Constructing and Solving Variational Image Registration Problems

Nathan D. Cahill
St Cross College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity Term 2009

Acknowledgements

A great number of people have helped and supported me in the years prior to and during my doctoral research. It is tempting to think that this number may be uncountable, as any attempt to list all of these people would surely omit someone important. Of course the list is actually countable, but this still leaves the problem of the Oxford Engineering Science Department's page limit.

First, I would like to thank my two supervisors: Alison Noble (Oxford) and Dave Hawkes (UCL). Alison's constant advice on focusing my work in conjunction with Dave's constant advice on shooting for the moon have served as an excellent primal-dual approach to doctoral research. Next, I would like to thank my examiners, Mike Brady and Gary Christensen for a lively viva. I enjoyed getting kicked around like a football.

I would also like to thank my previous industrial supervisor, Larry Ray, for much encouragement over the years, and for allowing me to participate in a joint project between UCL and Kodak European Research (and subsequently, Carestream Health) in order to remain employed while pursuing my doctoral research.

In conjunction with the joint UCL/Kodak project, I would like to thank Alan Payne, Graham Kiddle, and Hani Muammar, previously of Kodak European Research, for welcoming me to the UK, and for helping me learn to cope with simultaneously wearing industrial and academic hats. From the UCL side, I would like to thank John Hipwell and Christine Tanner for the enjoyable collaboration on project work, the many stimulating discussions and the periodic technical reviews of my research.

In the Oxford lab, I'd like to thank many of the previous and current DPhil students, postdocs, and lecturers, with whom I've enjoyed many hours in the lab as well as lively discussions in various pubs. These people include, among others: Julia Schnabel, Vicente Grau, Grace Vesom, Matthias Hofmann, Ramón Casero, Dom Van de Sompel, Mike Kadour,

Niranjan Joshi, Rohan Loveland, Rehan Ali, Kate Hibbs, Cat Kelly, and Mark Gooding.

In the various labs I've visited, I'd like to thank the following people for their hospitality and for some very stimulating conversations: Jan Modersitzki, Jeff Fessler, Matthias Hofmann, Stefan Heldmann, and Bernd Fischer. I'd like to especially thank Lena Gorelick and her family for opening up their home to me during an extended visit to the Weizmann Institute.

For a number of helpful discussions I've had that have helped me develop research ideas, I'd like to thank Bill Crum, Graeme Penney, Tim Carter, Torsten Rohlfing, Michael Unser, Fei Wang, and Baba Vemuri.

For everything I've learned about working successfully with others in an industrial research setting, I'd like to thank the following people who either mentored me or with whom I have had close collaborative and personal relationships: John D'Errico, Geoff Woolfe, John Spence, Shoupu Chen, Joe Revelli, Rick Simon, Gabriel Fielding, and Andy Gallagher.

For their close friendships, their empathy with the Oxford DPhil experience, and for our countless rounds of drinks at local pubs, I'd like to thank my Oxford flatmates Matt Marshall and Richard Branch. For her friendship and understanding of the American-in-Oxford experience, I'd like to thank Grace Vesom. For their support and encouragement in the pursuit of learning, I'd like to thank my parents, Chris and Bev Cahill. For their unwavering enthusiasm upon my every return to the U.S. and for their constant love of my wife and I, I'd like to thank Bella and Merlin. For logistical and emotional support for both my wife and I, I'd like to thank my mother-in-law, Joan Embt.

Finally, and most importantly, for her many years of taking care of all-that-is-not-thesis, for her loyalty and support in the face of my wild ideas (e.g., "Wouldn't it be cool to get a doctorate from Oxford?"), for her ability to keep me grounded and continually remind me of the important things in life, and for her love, I'd like to thank and dedicate this thesis to my wife, Kathy Embt.

Abstract

Nonrigid image registration has received much attention in the medical imaging and computer vision research communities, because it enables a wide variety of applications. Feature tracking, segmentation, classification, temporal image differencing, tumour growth estimation, and pharmacokinetic modeling are examples of the many tasks that are enhanced by the use of aligned imagery. Over the years, the medical imaging and computer vision communties have developed and refined image registration techniques in parallel, often based on similar assumptions or underlying paradigms.

