Title
Sparsity-Inducing Methods in Imaging Sciences and Partial Differential Equations

Author
Tran, Giang Thi Tra

Publication Date
2015

Peer reviewed|Thesis/dissertation

University of California
Los Angeles

Sparsity-Inducing Methods in Imaging Sciences and Partial Differential Equations

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Mathematics

by

Giang Thi Tra Tran

2015

Abstract of the Dissertation

Sparsity-Inducing Methods in Imaging Sciences and Partial Differential Equations

by

Giang Thi Tra Tran

Doctor of Philosophy in Mathematics

University of California, Los Angeles, 2015

Professor Stanley J. Osher, Chair

Sparsity has played a central role in many fields of applied mathematics such as signal processing, image processing, compressed sensing, and optimization. Theoretically, sparse solutions are of interested in these fields because they can be recovered exactly from ill-posed inverse problems. Numerically, methods for computing sparse solutions have fast and efficient implementations, making them extremely practical. In terms of modeling, sparsity is promoted through the addition of an L^1 norm (or related quantity) as a constraint or penalty in a variational model. This methodology is also related to various properties of solutions of partial differential equations (PDEs) including compact support sets for free boundary problems and sparse representation of the solution space. Recently, sparsity-inducing methods used in image processing and compressed sensing have been applied to computational PDEs and applied harmonic analysis.

Part I of the thesis focuses on the construction of efficient numerical schemes for PDEs with sparse structures. From the early theoretical work on variational inequalities, it was showed that PDEs with L^1 subdifferential terms have unique solutions with compact supports which can be considered to be sparse in the discrete sense. We considered an elliptic PDE derived from a variational principle or a parabolic PDE associated with the gradient flow of a convex functional. In

order to compute the solutions, fast computational schemes were introduced to solve the resulting minimization problem. Those methods can handle the multi-valued nature of the sub-gradient $\partial\|u\|_{L^1}$, which involve the proximal operator of the L^1 term. Although these methods are popular in imaging and data science, their applications in PDEs are limited.

The L^1-based methodology was also introduced to elliptic obstacle problems and related free boundary problems such as Hele-Shaw flow, two-phase membrane, and divisible sandpile. Our numerical methods are based on a reformulation of those PDEs in terms of L^1-like penalties on the associated variational problems. One advantage of the proposed methods is that the free boundary inherent in the obstacle problem arises naturally in the energy minimization without any need for problem specific or complicated discretization. Moreover, the numerical solution can be computed using fast and simple algorithms.

Part II of the thesis focuses on data decomposition methods to extract important features or recover the intrinsic properties from the original data. From the variational approaches, a unifying retinex framework was developed to decompose an image into two components with certain sparsity and fidelity priors. The unified formulation connects many retinex implementations into one model. Moreover, new retinex applications were introduced within a single framework which includes shadow detection, cartoon-texture decomposition, nonuniform illumination correction and color contrast enhancement.

The dissertation of Giang Thi Tra Tran is approved.

Christopher R. Anderson

Aichi Chien

Luminita A. Vese

Alan L. Yuille

Stanley J. Osher, Committee Chair

University of California, Los Angeles

2015

To my family.

Table of Contents

I Partial Differential Equations with Sparse Structures 1

1 Introduction to PDEs with Sparse Structures 2

2 PDEs with Compressed Solutions 6

2.1 Problem Formulation 6

2.2 Various Properties 7

2.3 Numerical Implementation 17

2.4 Computational Simulations 20

3 An L^1 Penalty Method for Obstacle Problems 37

3.1 Motivation 37

3.2 Obstacle Problem 39

3.3 Free Boundary Problem 41

3.4 Numerical Method 45

3.5 Computational Simulations 51

II Non-local Operators and Retinex 64

4 A Short Review of Retinex Implementations 65

5 Non-local Differential Operators 76

5.1 Non-local Differential Operators 76

5.2 Sparse Quasi-gradients 80

5.3 Computing the Weights 87

5.4 Closing the Gap Between Kernel and Variational Retinex 91

6 Non-local Retinex Model 93

6.1 Non-local Retinex Model 93

6.2 Numerical Optimization 95

6.3 Results I: Relations to Existing Models 100

6.4 Results II: New Perspectives 109

6.5 Adaptive Thresholding with Texture and Color 115

References 119

List of Figures

2.1 Numerical solution of the modified heat equation starting with an initial traveling wave profile 21

2.2 Convergence analysis of the modified heat equation, using the L^1, L^2, and L^∞ norms in space and L^∞ norm in time 22

2.3 Numerical solution starting with an initial traveling wave profile perturbed by uniformly random noise 23

2.4 A 1D simulation of the modified heat equation 24

2.5 Solutions of a modified 2D heat equation 29

2.6 The corresponding support set of the modified 2D heat equation . 30

2.7 Solution of the standard normalized graph Laplacian 31

2.8 Solution of a graph diffusion with the subgradient term 32

2.9 Our numerical approximation to the Signum-Gordon equation . . 32

2.10 The time evolution of an oscillatory compact soliton-like structure 33

2.11 A two-region sandpile problem 34

2.12 Iterative evolution of a sandpile problem I 34

2.13 Iterative evolution of a sandpile problem II 35

2.14 Iterative evolution of a sandpile problem III 36

2.15 Numerical solution of a sandpile problem 36

3.1 Numerical solutions of 1D obstacle problems 52

3.2 Errors between the numerical solutions and the analytic solutions of 1D obstacle problems versus number of iterations 52

3.3 Numerical solution of a 2D obstacle problem and the difference with the analytic solution 53

3.4 Numerical solution and its level curves of a 2D obstacle problem . 54
3.5 Level curves of an obstacle and our corresponding numerical solution 55
3.6 Error between the analytic and numerical solutions computed by our method and method from [93] vs. the number of iterations . . 56
3.7 Numerical solution of a nonlinear 1D obstacle problem 58
3.8 Numerical solution and the error between the numerical and the analytic solution of a 1D two-phase membrane 59
3.9 Numerical solution of a 1D two-phase membrane and the corresponding free boundary point 60
3.10 Numerical solution of a 2D two-phase membrane and the boundaries between regions 61
3.11 Free boundary of Hele-Shaw problems 62

4.1 Mondrian illusion 66
4.2 Relative and average reflectance 69

6.1 Logvinenko illusion and different Retinex decompositions 105
6.2 Unifying non-local Retinex model results 107
6.3 HSV color Retinex 108
6.4 Logvinenko illusion and new L^0-based Retinex decompositions . . 110
6.5 Shadow detection results of our non-local Retinex model 112
6.6 Shadow removal results of our proposed model 113
6.7 Cartoon-texture decomposition results 114
6.8 Color contrast enhancement result using adaptive thresholding . . 117

List of Tables

2.1 Error between our numerical and the analytic solutions of the Signum-Gordon Equation 26

3.1 Errors between the analytic and numerical solutions computed by our method and the method found in [93] for a 2D obstacle problem 57

3.2 The CPU times and number of iterations for our method and the method found in [93] 57

3.3 The error between the radius of the free boundary of our numerical solution and the exact radius of a Hele-Shaw problem 62

6.1 Retinex models' correspondences 101

Acknowledgments

First and foremost, I would like to thank my academic advisor, Stanley Osher, for his tremendous help, invaluable mathematical and professional guidance, and generous support throughout my doctoral program. It has been a great pleasure to work with him. His approach of doing research has a great influence on me. I want to thank him for being the best advisor and I am truly grateful.

I would also like to express my gratitude to the rest of my thesis committee Christopher Anderson, Aichi Chien, Luminita Vese and Alan Yuille for their time and suggestions. I thank Luminita Vese for her guidance and encouragement since the very beginning of my graduate studies, and for introducing me to my collaborators Yonggang Shi and Ivo Dinov at Laboratory of Neuro Imaging. I thank Christopher Anderson for teaching me numerical analysis and scientific computing. I am thankful to Aichi Chien for her consideration and her helpful conversations about medical imaging. I would like to thank Alan Yuille for his interest in my work.

I would also like to acknowledge my collaborators Russel Caflisch, William Feldman, Jerome Gilles, Hayden Schaeffer, Yonggang Shi, and Dominique Zosso. I am thankful to Dominique Zosso for helping me with coding and writing in my first applied math project, to Yonggang Shi for introducing me to very interesting topics in brain imaging, to Jerome Gilles for helpful conversations on wavelets and texture models, to Russel Caflisch for his invaluable advice on partial differential equations, to William Feldman for fruitful discussions on free boundary problems. I especially would like to thank my friend and collaborator, Hayden Schaeffer, for his consistent help, support and encouragement.

In addition, I would like to thank my classmates, officemates and friends for their support and for making my PhD life so colorful and joyful. I am also grateful for all the staffs at UCLA Math, IPAM and VEF for helping me with numerous

administrative matters.

Finally, I would like to thank my parents for their endless love and unconditional support. I also thank to my brother, Hieu Tran, for always keeping me in good spirits. Without their encouragement, I would not be going this far.

The research in this dissertation was supported by UC Lab Fees Award ID# 12-LR-23660, UC Lab 443948-B1-69763, ONR N00014-10-10221, ONR N00014-14-1-0444 and N000141110719, and Keck Funds 449041-PW-58414.

Vita

2006	B.S. in Mathematics Hanoi National University of Education, Vietnam
2009	M.S. in Mathematics University of Strasbourg, Strasbourg, France
2009–2015	Teaching and Research Assistant, Department of Mathematics University of California, Los Angeles
2009–2015	Vietnamese Education Foundation Fellowship
2011	M.A. in Mathematics University of California, Los Angeles
2014–2015	Beckenbach Dissertation Year Fellowship

Publications

G. Tran, H. Schaeffer, W. M. Feldman, and S. Osher. "An L^1 Penalty Method for General Obstacle Problems". SIAM Journal on Applied Mathematics, to appear.

R. Caflisch, S. Osher, H. Schaeffer, and G. Tran. "PDEs with Compressed Solutions". Communications in Mathematical Sciences, to appear.

D. Zosso, G. Tran, and S. Osher. "Non-local Retinex - A Unifying Framework

and Beyond". SIAM Journal on Imaging Sciences, 8(2), 787826.

D. Zosso, G. Tran, and S. Osher. "A Unifying Retinex Model Based on Non-local Differential Operators". S&T/SPIE Electronic Imaging: Computational Imaging XI, San Francisco, USA, 2013

Part I

Partial Differential Equations with Sparse Structures

CHAPTER 1

Introduction to PDEs with Sparse Structures

Sparsity has played a central role in recent developments in fields such as imaging science, data analysis and optimization. Examples include compressed sensing [32, 42], phase retrieval [30], robust principal component analysis [31, 41, 126], as well as many others. A key step in these examples is the use of an L^1-norm (or related quantity) as a constraint or penalty term in a variational formulation.

The use of sparse techniques in physical sciences and partial differential equations (PDEs) has been studied recently. For example, in [134, 97], L^1-regularized least squares were used to sparsely approximate the Fourier coefficients of the multiscale oscillatory solutions. In [109, 118, 119], new sets of spatially localized orthonormal functions were introduced to solve problems in quantum mechanics. Sparse solutions with respect to low-rank libraries are used in modeling and approximating dynamical systems, see for example [24].

Motivated by these works and by the early theoretical framework established in multivalued PDEs [20, 21, 22, 23], we investigated PDEs with L^1 subdifferential terms. We considered elliptic PDEs coming from a variational principle or parabolic PDEs coming from a gradient flow of a convex functional. Our starting point is the convex functional:

$$E(u) = \int \frac{1}{2}(\nabla u) \cdot M(\nabla u) - uf + \gamma|u|dx, \tag{1.1}$$

where $\gamma \geq 0$, $M = M(x)$ is a symmetric, positive definite matrix as a function of x, and $f = f(x)$ or $f = f(x,t)$ will be a specified function depending on x or (x,t). Define the partial differential operator $Au = -\nabla \cdot (M\nabla u)$. Minimization

of $E(u)$ for $f = f(x)$ leads to the following elliptic PDE

$$Au = f - \gamma p(u), \tag{1.2}$$

and gradient descent $\partial_t u = -\partial_u E(u)$, starting from initial data $g(x)$, leads to the following parabolic PDE

$$\begin{aligned} u_t + Au &= f - \gamma p(u) \\ u(x,0) &= g(x), \end{aligned} \tag{1.3}$$

in which $p(u)$ is a subgradient of $\|u\|_{L^1}$, *i.e.*, $\|v\|_1 \geq \|u\|_1 + \langle v - u, p(u)\rangle$, for any u and v, where $\langle\, , \rangle$ denotes the (L^1, L^∞) pairing.

In either case, the L^1 term in the convex functional leads to a subgradient term in the PDE. Fortunately, the subgradient term has a simple explicit form, so that the PDEs are amenable to analysis and computation. Moreover, by adding the subdifferential of L^1 to some typical PDEs, such as the heat equation and the graph diffusion, it can be shown that the solutions are compactly supported [22]. In addition, we have shown several theoretically insights on the behavior of the solutions, such as L^1-contraction, total variation diminishing, entropy condition, regularity and estimation of the support size. We also demonstrated that the subdifferential term can be related to some physical equations, such as divisible sandpile and signum-Gordon. Those properties are discussed at the beginning of Chapter 2.

Our main focus in Chapter 2 is to present fast computational schemes for these modified PDEs, where we apply the proximal gradient and Douglas-Rachford splitting algorithm to solve the resulting minimization problem. Both methods can handle the multivalued nature of the subgradient $\partial\|u\|_{L^1}$, which involve the proximal operator of the L^1 term. Although these methods are popular in imaging and data science, their applications in PDEs are limited.

On the other hand, L^1 optimization is also deeply connected to some physical properties of solutions to PDE. In the case where the PDE emits solutions which

are compactly supported (possible after a linear transformation), the underlying equation should have a discontinuous structure which is directly related to the L^1 subdifferential. Therefore, the appearance of an L^1 term in the energy formulation is expected. Indeed, many physical free boundary problems have compactly supported solutions. In Chapter 3, we showed that certain free boundary problems can be represented in an unconstrained variational form with the addition of an L^1 term.

The classical obstacle problem models the equilibrium state of an elastic membrane stretched atop of a physical obstacle with fixed boundary conditions. This has a direct mathematical interpretation as an energy minimization (*i.e.* the classical elastic energy of the membrane) with the addition of a constraint (*i.e.* the solutions are bounded below by the obstacle).

The original theory for obstacle problems centered around minimizations of the form:

$$\min_{u \in K} a(u,u) - \langle f, u \rangle \,,$$

where $a(-,-)$ is a bounded and coercive bilinear form on some Sobolev space V, $K = \{v \geq \varphi\}$ for some smooth φ, and $<,>$ is the standard L^2 inner product [131, 78, 51]. This minimization problem is equivalent to the problem of finding a $u \in K$ satisfying the variational inequality:

$$a(u, v-u) \;\geq\; \langle f, v-u \rangle \qquad \text{for all} \quad v \in K,$$

which can be considered as the Euler-Langrange equation for the constrained problem.

Over the years, there have been many numerical methods for solving various types of obstacle problems. A list of numerical methods for variational inequalities are discussed in [54, 150]. A vast majority of those algorithms use the weak variational inequality characterization to approximate the solutions numerically, see for example [64, 8, 144, 7, 65, 80, 81, 153, 38]. It is also possible to solve the

obstacle problem using the complementarity conditions [131, 61]. With the help of the level set method [114], the authors of [98] construct a method to locate the contact set of the obstacle problem, then find the solution to the obstacle problem directly without the need of the variational inequalities.

Alternative approaches use the constrained optimization formulation to construct appropriate algorithms, see for example [62, 135]. In those papers, a penalty method or a regularization formulation were proposed to encourage solutions to satisfy the constraint. However, the penalty method is not exact and the existence of solutions relies on regularizing the functional due to the lack of differentiability.

For the two-phase membrane problem, which is a double obstacle problem, the author of [17] introduces two algorithms. In the first method, the solution is split into two parts, a positive and a negative part, which results in a coupled system of PDE with matching conditions. In the second method, a finite element approach is done on a regularized version of the problem so as to avoid the non-differentiablity of the L^1-like functions.

In Chapter 3, we will use an L^1-like penalty on the original variational form of the obstacle problem:

$$\min_u a(u,u) - \langle f, u\rangle + \mu \int \max(\varphi - u, 0)\, dx,$$

which is an exact penalty for sufficiently large $\mu > 0$, see [50, 101]. For more details on general theoretical results including regularity of solutions for the obstacle and related free boundary problems, see for example, [28, 27, 26]. In particular, we provide some theoretical results on solutions of L^1 regularized variational methods to the solutions of obstacle problems with zero obstacle. We derive bounds on the exactness of the penalty formulation as well as construct a fast and simple algorithm to solve the non-differentiable unconstrained problem. Unlike other penalty methods, we do not require the penalty parameter to go to ∞ (for sufficiently smooth obstacles) and no regularization of the penalty is required.

CHAPTER 2

PDEs with Compressed Solutions

In this chapter, we first provide the general formulation of the problem. Then we review known results and present various properties of solutions to the modified PDEs. The numerical implementation and simulations are presented in Sections 2.3 and 2.4. This chapter was taken with slight modification from [29].

2.1 Problem Formulation

The problem we consider is to numerically solve the following PDE:

$$u_t + Au = f - \gamma p(u)$$
$$u(x,0) = g(x),$$

and to verify theoretical results. The difficulty with such equations is the multi-valued nature of the subgradient term. Fortunately for this type of equation, we can explicitly identified the subgradient as

$$p(u) = \begin{cases} \text{sign}(u) & \text{if } |u| > 0 \\ \underset{|q|\leq 1}{\text{argmin}} |f - \gamma q| & \text{if } u = 0. \end{cases} \tag{2.1}$$

Note that if $u = 0$ and $|f(x)| \leq \gamma$, then $p = f(x)/\gamma$. Equation 2.1 was proved in general in [33, 20]. It can be shown directly from Equations (1.2) and (1.3), as follows. For $u = 0$ in an open set, the left side of the equations is 0 so that $f(x) - \gamma p(u) = 0$, which is only possible if $f(x) \leq \gamma$ and $p(u) = f(x)/\gamma$. The value of $p(u)$ on a lower dimensional set does not matter, since the value of the forcing

terms on a lower dimensional set does not affect the solution u of the differential equations.

2.2 Various Properties

In this section we recall the established existence theory for the elliptic equation (1.2) and the parabolic equation (1.3), and provide some further insights to the behavior of solutions.

Review of Theoretical Results

Equation (1.2) is related to the general class of elliptic equation:

$$-\Delta u = F(u),$$

where F contains a discontinuous component. The existence and uniqueness of the solution u are studied in [79, 75, 40]. Solutions also satisfy the standard maximum and comparison principles given the correct sign of F. The solutions are compactly supported in both the elliptic and parabolic case, under some additional conditions [22, 23]. For the parabolic equation, the solutions are Lipschitz continuous and right differentiable in time. Furthermore, solutions exhibit finite speed of propagation [23]. More precisely, let $S(t)$ be the support set of $u(x,t)$, then for small times t:

- if $u(x,0)$ does not vanish on $\partial S(0)$, then

$$S(t) \subset S(0) + B(c\sqrt{t\log(t)}),$$

- if $u(x,0)$ and $\nabla u(x,0)$ vanishes on $\partial S(0)$, then

$$S(t) \subset S(0) + B(c\sqrt{t}),$$

where $B(r)$ is the ball of radius r centered at the origin. In a simple case, we can construct the exact bounds in order to verify the convergence of the method to a known solution.

At a number of places in the manuscript, we will simplify the presentation by assuming that $x \in \mathbb{R}^1$ and that $M = 1$, so that the elliptic PDE (1.2) becomes Laplace's equation with nonlinear forcing:

$$u_{xx} = -f + \gamma p(u), \tag{2.2}$$

and the parabolic PDE (1.3) becomes the heat equation with nonlinear forcing:

$$u_t - u_{xx} = f - \gamma p(u). \tag{2.3}$$

Support Size

Since it is known that the support is compact, we would like to estimate its size. First, observe that if $\gamma \geq \max |f|$, then the unique solution of Equation (1.2). is $u \equiv 0$. Indeed, if $u = 0$, since $\dfrac{f}{\gamma} \in [-1, 1]$, we can choose $p(u) = \dfrac{f}{\gamma}$ and Equation (1.2) is satisfied.

Now, take $\mathcal{S} = \text{supp}(u)$ and integrating both sides of Equation (1.2) gives us

$$\int_{\partial \mathcal{S}} M \nabla u \cdot N ds = -\int_{\mathcal{S}} f dx + \gamma \, \text{sign}(u) |\mathcal{S}|,$$

where N is the normal. On $\partial \mathcal{S}$, since $u = 0$, we have $\nabla u = \omega N$, for some scalar function $\omega : \partial \mathcal{S} \to \mathbb{R}$. In addition, since $u > 0$ in $\mathcal{S}$ and $u = 0$ on $\partial \mathcal{S}$, we have $\omega \leq 0$. Lastly, by assumption M is positive definite so the left hand side of Equation (A.3) is non-positive:

$$\int_{\partial \mathcal{S}} M \nabla u \cdot N ds = \int_{\partial \mathcal{S}} \omega M N \cdot N ds \leq 0$$

Therefore,

$$|\text{supp}(u)| \leq \gamma^{-1} \int_{\text{supp}(u)} |f| dx. \tag{2.4}$$

A slight modification of (2.4) shows that for any nonnegative α and β with $\alpha+\beta = 1$, we have

$$|\text{supp}(u)| \leq (\alpha\gamma)^{-1} \int (|f| - \beta\gamma)_+ dx. \tag{2.5}$$

In this inequality, the superscript + denotes the positive part; *i.e.*, $(x)_+ = \max(x, 0)$.

For the parabolic case, define the time dependent support set $\mathcal{S}(t) := \text{supp}(u(x,t))$. Differentiating the integral of u over $\mathcal{S}(t)$ and using the boundary conditions (*i.e.*, $u = 0$ on $\partial\mathcal{S}(t)$) yields:

$$\frac{d}{dt}\int_{\mathcal{S}(t)} u(x,t)dx = \int_{\mathcal{S}(t)} u_t dx = \int_{\mathcal{S}(t)} \nabla \cdot M\nabla u + f - \gamma p(u)\ dx.$$

Because of the divergence theorem and the fact that M is positive definite, we have

$$\frac{d}{dt}\int_{\mathcal{S}(t)} |u(x,t)|dx \leq \int_{\mathcal{S}(t)} |f|dx - \gamma|\mathcal{S}(t)|.$$

Integrating the expression in time yields the following bound on the support size:

$$|\,\text{supp}_{(x,t)}\, u(x,t)| \leq \int_{\mathcal{S}(t)} |g|dx + \iint_{\mathcal{S}(t)} |f|dx\, dt.$$

Also, similar to the elliptic case, we have

$$|\,\text{supp}_{(x,t)}\, u(x,t)| \leq (\alpha\gamma)^{-1} \left(\int |g|dx + \iint (|f| - \beta\gamma)_+ dx\, dt \right), \tag{2.6}$$

for any nonnegative α and β with $\alpha + \beta = 1$.

L^1 Contraction and Total Variation Diminishing

Let u and v be solutions of Equation (2.3) with initial data $g(x)$ and $h(x)$, respectively. First, note that for any subgradient p of a convex functional, we have

$$\text{sign}(u - v)(p(u) - p(v)) \geq 0. \tag{2.7}$$

We wish to show that the solutions are L^1 contractive and TVD by computing the following:

$$\begin{aligned}\frac{d}{dt}||u-v||_{L^1} &= \frac{d}{dt}\int\limits_{|u-v|>0} |u-v|dx \\ &= \int\limits_{|u-v|>0} \operatorname{sign}(u-v)(u_t-v_t)dx \\ &= \int\limits_{|u-v|>0} \operatorname{sign}(u-v)(u-v)_{xx} - \gamma \operatorname{sign}(u-v)(p(u)-p(v))dx.\end{aligned}$$

The first term is zero by the divergence theorem and the second term is negative by Equation (2.7), so we have $\frac{d}{dt}||u-v||_{L^1} \leq 0$, and thus the modified PDE is an L^1 contraction. Moreover, if we take $h(x) = g(x+\delta)$ for any $\delta > 0$ we have

$$\frac{d}{dt}\|u(x,t) - u(x+\delta,t)\|_{L^1} \leq 0.$$

Dividing the equation above by δ and taking the supremum over all δ, the following inequality holds:

$$\frac{d}{dt}||u||_{TV} \leq 0.$$

Therefore, Equation (2.3) is TVD.

Entropy Condition

The L^1 contraction and TVD results are directly analogous to those that are obtained by solving the viscosity regularized nonlinear conservation laws:

$$w_t^\epsilon = \epsilon\, w_{xx}^\epsilon - f(w^\epsilon)_x,$$

for $\epsilon > 0$. Then by letting $\epsilon \to 0$, one recovers the unique inviscid limit, see [87].

We can also easily obtain an "entropy inequality" in the same spirit. Consider the scaled modified heat equation:

$$u_t = \epsilon\, u_{xx} - \gamma p(u). \tag{2.8}$$

We deliberately put an ϵ in front of the diffusion term to emphasize the similarities to the theory of scalar conservation laws. The following argument holds in more general cases.

Let $K(u)$ be a convex function of u with subgradient $q(u)$. Multiplying Equation (2.8) by the subgradient (as in [87]) yields:

$$\frac{d}{dt}K(u) \leq \epsilon \frac{d^2}{dx^2}K(u) - \gamma q(u)p(u). \tag{2.9}$$

For example, if $K(u) = |u|$, then whenever $u \neq 0$, we have

$$|u|_t \leq \epsilon |u|_{xx} - \gamma. \tag{2.10}$$

We integrate Equation (2.9) over the region $\mathcal{S}(t)$, the support set of $u(x,t)$ defined in Section 2.2, to get

$$\frac{d}{dt}\int_{\mathcal{S}(t)} K(u)\,dx \leq -\gamma \int_{\mathcal{S}(t)} q(u)p(u)\,dx, \tag{2.11}$$

since the spatial gradient is zero along the boundary. By choosing $K(u) = \frac{1}{a}|u|^a$ for $a \geq 1$, Equation (2.11) provides L^a estimates of the solutions. Furthermore, if $K(u) = (u-c)_+$ for $c > 0$, then

$$\frac{d}{dt}\int_{\mathcal{S}_c^+(t)} (u-c)_+\,dx \leq -\gamma |\mathcal{S}_c^+(t)|, \tag{2.12}$$

where $\mathcal{S}_c^+(t)$ is the set of x for which $u(x) > c$.

Regularity

We can show that the solutions of the Laplace's equation (2.2) and of the heat equation (2.3) are smooth. Let Ω_+, Ω_-, and Ω_0 denote the sets $\{u > 0\}$, $\{u < 0\}$ and $\{u = 0\}$, respectively. Then the solution u of the Laplace's equation (2.2) can be represented by

$$u(x) = \int_{\Omega_+} G(x-y)(f(y) - \gamma)dy + \int_{\Omega_-} G(x-y)(f(y) + \gamma)dy, \tag{2.13}$$

and the solution of the heat equation (2.3) can be written as

$$u(x,t) = \int G(x-y,t)g(y)dy + \int_0^t \int_{\Omega_+(s)} G(x-y,t-s)(f(y)-\gamma)dyds + \int_0^t \int_{\Omega_-(s)} G(x-y,t-s)(f(y)+\gamma)dyds, \tag{2.14}$$

in which the Green's function $G(x,t)$ for the heat equation and the Green's function $G(x)$ for the Laplace's equation are given by

$$\begin{aligned} G(x) &= |x|/2, \\ G(x,t) &= (4\pi t)^{-1/2}\exp(-x^2/4t). \end{aligned} \tag{2.15}$$

From these formulas, if f is continuous, then one can see that u is $C^2(x)$ and $C^1(t)$ away from $u=0$ and that u is $C^1(x)$ everywhere.

Traveling Wave

To demonstrate finite speed of propagation, consider the 1D-traveling wave solution $u(x,t) = v(s)$ for $s = x - \sigma t$, of the Equation (2.3) with no forcing term. To be specific, we will assume that $v(s) \geq 0$ for $s \geq 0$ and $v(s) = 0$ for $s \leq 0$. We see that v must satisfy the ODE

$$v_{ss} + \sigma v_s - \gamma = 0, \tag{2.16}$$

subject to the conditions

$$v(0) = v'(0) = 0.$$

The general solution of Equation (2.16) is

$$v(s) = \begin{cases} \frac{\gamma}{\sigma}s + c_1 e^{-\sigma s} + c_2, & s \geq 0 \\ 0, \quad \text{otherwise.} \end{cases} \tag{2.17}$$

The boundary conditions imply

$$c_1 = -c_2 = \frac{\gamma}{\sigma^2},$$

so that the traveling wave solution of Equation (2.3) is

$$u(x,t) = \begin{cases} \frac{\gamma}{\sigma}(x-\sigma t) + \frac{\gamma}{\sigma^2}\left(e^{-\sigma(x-\sigma t)} - 1\right), & x \geq \sigma t \\ 0, \quad \text{otherwise.} \end{cases}$$

We see that in this case we have one sided support.

Remark 1. *This traveling wave solution is used as a reference solution to compute the error for our numerical scheme (see Section 2.4). Also, the simple analytic form shows that solutions with non-trivial support sets are easy to find in the modified PDEs.*

An Exact Solution

We construct the exact solution of Equation (2.2) with nonnegative force $f = (1+x^2)^{-3/2}$ and $\gamma \in [0,1]$. The exact solution is given explicitly by:

$$u = \begin{cases} -(1+x^2)^{1/2} + \frac{1}{2}\gamma x^2 + c, & |x| \leq a \\ 0, \quad |x| > a. \end{cases}$$

where,

$$c = \frac{\gamma + \gamma^{-1}}{2}, \quad a = \sqrt{\gamma^{-2} - 1}.$$

The boundary value a and constant c are determined so that $u(\pm a) = u_x(\pm a) = 0$. At the boundary of the support, $f(\pm a) = \gamma^3 < \gamma$. These results show that the solution is nonnegative for nonnegative f, and that having $|f(x)| \leq \gamma$ does not imply $p(u(x)) = \frac{f(x)}{\gamma}$.

Flux Condition

We provide here the natural boundary condition to Equation (2.3).

Lemma 1. *Let $u(x,t) \in C^0(C^1(\mathbb{R}); (0,T))$ and $u_t \in L^\infty(C^1(\mathbb{R}); (0,T))$ be a solution to*

$$u_t - u_{xx} = h(x,t,\gamma). \tag{2.18}$$

Assume that there exists a positive valued function $a \in C^1(0,T)$ *such that* $h = 0$ *for* $|x| > a(t)$*, the initial data* $g = 0$ *for* $|x| > a(0)$*, and the exterior mass,*

$$m(t) = \int_{a(t)}^{\infty} u(x,t)dx,$$

is conserved, then $u(a(t),t) = 0$ *and* $u_x(a(t),t) = 0$*.*

To derive this condition, consider the heat equation (2.18). Differentiate the one sided mass in time yields:

$$\begin{aligned}
\frac{dm}{dt} &= -u(a(t),t)a'(t) + \int_{a(t)}^{\infty} u_t(x,t)dx \\
&= -u(a(t),t)a'(t) + \int_{a(t)}^{\infty} u_{xx}(x,t)dx \\
&= -u(a(t),t)a'(t) - u_x(a(t),t) \\
&= -F(t),
\end{aligned}$$

in which F is the flux across the moving boundary $x = a(t)$.

We now can see that if the flux across a moving boundary $x = a(t)$ is zero (*i.e.* the mass is conserved), we have

$$F(t) = u(a(t),t)a'(t) + u_x(a(t),t) = 0. \tag{2.19}$$

This is the natural boundary condition for this problem. In the time-dependent region $\mathcal{F} = \{(x,t) : x > a(t)\}$, the initial data g, force h and and incoming flux F are all zero, so that the solution is identically zero. In particular, $u = u_x = 0$ on $x = \pm a(t)$.

