ &\quad \left( \delta^{(3)}(\phi_{n+1/2})/24 + [[\phi]]_n^2 \delta^{(5)}(\phi_{n+\xi})/1920 \right) \|\nabla \phi_{n+1/2}\|_{\epsilon,2} \, d\Omega \\ &\quad + \int_{\Omega} \frac{1}{\text{We}\Delta t_n} \delta'(\phi_{n+1/2}) [[\phi]]_n \frac{\epsilon^2}{\|\nabla \phi_{n+1/2}\|_{\epsilon,2}} \, d\Omega, \end{aligned} \quad (6.B.2b)$$

for some  $\xi \in (0, 1)$ .

**Remark 6.B.2.** *The semi-discrete convective method has the same energy evolution (6.B.2). For completeness we provide the convective method:*

Given  $\mathbf{u}_n, p_n$  and  $\phi_n$ , find  $\mathbf{u}_{n+1}, p_{n+1}$  and  $\phi_{n+1}$  such that:

$$\begin{aligned} \rho_{n+1/2} \left( \frac{[[\mathbf{u}]]_n}{\Delta t_n} + \mathbf{u}_{n+1/2} \cdot \nabla \mathbf{u}_{n+1/2} \right) + \nabla p_{n+1} - \nabla \cdot \boldsymbol{\tau}(\mathbf{u}_{n+1/2}) \\ + \frac{1}{\text{We}} \kappa(\phi_{n+1/2}) \mathbf{v}(\phi_{n+1/2}) \delta_{\Gamma}(\phi_{n+1/2}) - \frac{1}{\text{Fr}^2} \rho_{n+1/2} \mathbf{J} = 0, \end{aligned} \quad (6.B.3a)$$

$$\nabla \cdot \mathbf{u}_{n+1/2} = 0, \quad (6.B.3b)$$

$$\frac{[[\phi]]_n}{\Delta t_n} + \mathbf{u}_{n+1/2} \cdot \nabla \phi_{n+1/2} = 0, \quad (6.B.3c)$$

where  $\rho \equiv \rho(\phi)$  on the indicated time-level.

*Proof.* We give the proof for the conservative formulation, that of the convective formulation follows analogously. Multiplication of the continuity equation by  $q =$

$p_{n+1} - \rho_{n+1/2}(\frac{1}{2}\mathbf{u}_{n+1/2} \cdot \mathbf{u}_{n+1/2} + \frac{1}{\mathbb{F}\mathbb{R}^2}y)$  and the level-set equation by  $-(\llbracket \rho \rrbracket \frac{1}{2}\mathbf{u}_{n+1/2} \cdot \mathbf{u}_{n+1/2} - \frac{1}{\mathbb{F}\mathbb{R}^2}\llbracket \rho \rrbracket y + \frac{1}{\mathbb{W}\mathbb{e}}\kappa(\phi_{n+1/2}))\delta(\phi_{n+1/2})$  and subsequently integrating yields:

$$\int_{\Omega} (p_{n+1} - \rho_{n+1/2}(\frac{1}{2}\mathbf{u}_{n+1/2} \cdot \mathbf{u}_{n+1/2} - \frac{1}{\mathbb{F}\mathbb{R}^2}y)) \nabla \cdot \mathbf{u}_{n+1/2} \, d\Omega = 0, \quad (6.B.4a)$$

$$\begin{aligned} & - \int_{\Omega} \left( \frac{1}{2}\mathbf{u}_{n+1/2} \cdot \mathbf{u}_{n+1/2} - \frac{1}{\mathbb{F}\mathbb{R}^2}y \right) \\ & \times \left( \frac{\llbracket \rho \rrbracket_n}{\Delta t_n} + \mathbf{u}_{n+1/2} \cdot \nabla \rho_{n+1/2} \right) \, d\Omega \\ & - \int_{\Omega} \left( \frac{1}{\mathbb{W}\mathbb{e}}\kappa(\phi_{n+1/2})\delta(\phi_{n+1/2}) \right) \\ & \times \left( \frac{\llbracket \phi \rrbracket_n}{\Delta t_n} + \mathbf{u}_{n+1/2} \cdot \nabla \phi_{n+1/2} \right) \, d\Omega = 0. \end{aligned} \quad (6.B.4b)$$

We add the equations (6.B.4) and find:

$$\begin{aligned} & - \int_{\Omega} \left( \frac{1}{2}\mathbf{u}_{n+1/2} \cdot \mathbf{u}_{n+1/2} + \frac{1}{\mathbb{F}\mathbb{R}^2}y \right) \frac{\llbracket \rho \rrbracket_n}{\Delta t_n} \, d\Omega \\ & - \int_{\Omega} \frac{1}{\mathbb{W}\mathbb{e}}\kappa(\phi_{n+1/2})\delta(\phi_{n+1/2}) \frac{\llbracket \phi \rrbracket_n}{\Delta t_n} \, d\Omega = \\ & - \int_{\Omega} (p_{n+1} - \rho_{n+1/2}\frac{1}{2}\mathbf{u}_{n+1/2} \cdot \mathbf{u}_{n+1/2}) \nabla \cdot \mathbf{u}_{n+1/2} \, d\Omega \\ & + \int_{\Omega} \frac{1}{2}\mathbf{u}_{n+1/2} \cdot \mathbf{u}_{n+1/2} (\mathbf{u}_{n+1/2} \cdot \nabla \rho_{n+1/2}) \, d\Omega \\ & - \int_{\Omega} \frac{1}{\mathbb{F}\mathbb{R}^2}y (\mathbf{u}_{n+1/2} \cdot \nabla \rho_{n+1/2} + \rho_{n+1/2} \nabla \cdot \mathbf{u}_{n+1/2}) \, d\Omega \\ & + \int_{\Omega} \frac{1}{\mathbb{W}\mathbb{e}}\kappa(\phi_{n+1/2})\delta(\phi_{n+1/2}) \mathbf{u}_{n+1/2} \cdot \nabla \phi_{n+1/2} \, d\Omega. \end{aligned} \quad (6.B.5)$$

We take the second term on the left-hand side of (6.B.5) in isolation and perform integration by parts to get:

$$\begin{aligned} & - \int_{\Omega} \frac{1}{\mathbb{W}\mathbb{e}}\kappa(\phi_{n+1/2})\delta(\phi_{n+1/2}) \frac{\llbracket \phi \rrbracket_n}{\Delta t_n} \, d\Omega = \\ & \int_{\Omega} \frac{1}{\mathbb{W}\mathbb{e}} \nabla \left( \delta(\phi_{n+1/2}) \frac{\llbracket \phi \rrbracket_n}{\Delta t_n} \right) \cdot \frac{\nabla \phi_{n+1/2}}{\|\nabla \phi_{n+1/2}\|_{\epsilon,2}} \, d\Omega = \\ & \int_{\Omega} \frac{1}{\mathbb{W}\mathbb{e}\Delta t_n} \delta(\phi_{n+1/2}) \nabla \llbracket \phi \rrbracket_n \cdot \frac{\nabla \phi_{n+1/2}}{\|\nabla \phi_{n+1/2}\|_{\epsilon,2}} \, d\Omega \\ & + \int_{\Omega} \frac{1}{\mathbb{W}\mathbb{e}\Delta t_n} \delta'(\phi_{n+1/2}) \llbracket \phi \rrbracket_n \|\nabla \phi_{n+1/2}\|_{\epsilon,2} \, d\Omega \\ & - \int_{\Omega} \frac{1}{\mathbb{W}\mathbb{e}\Delta t_n} \delta'(\phi_{n+1/2}) \llbracket \phi \rrbracket_n \frac{\epsilon^2}{\|\nabla \phi_{n+1/2}\|_{\epsilon,2}} \, d\Omega. \end{aligned} \quad (6.B.6)$$

For the first term on the right-hand side we use

$$\nabla \llbracket \phi \rrbracket_n \cdot \frac{\nabla \phi_{n+1/2}}{\|\nabla \phi_{n+1/2}\|_{\epsilon,2}} = \llbracket \|\nabla \phi\|_{\epsilon,2} \rrbracket_n \frac{(\|\nabla \phi\|_{\epsilon,2})_{n+1/2}}{\|\nabla \phi_{n+1/2}\|_{\epsilon,2}}, \quad (6.B.7)$$

while for the second term employ a truncated Taylor series in the form:

$$\begin{aligned} \llbracket \delta(\phi) \rrbracket_n &= \llbracket \phi \rrbracket_n \delta^{(1)}(\phi_{n+1/2}) + \llbracket \phi \rrbracket_n^3 \delta^{(3)}(\phi_{n+1/2})/24 \\ &+ \llbracket \phi \rrbracket_n^5 \delta^{(5)}(\phi_{n+\xi})/1920, \end{aligned} \quad (6.B.8)$$

for some  $\xi \in (0, 1)$ . Substitution of (6.B.7)-(6.B.8) into (6.B.6) and reorganizing gives:

$$\begin{aligned}
& - \int_{\Omega} \frac{1}{\mathbb{W}e} \kappa(\phi_{n+1/2}) \delta(\phi_{n+1/2}) \frac{[\![\phi]\!]_n}{\Delta t_n} \, d\Omega \\
& = \frac{1}{\mathbb{W}e \Delta t_n} \int_{\Omega} \delta(\phi)_{n+1/2} [\![\|\nabla \phi\|_{\epsilon,2}]\!]_n + [\![\delta(\phi)]\!]_n (\|\nabla \phi\|_{\epsilon,2})_{n+1/2} \, d\Omega \\
& \quad + \frac{1}{\mathbb{W}e \Delta t_n} \int_{\Omega} [\![\delta(\phi)]\!]_n (\|\nabla \phi_{n+1/2}\|_{\epsilon,2} - (\|\nabla \phi\|_{\epsilon,2})_{n+1/2}) \, d\Omega \\
& \quad + \frac{1}{\mathbb{W}e \Delta t_n} \int_{\Omega} [\![\|\nabla \phi\|_{\epsilon,2}]\!]_n \left( \delta(\phi_{n+1/2}) \frac{\|\nabla \phi\|_{n+1/2}}{\|\nabla \phi_{n+1/2}\|_{\epsilon,2}} - \delta(\phi)_{n+1/2} \right) \, d\Omega \\
& \quad - \frac{1}{\mathbb{W}e \Delta t_n} \int_{\Omega} [\![\phi]\!]_n^3 \left( \delta^{(3)}(\phi_{n+1/2})/24 + [\![\phi]\!]_n^2 \delta^{(5)}(\phi_{n+\xi})/1920 \right) \\
& \quad \quad \times \|\nabla \phi_{n+1/2}\|_{\epsilon,2} \, d\Omega \\
& \quad - \int_{\Omega} \frac{1}{\mathbb{W}e \Delta t_n} \delta'(\phi_{n+1/2}) [\![\phi]\!]_n \frac{\epsilon^2}{\|\nabla \phi_{n+1/2}\|_{\epsilon,2}} \, d\Omega, \tag{6.B.9}
\end{aligned}$$

where the first term on the right-hand side represents the temporal change of surface energy (see Lemma 6.6.5):

$$\begin{aligned}
\frac{1}{\mathbb{W}e} \int_{\Omega} \frac{[\![\delta_{\Gamma}(\phi)]\!]_n}{\Delta t_n} \, d\Omega & = \frac{1}{\mathbb{W}e \Delta t_n} \int_{\Omega} \delta(\phi)_{n+1/2} [\![\|\nabla \phi\|_{\epsilon,2}]\!]_n \\
& \quad + [\![\delta(\phi)]\!]_n (\|\nabla \phi\|_{\epsilon,2})_{n+1/2} \, d\Omega. \tag{6.B.10}
\end{aligned}$$

Next we multiply the momentum equation by  $\mathbf{u}_{n+1/2}$  and subsequently integrate to get:

$$\begin{aligned}
& \int_{\Omega} \mathbf{u}_{n+1/2} \cdot \frac{[\![\rho \mathbf{u}]\!]_n}{\Delta t_n} \, d\Omega + \int_{\Omega} \mathbf{u}_{n+1/2} \nabla \cdot (\rho_{n+1/2} \mathbf{u}_{n+1/2} \otimes \mathbf{u}_{n+1/2}) \, d\Omega \\
& \quad + \int_{\Omega} \mathbf{u}_{n+1/2} \nabla p_{n+1} \, d\Omega + \int_{\Omega} \mathbf{u}_{n+1/2} \nabla \cdot \boldsymbol{\tau}(\mathbf{u}_{n+1/2}, \phi_{n+1/2}) \, d\Omega \\
& \quad + \int_{\Omega} \mathbf{u}_{n+1/2} \rho_{n+1/2} \frac{1}{\mathbb{F}r^2} \mathbf{j} \, d\Omega \\
& \quad + \int_{\Omega} \frac{1}{\mathbb{W}e} \kappa(\phi_{n+1/2}) \mathbf{u}_{n+1/2} \cdot \boldsymbol{\nu}(\phi_{n+1/2}) \delta_{\Gamma}(\phi_{n+1/2}) \, d\Omega = 0. \tag{6.B.11}
\end{aligned}$$