This thesis focuses on variational registration, which comprises a subset of nonrigid image registration. It is divided into chapters that are based on fundamental aspects of the variational registration problem: image dissimilarity measures, changing overlap regions, regularizers, and computational solution strategies. Key contributions include the development of local versions of standard dissimilarity measures, the handling of changing overlap regions in a manner that is insensitive to the amount of non-interesting background information, the combination of two standard taxonomies of regularizers, and the generalization of solution techniques based on Fourier methods and the Demons algorithm for use with many regularizers. To illustrate and validate the various contributions, two sets of example imagery are used: 3D CT, MR, and PET images of the brain as well as 3D CT images of lung cancer patients.

Contents

List of Figures

List of Tables

List of Abbreviations

Imaging Modalities

CT	computed tomography
MR	magnetic resonance
PET	positron emission tomography
US	ultrasound

Image Dissimilarity

BC	brightness constancy
BGC	brightness gradient constancy
BGMC	brightness gradient magnitude constancy
BHC	brightness Hessian constancy
BHDC	brightness Hessian determinant constancy
BLC	brightness Laplacian constancy
BMC	brightness morphology constancy
CCRE	cross cumulative residual entropy
CR	correlation ratio
CRE	cumulative residual entropy
CRECC	cumulative residual entropy correlation coefficient
ECC	entropy correlation coefficient
JCRE	joint cumulative residual entropy
M-*	modified version of * dissimilarity measure
MAD	mean of absolute differences
MI	mutual information
MSD	mean of squared differences
MSODL	mean squared orthogonal difference from linearity

MSVDL	mean squared vertical difference from linearity
NCCRE	normalized cross cumulative residual entropy
NMI	normalized mutual information
ODL	orthogonal difference from linearity
SAD	sum of absolute differences
SCC	squared correlation coefficient
SCCRE	symmetric cross cumulative residual entropy
SOCC	squared orthogonal correlation coefficient
SSD	sum of squared differences
VDL	vertical difference from linearity

Computational Solutions

AOS	additive operator splitting
DCT	discrete cosine transform
DFT	discrete Fourier transform
DST	discrete sine transform
FAS	full approximation scheme
FFT	fast Fourier transform
FMG	full multigrid
PDE	partial differential equation
SOR	successive overrelaxation

Registration Experiments

ATRE	aggregate target registration error
ECS	effective convolution steps
RIRE	Retrospective Image Registration Evaluation
TRE	target registration error
VOI	volume of interest

List of Symbols

Vectors, Matrices, and Images

$\mathbb{R}^n$	space of real-valued n-vectors
$\mathbb{R}^{m \times n}$	space of real-valued $(m \times n)$ matrices
$\mathbf{x}$ (lower case, bold)	vector in $\mathbb{R}^n$
x_j	j^{th} component of $\mathbf{x}$
$\mathbf{X}$ (upper case, bold)	matrix in $\mathbb{R}^{m \times n}$
$X_{j,k}$	(j, k) component of $\mathbf{X}$
$\mathbf{I}$	identity matrix
$\|\mathbf{a}\|$ or $\|\mathbf{A}\|$	vector or matrix norm
$\|\mathbf{A}\|$	matrix determinant
$\mathrm{tr}\,(\mathbf{A})$	matrix trace
$\langle \mathbf{A}, \mathbf{B} \rangle$	inner product $\mathbf{A}^{\mathrm{T}}\mathbf{B}$ (for $\mathbf{A}$, $\mathbf{B}$ matrices *or* vectors)
R or $R(\mathbf{x})$	scalar-valued reference image (evaluated at $\mathbf{x}$)
F or $F(\mathbf{x})$	scalar-valued floating image (evaluated at $\mathbf{x}$)
A, B	scalar-valued arbitrary images

Regions

Ω	subset of $\mathbb{R}^n$
$\partial\Omega$	boundary of Ω
Ω_A	region over which image A is defined
$\|\Omega\|$	area/volume/content of Ω
Θ	signed function, positive inside Ω, negative outside Ω
H	Heaviside function
H_ϵ	regularized Heaviside function

Deformations

$\mathbf{\Phi}$ or $\mathbf{\Phi}(\mathbf{x})$	vector-valued deformation field (evaluated at $\mathbf{x}$)
$\mathbf{u}$ or $\mathbf{u}(\mathbf{x})$	vector-valued displacement field (evaluated at $\mathbf{x}$)
$\mathbf{v}$ or $\mathbf{v}(\mathbf{x})$	vector-valued velocity field (evaluated at $\mathbf{x}$)
$\vec{\mathbf{n}}$	vector normal to boundary/surface
$F^{\mathbf{u}}$	floating image warped according to displacement $\mathbf{u}$