A Free Boundary Formula

In 1D, consider the following equation

$$u_t - u_{xx} = \begin{cases} f(x) - \gamma, & |x| < a(t) \\ 0, & |x| > a(t) \end{cases} \tag{2.20}$$
$$u(x,0) = 0,$$

which is equivalent to Equation (1.3) where the support set is parametrized. For simplicity assume that $f(x) = f(|x|)$ and f is a radially decreasing function with $f(|x|) \to 0$ as $|x| \to \infty$. Denote $a_0 \geq 0$ such that $f(a_0) = \gamma$ and assume that $f_x(a_0) \neq 0$. Then, the free boundary's endpoint is governed by (for small time t):

$$a(t) = a_0 + a_1\sqrt{t} + o(\sqrt{t}), \tag{2.21}$$

for some $a_1 \geq 0$. Indeed, we look for an increasing function $a(t)$ such that the exterior mass of Equation (2.20) is zero:

$$m(t) = \int_{a(t)}^{\infty} dx \int_0^t ds \int_{-a(s)}^{a(s)} G(x-y, t-s)(f(y) - \gamma)dy,$$

where we use the Green's formula for the heat equation, Equation (2.15), to represent u. Since $a(t)$ is an increasing function, we have

$$y \leq a(s) \leq a(t) \leq x.$$

Therefore, for t small, the Green's function $G(x-y, t-s)$ is sharply peaked near the point

$$y = a(t),\ s = t,\ x = a(t).$$

So we can replace $(f(y) - \gamma)$ by the first few terms in its Taylor expansion

$$f(y) - \gamma = (y - a_0)f_1 + \mathcal{O}((y-a_0)^2),$$

in which $f_1 = f_x(a_0)$. Also, since $G(x-y,t-s)$ decays exponentially as $y \to -\infty$, we replace the lower limit $y = -a(s)$ by $-\infty$. Now the mass can be approximated by

$$m(t) = f_1 \int_{a(t)}^{\infty} dx \int_0^t ds \int_{-\infty}^{a(s)} (y - a_0) G(x-y, t-s) dy$$

Next we show the existence of a_1 satisfying the following approximations

$$a(t) = a_0 + a_1\sqrt{t} + o(\sqrt{t}), \quad \text{and} \quad m(t) = 0.$$

We change the variables to

$$\begin{aligned} x =& x_1\sqrt{t} + a_0, && x_1 \in [a_1, \infty), \\ y =& y_1\sqrt{t} + a_0, && y_1 \in (-\infty, a_1\sqrt{s_1}\,], \\ s =& s_1 t, && s_1 \in [0,1], \end{aligned}$$

and

$$\begin{aligned} x_1 =& x_2 a_1, && x_2 \in [1, \infty), \\ y_1 =& y_2 a_1, && y_2 \in (-\infty, \sqrt{s_1}\,], \end{aligned}$$

and note that $G(x-y,t-s) = t^{-1/2} G(x_1 - y_1, 1-s_1) = t^{-1/2} G(a_1(x_2-y_2), 1-s_1)$. Then

$$\begin{aligned} m(t) =& f_1 t^2 \int_{a_1}^{\infty} dx_1 \int_0^1 ds_1 \int_{-\infty}^{a_1\sqrt{s_1}} y_1 G(x_1 - y_1, 1-s_1) dy_1 \\ =& a_1^3 f_1 t^2 \int_1^{\infty} dx_2 \int_0^1 ds_1 \int_{-\infty}^{\sqrt{s_1}} y_2 G(a_1(x_2 - y_2), 1-s_1) dy_2. \end{aligned}$$

Consider the rescaled masses $\widetilde{m_1}(a_1) = m(t)/(f_1 t^2)$ and $\widetilde{m_2}(a_1) = m(t)/(a_1^2 f_1 t^2)$; *i.e.*,

$$\begin{aligned} \widetilde{m_1}(a_1) &= \int_{a_1}^{\infty} dx_1 \int_0^1 ds_1 \int_{-\infty}^{a_1\sqrt{s_1}} y_1 G(x_1 - y_1, 1-s_1) dy_1, \\ \widetilde{m_2}(a_1) &= a_1 \int_1^{\infty} dx_2 \int_0^1 ds_1 \int_{-\infty}^{\sqrt{s_1}} y_2 G(a_1(x_2 - y_2), 1-s_1) dy_2. \end{aligned}$$

As $a_1 \to 0$, $\widetilde{m_1}(a_1)$ goes to

$$\widetilde{m_1}(0) = \int_0^\infty dx_1 \int_0^1 ds_1 \int_{-\infty}^0 y_1 G(x_1 - y_1, 1 - s_1) dy_1$$

with $\widetilde{m_1}(0) < 0$. This shows that $m(t) < 0$ for $a_1 = 0$.

On the other hand, for $a_1 \gg 1$, $a_1 G(a_1(x_2 - y_2), 1 - s_1)$ is approximately the Dirac delta function at $x_2 = y_2$, $s_1 = 1$. At this point, we have $y_2 > 0$, therefore $\widetilde{m_2}(a_1) > 0$. This shows that $m(t) > 0$ for large values of a_1. Thus there exists a positive value a_1 so that $m(t) = 0$.

A similar result holds for zero force and non-zero (finitely supported) initial data.

2.3 Numerical Implementation

Given an elliptic operator A, we would like to solve problems of the form:

$$Au + \partial \|u\|_{L^1} \ni f$$

or

$$u_t + Au + \partial \|u\|_{L^1} \ni f$$

which corresponds to the elliptic or parabolic equations, respectively. We will present two methods to do so. The first scheme is semi-implicit (also known as implicit-explicit or proximal gradient method), where the subgradient term is discretized forward in time and the diffusion term is lagged. We apply this method to solve the time dependent equations. The second scheme is the Douglas-Rachford method, which we use to solve both the elliptic problem and the parabolic problem. Both methods can handle the multivalued nature of the subgradient $\partial \|u\|_{L^1}$. In this section, we denote h and τ the space and time steps of the finite difference schemes.

Implicit-Explicit Scheme (Proximal Gradient Method)

From the numerical perspective, the multivalued term $\partial\|u\|_{L^1}$ is the main source of difficulties, since the value is ambiguous. However, an operator of the form $I+\sigma\,\partial F$ (where F is convex) has an easy-to-compute inverse. The inverse operator $(I+\sigma\,\partial F)^{-1}$, also known as the resolvent or proximal operator, $\text{prox}_{\sigma F}(\cdot)$, can be found by solving the following optimization:

$$(I+\sigma\,\partial F)^{-1}(z) = \underset{v}{\text{argmin}}\ \frac{1}{2}||v-z||_{L^2}^2 + \sigma F(v). \tag{2.22}$$

For example, if $F(u) = ||u||_{L^1}$ and thus $\partial F(u) = \partial\|u\|_{L^1}$, we have:

$$\begin{aligned}(I+\sigma\,\partial\|\cdot\|_{L^1})^{-1}(z) &= \underset{v}{\text{argmin}}\ \frac{1}{2}||v-z||_{L^2}^2 + \sigma||v||_{L^1} \\ &= S(v,\sigma),\end{aligned}$$

where the shrink operator, S, is defined point-wise as $S(v,\sigma) := \max(|v|-\sigma,0)\dfrac{v}{|v|}$.

Using the proximal operator, we will write the discretization of Equation (1.3) in a semi-implicit form. We first discretize Equation (1.3) in time, then to apply the proximal gradient method, the subdifferential term is evaluated at time step $n+1$, yielding:

$$u^{n+1} - u^n + \tau A u^n + \tau\partial\|u^{n+1}\|_{L^1} \ni \tau f, \tag{2.23}$$

where $\tau > 0$ is the time step. The resulting iterative scheme is:

$$u^{n+1} = S(u^n - \tau A u^n + \tau f, \tau). \tag{2.24}$$

For example, for the heat equation, where $A = -\Delta$ and is discretized using five-point stencil, the iterative scheme is:

$$u^{n+1} = S(u^n + \tau\Delta u^n + \tau f, \tau), \tag{2.25}$$

and is convergent given $\tau \leq \frac{h^2}{4}$. This scheme has the same complexity as the corresponding standard explicit methods for PDEs.

Alternating Direction Implicit (Douglas-Rachford) Method

The Douglas-Rachford algorithm for nonlinear multivalued evolution equation was studied in [93]. Denote $Bu := \partial\|u\|_{L^1}$, the iterative scheme for Equation (1.3) is

$$u^{n+1} = (I + \tau B)^{-1}\left[(I + \tau A)^{-1}(I - \tau B) + \tau B\right] u^n, \tag{2.26}$$

which can be rewritten as:

$$\begin{aligned} u^{n+1} &= (I + \tau B)^{-1}\tilde{u}^n \\ \tilde{u}^{n+1} &= \tilde{u}^n + (I + \tau A)^{-1}(2u^{n+1} - \tilde{u}^n) - u^{n+1}. \end{aligned} \tag{2.27}$$

It was shown that the method is unconditionally stable and convergent for all $\tau > 0$ [37, 93, 136]. Also, note that the iterates u^n converges to a solution of the stationary equation (1.2). For the sandpile problem [91], which is discussed in Section 2.4, the operators A and B are chosen specifically as follows:

$$Au = -\Delta u - f, \quad Bu = \partial\|u\|_{L^1}, \tag{2.28}$$

so that the operation for u^{n+1} in the iterative process, Equation (2.27), is a shrink. The corresponding proximal operators are

$$\begin{aligned} \text{prox}_{\tau F}(z) &= (I + \tau A)^{-1}(z) = (I - \tau\Delta)^{-1}(z + \tau f) \\ \text{prox}_{\tau G}(z) &= (I + \tau B)^{-1}(z) = S(z, \tau), \end{aligned}$$

where $F(z) = \frac{1}{2}\|\nabla z\|_{L^2}^2 - \langle f, z\rangle$ and $G(z) = \|z\|_{L^1}$. To compute $(I - \tau\Delta)^{-1}$ numerically, we use the FFT, where the discrete Laplacian $\Delta_h u$ is viewed as the convolution of u with the finite difference stencil.

Remark 2. *Since the shrink operator is the last step of the iterative process, this method provides a numerically well-defined support set for u, making it easier to locate the free boundary.*

2.4 Computational Simulations

In this section we show convergence of our numerical scheme to known solutions, approximations to the support set evolution, and numerical solutions for higher dimension.

Numerical Convergence

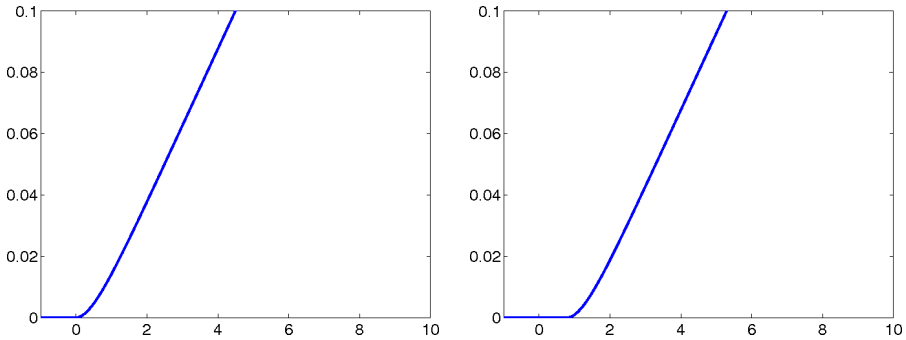
In Figure 2.1, we solve Equation (2.3) (with $\gamma = 0.05$) using the implicit-explicit scheme (Equation (2.25)). The initial data is taken to be the traveling wave profile (Equation (2.17)) with speed $\sigma = 2$. The numerical solution has the correct support set and speed of propagation, validating the traveling wave solution as well as the numerical method.

This is further confirmed in Figure 2.2, where the numerical solution is compared to the exact solution. To compute the error, we use the following norms:

$$\text{Error}_q(h) = \max_n ||u_h^n - u_{exact}||_q,$$

where $q = 1, 2, \infty$ and u_h^n is the solution at t_n with space resolution h. The errors in these three norms are plotted along side the line representing the second order (dashed line) convergence.

To test the stability of these traveling wave solutions, we initialize our numerical scheme with the traveling wave profile perturbed by uniformly random noise sampled from $[0, 0.05]$. The time evolution is shown in Figure 2.3. In a short time, the Laplacian term dominates the evolution, which is expected. The solution gradually smoothes down to a new traveling wave profile and begins to translate at the expected speed. This shows that the traveling wave solution is an attracting solution, at least locally.



(a) $t = 0$ (b) $t = 0.398$

(c) $t = 1.594$ (d) $t = 2.390$

Figure 2.1: Numerical solution starting with an initial traveling wave profile with $\sigma = 2$ and $\gamma = 0.05$ computed using 500 grid points.

One Dimensional Heat Equation

In Figure 2.4, the plot shows the modified heat equation (Equation (2.3)) with zero initial data and force $f(x) = 2e^{-5x^2}$. The solutions evolves upward in time with their support sets marked by red circles. We see that the computed solutions are indeed compactly supported in space, as the theory states. The corresponding table provides a least squares fit to estimate the coefficient a_1 from Equation (2.21) under grid refinement. We see that the coefficient a_1 approaches the value 1 quickly within some small approximation error, which is used to verify that our numerical approximation is valid.

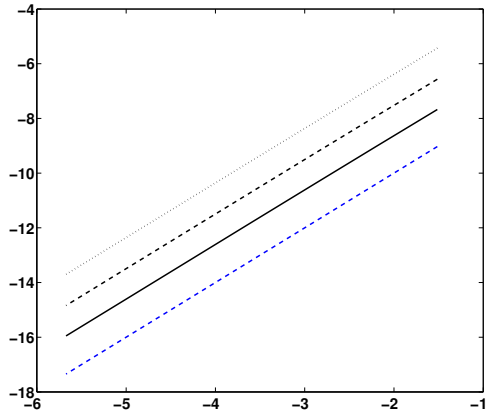


Figure 2.2: Convergence analysis using the L^1 (dotted line), L^2 (dashed line), and L^∞ (solid line) norms in space and L^∞ norm in time. The x-axis is the *log* of the grid resolution h and the y-axis is the *log* of the Error. The blue dashed line represents second order convergence.

Two Dimensional Heat Equation

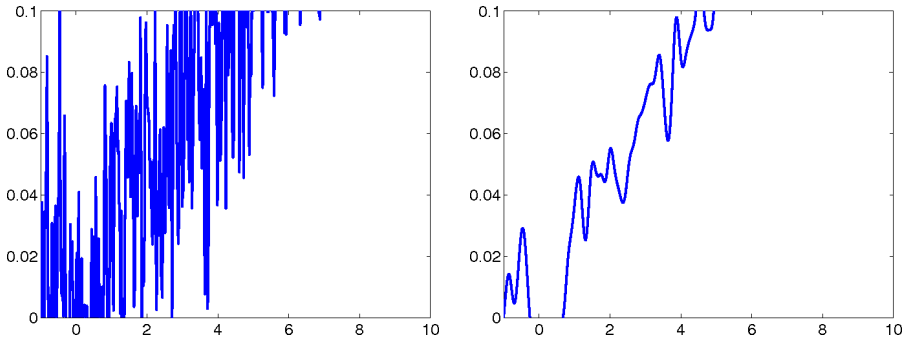
In Figure 2.5, we compute the solution of Equation (2.3) with $\gamma = 2$ and $f = 0$. In this case, we apply the parabolic Douglas-Rachford algorithm, which allows for larger time-steps. The initial data is a smoothed indicator function on the star shaped domain. In Figure 2.6, the corresponding support set of Figure 2.5 is shown. The support set grows outward to a maximum size and retracts inward as the solution decays to zero. The solution is identically zero at time $t = 0.1152$.

Graph Diffusion

In higher dimensions, we can consider the standard normalized diffusion equation:

$$\begin{aligned} u_t = L_g u &:= -\left(\mathrm{I} - D^{-1/2} A D^{-1/2}\right) u \\ u(x,0) &= g(x), \end{aligned} \tag{2.29}$$

where L_g is the graph Laplacian, A is the adjacency matrix, and D is the degree matrix. For more on the graph Laplacian, see [36, 151].



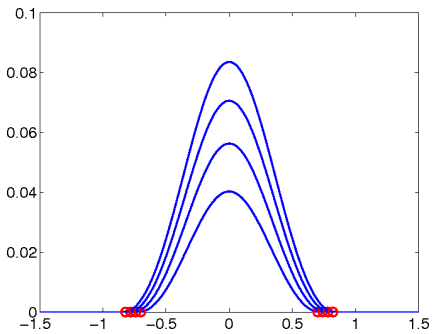
(a) $t = 0$ (b) $t = 0.004$

(c) $t = 0.040$ (d) $t = 0.398$

(e) $t = 1.594$

Figure 2.3: Numerical solution starting with an initial traveling wave profile perturbed by uniformly random noise sampled from $[0, 0.05]$ with $\sigma = 2$ and $\gamma = 0.05$. This is solved on a grid of 500 points.

In Figure 2.7, the points represent the projection of vectors from $\mathbb{R}^{100}$ and each point is connected to many others in a non-local fashion. For the initial data, we concentrate the mass on one point in the far left, specifically, the $u(x_j, 0) = \delta_{j,1000}$



Number of grid points	Estimate of a_1
256	0.948
512	0.979
1024	0.985
2048	0.991
4096	0.995
8192	0.997
16384	0.997

Figure 2.4: The graph is a 1D simulation of the heat equation with the subgradient term, zero initial data, and a Gaussian forcing function centered around zero, $f(x) = 2e^{-5x^2}$. The solutions are growing upward in time and their support sets are marked by red circles. The table shows the estimate of the coefficient a_1 from Equation (2.21) under grid refinement.

where $\delta_{j,k}$ is the Kronecker delta function. As the system evolves governed by Equation (2.29), the solution becomes strictly positive quickly.

The modified equation is:

$$\begin{aligned} u_t &= -\left(\mathrm{I} - D^{-1/2}AD^{-1/2}\right)u - \gamma p(u), \\ u(x,0) &= g(x). \end{aligned} \tag{2.30}$$

In Figure 2.8, we begin with the same initial condition and see that over time the support set does not grow past a bounded region if u evolves as in (2.30). Therefore, numerically we show that the support is of finite size for the case of graph diffusion. In Figure 2.8(d), the solution begins to decay to zero which causes its support set to retract towards the initial support before vanishing.

Signum-Gordon Equation

The signum-Gordon equation has an interpretation as an approximation to certain physical models [4, 3, 5]. The equation takes the form of a second order nonlinear hyperbolic equation:

$$\begin{aligned} u_{tt} - \Delta u &= -\text{sign}(u) \\ u(x,0) &= g_1(x) \\ u_t(x,0) &= g_2(x), \end{aligned} \tag{2.31}$$

and exhibits both compactly supported traveling waves and oscillatory (stationary) soliton-like structures. This equation can be derived from the Lagrangian with the following L^1 potential:

$$L = \text{Kinetic} - \text{Potential} = \frac{1}{2}|u_t|^2 - \frac{1}{2}|\nabla u|^2 - |u|.$$

The equation of motion can be derived from the Lagrangian:

$$\begin{aligned} u_{tt} - \Delta u &= -p(u) \\ u(x,0) &= g_1(x) \\ u_t(x,0) &= g_2(x), \end{aligned}$$

which is the same as Equation (2.31) by replacing the sign(u) term with the subgradient $p(u)$.

To discretize the problem, we apply the ideas from the proximal gradient method, by placing $p(u)$ in the future:

$$u^{n+1} - 2u^n + u^{n-1} - \tau^2 \Delta u^n = -\tau^2 p(u^{n+1}),$$

and thus,

$$u^{n+1} = S(2u^n - u^{n-1} + \tau^2 \Delta u^n, \tau^2).$$

In Figure 2.9, we plot our numerical approximation to the traveling wave solution found in [3]. Since the traveling wave profile is also known analytically,

we show numerical convergence of our scheme as $h \to 0^+$ (see Table 1). Also, in Figure 2.10, we show the time evolution of an oscillatory compact soliton-like structure which appears in [4, 5]. These examples show the range of behaviors that appear via the addition of an L^1 subgradient term.

Table 2.1: Error between our numerical solution and the analytic solution of the Signum-Gordon Equation.

Grid Size	128	256	512	1024	2048	4096	8192
L^2-Error	0.4601	0.2319	0.1133	0.0569	0.0284	0.0143	0.0072

Divisible Sandpile

As a model for self-assembly and internal diffusion limited aggregation, the sandpile problem has received attention recently [121, 91, 46, 89, 90]. The problem is posed discretely, but has the following continuous formulation for the divisible sandpile problem [91, 46, 92]:

$$\Delta u = 1 - f, \quad \text{if } u \geq 0, \tag{2.32}$$

where f is some non-negative external force. By multiplying Equation (2.32) with u and integrating over $\mathbb{R}^2$, the associated variational energy is:

$$\min_u \int_{u \geq 0} \frac{1}{2}|\nabla u|^2 + u - uf \ \mathrm{dx}. \tag{2.33}$$

There are several choices for relaxing the constraint $u \geq 0$, in particular, we use the following:

$$\min_u \int \frac{1}{2}|\nabla u|^2 + |u| - uf \ \mathrm{dx}. \tag{2.34}$$

It can be shown (via maximum principle) that for $f \geq 0$ the solution of Equation (2.34) and Equation (2.33) are the same. The Euler-Lagrange equation for

the L^1 sandpile problem is:

$$\Delta u = p(u) - f, \tag{2.35}$$

and is solved numerically via the Douglas-Rachford algorithm (see Equation (2.27)). Note that if the external force is a finite sum of characteristic functions $f = \sum_{j=1}^{N} \alpha_j \chi_{S_j}$ where S_j are compact sets and $\alpha_j \geq 0$, then by integrating Equation (2.35) over $\mathbb{R}^2$ we get:

$$|\text{supp}(u)| = \sum_{j=0}^{N} \alpha_j |S_j|, \tag{2.36}$$

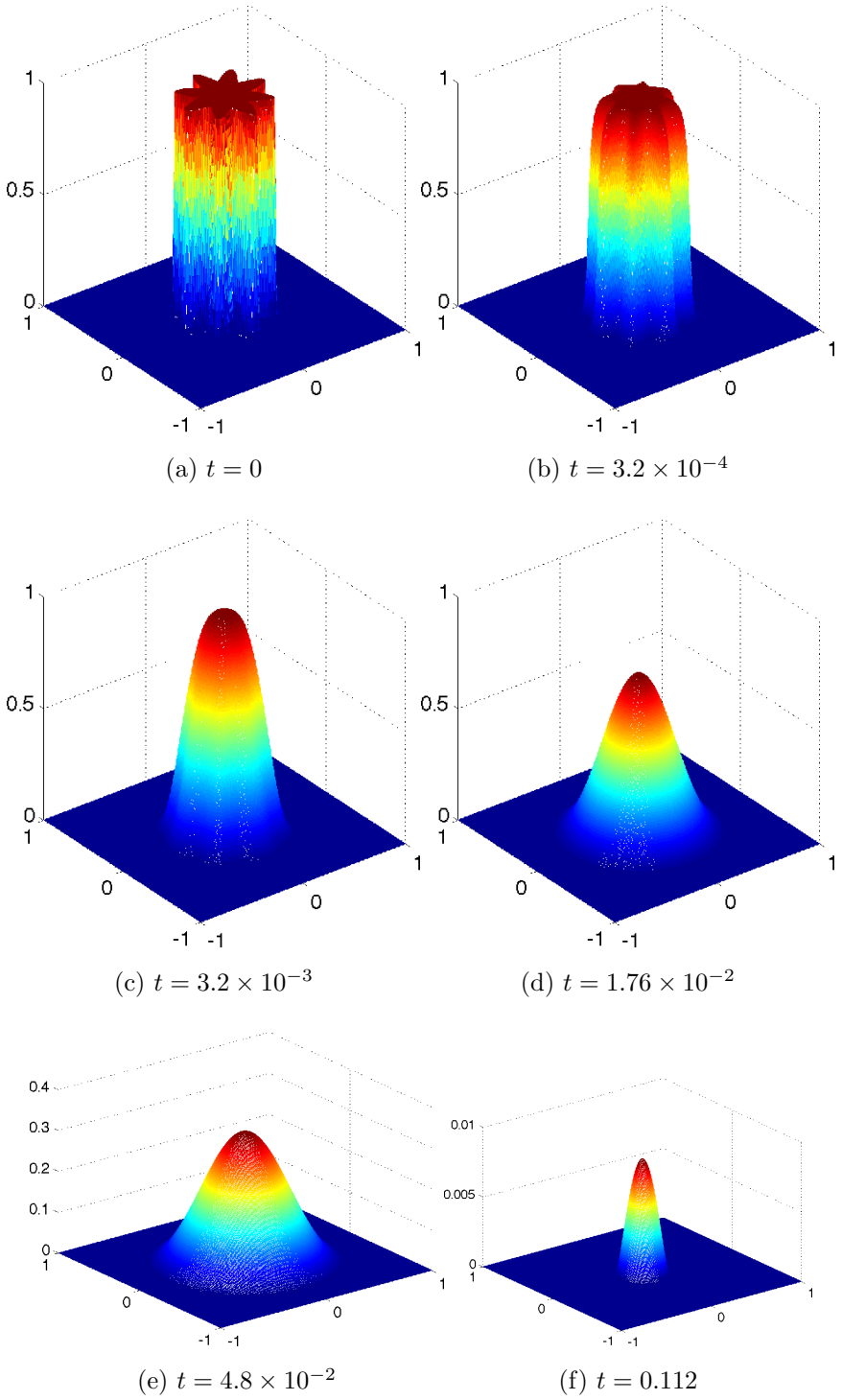
since $u \geq 0$ and $\text{supp}(u)$ is compact. This is referred to as the preservation of mass.

In Figure 2.11, we take $f = \chi_{S_1} + \chi_{S_2}$, where S_1 and S_2 are the two overlapping square domains (on the left). The support set of u, given in Figure 2.11 (right), agrees with direct numerical simulation of the discrete sandpile problem. The direct simulation follows a topping rule described in [91].

In Figures 2.12-2.14, we take $f = \alpha \chi_S$ where S is the shape given in Figures 2.12-2.14 (the top left), and $\alpha = 2.0, 1.2$ and 1.5, respectively. The support set of u is given in Figures 2.12-2.14 (the bottom right) with intermediate calculation shown in Figures 2.12-2.14 (the remaining plots). To verify that the solutions from our algorithm correspond to the correct solutions for the sandpile problem, we use the mass conservation property, Equation (2.36). Unlike direct simulation, our method also calculates the function u as shown in Figure 2.15. One of the benefits of our approach is that the solutions can be computed quickly, for example, our method is at least 8 times faster than direct simulation (76 seconds vs. 652 seconds) at approximating the solution found in Figure 2.15.

Conclusion

By adding the subdifferential of L^1 to certain PDEs, we have shown (numerically and theoretically) various properties of the solutions. These problems arise from physical models as well as exact relaxation of other PDEs, and could provide useful tools in computing fast approximations to nonlinear problems with a compactly supported free boundary. This is all in the spirit of borrowing the key idea from compressed sensing, that L^1 regularization implies sparsity of discrete systems and transferring it to classical problems in PDEs.

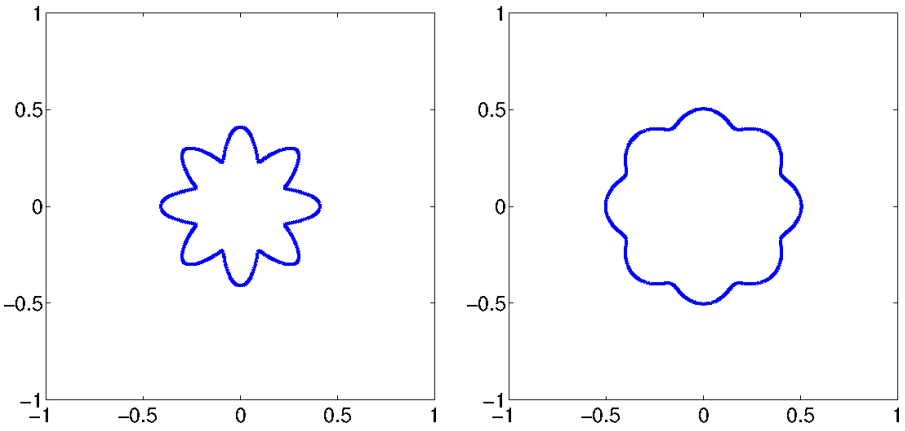


(a) $t = 0$ (b) $t = 3.2 \times 10^{-4}$

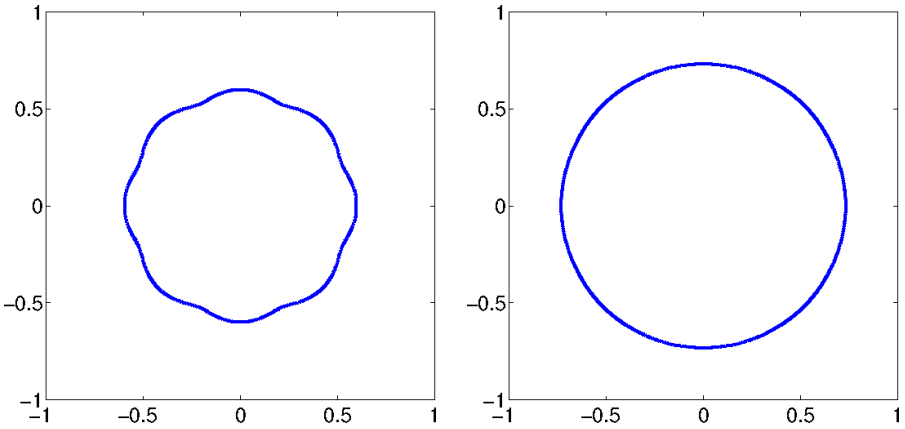
(c) $t = 3.2 \times 10^{-3}$ (d) $t = 1.76 \times 10^{-2}$

(e) $t = 4.8 \times 10^{-2}$ (f) $t = 0.112$

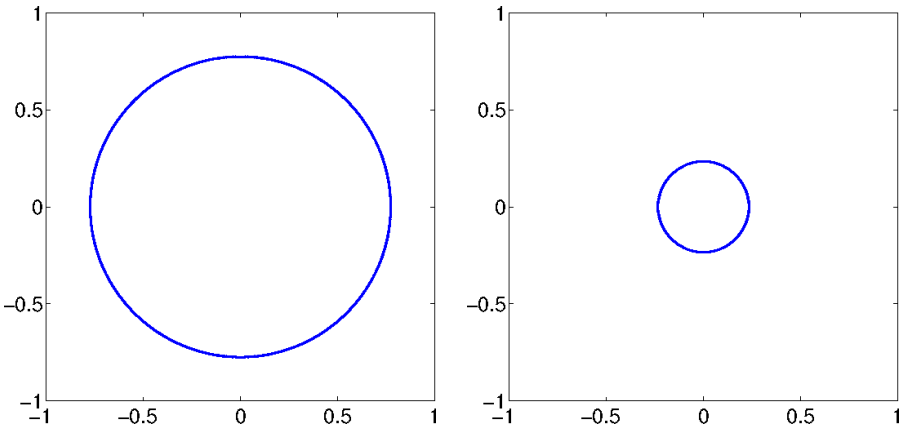
Figure 2.5: Solutions of the initial value problem (no forcing term) computed on a 500 by 500 grid with $\gamma = 2$ at times indicated. The solution smoothes out and decays to zero.



(a) $t = 0$ (b) $t = 3.2 \times 10^{-4}$

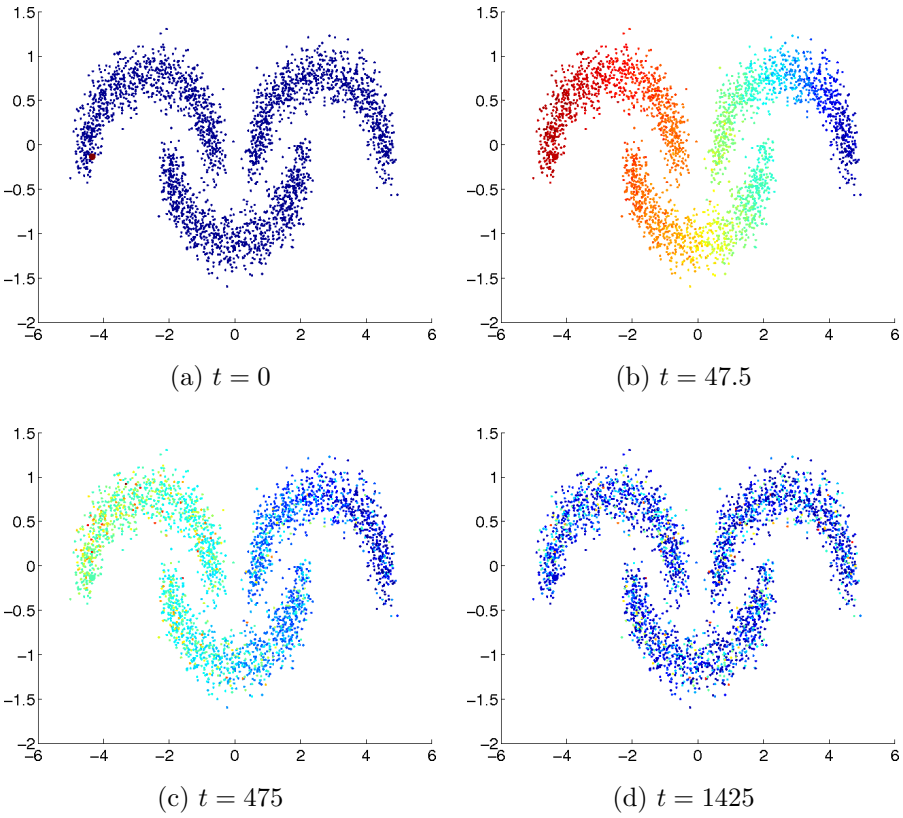


(c) $t = 3.2 \times 10^{-3}$ (d) $t = 1.76 \times 10^{-2}$



(e) $t = 4.8 \times 10^{-2}$ (f) $t = 0.112$

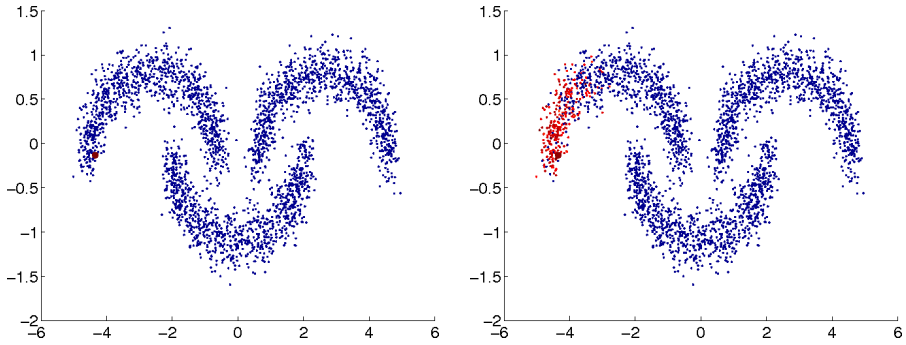
Figure 2.6: Support set of the initial value problem in Figure 2.5. The support set grows outward to a maximum size and retracts inward as the solution decays to zero.