The time-derivative term may be written as

$$\begin{aligned}
\int_{\Omega} \mathbf{u}_{n+1/2} \cdot \frac{[\![\rho \mathbf{u}]\!]_n}{\Delta t_n} \, d\Omega & = \Delta t_n^{-1} \int_{\Omega} \frac{1}{2} \rho_{n+1} \|\mathbf{u}_{n+1/2}\|^2 - \frac{1}{2} \rho_n \|\mathbf{u}_n\|^2 \, d\Omega \\
& \quad + \Delta t_n^{-1} \int_{\Omega} \frac{1}{2} (\rho_{n+1} - \rho_n) \mathbf{u}_n \cdot \mathbf{u}_{n+1} \, d\Omega. \tag{6.B.12}
\end{aligned}$$

Writing the convective term into skew-symmetric form and working out the product gives:

$$\begin{aligned}
& \int_{\Omega} \mathbf{u}_{n+1/2} \nabla \cdot (\rho_{n+1/2} \mathbf{u}_{n+1/2} \otimes \mathbf{u}_{n+1/2}) \, d\Omega = \\
& \quad \int_{\Omega} \frac{1}{2} \|\mathbf{u}_{n+1/2}\|^2 \mathbf{u}_{n+1/2} \cdot \nabla \rho_{n+1/2} \, d\Omega \\
& \quad + \int_{\Omega} \frac{1}{2} \|\mathbf{u}_{n+1/2}\|^2 \rho_{n+1/2} \nabla \cdot \mathbf{u}_{n+1/2} \, d\Omega. \tag{6.B.13}
\end{aligned}$$

Substitution of (6.B.12)-(6.B.13) into (6.B.11) and performing integration by parts gives:

$$\begin{aligned}
& \Delta t_n^{-1} \int_{\Omega} \frac{1}{2} \rho_{n+1} \|\mathbf{u}_{n+1/2}\|^2 - \frac{1}{2} \rho_n \|\mathbf{u}_n\|^2 \, d\Omega = \\
& \quad - \int_{\Omega} \frac{1}{2} \|\mathbf{u}_{n+1/2}\|^2 \mathbf{u}_{n+1/2} \cdot \nabla \rho_{n+1/2} \, d\Omega \\
& \quad - \int_{\Omega} \frac{1}{2} \|\mathbf{u}_{n+1/2}\|^2 \rho_{n+1/2} \nabla \cdot \mathbf{u}_{n+1/2} \, d\Omega \\
& - \int_{\Omega} \mathbf{u}_{n+1/2} \nabla p_{n+1} \, d\Omega - \int_{\Omega} \mathbf{u}_{n+1/2} \rho_{n+1/2} \frac{1}{\mathbb{F}\mathbb{r}^2} \mathbf{j} \, d\Omega \\
& \quad + \int_{\Omega} \nabla \mathbf{u}_{n+1/2} : \boldsymbol{\tau}(\mathbf{u}_{n+1/2}, \phi_{n+1/2}) \, d\Omega \\
& - \int_{\Omega} \frac{1}{\mathbb{W}\mathbb{e}} \kappa(\phi_{n+1/2}) \mathbf{u}_{n+1/2} \cdot \boldsymbol{\nu}(\phi_{n+1/2}) \delta_{\Gamma}(\phi_{n+1/2}) \, d\Omega \\
& \quad - \Delta t_n^{-1} \int_{\Omega} \frac{1}{2} (\rho_{n+1} - \rho_n) \mathbf{u}_n \cdot \mathbf{u}_{n+1} \, d\Omega. \tag{6.B.14}
\end{aligned}$$

Addition of (6.B.5) and (6.B.14) while using (6.B.9)-(6.B.10) gives:

$$\begin{aligned}
& \Delta t_n^{-1} \int_{\Omega} \frac{1}{2} \rho_{n+1} \|\mathbf{u}_{n+1}\|^2 + \frac{1}{\mathbb{F}\mathbb{r}^2} \gamma \rho_{n+1} + \frac{1}{\mathbb{W}\mathbb{e}} \delta_{\Gamma}(\phi_{n+1}) \, d\Omega \\
& \quad - \Delta t_n^{-1} \int_{\Omega} \frac{1}{2} \rho_n \|\mathbf{u}_n\|^2 + \frac{1}{\mathbb{F}\mathbb{r}^2} \gamma \rho_n + \frac{1}{\mathbb{W}\mathbb{e}} \delta_{\Gamma}(\phi_n) \, d\Omega \\
& \quad = \int_{\Omega} \nabla \mathbf{u}_{n+1/2} : \boldsymbol{\tau}(\mathbf{u}_{n+1/2}, \phi_{n+1/2}) \, d\Omega \\
& \quad \quad \quad + \text{error}, \tag{6.B.15}
\end{aligned}$$

with error defined in (6.B.2b). Recognizing the left-hand side as the change in energy completes the proof.  $\square$



---

## CONCLUSIONS AND FUTURE WORK

---

The overarching theme in this dissertation has been the development of finite element isogeometric methods that inherit the stability properties of the underlying mathematical model. The emphasis lies on the construction of techniques applicable to free surface flow simulations. Stability issues in two-fluid simulations typically arise from either (i) the multiscale formulation, (ii) sharp layers or (iii) the interface separating the fluid. In order to address these problems, we had formulated the objectives:

- (i) Develop an energy-stable finite element method for turbulent flow.
- (ii) Derive a discontinuity capturing mechanism from the underlying physical system.
- (iii) Construct an energy-dissipative, maximum-principle satisfying numerical method for the simulation of free surface flows.

Each of these objectives is linked to one part of the thesis. The aim of this chapter is to outline the main conclusions of the research and to suggest possible future research directions.

### 7.1 CONCLUSIONS

We summarize the main conclusions that can be drawn from each of the three parts below.

#### 7.1.1 *Part I: Energy-dissipative multiscale formulations*

Stabilized finite element methods provide a way to stabilize the finite element solution. These methods may be derived from variational multiscale (VMS) analysis. This procedure provides a way to account for the small-scales: that part of the solution that does not fit on the mesh. This methodology provides nodally exact solutions for some linear one-dimensional model problems. However, complications arise when the mathematical models are time-dependent or nonlinear. The reason

lies in the selection of a splitting operator and in the construction of a model for the small-scales. These components do not fit in the rigorous multiscale framework that the VMS methodology offers as for the linear model problem. However, existing VMS methods still yield good numerical solutions for these problems. Unfortunately, these solutions are not energy-stable. The aim of this part of the dissertation has been to rectify this situation.

The standard small-scale models have a static nature: the time-derivative of the small-scale component is neglected in the eventual formulation. The main reason for this approximation is that it leads to a conceptually easier formulation. However, this step is not properly justified and as such upsets the energy behavior of the system. The remedy for this problem is to use dynamic small-scales. Dynamic small-scale models keep the small-scale time-derivative as a separate term in the model. A consequence is that the formulation consists of a large- and a small-scale part that are treated as separate equations.

The small-scale models contain an algebraic stabilization parameter, often denoted  $\tau$ . The parameter  $\tau$  is exact in the special case of stationary, linear, one-dimensional convection-diffusion equations with constant coefficients and linear elements. However, in the time-dependent, nonlinear, multidimensional case,  $\tau$  contains an approximation. It turns out that the standard approach does not comply with the energetic stability of the underlying model in this case. Numerical results show both positive and negative small-scale contributions to energy dissipation for the standard methods. The analysis reveals that energetic-stability can be retrieved when employing the  $H_0^1$ -projection operator. It is well-known this is a good projector for the convection-diffusion equations and it reduces the formulation to standard Galerkin in absence of convection. However, the projection operator is often not explicitly specified when deriving a VMS formulation. To arrive at an energy-stable method one can *assume*  $H_0^1$ -orthogonality converting the VMS formulation into a Galerkin/least-squares (GLS) formulation. An alternative, one may also *enforce* the  $H_0^1$ -projection operator via a separate equation.

Stabilized methods for the incompressible Navier-Stokes equations introduce some extra complications in terms of energetic stability. One wishes to retrieve the Galerkin method for laminar flows since this method is well-established in that regime. This suggests that the natural orthogonality for the incompressible Navier-Stokes equations is induced by the Stokes equations. Employing the Stokes projector significantly reduces the complexity of the formulation. It appears that a convenient property for energetic-stability is a divergence-conforming discretization. A certain selection of NURBS-spaces in the isogeometric method provide a way to obtain this. This is the road we take. To arrive at the eventual energy-stable numerical method we enforce divergence-free small-scale velocities fields and use a GLS-type formulation. The new numerical method shows that solutions improve, in terms of energy behavior, upon those obtained with the standard VMS method.

In conclusion, standard multiscale formulations employ several approximation steps that may upset the energetic-stability of the method. The multiscale framework itself provides the possibility to carefully select a combination of a projection operator and a small-scale model that does not upset the energetic-dissipative structure of the underlying model. The orthogonality implied by the multiscale projector can either be assumed or enforced.

### 7.1.2 Part II: A framework for discontinuity capturing methods

Discontinuity capturing methods aim to reduce or eliminate spurious oscillations in numerical solutions. Since the 1980's and many different forms of discontinuity capturing devices have been suggested. These methods have a large *ad hoc* component which gives an unsatisfying feeling and implies that there is room for improvement. Moreover, the fact that (very) good results have been obtained with this technique suggests that there should be an underlying theoretical foundation. Apart from the lacking profound understanding of the origin of the instabilities, it was unclear how the discontinuity capturing term enters the finite element formulation. The purpose of this part of the dissertation has been to resolve these issues.

The standard entropy solutions provide a stability concept which is based on the solution itself. On the other hand, it is well-known that spurious oscillations typically appear near sharp layers in solution profiles. Since sharp layers are characterized by large gradients we may identify the origin of these oscillations by looking into the behavior of the gradient of the solution. These observations suggest to develop a stability concept based on the gradient of the solution. This has led to a new stability concept for nonlinear conservation laws called *variation entropy theory*.

Variation entropy theory provides an entropy stability concept based on the gradient of the solution instead of on the solution itself. The associated *variation entropy solutions* are those solution that, apart from solving the conservation law, satisfy the so-called *variation entropy condition*. Harming this condition is the source of the instabilities.

Now that the origin of the instabilities has been established, the question is how to use this insight to construct a numerical method with a discontinuity capturing mechanism. The answer is variational multiscale analysis. The VMS approach provides a way to incorporate the effect of the missing scales into a numerical method. In the VMS methodology the user may select the optimality projector to split the solution into large- and small-scales. This is where the variation entropy condition pops up: to reduce oscillations near sharp layers one chooses variation entropy solutions as the optimal solutions.

Following this path naturally leads to a stabilized method with a discontinuity capturing mechanism. The new discontinuity capturing diffusion does not use any *ad hoc* devices and is based on the variation entropy condition. In absence of sharp layers the variation entropy condition is fulfilled, and in these regions the discontinuity capturing term vanishes. Near sharp layers, however, the variation entropy condition might be harmed. If so, the discontinuity capturing term switches on; dissipation based on the amount of variation entropy production is added to the formulation. The discontinuity capturing term provides diffusion in the direction identified by the change of the variation entropy. This implies that diffusion is not added at the sharp layer but right next to it. The corresponding numerical results show that, compared to other discontinuity capturing methods, the new discontinuity method is more localized *near* sharp layers. These are the region where Gibbs oscillation typically pop up.

In conclusion, these insights provide a profound understanding of the origin of the spurious oscillations and the discontinuity capturing mechanisms. Additionally, this framework sheds light on the TVD (total variation diminishing) property since

the TVD property may be derived from variation entropy theory. The derivation of a discontinuity capturing operator consists of two ingredients: variation entropy theory and variational multiscale analysis. Variation entropy theory tells us where to add viscosity and VMS tells us how to add the viscosity. In Figure 7.1 we visualize the relations between the various concepts.