Functionals

$\mathcal{J}(R, F^{\mathbf{u}})$	dissimilarity measure
$\mathcal{R}(\mathbf{u})$	regularizer
$\mathcal{S}(\mathbf{u})$ or $\mathcal{S}(R, F, \mathbf{u})$	functional of $\mathbf{u}$ (possibly dependent on R, F)
$d\mathcal{S}(\mathbf{u}; \mathbf{w})$ or $d\mathcal{S}(R, F, \mathbf{u}; \mathbf{w})$	Gâteaux derivative of $\mathcal{S}(\mathbf{u})$ in the direction $\mathbf{w}$
$\mathcal{A}$	partial differential operator arising from $d\mathcal{R}(\mathbf{u}; \mathbf{w})$
$\mathcal{B}$	boundary conditions arising from $d\mathcal{R}(\mathbf{u}; \mathbf{w})$
$\hat{\mathcal{A}}$	$\mathcal{A}$ extended to include boundary conditions
$\mathcal{P}$	force vector arising from $d\mathcal{J}(R, F^{\mathbf{u}}; \mathbf{w})$

Differential operators

$\partial_{x_j} f(\mathbf{x})$	partial derivative of f with respect to j^{th} component of $\mathbf{x}$
$\nabla f(\mathbf{x})$	gradient vector of f with respect to $\mathbf{x}$
$\nabla \mathbf{f}(\mathbf{x})$	Jacobian matrix of $\mathbf{f}$ with respect to $\mathbf{x}$
$\nabla \nabla^{\mathrm{T}} f(\mathbf{x})$	Hessian of f with respect to $\mathbf{x}$
$\nabla \mathbf{f}(\mathbf{x})$	Jacobian matrix of $\mathbf{f}$ with respect to $\mathbf{x}$
$\nabla \cdot \mathbf{f}(\mathbf{x})$ or $\operatorname{div} \mathbf{f}(\mathbf{x})$	divergence of $\mathbf{f}$ with respect to $\mathbf{x}$
$\operatorname{rot} \mathbf{f}(\mathbf{x})$	rotational component of $\mathbf{f}$ with respect to $\mathbf{x}$
$\operatorname{curl} \mathbf{f}(\mathbf{x})$	curl of $\mathbf{f}$ with respect to $\mathbf{x}$
$\Delta f(\mathbf{x})$ or $\Delta \mathbf{f}(\mathbf{x})$	Laplacian of f (or $\mathbf{f}$) with respect to $\mathbf{x}$

Distributions (Global)

$p_A(a)$	probability density function of A at a
$p_{A,B}(a, b)$	joint p.d.f. of A and B at (a, b)
$P_A(a)$	cumulative distribution function of A at a

$P_{A,B}(a,b)$	joint c.d.f. of A and B at (a,b)
$E[A]$, $\bar{A}$, or μ_A	expected value of A
$E[B\|A]$ or $\mu_{B\|A}$	conditional expectation of B given A
$\mathrm{Var}(A)$ or σ_A^2	variance of A
$\mathrm{Var}(B\|A)$ or $\sigma_{B\|A}^2$	conditional variance of B given A
$\mathrm{Cov}(A,B)$ or $\sigma_{A,B}$	covariance of A and B
$H(A)$	differential entropy of A
$H(A,B)$	differential joint entropy of A and B
$\varepsilon(A)$	cumulative residual entropy (CRE) of A
$\varepsilon(A,B)$	joint cumulative residual entropy (JCRE) of A and B

Distributions (Local)