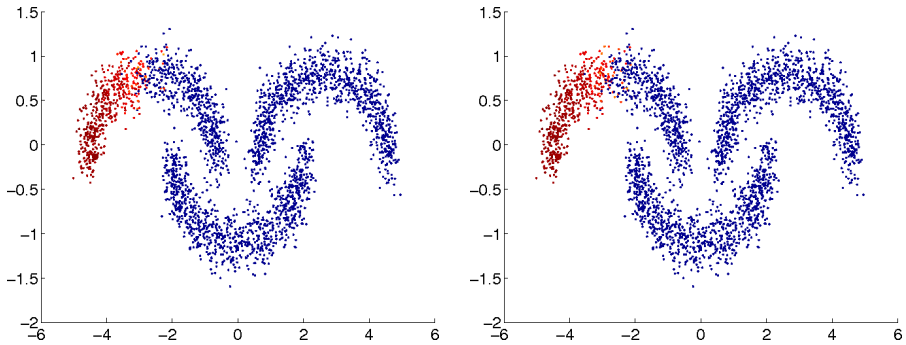
(a) $t = 0$ (b) $t = 47.5$

(c) $t = 475$ (d) $t = 1425$

Figure 2.7: Solution of the initial value problem diffusing standard normalized graph Laplacian.



(a) $t = 0$ (b) $t = 2.85$



(c) $t = 6.65$ (d) $t = 11.4$

Figure 2.8: Solution of the initial value problem with the subgradient term, $\gamma = 5 \times 10^{-5}$.

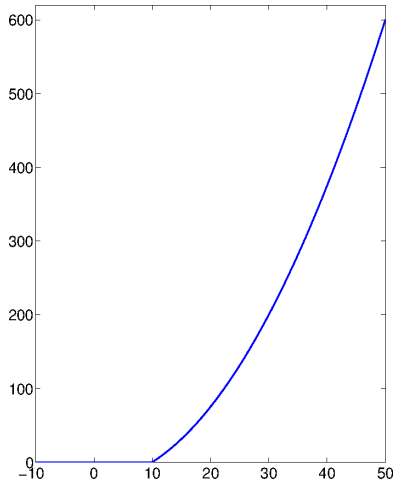


Figure 2.9: Our numerical approximation to a compact traveling wave solution to the Signum-Gordon equation.

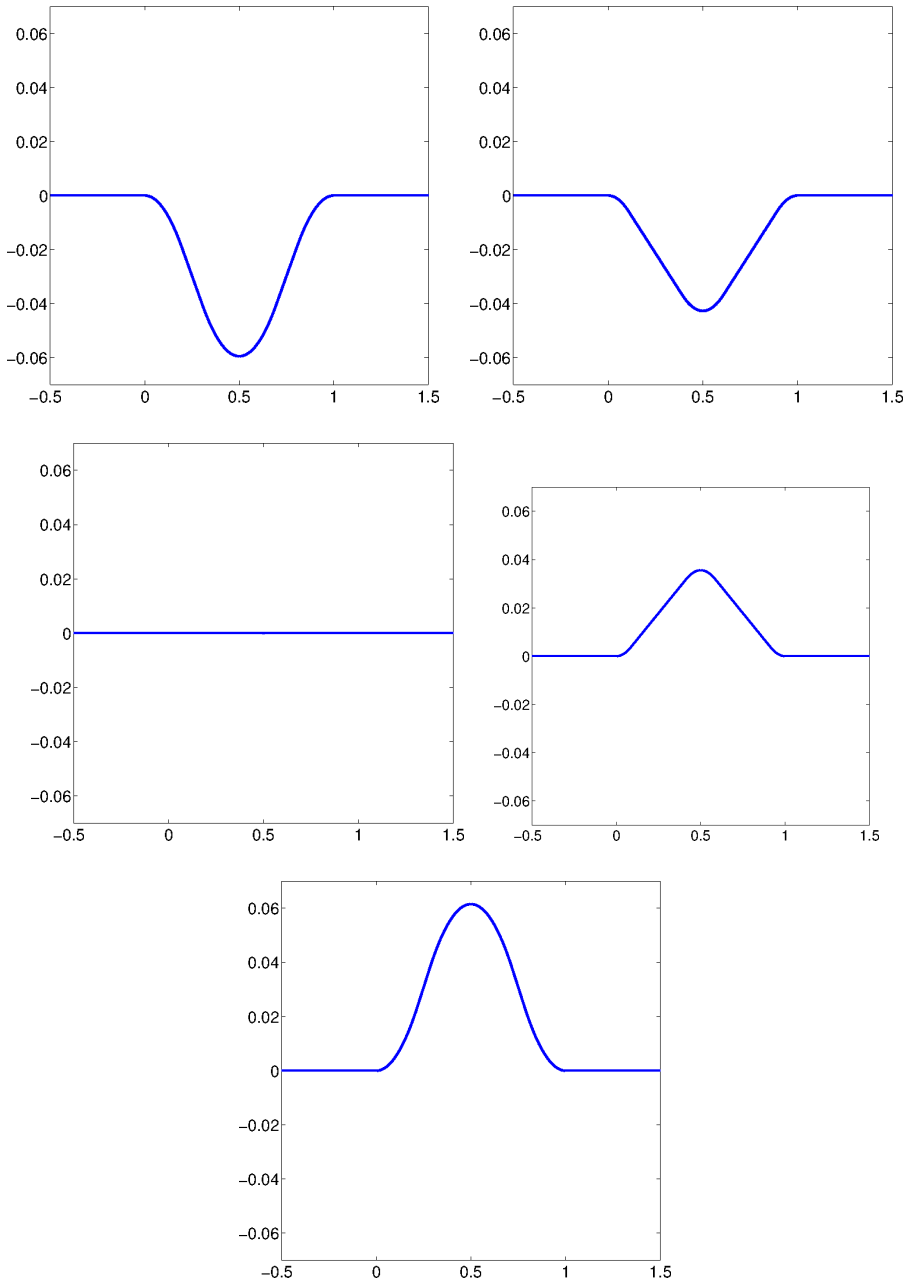


Figure 2.10: The dynamics of an oscilatory compact solution of the Signum-Gordon equation.

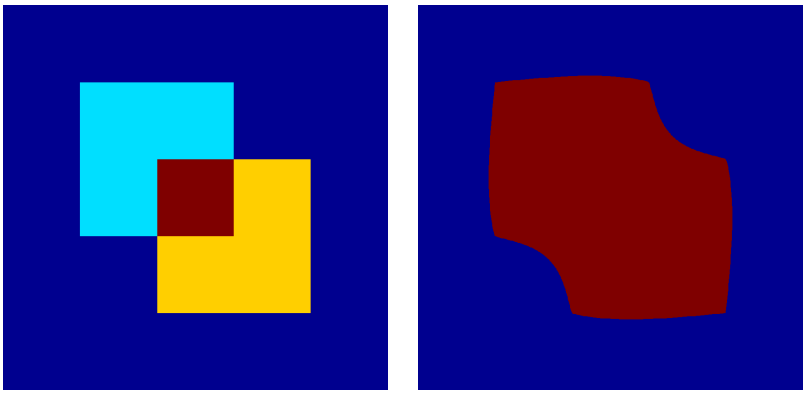


Figure 2.11: A two-region sandpile problem, where each of the larger squares defines the set S_j, for $j = 1, 2$. The darker blue region has no mass, the lighter blue and yellow region has a mass density of 1, and the red (overlap) region has a density of 2. On the left, the region of positive mass is displayed.

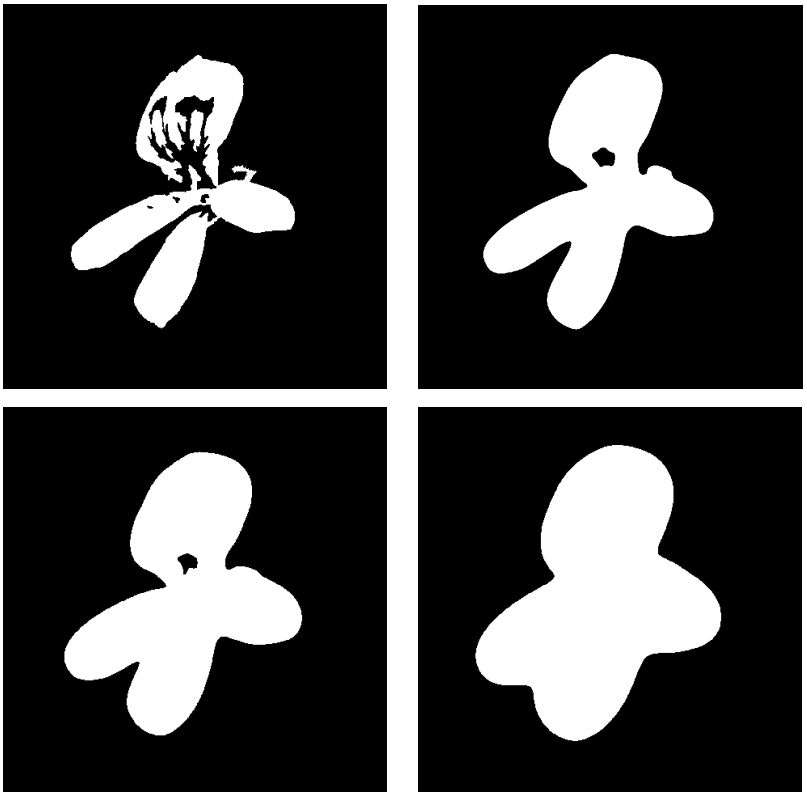


Figure 2.12: The iterative evolution of our sandpile problem algorithm applied to a flower-shaped region S on the top left with $f = 2\chi_S$. The final state appears in the bottom right corner.

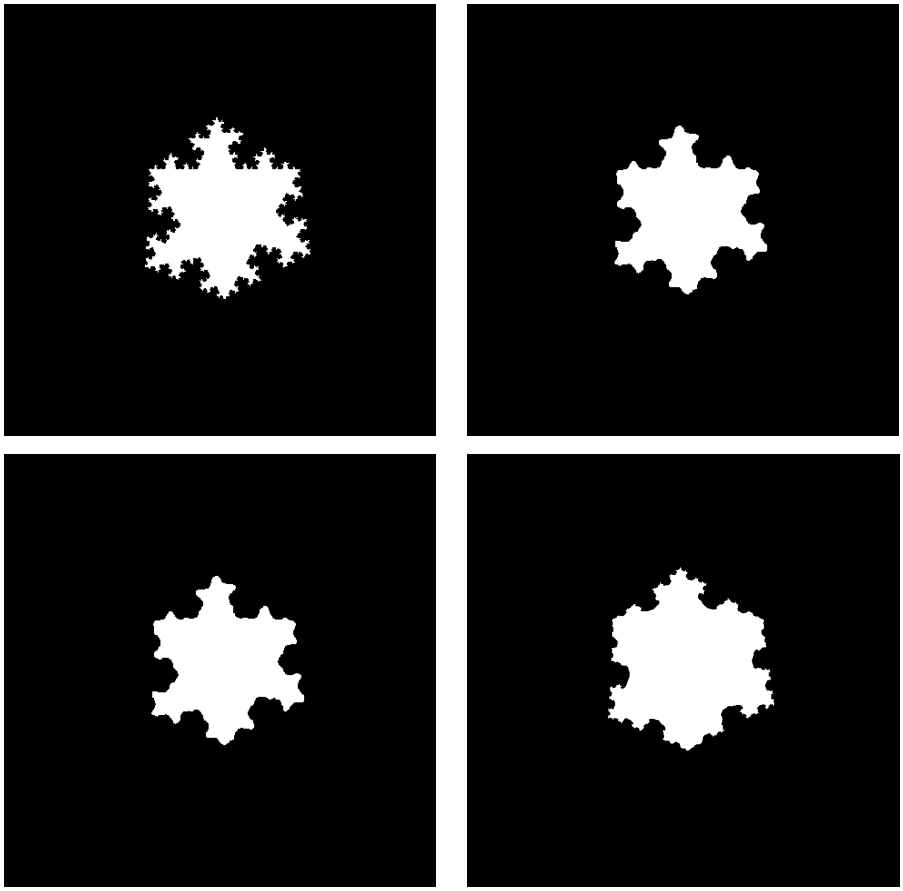


Figure 2.13: The iterative evolution of our sandpile problem algorithm applied to the fractal region S on the top left with $f = 1.2\chi_S$. The final state appears in the bottom right corner.

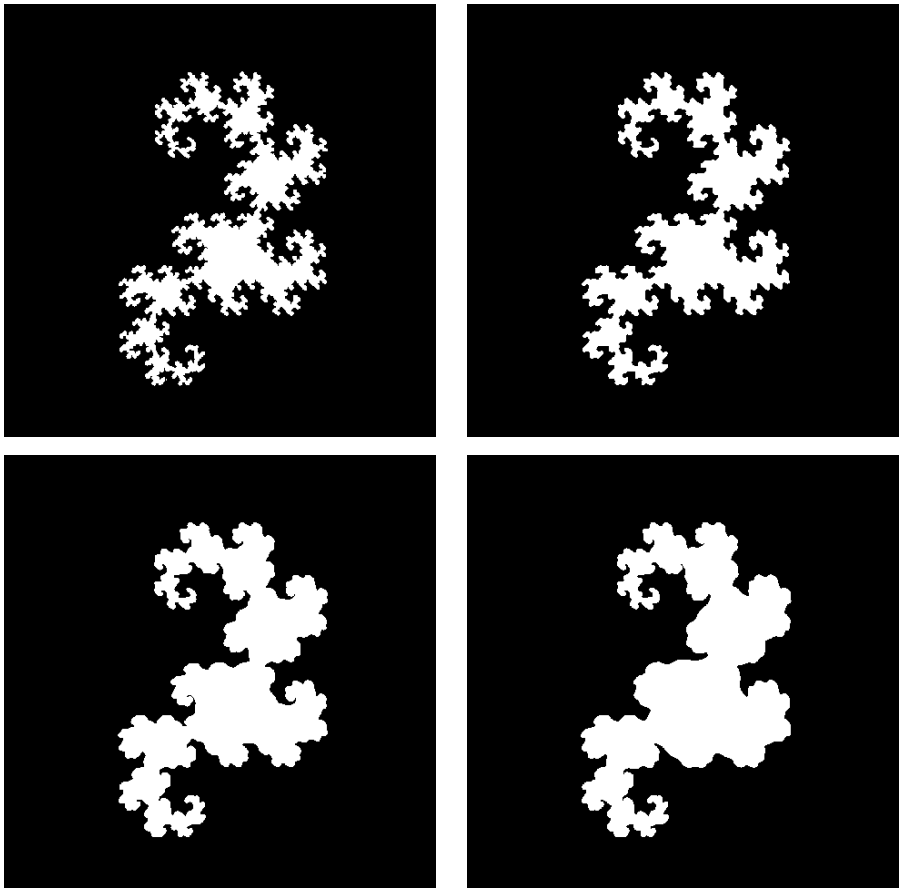


Figure 2.14: The iterative evolution of our sandpile problem algorithm applied to the fractal region S on the top left with $f = 1.5\chi_S$. The final state appears in the bottom right corner.

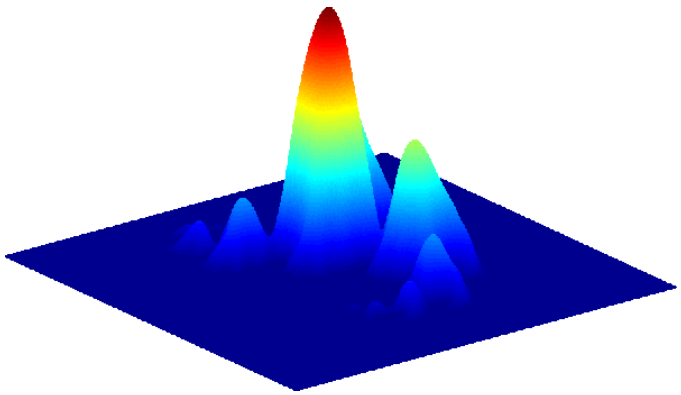


Figure 2.15: The solution u from Figure 2.14 (bottom right).

CHAPTER 3

An L^1 Penalty Method for Obstacle Problems

In this chapter, we illustrate the use of an L^1-penalty method to construct an unconstrained problem whose solutions correspond exactly to those of the obstacle problems. We provide a lower bound on the value of the penalty parameter and use this to guide our numerical calculations. We then present several experimental results showing the applicability of our method to various physical problems. This chapter was taken with slight modification from [149].

3.1 Motivation

In this section, we motivate the use of L^1 based optimization for obstacle problems by establishing a connection between solutions of an L^1 penalized variational method and the solutions of obstacle problems with zero obstacle. These problems were considered in [22, 29] and can be used for finding compactly supported functions. Given $f \in L^2(\mathbb{R}^d)$ and $\mu \geq 0$, consider the following functional defined for $v \in H^1(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$,

$$\mathcal{J}(v) = \int_{\mathbb{R}^d} \frac{1}{2}|\nabla v|^2 - fv + \mu|v|\, dx. \tag{3.1}$$

Then for all test functions $\psi \in H^1(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$, its unique minimizer u,

$$u = \operatorname{argmin}\{\mathcal{J}(v) \mid v \in H^1(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)\}, \tag{3.2}$$

satisfies the following equation:

$$\int_{\mathbb{R}^d} \nabla u \cdot \nabla \psi - f\psi + \mu p(u)\psi\, dx = 0, \tag{3.3}$$

where $p(u)$ is an element of the subdifferential of the L^1 term in Equation (3.1) and can be identified by (see [29]):

$$p(u) = \begin{cases} \text{sign}(u) & \text{if} \quad u \neq 0 \\ -f/\mu & \text{if} \quad u = 0. \end{cases}$$

We also consider the solution of the following obstacle problem,

$$\bar{u} = \text{argmin}\{\mathcal{J}(v) \mid v \in H^1(\mathbb{R}^d) \cap L^1(\mathbb{R}^d) \ \text{ and } \ v \geq 0\}. \tag{3.4}$$

As a minimizer, $\bar{u}$ satisfies the variational inequality

$$\int_{R^d} \nabla \bar{u} \cdot \nabla \psi - f\psi + \mu\psi \, dx \geq 0, \tag{3.5}$$

for all test functions $\psi \in H^1(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ with $\psi \geq 0$. One can analogously define $\underline{u}$ as the minimizer of $\mathcal{J}$ over $v \in H^1(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ with $v \leq 0$. We will show that the solutions to the variational problems above, u, $\bar{u}$ and $\underline{u}$, are related. For the rest of the paper, we denote

$$u_+ := \max(u, 0) \quad \text{and} \quad u_- := \min(u, 0).$$

Theorem 3. *Let u and $\bar{u}$ be the solutions of Equations* (3.2) *and* (3.4)*, respectively, then $\bar{u} = u_+$. Moreover, if $f \geq 0$ then $\bar{u} = u$. Similarly, we have $\underline{u} = u_-$ and if $f \leq 0$ then $\underline{u} = u$.*

Proof. Let $w \geq 0$ be a solution of the variational inequality:

$$\int_{\mathbb{R}^d} \nabla w \cdot \nabla \psi - f\psi + \mu\psi \, dx \geq 0, \tag{3.6}$$

for all $\psi \in H^1(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ with $\psi \geq 0$. Next, since $(u - w)_+$ is a valid test function for Equation (3.3) the following holds:

$$0 = \int_{\mathbb{R}^d} \nabla u \cdot \nabla (u - w)_+ - f(u - w)_+ + \mu p(u)(u - w)_+ \, dx.$$

Since $p(u) = 1$ on $\{(u - w)_+ \neq 0\} \subset \{u > 0\}$, we have

$$0 = \int_{\mathbb{R}^d} \nabla (u - w) \cdot \nabla (u - w)_+ + \nabla w \cdot \nabla (u - w)_+ - f(u - w)_+ + \mu (u - w)_+ \, dx.$$

Note that $(u-w)_+$ is also a valid test function for Equation (3.6), so the sum of the last three terms in the above equation is non-negative. Therefore

$$0 \geq \int_{\mathbb{R}^d} \nabla(u-w) \cdot \nabla(u-w)_+ \, dx = \int_{\mathbb{R}^d} |\nabla(u-w)_+|^2 \, dx.$$

Thus $(u-w)_+ = c$ a.e., for some non-negative constant c. Since $(u-w)_+ \in L^1(\mathbb{R}^d)$, we have $c = 0$, which means $u_+ \leq w$. In particular, since $\bar{u}$ is also a supersolution of (3.6), we have $u_+ \leq \bar{u}$.

As for $\bar{u}$, after noting that for any ε the perturbation $\bar{u} - \varepsilon(\bar{u}-w)_+$ is an admissible function in the minimization (3.4), a similar calculation shows that

$$0 \leq \left. \frac{d}{d\varepsilon} \right|_{\varepsilon=0} \mathcal{J}(\bar{u} - \varepsilon(\bar{u}-w)_+) \leq -\int_{\mathbb{R}^d} |\nabla(\bar{u}-w)_+|^2 dx.$$

Taking the derivative in ε above can be justified by writing out the difference quotients. Using the same argument as before, we conclude that $\bar{u} \leq w$. Finally, to prove that $u_+ \geq \bar{u}$, we will show that u_+ is also a supersolution of (3.6). Indeed, since $-f + \mu p(u_+) \leq -f + \mu$ so as long as $\psi \in H^1(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ is nonnegative,

$$\int_{\mathbb{R}^d} \nabla u_+ \cdot \nabla \psi + (-f+\mu)\psi \, dx \geq \int_{\mathbb{R}^d} \nabla u_+ \cdot \nabla \psi - f\psi + \mu p(u_+)\psi \, dx = 0.$$

We have proven that $\bar{u} = u_+$. In particular, if $f > 0$, one can show that u is non-negative [22, 29]. In this case we have $\bar{u} = u$. This completes the proof. □

3.2 Obstacle Problem

In this section, we recall the classical obstacle problem as well as its penalty formulation which contains an L^1-like term. It is shown in [50, 101] that if the penalty parameter is large enough, the solution of the penalized problem is identical to the solution of the constrained optimization problem (the obstacle problem in our case). In addition, we provide a lower bound on the size of the penalty parameter of the unconstrained problem as a function of the obstacle.

We will consider the obstacle problem in a bounded domain $\Omega \subset \mathbb{R}^d$ with Dirichlet boundary conditions. The results we describe in this case will also hold on the whole space $\mathbb{R}^d$ under similar assumptions. Consider the problem of minimizing the Dirichlet energy

$$\mathcal{J}(v) = \int_\Omega \frac{1}{2}|\nabla v|^2 \, dx, \tag{3.7}$$

among all functions v such that $v - g \in H_0^1(\Omega)$ and $v \geq \varphi$, where $\varphi : \Omega \to \mathbb{R}$ is a given smooth function, called the obstacle, which has $\varphi \leq g$ on $\partial\Omega$. Its unique minimizer $\bar{u}$ satisfies the complementarity problem [131]:

$$-\Delta\bar{u} \geq 0, \quad \bar{u} \geq \varphi\,, \quad (-\Delta\bar{u})(\bar{u} - \varphi) = 0, \quad \bar{u} - g \in H_0^1(\Omega).$$

Let u_μ be the unique minimizer in $H_0^1(\Omega)$ of

$$\mathcal{J}_\mu(v) = \int_\Omega \frac{1}{2}|\nabla v|^2 + \mu(\varphi - v)_+ \, dx. \tag{3.8}$$

In [50, 101], the authors showed that $u_\mu = \bar{u}$, for μ large enough and provided a lower bound for μ which is the L^∞-norm of any dual optimal multiplier of (3.7). That is if $\mu \geq -\Delta v$, for any dual optimal multiplier $v \geq \varphi$ of the optimization (3.7), then $u_\mu = \bar{u}$. Here we provide a concrete lower bound for μ, which can also be derived from Theorem 3.

Theorem 4. *Let u and u_μ be the optimal minimizers of Equations* (3.7) *and* (3.8)*, respectively. Then for any μ such that $-\Delta\varphi \leq \mu$ we have $u_\mu = u$.*

Proof. For any $v \in H_0^1(\Omega)$, define $w = v + (\varphi - v)_+$. Then w is a valid test function for (3.7), i.e., $w \geq \varphi$. Compute

$$\begin{aligned}
\mathcal{J}_\mu(w) &= \int_\Omega \frac{1}{2}|\nabla v|^2 + \nabla(\varphi - v)_+ \cdot \nabla v + \frac{1}{2}|\nabla(\varphi - v)_+|^2 \, dx \\
&= \int_\Omega \frac{1}{2}|\nabla v|^2 + \nabla(\varphi - v)_+ \cdot \nabla\varphi - \frac{1}{2}|\nabla(\varphi - v)_+|^2 \, dx \\
&\leq \int_\Omega \frac{1}{2}|\nabla v|^2 + \mu(\varphi - v)_+ - \int_\Omega \frac{1}{2}|\nabla(\varphi - v)_+|^2 \, dx \\
&= \mathcal{J}_\mu(v) - \int_\Omega \frac{1}{2}|\nabla(\varphi - v)_+|^2 \, dx.
\end{aligned}$$

The inequality in the third line holds since $-\Delta\varphi \leq \mu$ in the weak sense. Therefore, $\mathcal{J}_\mu(v + (\varphi - v)_+) < \mathcal{J}_\mu(v)$ unless $\nabla(\varphi - v)_+$ is zero, which implies $(\varphi - v)_+ = 0$ since $(\varphi - v)_+ \in H_0^1(\Omega)$. In particular, we have

$$\mathcal{J}_\mu(u_\mu + (\varphi - u_\mu)_+) \leq \mathcal{J}_\mu(u_\mu).$$

Since u_μ is the uniqueness minimizer of (3.8), $(\varphi - u_\mu)_+ = 0$ which means $u_\mu \geq \varphi$ is a valid test function for (3.7). In addition, we observe

$$\mathcal{J}(u_\mu) = \mathcal{J}_\mu(u_\mu) \leq \mathcal{J}_\mu(u) = \mathcal{J}(u).$$

Since u is the unique minimizer of $\mathcal{J}$, we conclude that $u = u_\mu$. □

Remark 5. *It is worth noting that in the numerical experiments provided in this work, the smaller the value of μ is, the faster the iterative scheme converges to the steady state. Therefore, an explicit lower bound on μ greatly improves the convergence rate of the method.*

3.3 Free Boundary Problem

In this section, we show how to transfer a class of free boundary problems into a form where the methodology of Sections 3.1 and 3.2 can be directly applied. We emphasize that for these problems our primary interest is in the location of the free boundary $\partial\{u > 0\}$ as opposed to the solution itself. For a concrete example, we will focus our attention on the Hele-Shaw model.

Turning a Class of Free Boundary into an Obstacle

Consider the solution $u \geq 0$ of the following free boundary problem in $\mathbb{R}^d$:

$$\begin{cases} -\Delta u = f - \gamma & \text{in} \quad \{u > 0\} \\ u = |\nabla u| = 0 & \text{on} \quad \partial\{u > 0\}, \end{cases} \tag{3.9}$$

with some given source function f and constant γ. In this form, we can see the connection to an L^1-minimization problem (Equation (3.1) with $\gamma = \mu$). In general, this can be difficult to solve numerically because of the free boundary $\partial\{u > 0\}$. We will show that our method naturally treats the free boundary conditions thereby avoiding any difficulty in directly tracking or approximating it. The details are described below.

First let us define the obstacle:

$$\varphi := -\tfrac{\gamma}{2d}|x|^2 - (-\Delta)^{-1} f(x).$$

Here $(-\Delta)^{-1}$ is shorthand for convolution with the Newtonian potential in $\mathbb{R}^d$. Then the function $w = u + \varphi$ will be the least super harmonic majorant of φ in $\mathbb{R}^d$, that is, it solves the free boundary problem:

$$\begin{cases} -\Delta w = 0 & \text{in} \quad \{w > \varphi\} \\ \nabla w = \nabla \varphi & \text{on} \quad \partial\{w > \varphi\}. \end{cases} \tag{3.10}$$

By transforming the PDE (3.9), we replace the source term with an obstacle. Indeed, the solution w of Equation (3.10) is the unique minimizer of the following optimization problem:

$$w = \operatorname*{argmin}_{v \in H^1(\mathbb{R}^d)} \int \frac{1}{2}|\nabla v|^2 + \mu(\varphi - v)_+ \, dx, \tag{3.11}$$

for some parameter μ. Therefore, by finding the solution to the unconstrained optimization problem (Equation (3.11)), we can locate the free boundary to the original problem directly.

Hele-Shaw

Let us recall the classical Hele-Shaw problem with a free boundary. Let $K \subset \mathbb{R}^d$ be a compact set and $\Omega_0 \supset K$ be open and bounded. Fluid initially occupies Ω_0 and is injected at a constant rate 1 per unit length through the surface K. The fluid expands and occupies the region Ω_t with the free boundary Γ_t. Let

$p(x,t) : \mathbb{R}^d \times [0,\infty) \to \mathbb{R}$ be the pressure of the fluid. For simplicity we consider a slight variant of the Hele-Shaw model where $p(x,t)$ instead of its normal derivative is equal to 1 on ∂K, see [71]. Then the time integral of p, $u(x,t) = \int_0^t p(x,\tau)d\tau$ formally satisfies (see also [45, 60]),

$$\begin{cases} -\Delta u = \chi_{\Omega_0} - 1 & \text{in} \quad \Omega_t(u) \setminus K \\ u = t & \text{on} \quad K \\ u = |\nabla u| = 0 & \text{on} \quad \Gamma_t(u). \end{cases} \tag{3.12}$$

Note that $\Omega_t(u) = \{u > 0\}$. We are free to solve Equation (3.12) only since the free boundary is the same as the free boundary of the pressure p. Here we consider the stable flow examples. For an example of a numerical method to solve the unstable Hele-Shaw flow (with the known fingering effect), see [67].

Let us define the obstacle:

$$\varphi_0 := -\tfrac{1}{2d}|x|^2 - (-\Delta)^{-1}\chi_{\Omega_0}.$$

Similar to Section 3.3, the function $w = u + \varphi_0$ solves

$$w = \operatorname*{argmin}_{v \in \mathcal{V}_t} \int_{\mathbb{R}^d \setminus K} \frac{1}{2}|\nabla v|^2 + \mu(\varphi_0 - v)_+ \, dx,$$

where the admissible set is defined as follows

$$\mathcal{V}_t = \{(v - \varphi_0) \in (H^1 \cap L^1)(\mathbb{R}^d \setminus K) : v = \varphi_0 + t \ \text{ on } \ \partial K\}.$$

For computational purposes, it is desirable to avoid solving a minimization problem in a possibly complicated domain $\mathbb{R}^d \setminus K$. We formulate a penalization procedure to include the boundary condition on K as a second obstacle. To do so, we define new obstacles $\varphi_1 \le \varphi_2$,

$$\varphi_1 = \varphi_0 + t\chi_K \ \text{ and } \ \varphi_2 = \varphi_1\chi_K + t(1 - \chi_K)$$

with the associated double penalized energy,

$$\mathcal{J}_\gamma(v) = \int_{\mathbb{R}^d} \frac{1}{2}|\nabla v|^2 + \gamma_1(\varphi_1 - v)_+ - \gamma_2(\varphi_2 - v)_- \, dx, \tag{3.13}$$

for some parameters $\gamma_1, \gamma_2 > 0$. Note that the true solution of (3.12) has $u \leq t$ in $\mathbb{R}^d$ and $\varphi_1 \leq 0$ thus $v = u + \varphi_1 \leq t$ as well. Thus the true solution satisfies $v \leq \varphi_2$ in $\mathbb{R}^d$.

Since φ is not smooth, the argument of the previous section, namely that $-\Delta\varphi - \gamma_1$ is subharmonic for γ_1 sufficiently large, is not directly applied. However, we can build a smooth approximation for the obstacle using mollifier. Heuristically when one minimizes a discretization of $\mathcal{J}_\gamma$ with grid spacing h, the minimizer of the discretization is as good an approximation to $\mathcal{J}_\gamma$ as it is to,

$$\mathcal{J}_\gamma^h(v) = \int_{\mathbb{R}^d} |\nabla v|^2 + \gamma_1(\rho_h * \varphi_1 - v)_+ - \gamma_2(\rho_h * \varphi_2 - v)_- \, dx,$$

where $\rho_h = h^{-d}\rho(h^{-1}x)$ with $\rho \in C^\infty(\mathbb{R}^d)$ being a standard mollifier. Note that since mollifying preserves ordering, $\rho_h * \varphi_1 \leq \rho_h * \varphi_2$. Now one can estimate:

$$\|\Delta\rho_h * \varphi_1\|_{L^\infty} \leq \|\Delta\varphi_0\|_{L^\infty} + t\|\Delta\rho_h\|_{L^1}\|\chi_K\|_{L^\infty} \leq 1 + th^{-2}\|\Delta\rho\|_{L^1(\mathbb{R}^d)}.$$

For the mollified functional $\mathcal{J}_\gamma^h$ as long as

$$\gamma_1 \geq th^{-2}\|\Delta\rho\|_{L^1(\mathbb{R}^d)} + 1,$$

the global minimizer solves the obstacle problem with $\rho_h * \varphi_1$ as the obstacle from below. A similar argument holds for γ_2. We are using a slight extension of the result of Section 3.2 to include an obstacle from above and from below.