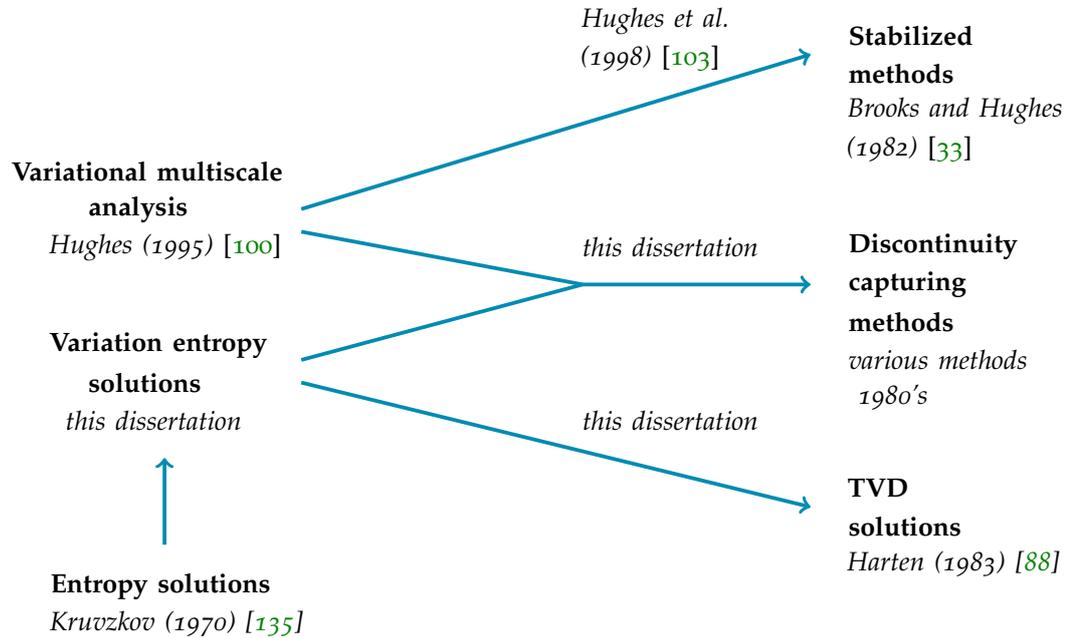


Figure 7.1: The connection between the various concepts in stabilized methods.

### 7.1.3 Part III: An energy-dissipative method for free surface flow

Level-set methods have proven to be powerful methods for the simulation of complex (flow) phenomena in which a large amount of topological changes occur. One of the key features is the satisfaction of the maximum principle for the dependent fields, which are for flow simulations the density and the viscosity. Unfortunately, the level-set methods have some deficiencies. In this part of the dissertation we have rectified one of these issues being the possibility of artificial energy creation.

Starting from a sharp-interface model of the incompressible Navier-Stokes equations, we have derived the standard level-set regularized-interface model. A standard regularization procedure used for this model makes the model energy-inconsistent with the sharp-interface model. More importantly, standard spatial and temporal discretization procedures are not energy-stable. The reason is that both the discretization in space and time require the selection of certain weighting functions that are not available in the standard setting. The key remedy is to create extra ‘space’ via the concept of *functional entropy variables*. This introduces a new variable, which represents the unavailable test function, into the model. The new variable is coupled to the existing model via the surface tension term. This concept turns out to resolve each of the three problems mentioned above: the regularization and the test function choice for spatial and temporal discretization. Similar as for the

methodology developed in Part I, it turns out essential to have divergence-free discrete velocity field. As such, this is where isogeometric analysis naturally comes into play. The fully-discrete energy-dissipative methodology follows when using a novel perturbed mid-point scheme for the temporal discretization.

In conclusion, the standard level-set model for the incompressible Navier-Stokes equations with surface tension is too restrictive to obtain a guaranteed, locally energy-dissipative discretization. Functional entropy variables provide a possibility to circumvent this and allow to develop an energy-dissipative discretization.

## 7.2 FUTURE WORK

The main goal of this thesis is to develop algorithms that inherit the stability properties of the underlying mathematical model with particular emphasize on techniques relevant for free surface flow. Various interesting methods have been suggested and their applicability has been shown. This has several opened doors for the development of improved techniques and for applying the tailored methods to other relevant problems. We note that instabilities emerging from boundary conditions have not been addressed in this thesis. This has been the topic of another work [179]. Below we suggest some possibilities and first thoughts for future work.

- A first suggestion would be to construct a stabilized energy-dissipative method for two-fluid flow. To establish this one could look into merging the techniques developed in Part I and III. We note that recently a new stabilized energy-dissipative method has been proposed by J.A. Evans et al. [71]. This method goes for a large extend along the same lines as the techniques developed in Part I and also uses the Stokes projector. This method might be useful alternative for the method of Part I.
- The established discontinuity capturing mechanism in Part II may be seen as a penalty method. In order to guarantee satisfaction of the variation entropy condition one may try to enforce this via a Lagrange multiplier construction.
- The framework in Part II has been established for scalar conservation laws. We anticipate that one can directly apply discontinuity capturing devices inspired by this framework to each equation of a system of conservation laws. However, rigorously establishing the extension to systems of conservation laws is not trivial and requires some additional effort.
- Even though the level-set method has now been equipped with the energy-stability property several deficiencies remain. We mention the absence of local mass conservation and the seemingly artificial redistancing procedures. The first issue is simply the consequence of the fact that the regularized Dirac delta is not a member of the discretization space. Perhaps the ideas used by functional entropy variables may turn out useful here. The reason is that this framework provides a way to consistently convert the model into a form which can circumvent the need to select a particular weighting function. Concerning redistancing methods, we hope that an ‘ideal’ level-set method does not require these procedures at all. Until that is available, we suggest to look into redistancing methods that do not upset the thermodynamical

structure of the method. One may need to introduce a redistancing energy to achieve this.

As a last remark we wish to emphasize that the developed *variation entropy concept* is a general theory and its applicability is not limited to a certain discretization method. As such, it may have profound impact for the design of many discretization methods for problems in which shock waves or sharp layers occur. An interesting class of methods in this regard are the discontinuous Galerkin methods. Recently, S.K.F. Stoter et al. [178, 180] have shown these methods may be derived by using discontinuous approximation spaces in the variational multiscale method. Perhaps the framework of S.K.F. Stoter et al. can be combined with the strategy used in Part II of this thesis.

---

## BIBLIOGRAPHY

---

- [1] T. Abadie, J. Aubin, and D. Legendre. “On the combined effects of surface tension force calculation and interface advection on spurious currents within Volume of Fluid and Level Set frameworks.” *Journal of Computational Physics* 297 (2015), pp. 611–636.
- [2] I. Akkerman. “Monotone level-sets on arbitrary meshes without redistancing.” *Computers & Fluids* 146 (2017), pp. 74–85.
- [3] I. Akkerman, Y. Bazilevs, D.J. Benson, M.W. Farthing, and C.E. Kees. “Free-Surface Flow and Fluid-Object Interaction Modeling With Emphasis on Ship Hydrodynamics.” *Journal of Applied Mechanics* 79 (2012), p. 010905.
- [4] I. Akkerman, Y. Bazilevs, V.M. Calo, T.J.R. Hughes, and S. Hulshoff. “The role of continuity in residual-based variational multiscale modeling of turbulence.” *Computational Mechanics* 41 (2008), pp. 371–378.
- [5] I. Akkerman, Y. Bazilevs, C. Kees, and M. Farthing. “Isogeometric Analysis of Free-Surface Flow.” *Journal of Computational Physics* 230 (2011), pp. 4137–4152.
- [6] I. Akkerman and M.F.P. ten Eikelder. “Toward free-surface flow simulations with correct energy evolution: an isogeometric level-set approach with monolithic time-integration.” *Computers & Fluids* 181 (2019), pp. 77–89.
- [7] D.N. Arnold, F. Brezzi, B. Cockburn, and L.D. Marini. “Unified analysis of Discontinuous Galerkin methods for elliptic problems.” *SIAM Journal of Numerical Analysis* 39 (2002), pp. 1749–1779.
- [8] E. Audusse and M.-O. Bristeau. “A well-balanced positivity preserving “second-order” scheme for shallow water flows on unstructured meshes.” *Journal of Computational Physics* 206.1 (2005), pp. 311–333.
- [9] I. Babuška. “The finite element method with Lagrangian multipliers.” *Numerische Mathematik* 20.3 (1973), pp. 179–192.
- [10] M.R. Baer and J.W. Nunziato. “A two-phase mixture theory for the deflagration-to-detonation transition (DDT) in reactive granular materials.” *International journal of multiphase flow* 12 (1986), pp. 861–889.
- [11] S. Balay, W. Gropp, L. C. McInnes, and B. Smith. “PETSc 2.0 Users Manual.” Mathematics and Computer Science Division, Argonne National Laboratory <http://www.mcs.anl.gov/petsc>. 2000.

- [12] S. Balay, W.D. Gropp, L. Curfman McInnes, and B. F. Smith. "Efficient Management of Parallelism in Object Oriented Numerical Software Libraries." *Modern Software Tools in Scientific Computing*. Ed. by E. Arge, A. M. Bruaset, and H. P. Langtangen. Birkhäuser Press, 1997, pp. 163–202.
- [13] E. Bänsch. "Finite element discretization of the Navier–Stokes equations with a free capillary surface." *Numerische Mathematik* 88 (2001), pp. 203–235.
- [14] C. Bardos, A.-Y. LeRoux, and J.-C. Nédélec. "First order quasilinear equations with boundary conditions." *Communications in partial differential equations* 4 (1979), pp. 1017–1034.
- [15] C. Bayona-Roa, R. Codina, and J. Baiges. "Variational multiscale error estimators for the adaptive mesh refinement of compressible flow simulations." *Computer Methods in Applied Mechanics and Engineering* 337 (2018), pp. 501–526.
- [16] Y. Bazilevs and I. Akkerman. "Large eddy simulation of turbulent Taylor-Couette flow using isogeometric analysis and the residual-based variational multiscale method." *Journal of Computational Physics* 229 (2010), pp. 3402–3414.
- [17] Y. Bazilevs, V.M. Calo, J.A. Cottrel, T.J.R. Hughes, A. Reali, and G. Scovazzi. "Variational multiscale residual-based turbulence modeling for large eddy simulation of incompressible flows." *Computer Methods in Applied Mechanics and Engineering* 197 (2007), pp. 173–201.
- [18] Y. Bazilevs, V.M. Calo, T.J.R. Hughes, and Y. Zhang. "Isogeometric fluid-structure interaction: theory, algorithms, and computations." *Computational mechanics* 43 (2008), pp. 3–37.
- [19] Y. Bazilevs, V.M. Calo, T.E. Tezduyar, and T.J.R. Hughes. "YZ $\beta$  Discontinuity-Capturing for Advection-Dominated Processes with Application to Arterial Drug Delivery." *International Journal of Numerical Methods in Fluids* 54 (2007), pp. 593–608.
- [20] Y. Bazilevs, V.M. Calo, Y. Zhang, and T.J.R. Hughes. "Isogeometric fluid-structure interaction analysis with applications to arterial blood flow." *Computational Mechanics* 38 (2006), pp. 310–322.
- [21] Y. Bazilevs, V.M. Calo, Y. Zhang, and T.J.R. Hughes. "Isogeometric fluid-structure interaction analysis with applications to arterial blood flow." *Computational Mechanics* 38 (4-5 2006), pp. 310–322.
- [22] Y. Bazilevs, J.R. Gohean, T.J.R. Hughes, R.D. Moser, and Y. Zhang. "Patient-Specific Isogeometric Fluid-Structure Interaction Analysis of Thoracic Aortic Blood Flow due to Implantation of the Jarvik 2000 Left Ventricular Assist Device." *Computer Methods in Applied Mechanics and Engineering* 198 (2009), pp. 3534–3550.
- [23] Y. Bazilevs, M-C Hsu, and M.A. Scott. "Isogeometric fluid-structure interaction analysis with emphasis on non-matching discretizations, and with application to wind turbines." *Computer Methods in Applied Mechanics and Engineering* 249 (2012), pp. 28–41.