ω_γ	weighting function of width γ
$p_A^\gamma(a;\mathbf{x})$	local p.d.f., centered at $\mathbf{x}$, evaluated at a
$p_{A,B}^\gamma(a,b;\mathbf{x})$	local joint p.d.f., centered at $\mathbf{x}$, evaluated at (a,b)
$P_A^\gamma(a;\mathbf{x})$	local c.d.f., centered at $\mathbf{x}$, evaluated at a
$P_{A,B}^\gamma(a,b;\mathbf{x})$	local joint c.d.f., centered at $\mathbf{x}$, evaluated at (a,b)
$E_\gamma[A;\mathbf{x}]$	local expected value, centered at $\mathbf{x}$
$E_\gamma[B\|A;\mathbf{x}]$	local conditional expectation, centered at $\mathbf{x}$
$\mathrm{Var}_\gamma(A;\mathbf{x})$ or σ_A^2	local variance, centered at $\mathbf{x}$
$\mathrm{Var}_\gamma(B\|A;\mathbf{x})$ or $\sigma_{B\|A}^2$	local conditional variance, centered at $\mathbf{x}$
$\mathrm{Cov}_\gamma(A,B;\mathbf{x})$ or $\sigma_{A,B}$	local covariance, centered at $\mathbf{x}$
$H_\gamma(A;\mathbf{x})$	local differential entropy, centered at $\mathbf{x}$
$H_\gamma(A,B;\mathbf{x})$	local differential joint entropy, centered at $\mathbf{x}$
$\varepsilon_\gamma(A;\mathbf{x})$	local CRE, centered at $\mathbf{x}$
$\varepsilon_\gamma(A,B;\mathbf{x})$	local JCRE, centered at $\mathbf{x}$

Distributions (Regions)

$p_A^\Omega(a)$	p.d.f. in Ω at a
$p_{A,B}^\Omega(a,b)$	joint p.d.f. in Ω at (a,b)
$P_A^\Omega(a)$	c.d.f in Ω at a
$P_{A,B}^\Omega(a,b)$	joint c.d.f. in Ω at (a,b)

$p_A^{\Omega,\gamma}(a;\mathbf{x})$	local p.d.f. in Ω, centered at $\mathbf{x}$
$p_{A,B}^{\Omega,\gamma}(a,b;\mathbf{x})$	local joint p.d.f. in Ω, centered at $\mathbf{x}$
$P_A^{\Omega,\gamma}(a;\mathbf{x})$	local c.d.f in Ω, centered at $\mathbf{x}$
$P_{A,B}^{\Omega,\gamma}(a,b;\mathbf{x})$	local joint c.d.f. in Ω, centered at $\mathbf{x}$
$E_\Omega[A]$	expected value in Ω
$E_\Omega[B\|A]$	conditional expectation in Ω
$\mathrm{Var}_\Omega(A)$	variance in Ω
$\mathrm{Var}_\Omega(B\|A)$	conditional variance in Ω
$\mathrm{Cov}_\Omega(A,B)$	covariance in Ω
$\tilde{P}_A(a)$	c.d.f. of A in background region
$\tilde{P}_{A,B}(a,b)$	joint c.d.f. of A and B in background region
$\tilde{\mu}_A$	expected value of A in background region
$\tilde{\sigma}_A^2$	variance of A in background region
$\tilde{\sigma}_{A,B}$	covariance of A and B in background region

PDE Solutions

$\mathbf{u}^{(k)}(\mathbf{x})$	k^{th} estimate of vector field $\mathbf{u}(\mathbf{x})$
$\mathbf{u}(\mathbf{x},t)$	vector field defined at position $\mathbf{x}$ and time t
$F_\mathbf{m}$ or $R_\mathbf{m}$	images F or F evaluated at lattice position $\mathbf{m}$
$\mathbf{u}_\mathbf{m}$	vector field $\mathbf{u}$ evaluated at lattice position $\mathbf{m}$
$\mathbf{e}_j$	j^{th} column of identity matrix
$\check{\nabla}$	discrete gradient
$\check{\nabla}\cdot$ or $\check{\mathrm{div}}$	discrete divergence
$\check{\Delta}$	discrete Laplacian
τ	time step
h_j	lattice spacing in j^{th} dimension
Θ^ℓ, ω_ℓ	eigenfunctions and eigenvalues of discrete operator
$\hat{\mathbf{u}}_\ell$	coefficient of eigenfunction expansion of $\mathbf{u}$
$\circ$	function/operator composition

Introduction

Successfully aligning images from a single modality or from multiple modalities enables subsequent processing, analysis, and/or visualization that can aid in the screening, diagnosis, prognosis, treatment, and monitoring of disease. In the past two decades, a vast amount of research has been performed to develop various models and computational techniques for image registration across a wide spectrum of applications. This thesis explores the approaches that have been developed for nonrigid registration using variational techniques. Furthermore, it attempts to provide a unifying framework that will allow the experienced reader to be able to determine how best to construct and solve the variational registration problem when faced with a new application. It provides contributions in four main chapters, each based on a fundamental aspect of the registration problem: image dissimilarity measures, changing overlap regions, regularization techniques, and efficient numerical solutions.