Remark 6. *The solution of Equation* (3.12) *can also be viewed as the minimizer of*

$$\mathcal{J}(v) = \int_{\mathbb{R}^d \setminus K} \frac{1}{2}|\nabla v|^2 - \chi_{\Omega_0} v + |v| \, dx,$$

over the admissible set,

$$\mathcal{V}_t = \{v \in (H^1 \cap L^1)(\mathbb{R}^d \setminus K) : v = t \quad on \quad \partial K\}.$$

Let us call

$$\tilde{u}(\cdot, t) = \text{argmin}\{\mathcal{J}(v) : v \in \mathcal{V}_t\},$$

then as in the Section 3.1, $\tilde{u}$ will be the same as the solution of the obstacle problem (3.12) *obtained as the infimal non-negative supersolution. Simulations based on this observation yield similar results to those of the penalized energy.*

3.4 Numerical Method

For the numerical method, we employ the energy minimization formulation and discretize the energy using a uniform fixed grid on a given domain. The energies are convex, so we construct an algorithm via [56] to decouple the problem into an explicit part and a strictly convex part. In the explicit part, the optimal value can be computed directly using shrink-like operators. For the strictly convex part, we can use either a conjugate gradient method or an accelerated gradient descent method to quickly solve the subproblem. The detailed algorithm and its construction are described here. Note that for each problem, there could be slight variations in the algorithm, which we will explain in each subsection.

For a domain $\Omega \subset \mathbb{R}^d$, we approximate functions $u \in H^1(\Omega) \cap L^1(\Omega)$ by taking an N-point uniformly spaced discretization of the domain and defining the discrete approximation of u by $u_h \in \mathbb{R}^N$, where $h > 0$ is the space step. The discrete spatial gradient is defined by

$$\nabla_h u_h := \left[\frac{(u_h)_{i+1,j} - (u_h)_{i,j}}{h}, \frac{(u_h)_{i,j+1} - (u_h)_{i,j}}{h} \right]$$

but any standard finite difference scheme can be used. Consider the following discrete energy

$$\min_{u \geq \varphi} F(\nabla_h u),$$

where $F : \mathbb{R}^{Nd} \to \mathbb{R}$ is a convex function of its input. For simplicity we drop the subscript h on the functions. To solve this problem, we first convert it into an unconstrained problem by using the penalty method:

$$\min_u F(\nabla_h u) + \mu \|(\varphi - u)_+\|_{L^1},$$

for some parameter $\mu > 0$. Here the L^1 norm, $||\cdot||_{L^1}$, is approximated using the quadrature rule

$$\int_\Omega v\,dx = \sum_{j=1}^{N} v_j\,h^d.$$

Since $(\cdot)_+$ is not differentiable, we construct an equivalent minimization problem using [56]. We first introduce an auxiliary variable $v = \varphi - u$ then apply the Bregman iteration:

$$\begin{cases} (u^{k+1}, v^{k+1}) & = \underset{u,v}{\operatorname{argmin}} F(\nabla_h u) + \mu\,|v_+| + \frac{\lambda}{2}||v - \varphi + u + b^k||_2^2, \\ b^{k+1} & = b^k + u^{k+1} + v^{k+1} - \varphi. \end{cases}$$

Now we can efficiently solve the minimization by splitting it into two subproblems with respect to u and v:

$$\begin{cases} \text{Step 1:} \quad u^{n+1} & = \underset{u}{\operatorname{argmin}}\, \mathcal{F}(u) = F(\nabla_h u) + \frac{\lambda}{2}||v^n - \varphi + u + b^n||_2^2, \\ \text{Step 2:} \quad v^{n+1} & = \underset{v}{\operatorname{argmin}}\, \mu\,|v_+| + \frac{\lambda}{2}||v - \varphi + u^{n+1} + b^n||_2^2. \end{cases}$$

The solution for v is given explicitly:

$$v = S_+\left(\varphi - u^{n+1} - b^n,\ \frac{\mu}{\lambda}\right),$$

where $S_+(z, c) := (z - c)$ if $z > c$, z if $z < 0$, and 0 otherwise.

To solve the u subproblem, we consider two cases for the first variation, G, of F. If G is linear, for example taking the continuous functional $F(\nabla u) = \frac{1}{2}\int \nabla u \cdot A\nabla u\,dx$ and A is positive semi-definite, then the first variation is:

$$(\lambda I - \nabla \cdot A\nabla)u = \lambda(\varphi - v^n - b^n),$$

which can be solved by using the conjugate gradient method. In the case where G is non-linear, for example taking the continuous functional $F(\nabla u) = \int \sqrt{1 + |\nabla u|^2}\,dx$, we leverage the strict convexity of the functional to quickly solve the substep by

using Nesterov's acceleration method [110]. The resulting scheme for u is as follows:

$$\begin{cases} w^k & = U^k + \frac{\sqrt{L}-\sqrt{\lambda}}{\sqrt{L}+\sqrt{\lambda}} \left(U^k - U^{k-1}\right) \\ U^{k+1} & = w^k - \tau(G(w^k) + \lambda(v^n - \varphi + w^k + b^n)), \end{cases} \tag{3.14}$$

where $\tau > 0$ is a psuedo-time step, L is the Lipschitz norm of $\mathcal{F}$, and w is an auxiliary variable. This scheme has the following convergence bound:

$$\mathcal{F}(U^k) - \mathcal{F}(U^*) \leq 2\left(1 - \sqrt{\frac{\lambda}{L}}\right)^k \left(\mathcal{F}(U^0) - \mathcal{F}(U^*)\right),$$

where $u^n = U^0$, $u^{n+1} = U^*$, and U^* is the steady state solution of Equation (3.14). Both algorithms are summarized below.

Algorithm (Linear)

Given: u^0, b^0, tol and parameters λ, μ

while $||u^n - u^{n-1}||_\infty > tol$ **do**

$u^{n+1} = (I - \lambda^{-1}G)^{-1}(\varphi - v^n - b^n)$

$v^{n+1} = S_+\left(\varphi - u^{n+1} - b^n, \frac{\mu}{\lambda}\right)$

$b^{n+1} = b^n + u^{n+1} + v^{n+1} - \varphi$

end while

Algorithm (Non-linear)

Given: u^0, b^0, tol and parameters λ, μ

while $||u^n - u^{n-1}||_\infty > tol$ **do**

$U^0 = u^n$

while $||U^k - U^{k-1}||_\infty > tol$ **do**

$w^k = U^k + \frac{\sqrt{L}-\sqrt{\lambda}}{\sqrt{L}+\sqrt{\lambda}} (U^k - U^{k-1})$

$U^{k+1} = w^k - \tau(G(w^k) + \lambda(v^n - \varphi + w^k + b^n))$

end while

$u^{n+1} = U^*$

$v^{n+1} = S_+ \left(\varphi - u^{n+1} - b^n, \frac{\mu}{\lambda}\right)$

$b^{n+1} = b^n + u^{n+1} + v^{n+1} - \varphi$

end while

Obstacle Problem

Given an obstacle $\varphi : \mathbb{R}^d \to \mathbb{R}$ and μ satisfying the condition from Theorem 4, we solve the following obstacle problem:

$$\min_u \int \frac{1}{2}|\nabla u|^2 + \mu(\varphi - u)_+ \, dx.$$

The corresponding discrete problem (where we factor the h^d constant) is given by:

$$\min_u \sum_j \frac{1}{2}|\nabla_h u_j|^2 + \mu(\varphi_j - u_j)_+ \,,$$

where $u_j = u(x_j)$ and $h > 0$ is the uniform grid spacing. Since the functional is quadratic, the Euler-Lagrange equation for the subproblem in terms of u satisfies a Poisson equation:

$$(\lambda I - \Delta_h)u = \lambda(\varphi - v + b), \tag{3.15}$$

with Dirichlet boundary conditions which will be specified for each problem. To approximate the solution of the linear system, we use a few iterations of the conjugate gradient method. It was noted in [56] that full convergence is not necessary within the main iterations, thus we are not required to solve Equation (3.15) exactly.

Two-Phase Membrane Problem

Consider the following optimization problem arising from finding the equilibrium state of a thin film:

$$\min_u \int \frac{1}{2}|\nabla u|^2 + \mu_1 u_+ - \mu_2 u_- \, dx,$$

for some positive and continuous Lipschitz functions $\mu_1(x)$ and $\mu_2(x)$. The corresponding Euler-Lagrange equation is:

$$\Delta u = \mu_1 \chi_{\{u>0\}} - \mu_2 \chi_{\{u<0\}}.$$

The regularity of this problem was studied in [137, 122]. Here we are concerned with the numerical approximation of this problem as well as computing the zero level set.

The corresponding discrete minimization problem is given by:

$$\min_u \sum_j \frac{1}{2}|\nabla_h u_j|^2 + \mu_1 (u_j)_+ - \mu_2 (u_j)_- \, .$$

Now one can apply the split Bregman method by introducing two auxiliary variables $v_1 = u_+$ and $v_2 = u_-$. However, the algorithm can be further simplified by using the following relations:

$$u_+ = \frac{u + |u|}{2}, \quad \text{and} \quad u_- = \frac{u - |u|}{2}.$$

Now we can rewrite the problem as:

$$\min_u \sum_j \frac{1}{2}|\nabla_h u_j|^2 + \alpha u_j + \beta |u_j|,$$

where $\alpha = \frac{\mu_1 - \mu_2}{2}$ and $\beta = \frac{\mu_1 + \mu_2}{2}$. In this form, we have a slightly different numerical scheme. As before, the splitting leads to:

$$\min_u \sum_j \frac{1}{2}|\nabla_h u_j|^2 + \alpha u_j + \beta|v_j| + \frac{\lambda}{2}|v_j - u_j - b_j|^2,$$

and iterative scheme is written as follows

$$\begin{aligned} u^{n+1} &= (\lambda I - \Delta_h)^{-1}(\lambda(v^n - b^n) - \alpha) \\ v^{n+1} &= S\left(u^{n+1} + b^n,\, \frac{\beta}{\lambda}\right) \\ b^{n+1} &= b^n + u^{n+1} - v^{n+1}, \end{aligned}$$

where the shrink function is defined as $S(z, c) := (|z| - c)_+ \operatorname{sign}(z)$.

Hele-Shaw

As described in Section 3.3, we minimize the obstacle problem transformation of the Hele-Shaw flow:

$$\min \int \frac{1}{2}|\nabla u|^2 + \gamma_1(\varphi - u)_+ - \gamma_2(t\chi_K - u)_- \, dx,$$

with φ given by Equation (3.3). The corresponding discretization problem is:

$$\min_u \sum_j \frac{1}{2}|\nabla_h u_j|^2 + \gamma_1(\varphi_j - u_j)_+ - \gamma_2(t\,\chi_{K,j} - u_j)_- \,.$$

Once again, we construct an equivalent minimization problem by introducing two auxiliary variables $v_1 = \varphi - u$ and $v_2 = u - t\chi_K$. For convenience, we drop the subscript j in all terms:

$$\min_{u,v_1,v_2,b_1,b_2} \sum_j \frac{1}{2}|\nabla_h u|^2 + \gamma_1(v_1)_+ + \gamma_2(v_2)_+ + \frac{\lambda_1}{2}(v_1 - \varphi + u + b_1)^2 + \frac{\lambda_2}{2}(v_2 - u + t\chi_K + b_2)^2.$$

The iterative scheme is written as follows:

$$
\begin{aligned}
u^{n+1} &= \left((\lambda_1+\lambda_2)I - \Delta_h\right)^{-1}\left(\lambda_1(\varphi - v_1^n - b_1^n) + \lambda_2(v_2 + t\chi_K + b_2)\right)\\
v_1^{n+1} &= S_+\left(\varphi - u^{n+1} - b_1^n,\ \frac{\gamma_1}{\lambda_1}\right)\\
v_2^{n+1} &= S_+\left(u^{n+1} - t\chi_K - b_2^n,\ \frac{\gamma_2}{\lambda_2}\right)\\
b_1^{n+1} &= b_1^n + v_1^{n+1} - \varphi + u^{n+1}\\
b_2^{n+1} &= b_2^n + v_2^{n+1} - u^{n+1} + t\chi_K.
\end{aligned}
$$

Each substep is either a linear system of equations or an explicit update using the shrink-like operators, making it easy to solve.

3.5 Computational Simulations

In this section, we apply the methods from Section 3.4 to various examples. The iterative schemes stop when the difference between two consecutive iterations in the L^∞ norm is less than a set tolerance, *tol*. We will specify the tolerance parameter for each problem. In general, *tol* is set to be Ch^2, for some constant C. Unless otherwise specified, the solutions are zero on the boundary.

Obstacle Problem

For our first examples, we show some numerical results for the minimization problem:

$$\min_u \int \frac{1}{2}|\nabla u|^2 + \mu(\varphi - u)_+\, dx,$$

with different types of obstacles. In particular, consider the following 1D obstacles:

$$
\varphi_1(x) := \begin{cases} 100x^2 & \text{for} \quad 0 \le x \le 0.25 \\ 100x(1-x) - 12.5 & \text{for} \quad 0.25 \le x \le 0.5 \\ \varphi_1(1-x) & \text{for} \quad 0.5 \le x \le 1.0, \end{cases} \tag{3.16}
$$

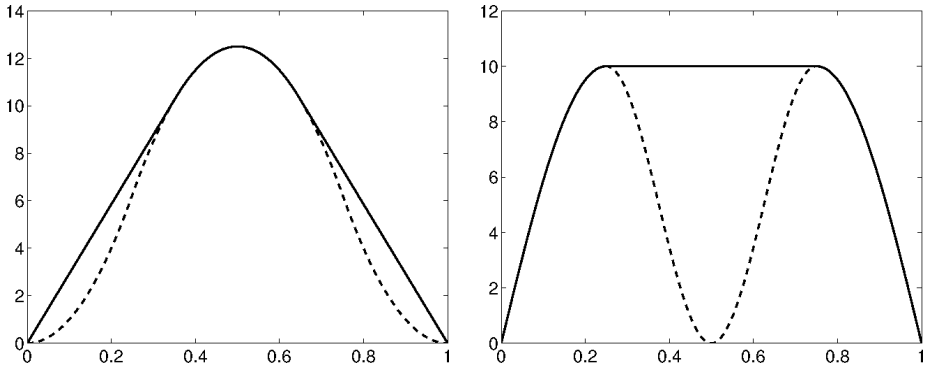


Figure 3.1: The dashed curves are the obstacles and the black ones are our numerical solutions associated with Equations (3.16) (left) and (3.17) (right) after 50 iterations. The grid size is 256, the parameters are $(\mu, \lambda) = (300, 45)$ and $(2.5 \times 10^4, 250)$, respectively.

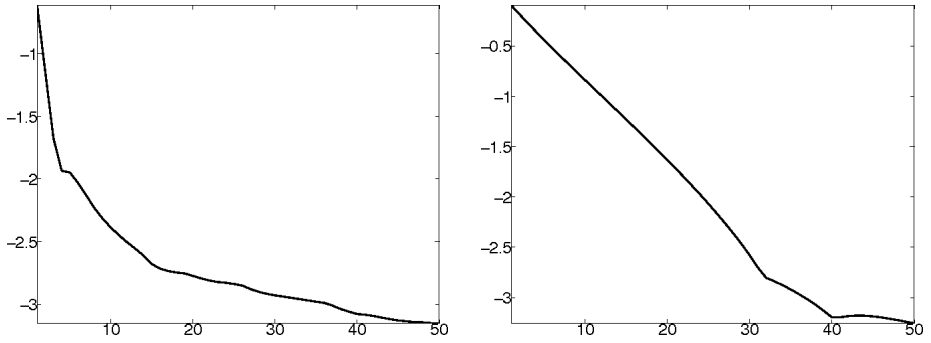


Figure 3.2: The plots show the error between our numerical solutions (from Fig. 3.1) and the analytic solutions versus number of iterations. More precisely, the error is measured by $E^n = ||u_h^n - u_{exact}||_{L^\infty}$ and we plot $\log(E^n)$ versus n.

and

$$\varphi_2(x) := \begin{cases} 10\sin(2\pi x) & \text{for} \quad 0 \le x \le 0.25 \\ 5\cos(\pi(4x-1)) + 5 & \text{for} \quad 0.25 \le x \le 0.5 \\ \varphi_2(1-x) & \text{for} \quad 0.5 \le x \le 1.0. \end{cases} \tag{3.17}$$

In both cases, the parameter μ is determined discretely (see Theorem 4) and u is initialized using the obstacle, *i.e.* $u^0 = \varphi$. The results are shown in Fig. 3.1. In both cases the numerical solutions are linear away from their corresponding

obstacles, which agrees with the analytic solutions:

$$u_{1,exact}(x) = \begin{cases} (100 - 50\sqrt{2})x & \text{for} \quad 0 \le x \le \frac{1}{2\sqrt{2}} \\ 100x(1-x) - 12.5 & \text{for} \quad \frac{1}{2\sqrt{2}} \le x \le 0.5 \\ u_{1,exact}(1-x) & \text{for} \quad 0.5 \le x \le 1.0, \end{cases}$$

and

$$u_{2,exact}(x) = \begin{cases} 10\sin(2\pi x) & \text{for} \quad 0 \le x \le 0.25 \\ 10 & \text{for} \quad 0.25 \le x \le 0.5 \\ u_{2,exact}(1-x) & \text{for} \quad 0.5 \le x \le 1.0. \end{cases}$$

These simple examples are used to verify the numerical convergence of our method. The errors between the analytic solutions and the numerical solutions versus the number of iterations associated to obstacle problems (3.16) and (3.17) are shown in Fig. 3.2. Notice that the numerical scheme has nearly exponential error decay in the beginning.

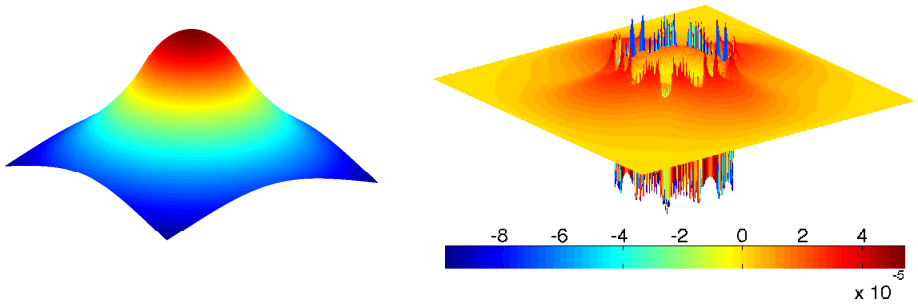


Figure 3.3: The plots above are our numerical solution (left) and the difference with the analytic solution (right) associated with Equation (3.18). The grid size is 256 by 256, the parameters are $(\mu, \lambda) = (10/h^2, 20.3)$, and $tol = 10^{-6}$.

Next, we consider a 2D problem on the domain $\Omega = [-2, 2] \times [-2, 2]$ with the

following obstacle:

$$\varphi(x,y) = \begin{cases} \sqrt{1-x^2-y^2}, & \text{for} \quad x^2+y^2 \leq 1 \\ -1, & \text{otherwise.} \end{cases} \tag{3.18}$$

Since the obstacle is radial symmetric, the analytical solution can be solved directly:

$$u(x,y) = \begin{cases} \sqrt{1-x^2-y^2}, & \text{for} \quad r \leq r^* \\ -(r^*)^2 \log(r/2)/\sqrt{1-(r*)^2}, & \text{for} \quad r \geq r^*, \end{cases}$$

where $r = \sqrt{x^2+y^2}$, and r^* is the solution of

$$(r^*)^2(1-\log(r^*/2)) = 1.$$

Our numerical solution and the difference with the analytic solution are presented in Fig. 3.3. For comparison see [98]. We can see that the error is concentrated along the contact set, where the function is no longer C^2, and is relatively small everywhere else.

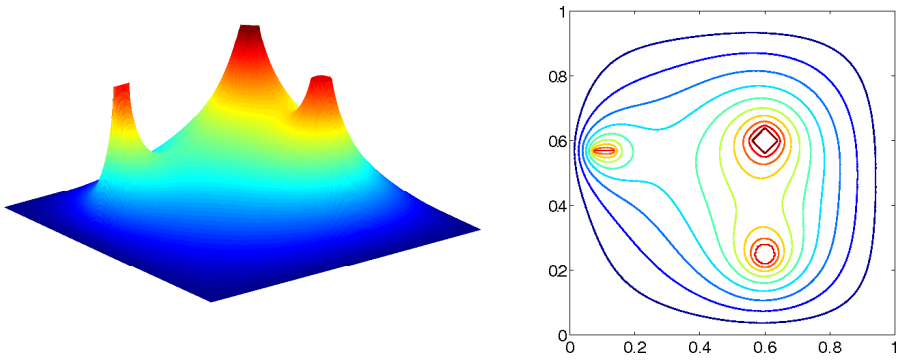


Figure 3.4: The plots above are our numerical solution (left) and its level curves (right) of the obstacle problem associated with Equation (3.19). The grid size is 256 by 256, the parameters are $(\mu, \lambda) = (6.5 \times 10^5, 1.3 \times 10^4)$, and $tol = 5 \times 10^{-4}$.

Next, to examine the behavior of non-smooth obstacles, we consider:

$$\varphi_3(x,y) = \begin{cases} 5.0, & \text{for} \quad |x-0.6| + |y-0.6| < 0.04 \\ 4.5, & \text{for} \quad (x-0.6)^2 + (y-0.25)^2 < 0.001 \\ 4.5, & \text{for} \quad y = 0.57 \ \text{ and } \ 0.075 < x < 0.13 \\ 0, & \text{otherwise} \end{cases} \tag{3.19}$$

which consists of different disjoint shapes inside the domain $[0,1] \times [0,1]$. The numerical result and its level curves are shown in Fig. 3.4. One can see that the solution is smooth away from the obstacle and agrees well with the obstacle on its support set.

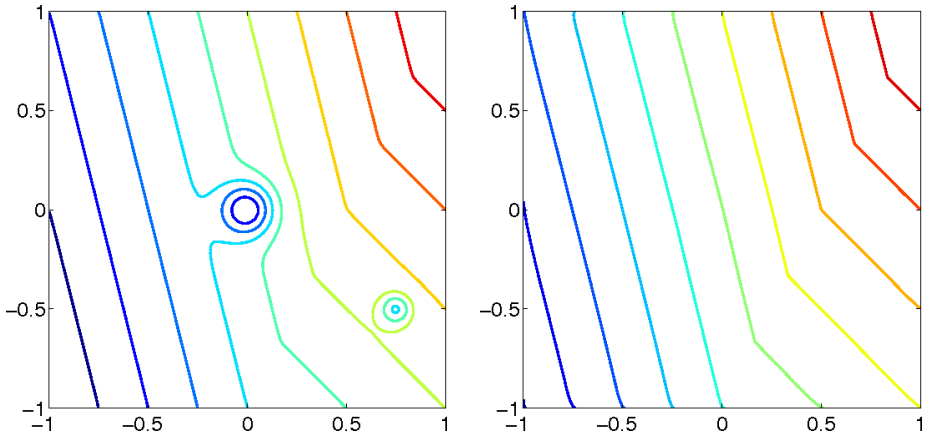


Figure 3.5: The plots above are the level curves of the obstacle (left) and our numerical solution (right) associated with Equation (3.20). The grid size is 256 by 256, the parameters are $(\mu, \lambda) = (10^5, 5 \times 10^3)$, and $tol = 5 \times 10^{-4}$.

Finally, in Fig. 3.5, an obstacle consisting of two intersect planes with a bump on each plane in the domain $[-1,1] \times [-1,1]$ is examined:

$$\varphi_4 = \min(x+y-2, 2x+0.5y-2.5) - 2e^{-60(x^2+y^2)} - 1.5e^{-200((x-0.75)^2+(y+0.5)^2)} \tag{3.20}$$

In this case, the solution agrees with the obstacle in a large portion of the domain. The analytic solution is given by the two intersecting planes, which can be seen by the linear level curves.

Comparison

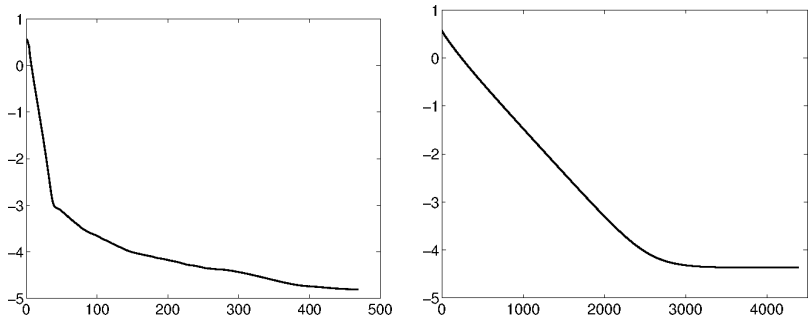


Figure 3.6: The plots correspond to the log error between analytic solution and numerical solution computed by our method (left) and method from [93] (right) vs. the number of iterations. Both algorithms are applied to Equation (3.18) on a grid of size 256 by 256, with the error measured in the L^2 norm.

We compare our method to the one found in [93], which uses an indicator function to enforce the obstacle inequality. The addition of an indicator function to the variational problem may be found in many of the models cited in the introduction of this paper. We choose to compare directly to the method from [93], since their algorithm also uses operator splitting (in particular the Douglas-Rachford algorithm) which is closely related to the algorithm we use here. Also, their method can be seen as a representative for indicator based methods found in the literature.

In Table 3.1, we calculate the lowest achievable errors for our method and the method in [93] applied to the problem found in Figure 3.4. The errors for our method, in the L^2 and L^∞ norm, are lower than the errors for the method used in [93]. Additionally, our method seems to converge at a faster rate. This is likely due to the differences in continuity between the penalty functions. In Table 3.2 the time and iteration complexity are examined. Note that although the complexities are similar, which is expected for these two algorithms, the absolute CPU times and number of iterations are lower in our method.

Table 3.1: The errors for our method and the method found in [93] are calculated for the obstacle in Figure 3.4. The error for our method, in the L^2 and L^∞ norm, is lower than the error for the method used in [93].

Grid Size	L^2 Error (Our)	L^2 Error ([93])	L^∞ Error (Our)	L^∞ Error ([93]))
64×64	6.83×10^{-4}	1.06×10^{-3}	1.21×10^{-3}	1.21×10^{-3}
128×128	1.11×10^{-4}	2.29×10^{-4}	2.71×10^{-4}	3.36×10^{-4}
256×256	1.56×10^{-5}	4.26×10^{-5}	4.7×10^{-5}	7.34×10^{-5}
Order	2.7	2.3	2.3	2.0

Table 3.2: The CPU times and number of iterations for our method and the method found in [93] to achieve the errors in Table 3.1 are shown here. The complexity is measured as a function of the number of nodes used in the discretization.

Grid Size	Time (ours)	Time ([93])	Iterations (ours)	Iterations ([93])
64×64	20.8 sec	55.1 sec	315	3508
128×128	57.7 sec	193.7 sec	337	3997
256×256	222.3 sec	1158.3 sec	469	4383
Complexity	0.85	1.1	0.14	0.08

Also, the error between the analytic solution and the numerical solution computed by both methods versus the number of iterations are shown in Figure 3.6. Our method quickly achieves a relatively low error (10^{-3}) compared to method in [93]. A partial explanation for this fast initial decay can be found in [56]. The more rapid convergence of our method might be due to the "error forgetting" property of L^1 regularization combined with Bregman iteration, which is analyzed in [160].

Nonlinear Obstacle

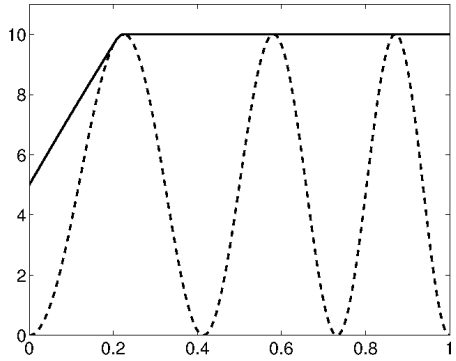


Figure 3.7: The dashed curve is the obstacle and the black one is the numerical solution of the nonlinear obstacle problem (Equation (3.21)). The grid size is 512, $\lambda = 5.3, \mu = 1.1 \times 10^3, \tau = 1/L = h^2/2$.

We would like to show that the methodology here can be easily applied to nonlinear problems, so as a proof of concept we minimize the surface tension:

$$\min_{v \geq \varphi} \int_0^1 \sqrt{1 + |\nabla v|^2}\, dx,$$

which is the energy associated with the classical model of stretching an elastic membrane over a fixed obstacle. The obstacle φ is given by the oscillatory function:

$$\varphi = 10 \sin^2(\pi(x+1)^2), \quad x \in [0,1]. \tag{3.21}$$

The boundary data for this example is taken to be $u(0) = 5$ and $u(1) = 10$. The numerical solution is linear away from the contact set as can be seen in Fig. 3.7.

Two-phase Membrane Problem

We examine the two-phase membrane problem:

$$\min_u \int \frac{1}{2}|\nabla u|^2 + \mu_1 u_+ - \mu_2 u_-\, dx,$$

with different sets of (μ_1, μ_2) and boundary conditions. First, in the symmetric case, we consider the following 1D problem:

$$u'' = 8\chi_{\{u>0\}} - 8\chi_{\{u<0\}} \quad \text{with} \quad u(1) = 1, \quad u(-1) = -1, \tag{3.22}$$

whose analytic solution is given by:

$$u(x) = \begin{cases} -4x^2 - 4x - 1 & \text{for} \quad -1 \le x \le -0.5 \\ 0 & \text{for} \quad -0.5 \le x \le 0.5 \\ 4x^2 - 4x + 1 & \text{for} \quad 0.5 \le x \le 1. \end{cases}$$

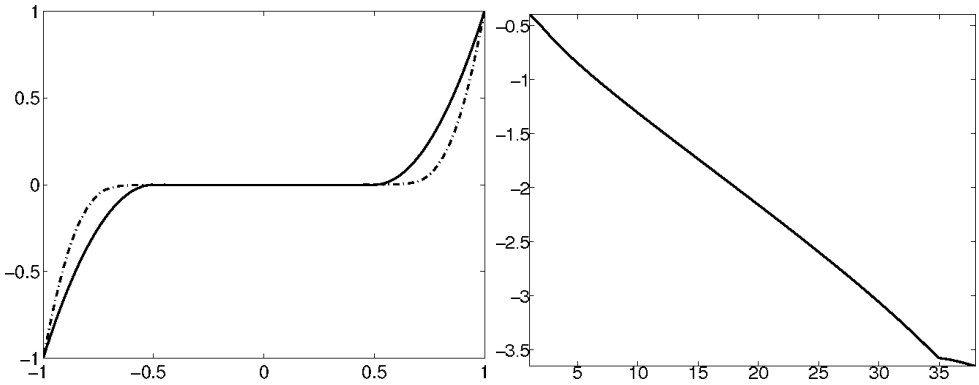


Figure 3.8: Left: The black curve is the final numerical solution, the dotted one is the numerical solution after 3 iterations of the two-phase membrane associated with Equation (3.22). The grid size is 512, $\lambda = 204.8$ and $tol = 5 \times 10^{-5}$. Right: The log error in L^∞-norm between the numerical and the analytic solutions.

In Fig. 3.8 (left), we plot our numerical solution at the third iteration and the final state. Within a few iterations, our numerical method is able to locate the correct zero set. The error versus the number of iterations is shown in Fig. 3.8 (right), and converges nearly exponentially. For comparison of the numerical results, see [17].

Next we consider a non-symmetric equation:

$$u'' = 2\chi_{\{u>0\}} - \chi_{\{u<0\}} \quad \text{with} \quad u(1) = 1, \quad u(-1) = -1, \tag{3.23}$$

The calculated free boundary is at the point $x \approx 0.141$, which was also observed in [137] (see Fig. 3.9).

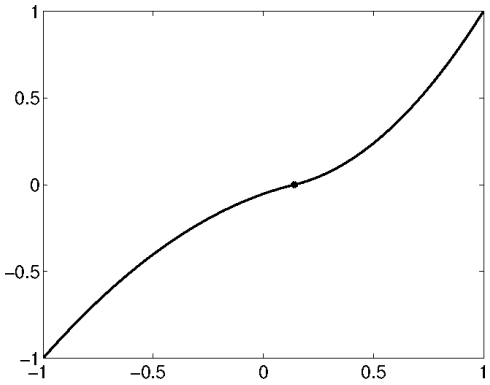


Figure 3.9: The plot above is our numerical solution of the two-phase membrane associated with Equation (3.23). The free boundary point (the dot) is located at $x \approx 0.141$. The grid size is 2^{12}, $\lambda = 3072$, and $tol = 5 \times 10^{-7}$.