- [24] Y. Bazilevs, K. Takizawa, and T.E. Tezduyar. "New directions and challenging computations in fluid dynamics modeling with stabilized and multiscale methods." *Mathematical Models and Methods in Applied Sciences* 25.12 (2015), pp. 2217–2226.
- [25] Y. Bazilevs, K. Takizawa, T.E. Tezduyar, M.-C. Hsu, Y. Otoguro, H. Mochizuki, and M.C.H. Wu. "Wind Turbine and Turbomachinery Computational Analysis with the ALE and Space-Time Variational Multiscale Methods and Isogeometric Discretization." *Journal of Advanced Engineering and Computation* 4.1 (2020), pp. 1–32.
- [26] D.J. Benson, Y. Bazilevs, M.-C. Hsu, and T.J.R. Hughes. "Isogeometric Shell Analysis: The Reissner-Mindlin Shell." *Computer Methods in Applied Mechanics and Engineering* 199 (2010), pp. 276–289.
- [27] L.E. Blumenson. "A derivation of n-dimensional spherical coordinates." *The American Mathematical Monthly* 67 (1960), pp. 63–66.
- [28] A. Bonito, J.-L. Guermond, and B. Popov. "Stability analysis of explicit entropy viscosity methods for non-linear scalar conservation equations." *Mathematics of Computation* 83 (2014), pp. 1039–1062.
- [29] Z. Bontinck, J. Corno, H. De Gerssem, S. Kurz, A. Pels, S. Schöps, F. Wolf, C. de Falco, J. Dölz, R. Vázquez, et al. "Recent advances of isogeometric analysis in computational electromagnetics." *ICS Newsletter* 24 (2017).
- [30] M.E. Brachet, D.I. Meiron, S.A. Orszag, B.G. Nickel, R.H. Morf, and U. Frisch. "Small-scale structure of the Taylor–Green vortex." *Journal of Fluid Mechanics* 130 (1983), pp. 411–452.
- [31] J.U. Brackbill, D.B. Kothe, and C. Zemach. "A continuum method for modeling surface tension." *Journal of computational physics* 100 (1992), pp. 335–354.
- [32] F. Brezzi. "On the existence, uniqueness and approximation of saddle-point problems arising from Lagrangian multipliers." *Publications mathématiques et informatique de Rennes* S4 (1974), pp. 1–26.
- [33] A.N. Brooks and T.J.R. Hughes. "Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations." *Computer Methods in Applied Mechanics and Engineering* 32 (1982), pp. 199–259.
- [34] A. Buffa, C. De Falco, and G. Sangalli. "IsoGeometric Analysis: Stable elements for the 2D Stokes equation." *International Journal for Numerical Methods in Fluids* 65 (2011), pp. 1407–1422.
- [35] A. Buffa, J. Rivas, G. Sangalli, and R. Vázquez. "Isogeometric Discrete Differential Forms in Three Dimensions." *SIAM Journal on Numerical Analysis* 49 (2011), pp. 818–844.
- [36] A. Buffa, G. Sangalli, and R. Vázquez. "Isogeometric analysis in electromagnetics: B-splines approximation." *Computer Methods in Applied Mechanics and Engineering* 199 (2010), pp. 1143–1152.

- [37] G.C. Buscaglia and R.F. Ausas. "Variational formulations for surface tension, capillarity and wetting." *Computer Methods in Applied Mechanics and Engineering* 200 (2011), pp. 3011–3025.
- [38] J.W. Cahn and J.E. Hilliard. "Free energy of a nonuniform system. I. Interfacial free energy." *The Journal of chemical physics* 28 (1958), pp. 258–267.
- [39] J.W. Cahn and J.E. Hilliard. "Free energy of a nonuniform system. III. Nucleation in a two-component incompressible fluid." *The Journal of chemical physics* 31 (1959), pp. 688–699.
- [40] F. Chalot, T.J.R. Hughes, and F. Shakib. "Symmetrization of conservation laws with entropy for high-temperature hypersonic computations." *Computing systems in engineering* 1 (1990), pp. 495–521.
- [41] Y.C. Chang, T.Y. Hou, B. Merriman, and S. Osher. "A level set formulation of Eulerian interface capturing methods for incompressible fluid flows." *Journal of computational Physics* 124 (1996), pp. 449–464.
- [42] S. Cho and S.-H. Ha. "Isogeometric shape design optimization: exact geometry and enhanced sensitivity." *Structural and Multidisciplinary Optimization* 38 (2009), p. 53.
- [43] I. Christov and B. Popov. "New non-oscillatory central schemes on unstructured triangulations for hyperbolic systems of conservation laws." *Journal of Computational Physics* 227 (2008), pp. 5736–5757.
- [44] J. Chung and G. M. Hulbert. "A time integration algorithm for structural dynamics with improved numerical dissipation: The generalized- $\alpha$  method." *Journal of Applied Mechanics* 60 (1993), pp. 371–75.
- [45] B. Cockburn, G.E. Karniadakis, and C.-W. Shu. "The development of discontinuous Galerkin methods." *Discontinuous Galerkin Methods*. Springer, 2000, pp. 3–50.
- [46] R. Codina. "Stabilized finite element approximation of transient incompressible flows using orthogonal subscales." *Computer Methods in Applied Mechanics and Engineering* 191 (2002), pp. 4295–4321.
- [47] R. Codina, S. Badia, J. Baiges, and J. Principe. "Variational multiscale methods in computational fluid dynamics." *Encyclopedia of computational mechanics*, E. Stein, R. de Borst, and T.J.R. Hughes, Eds., Wiley Online Library (2017).
- [48] R. Codina, J. Principe, O. Guasch, and S. Badia. "Time dependent subscales in the stabilized finite element approximation of incompressible flow problems." *Computer Methods in Applied Mechanics and Engineering* 196 (2007), pp. 2413–2430.
- [49] O. Colomés, S. Badia, R. Codina, and J. Principe. "Assessment of variational multiscale models for the large eddy simulation of turbulent incompressible flows." *Computer Methods in Applied Mechanics and Engineering* 285 (2015), pp. 32–63.
- [50] J.A. Cottrell, T.J.R. Hughes, and Y. Bazilevs. *Isogeometric analysis: toward integration of CAD and FEA*. John Wiley & Sons, 2009.

- [51] J.A. Cottrell, A. Reali, Y. Bazilevs, and T.J.R. Hughes. "Isogeometric analysis of structural vibrations." *Computer methods in applied mechanics and engineering* 195 (2006), pp. 5257–5296.
- [52] R. Courant. "Variational methods for the solution of problems of equilibrium and vibrations." *Verlag nicht ermittelbar* (1943).
- [53] F. Daude and P. Galon. "On the computation of the Baer–Nunziato model using ALE formulation with HLL- and HLLC-type solvers towards fluid–structure interactions." *Journal of Computational Physics* 304 (2016), pp. 189–230.
- [54] P.A.B. De Sampaio and A.L.G.A. Coutinho. "A natural derivation of discontinuity capturing operator for convection–diffusion problems." *Computer methods in applied mechanics and engineering* 190 (2001), pp. 6291–6308.
- [55] L. Demkowicz. "Babuška  $\Leftrightarrow$  Brezzi?" *Technical report, ICES* (2006).
- [56] J. Douglas and J.P. Wang. "An absolutely stabilized finite element method for the Stokes problem." *Mathematics of computation* 52.186 (1989), pp. 495–508.
- [57] M.F.P. ten Eikelder and I. Akkerman. "Correct energy evolution of stabilized formulations: The relation between VMS, SUPG and GLS via dynamic orthogonal small-scales and isogeometric analysis. I: The convective–diffusive context." *Computer Methods in Applied Mechanics and Engineering* 331 (2018), pp. 259–280.
- [58] M.F.P. ten Eikelder and I. Akkerman. "Correct energy evolution of stabilized formulations: The relation between VMS, SUPG and GLS via dynamic orthogonal small-scales and isogeometric analysis. II: The incompressible Navier–Stokes equations." *Computer Methods in Applied Mechanics and Engineering* 340 (2018), pp. 1135–1159.
- [59] M.F.P. ten Eikelder and I. Akkerman. "Variation entropy: a continuous local generalization of the TVD property using entropy principles." *Computer Methods in Applied Mechanics and Engineering* 355 (2019), pp. 261–283.
- [60] M.F.P. ten Eikelder and I. Akkerman. "An energy-dissipative level-set method for the incompressible Navier–Stokes equations with surface tension using functional entropy variables." *under review* (2020).
- [61] M.F.P. ten Eikelder, Y. Bazilevs, and I. Akkerman. "A theoretical framework for discontinuity capturing: Joining variational multiscale analysis and variation entropy theory." *Computer Methods in Applied Mechanics and Engineering* 359 (2020), p. 112664.
- [62] M.F.P. ten Eikelder, F. Daude, and B. Koren. "A Lagrange-projection-like numerical scheme for mixed acoustic-convective two-phase flows." *ASME 2016 Pressure Vessels and Piping Conference*. American Society of Mechanical Engineers Digital Collection. 2016.
- [63] M.F.P. ten Eikelder, F. Daude, B. Koren, and A.S. Tijsseling. "An acoustic-convective splitting-based approach for the Kapila two-phase flow model." *Journal of Computational Physics* 331 (2017), pp. 188–208.

- [64] S. Elgeti and H. Sauerland. "Deforming fluid domains within the finite element method: five mesh-based tracking methods in comparison." *Archives of Computational Methods in Engineering* 23 (2016), pp. 323–361.
- [65] J.A. Evans. "Divergence-free B-spline discretizations for viscous incompressible flows." Ph.D. Thesis. The University of Texas at Austin, 2011.
- [66] J.A. Evans, Y. Bazilevs, I. Babuška, and T.J.R. Hughes. "N-widths, sup-infs, and optimality ratios for the k-version of the isogeometric finite element method." *Computer Methods in Applied Mechanics and Engineering* 198 (2009), pp. 1726–1741.
- [67] J.A. Evans and T.J.R. Hughes. "Variational multiscale analysis: A new link between flux correction, total variation, and constrained optimization." *ICES Report 2010-35, The University of Texas at Austin* (2010).
- [68] J.A. Evans and T.J.R. Hughes. "Isogeometric divergence-conforming B-splines for the steady Navier–Stokes equations." *Mathematical Models and Methods in Applied Sciences* 23 (2013), pp. 1421–1478.
- [69] J.A. Evans and T.J.R. Hughes. "Isogeometric divergence-conforming B-splines for the unsteady Navier–Stokes equations." *Journal of Computational Physics* 241 (2013), pp. 141–167.
- [70] J.A. Evans, T.J.R. Hughes, and G. Sangalli. "Enforcement of constraints and maximum principles in the variational multiscale method." *Computer Methods in Applied Mechanics and Engineering* 199 (2009), pp. 61–76.
- [71] J.A. Evans, D. Kamensky, and Y. Bazilevs. "Variational multiscale modeling with discretely divergence-free subscales." *Computers & Mathematics with Applications* (2020).
- [72] L.C. Evans. *Partial Differential Equations (Graduate Studies in Mathematics, Vol. 19)*. Providence, RI: American Mathematical Society, 2002.
- [73] L. P. Franca and S. Frey. "Stabilized finite element methods: II. The incompressible Navier-Stokes equations." *Computer Methods in Applied Mechanics and Engineering* 99 (1992), pp. 209–233.
- [74] L. P. Franca, S. Frey, and T.J.R. Hughes. "Stabilized finite element methods: I. Application to the advective-diffusive model." *Computer Methods in Applied Mechanics and Engineering* 95 (1992), pp. 253–276.
- [75] M.M. Francois, S.J. Cummins, E.D. Dendy, D.B. Kothe, J.M. Sicilian, and M.W. Williams. "A balanced-force algorithm for continuous and sharp interfacial surface tension models within a volume tracking framework." *Journal of Computational Physics* 213 (2006), pp. 141–173.
- [76] K.O. Friedrichs and P.D. Lax. "Systems of conservation equations with a convex extension." *Proceedings of the National Academy of Sciences* 68 (1971), pp. 1686–1688.
- [77] A.C. Galeao and E.G. Dutra do Carmo. "A consistent approximate upwind Petrov–Galerkin method for convection-dominated problems." *Computer Methods in Applied Mechanics and Engineering* 68 (1988), pp. 83–95.