Thesis Layout

This thesis contains six chapters: a first chapter presenting an overview of the nonrigid registration problem and the mathematical preliminaries necessary to understand the variational registration problem; four main chapters detailing image dissimilarity measures, changing overlap regions, regularization techniques, and efficient numerical solutions; and a final chapter describing conclusions and future work. This section provides summary descriptions of each chapter.

Chapter 1 describes the context of variational registration within the nonrigid registration problem in medical imaging. It explores how the variational image registration

problem can be described in terms of underlying components including the dissimilarity measure, overlap region, regularizer, and solution technique. Next, it provides the mathematical preliminaries necessary to understand the remaining chapters in this thesis. Finally, it mentions some of the limitations and challenges that the variational approach is likely to encounter in practice.

Chapter 2 describes a wide variety of image dissimilarity measures that have been proposed and used in practice. It classifies the dissimilarity measures according to the underlying assumption about the relationship between the images being registered, and it proposes three new dissimilarity measures for inclusion in this classification. Next, the chapter builds on Hermosillo's idea [50] of constructing local versions of some specific dissimilarity measures, illustrating how to construct local versions of *all* of the dissimilarity measures. Finally, this chapter presents Gâteaux derivatives of all of the dissimilarity measures (both the global and local versions), which are required for variational registration.

Chapter 3 discusses problems that emerge over the course of registration due to the changing nature of the *overlap region*, or the region of intersection of valid data in both images. One problem is that Gâteaux derivatives are typically derived under the simplifying assumption that the overlap region remains constant. To address this problem, the chapter proposes a technique in the spirit of Chan and Vese's active contours [25], by representing the regions of interest in both images by Heaviside functions that deform over the course of registration. A second problem is that many dissimilarity measures exhibit a non-intuitive sensitivity when background (non-interesting information) is the only content moving in and out of the changing overlap region. This problem is addressed by using image statistics in the background regions to modify the dissimilarity measures. The proposed solutions to both problems change the form of each dissimilarity measure, and therefore this chapter shows how the Gâteaux derivatives presented in the previous chapter must be modified for use in variational registration.

Chapter 4 focuses on regularizers that must be added to the dissimilarity measure in order to ensure the existence of a unique solution to the variational registration problem. It describes two taxonomies of regularizers, the *quadratic* taxonomy [63] of homogeneous regularizers (including diffusion, elastic, curvature, and fluid) that is commonly used in medical image registration, and the *Weickert* taxonomy [95] of nonhomogeneous regularizers (including isotropic, anisotropic, image-driven, and flow-driven) that is used in optical flow. By pointing out that both taxonomies start with the diffusion (or Horn-Schunck [51]) regularizer as their base, this chapter shows how these taxonomies can be combined in order to construct nonhomogeneous versions of any of the quadratic regularizers. Gâteaux derivatives of every regularizer in the proposed combined taxonomy are presented, enabling construction of the PDE systems arising in variational registration.

Chapter 5 illustrates strategies for solving the PDE systems arising from the Gâteaux derivatives of the dissimilarity measures and regularizers. It shows how semi-implicit discretizations in conjunction with fixed-point or steepest descent algorithms are used to transform any PDE system into a sequence of linear algebraic equations that can be solved in succession. After reviewing some common iterative techniques that can be applied for any regularizer, the chapter then focuses on two rapid strategies that can be exploited for homogeneous regularizers; namely, Fourier methods and successive Gaussian convolution. For both cases, prior research [29, 46, 40, 63, 87] has shown that these strategies are useful for a few specific homogeneous regularizers. This chapter shows how they can be extended for use with *any* homogeneous regularizer.

Chapter 6 summarizes the key contributions of each of the previous chapters. It also presents some ideas for future work, including learning optimal dissimilarity measures for particular registration applications, exploring the possibility of rapid solution techniques for nonhomogeneous regularizers, and ensuring that deformation fields computed using arbitrary regularizers are diffeomorphic.