For an example in 2D, we set $\mu_1 = \mu_2 = 1$ with Dirichlet boundary condition g given by:

$$g(x,y) = \begin{cases} (1-x)^2/4 & -1 \le x \le 1 \text{ and } \ y = 1 \\ -(1-x)^2/4 & -1 \le x \le 1 \text{ and } \ y = -1 \\ y^2 & 0 \le y \le 1 \text{ and } \ x = -1 \\ -y^2 & -1 \le y \le 0 \text{ and } \ x = -1 \\ 0 & -1 \le y \le 1 \text{ and } \ x = 1. \end{cases} \tag{3.24}$$

In this case, the zero set has non-zero measure, see Fig. 3.10 (right). The boundary between the regions $\{u > 0\}$, $\{u < 0\}$ and $\{u = 0\}$ contains a branching point, which we are able to resolve numerically.

Hele-Shaw

We present three examples of the Hele-Shaw problem:

$$\min_u \int \frac{1}{2}|\nabla u|^2 + \gamma_1(\varphi - u)_+ - \gamma_2(t\chi_K - u)_- \, dx,$$

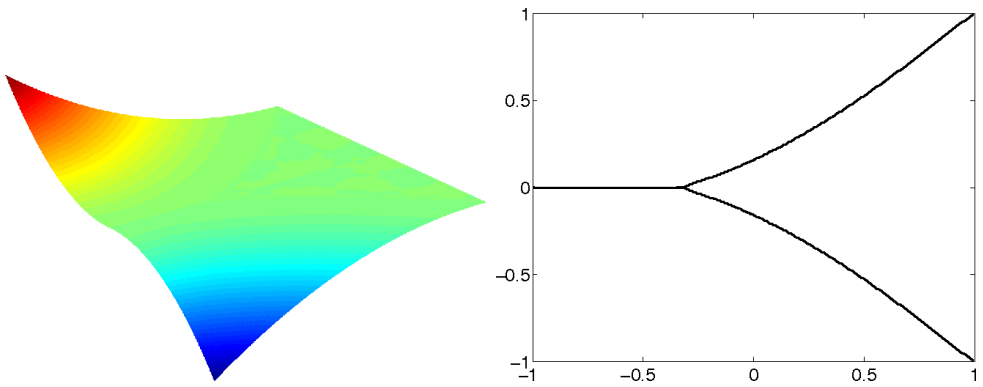


Figure 3.10: Left: Our numerical solution associated with Equation (3.24). Right: the boundaries between the regions $\{u > 0\}$ (top), $\{u < 0\}$ (bottom) and $\{u = 0\}$. The grid size is 256 by 256, $\lambda = 100.0$ and $tol = 10^{-6}$.

with different sets of (K, Ω_0). The parameters are fixed at $\gamma_1 = \gamma_2 = 1.5\times10^4$, $\lambda_1 = \lambda_2 = 150$. The free boundary starts moving from Ω_0.

To validate our numerical scheme, in the first example we compare the boundary of our numerical solution and that of the analytic solution. In particular, when both K and Ω_0 are circles centered at the origin:

$$K = \{(x, y) \in [-5, 5]^2 \mid x^2 + y^2 \leq 1\}, \quad \Omega_0 = \{(x, y) \in [-5, 5]^2 \mid x^2 + y^2 \leq 2\}, \tag{3.25}$$

the free boundary remains a circle centered at the origin for all time. Thus the radius of Ω, R_{exact}, can be calculated explicitly. In Table 6.1, we compute the error between the radius of the free boundary of our numerical solution and the analytic solution at time $t = 0.25$ using different grid sizes. The experimental error in the radius is about $\mathcal{O}(h^{0.8})$, which is expected for a low dimensional structure.

Next, we present two numerical results for more complicated cases of (K, Ω_0). In Fig. 3.11 (left), the free boundary $\partial\Omega$ is pinned at the two acute vertices along $\partial\Omega_0$. As expected, the free boundary opens up to right angles then smooth out and move away from Ω_0. For more details on this short time behavior as well as singularities in the Hele-Shaw model see [71, 145, 112]. Finally, in Fig. 3.11

Table 3.3: The error between the radius of the free boundary of our numerical solution and the exact radius associated with (K, Ω_0) defined in Equation (3.25) at time $t = 0.25$. The parameters are fixed at $\gamma_1 = \gamma_2 = 1.5 \times 10^4, \lambda_1 = \lambda_2 = 150$, $tol = 10^{-6}$. The convergence rate is approximately $\mathcal{O}(h^{0.8})$.

Grid Size	128×128	256×256	512×512	1024×1024
Error (radius)	0.0238	0.0124	0.0083	0.0044

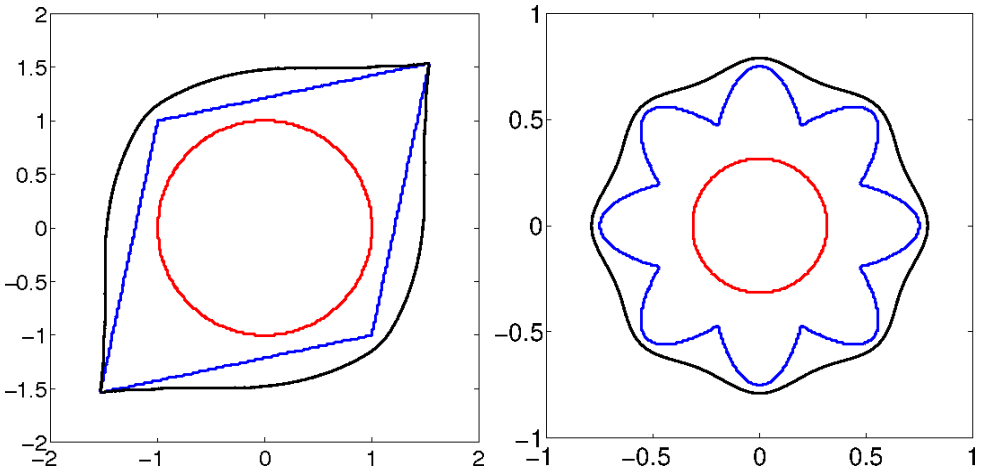


Figure 3.11: From inside to outside: boundaries of the sets K, Ω_0 and the free boundary of the Hele-Shaw problem. The grid size is 256 by 256, $\gamma_1 = \gamma_2 = 1.5 \times 10^4, \lambda_1 = \lambda_2 = 150, tol = 10^{-5}$, and time $t = 0.1$ and $t = 0.06$, respectively.

(right), we take the boundary of Ω_0 to be smooth but concave. The free boundary moves away from the initial state and begins to smooth out.

Conclusion

Using an L^1-penalty method, we are able to construct an unconstrained problem whose solutions correspond exactly to those of the obstacle problems. We provide a lower bound on the value of the penalty parameter and use this to guide

our numerical calculations. We present several experiment results showing the applicability of our method to various physical problems.

Part II

Non-local Operators and Retinex

CHAPTER 4

A Short Review of Retinex Implementations

Introduction

Retinex is a theory on the human visual perception [82, 86, 85]. It was an attempt at explaining how a combination of processes supposedly taking place both in the retina and the cortex is capable of adaptively coping with illumination that varies spatially. The fundamental observation is the insensitivity of human visual perception with respect to a slowly varying illumination on a Mondrian-like scene, see Figure 4.1.

The original Retinex computational algorithm was a model of human vision. The input was the array of scene *radiances*. The algorithm's output was the array of calculated *appearances*. The experimental data used in [86] was a Black and White Mondrian (a flat array of uniform reflectance rectangles in a smooth gradient of illumination). Here, appearance does not correlate with radiance. Since gradients in illumination are nearly invisible, appearance shows moderately good correlation with *reflectances*. More generally, however, appearance correlates with edge ratios in radiance, not surface reflectance. That is not a problem, because the original Retinex goal is to calculate *appearances*. 3-D Mondrians [105] use uniform reflectance surfaces in non-uniform illuminations that include edges in illumination. Here, there is very little average correlation with appearance and surface reflectance. Some areas show high correlation; other show no correlation. Land and McCann's model of human vision calculates *appearance* both when it

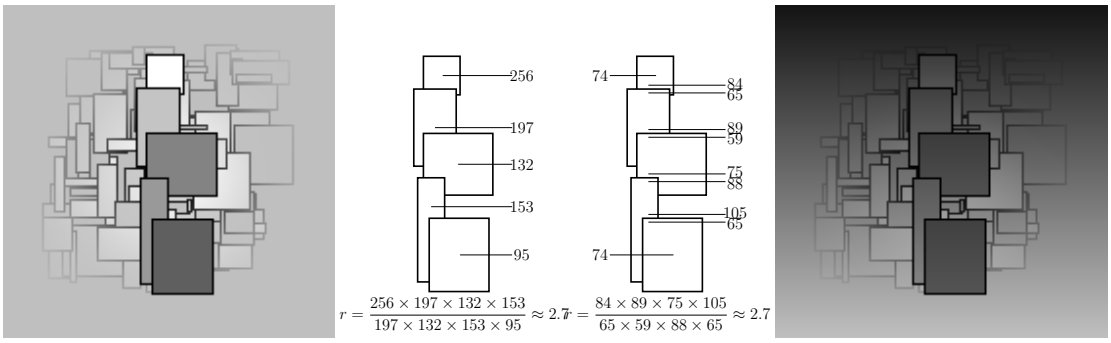


Figure 4.1: **Mondrian illusion**. **Left**: Under uniform illumination, the relative intensity ratio between the top-most and bottom-most patch in this synthetic "Mondrian" image is obvious. **Right**: Albeit these patches have the same absolute gray-level under smoothly varying illumination, their underlying relative reflectance ratio, 2.7, is still perceived by humans, and it can be recovered by multiplying the local intensity ratios at the discontinuities along any path joining the two patches, effectively discarding gradients due to lighting.

correlates with reflectance, and when it does not. Land and McCann's Retinex is not intended, however, to identify *illumination* (nor *reflectance*).

Inspired by the initial ideas of Retinex, other researchers picked up the model, but the goal shifted towards reflectance-illumination decomposition (given the connection with the original Land and McCann work, it is also coined the "Retinex-problem"), contrast enhancement, and image aesthetics. It is important to note that this new goal differs from the original Retinex motivation, and from the human visual system. Moreover, the aspired reflectance-illumination separation does not even have to be physically plausible, depending on the requirements and priors of the specific decomposition application. For the rest of the thesis, the terms illumination and reflectance, are thus greatly relaxed with respect to the strict physical definitions.

Among models for enhancement and noise removal—such as [152, 88, 129, 130],

for an overview see [1]—Retinex has received particular attention because it is deeply connected with psycho-visual empirical evidence of human perception. However, depending on the application, the various Retinex assumptions are given different importance, and resulting implementations vary significantly. In this thesis, we even include shadow-removal as potential Retinex-application. Important differences are also found in the respective pre- and post-processing steps for each application, such as calibration or gamma correction. For an overview, we refer to [106]. In this thesis, we focus on the more recent interpretation of the "Retinex-problem" and largely refrain from extensive pre- and post-processing.

We tentatively classify models and algorithms in five main classes: original Retinex formulation, threshold-based PDE-models, reset-based random walk and kernel-based methods, reset- and threshold-free center-surround models, variational Retinex. While fundamental connections and equivalences have been shown between random-walk, threshold, and the original Retinex formulation [68, 108], a divide still exists between recent kernel-based Retinex such as [10] and variational models like [76]. We will show in Chapter 5 how threshold-based, kernel-based, and center-surround Retinex can be expressed as variational models by making use of non-local generalizations of differential operators using particular kernels. It is to note that similar connections between variational denoising models, PDE schemes, and short-time kernels have been established in [141].

Part II of this thesis was taken with slight modification from the article "Non-local Retinex - a unifying framework and beyond", published in *SIAM Journal on Imaging Sciences*, 8(2), 787826.

Original Retinex Algorithm

Land formalized the reflectance ratios, by summing thresholded log-ratios over continuous paths between two pixels [83]. He defines the relative reflectance of

pixel i to j as (see figure 4.2):

$$R(i,j) = \sum_k \delta_\tau \log \frac{I_{k+1}}{I_k} \qquad \text{with} \qquad \delta_\tau(\cdot) := \begin{cases} \cdot & \text{if } |\cdot| \geq \tau \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

where δ_τ denotes hard thresholding, and I_k is the intensity of the k-th pixel along the path from i to j. The thresholding makes sure, that only sharp intensity transitions are included in the sum, and the slow drift due to smoothly varying illumination is eliminated. Provided that the thresholding yields perfect distinction between illumination gradients and actual feature edges, the relative reflectance of i to j is invariant to the path connecting them. The average relative reflectance at i is then estimated as

$$\bar{R}(i) = \mathbb{E}_j \left[R(i,j) \right] = \frac{1}{N} \sum_{j=1}^{N} R(i,j) \tag{4.2}$$

However, "the ultimate purpose is to describe any area by relating its reflectance to a single, standard, high reflectance somewhere in the Mondrian or to several equally high reflectances" [86]. Instead of localizing the highest reflectance in a preprocessing step, which seemed biologically unplausible, it was proposed to estimate the maximum reflectance directly while performing the sequential sum along each path. Indeed, whenever the intermediate sequential sum from j up to $k+1$, i.e. the relative reflectance of pixel $k+1$ to j, becomes positiv—equivalent to a sequential product bigger than 1—, one has reached a new maximum reflectance, and the sequential sum is reset, with I_{k+1} as new reference intensity. Due to the presence of the thresholding operator, the final reference pixel does not necessarily coincide with the brightest pixel along the path. For a mathematical definition and analysis of this reset mechanism, see [125].

There has been quite some debate about the respective role and importance of both threshold and reset in the Retinex, including by McCann himself [104]. The criterion will serve us dividing the many Retinex implementations in two broad classes: threshold-based versus reset-based. A third class of implementations is

a) Relative reflectance of i to j b) Average reflectance of i

Figure 4.2: **Relative and average reflectance**. **a)** The relative reflectance of pixel i to pixel j is computed by accumulating all *important* local intensity gradients along any path γ joining these pixels. **b)** The average reflectance of i is then obtained as the average relative reflectance of i to different j.

based on an alternative technique proposed by Land, which determines lightness as ratio of the local intensity compared to the average intensity of its immediate (circular) surroundings, without neither thresholding, nor reset [84]. A forth class, finally, extracts the reflectance and illumination information variationally, by optimizing different energy functionals.

Threshold-based Retinex Implementations (PDE)

In 1974, Horn proposed a mathematical alternative to the Retinex algorithm that differs substantially in form [66]. He essentially stripped the Retinex algorithm down to a smoothness prior on the illumination field, and thus to a thresholding on intensity derivatives. He poses the problem of recovering the underlying reflectance R, which multiplied by the illumination B resulted in the observed I:

$$I(x, y) = B(x, y)R(x, y) \tag{4.3}$$

By taking the logarithm, $i := \log(I)$ etc., we obtain an additive impact of illumination:

$$i(x, y) = b(x, y) + r(x, y) \tag{4.4}$$

Since the illumination b is supposed to be varying smoothly, the spatial derivatives of the observed intensity are mostly due to edges in the reflectance r. Horn considered the scalar Laplacian operator: Δb will be finite everywhere, while Δr will be zero except at each edge separating regions [66]. Therefore, discarding the finite parts of the observed intensity Laplacian is supposed to yield the Laplacian of the reflectance (Poisson equation):

$$\Delta r = \delta_\tau \Delta i \tag{4.5}$$

A tight mathematical connection between Land's and Horn's computations, on the basis of Green's formula, has been shown in work by Hurlbert [68]. A fully discrete alternative to Horn's convolution and inversion scheme was proposed by [102]. There, an equivalent two-stage algorithm was introduced, which achieves inversion at very low computational cost, feasible in terms of neural networks. The steps are first local contrast computation and thresholding (δ_τ), then inversion, as follows:

$$c(x) = \delta_\tau \left(i(x) - \sum_{y \neq x} w(x,y) i(y) \right) \tag{4.6}$$

$$r(x) = c(x) + \sum_{y \neq x} w(x,y) r(y) \tag{4.7}$$

where $w(x,y)$ are the weights associated with the finite differences stencil for the discrete Laplacian operator.

While the first step is straightforward, the second step realizes inversion through feedback in terms of neural networks: $r(y)$ is Retinex output *computed in parallel*, and the system (hopefully) converges to a steady state solution.

Horn's model has been strongly backed up by a much more recent paper by Morel [108], where the authors show a very tight connection between Horn's Laplacian thresholding and Land's original, resetless Retinex algorithm. Indeed, "*if* the Retinex paths are interpreted as symmetric random walks, then Retinex is equivalent to a Neumann problem for a linear Poisson equation" [108]. The main

difference between Horn and Morel concerns the argument of the hard thresholding operator: while Horn thresholds the scalar Laplacian, Morel thresholds the components of the gradient prior to computing the divergence. *De facto*, Morel thus effectively solves an L^2-gradient fitting problem:

$$\hat{r} = \underset{r}{\operatorname{argmin}} \left\{ \|\nabla r - \delta_\tau \nabla i\|_2^2 \right\} \tag{4.8}$$

We refer to this basic model as L^2-*Retinex*. Note that reconstruction from thresholded gradient has earlier been proposed by Blake [14, 15, 18]. More recently, the L^1-equivalent thresholded-gradient fidelity Retinex has been proposed: The L^1-*Retinex* minimizes the isotropic L^1-distance [95]:

$$\hat{r} = \underset{r}{\operatorname{argmin}} \left\{ \|\nabla r - \delta_\tau \nabla i\|_1 \right\} \tag{4.9}$$

Reset-based Retinex Implementations (Random Walk)

Moving away from thresholding and relying purely on the reset mechanism, Frankle and McCann have patented their Retinex algorithm [49]. The Frankle-McCann algorithm replaces sequential products along paths by pairwise pixel ratios sampled along discrete spirals. Long-distance interactions are computed first, then the sampling progressively approaches the center pixel while decreasing the spacing. At each step, the lightness estimate is updated with a ratio-product-reset-average operation [52]. More recent variants of the algorithm mainly involve multiresolution image pyramids [103, 52], different sampling patterns [161, 69], or ratio modifiers [140].

Two of the main drawbacks of the Frankle-McCann algorithm are the strong dependence on the path length of the spiral pattern (represented as number of iterations), and the appearance of asymmetric halos due to the anisotropic sampling pattern. In order to avoid these issues, Provenzi et al. replace the path-based sampling pattern by a repeated sampling through random sprays [123]. Indeed,

if the threshold is removed from the Retinex formulation, then the reset reduces the relative reflectance, computed using a specific path γ_k, to the ratio of central pixel $I(i)$ and brightest pixel $I(x)$ along that same path γ_k [125]:

$$R(i,j) = \log \frac{I(i)}{I(x)}$$
$$R(i) = \frac{1}{N} \sum_{k=1}^{N} \log \frac{I(i)}{I(x_k)}$$

where x_k is the brightest pixel encountered along γ_k. Therefore, many paths become redundant, and the maxima $I(x_k)$ can be sampled alternatively. It was suggested that averaging repeated random spray sampling of x_k directly, with radially decreasing sampling density, can substitute the path-based filtering efficiently. On the downside, such sampling of extrema exhibits high variance, and the recovered reflectances are typically noisy. More robust estimators, such as high-percentage quantiles rather than the most extreme sample, results in smoother illumination estimation and therefore lower noise.

Beyond, the (white-patch) random spray Retinex was combined with a (gray-world) model used for automatic color equalization (ACE) [129, 124]. Eventually, the random spray sampling was replaced by a kernel, representing the sampling density of the random spray in the limit case [10, 120]:

$$R(i) = \sum_{j:I(j)\geq I(i)} w(i,j) f\left(\frac{I(i)}{I(j)}\right) + \sum_{j:I(j)<I(i)} w(i,j)$$

where $w(i,j)$ is the kernel, representing the probability density of picking a pixel j in the neighborhood of i [10]. Here, we find the ratio modifier f as a generalization of the log-ratio seen until now. Such a ratio-modifier has previously been introduced by Sobol [140].

Center-surround Retinex Implementations

A simple alternative to threshold/reset based Retinex algorithms was proposed by Land based on findings of lateral inhibition [84]. The alternative consists in determining the local lightness (reflectance) as the ratio between local intensity and an average of its close surroundings. Land realized that this much simpler model could reproduce all Retinex behavior modeled so far, and had the additional competence of generating empirically perceived Mach bands [128, 43].The fundamental idea is again that the low-frequency components are due to illumination, while the high-frequency details are features in the reflectance.

10 years later, only, the idea was picked up and formulated as single- and multi-scale center-surround Retinex [74, 73, 127]. The single-scale Retinex is given by

$$R(i) = \log I(i) - \log \left[F * I\right](i) \tag{4.10}$$

where F is a Gaussian kernel. The multi-scale Retinex is then simply the combination of different single-scale retinexes:

$$R(i) = \log I(i) - \sum_n w_n \log \left[F_n * I\right](i) \tag{4.11}$$

where $w_n : \sum_n w_n = 1$ are the weights of each scale, and F_n are Gaussian kernels of different scale.

Starting from the single-scale retinex (4.10), by changing the order of log and Gaussian convolution, one gets homomorphic filtering

$$R(i) = \log I(i) - \left[F * \log I\right](i) \tag{4.12}$$

which in turn can be identified as a special case of (resetless) kernel-Retinex, with the kernel $w(i,j) \equiv F$ and ratio modifier $f \equiv \log$:

$$R(i) = \sum_j w(i,j) \log\left(\frac{I(i)}{I(j)}\right) = \log I(i) - \sum_j w(i,j) \log I(j) \tag{4.13}$$

Variational Retinex

A whole family of variational Retinex models handles the regularity priors on the reflectance and illumination parts of the Retinex decomposition in a more explicit way. First of its kind, the variational framework by Kimmel introduces competing H^1 smoothness priors on both the illumination and reflectance fields, as well as a quadratic fidelity prior between illumination and observed intensity [76]. In addition, illumination is constrained to be bigger or equal to observed intensity, i.e. the reflectance is limited by an upper bound:

$$\min_b \left\{ \int_\Omega |\nabla b|^2 + \alpha(b-i)^2 + \beta|\nabla b - \nabla i|^2 dxdy \right\} \quad \text{s.t.} \quad b \geq i, \quad \langle \nabla b, \vec{n} \rangle = 0 \text{ on } \partial\Omega. \tag{4.14}$$

Kimmel solves this quadratic programming problem using projected normalized steepest descent at multiple resolutions [76]. Here, we rewrite the problem slightly, optimizing for the reflectance rather than the illumination, by substituting according to the coherence condition $i = b + r$:

$$\min_r \left\{ \|\nabla r - \nabla i\|_2^2 + \alpha\|r\|_2^2 + \beta\|\nabla r\|_2^2 \right\} \quad \text{s.t.} \quad r \leq 0, \quad \langle \nabla r, \vec{n} \rangle = 0 \text{ on } \partial\Omega. \tag{4.15}$$

This form makes clear that variational Retinex is an optimization between reflectance gradient fidelity and some sparsity penalties.

Subsequently, variations of this variational Retinex model have been proposed, mainly involving different norms for the fidelity and sparsity terms, and dropping the asymmetry constraint $r \leq 0$. First, Ma and Osher have dropped a few terms and replace H^1 smoothness of the reflectance by a TV-prior [96]:

$$\min_r \left\{ \|\nabla r - \nabla i\|_2^2 + 2\lambda\|\nabla r\|_1 \right\} \tag{4.16}$$

As a complication, instead of the local TV prior, they also make use of non-local total variation. Further, Ng and Wang introduce an L^2-fidelity prior between

reflectance and intensity [111]:

$$\min_r \left\{ \|\nabla r - \nabla i\|_2^2 + \alpha \|r - i\|_2^2 + 2\lambda \|\nabla r\|_1 \right\} \tag{4.17}$$

Chen et al. have used a TV-L^1-based variational Retinex approach, which they call logarithmic total variation (LTV), for illumination normalized face detection [35]:

$$\min_r \left\{ \|\nabla r - \nabla i\|_1 + \alpha \|r\|_1 \right\} \tag{4.18}$$

Remark 7. *At this point it is worthwhile noting, that both the L^2- and L^1-Retinex [108, 95] have a threshold-free variational equivalent.*

Indeed, the hard threshold on the intensity gradient can be seen as a contraction of an L^0-sparsity prior on the gradients of the reflectance (see §5.2 for details):

$$\min_r \left\{ \|\nabla r - \delta_\tau \nabla i\|_2^2 \right\} = \min_r \left\{ \left\| \nabla r - \operatorname*{argmin}_{\vec{q}} \left\{ \|\vec{q} - \nabla i\|_2^2 + \tau^2 \|\vec{q}\|_0 \right\} \right\|_2^2 \right\} \tag{4.19}$$

which is a relaxed version of the more complicated problem

$$\min_r \left\{ \|\vec{q} - \nabla i\|_2^2 + \tau^2 \|\vec{q}\|_0 \right\} \quad \text{s.t.} \quad \nabla r = \vec{q} \tag{4.20}$$

This connection between basic threshold and a related variational problem makes it intuitive, that other variational models, such as the TV-Retinex [96], can be retro-fit into a threshold based Poisson-problem, as well. *This is the fundamental insight leading to the non-local unification proposed in this paper.*

CHAPTER 5

Non-local Differential Operators

In this chapter, we recall and give a few definitions of non-local differential operators, which we need in order to cast existing kernel-based Retinex methods into a variational framework, and based on which we will propose our unifying Retinex framework in Chapter 6.

5.1 Non-local Differential Operators

Product and Norms

First, we give the definitions of particular products and norms of scalars and non-local vectors. To begin with, we require appropriate inner products.

Definition 1. *For scalars $i : \Omega \to \mathbb{R}$, we choose:*

$$\langle i, j \rangle := \int_{\Omega} i(x) j(x) dx, \tag{5.1}$$

which is the common L^2 inner product.

Definition 2. *Accordingly, we introduce the following inner product for vectors $\vec{v} : \Omega \to \Omega \times \Omega$:*

$$\langle \vec{u}, \vec{v} \rangle := \int_{\Omega \times \Omega} u(x, y) v(x, y) dx dy. \tag{5.2}$$

Definition 3. *The associated L^2 norms are respectively for scalars $i : \Omega \to \mathbb{R}$:*

$$\|i\|_2 := \sqrt{\langle i, i \rangle} = \sqrt{\int_{\Omega} i(x)^2 dx}, \tag{5.3}$$

and for vectors $\vec{v} : \Omega \to \Omega \times \Omega$*:*

$$\|\vec{v}\|_2 := \sqrt{\langle \vec{v}, \vec{v} \rangle} = \sqrt{\int_{\Omega \times \Omega} v(x,y)^2 dxdy}\,. \tag{5.4}$$

Definition 4. *Similarly, the* L^1*-norm of the vector* $\vec{v}$*,* $\|\vec{v}\|_1 : \Omega \times \Omega \to \mathbb{R}$*, is defined as*

$$\|\vec{v}\|_1 := \int_{\Omega \times \Omega} |v(x,y)| dxdy. \tag{5.5}$$

Definition 5. *Let* w *be a non-negative weighting function and* $\vec{v}$ *a vector. The weighted* L^0*-"norm" of the vector* $\vec{v}$*,* $\|\vec{v}\|_{0,w} : \Omega \times \Omega \to \mathbb{R}$*, is defined as*

$$\|\vec{v}\|_{0,w} := \int_{\Omega \times \Omega} w(x,y)(1 - \delta(v(x,y))) dxdy, \tag{5.6}$$

where δ *is the Dirac distribution. This functional emphasizes the* (L^0) *sparsity of important vector components.*

Definition 6. *Further, point-wise multiplication is written for scalars* i *and* j *as*

$$(i \cdot j)(x) := i(x) j(x), \quad x \in \Omega, \tag{5.7}$$

and for vectors $\vec{u}$ *and* $\vec{v}$ *as:*

$$(\vec{u} \cdot \vec{v})(x,y) := u(x,y) v(x,y), \quad x, y \in \Omega. \tag{5.8}$$

Differential Operators

We recall some definitions of non-local operators [53].

Definition 7. *Let* $\Omega \in \mathbb{R}^n$*,* $x \in \Omega$*,* $i(x)$ *be a real function* $i : \Omega \to \mathbb{R}$*. We define the non-local gradient of this function as the vector of all partial derivatives,* $\nabla_w i : \Omega \to \Omega \times \Omega$*:*

$$(\nabla_w i)(x,y) := \sqrt{w(x,y)}(i(y) - i(x)), \quad x, y \in \Omega, \tag{5.9}$$

for some non-negative weights $w(x,y)$*.*

Remark 8. *Note that this definition of non-local gradients amounts to a point-wise multiplication between the vector of all finite differences, $\vec{d}_i$, and the weight-vector $\vec{\omega}$:*

$$(\nabla_w i) \equiv \vec{\omega} \cdot \vec{d}_i \quad \begin{cases} \vec{\omega} : \omega(x,y) = \sqrt{w(x,y)} & x,y \in \Omega \\ \vec{d}_i : d_i(x,y) = i(y) - i(x) & x,y \in \Omega \end{cases} \tag{5.10}$$

Thus we will call any such point-wise product between a weighting function and another vector, say, $\vec{q}$, a quasi-gradient*, as it shares the form of an actual gradient up to the fact that the vector $\vec{q}$ does not actually stem from non-local finite differences.*

Definition 8. *The associated divergence of a vector $\vec{v} \in \Omega \times \Omega$, namely $\div_w \vec{v} : \Omega \times \Omega \to \Omega$, is then defined as the negative adjoint under the above inner products:*

$$\langle \nabla_w i, \vec{v} \rangle = \langle i, - \div_w \vec{v} \rangle \,. \tag{5.11}$$

The expression for the divergence is easily found as

$$(\div_w \vec{v})(x) := \int_\Omega \sqrt{w(x,y)} v(x,y) - \sqrt{w(y,x)} v(y,x) dy. \tag{5.12}$$

Definition 9. *The non-local Laplacian, $\Delta_w i : \Omega \to \Omega$ is defined as the composition of non-local divergence and non-local gradient:*

$$(\Delta_w i)(x) := (\div_w (\nabla_w i))(x) = \int_\Omega (w(x,y) + w(y,x))(i(y) - i(x)) dy. \tag{5.13}$$

Note that if $w_s(x,y)$ is a symmetric weighting function, then the associated Laplacian can be simplified as:

$$(\Delta_{w_s} i)(x) = (\div_{w_s} (\nabla_{w_s} i))(x) = 2 \int_\Omega w_s(x,y)(i(y) - i(x)) dy, \tag{5.14}$$

where the Laplacian now differs from the regular graph Laplacian by a factor 2.

Filtered Gradients

Based on the non-local differential operators defined above, we now introduce *filtered gradients*, by making use of a filter function f acting on the scalar differences.

Definition 10. *Let $f : \mathbb{R} \to \mathbb{R}$ be a real-valued distortion function applied to the finite differences. We define filtered non-local gradients, $\nabla_{w,f} i : \Omega \to \Omega \times \Omega$, as the quasi-gradients obtained as follows:*

$$(\nabla_{w,f} i)(x, y) := \sqrt{w(x, y)} f(i(y) - i(x)), \quad x, y \in \Omega. \tag{5.15}$$

Definition 11. *We call $\Delta_{w,f}$ the filtered non-local Laplacian obtained by applying the (regular) divergence to filtered gradients*

$$(\Delta_{w,f})(x) := (\div_w(\nabla_{w,f} i))(x) = \int_\Omega w(x, y) f(i(y) - i(x)) - w(y, x) f(i(x) - i(y)) dy. \tag{5.16}$$

Therefore, if f_s is a symmetric real-valued function, *i.e.*, $f_s(z) = f_s(-z)$, and the weight $w_s(\cdot, \cdot)$ is symmetric, then the associated filtered non-local Laplacian Δ_{w_s,f_s} is always zero. On the other hand, if f_a is an anti-symmetric real-valued function, the filtered non-local Laplacian is given by:

$$(\Delta_{w_s,f_a})(x) = 2 \int_\Omega w_s(x, y) f_a(i(y) - i(x)) dy. \tag{5.17}$$

Non-stationary Filtering

In the above definitions, the filter function f was stationary, *i.e.*, independent of the location. Now we introduce a non-stationary filtering function.

Definition 12. *Let $f^{ns} : \mathbb{R} \times \Omega \times \Omega \to \mathbb{R}$ be a non-stationary real-valued distortion function applied to the finite differences. We define non-stationary filtered non-local gradients, $\nabla_{w,f^{ns}} i : \Omega \to \Omega \times \Omega$, as follows:*

$$(\nabla_{w,f^{ns}} i)(x, y) := \sqrt{w(x, y)} f^{ns}(i(y) - i(x), x, y), \quad x, y \in \Omega. \tag{5.18}$$

Definition 13. *We call $\Delta_{w,f^{ns}}$ the non-stationary filtered non-local Laplacian obtained by applying the (regular) divergence to non-stationary filtered gradients*

$$\begin{aligned}(\Delta_{w,f^{ns}})(x) &:= (\div_w(\nabla_{w,f^{ns}} i))(x) \\ &= \int_\Omega w(x,y) f^{ns}(i(y)-i(x),x,y) - w(y,x) f^{ns}(i(x)-i(y),y,x) dy.\end{aligned} \tag{5.19}$$

Again, if f is a cross-symmetric non-stationary real-valued filter function, *i.e.*, $f(z,x,y) = -f(-z,y,x)$, and the weight $w_s(\cdot,\cdot)$ is symmetric, then the associated non-stationary filtered non-local Laplacian $\Delta_{w_s,f}$ is given by:

$$(\Delta_{w_s,f})(x) = 2\int_\Omega w_s(x,y) f(i(y)-i(x),x,y) dy. \tag{5.20}$$

5.2 Sparse Quasi-gradients

Based on the above definitions of non-stationary filtered gradients, we can now make a formal connection between particular types of sparse gradient approximations and related thresholding filter functions.