- [78] H. Garcke, B. Nestler, and B. Stoth. "On anisotropic order parameter models for multi-phase systems and their sharp interface limits." *Physica D: Nonlinear Phenomena* 115 (1998), pp. 87–108.
- [79] J. Giesselmann, C. Makridakis, and T. Pryer. "Energy consistent discontinuous Galerkin methods for the Navier–Stokes–Korteweg system." *Mathematics of Computation* 83 (2014), pp. 2071–2099.
- [80] H. Gómez, V.M. Calo, Y. Bazilevs, and T.J.R. Hughes. "Isogeometric analysis of the Cahn–Hilliard phase-field model." *Computer methods in applied mechanics and engineering* 197 (2008), pp. 4333–4352.
- [81] H. Gomez, T.J.R. Hughes, X. Nogueira, and V.M. Calo. "Isogeometric analysis of the isothermal Navier-Stokes-Korteweg equations." *Computer Methods in Applied Mechanics and Engineering* 199 (2010), pp. 1828–1840.
- [82] H. Gomez, A. Reali, and G. Sangalli. "Accurate, efficient, and (iso) geometrically flexible collocation methods for phase-field models." *Journal of Computational Physics* 262 (2014), pp. 153–171.
- [83] S. Gross and A. Reusken. "Finite element discretization error analysis of a surface tension force in two-phase incompressible flows." *SIAM journal on numerical analysis* 45 (2007), pp. 1679–1700.
- [84] J.-L. Guermond and M. Nazarov. "A maximum-principle preserving Co finite element method for scalar conservation equations." *Computer Methods in Applied Mechanics and Engineering* 272 (2014), pp. 198–213.
- [85] J.-L. Guermond, R. Pasquetti, and B. Popov. "Entropy viscosity method for nonlinear conservation laws." *Journal of Computational Physics* 230 (2011), pp. 4248–4267.
- [86] I. Harari and T.J.R. Hughes. "What are  $C$  and  $h$ ?: Inequalities for the analysis and design of finite element methods." *Computer Methods in Applied Mechanics and Engineering* 97 (1992), pp. 157–192.
- [87] F.H. Harlow and J.E. Welch. "Numerical calculation of time-dependent viscous incompressible flow of fluid with free surface." *The physics of fluids* 8 (1965), pp. 2182–2189.
- [88] A. Harten. "High resolution schemes for hyperbolic conservation laws." *Journal of computational physics* 49 (1983), pp. 357–393.
- [89] A. Harten. "On the symmetric form of systems of conservation laws with entropy." *Journal of computational physics* 49 (1983), pp. 151–164.
- [90] A. Harten. "On a class of high resolution total-variation-stable finite-difference schemes." *SIAM Journal on Numerical Analysis* 21 (1984), pp. 1–23.
- [91] G. Hauke and T.J.R. Hughes. "A comparative study of different sets of variables for solving compressible and incompressible flows." *Computer Methods in Applied Mechanics and Engineering* 153.1-2 (1998), pp. 1–44.
- [92] A.J. Herrema, J. Kiendl, and M.-C. Hsu. "A framework for isogeometric-analysis-based optimization of wind turbine blade structures." *Wind Energy* 22 (2019), pp. 153–170.

- [93] C.W. Hirt and B.D. Nichols. "Volume of fluid (VOF) method for the dynamics of free boundaries." *Journal of computational physics* 39 (1981), pp. 201–225.
- [94] J. Holmen, T.J.R. Hughes, A.A. Oberai, and G.N. Wells. "Sensitivity of the scale partition for variational multiscale LES of channel flow." *Physics of Fluids* 16 (2004), pp. 824–827.
- [95] L. Hörmander. *The analysis of linear partial differential operators I: Distribution theory and Fourier analysis*. Springer, 2015.
- [96] A. Hrennikoff. "Solution of problems of elasticity by the framework method." *Journal of Applied Mechanics* 8 (1941), 169–175.
- [97] M.-C. Hsu, D. Kamensky, Y. Bazilevs, M.S. Sacks, and T.J.R. Hughes. "Fluid–structure interaction analysis of bioprosthetic heart valves: significance of arterial wall deformation." *Computational mechanics* 54 (2014), pp. 1055–1071.
- [98] M.-C. Hsu, D. Kamensky, F. Xu, J. Kiendl, C. Wang, M.C.H. Wu, J. Mineroff, A. Reali, Y. Bazilevs, and M.S. Sacks. "Dynamic and fluid–structure interaction simulations of bioprosthetic heart valves using parametric design with T-splines and Fung-type material models." *Computational mechanics* 55 (2015), pp. 1211–1225.
- [99] M.C. Hsu, Y. Bazilevs, V.M. Calo, T.E. Tezduyar, and T.J.R. Hughes. "Improving stability of multiscale formulations of fluid flow at small time steps." *Computer Methods in Applied Mechanics and Engineering* 199 (2010), pp. 828–840.
- [100] T.J.R. Hughes. "Multiscale phenomena: Green’s functions, the Dirichlet-to-Neumann formulation, subgrid scale models, bubbles and the origins of stabilized methods." *Computer Methods in Applied Mechanics and Engineering* 127 (1995), pp. 387–401.
- [101] T.J.R. Hughes, J.A. Cottrell, and Y. Bazilevs. "Isogeometric analysis: CAD, finite elements, NURBS, exact geometry, and mesh refinement." *Computer Methods in Applied Mechanics and Engineering* 194 (2005), pp. 4135–4195.
- [102] T.J.R. Hughes, G. Engel, L. Mazzei, and M. Larson. "The Continuous Galerkin Method is Locally Conservative." *Journal of Computational Physics* 163 (2000), pp. 467–488.
- [103] T.J.R. Hughes, G. Feijóo., L. Mazzei, and J. B. Quincy. "The variational multiscale method – A paradigm for computational mechanics." *Computer Methods in Applied Mechanics and Engineering* 166 (1998), pp. 3–24.
- [104] T.J.R. Hughes and L. P. Franca. "A new finite element formulation for fluid dynamics: VII. The Stokes problem with various well-posed boundary conditions: Symmetric formulations that converge for all velocity/pressure spaces." *Computer Methods in Applied Mechanics and Engineering* 65 (1987), pp. 85–96.
- [105] T.J.R. Hughes, L. P. Franca, and M. Balestra. "A new finite element formulation for fluid dynamics: V. A stable Petrov-Galerkin formulation of the Stokes problem accommodating equal-order interpolations." *Computer Methods in Applied Mechanics and Engineering* 59 (1986), pp. 85–99.

- [106] T.J.R. Hughes, L.P. Franca, and G.M. Hulbert. "A new finite element formulation for computational fluid dynamics: VIII. The Galerkin/least-squares method for advective-diffusive equations." *Computer Methods in Applied Mechanics and Engineering* 73 (1989), pp. 173–189.
- [107] T.J.R. Hughes, L.P. Franca, and M. Mallet. "A new finite element formulation for computational fluid dynamics: I. Symmetric forms of the compressible Euler and Navier-Stokes equations and the second law of thermodynamics." *Computer Methods in Applied Mechanics and Engineering* 54 (1986), pp. 223–234.
- [108] T.J.R. Hughes, L.P. Franca, and M. Mallet. "A new finite element formulation for fluid dynamics: VI. Convergence analysis of the generalized SUPG formulation for linear time-dependent multidimensional advective-diffusive systems." *Computer Methods in Applied Mechanics and Engineering* 63 (1987), pp. 97–112.
- [109] T.J.R. Hughes, W.K. Liu, and T.K. Zimmermann. "Lagrangian-Eulerian finite element formulation for incompressible viscous flows." *Computer methods in applied mechanics and engineering* 29 (1981), pp. 329–349.
- [110] T.J.R. Hughes and M. Mallet. "A new finite element formulation for computational fluid dynamics: IV. A discontinuity-capturing operator for multidimensional advective-diffusive systems." *Computer Methods in Applied Mechanics and Engineering* 58 (1986), pp. 329–336.
- [111] T.J.R. Hughes and M. Mallet. "A new finite element formulation for fluid dynamics: III. The generalized streamline operator for multidimensional advective-diffusive systems." *Computer Methods in Applied Mechanics and Engineering* 58 (1986), pp. 305–328.
- [112] T.J.R. Hughes, M. Mallet, and A. Mizukami. "A new finite element formulation for computational fluid dynamics: II. Beyond SUPG." *Computer Methods in Applied Mechanics and Engineering* 54 (1986), pp. 341–355.
- [113] T.J.R. Hughes and G. Sangalli. "Variational Multiscale Analysis: the Fine-scale Green's Function, Projection, Optimization, Localization, and Stabilized Methods." *SIAM Journal of Numerical Analysis* 45 (2007), pp. 539–557.
- [114] T.J.R. Hughes, G. Scovazzi, and T.E. Tezduyar. "Stabilized methods for compressible flows." *Journal of Scientific Computing* 43 (2010), pp. 343–368.
- [115] T.J.R. Hughes and J. Stewart. "A space-time formulation for multiscale phenomena." *J. Comput. Appl. Math.* 74 (1996), pp. 217–229.
- [116] T.J.R. Hughes and T.E. Tezduyar. "Finite Element Methods for First-order Hyperbolic Systems with Particular Emphasis on the Compressible Euler Equations." *Computer Methods in Applied Mechanics and Engineering* 45 (1984), pp. 217–284.
- [117] T.J.R. Hughes and G.N. Wells. "Conservation properties for the Galerkin and stabilised forms of the advection-diffusion and incompressible Navier-Stokes equations." *Computer Methods in Applied Mechanics and Engineering* 194 (2005), pp. 1141–1159.
- [118] A. Jameson. "Computational aerodynamics for aircraft design." *Science* 245 (1989), pp. 361–371.

- [119] A. Jameson. "Positive schemes and shock modelling for compressible flows." *International Journal for Numerical Methods in Fluids* 20 (1995), pp. 743–776.
- [120] K.E. Jansen, S.S. Collis, C. Whiting, and F. Shaki. "A better consistency for low-order stabilized finite element methods." *Computer methods in applied mechanics and engineering* 174 (1999), pp. 153–170.
- [121] V. John and P. Knobloch. "On spurious oscillations at layers diminishing (SOLD) methods for convection-diffusion equations: Part I - A review." *Computer Methods in Applied Mechanics and Engineering* 196 (2007), pp. 2197–2215.
- [122] V. John and P. Knobloch. "On spurious oscillations at layers diminishing (SOLD) methods for convection–diffusion equations: Part II–Analysis for P<sub>1</sub> and Q<sub>1</sub> finite elements." *Computer Methods in Applied Mechanics and Engineering* 197 (2008), pp. 1997–2014.
- [123] D. Juric and G. Tryggvason. "Computations of boiling flows." *International journal of multiphase flow* 24 (1998), pp. 387–410.
- [124] D. Kamensky and Y. Bazilevs. "tIGAr: Automating isogeometric analysis with FEniCS." *Computer Methods in Applied Mechanics and Engineering* 344 (2019), pp. 477–498.
- [125] D. Kamensky, M.-C. Hsu, D. Schillinger, J.A. Evans, A. Aggarwal, Y. Bazilevs, M.S. Sacks, and T.J.R. Hughes. "An immersogeometric variational framework for fluid–structure interaction: Application to bioprosthetic heart valves." *Computer methods in applied mechanics and engineering* 284 (2015), pp. 1005–1053.
- [126] A.K. Kapila, R. Menikoff, J.B. Bdzil, S.F. Son, and D.S. Stewart. "Two-phase modeling of deflagration-to-detonation transition in granular materials: Reduced equations." *Physics of fluids* 13 (2001), pp. 3002–3024.
- [127] K.H. Karlsen, K. Brudal, H.K. Dahle, S. Evje, and K.-A. Lie. "The corrected operator splitting approach applied to a nonlinear advection-diffusion problem." *Computer Methods in Applied Mechanics and Engineering* 167 (1998), pp. 239–260.
- [128] J. Kiendl, Y. Bazilevs, M.-C. Hsu, R. Wüchner, and K.-U. Bletzinger. "The bending strip method for isogeometric analysis of Kirchhoff–Love shell structures comprised of multiple patches." *Computer Methods in Applied Mechanics and Engineering* 199 (2010), pp. 2403–2416.
- [129] J. Kiendl, K.-U. Bletzinger, J. Linhard, and R. Wüchner. "Isogeometric shell analysis with Kirchhoff–Love elements." *Computer Methods in Applied Mechanics and Engineering* 198 (2009), pp. 3902–3914.
- [130] P.M. Knupp. "Algebraic mesh quality metrics." *SIAM journal on scientific computing* 23 (2001), pp. 193–218.
- [131] B. Koobus and C. Farhat. "A variational multiscale method for the large eddy simulation of compressible turbulent flows on unstructured meshes – application to vortex shedding." *Computer Methods in Applied Mechanics and Engineering* 193 (2004), pp. 1367–1383.