In this thesis, two different clinical problems are used to illustrate and compare various aspects of the registration problem. In the problem examined in Chapter 3, CT, PET, and MR brain images from the Retrospective Image Registration Evaluation project [97] are used to compare a number of dissimilarity measures with each other and with their modified versions. In the problem examined in Chapter 5, serial sets of CT chest images of patients with lung nodules are used to compare registration accuracy and speed when various regularizers and computational algorithms are compared.

Novel Contributions

This thesis presents novel contributions in each of the four main chapters. These contributions are summarized here, according to chapter.

Chapter 2: Dissimilarity Measures

- Presents new dissimilarity measures, including squared orthogonal correlation coefficient (SOCC), normalized cross cumulative residual entropy (NCCRE), and cumulative residual entropy correlation coefficient (CRECC).
- Generalizes Hermosillo's technique [50] to derive local versions of *all* dissimilarity measures.
- Presents Gâteaux derivatives for global and local versions of *all* dissimilarity measures.

Chapter 3: Changing Overlap

- Proposes technique for incorporating varying overlap regions into dissimilarity measures and Gâteaux derivatives.
- Develops approach for modifying dissimilarity measures based on image background statistics in order to avoid the overlap sensitivity problem.
- Illustrates how to compute modified Gâteaux derivatives for incorporation into variational registration.

Chapter 4: Regularizers

- Presents second-order elastic regularizer, extending the quadratic taxonomy [63] of regularizers.
- Generalizes Weickert taxonomy [95] of regularizers to enable weighting functions/matrices to vary according to the underlying deformation.
- Combines the ideas behind the quadratic and Weickert taxonomies to present nonhomogeneous versions of the elastic, curvature, and second-order elastic regularizers.
- Presents Gâteaux derivatives for all regularizers in the combined taxonomy.

Chapter 5: Solving the Variational Registration Problem

- Presents a framework for approximating the variational registration problem with *any* of the proposed regularizers (linear or nonlinear) as a succession of linear systems of algebraic equations.
- Extends the idea [40] of rapidly solving the curvature registration problem with the discrete cosine transform to solving the variational registration problem with *any* homogeneous regularizer with Dirichlet, Neumann, or periodic boundary conditions in a fixed-point or steepest descent style iteration using Fourier methods.
- Extends the idea of successive Gaussian filtering presented by Thirion [87] to solving the variational registration problem with *any* homogeneous regularizer.

Related Publications

Various portions of this thesis, as well as preliminary and/or closely related work have been published in the following:

1. Cahill, N.D., J.A. Noble, D.J. Hawkes and L.A. Ray. Fast Fluid Registration with Dirichlet Boundary Conditions. *Proc. International Symposium on Biomedical Imaging*, pp. 712-715, April 2007.

2. Cahill, N.D., J.A. Noble and D.J. Hawkes. Fourier Methods for Nonparametric Image Registration. *Proc. CVPR Workshop on Image Registration and Fusion*, pp. 1-8, June 2007.

3. Cahill, N.D., G. Vesom, L. Gorelick, J. Brady, J.A. Noble and J.M. Brady. Investigating Implicit Shape Representations for Alignment of Livers from Serial CT Examinations. *Proc. International Symposium on Biomedical Imaging*, pp. 776-779, May 2008.

4. Cahill, N.D., J.A. Schnabel, J.A. Noble and D.J. Hawkes. Revisiting Overlap Invariance in Medical Image Alignment. *Proc. Mathematical Methods in Biomedical Image Analysis*, pp. 1-8, June 2008.

5. Cahill, N.D., J.A. Schnabel, J.A. Noble and D.J. Hawkes. Overlap Invariance of Cumulative Residual Entropy Measures for Multimodal Image Alignment. *Proc. SPIE Medical Imaging: Image Processing*, February 2009.

6. Lee, D., M. Hofmann, F. Steinke, Y. Altun, N.D. Cahill and B. Schölkopf. Learning the Similarity Measure for Multi-modal Image Registration. *Proc. Computer Vision and Pattern Recognition*, June 2009.

7. Cahill, N.D., J.A. Noble and D.J. Hawkes. Demons Algorithms for Fluid and Curvature Registration. *Proc. International Symposium on Biomedical Imaging*, June 2009.

8. Cahill, N.D., J.A. Noble and D.J. Hawkes. A Demons Algorithm for Image Registration with Locally Adaptive Regularization. *Proc. Medical Image Computing and Computer Assisted Intervention*, September 2009.