L^0-Sparse Quasi-gradients

Lemma 2. *Let w_1 and w_2 be two non-negative weighting functions. We look for a vector $\vec{q}_{L^0}$ which is L^0 sparse as weighted by w_1, while the quasi-gradient $\sqrt{w_2}\cdot\vec{q}_{L^0}$ remains close to the observed gradients $\nabla_{w_2} i$. This is the solution of the following optimization problem*

$$\vec{q}_{L^0} = \operatorname*{argmin}_{\vec{q}} \left\{ \lambda^2 \|\vec{q}\|_{0,w_1} + \|\sqrt{w_2}\cdot\vec{q} - \nabla_{w_2} i\|_2^2 \right\}, \tag{5.21}$$

which has a closed form as a component-wise hard-thresholding applied to the non-local finite differences:

$$q_{L^0}(x,y) = S_\tau^h(i(y)-i(x)), \quad \textit{where} \quad \tau = \lambda\sqrt{\frac{w_1(x,y)}{w_2(x,y)}} \quad \textit{and} \quad S_\tau^h(z) = \begin{cases} 0 & |z| \le \tau \\ z & \textit{otherwise} \end{cases} \tag{5.22}$$

Proof. We first proceed by rewriting the optimization problem component-wise, and explicitly using the definition of the L^0-cost function:

$$q_{L^0}(x,y) = \underset{q\in\mathbb{R}}{\operatorname{argmin}} \left\{ \lambda^2 w_1(x,y)(1-\delta(q)) + w_2(x,y)\left(q-(i(y)-i(x))\right)^2 \right\}, \quad \forall (x,y) \in \Omega\times\Omega. \tag{5.23}$$

It is easy to see that the sub-differential of the expression to be minimized contains 0 in at most two points, *i.e.*, there are at most two local minima (just one if they coincide), namely:

$$q_{L^0}(x,y) \in \{0, i(y)-i(x)\}. \tag{5.24}$$

Since the expression goes to $+\infty$ for $q \to \pm\infty$, the global minimum is determined by comparing the cost associated with just these two candidates:

$$q_{L^0}(x,y) = \begin{cases} 0 & w_2(x,y)(i(y)-i(x))^2 \le \lambda^2 w_1(x,y) \\ i(y)-i(x) & \text{otherwise,} \end{cases} \qquad \forall (x,y) \in \Omega\times\Omega. \tag{5.25}$$

Identifying this expression with hard-thresholding completes the proof. □

Remark 9. *Let w_3 be another weighting function, based on w_1 and w_2 as*

$$w_3(x,y) = \max\left(w_1(x,y), w_2(x,y)\right) \quad x,y \in \Omega \tag{5.26}$$

The quasi-gradient $\sqrt{w_3}\cdot\vec{q}_{L^0}$ is an instance of non-stationary filtered non-local gradient:

$$\left(\sqrt{w_3}\cdot\vec{q}_{L^0}\right)(x,y) = (\nabla_{w_3,f^{ns}} i)(x,y) = \sqrt{w_3(x,y)} f^{ns}(i(y)-i(x),x,y) \tag{5.27}$$

with

$$f^{ns} = S^h_\tau \quad \textit{and} \quad \tau(x,y) = \lambda\sqrt{\frac{w_1(x,y)}{w_2(x,y)}}. \tag{5.28}$$

Note that other weights w_3 could be chosen, e.g., $w_3 = w_1 + w_2$. It is wise, however, to ensure $\operatorname{supp}(w_3) = \operatorname{supp}(w_1)\cup\operatorname{supp}(w_2)$ to avoid ill-defined thresholds.

L^1/TV Sparse Quasi-gradients

After making the relation between hard-thresholded gradients and L^0-sparsity, we now highlight the similar connection between soft-thresholding (shrinkage) and L^1/TV-sparsity.

Lemma 3. *Let w_1 and w_2 be two non-negative weighting functions. We look for a vector $\vec{q}_{TV}$ such that the quasi-gradient $\sqrt{w_1} \cdot \vec{q}_{TV}$ has small L^1-norm, while the quasi-gradient $\sqrt{w_2} \cdot \vec{q}_{TV}$ remains close to the observed gradients $\nabla_{w_2} i$. More precisely, we are interested in the solution of the following convex optimization problem:*

$$\vec{q}_{TV} = \underset{\vec{q}}{\operatorname{argmin}} \left\{ 2\lambda \| \sqrt{w_1} \cdot \vec{q} \|_1 + \| \sqrt{w_2} \cdot \vec{q} - \nabla_{w_2} i \|_2^2 \right\}. \tag{5.29}$$

It is found as component-wise soft-thresholding applied to the non-local finite differences:

$$q_{TV}(x, y) = S_\tau^s(i(y) - i(x)), \tag{5.30}$$

where

$$\tau = \lambda \frac{\sqrt{w_1(x, y)}}{w_2(x, y)} \quad \textit{and} \quad S_\tau^s(z) = \begin{cases} z + \tau & z < -\tau \\ 0 & |z| \leq \tau \\ z - \tau & z > \tau \end{cases}$$

Proof. For shorter notation, let us write $\tau := \lambda \frac{\sqrt{w_1(x,y)}}{w_2(x,y)} \geq 0$, and $z := i(y) - i(x)$. Again we rewrite the problem component-wise:

$$q_{TV}(x, y) = \underset{q}{\operatorname{argmin}} \left\{ f(q) := 2\tau |q| + (q - z)^2 \right\}, \qquad \forall (x, y) \in \Omega \times \Omega. \tag{5.31}$$

The cost functional $f(q)$ is non-differentiable at $q = 0$, and it grows to infinity for $q \to \pm\infty$. As a consequence, it is minimized for a finite q^* such that the functional's sub-gradient contains 0 at that point, $\partial_q f(q^*) \ni 0$. The sub-differential is

easily computed as

$$\partial_q f(q) = \begin{cases} 2(\tau + q - z) & q > 0 \\ [-\tau - z, \tau - z] & q = 0 \\ 2(-\tau + q - z) & q < 0. \end{cases} \tag{5.32}$$

There are, thus, three cases to consider:

Case 1a: $z > \tau$ Clearly, in this case 0 is not in the sub-differential at $q = 0$. Also, the differential does not vanish for any negative q. Instead, $q = z - \tau$ is the only minimizer.

Case 1b: $z < -\tau$ Similar to case 1a, except that the only minimizer is found for $q = z + \tau$.

Case 2: $z \in [-\tau, \tau]$ In this case, $0 \in \partial_q f(q = 0)$, while no $q \neq 0$ leads to a vanishing derivative. The minimizer is thus given by $q = 0$. In addition, we have:

$$f(0) = z^2 \tag{5.33}$$

$$f(q > 0) - f(0) = 2\tau q + (z - q)^2 - z^2 = 2(\tau - z)q + q^2 > 0 \tag{5.34}$$

$$f(q < 0) - f(0) = -2\tau q + (z - q)^2 - z^2 = 2(-\tau - z)q + q^2 > 0, \tag{5.35}$$

which confirms $q = 0$ as only minimizer in this case. Identifying the three cases and their minimizers with soft-thresholding (shrinkage) completes the proof. □

Remark 10. *Let again w_3 be the weighting function such that $w_3 = \max(w_1, w_2)$. The quasi-gradient $\sqrt{w_3} \cdot \vec{q}_{TV}$ is an instance of non-stationary filtered non-local gradient:*

$$\left(\sqrt{w_3} \cdot \vec{q}_{TV}\right)(x, y) = (\nabla_{w_3, f^{ns}} i)(x, y) = \sqrt{w_3(x, y)} f^{ns}(i(y) - i(x), x, y) \tag{5.36}$$

with

$$f^{ns} = S^s_\tau \quad \textit{and} \quad \tau(x, y) = \lambda \frac{\sqrt{w_1(x, y)}}{w_2(x, y)} \tag{5.37}$$

L^2/H^1 Sparse Quasi-gradients

Lemma 4. *Let w_1 and w_2 be two non-negative weighting functions. We look for a vector $\vec{q}_{H^1}$ such that the quasi-gradient $\sqrt{w_1} \cdot \vec{q}_{H^1}$ has small L^2-norm, while the quasi-gradient $\sqrt{w_2} \cdot \vec{q}_{H^1}$ remains close to the observed gradients $\nabla_{w_2} i$. This corresponds to the following quadratic, convex optimization problem:*

$$\vec{q}_{H^1} = \underset{\vec{q}}{\operatorname{argmin}} \left\{ \lambda \|\sqrt{w_1} \cdot \vec{q}\|_2^2 + \|\sqrt{w_2} \cdot \vec{q} - \nabla_{w_2} i\|_2^2 \right\} \tag{5.38}$$

Its solution is found as scaling applied to the non-local differences:

$$q_{H^1}(x,y) = S_\tau^u(i(y) - i(x)), \quad \textit{where} \quad \tau = \lambda \frac{w_1(x,y)}{w_2(x,y)} \quad \textit{and} \quad S_\tau^u(z) = \frac{z}{1+\tau} \tag{5.39}$$

Proof. The cost function is fully differentiable and can be solved point-wise:

$$q_{H^1}(x,y) = \underset{q}{\operatorname{argmin}} \left\{ \lambda w_1 q^2 + w_2 (q - (i(y) - i(x)))^2 \right\}, \qquad \forall (x,y) \in \Omega \times \Omega. \tag{5.40}$$

Since the cost functional approaches $+\infty$ for $q \to \pm\infty$, the minimum is achieved for vanishing first variation. Optimality thus requires from a minimizer q^*:

$$\left(\lambda \frac{w_1(x,y)}{w_2(x,y)} + 1 \right) q^* = i(y) - i(x), \tag{5.41}$$

from which the lemma is directly obtained. □

Remark 11. *Let again w_3 be the weighting function such that $w_3 = \max(w_1, w_2)$. The quasi-gradient $\sqrt{w_3} \cdot \vec{q}_{H^1}$ is an instance of non-stationary filtered non-local gradient:*

$$\left(\sqrt{w_3} \cdot \vec{q}_{H^1}\right)(x,y) = (\nabla_{w_3, f^{ns}} i)(x,y) = \sqrt{w_3(x,y)} f^{ns}(i(y) - i(x), x, y) \tag{5.42}$$

with

$$f^{ns} = S_\tau^u \quad \textit{and} \quad \tau(x,y) = \lambda \frac{w_1(x,y)}{w_2(x,y)} \tag{5.43}$$

Remark 12. *For the particular choice of identical weights $w = w_1 \equiv w_2$, we have $w_3 = w$, and the thresholdings in both L^0 and L^2/H^1 sparse quasi-gradients become a constant in space, that is $\tau(x, y) = \lambda$ uniformly, which makes the filtering function stationary and anti-symmetric. If the weights are binary, then this extends to the L^1/TV case, as well. On the other hand, if the weights controlling the sparsity and the fidelity are different, then those thresholding functions τ vary spatially, which results in* adaptive thresholding. *We will further discuss this in Section 6.5.*

TV Augmented Quasi-gradients

The last case deals with a gradient enhancement:

Lemma 5. *Let w_1 and w_2 be two non-negative weighting functions. We look for a vector $\vec{q}^{TV}$ such that the quasi-gradient $\sqrt{w_1} \cdot \vec{q}^{TV}$ has* increased *L^1-norm, while the quasi-gradient $\sqrt{w_2} \cdot \vec{q}^{TV}$ remains close to the observed gradients $\nabla_{w_2} i$. It's the solution of the following optimization problem:*

$$\vec{q}^{TV} = \underset{\vec{q}}{\operatorname{argmin}} \left\{ -2\lambda \| \sqrt{w_1} \cdot \vec{q} \|_1 + \| \sqrt{w_2} \cdot \vec{q} - \nabla_{w_2} i \|_2^2 \right\} \tag{5.44}$$

and is found as unshrinkage *applied to the non-local differences:*

$$q^{TV}(x, y) = S^s_{-\tau}(i(y) - i(x)), \tag{5.45}$$

where

$$\tau = \lambda \frac{\sqrt{w_1(x, y)}}{w_2(x, y)} \quad \textit{and} \quad S^s_{-\tau}(z) = \begin{cases} z + \tau & z > 0 \\ 0 & z = 0 \\ z - \tau & z < 0 \end{cases}$$

Proof. In this case again, we proceed by rewriting the problem component-wise, and substituting $\tau := \lambda \frac{\sqrt{w_1(x,y)}}{w_2(x,y)}$ and $z := i(y) - i(x)$:

$$q^{TV}(x, y) = \underset{q}{\operatorname{argmin}} \left\{ f(q) := -2\tau|q| + (q - z)^2 \right\}, \qquad \forall (x, y) \in \Omega \times \Omega. \tag{5.46}$$

The cost function $f(q)$ is non-differentiable at $q = 0$, and $f(q) \to \infty$ as $q \to \pm\infty$. The subgradient of f is:

$$\partial_q f(q) = \begin{cases} 2(-\tau + q - z) & q > 0 \\ \varnothing & q = 0 \\ 2(\tau + q - z) & q < 0. \end{cases} \tag{5.47}$$

There are now 5 cases to be considered:

Case 1a: $z \geq \tau$ The only vanishing subgradient is found for $q = z + \tau$.

Case 1b: $z \in (0, \tau)$ There are two locations with vanishing subgradient, namely $q = z \pm \tau$, corresponding to two local minima, one of which is the minimizer. The global minimum is found by evaluating the cost function at these two locations:

$$f(z + \tau) = -2\tau(z + \tau) + \tau^2 = -2\tau z - \tau^2 \tag{5.48}$$

$$f(z - \tau) = 2\tau(z - \tau) + (-\tau)^2 = +2\tau z - \tau^2, \tag{5.49}$$

therefore $f(z + \tau) < f(z - \tau)$, and the global minimizer is found as $q = z + \tau$.

Case 2a: $z \leq -\tau$ The only vanishing subgradient is found for $q = z - \tau$.

Case 2b: $z \in (-\tau, 0)$ As in case 1b, there are two local minima, at $q = z \pm \tau$. This time, $f(z - \tau) < f(z + \tau)$ and $q = z - \tau$ is the unique minimizer.

Case 3: $z = 0$ In this case, there are two equal minimizers, $q = \pm\tau$, since $f(\tau) = f(-\tau) = -\tau^2$, here.

The cases 1a through 2b combine into the gradient enhancement of the lemma. For the single point $z = 0$ we are left with an ambiguity between two minimizers. *In order to resolve this ambiguity, we deliberately set the solution to* $q = 0$ *for*

$z = 0$, *albeit this is clearly not a minimizer, but a local maximum instead.* This choice is primarily motivated by the desire to obtain an odd filtering function, and since no image gradients are to be created out of nowhere. □

Remark 13. *Let again w_3 be the weighting function such that $w_3 = \max(w_1, w_2)$. The quasi-gradient $\sqrt{w_3} \cdot \vec{q}^{TV}$ is an instance of non-stationary filtered non-local gradient:*

$$\left(\sqrt{w_3} \cdot \vec{q}^{TV}\right)(x,y) = (\nabla_{w_3, f^{ns}} i)(x,y) = \sqrt{w_3(x,y)} f^{ns}(i(y) - i(x), x, y) \tag{5.50}$$

with

$$f^{ns} = S^s_{-\tau} \quad \textit{and} \quad \tau(x,y) = \lambda \frac{\sqrt{w_1(x,y)}}{w_2(x,y)} \tag{5.51}$$

5.3 Computing the Weights

At this point, it is worthwhile spending some time on different choices for the weight vectors used in the non-local differential operators. We will discuss common choices such as local weights, patch-based non-local weights, semi-local Gaussian kernels, and finally we suggest the use of cosine-based distances.

Local Weights

It is possible to construct a local weight vector that reproduces the finite difference scheme of standard local differential operators.

Definition 14. *Indeed, we construct two-dimensional local weights between two points x and y as*

$$w_\epsilon(x,y) := \frac{1}{\epsilon}\left[\delta\left(x - y + \begin{pmatrix}\epsilon\\0\end{pmatrix}\right) + \delta\left(x - y + \begin{pmatrix}0\\\epsilon\end{pmatrix}\right)\right], \tag{5.52}$$

where δ is the Dirac distribution.

Then, in the limit $\epsilon \to 0$, the non-local derivative approaches the standard gradient:

$$\lim_{\epsilon\to 0}(\nabla_{w_\epsilon} u)(x) = \lim_{\epsilon\to 0}\frac{1}{\epsilon}\begin{pmatrix} u\left(x+\begin{pmatrix}\epsilon\\0\end{pmatrix}\right) - u(x) \\ u\left(x+\begin{pmatrix}0\\\epsilon\end{pmatrix}\right) - u(x)\end{pmatrix} = (\nabla u)(x) \tag{5.53}$$

(Here, in an abuse of notation, $(\nabla_{w_\epsilon} u)(x) \in \mathbb{R}^{|\Omega|}$ denotes the collection of $(\nabla_{w_\epsilon} u)(x,y)$ for all $y \in \Omega$, of which only two elements corresponding to $y = x + (\epsilon, 0)^T$ and $y = x + (0,\epsilon)^T$ are non-zero.)

Semi-local Gaussian Kernel

The Gaussian kernel corresponds to a symmetrical extension and mollification of local weights.

Definition 15. *The isotropic 2-dimensional normalized Gaussian kernel is given by:*

$$w_g(x,y) := \frac{1}{2\pi\sigma^2} e^{-\frac{d(x,y)^2}{2\sigma^2}}, \tag{5.54}$$

where $d(x,y)$ denotes the Euclidean distance between points x and y, and σ is a scale factor.

Patch-based Non-local Weights

The non-local weights commonly used in imaging have been introduced in [25].

Definition 16. *Based on an image $i(x)$, we define non-local weights as*

$$w_{nl}(x,y) := e^{-\frac{d(I(x),I(y))^2}{h^2}} \quad \textit{with} \quad d^2(I(x),I(y)) := \int_\Omega G_\sigma(t)(i(y+t)-i(x+t))^2 dt, \tag{5.55}$$

where $h > 0$ is a scale parameter and G_σ is a Gaussian window with standard deviation σ.

This weight is close to zero if the regions (patches) around x and y, $I(x)$ and $I(y)$, have an important Gaussian-weighted L^2-distance $d(I(x), I(y))$. Practically, for each pixel x we only calculate $w_{nl}(x, y)$ for y in a small search window centered at x, keep only a few, large coefficients and discard the rest [25, 53, 96]. These weights are generally used today, and little attention is typically paid to where they come from.

However, an important connection has been shown by [2], where the exponential weights have been connected with maximum-entropy distributions known from statistical mechanics [70, 57]. Indeed, the Gaussian weights are maximum-entropy weights minimizing the non-local H^1-energy based on patch-distances $d(I(x), I(y))$. This argument suggests, that different weights are optimal when functionals other than non-local H^1 are considered, such as non-local TV or non-local L^0, corresponding to different choices of distances d.

Remark 14. *Combining patch-distances in both space and intensity leads to the bilateral filter or Beltrami diffusion in patch-space [148, 9, 132, 154], as used for shadow removal in [158].*

Cosine-distances

Indeed, it must appear wrong to use (illumination-sensitive) non-local weights based on the input image, to measure the non-local H^1 of the illumination invariant reflectance (or its fidelity to filtered gradients). Instead of computing and updating the weights on the current estimate of the reflectance, we suggest using illumination-robust weights stemming from the input image as a proxy, by normalizing patches by their L^2-norm before computing distances. Such normalized L^2-distance can be shown to be equivalent to a cosine distance:

$$d(I(x)/\|I(x)\|_2, I(y)/\|I(y)\|_2) = 1 - \langle I(x)/\|I(x)\|_2, I(y)/\|I(y)\|_2 \rangle = 1 - \cos \alpha_{xy}, \tag{5.56}$$

where $\alpha_{xy} := \angle(I(x), I(y))$ is the angle between two patches.

Further, exponentials of such cosine distances have the same low order terms in their series expansion as powers of the cosine:

$$e^{-\lambda(1-\cos\alpha_{xy})} \approx \cos^{\lambda}\alpha_{xy} \qquad \alpha_{xy} \ll 1 \tag{5.57}$$

Remark 15. *The use of L^2-normalization amplifies the noise in low-intensity regions.*

Color Distances

In a natural image, objects of similar material may have similar texture and thus small patch-angle. However, more likely pixels of the same material will have a very similar hue. Here, cosine distances come in very handy. Indeed, to measure the similarity in material between pixels, we use the cosine distance in RGB space. This comes from the observation that two pixels in RGB space have the same hue if the angle between the two corresponding RGB vectors is small—irrespective of their strength of illumination, encoded as magnitude. Given any two pixels x, y with the corresponding RGB values $(I_1(x), I_2(x), I_3(x))$ and $(I_1(y), I_2(y), I_3(y))$, the angle between those two vectors, α_{xy}, is computed via

$$\cos(\alpha_{xy}) := \frac{\langle I(x), I(y)\rangle}{\|I(x)\|\|I(y)\|} = \frac{I_1(x)I_1(y) + I_2(x)I_2(y) + I_3(x)I_3(y)}{\sqrt{I_1^2(x) + I_2^2(x) + I_3^2(x)}\sqrt{I_1^2(y) + I_2^2(y) + I_3^2(y)}} \tag{5.58}$$

Definition 17. *Then we define the color/hue-based weight as follows:*

$$w_c(x,y) = \begin{cases} \cos(\alpha_{xy}) & \text{if } \cos(\alpha_{xy}) \geq c \\ 0 & \text{otherwise,} \end{cases} \tag{5.59}$$

where $0 < c < 1$ is a parameter close to one.

This definition of color weight in RGB space can be directly generalized to weights based on cosine-distances in hyperspectral images [6].

5.4 Closing the Gap Between Kernel and Variational Retinex

The link between PDE-Retinex (thresholded Poisson equations) and variational models has been shown before. The definitions of non-local differential operators will now allow us to make a formal connection between semi-local kernel-based and center-surround Retinex models on one hand, and local variational as well as threshold-based models on the other hand. This will eventually provide a big umbrella under which all Retinex methods can be classified.

Homomorphic Filtering as Variational Problem

We have already mentioned that the homomorphic filtering Retinex can be rewritten as a Gaussian-kernel $w_g(x,y)$ based computation of the following form:

$$r(x) = i(x) - \sum_y w_g(x,y) i(y) = -\sum_y w_g(x,y) \left(i(y) - i(x)\right), \tag{5.60}$$

provided that the Gaussian kernel is normalized, *i.e.*, $\sum_y w_g(x,y) = 1$. The second sum now clearly identifies with our definition of non-local Laplacian (see Definition 9 and (5.14)), and we may thus also write:

$$r(x) + \frac{1}{2}\Delta_{w_g} i(x) = 0. \tag{5.61}$$

This in turn is the Euler-Lagrange equation corresponding to the following convex optimization problem:

$$\min_r \left\{ \underbrace{\left\|\nabla_{w_g} r - \nabla_{w_g} i\right\|_2^2}_{\text{A}} - \underbrace{\left\|\nabla_{w_g} r\right\|_2^2}_{\text{B}} + \underbrace{2\left\|r\right\|_2^2}_{\text{C}} \right\}. \tag{5.62}$$

The interpretation of this variational model is as follows: We look for a reflectance r whose gradients are similar to those of i (A), but enhanced (B), while having minimal energy (C).

Perceptual Contrast Enhancement and Non-local Derivatives

In their award-winning model, Bertalmìo and colleagues have used their kernel-based lightness estimate together with a grey-world prior and a fidelity constraint to build a "perceptually inspired variational framework" for image enhancement [10, 120]. Their anti-symmetrized kernel-based Retinex has a variational formulation, which is very close to the ACE model [11], namely:

$$\min_R \left\{ \int_\Omega \left[\alpha \left(R(x) - \frac{1}{2} \right)^2 + \beta \left(R(x) - I(x) \right)^2 \right] dx + C^{\frac{\min}{\max}}_{\sqrt{w}}(R) \right\}. \tag{5.63}$$

where $C^{\frac{\min}{\max}}_{\sqrt{w}}(R)$ is a contrast function. For particular, but reasonable choices of the contrast functions, the contrast term can be written as

$$C^{\frac{\min}{\max}}_{\sqrt{w}}(R) \equiv -\int_{\Omega\times\Omega} \sqrt{w(x,y)} |f_a(r(y) - r(x))| dxdy = -\|\nabla_{w,f_a} r\|_1 . \tag{5.64}$$

In particular, f_a may be the identity. Thus, we rewrite the perceptual contrast enhancement in terms of non-local derivatives as follows:

$$\min_{r=\log(R)} \left\{ \alpha \left\| R - \frac{1}{2} \right\|_2^2 + \beta \|R - I\|_2^2 - \|\nabla_w r\|_1 \right\}. \tag{5.65}$$

where the first term represents the gray-world prior, the second is a fidelity term with respect to the observed intensity, and the contrast term increases non-local TV of the reflectance.

CHAPTER 6

Non-local Retinex Model

In this chapter, we introduce our unifying non-local Retinex model. We will first show show how the existing Retinex implementations can be reproduced in our proposed model. Moreover, our proposed framework also yields new results in shadow detection, color contrast enhancement and cartoon-texture decomposition. Part II of the thesis was taken with slight modification from [164, 165].

6.1 Non-local Retinex Model

All those Retinex models that we have seen so far have a variational cousin, potentially through the use of non-local differential operators. Even more, these variational counterparts all share a very similar structure: the energy typically comprises one or two fidelity terms (image and/or its gradient), as well as sparsity priors or alternatively, through negation, enhancement terms.

Also, we have shown that this type of variational problem can be retrofitted into a Horn/Morel-style gradient-fidelity problem (PDE-Retinex), potentially adding further terms. In particular, we have shown in the previous section, how different gradient sparsity and fidelity terms translate into different associated thresholding functions.

Here, we want to formulate this retrofitted PDE-Retinex model as a general recipe. We tackle the Retinex problem in a two step approach:

1) Gradient filtering We realize that the reflectance obeys both to some gradient sparsity priors and some gradient fidelity priors. In a first step, we thus look for an optimal quasi-gradient that best satisfies those two constraints. This quasi-gradient is obtained as filtered gradient of the observed image $\nabla_{w,f} i$. Here, we write w as short for w_3, since both w_1 and w_2 are entirely hidden within the generic gradient filter f, inspired by the threshold employed by Horn and Morel [66, 108].

2) Gradient fitting However, the resulting quasi-gradient almost certainly is not a valid gradient by itself, and we wish to fit a reflectance, whose gradient comes closest to the quasi-gradient determined in the first step, while possibly respecting some additional constraints:

$$\hat{r} = \underset{r}{\operatorname{argmin}} \left\{ \|\nabla_w r - \nabla_{w,f} i\|_p^p + \alpha \|r\|_2^2 + \beta \|r - i\|_2^2 \right\} \tag{6.1}$$

The sparsity and gradient fidelity terms of the first step will determine the exact filter function f to be used, while the sparsity/smoothness priors on the illumination will essentially govern the gradient fidelity norm p of the second step.

The interest of such a two step procedure is manifold: First, each step, *i.e.*, thresholding the input gradient, followed by a gradient fitting is relatively simple to compute, compared to the non-compacted variational model. Further, the computational tools required to solve the gradient reconstruction step become independent of the gradient sparsity imposed.

Finally, this two step procedure is fully compatible with Marr's theory of lightness computation in the (primate) retina, given in (4.6) and (4.7) [102]. Beyond, it particularly neatly fits into Wilson-Cowan equations, modeling large-scale activity in cortical neural populations [155, 156], as will be shown below.

6.2 Numerical Optimization

In this section, we present basic solvers used for proof of concept. Note that solvers with better efficiency could easily be devised, but doing so is beyond the scope of this manuscript.

L^2 Gradient Fidelity

The L^2-based problem is differentiable and we propose to solve its Euler-Lagrange equations. The energy of the L^2 gradient-fidelity non-local Retinex is

$$J(r) = \|\nabla_w r - \nabla_{w,f} i\|_2^2 + \alpha \|r\|_2^2 + \beta \|r - i\|_2^2$$

The corresponding Euler-Lagrange equations are

$$0 = 2(-\Delta_w \hat{r} + \Delta_{w,f} i + \alpha \hat{r} + \beta(\hat{r} - i)), \tag{6.2}$$

and we recover an estimate of the reflectance $\hat{r}$ as

$$\hat{r} = ((\alpha + \beta)I - L)^{-1} (\beta i - \Delta_{w,f} i), \tag{6.3}$$

where I is the identity matrix, and L is the Laplacian matrix derived from the weights $w(x, y)$:

$$L_{xy} = \begin{cases} w(x,y) + w(y,x) & x \neq y \\ -\sum_{z \neq x} w(x,z) + w(z,x) & x = y. \end{cases} \tag{6.4}$$

Remark 16. *Since the graph Laplacian L is negative semi-definite, the operator $(\alpha + \beta)I - L$ is diagonally dominant. Therefore, we can solve for $\hat{r}$ either by a Gauss-Seidel algorithm or conjugate gradient method with a few iterations.*

Remark 17. *The problem at hand strongly resembles the L^2 statistical ranking problem [72, 63, 116], from which alternative optimization strategies could be inspired. Also, in some cases the quadratic programming scheme of [76], and its efficient approximations in [44], apply.*

Wilson-Cowan Equations

Originally, the Wilson-Cowan equations [155, 156] provided a description of the (temporal) evolution of the coarse-scale, mean activity of a population of both inhibitory and excitatory neurons in the cortex [12]. The equations have later been generalized in order to model the spatio-temporal distribution and patterns of excitation in the visual cortex (V1) [19]. A first connection between the Wilson-Cowan equations and Retinex was proposed by Cowan and Bressloff, where it was shown that Marr's Retinex model could actually be written in terms of Wilson-Cowan equations, and thus be implemented by corresponding neuron populations [102, 39].

Definition 18. *The relevant descriptor of mean activity, $a(x, \phi, t)$, at cortical coordinates x, orientation preference ϕ is given by the following integro-differential equation (IDE) [19, 12]:*

$$\frac{\partial a(x,\phi,t)}{\partial t} = -\nu a(x,\phi,t) + \mu \int_o^{\pi} \int_{\Omega} w(x,\phi,y,\varphi)\sigma\left[a(y,\varphi,t)\right] dyd\varphi + \lambda h(x,\phi,t) \tag{6.5}$$

where $h(x, \phi, t)$ is the external stimulus, w is a kernel that decays with the distances $|x - y|$ and $|\phi - \varphi|$, and σ is a sigmoid function.

Definition 19. *Like previous authors, we ignore the orientation ϕ and assume the external stimuli to be constant in time:*

$$\frac{\partial a(x,t)}{\partial t} = -\nu a(x,t) + \mu \int_{\Omega} w(x,y)\sigma\left[a(y,t)\right] dy + \lambda h(x) \tag{6.6}$$

Lemma 6. *The proposed L^2-based non-local Retinex model is a steady-state of the above orientation-insensitive Wilson-Cowan equations.*

Proof. First, let us write the gradient-descent equation associated with the first

variation

$$\frac{\partial r(x,t)}{\partial t} = 2\int_\Omega w(x,y)(r(y,t) - r(x,t))dy \\ - 2\int_\Omega w(x,y)f(i(y) - i(x))dy - 2\alpha r(x,t) - 2\beta(r(x,t) - i(x)).$$

After rearrangement,

$$\frac{\partial r(x,t)}{\partial t} = -2\left(\alpha + \beta + \int_\Omega w(x,y)dy\right) r(x,t) + 2\int_\Omega w(x,y)r(y,t)dy \\ + \beta i(x) - 2\int_\Omega w(x,y)f(i(y) - i(x))dy,$$

the individual terms can easily be identified with the elements of the Wilson-Cowan equation:

$$\nu = 2\alpha + 2\beta + 2\int_\Omega w(x,y)dy$$

$$\mu = 2$$

$$\sigma[r] = r$$

$$\lambda h(x) = \beta i(x) - 2\int_\Omega w(x,y)f(i(y) - i(x))dy,$$

where the first coefficient can easily be made constant by imposing kernel normality, *i.e.*, $\forall x: \ \int_\Omega w(x,y)dy = 1$. Further, the central term clearly identifies as lateral excitation, whereas the input stimulus $i(x)$ is affected by lateral inhibition through the (thresholded) gradients $\int_\Omega w(x,y)f(i(y) - i(x))dy$.