- [132] B. Koren. "A robust upwind discretization method for advection, diffusion and source terms." *Numerical methods for advection-diffusion problems*. Vieweg, 1993.
- [133] A. Korobenko, X. Deng, J. Yan, and Y. Bazilevs. "Recent Advances in Fluid–Structure Interaction Simulations of Wind Turbines." *Advances in Computational Fluid-Structure Interaction and Flow Simulation*. Springer, 2016, pp. 489–500.
- [134] D.J. Korteweg. "Sur la forme que prennent les équations du mouvements des fluides si l'on tient compte des forces capillaires causées par des variations de densité considérables mais connues et sur la théorie de la capillarité dans l'hypothèse d'une variation continue de la densité." *Archives Néerlandaises des Sciences exactes et naturelles* 6 (1901), pp. 1–24.
- [135] S.N. Kružkov. "First order quasilinear equations in several independent variables." *Mathematics of the USSR-Sbornik* 10 (1970), p. 217.
- [136] C. Kublik and R. Tsai. "Integration over curves and surfaces defined by the closest point mapping." *Research in the mathematical sciences* 3 (2016), p. 3.
- [137] A. Kurganov, G. Petrova, and B. Popov. "Adaptive semidiscrete central-upwind schemes for nonconvex hyperbolic conservation laws." *SIAM Journal on Scientific Computing* 29 (2007), pp. 2381–2401.
- [138] G.J. Le Beau, S.E. Ray, S.K. Aliabadi, and T.E. Tezduyar. "SUPG finite element computation of compressible flows with the entropy and conservation variables formulations." *Computer Methods in Applied Mechanics and Engineering* 104 (1993), pp. 397–422.
- [139] G.J. Le Beau and T.E. Tezduyar. *Finite element computation of compressible flows with the SUPG formulation*. Army High Performance Computing Research Center, 1991.
- [140] R.J. LeVeque. *Numerical Methods for Conservation Laws*. Birkhauser Verlag, Boston, 1992.
- [141] E. F. Lins, R. N. Elias, F. A. Rochinha, and A. L. G. A. Coutinho. "Residual-based variational multiscale simulation of free surface flows." *Computational Mechanics* 46 (2010), pp. 545–557.
- [142] S. Lipton, J.A. Evans, Y. Bazilevs, T. Elguedj, and T.J.R. Hughes. "Robustness of isogeometric structural discretizations under severe mesh distortion." *Computer Methods in Applied Mechanics and Engineering* 199 (2010), pp. 357–373.
- [143] J. Liu. "Thermodynamically consistent modeling and simulation of multi-phase flows." Ph.D. Thesis, Department of Computational Science, Engineering, and Mathematics. The University of Texas at Austin, 2014.
- [144] J. Liu, H. Gomez, J.A. Evans, T.J.R. Hughes, and C.M. Landis. "Functional entropy variables: a new methodology for deriving thermodynamically consistent algorithms for complex fluids, with particular reference to the isothermal Navier–Stokes–Korteweg equations." *Journal of Computational Physics* 248 (2013), pp. 47–86.

- [145] J. Liu, C.M. Landis, H. Gomez, and T.J.R. Hughes. "Liquid–vapor phase transition: Thermomechanical theory, entropy stable numerical formulation, and boiling simulations." *Computer Methods in Applied Mechanics and Engineering* 297 (2015), pp. 476–553.
- [146] D. Lörst ad and L. Fuchs. "High-order surface tension VOF-model for 3D bubble flows with high density ratio." *Journal of computational physics* 200.1 (2004), pp. 153–176.
- [147] A. Masud and A.A. Al-Naseem. "Variationally derived discontinuity capturing methods: Fine scale models with embedded weak and strong discontinuities." *Computer Methods in Applied Mechanics and Engineering* 340 (2018), pp. 1102–1134.
- [148] O. Mehmetoglu and B. Popov. "Maximum principle and convergence of central schemes based on slope limiters." *Mathematics of Computation* 81 (2012), pp. 219–231.
- [149] G. Moutsanidis, D. Kamensky, J.S. Chen, and Y. Bazilevs. "Hyperbolic phase field modeling of brittle fracture: Part II—immersed IGA–RKPM coupling for air-blast–structure interaction." *Journal of the Mechanics and Physics of Solids* 121 (2018), pp. 114–132.
- [150] A. Murrone and H. Guillard. "A five equation reduced model for compressible two phase flow problems." *Journal of Computational Physics* 202 (2005), pp. 664–698.
- [151] S. Nagrath, K. E. Jansen, and R. T. Jr. Lahey. "Computation of incompressible bubble dynamics with a stabilized finite element level set method." *Computer Methods in Applied Mechanics and Engineering* 194 (2005), pp. 4565–4587.
- [152] V.P. Nguyen, C. Anitescu, S.P.A. Bordas, and T. Rabczuk. "Isogeometric analysis: an overview and computer implementation aspects." *Mathematics and Computers in Simulation* 117 (2015), pp. 89–116.
- [153] B.D. Nichols and C.W. Hirt. "Improved free surface boundary conditions for numerical incompressible-flow calculations." *Journal of computational physics* 8 (1971), pp. 434–448.
- [154] T.M. Opstal, J. Yan, J.A. Evans, T. Kvamsdal, and Y. Bazilevs. "Isogeometric divergence-conforming variational multiscale formulation of incompressible turbulent flows." *Computer Methods in Applied Mechanics and Engineering* 361 (2017), pp. 859–879.
- [155] S. Osher and R.P. Fedkiw. "Level set methods: an overview and some recent results." *Journal of Computational physics* 169 (2001), pp. 463–502.
- [156] S. Osher and J.A. Sethian. "Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations." *Journal of computational physics* 79 (1988), pp. 12–49.
- [157] J.E. Pilliod Jr and E.G. Puckett. "Second-order accurate volume-of-fluid algorithms for tracking material interfaces." *Journal of Computational Physics* 199 (2004), pp. 465–502.
- [158] S. Popinet. "Numerical models of surface tension." *Annual Review of Fluid Mechanics* 50 (2018), pp. 49–75.

- [159] J. Principe, R. Codina, and F. Henke. "The dissipative structure of variational multiscale methods for incompressible flows." *Computer Methods in Applied Mechanics and Engineering* 199 (2010), pp. 791–801.
- [160] J. Prüss and G. Simonett. "On the two-phase Navier-Stokes equations with surface tension." *Interfaces Free Bound* 12 (2010), pp. 311–345.
- [161] J. Qian, G. Tryggvason, and C.K. Law. "A front tracking method for the motion of premixed flames." *Journal of Computational Physics* 144 (1998), pp. 52–69.
- [162] X. Qian. "Full analytical sensitivities in NURBS based isogeometric shape optimization." *Computer Methods in Applied Mechanics and Engineering* 199 (2010), pp. 2059–2071.
- [163] S. Ramakrishnan and S.S. Collis. "Turbulence control simulation using the variational multiscale method." *AIAA Journal* 42 (2004), pp. 745–753.
- [164] W.H. Reed and T.R. Hill. *Triangular mesh methods for the neutron transport equation*. Tech. rep. Los Alamos Scientific Lab., N. Mex. (USA), 1973.
- [165] G. Sangalli. "Construction of a natural norm for the convection-diffusion-reaction operator." *Bollettino dell'Unione Matematica Italiana* 7 (2004), pp. 335–355.
- [166] R. Saurel, F. Petitpas, and R.A. Berry. "Simple and efficient relaxation methods for interfaces separating compressible fluids, cavitating flows and shocks in multiphase mixtures." *Journal of Computational Physics* 228 (2009), pp. 1678–1712.
- [167] I. Seric, S. Afkhami, and L. Kondic. "Direct numerical simulation of variable surface tension flows using a volume-of-fluid method." *Journal of Computational Physics* 352 (2018), pp. 615–636.
- [168] J. A. Sethian. "Evolution, Implementation, and Application of Level Set and Fast Marching Methods for Advancing Fronts." *Journal of Computational Physics* 169 (2001), pp. 503–555.
- [169] J.A. Sethian. *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*. Vol. 3. Cambridge university press, 1999.
- [170] F. Shakib and T.J.R. Hughes. "A new finite element formulation for computational fluid dynamics: IX. Fourier analysis of space-time Galerkin/least-squares algorithms." *Computer Methods in Applied Mechanics and Engineering* 87 (1991), pp. 35–58.
- [171] F. Shakib, T.J.R. Hughes, and Z. Johan. "A new finite element formulation for computational fluid dynamics: X. The compressible Euler and Navier-Stokes equations." *Computer Methods in Applied Mechanics and Engineering* 89 (1991), pp. 141–219.
- [172] H. Shen, C.-Y. Wen, M. Parsani, and C.-W. Shu. "Maximum-principle-satisfying space-time conservation element and solution element scheme applied to compressible multifluids." *Journal of Computational Physics* 330 (2017), pp. 668–692.

- [173] M. Shokrpour Roudbari and E.H. van Brummelen. "Binary-fluid–solid interaction based on the Navier–Stokes–Korteweg equations." *Mathematical Models and Methods in Applied Sciences* 29 (2019), pp. 995–1036.
- [174] M. Shokrpour Roudbari, G. Şimşek, E.H. van Brummelen, and K.G. van der Zee. "Diffuse-interface two-phase flow models with different densities: A new quasi-incompressible form and a linear energy-stable method." *Mathematical Models and Methods in Applied Sciences* 28 (2018), pp. 733–770.
- [175] C.-W. Shu. "Total-variation-diminishing time discretizations." *SIAM Journal on Scientific and Statistical Computing* 9 (1988), pp. 1073–1084.
- [176] J. Sokolowski and J.-P. Zolésio. "Introduction to shape optimization." *Introduction to Shape Optimization*. Springer, 1992.
- [177] H.A. Stone. "A simple derivation of the time-dependent convective-diffusion equation for surfactant transport along a deforming interface." *Physics of Fluids A: Fluid Dynamics* 2 (1990), pp. 111–112.
- [178] S.K.F. Stoter. "The variational multiscale method for discontinuous Galerkin type finite element formulations." Ph.D. Thesis, Department of Civil, Environmental and Geo-engineering. University of Minnesota, 2019.
- [179] S.K.F. Stoter, M.F.P. ten Eikelder, F. de Prenter, I. Akkerman, E.H. van Brummelen, C.V. Verhoosel, and D. Schillinger. "A unification of variational multiscale analysis and Nitsche's method, and a resulting boundary layer fine-scale model." *under review* (2019).
- [180] S.K.F. Stoter, S.R. Turteltaub, S.J. Hulshoff, and D. Schillinger. "Residual-based variational multiscale modeling in a discontinuous Galerkin framework." *Multiscale Modeling & Simulation* 16.3 (2018), pp. 1333–1364.
- [181] M. Sussman, P. Smereka, and S.J. Osher. "A level set approach for computing solutions to incompressible two-phase flows." *Journal of Computational Physics* 114 (1994), pp. 146–159.
- [182] P.K. Sweby. "High resolution schemes using flux limiters for hyperbolic conservation laws." *SIAM journal on numerical analysis* 21 (1984), pp. 995–1011.
- [183] E. Tadmor. "Entropy functions for symmetric systems of conservation laws." *Journal of Mathematical Analysis and Applications* 122 (1983), pp. 355–359.
- [184] E. Tadmor. "Skew-selfadjoint form for systems of conservation laws." *Journal of Mathematical Analysis and Applications* 103 (1984), pp. 428–442.
- [185] T.E. Tezduyar. "Finite element methods for fluid dynamics with moving boundaries and interfaces." *Encyclopedia of computational mechanics* (2004).
- [186] T.E. Tezduyar, M. Behr, S. Mittal, and J. Liou. "A new strategy for finite element computations involving moving boundaries and interfaces—the deforming-spatial-domain/space-time procedure: II. Computation of free-surface flows, two-liquid flows, and flows with drifting cylinders." *Computer methods in applied mechanics and engineering* 94 (1992), pp. 353–371.

- [187] T.E. Tezduyar and T.J.R. Hughes. "Development of time-accurate finite element techniques for first-order hyperbolic systems with particular emphasis on the compressible Euler equations." *NASA Technical Report NASA-CR-204772*, NASA (1982).
- [188] T.E. Tezduyar and Y. Osawa. "Finite Element Stabilization Parameters Computed from Element Matrices and Vectors." *Computer Methods in Applied Mechanics and Engineering* 190 (2000), pp. 411–430.
- [189] T.E. Tezduyar and M. Senga. "SUPG Finite Element Computation of Inviscid Supersonic Flows with  $YZ\beta$  Shock-Capturing." *Computers & Fluids* 36 (2007), pp. 147–159.
- [190] T.E. Tezduyar, M. Senga, and D. Vicker. "Computation of Inviscid Supersonic Flows around Cylinders and Spheres with the SUPG Formulation and  $YZ\beta$  Shock-Capturing." *Computational Mechanics* 38 (2006), pp. 469–481.
- [191] T.Q. Thai, T. Rabczuk, Y. Bazilevs, and G. Meschke. "A higher-order stress-based gradient-enhanced damage model based on isogeometric analysis." *Computer Methods in Applied Mechanics and Engineering* 304 (2016), pp. 584–604.
- [192] E.F. Toro and S.J. Billett. "Centred TVD schemes for hyperbolic conservation laws." *IMA Journal of Numerical Analysis* 20 (2000), pp. 47–79.
- [193] G. Tryggvason, B. Bunner, A. Esmaeeli, D. Juric, N. Al-Rawahi, W. Tauber, J. Han, S. Nas, and Y.-J. Jan. "A front-tracking method for the computations of multiphase flow." *Journal of computational physics* 169 (2001), pp. 708–759.
- [194] S.O. Unverdi and G. Tryggvason. "A front-tracking method for viscous, incompressible, multi-fluid flows." *Journal of Computational Physics* 100 (1992), pp. 25–37.
- [195] B. Van Leer. "Towards the ultimate conservative difference scheme. V. A second-order sequel to Godunov's method." *Journal of computational Physics* 32 (1979), pp. 101–136.
- [196] B. Van Leer. "Towards the ultimate conservative difference scheme." *Journal of Computational Physics* 135 (1997), pp. 229–248.
- [197] T. Waclawczyk. "On a relation between the volume of fluid, level-set and phase field interface models." *International Journal of Multiphase Flow* 97 (2017), pp. 60–77.
- [198] W.A. Wall, M.A. Frenzel, and C. Cyron. "Isogeometric structural shape optimization." *Computer methods in applied mechanics and engineering* 197 (2008), pp. 2976–2988.
- [199] Z. Wang and A.A. Oberai. "Spectral analysis of the dissipation of the residual-based variational multiscale method." *Computer Methods in Applied Mechanics and Engineering* 199 (2010), pp. 810–818.
- [200] C. H. Whiting and K. E. Jansen. "A stabilized finite element method for the incompressible Navier-Stokes equations using a hierarchical basis." *International Journal of Numerical Methods in Fluids* 35 (2001), pp. 93–116.

- [201] MW Williams, DB Kothe, and EG Puckett. "Accuracy and convergence of continuum surface tension models." *Fluid Dynamics at Interfaces, Cambridge University Press, Cambridge* (1998), pp. 294–305.
- [202] J. Yan, A. Korobenko, A.E. Tejada-Martínez, R. Golshan, and Y. Bazilevs. "A new variational multiscale formulation for stratified incompressible turbulent flows." *Computers & Fluids* 158 (2017), pp. 150–156.
- [203] J. Yan, S. Lin, Y. Bazilevs, and G.J. Wagner. "Isogeometric analysis of multi-phase flows with surface tension and with application to dynamics of rising bubbles." *Computers & Fluids* 179 (2019), pp. 777–789.
- [204] X. Zhang and C.-W. Shu. "A genuinely high order total variation diminishing scheme for one-dimensional scalar conservation laws." *SIAM Journal on Numerical Analysis* 48 (2010), pp. 772–795.
- [205] Y. Zhang, Y. Bazilevs, S. Goswami, C. L. Bajaj, and T.J.R. Hughes. "Patient-specific vascular NURBS modeling for isogeometric analysis of blood flow." *Computer methods in applied mechanics and engineering* 196.29-30 (2007), pp. 2943–2959.
- [206] V. Zingan, J.-L. Guermond, J. Morel, and B. Popov. "Implementation of the entropy viscosity method with the discontinuous Galerkin method." *Computer Methods in Applied Mechanics and Engineering* 253 (2013), pp. 479–490.

---

## SUMMARY

---

Numerical procedures and simulation techniques in science and engineering have progressed significantly during the last decades. The *finite element method* plays an important role in this development and has gained popularity in many fields including fluid mechanics. A recent finite element solution strategy is *isogeometric analysis*. Isogeometric analysis replaces the usual finite element basis functions by higher-order splines. This leads to significantly more accurate results and equips the numerical method with several desirable properties.

By naively applying the finite element isogeometric method one may obtain solutions that are seriously perturbed and are as such not physically relevant. The reason is often linked to the stability of the method; a finite element method is not *a priori* stable. The overall objective of this thesis is centered around this point. The aim is to develop numerical techniques that *inherit* the stability properties of the underlying physical system. In particular we are interested in finite element techniques that can be applied to free-surface flow simulations. Stability issues in free-surface flow computations may already appear in single-fluid flow problems. Other causes of instabilities are steep layers or discontinuities and instabilities arising from the numerical treatment of the interface that separates the fluids. This thesis addresses each of these topics.

Several *stabilized finite element methods* have been proposed to stabilize standard finite element solutions. These methodologies are provably stable for model problems and yield significantly improved solutions for more realistic problems. Despite that the proposed stabilized methods often show good behavior, they are generally not energetically stable for the more advanced problems.

In the first part of this dissertation we remedy this discrepancy for single-fluid flow with the aid of *variational multiscale analysis*. This technology provides a way to derive multiscale models and methods that satisfy certain objectives *ab initio*. The concept is to split the solution into large-scales that live on the mesh and small-scales of which their effect is modeled. As such, the two main ingredients of this paradigm are the projection operator that dictates the scale separation and the small-scale model. The idea is to select a projection operator that renders the multiscale model energy-dissipative. This design criterion leads to projectors that have a natural connection to the mathematical model, e.g., the Stokes projector for the incompressible Navier-Stokes equations. Concerning the small-scale model, we note that standard small-scale models have a static form which may seem a bit strange when dealing with dynamical problems. A dynamical form is arguably more suitable and permits to associate the eventual method with an energy that dissipates in time. Lastly, we note that it is convenient to have a divergence-conforming

discretization in order to obtain energy-dissipation. This is naturally provided by the isogeometric discretization.

Stabilized multiscale formulations produce oscillation-free solutions in absence of steep layers in the solution profiles. However, when sharp layers appear, these methods do not preclude oscillations. Even though these oscillations are generally small in size, they might render the simulation unfeasible; e.g., a negative density is nonphysical. Therefore, the stabilized finite element method is often augmented a term that introduces artificial diffusion. These mesh-dependent residual-based terms, known as *discontinuity capturing terms*, significantly improve the solution quality and reduce the oscillations. Unfortunately, these terms are known to have a large *ad hoc* character and a derivation is missing. As such, since the initiation of stabilized methods many different forms have been suggested. This gives a somewhat dissatisfying feeling and suggests room for improvement.

In the second part of this dissertation we provide the missing derivation using variational multiscale analysis. To do so, we first introduce a new concept called *variation entropy theory*. This theory introduces a new stability concept that grants a profound understanding of oscillations near sharp layers. Employing this stability concept in variational multiscale analysis naturally leads to a class of consistent discontinuity capturing mechanisms. The variation entropy theory may be understood as a way to localize the oscillations whereas variational multiscale analysis is the paradigm that tells us how to use this information to construct a discontinuity capturing method.

Free-surface simulations for maritime applications are often performed using the *level-set method*. In this method the interface separating the fluids is represented by the zero level-set of a higher-dimensional field. Level-set simulations generally use a diffuse-interface model that results from a mesh-dependent smearing of the interface. The key features of this approach are its ability to naturally deal with topological changes and the fact that the density and viscosity satisfy the (standard) maximum principle. A deficiency of the level-set method is that it may create artificial energy.

The last part of this thesis cures this imperfection for the incompressible Navier-Stokes equations with surface tension. We present a novel fully-discrete energy-dissipative level-set method. Standard finite element methods are too restrictive to inhibit energy-instabilities in a level-set formulation. *Functional entropy variables* provide the means to circumvent this limitation by discretizing a modified model. The modified model introduces a new variable that renders the model less restrictive. This, in combination with isogeometric analysis, allows to derive an energy-dissipative discretization. A new time-integration scheme, based on the perturbed midpoint rule, eventually yields the fully-discrete energy-dissipative method.

In summary, this dissertation sheds light on stability issues in existing isogeometric finite element methods for free-surface problems and offers resolutions for these issues.

---

## SAMENVATTING

---

Numerieke procedures en simulatietechnieken in wetenschap en techniek zijn de afgelopen decennia aanzienlijk verbeterd. De *eindige-elementenmethode* speelt een belangrijke rol in deze ontwikkeling en is populair geworden op vele gebieden, waaronder vloeistofmechanica. Een recente oplossingsmethode binnen de eindige elementen methoden is *isogeometrische analyse*. Isogeometrische analyse vervangt de gebruikelijke basis functies in eindige-elementen methoden door splines van hogere orde. Dit leidt tot aanzienlijk nauwkeurigere resultaten en voorziet de numerieke methode van verschillende gewenste eigenschappen.

Door naïef gebruik te maken van isogeometrische eindige elementen methoden kan men tot oplossingen komen die ernstig verstoord zijn (d.w.z. grote fouten bevatten), en als zodanig niet fysisch relevant zijn. De reden hangt vaak samen met de stabiliteit van de methode; een eindige-elementenmethode is niet *a priori* stabiel. De algemene doelstelling van dit proefschrift is rond dit punt gecentreerd. Het doel is numerieke technieken te ontwikkelen die de stabiliteitseigenschappen van het onderliggende fysische systeem *overnemen*. In het bijzonder zijn we geïnteresseerd in eindige-elemententechnieken die kunnen worden toegepast op simulaties van vrije oppervlakte stromingen. Stabiliteitsproblemen bij berekeningen van vrije oppervlakte stromen kunnen al optreden in het deel van de eindige-elementenmethode dat zich bezighoudt met een enkele vloeistofstroom. Andere oorzaken van instabiliteiten zijn steile lagen of discontinuïteiten en instabiliteiten als gevolg van de numerieke behandeling van het oppervlak dat de vloeistoffen scheidt. Dit proefschrift behandelt elk van deze onderwerpen.

Er zijn verschillende *gestabiliseerde eindige-elementenmethoden* voorgesteld om standaard eindige-elementoplossingen te stabiliseren. Deze methodologieën zijn aantoonbaar stabiel voor modelproblemen en leveren aanzienlijk verbeterde oplossingen op voor meer realistische problemen. Ondanks dat de voorgestelde gestabiliseerde methoden vaak goed gedrag vertonen, zijn ze over het algemeen niet energetisch stabiel voor de meer geavanceerde problemen.

In het eerste deel van dit proefschrift verhelpen we deze discrepantie voor stromingen met één vloeistof aan de hand van *variatierekening met meerdere schalen*. Deze technologie biedt een manier om modellen en methoden met meerdere schalen af te leiden die *ab initio* aan bepaalde doelstellingen voldoen. Het concept is om de oplossing te splitsen in grote schalen die op het rooster bestaan en kleine schalen waarvan het effect wordt gemodelleerd. Als zodanig zijn de twee belangrijkste ingrediënten van dit paradigma de projectie operatie, die de scheiding van de schalen bepaalt, en het model voor de kleine schalen. Het idee is om een projectie operatie te selecteren die het meerdere-schalen model energie dissipatief maakt. Dit criterium leidt tot

projecties die een natuurlijke connectie hebben met het wiskundige model; zo resulteren de onsamendrukbare Navier-Stokes-vergelijkingen in de Stokes-projectie. Wat betreft het kleinschalige model, merken we op dat standaard kleinschalige modellen een statische vorm hebben die minder goed past bij dynamische problemen. Een dynamische vorm is mogelijk geschikter en maakt het mogelijk de uiteindelijke methode te associëren met een energie die in de tijd afneemt. Als laatste merken we op dat een divergentie-conformerende discretizatie method handig is om energie-dissipatie te verkrijgen. Hier komt isogeometrische analyse goed van pas; het levert een natuurlijke manier om divergentievrije velden te construeren.