As usual, the solution of the initial minimization problem is associated with the steady state of its gradient descent, and since the gradient descent corresponds to a Wilson-Cowan IDE, therefore, the L^2 non-local gradient Retinex is a steady state of the corresponding Wilson-Cowan IDE. □

Remark 18. *It is to note that a similar claim was made in [12], where the perceptually inspired variational Retinex model is associated with a Wilson-Cowan IDE. However, there, the claim involves the sigmoid function σ to be both spatially varying and depending on $r(x,t)$; a complication which is not required in the Retinex IDE proposed here.*

L^1 Gradient Fidelity

Let us now consider the optimization of the proposed gradient fitting after filtering in the $p = 1$ case. We would like to point out that this problem is formally equivalent to the L^1 statistical ranking problem, for which efficient graph-cuts-based solvers have been developed [117]. Here, as simpler yet less efficient alternative for illustrative purposes we might as well explore a more intuitively accessible split Bregman/augmented Lagrangian based approach, as outlined in [13, 55, 113].

The L^1-based problem writes:

$$\min_r \left\{ \|\nabla_w r - \nabla_{w,f} i\|_1 + \alpha \|r\|_2^2 + \beta \|r - i\|_2^2 \right\}, \tag{6.7}$$

which we may split into the following linearly constraint minimization problem over two variables:

$$\min_{r,e} \left\{ \|e - \nabla_{w,f} i\|_1 + \alpha \|r\|_2^2 + \beta \|r - i\|_2^2 \right\} \quad \text{s.t.} \quad e = \nabla_w r. \tag{6.8}$$

We address the constraint by introducing the following augmented Lagrangian, which includes the constraint as both a quadratic penalty and Lagrangian multiplier term:

$$AL(r, e, \mu) = \|e - \nabla_{w,f} i\|_1 + \alpha \|r\|_2^2 + \beta \|r - i\|_2^2 + \rho \|\nabla_w r - e\|_2^2 + 2 \langle \mu, \nabla_w r - e \rangle \tag{6.9}$$

The L^1 minimization problem can now be solved by iteratively finding a saddle point to this augmented Lagrangian, iteratively along the different problem dimensions (ADMM). This process essentially involves iteration of three steps: solving

the L^2 minimization in r, shrinkage of e, updating the Lagrangian multiplier μ:

$$\begin{cases} r^{k+1} = \underset{r}{\operatorname{argmin}} \left\{ \alpha \left\| r \right\|_2^2 + \beta \left\| r - i \right\|_2^2 + \rho \left\| \nabla_w r - e^k + \dfrac{\mu^k}{\rho} \right\|_2^2 \right\} \\ \qquad = \left((\alpha + \beta) I - \rho L \right)^{-1} \left(\beta i - \rho \div_w \left(e^k - \dfrac{\mu^k}{\rho} \right) \right) & (L^2\text{-Retinex}) \\ e^{k+1} = \underset{r}{\operatorname{argmin}} \left\{ \left\| e - \nabla_{w,f} i \right\|_1 + \rho \left\| e - \nabla_w r^{k+1} - \dfrac{\mu^k}{\rho} \right\|_2^2 \right\} \\ \qquad = S^s_{1/2\rho} \left(\nabla_w r^{k+1} - \nabla_{w,f} i + \dfrac{\mu^k}{\rho} \right) + \nabla_{w,f} i & (\text{Shrinkage}) \\ \mu^{k+1} = \mu^k + \rho \left(\nabla_w r^{k+1} - e^{k+1} \right) & (\text{Dual ascent}) \end{cases} \tag{6.10}$$

Remark 19. *The most time consuming part is the update r^{k+1}. To speed up, we use a fixed ρ for every iteration as suggested by the split Bregman method [56], resulting in repeated inversion using the same system. Also, since the system matrix $((\alpha+\beta)I-\rho L)$ is sparse and strictly positive, we can use the Gauss- Seidel or conjugate gradient methods to solve r^{k+1} with a few iterations.*

L^0 Gradient Fidelity

Finally, let us look at the non-convex L^0-based gradient-fidelity optimization problem. This L^0-based problem writes:

$$\min_r \left\{ \left\| \nabla_w r - \nabla_{w,f} i \right\|_0 + \alpha \left\| r \right\|_2^2 + \beta \left\| r - i \right\|_2^2 \right\}. \tag{6.11}$$

This problem is hard to solve. We propose to treat it very similarly to the above L^1-problem and first split it into the following linearly constraint minimization problem over two variables:

$$\min_{r,e} \left\{ \left\| e - \nabla_{w,f} i \right\|_0 + \alpha \left\| r \right\|_2^2 + \beta \left\| r - i \right\|_2^2 \right\} \quad \text{s.t.} \quad e = \nabla_w r. \tag{6.12}$$

The resulting sub-optimization problems are now easy to solve, but we have no guarantee for overall convergence. Indeed, we address this problem in analogy

to the L^1 gradient fidelity, introducing both quadratic penalty and Lagrangian multiplier, and solve iteratively along each direction, where the only change occurs in the e-step:

$$\begin{cases} r^{k+1} = \underset{r}{\operatorname{argmin}} \left\{ \alpha \left\| r \right\|_2^2 + \beta \left\| r - i \right\|_2^2 + \rho^k \left\| \nabla_w r - e^k + \dfrac{\mu^k}{\rho^k} \right\|_2^2 \right\} \\ \qquad = \left((\alpha + \beta) I - \rho^k L \right)^{-1} \left(\beta i - \rho^k \div_w \left(e - \dfrac{\mu^k}{\rho^k} \right) \right) & (L^2\text{-Retinex}) \\ e^{k+1} = \underset{r}{\operatorname{argmin}} \left\{ \left\| e - \nabla_{w,f} i \right\|_0 + \rho^k \left\| e - \nabla_w r^{k+1} - \dfrac{\mu^k}{\rho^k} \right\|_2^2 \right\} \\ \qquad = S^h_{1/\sqrt{\rho^k}} \left(\nabla_w r^{k+1} - \nabla_{w,f} i + \dfrac{\mu^k}{\rho^k} \right) + \nabla_{w,f} i & (\text{Thresholding}) \\ \mu^{k+1} = \mu^k + \rho^k \left(\nabla_w r^{k+1} - e^{k+1} \right) & (\text{Dual ascent}) \\ \rho^{k+1} = \rho^k \cdot s & (\text{Step reduction}) \end{cases} \tag{6.13}$$

where $s > 1$ is a constant that essentially reduces the step size at each iteration and causes the iterative process to stabilize eventually.

Remark 20. *Note that the L^0-optimization problem is not convex. In order to enforce convergence, we add one more step to the iterative process, that is decreasing the effective step-size of the variable-split at every iteration. As a ramification, however, the matrix $(\alpha + \beta)I - \rho L$ is not constant, and the r^{k+1}-update is less efficient. Moreover, such system-matrix becomes singular at some point. Numerically, we will converge sooner and stop before that. The proof of the convergence of this scheme is still an open question. However, we can see later that despite rigorous theory, this model offers interesting new applications of Retinex in shadow detection with interesting, stable results.*

6.3 Results I: Relations to Existing Models

The first results section is dedicated to demonstrate the unifying power of the proposed non-local two-step Retinex model.

Table 6.1: Filtered non-local gradient-fidelity based approximations to existing Retinex models. Both Poisson PDE [108] and L^1-Retinex [95] employ gradient filtering natively. For the other methods, the filtered gradient reproduces a gradient sparsity term.

Model	**Norm** p	**Weights** w	**Filter** f_a	**Additional terms**
Poisson PDE [108]	L^2	local	S^h_λ	—
L^1-Retinex [95]	L^1	local	S^h_λ	—
TV-Retinex [96]	L^2	local	S^s_λ	—
Variational Retinex [76]	L^2	local	S^u_λ	$\alpha \|r\|_2^2 \quad (r \leq 0)$
TV-Retinex [111]	L^2	local	S^s_λ	$\beta \|r - i\|_2^2$
TV-L^1 [35]	L^1	local	—	$\alpha \|r\|_2^2$ ($\alpha \|r\|_1$)
Random walk/Kernel [120]	L^2	Gaussian	$S^s_{-\lambda}$	$\alpha \|r\|_2^2 + \beta \|r - i\|_2^2$

In the following paragraphs, we want to show how the existing Retinex implementations can be reproduced in our proposed, fundamental non-local Retinex model. In all these models, we can restrict ourselves to identical weights $w = w_1 = w_2 = w_3$. The different correspondences are summarized in table 6.1.

Poisson

The PDE version of Retinex [108] can be derived exactly from the L^2-version, $p = 2$, of the proposed Retinex model, under local weights w_l and gradient thresholding $f_a = S^h_\lambda$. Indeed:

$$\hat{r} = \underset{r}{\operatorname{argmin}} \left\{ \|\nabla_{w_l} r - \nabla_{w_l, f_a} i\|_2^2 \right\} \quad (L^2\text{-Retinex}) \tag{6.14}$$

implies the Euler-Lagrange equations

$$(\Delta\hat{r})(x,y) = (\Delta_{w_l,f_a} i)(x,y) \quad x,y \in \Omega. \quad \text{(Poisson PDE Retinex)} \tag{6.15}$$

L^1-Retinex

The next close relative of the proposed non-local Retinex model is its local L^1 predecessor, L^1-Retinex [95]. The closest match to L^1-Retinex in the proposed framework is obtained if we choose the weights $w_l(x,y)$ such as to reproduce the well-known local finite differences differential operators, gradient filtering $f_a = S^h_\lambda$, and with $p = 1$:

$$\hat{r} = \underset{r}{\operatorname{argmin}} \left\{ \|\nabla_{w_l} r - \nabla_{w_l,f_a} i\|_1 \right\} = \underset{r}{\operatorname{argmin}} \left\{ \int_\Omega \sum_{k=1}^{n} \left| \nabla_k r(x) - S^h_\lambda(\nabla_k i(x)) \right| dx \right\} \tag{6.16}$$

which is the anisotropic L^1-distance for local gradient fidelity.

TV Regularized Retinex

In [96], the authors propose to solve directly for an image, whose gradient is close to the observed gradient in L^2, while minimizing isotropic TV:

$$\hat{r} = \underset{r}{\operatorname{argmin}} \left\{ \|\nabla r - \nabla i\|_2^2 + 2\lambda \| |\nabla r|_2 \|_1 \right\}, \quad \text{(TV Retinex)} \tag{6.17}$$

This is the constraint L^1-relaxed gradient sparsity assumption, solved through Bregman iterations:

$$\hat{r} = \underset{r}{\operatorname{argmin}} \left\{ \|q - \nabla i\|_2^2 + 2\lambda \| |q|_2 \|_1 \right\} \quad \text{s.t.} \quad q = \nabla r \quad \text{(Bregman TV Retinex)} \tag{6.18}$$

A similar model can be obtained through the proposed general Retinex model by employing soft-shrinkage gradient filtering, $f_a = S^s_\lambda$, to which the according potential is recovered:

$$\hat{r} = \underset{r}{\operatorname{argmin}} \left\{ \|\nabla_{w_l} r - \nabla_{w_l,f_a} i\|_2^2 \right\}. \tag{6.19}$$

Again, the main difference is the use of anisotropic TV through gradient filtering in the proposed framework.

H^1+L^2 Regularized

The variational Retinex model in [76] can be rewritten exactly as

$$\hat{r} = \underset{r}{\operatorname{argmin}} \left\{ \|\nabla r - \nabla i\|_2^2 + \alpha \|r\|_2^2 + \lambda \|\nabla r\|_2^2 \right\} \quad \text{s.t.} \quad r \leq 0 \quad (H^1/L^2 \text{ Retinex}) \tag{6.20}$$

In [76], the authors motivate the L^2 term mainly as "a regularization of the problem that makes it better conditioned", and they state that "in practice this penalty term should be weak [...] and α should therefore be very small." The constraint $r < 0$ corresponds to the reset in the original Retinex theory. The constraint and L^2 norm together push the reflectance close to white.

We may find a similar problem within the proposed framework, where we choose uniform gradient scaling $f_a = S_\lambda^u$ and omit the clipping constraint:

$$\hat{r} = \underset{r}{\operatorname{argmin}} \left\{ \|\nabla_{w_l} r - \nabla_{w_l, f_a} i\|_2^2 + \alpha \|r\|_2^2 \right\}. \tag{6.21}$$

TV+L^2 Regularized

Recently, a mixture of TV regularized and Kimmel's variational approach was proposed [111]. This model essentially boils down to:

$$\hat{r} = \underset{r}{\operatorname{argmin}} \left\{ \|\nabla r - \nabla i\|_2^2 + \beta \|r - i\|_2^2 + 2\lambda \|\nabla r\|_1 \right\} \quad (\text{TV}/L^2 \text{ Retinex}) \tag{6.22}$$

Again, we may approximate this model with a similar energy based on similarity to filtered gradients, with $f_a = S_\lambda^s$:

$$\hat{r} = \underset{r}{\operatorname{argmin}} \left\{ \|\nabla_{w_l} r - \nabla_{w_l, f_a} i\|_2^2 + \beta \|r - i\|_2^2 \right\}. \tag{6.23}$$

The main difference is the use of anisotropic TV through gradient filtering in the proposed framework.

TV-L^1

The "logarithmic total variation" (LTV) model was suggested for extraction of illumination invariant features for face recognition [35]. It is defined as an TV-L^1 based on the logarithmic input image and its logarithmic illumination:

$$\hat{r} = \underset{r}{\operatorname{argmin}} \left\{ \|\nabla r - \nabla i\|_1 + \alpha \|r\|_1 \right\} \quad \text{(TV-}L^1\text{)} \tag{6.24}$$

Its equivalent in the proposed framework is found by relaxing the second term to an L^2-norm, *i.e.*, TV-L^2 Retinex.

Bertalmío Model

To approximate the perceptually inspired Retinex model through our proposed general framework, we replace or complete the intensity fidelity by a gradient fidelity:

$$\hat{r} = \underset{r}{\operatorname{argmin}} \left\{ \alpha \|r\|_2^2 + \beta \|r - i\|_2^2 + \|\nabla_w r - \nabla_w i\|_2^2 - 2\lambda \|\nabla_w r\|_1 \right\} \text{ (-TV+}L^2\text{ Retinex)} \tag{6.25}$$

This is essentially homomorphic filtering with TV in place of H^1. Again, we may now substitute by incorporating the TV-enhancement term as an input-gradient filter $f_a = S^s_{-\lambda}$:

$$\hat{r} = \underset{r}{\operatorname{argmin}} \left\{ \|\nabla_w r - \nabla_{w,f_a} i\|_2^2 + \alpha \|r\|_2^2 + \beta \|r - i\|_2^2 \right\} \tag{6.26}$$

Next, we will provide a range of results obtained using the proposed non-local Retinex model to illustrate the broad spectrum of Retinex flavors it includes.

Examples: The Logvinenko Illusion

We have applied the whole range of Retinex "modes" retrofitted above to existing Retinex implementations to a single common test image extracted from the

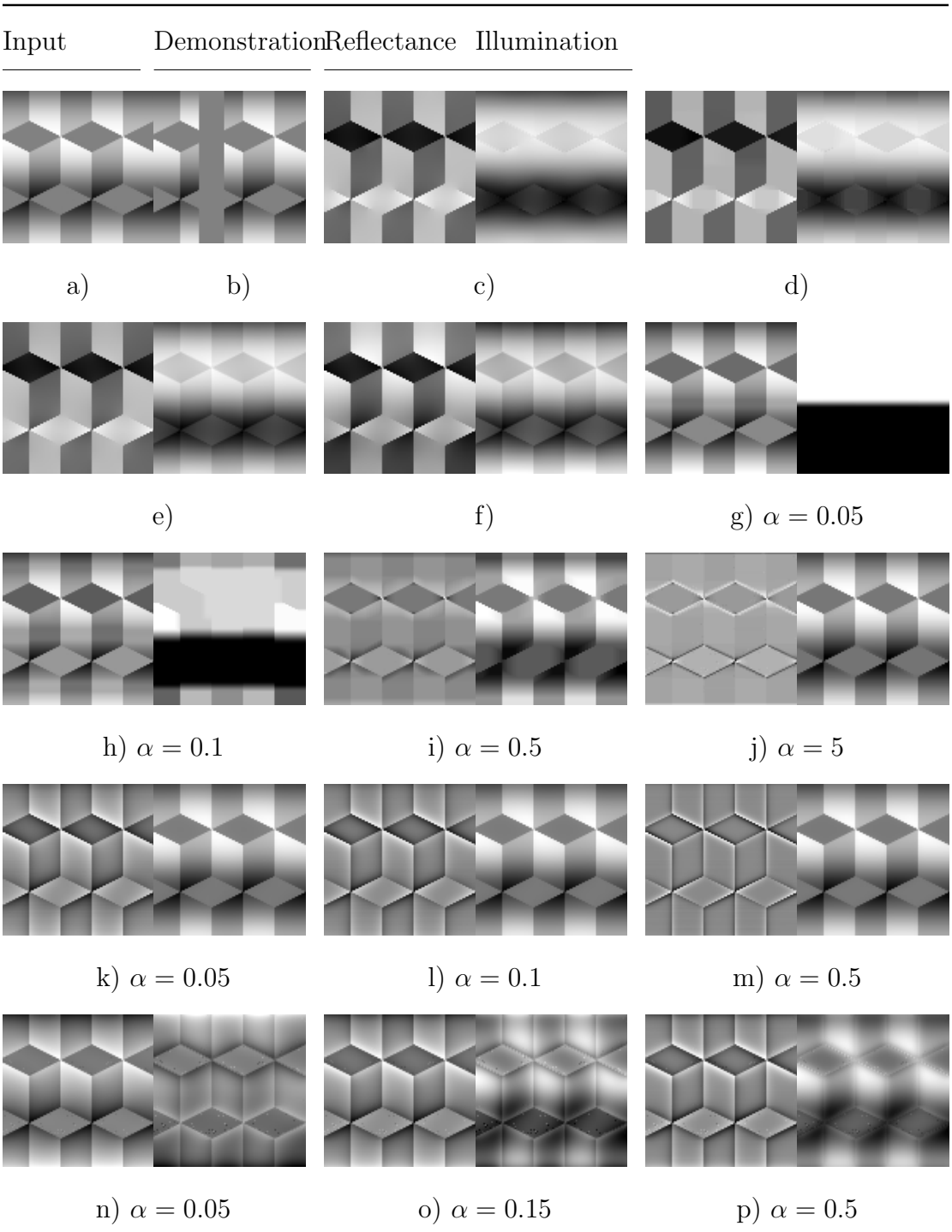


Figure 6.1: Logvinenko illusion and different Retinex decompositions

Logvinenko illusion pattern [94]. The test image is shown in Fig. 6.1a). The illusion consists of the following: due to the suggested smoothly varying lighting, the oblique grey diamonds of the upper row appear darker than the diamonds of the lower row. However, as shown in the adjacent Fig. 6.1b), their actual intensity is exactly equal. In this example, the Retinex model is expected to separate the almost smooth shading from the rough checkerboard-like reflectance, thereby truly rendering the two rows of diamonds at different reflectances.

The first model, L^2-Retinex equivalent to [108], produces the standard result in Fig. 6.1c). It can be clearly seen that in particular the lower row of diamonds is not recovered completely flat, since the illumination is not smooth everywhere. The related L^1-Retinex in Fig. 6.1d), inspired by [95], suffers from very similar artifacts. In Fig. 6.1e) we show the results of our model with parameters set to correspond to TV-regularized Retinex [96], resulting in less artifacts. Adding an L^2 fidelity-constraint ($\beta > 0$), as in [111], injects more of the initial shading into the estimated reflectance, see Fig. 6.1f).

The TV-L^1-inspired model [35] is in our case an TV-L^2 model for illumination recovery, where the TV-sparsity of the extracted illumination is tuned by the parameter α. It is clearly appreciated in Fig. 6.1g–j) that the impact of the parameter is quite severe, with higher values corresponding to the output desired for illumination invariant feature extraction. The choice of parameters inspired by Kimmel's Retinex formulation yields the output shown in Fig. 6.1k–m), which corresponds well to the behavior expected from [76]. The parameter α controls the degree of dynamic range compression applied, *i.e.*, the dominance of local contrast enhancement. Finally, in Fig. 6.1n–p) we provide the output produced by model parameters mimicking Bertalmío's perceptually inspired Retinex [10]. Here, the unshrinking of the gradients has the unpleasant effect of amplifying pixel noise.

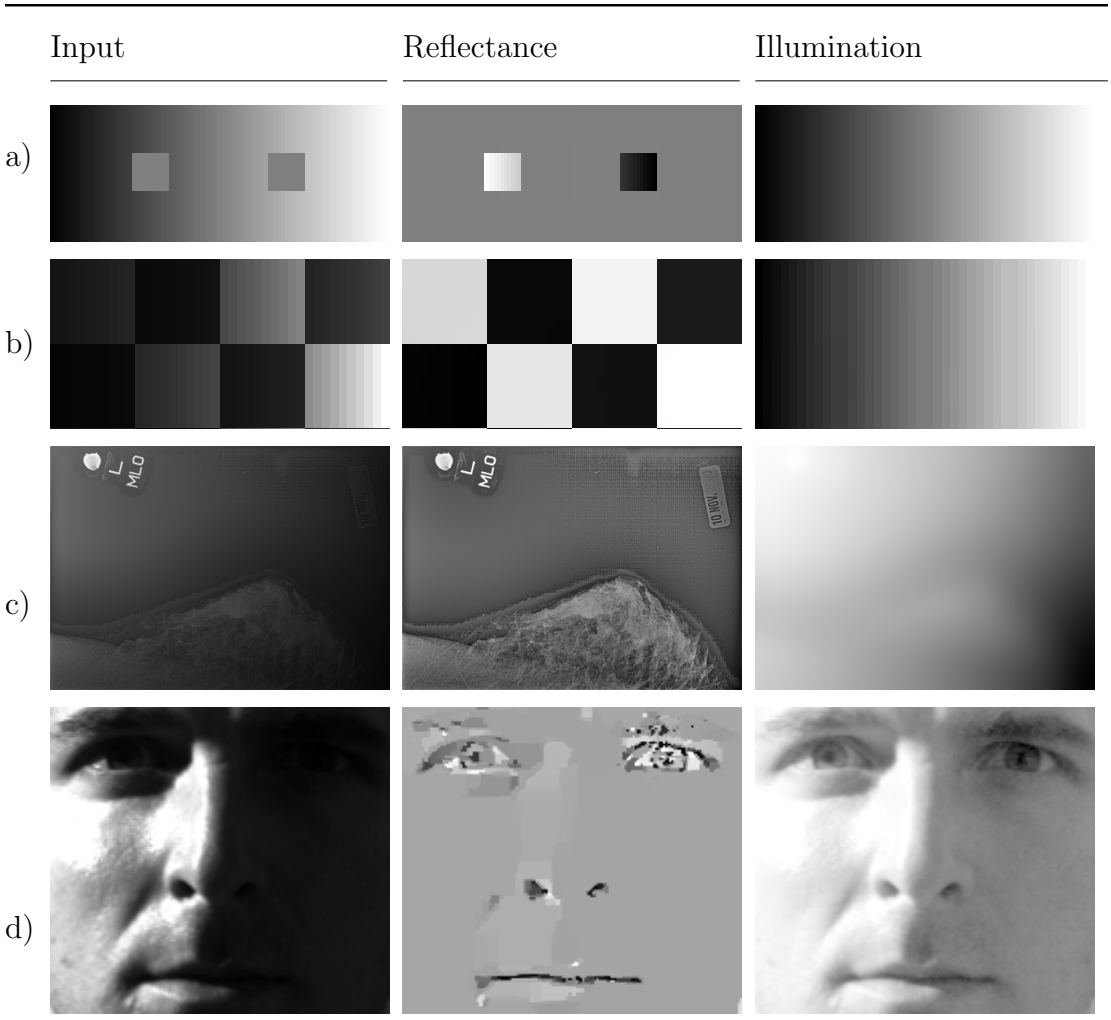


Figure 6.2: Unifying non-local Retinex model. **a–b)** Our unifying model decomposes an input image into underlying reflectance and estimated illumination, and successfully reproduces basic Retinex behavior (b/w squares and checkerboard, respectively). **c)** The same model allows dynamic range compression and local contrast enhancement (here: radiography), as well as **d)** illumination-invariant feature extraction, e.g. for face detection.

Input	Reflectance	Illumination (grey)	Output
			

Figure 6.3: HSV color Retinex. Retinex is performed on the value-channel of a color image in HSV space. The final output is obtained by applying post-processing enhancement as in [76].

Examples: Other Applications

To show the "unifying power" of the proposed Retinex model, we provide a few more illustrating examples in Fig. 6.2.

First, in Fig. 6.2a–b) two standard results are shown, that are based on L^2-gradient thresholding. As expected, we succeed in separating the smooth gradient illumination from the sharp reflectance features.

The third example in Fig. 6.2c) is an (artificially) unevenly exposed radiography, where important features are masked due to the great dynamic range. We perform center-surround-like exposure correction and dynamic range compression simply by choosing $p = 2$, wide Gaussian kernel based weights, no thresholding, and $\alpha = 0.01$.

The forth example is based on the challenge of illumination-invariant feature extraction for face recognition [35]. For this, we use $p = 1$, soft-thresholding $\lambda = 0.2$, and dynamic range compression $\alpha = 5$.

In Fig. 6.3 we provide an example inspired by Kimmel's variational Retinex [76], where color images are involved. Here, Retinex is applied to the V-channel

(lightness) of the color image (as opposed to separately on each RGB channel, as originally proposed), and we use matching parameters in our proposed framework. After application of their suggested postprocessing steps, we obtain images very similar to Kimmel's original results.

6.4 Results II: New Perspectives

Beyond reproducing existing Retinex models, our proposed framework also has the potential to yield new results thanks to it generalizing power. In this section, we explore a few new possibilities offered by choosing new sets of parameters, in particular based on $p = 0$ gradient fidelity, with applications to shadow detection and removal, and cartoon-texture decomposition.

L^0 Gradient Fidelity

In Fig. 6.1 we have shown a series of decomposition results obtained with different model configurations. The best results in terms of piecewise constant reflectance versus illumination have been achieved with the basic hard-thresholding models (L^2- and L^1-Retinex), as well as the soft-thresholding based TV-Retinex. However, all these models suffer from artifacts of illumination estimation at the edges and corners of the flat diamonds, where illumination smoothness is not a stringent enough prior. Therefore, we propose seeking for further illumination gradient penalty by choosing $p = 0$, corresponding to L^0 gradient fidelity (as opposed to TV- or H^1-sparsity of the illumination). In Fig. 6.4 we show a few results where we make use of TV-regularization of the reflectance (soft-thresholding). In particular in combination with Gaussian kernel weights, the decomposition exhibits less artifacts than previous results, see Fig. 6.4b).

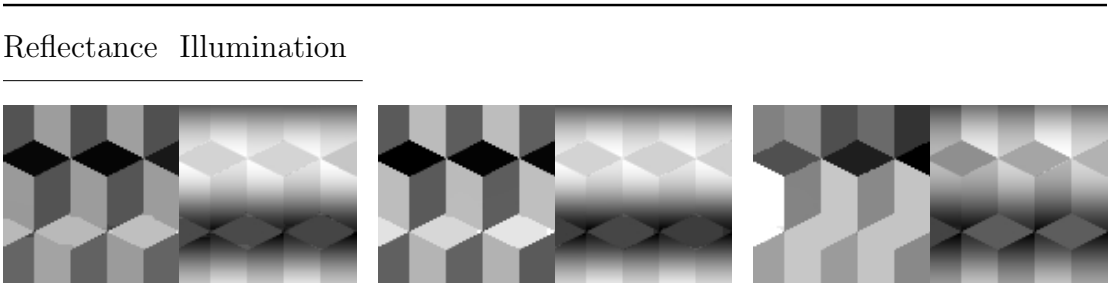


Figure 6.4: Logvinenko illusion and new L^0-based Retinex decompositions. Soft-thresholding, $p = 0$, $\alpha = \beta = 0$. **a)** Local weights. **b)** Narrow Gaussian kernel weights $\sigma^2 = 1$, $\lambda = 0.33$. **c)** Wider Gaussian kernel weights $\sigma^2 = 2$, $\lambda = 0.8$.

Shadow Detection in Natural Images

Shadow removal from a single (natural) image plays an important role in many computer vision algorithms. Most methods are based on a two-step procedure: first detect shadows, and then reconstruct shadow-free images. Shadow detection can be based on features such as intensity, gradients or texture, and even make use of supervision or training data [163, 59, 133]. Once the shadow regions have been reliably detected, several techniques aim at reconstructing shadow-free images, through matting, inpainting, or Poisson editing [133, 157].

Here, we explore the applications of the proposed unified Retinex model for shadow detection, first, and removal from a single image, second. We propose to use the L^0 gradient fidelity criterion combined with dynamic range compression, without any gradient thresholding. The unfiltered L^0 gradient fidelity is a strong prior on illumination gradient sparsity, while the dynamic range compression tends to take large intensity modulations out of the reflectance, and balance the mean intensities of inside- and outside-shadow regions. Our model can detect shadows in monochromatic and color images. We show a few example results for shadow detection in Fig. 6.5. We believe that the proposed model can largely compete

with the recent state-of-the-art shadow detection scheme proposed in [58, 59].

Moreover, to some extent our model also provides a shadow-free reflectance estimate at the same time. Reflectance output (*i.e.*, after shadow removal) is illustrated in Fig. 6.6. However, in most natural scenes, the actual border between shaded and unshaded regions is rather smooth, called the penumbra, which is due to the spatial extent of the light source. Hence the estimated shadow boundary in the proposed model is consistently overly sharp, and the estimated shadow-free reflectance image includes artifacts, see Fig. 6.6c). This problem can partially be tackled by smoothing the estimated illumination field in post-processing, as shown in Fig. 6.6d) (this is not equivalent to employing an L^2 gradient sparsity in the first place). A noticeable difference in texture is still visible, however, due to the missing specular highlights in the shadowed region, exclusively lit by ambient light.

Shadow Removal in Color Images

If the images are treated as color images, however, a few shortcomings of the simple shadow-removal model become obvious, beyond the penumbra-issue. In Fig. 6.6e) we show the output of Retinex being applied to the lightness channel in HSV-space only. Since the shadowed region was lit by (sky-blueish) ambient light only, compared to warmer direct sun light, the colorcast after intensity correction becomes really striking. If, in contrast, we perform Retinex on all three RGB channels independently, the colorcast can be successfully avoided, see Fig. 6.6f). However, since the three channels are not coupled, the respective shadow-boundaries differ slightly, creating local color-artifacts.

The observed issues are just a manifestation of a more fundamental Retinex problem when dealing with color images: namely the correct choice of color space and channels in which to perform Retinex. Traditionally, Retinex theories con-

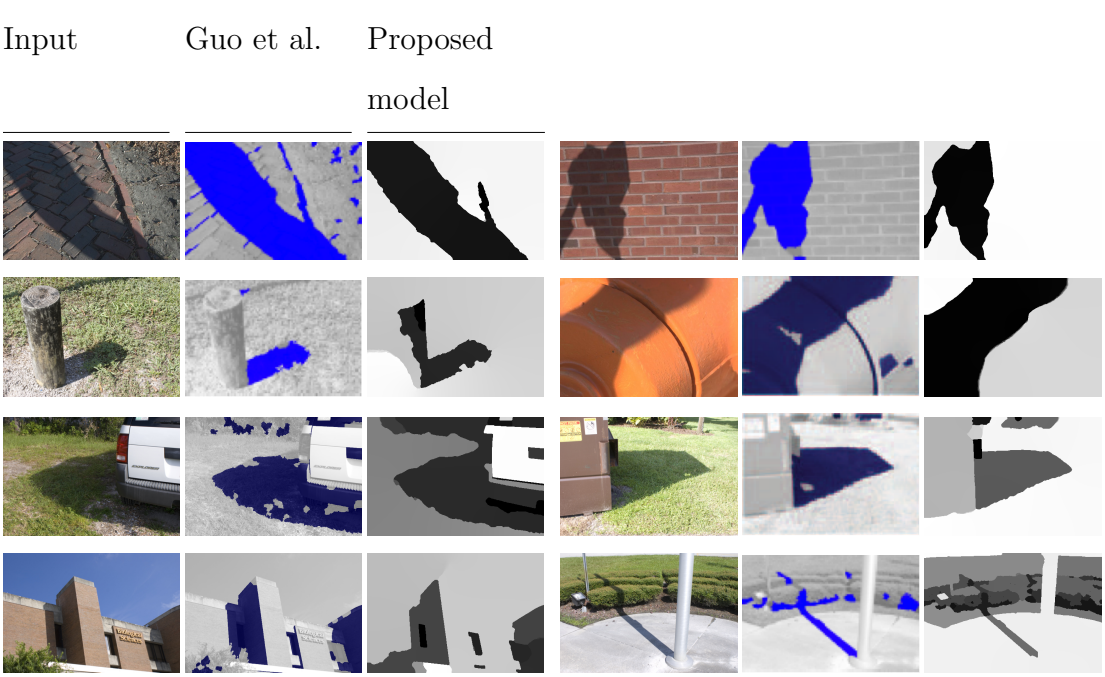


Figure 6.5: Shadow detection results. We compare the shadow detection results (illumination output) of our proposed model against the recently published results (blue mask) from [58, 59]. The results of the first row are very comparable, while we believe the examples of the second row are in favor of the proposed model. Indeed, our illumination output may be "multilevel" rather than just binary, and therefore better reflect the different nuances of shade in natural images (pole). On the other hand, our approach is less subject to local artifacts and produces more coherent shadow estimates.

Figure 6.6: Shadow removal results. **a)** Input image. **b)** Recently published results from Guo et al. [58, 59]. **c)** Reflectance output of input image reduced to grayscale. The sharp boundary of the detected shadow region creates artifacts in the penumbra. **d)** The artifacts are almost entirely removed by smoothing the estimated illumination in post-processing. **e)** If the Retinex model is applied only to the V-channel of the color image in HSV-space, then strong colorcast becomes apparent, due to different lighting color for direct and ambient light. **f)** The colorcast is avoided by correcting all three RGB channels (colorbalancing). However, local artifacts appear due to inconsistent shadow region boundaries in the three individual channels.

sider RGB-images and treat the color channels independently in order to achieve color constancy (color normalization) [86]. To some purists, this is the one and only right way. Other authors have suggested to perform illumination correction by just working on the V-channel (lightness) of images in HSV color space—Retinex is expected to correct the amount of lighting but should conserve the general tone-trend in an image. Both approaches can yield unsatisfactory results in some situations. In particular, HSV-Retinex is unable to normalize differences between different lighting-temperatures (for example, blueish ambient lighting versus directly lit parts of a scene). It was alternatively proposed to perform color correction in CIELAB colorspace, where Retinex again works on the lightness channel, while a co-correction is performed on the chroma-channels based on the estimated change of illumination [147, 146].

An intermediate compromise between channel-wise RGB and lightness-only

Figure 6.7: Cartoon-texture decomposition. For important α, the L^0 ($p = 0$) model separates texture (reflectance) from cartoon (illumination). Gaussian weights w_g with $\sigma^2 = 2$, $\lambda = 0.15$. **left:** $\alpha = 250$ **right:** $\alpha = 100$.

HSV Retinex could be devised as follows. The main advantage of RGB-Retinex is the capability of color-normalizing ("greying") the respective shadow/light regions in a scene independently. Its main drawback is the lack of coupling between shadow boundaries in the three color channels. For HSV-retinex the situation is exactly opposite. A potential solution could perform channel-wise Retinex in RGB, but using a coupling term that encourages shadow boundaries to be collocated in all three color channels. Such a goal could be achieved by replacing the current channel-wise gradient fidelity by a grouped gradient fidelity derived from group sparsity [99, 100]. More advanced, another gradient fidelity functional involving stronger channel coupling could be employed, such as the Color Beltrami energy, that aligns gradients across color channels [77].

Cartoon-texture Decomposition

The separation of an image into a piecewise regular component (cartoon) and its high-frequency parts (texture) is generally referred to as cartoon-texture decomposition [159, 115, 142]. If we give even more importance to dynamic range compression, then our proposed L^0 gradient-fidelity based Retinex model can be used to this very same end. Indeed, the "reflectance" will only contain the texture of the image, whereas all larger scale intensity patches will be attributed to illumination (cartoon part). The scale of separation is determined by the weight of

the dynamic range compression, α. In Fig. 6.7 we show results of cartoon-texture decompositions of two natural images.

6.5 Adaptive Thresholding with Texture and Color

In the preceding examples and comparisons, we always considered equal weights for gradient sparsity and fidelity, $w = w_1 = w_2$. However, doing so is not a requirement of the method, and choosing alternative weights for the two terms tremendously broadens the spectrum of possibilities. The fundamental principle is as follows: the reflectance of two pieces of same material is likely to be similar. If the non-local weights are constructed in a way that strongly connects same-material pixels, then the reflectance should have low non-local gradient magnitude. The first milestone is the definition of a suitable "material-distance" and the construction of the associated weights-graph. In the following, we quickly want to discuss two possible routes, namely the use texture and color-based weight graphs, as introduced in Section 5.3, and their connections to existing methods.

Texture-based Non-local Sparsity

Here, the assumption is that pixels belonging to objects of the same material (and thus same reflectance) are characterized by local image structure (image patches) that are similar to (at least some) other, distant patches of the same material, and dissimilar to patches around pixels of different material.

This idea of texture sparsity in reflectance has been employed in non-local retinex already, to different extents [138, 143, 96].

In the proposed, unified model, this idea could be easily implemented as follows. The reflectance gradient is expected to be non-locally sparse, therefore we would pick $w_1 = w_{nl}$. The gradient fidelity, however is required locally, thus one chooses local weights $w_2 = w_\epsilon$, or mollified semilocal weights $w_2 = w_g$. This

choice, together with an appropriate norm on both terms leads to one of the *non-stationary* gradient filters exposed in Section 5.2. In order to improve the material-similarity weights based on texture distances, illumination normalized cosine distances may be used in lieu of the classical non-local weights w_{nl}.

Chroma-conditioned Gradient Thresholding

The argument of texture-based material (dis-)similarity can easily be extended into a color-based (dis-)similarity metric, by using color-distances (see section 5.3). As pointed out earlier, classical Retinex treats the color channels independently and ignores any relations between them. In shadow-removal applications, however, this is clearly a shortcoming, since the presence of a shadow can be expected to similarly affect all color channels, and it is reasonable to exploit this interaction.

The role of threshold in the original Retinex theory is "to remove the effects of nonuniform illumination over the scene" [83]. However, those gradients are not necessarily small, for example the ones cross the shadow edges [47]. Explicitly, a large gradient belongs to reflectance if the material is different on both sides but to illumination if the material is the same. Such similar conditional thresholdings have been used in shadow removal [47, 48], and appeared in intrinsic image decomposition [16, 139, 107, 162, 34]. All these models aim at enforcing reflectance gradient sparsity between pixels that have similar hue, since these are believed to belong to objects of the same material and thus similar reflectance. Conversely, reflectance gradients are conserved only, if they are "motivated" by a hue gradient, suggesting a material boundary.

Such a method can easily be implemented within the non-local Retinex framework presented in this paper. Indeed, hue constancy requires sparsity of reflectance under illumination invariant but hue-sensitive weights. Thus one would choose $w_1 = w_c$. In contrast, the gradient fidelity weights can again be chosen

Figure 6.8: The original, our retinex output and our illumination layers using adaptive thresholding, $\lambda = 0.015, \alpha = 0.4, \beta = 0.6$.

locally, $w_2 = w_\epsilon$, or semilocally, $w_2 = w_g$. The resulting non-stationary filter function thresholds the observed gradients according to these criteria, and then a reflectance gradient is fitted to this thresholded target.

Conclusion

In the last three chapters, we have provided an overview of Retinex implementations existing in literature, and unified them within a single computational framework. Using non-local differential operators, the proposed unifying framework is a generalization of the well-known threshold-based two-step Retinex implementations. We also presented numerical implementations to solve the proposed non-local Retinex model in different configurations using variable splitting and the alternate direction method of multipliers.

Using the proposed Retinex functional, we are able to expose relations with all major classes of existing Retinex implementations, such as kernel and variational Retinex, perceptual-contrast enhancement, and threshold-based PDE-Retinex. Moreover, our proposed framework offers potential for new forms of Retinex. In particular, the L^0-based non-local Retinex model produces interesting results in shadow detection and removal. On the other hand, an important, yet largely unexplored, property of the proposed non-local Retinex model is the use of hybrid weights in the thresholding part; more specifically, different weight functions

can be employed in the gradient sparsity and gradient fidelity component, leading to a gradient filtering function that is spatially varying (non-stationary, adaptive thresholding). In this thesis, we presented one example of this form which is the L^1-based non-local Retinex model with adaptive thresholdings and its application in color contrast enhancement.

References

[1] V. Agarwal, B. R. Abidi, A. Koschan, and M. A. Abidi. An overview of color constancy algorithms. *Journal of Pattern Recognition Research*, 1(1):42–54, 2006.

[2] P. Arias, G. Facciolo, V. Caselles, and G. Sapiro. A variational framework for exemplar-based image inpainting. *International journal of computer vision*, 93(3):319–347, 2011.

[3] H. Arodź, P. Klimas, and T. Tyranowski. Scaling, self-similar solutions and shock waves for V-shaped field potentials. *Physical Review E*, 73(4):046609, 2006.

[4] H. Arodź, P. Klimas, and T. Tyranowski. Compact oscillons in the signum-Gordon model. *Physical Review D*, 77(4):047701, 2008.

[5] H. Arodź and Z. Świerczyński. Swaying oscillons in the signum-Gordon model. *Physical Review D*, 84(6):067701, 2011.

[6] E. A. Ashton, B. D. Wemett, R. A. Leathers, and T. V. Downes. A novel method for illumination suppression in hyperspectral images. In *SPIE Defense and Security Symposium*, pages 69660C–69660C. International Society for Optics and Photonics, 2008.

[7] L. Badea. Convergence rate of a multiplicative Schwarz method for strongly nonlinear variational inequalities. In *Analysis and optimization of differential systems*, pages 31–41. Springer, 2003.

[8] L. Badea, X.-C. Tai, and J. Wang. Convergence rate analysis of a multiplicative Schwarz method for variational inequalities. *SIAM Journal on Numerical Analysis*, 41(3):1052–1073, 2003.

[9] D. Barash. Fundamental relationship between bilateral filtering, adaptive smoothing, and the nonlinear diffusion equation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(6):844–847, 2002.

[10] M. Bertalmío, V. Caselles, and E. Provenzi. Issues about retinex theory and contrast enhancement. *International Journal of Computer Vision*, 83(1):101–119, 2009.

[11] M. Bertalmío, V. Caselles, E. Provenzi, and A. Rizzi. Perceptual color correction through variational techniques. *Image Processing, IEEE Transactions on*, 16(4):1058–1072, 2007.

[12] M. Bertalmío and J. D. Cowan. Implementing the retinex algorithm with Wilson–Cowan equations. *Journal of Physiology-Paris*, 103(1):69–72, 2009.

[13] D. P. Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Academic press, 2014.

[14] A. Blake. Boundary conditions for lightness computation in Mondrian world. *Computer Vision, Graphics, and Image Processing*, 32(3):314–327, 1985.

[15] A. Blake. On lightness computation in Mondrian world. *in Central and Peripheral Mechanisms of Colour Vision, D. Ottoson and S. Zeki, eds., MacMillan, New York*, pages 45–49, 1985.

[16] A. Bousseau, S. Paris, and F. Durand. User-assisted intrinsic images. In *ACM Transactions on Graphics (TOG)*, volume 28, pages 130: 1–10. ACM, 2009.

[17] F. Bozorgnia. Numerical solutions of a two-phase membrane problem. *Applied numerical mathematics*, 61(1):92–107, 2011.

[18] G. Brelstaff and A. Blake. Computing lightness. *Pattern Recognition Letters*, 5(2):129–138, 1987.

[19] P. C. Bressloff, J. D. Cowan, M. Golubitsky, P. J. Thomas, and M. C. Wiener. Geometric visual hallucinations, euclidean symmetry and the functional architecture of striate cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 356(1407):299–330, 2001.

[20] H. Brezis. *Operateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, volume 5. Elsevier, 1973.

[21] H. Brezis. Monotone operators, nonlinear semigroups and applications. In *Proceedings of the International Congress of Mathematicians (Vancouver, BC, 1974)*, volume 2, pages 249–255, 1974.

[22] H. Brezis. Solutions with compact support of variational inequalities. *Russian Mathematical Surveys*, 29(2):103–108, 1974.

[23] H. Brézis and A. Friedman. Estimates on the support of solutions of parabolic variational inequalities. *Illinois Journal of Mathematics*, 20(1):82–97, 1976.

[24] S. L. Brunton, J. H. Tu, I. Bright, and J. N. Kutz. Compressive sensing and low-rank libraries for classification of bifurcation regimes in nonlinear dynamical systems. *SIAM Journal on Applied Dynamical Systems*, 13(4):1716–1732, 2014.

[25] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 60–65. IEEE, 2005.

[26] L. Caffarelli. The obstacle problem, lezioni fermiane. *Accademia Nazionale dei Lincei, Rome*, 1998.

[27] L. A. Caffarelli. The regularity of elliptic and parabolic free boundaries. *Bulletin of the American Mathematical Society*, 82(4):616–618, 1976.

[28] L. A. Caffarelli. The obstacle problem revisited. *Journal of Fourier Analysis and Applications*, 4(4):383–402, 1998.

[29] R. E. Caflisch, S. J. Osher, H. Schaeffer, and G. Tran. PDEs with compressed solutions. *Communications in Mathematical Sciences*, To Appear.

[30] E. J. Candes, Y. C. Eldar, T. Strohmer, and V. Voroninski. Phase retrieval via matrix completion. *SIAM Journal on Imaging Sciences*, 6(1):199–225, 2013.

[31] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011.

[32] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, 2006.

[33] J. A.-A. Cellina and J. P. Aubin. *Differential inclusions. Set-valued maps and viability theory*. Springer-Verlag New York, Inc, 1984.

[34] C.-H. Chang, Y.-T. Cheng, and Y.-Y. Chuang. A non-local sparse model for intrinsic images. In *Pattern Recognition (ACPR), 2013 2nd IAPR Asian Conference on*, pages 100–104. IEEE, 2013.

[35] T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, and T. S. Huang. Total variation models for variable lighting face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(9):1519–1524, 2006.

[36] F. R. Chung. *Spectral graph theory*, volume 92. American Mathematical Soc., 1997.

[37] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In *Fixed-point algorithms for inverse problems in science and engineering*, pages 185–212. Springer, 2011.

[38] F. Conrad, R. Herbin, and H. D. Mittelmann. Approximation of obstacle problems by continuation methods. *SIAM journal on numerical analysis*, 25(6):1409–1431, 1988.

[39] J. D. Cowan and P. C. Bressloff. Visual cortex and the retinex algorithm. In *Electronic Imaging 2002*, pages 278–285. International Society for Optics and Photonics, 2002.

[40] M. G. Crandall and A. Pazy. Semi-groups of nonlinear contractions and dissipative sets. *Journal of functional analysis*, 3(3):376–418, 1969.

[41] A. d'Aspremont, L. El Ghaoui, M. I. Jordan, and G. R. Lanckriet. A direct formulation for sparse PCA using semidefinite programming. *SIAM review*, 49(3):434–448, 2007.

[42] D. L. Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.

[43] D. M. Eagleman. Visual illusions and neurobiology. *Nature Reviews Neuroscience*, 2(12):920–926, 2001.

[44] M. Elad, R. Kimmel, D. Shaked, and R. Keshet. Reduced complexity retinex algorithm via the variational approach. *Journal of visual communication and image representation*, 14(4):369–388, 2003.

[45] C. M. Elliott and V. Janovskỳ. A variational inequality approach to Hele-Shaw flow with a moving boundary. *Proceedings of the Royal Society of Edinburgh: Section A Mathematics*, 88(1-2):93–107, 1981.

[46] A. Fey, L. Levine, and Y. Peres. Growth rates and explosions in sandpiles. *Journal of Statistical Physics*, 138(1-3):143–159, 2010.

[47] G. D. Finlayson, S. D. Hordley, and M. S. Drew. Removing shadows from images using retinex. In *Color and Imaging Conference*, number 1, pages 73–79. Society for Imaging Science and Technology, 2002.

[48] G. D. Finlayson, S. D. Hordley, C. Lu, and M. S. Drew. On the removal of shadows from images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(1):59–68, 2006.

[49] J. A. Frankle and J. J. McCann. Method and apparatus for lightness imaging, May 17 1983. US Patent 4,384,336.

[50] M. P. Friedlander and P. Tseng. Exact regularization of convex programs. *SIAM Journal on Optimization*, 18(4):1326–1350, 2007.

[51] A. Friedman. *Variational principles and free-boundary problems*. Courier Corporation, 2010.

[52] B. Funt, F. Ciurea, and J. McCann. Retinex in MATLAB. *Journal of electronic imaging*, 13(1):48–57, 2004.

[53] G. Gilboa and S. Osher. Nonlocal operators with applications to image processing. *Multiscale Modeling & Simulation*, 7(3):1005–1028, 2008.

[54] R. Glowinski. *Numerical methods for nonlinear variational problems*, volume 4. New York: Springer-Verlag, 1984.

[55] R. Glowinski and P. Le Tallec. *Augmented Lagrangian and operator-splitting methods in nonlinear mechanics*, volume 9. SIAM, 1989.

[56] T. Goldstein and S. Osher. The split Bregman method for L1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

[57] S. Guiasu and A. Shenitzer. The principle of maximum entropy. *The mathematical intelligencer*, 7(1):42–48, 1985.

[58] R. Guo, Q. Dai, and D. Hoiem. Single-image shadow detection and removal using paired regions. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 2033–2040. IEEE, 2011.

[59] R. Guo, Q. Dai, and D. Hoiem. Paired regions for shadow detection and removal. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(12):2956–2967, 2013.

[60] B. Gustafsson. Applications of variational inequalities to a moving boundary problem for Hele Shaw flows. *SIAM journal on mathematical analysis*, 16(2):279–300, 1985.

[61] M. Hintermüller, K. Ito, and K. Kunisch. The primal-dual active set strategy as a semismooth Newton method. *SIAM Journal on Optimization*, 13(3):865–888, 2002.

[62] M. Hintermüller, V. A. Kovtunenko, and K. Kunisch. Obstacle problems with cohesion: a hemivariational inequality approach and its efficient numerical solution. *SIAM Journal on Optimization*, 21(2):491–516, 2011.

[63] A. N. Hirani, K. Kalyanaraman, and S. Watts. Least squares ranking on graphs. *arXiv preprint arXiv:1011.1716*, 2010.

[64] R. H. Hoppe. Multigrid algorithms for variational inequalities. *SIAM journal on numerical analysis*, 24(5):1046–1065, 1987.

[65] R. H. Hoppe and R. Kornhuber. Adaptive multilevel methods for obstacle problems. *SIAM journal on numerical analysis*, 31(2):301–323, 1994.

[66] B. K. Horn. Determining lightness from an image. *Computer graphics and image processing*, 3(4):277–299, 1974.

[67] T. Y. Hou, Z. Li, S. Osher, and H. Zhao. A hybrid method for moving interface problems with application to the Hele–Shaw flow. *Journal of Computational Physics*, 134(2):236–252, 1997.

[68] A. Hurlbert. Formal connections between lightness algorithms. *JOSA A*, 3(10):1684–1693, 1986.

[69] D.-G. Hwang, W.-R. Lee, Y.-J. Oh, and B.-M. Jun. Frankle-McCann retinex by shuffling. In *Convergence and Hybrid Information Technology*, pages 381–388. Springer, 2012.

[70] E. T. Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.

[71] D. Jerison and I. Kim. The one-phase Hele-Shaw problem with singularities. *The Journal of Geometric Analysis*, 15(4):641–667, 2005.

[72] X. Jiang, L.-H. Lim, Y. Yao, and Y. Ye. Statistical ranking and combinatorial Hodge theory. *Mathematical Programming*, 127(1):203–244, 2011.

[73] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *Image Processing, IEEE Transactions on*, 6(7):965–976, 1997.

[74] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell. Properties and performance of a center/surround retinex. *Image Processing, IEEE Transactions on*, 6(3):451–462, 1997.

[75] T. Kato. Accretive operators and nonlinear evolution equations in Banach spaces. In *Proc. Symp. in Pure Math*, volume 18, pages 138–161, 1970.

[76] R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel. A variational framework for retinex. *International Journal of computer vision*, 52(1):7–23, 2003.

[77] R. Kimmel, R. Malladi, and N. Sochen. Images as embedded maps and minimal surfaces: movies, color, texture, and volumetric medical images. *International Journal of Computer Vision*, 39(2):111–129, 2000.

[78] D. Kinderlehrer and G. Stampacchia. *An introduction to variational inequalities and their applications*, volume 31. SIAM, 1980.

[79] Y. Komura. Differentiability of nonlinear semigroups. *Journal of the Mathematical Society of Japan*, 21(3):375–402, 1969.

[80] R. Kornhuber. Monotone multigrid methods for elliptic variational inequalities i. *Numerische Mathematik*, 69(2):167–184, 1994.

[81] R. Kornhuber. Monotone multigrid methods for elliptic variational inequalities ii. *Numerische Mathematik*, 72(4):481–499, 1996.

[82] E. H. Land. The retinex. *American Scientist*, pages 247–264, 1964.

[83] E. H. Land. Recent advances in retinex theory and some implications for cortical computations: color vision and the natural image. *Proceedings of the National Academy of Sciences of the United States of America*, 80(16):5163, 1983.

[84] E. H. Land. An alternative technique for the computation of the designator in the retinex theory of color vision. *Proceedings of the National Academy of Sciences*, 83(10):3078–3080, 1986.

[85] E. H. Land et al. *The retinex theory of color vision.* Scientific America., 1977.

[86] E. H. Land and J. McCann. Lightness and retinex theory. *JOSA*, 61(1):1–11, 1971.

[87] P. D. Lax. *Hyperbolic systems of conservation laws and the mathematical theory of shock waves*, volume 11. SIAM, 1973.

[88] J.-S. Lee. Digital image smoothing and the sigma filter. *Computer Vision, Graphics, and Image Processing*, 24(2):255–269, 1983.

[89] L. Levine, W. Pegden, and C. K. Smart. Apollonian structure in the Abelian sandpile. *arXiv preprint arXiv:1208.4839*, 2012.

[90] L. Levine, W. Pegden, and C. K. Smart. The apollonian structure of integer superharmonic matrices. *arXiv preprint arXiv:1309.3267*, 2013.

[91] L. Levine and Y. Peres. Strong spherical asymptotics for rotor-router aggregation and the divisible sandpile. *Potential Analysis*, 30(1):1–27, 2009.

[92] L. Levine and Y. Peres. Scaling limits for internal aggregation models with multiple sources. *Journal d'Analyse Mathématique*, 111(1):151–219, 2010.

[93] P.-L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.

[94] A. D. Logvinenko. Lightness induction revisited. *PERCEPTION-LONDON-*, 28:803–816, 1999.

[95] W. Ma, J.-M. Morel, S. Osher, and A. Chien. An L1-based variational model for retinex theory and its application to medical images. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 153–160. IEEE, 2011.

[96] W. Ma and S. Osher. A TV Bregman iterative model of retinex theory. *UCLA CAM Report*, pages 10–13, 2010.

[97] A. Mackey, H. Schaeffer, and S. Osher. On the compressive spectral method. *Multiscale Modeling & Simulation*, 12(4):1800–1827, 2014.

[98] K. Majava and X.-C. Tai. A level set method for solving free boundary problems associated with obstacles. *Int. J. Numer. Anal. Model*, 1(2):157–171, 2004.

[99] A. Majumdar and R. K. Ward. Compressive color imaging with group-sparsity on analysis prior. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 1337–1340. IEEE, 2010.

[100] A. Majumdar and R. K. Ward. Non-convex group sparsity: application to color imaging. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 469–472. IEEE, 2010.

[101] O. Mangasarian. Sufficiency of exact penalty minimization. *SIAM journal on Control and Optimization*, 23(1):30–37, 1985.

[102] D. Marr. The computation of lightness by the primate retina. *Vision Research*, 14(12):1377–1388, 1974.

[103] J. McCann. Lessons learned from mondrians applied to real images and color gamuts. In *Color and Imaging Conference*, number 1, pages 1–8. Society for Imaging Science and Technology, 1999.

[104] J. J. McCann. Capturing a black cat in shade: past and present of retinex color appearance models. *Journal of Electronic Imaging*, 13(1):36–47, 2004.

[105] J. J. McCann, C. Parraman, and A. Rizzi. Reflectance, illumination, and appearance in color constancy. *Frontiers in psychology*, 5, 2014.

[106] J. J. McCann and A. Rizzi. *The art and science of HDR imaging*, volume 26. John Wiley & Sons, 2011.

[107] J. M. Morel. Personal communication, 2012.

[108] J. M. Morel, A. B. Petro, and C. Sbert. A PDE formalization of retinex theory. *Image Processing, IEEE Transactions on*, 19(11):2825–2837, 2010.

[109] L. J. Nelson, G. L. Hart, F. Zhou, and V. Ozoliņš. Compressive sensing as a paradigm for building physics models. *Physical Review B*, 87(3):035125, 2013.

[110] Y. Nesterov. Introductory lectures on convex optimization: a basic course. 2004.

[111] M. K. Ng and W. Wang. A total variation model for retinex. *SIAM Journal on Imaging Sciences*, 4(1):345–365, 2011.

[112] Q. Nie and F.-R. Tian. Singularities in Hele–Shaw flows driven by a multipole. *SIAM Journal on Applied Mathematics*, 62(2):385–406, 2001.

[113] J. Nocedal and S. J. Wright. Penalty and augmented Lagrangian methods. *Numerical Optimization*, pages 497–528, 2006.

[114] S. Osher and J. A. Sethian. Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. *Journal of computational physics*, 79(1):12–49, 1988.

[115] S. Osher, A. Solé, and L. Vese. Image decomposition and restoration using total variation minimization and the H1. *Multiscale Modeling & Simulation*, 1(3):349–370, 2003.

[116] B. Osting, C. Brune, and S. Osher. Optimal data collection for improved rankings expose well-connected graphs. *stat*, 1050:26, 2012.

[117] B. Osting, J. Darbon, and S. Osher. Statistical ranking using the l1-norm on graphs. AIMS, 2012.

[118] V. Ozoliņš, R. Lai, R. Caflisch, and S. Osher. Compressed modes for variational problems in mathematics and physics. *Proceedings of the National Academy of Sciences*, 110(46):18368–18373, 2013.

[119] V. Ozoliņš, R. Lai, R. Caflisch, and S. Osher. Compressed plane waves yield a compactly supported multiresolution basis for the laplace operator. *Proceedings of the National Academy of Sciences*, 111(5):1691–1696, 2014.

[120] R. Palma-Amestoy, E. Provenzi, M. Bertalmío, and V. Caselles. A perceptually inspired variational framework for color enhancement. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(3):458–474, 2009.

[121] W. Pegden and C. K. Smart. Convergence of the abelian sandpile. *Duke Mathematical Journal*, 162(4):627–642, 2013.

[122] A. Petrosyan, H. Shahgholian, and N. N. Ural'ceva. *Regularity of free boundaries in obstacle-type problems*, volume 136. American Mathematical Soc., 2012.

[123] E. Provenzi, M. Fierro, A. Rizzi, L. De Carli, D. Gadia, and D. Marini. Random spray retinex: a new retinex implementation to investigate the local properties of the model. *Image Processing, IEEE Transactions on*, 16(1):162–171, 2007.

[124] E. Provenzi, C. Gatta, M. Fierro, and A. Rizzi. A spatially variant white-patch and gray-world method for color image enhancement driven by local contrast. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(10):1757–1770, 2008.

[125] E. Provenzi, D. Marini, L. De Carli, and A. Rizzi. Mathematical definition and analysis of the retinex algorithm. *JOSA A*, 22(12):2613–2621, 2005.

[126] X. Qi, R. Luo, and H. Zhao. Sparse principal component analysis by choice of norm. *Journal of multivariate analysis*, 114:127–160, 2013.

[127] Z.-u. Rahman, D. J. Jobson, and G. A. Woodell. Retinex processing for automatic image enhancement. *Journal of Electronic Imaging*, 13(1):100–110, 2004.

[128] F. Ratliff. *Mach bands: quantitative studies on neural networks in the retina.* Holden-Day, 1965.

[129] A. Rizzi, C. Gatta, and D. Marini. A new algorithm for unsupervised global and local color correction. *Pattern Recognition Letters*, 24(11):1663–1677, 2003.

[130] A. Rizzi, C. Gatta, and D. Marini. From retinex to automatic color equalization: issues in developing a new algorithm for unsupervised color equalization. *Journal of Electronic Imaging*, 13(1):75–84, 2004.

[131] J.-F. Rodrigues. *Obstacle problems in mathematical physics*. Elsevier, 1987.

[132] A. Roussos and P. Maragos. Tensor-based image diffusions derived from generalizations of the total variation and Beltrami functionals. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 4141–4144. IEEE, 2010.

[133] A. Sanin, C. Sanderson, and B. C. Lovell. Shadow detection: A survey and comparative evaluation of recent methods. *Pattern recognition*, 45(4):1684–1695, 2012.

[134] H. Schaeffer, R. Caflisch, C. D. Hauck, and S. Osher. Sparse dynamics for partial differential equations. *Proceedings of the National Academy of Sciences*, 110(17):6634–6639, 2013.

[135] R. Scholz. Numerical solution of the obstacle problem by the penalty method. *Computing*, 32(4):297–306, 1984.

[136] S. Setzer. Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage. In *Scale space and variational methods in computer vision*, pages 464–476. Springer, 2009.

[137] H. Shahgholian, N. Uraltseva, and G. S. Weiss. The two-phase membrane problem regularity of the free boundaries in higher dimensions. *International mathematics research notices*, 2007:rnm026, 2007.

[138] L. Shen, P. Tan, and S. Lin. Intrinsic image decomposition with non-local texture cues. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–7. IEEE, 2008.

[139] L. Shen and C. Yeo. Intrinsic images decomposition using a local and global sparse representation of reflectance. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 697–704. IEEE, 2011.

[140] R. Sobol. Improving the retinex algorithm for rendering wide dynamic range photographs. *Journal of Electronic Imaging*, 13(1):65–74, 2004.

[141] N. Sochen, R. Kimmel, and A. M. Bruckstein. Diffusions and confusions in signal and image processing. *Journal of Mathematical Imaging and Vision*, 14(3):195–209, 2001.

[142] J.-L. Starck, M. Elad, and D. L. Donoho. Image decomposition via the combination of sparse representations and a variational approach. *Image Processing, IEEE Transactions on*, 14(10):1570–1582, 2005.

[143] V. Štruc and N. Pavešić. *Illumination invariant face recognition by non-local smoothing*. Springer, 2009.

[144] X.-C. Tai. Rate of convergence for some constraint decomposition methods for nonlinear variational inequalities. *Numerische Mathematik*, 93(4):755–786, 2003.

[145] F. R. Tian and Q. Nie. Singularities in Hele–Shaw flows. *SIAM Journal on Applied Mathematics*, 58(1):34–54, 1998.

[146] M. Toda and M. Tsukada. High dynamic range rendering method for YUV images with global luminance correction. In *Consumer Electronics (ICCE), 2011 IEEE International Conference on*, pages 255–256. IEEE, 2011.

[147] M. Toda, M. Tsukada, A. Inoue, and T. Suzuki. High dynamic range rendering for YUV images with a constraint on perceptual chroma preservation. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 1817–1820. IEEE, 2009.

[148] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Computer Vision, 1998. Sixth International Conference on*, pages 839–846. IEEE, 1998.

[149] G. Tran, H. Schaeffer, W. M. Feldman, and S. J. Osher. An L1 penalty method for general obstacle problems. *SIAM Journal on Applied Mathematics*, To Appear.

[150] R. Tremolieres, J.-L. Lions, and R. Glowinski. *Numerical analysis of variational inequalities*. Elsevier, 2011.

[151] U. Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

[152] R. Wallis. An approach to the space variant restoration and enhancement of images. In *Proc. of symp. on current mathematical problems in image science, naval postgraduate school, Monterey CA, USA, November*, pages 329–340, 1976.

[153] A. Weiss and B. Wohlmuth. A posteriori error estimator and error control for contact problems. *Mathematics of Computation*, 78(267):1237–1267, 2009.

[154] A. Wetzler and R. Kimmel. Efficient Beltrami flow in patch-space. In *Scale Space and Variational Methods in Computer Vision*, pages 134–143. Springer, 2012.

[155] H. R. Wilson and J. D. Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical journal*, 12(1):1, 1972.

[156] H. R. Wilson and J. D. Cowan. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, 13(2):55–80, 1973.

[157] T.-P. Wu, C.-K. Tang, M. S. Brown, and H.-Y. Shum. Natural shadow matting. *ACM Transactions on Graphics (TOG)*, 26(2):8, 2007.

[158] Q. Yang, K.-H. Tan, and N. Ahuja. Shadow removal using bilateral filtering. *Image Processing, IEEE Transactions on*, 21(10):4361–4368, 2012.

[159] W. Yin, D. Goldfarb, and S. Osher. Image cartoon-texture decomposition and feature selection using the total variation regularized L1 functional. In *Variational, Geometric, and Level Set Methods in Computer Vision*, pages 73–84. Springer, 2005.

[160] W. Yin and S. Osher. Error forgetting of Bregman iteration. *Journal of Scientific Computing*, 54(2-3):684–695, 2013.

[161] W.-C. Zhang, X.-J. An, and S.-D. Pan. An improved recursive retinex for rendering high dynamic range photographs. In *Wavelet Analysis and Pattern Recognition (ICWAPR), 2011 International Conference on*, pages 121–125. IEEE, 2011.

[162] Q. Zhao, P. Tan, Q. Dai, L. Shen, E. Wu, and S. Lin. A closed-form solution to retinex with nonlocal texture constraints. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(7):1437–1444, 2012.

[163] J. Zhu, K. G. Samuel, S. Z. Masood, and M. F. Tappen. Learning to recognize shadows in monochromatic natural images. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 223–230. IEEE, 2010.

[164] D. Zosso, G. Tran, and S. Osher. A unifying retinex model based on non-local differential operators. In *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2013.

[165] D. Zosso, G. Tran, and S. Osher. Non-local retinex - a unifying framework and beyond. *SIAM Journal on Imaging Sciences*, 8(2):787–826, 2015.