Gestabiliseerde formuleringen met meerdere schalen produceren oscillatievrije oplossingen in de afwezigheid van steile lagen in de oplossingsprofielen. Wanneer er echter scherpe lagen verschijnen, sluiten deze methoden oscillaties niet uit. Hoewel deze oscillaties over het algemeen klein zijn, kunnen ze de simulatie onbruikbaar maken; bijvoorbeeld een negatieve dichtheid is een niet-fysisch resultaat. Daarom wordt de gestabiliseerde eindige-elementenmethode vaak aangevuld met een term die kunstmatige diffusie introduceert. Deze rooster-afhankelijke, op residuen gebaseerde termen, bekend als *discontinuïteit afovangende termen*, verbeteren de kwaliteit van de oplossing aanzienlijk en verminderen de oscillaties. Helaas is bekend dat deze termen een groot *ad hoc* gedeelte hebben en dat een afleiding ervan ontbreekt. Als zodanig zijn er sinds het begin van gestabiliseerde methoden veel verschillende vormen voor deze termen voorgesteld. Dit geeft een wat onbevredigend gevoel en suggereert dat er ruimte voor verbetering is.

In het tweede deel van dit proefschrift geven we deze ontbrekende afleiding met behulp van variatierekening met meerdere schalen. Om dit te doen, introduceren we eerst een nieuw concept genaamd *variatie entropie theorie*. Deze theorie introduceert een nieuw stabiliteitsconcept dat een diepgaand begrip geeft van oscillaties nabij scherpe lagen. Het gebruik van dit stabiliteitsconcept in variatierekening met meerdere schalen leidt op een natuurlijke manier tot een klasse van consistente mechanismen voor het afvangen van de discontinuïteit. De variatie entropie theorie kan worden begrepen als een manier om de oscillaties te lokaliseren, terwijl variationele multischaalanalyse het paradigma is dat ons vertelt hoe we deze informatie kunnen gebruiken om een discontinuïteit afovangende methode te construeren.

Vrije oppervlakte simulaties voor maritieme toepassingen worden vaak uitgevoerd met de *niveauperzamelingen-methode*. Bij deze methode wordt het oppervlak dat de vloeistoffen scheidt weergegeven door de nul-niveauperzameling van een hogerdimensionaal veld. Numerieke simulaties aan de hand van deze techniek gebruiken over het algemeen een diffuus-interface model. Dit model resulteert wanneer het oppervlak rooster-afhankelijk wordt uitgesmeerd. De belangrijkste kenmerken van deze benadering zijn het vermogen om op natuurlijke wijze om te gaan met topologische veranderingen en het feit dat de dichtheid en viscositeit voldoen aan het maximumsprincipe. Een tekortkoming van de niveauperzamelingen-methode is dat deze kunstmatige energie kan creëren.

Het laatste deel van dit proefschrift verhelpt deze onvolkomenheid voor de onsamendrukbare Navier-Stokes-vergelijkingen met oppervlaktetensioning. We presenteren een nieuwe, volledig-discreet, energie-dissipatieve methode, die gebruikt maakt van de niveauperzamelingen-methode. Standaard eindige-elementenmethoden zijn te beperkend om energie instabiliteiten te voorkomen in een formulering die

gebruikt maakt van niveauverzamelingen. *Functionele entropie-variabelen* bieden de middelen om deze beperking te omzeilen door een aangepast model te discretiseren. Het gewijzigde model introduceert een nieuwe variabele die het model minder restrictief maakt. Dit, in combinatie met isogeometrische analyse, maakt het mogelijk een energie-dissipatieve discretisatie af te leiden. Een nieuw tijdintegratieschema, gebaseerd op de verstoorte middelpuntregel, levert uiteindelijk de volledig discrete energie-dissipatieve methode op.

Samenvattend belicht dit proefschrift stabiliteitsproblemen in bestaande isogeometrische eindige-elementenmethoden voor problemen met het vrije oppervlakken, en biedt het oplossingen die deze problemen verhelpen.



---

## PUBLICATIONS

---

### Journal articles

- *As part of this thesis:*

- (1). M.F.P. ten Eikelder and I. Akkerman, *Correct energy evolution of stabilized formulations: The relation between VMS, SUPG and GLS via dynamic orthogonal small-scales and isogeometric analysis. I: The convective-diffusive context*, *Comput. Meth. Appl. Mech. Engrg.* 331 (2018) 259-280.
- (2). M.F.P. ten Eikelder and I. Akkerman, *Correct energy evolution of stabilized formulations: The relation between VMS, SUPG and GLS via dynamic orthogonal small-scales and isogeometric analysis. II: The incompressible Navier-Stokes equations*, *Comput. Meth. Appl. Mech. Engrg.* 340 (2018) 1135-1159.
- (3). M.F.P. ten Eikelder and I. Akkerman, *Variation entropy: a continuous local generalization of the TVD property using entropy principles.*, *Comput. Meth. Appl. Mech. Engrg.* 355 (2019) 261-283.
- (4) M.F.P. ten Eikelder, Y. Bazilevs and I. Akkerman, *A theoretical framework for discontinuity capturing: Joining variational multiscale analysis and variation entropy theory*, *Comput. Meth. Appl. Mech. Engrg.* 359 (2020) 112664.
- (5) M.F.P. ten Eikelder and I. Akkerman, *An energy-dissipative level-set method for the incompressible Navier-Stokes equations with surface tension using functional entropy variables*, under review (2020).

- *Other journal articles:*

- (6). M.F.P. ten Eikelder, F. Daude, B. Koren and A.S. Tijsseling, *An acoustic-convective splitting based approach for the Kapila two-phase flow model*, *Journal Comput. Phys.* 331 (2017) 188-208.
- (7). I. Akkerman and M.F.P. ten Eikelder, *Toward free-surface flow simulations with correct energy evolution: an isogeometric level-set approach with monolithic time-integration*, *Comput. Fluids* 181 (2019) 77-89.
- (8). I. Akkerman, M.F.P. ten Eikelder, *Isogeometric analysis of linear free-surface potential flow*, *Ocean Engrg.* 201 (2020) 107114.

- (9). S.F. Stoter, M.F.P. ten Eikelder, F. De Prenter, I. Akkerman, E.H. van Brummelen, C.V. Verhoosel and D. Schillinger, *Unification of variational multiscale analysis and Nitsche's method, and a resulting boundary layer fine-scale model*, under review (2019).

#### Conferences attended with papers

- (1). M.F.P. ten Eikelder, F. Daude and B. Koren, A Lagrange-Projection-Like Numerical Scheme for Mixed Acoustic-Convective Two-Phase Flows, Proceedings of the ASME 2016 Pressure Vessels & Piping Conference - PVP 2016, Vancouver, Canada (July 2016), paper PVP2016-63539
- (2). M.F.P. ten Eikelder, I. Akkerman and R. van 't Veer, Correct energy behavior in two-fluid flow, Proceedings of the 20th Numerical Towing Tank Symposium - NuTTS'17, Wageningen, the Netherlands (October 2017)

#### Selected talks:

- (1). M.F.P. ten Eikelder and I. Akkerman, Ensured Energy Decay in Stabilized Methods using Dynamic Orthogonal Subgrid-Scales and Isogeometric Analysis, 19th International Conference on Finite Elements in Flow Problems - FEF 2017, Rome, Italy (April 7, 2017)
- (2). M.F.P. ten Eikelder and I. Akkerman, Correct energy behavior in stabilized finite element methods, Invited talk, CASA Colloquium, Eindhoven University of Technology, Eindhoven, The Netherlands (January 31, 2018)
- (3). M.F.P. ten Eikelder and I. Akkerman, Correct Energy Evolution of Stabilized Methods: Dynamic Orthogonal Small-Scales and Isogeometric Analysis, Isogeometric Analysis and Applications - IGAA 2018, Delft, The Netherlands (April 23, 2018)
- (4). M.F.P. ten Eikelder and I. Akkerman, Correct Energy Behavior of Stabilized Formulations, The 13th World Congress in Computational Mechanics - WCCM 2018, New York, USA (July 27, 2018)
- (5). M.F.P. ten Eikelder, Y. Bazilevs and I. Akkerman, Variation entropy, A theoretical VMS framework for discontinuity capturing Invited talk, School of Engineering, Brown university, Providence, USA (March 28, 2019)
- (6). M.F.P. ten Eikelder, I. Akkerman and Y. Bazilevs, The Variational Multiscale Method is the Underlying Theoretical Framework of Discontinuity Capturing, Finite Elements in Flow problems - FEF 2019, Chicago, USA (April 1, 2019)
- (7). M.F.P. ten Eikelder, Y. Bazilevs and I. Akkerman, A theoretical framework for discontinuity capturing, Merging VMS and variation entropy, Invited talk, Leibniz Universität Hannover, Hannover, Germany (April 17, 2019)

#### Posters:

- (1). M.F.P. ten Eikelder, F. Daude and B. Koren, A Lagrange-Projection-Like Numerical Scheme for Mixed Acoustic-Convective Two-Phase Flows, ASME 2016 Pressure Vessels & Piping Conference - PVP 2016, Vancouver, Canada (July 2016)

- (2). M.F.P. ten Eikelder and I. Akkerman, A Correct Energy Decay in Stabilized Methods using Dynamic Orthogonal Subgrid-Scales, The forty-first Woudschoten Conference 2016 Zeist, The Netherlands (October, 2016)



---

## ACKNOWLEDGMENTS

---

First of all I would like to express my thanks to my supervisor, **Ido Akkerman**, who has been without any doubt scientifically the most influential person. Thank you very much for pleasantly guiding me the entire journey. Thank you for introducing me to the beautiful world of stabilized finite element methods. Your creative mind and broad knowledge gave me a lot of material to think about. I believe that our collaboration has been and will remain very fruitful.

Second, I would like to thank my promotor **René Huijismans** for his absolute trust and for giving me all the freedom.

In addition I also want to thank some researchers from other universities. In particular, I wish to thank **Yuri Bazilevs** for the collaboration on the multiscale discontinuity capturing work. Thank you for arranging my visit to your research group at School of Engineering, Brown University. I want to thank **David Kamensky** for the discussions on the multiscale discontinuity capturing work and for providing assistance with the open source library TIGAR.

Furthermore, I would like to thank **Stein Stoter, Frits de Prenter, Dominik Schillinger, Harald van Brummelen** and **Clemens Verhoosel** for the collaboration on the multiscale boundary layer work (which is not part of this thesis). Stein and Dominik, thank you for hosting my joint visit with Frits to the Institut für Baumechanik und Numerische Mechanik, Leibniz Universität Hannover. It definitely led to a better understanding of variational multiscale analysis.

Besides, I want to thank my colleagues and friends at Delft for making my PhD journey a pleasant time. **Nikos, Bart, Themis, Jun, Xiaobo, Hugo, Martin, Abdu, Maarten** and others, thank you for our tea/coffee breaks, the nice dinners, activities and trips, and thank you **Reinier** for co-organizing the weekly group lectures.

Thanks goes to my other friends, running friends and in particular my running coaches **Koos** and **Eric** for providing most enjoyable and tough running trainings. It gave me the opportunity to distract me, in a positive way, from my research work.

Lastly, I wish to express my thanks to **my family** for their unconditional and most appreciated support. I could not have asked for more. Most importantly, **Wenhua**, thank you for always being there for me and for bringing a smile to my face.

*Marco ten Eikelder  
Delft, May 2020*



---

## CURRICULUM VITAE

---

Marco ten Eikelder was born in Eindhoven on July 31, 1992. He attended the Lorentz Casimir Lyceum in Eindhoven from 2004 until 2010. He obtained his Bachelor and Master of Science (both with honors) in applied mathematics at the Eindhoven University of Technology in 2013 and 2015 respectively. In 2013 he was awarded a scholarship for his Master's program consisting of two years tuition fee. His Master's program included a two-month research visit to Indian Institute of Technology, Kanpur, India. In addition he completed a master honors track program. He performed his master thesis research at Électricité de France (eDF) in Paris where he worked on the simulation of compressible two-phase flows.

From 2016 to 2020 he performed his PhD research under the supervision of Ido Akkerman on stabilized finite element isogeometric methods. The results of his PhD research are included in this thesis.