9. Hofmann, M., B. Schölkopf, I. Bezrukov and N.D. Cahill. Incorporating Prior Knowledge on Class Probabilities into Intermodality Image Registration. *Proc. MICCAI Workshop on Probabilistic Methods in Medical Image Analysis*, September 2009.

Chapter 1

Preliminaries

This chapter provides the mathematical preliminaries necessary for posing and solving the nonparametric image registration problem. It first describes how to pose the nonparametric image registration problem as the minimization of a functional comprising a dissimilarity measure and a regularizer. Then, it describes how to use calculus of variations to transform the registration problem into one of solving a set of partial differential equations known as the Euler-Lagrange equations. Next, it mentions some of the inherent limitations of variational methods for the task of image registration, and it suggests strategies for overcoming those limitations in practice. Finally, it describes how the contributions in the remaining chapters of this thesis address components of the variational image registration problem.

1.1 Posing the Nonparametric Image Registration Problem

Consider two images, a reference image R and a floating image F, both as functions in $\mathbb{R}^n$. Define a *deformation* $\mathbf{\Phi} : \mathbb{R}^n \mapsto \mathbb{R}^n$ by:

$$\mathbf{\Phi}(\mathbf{x}) = \mathbf{x} - \mathbf{u}(\mathbf{x}), \tag{1.1}$$

and call $\mathbf{u}$ the *displacement*. The objective of registration is to find a displacement that minimizes a dissimilarity measure $\mathcal{I}$ that measures the dissimilarity between the reference image R and the deformed floating image $F^{\mathbf{u}} := F(\mathbf{\Phi})$:

$$\min_{\mathbf{u}} \quad \mathcal{I}(R, F^{\mathbf{u}}). \tag{1.2}$$

In order to ensure that the minimization problem is well-posed, a regularizing term $\mathcal{R}$ must be added to the dissimilarity measure. The regularizer enforces smoothness in the

deformation, and in simple cases is typically chosen to be a quadratic form:

$$\mathcal{R}(\mathbf{u}) = \frac{1}{2}\, a\langle \mathbf{u}, \mathbf{u} \rangle, \tag{1.3}$$

which is based on the symmetric positive semidefinite bilinear form $a\langle \mathbf{u}, \mathbf{w} \rangle$.

Combining the regularizer with the dissimilarity measure enables us to state the general form of the registration problem:

$$\min_{\mathbf{u}} \qquad \mathcal{S}(R, F, \mathbf{u}) := \mathcal{R}(\mathbf{u}) + \alpha \mathcal{J}(R, F^{\mathbf{u}})\,, \tag{1.4}$$

where α is a positive parameter that allows a tradeoff between the influence of the regularizer and dissimilarity measure.

Practically speaking, analytical solutions to the minimization problem (1.4) are difficult or impossible to identify. The only recourse is to approximate the solution on a discrete grid. Discretization can be performed prior to the optimization step (the *discretize-optimize* approach), or after the optimization step (the *optimize-discretize* approach). The discretize-optimize approach yields a large-scale optimization problem that can have a huge number of parameters, especially for 3-D volumetric registration. Therefore, we focus in this thesis on the optimize-discretize approach; the *optimize* step involves identifying the necessary conditions for a minimum (which can be done using variational methods), yielding a system of partial differential equations, and the *discretize* step involves discretizing the resulting PDE system in a manner that can be efficiently solved.

1.2 Variational Minimization

We begin with a quick review of a concept from calculus: suppose we want to minimize a function $f(\mathbf{x})$ that we assume is twice differentiable. Intuitively, it must be true that any local minimum of f necessarily has a vanishing derivative in every direction. Now, the derivative of f in the direction $\mathbf{s}$ is given by $\mathbf{s}^{\mathrm{T}} \nabla f$, so in order for this quantity to vanish in every direction s, it must be true that $\nabla f = 0$. This condition, that $\nabla f = 0$, is then a necessary condition for a minimum.

In order to solve a functional minimization problem such as (1.4), we can use a similar type of reasoning. It must be true that any local minimum of $\mathcal{S}(\mathbf{u})$ necessarily has a

vanishing derivative in every direction. The derivative of $\mathcal{S}(\mathbf{u})$ in the direction of $\mathbf{w}$ is given by the Gâteaux derivative $d\mathcal{S}(\mathbf{u};\mathbf{w})$. As will be shown in Chapters 2, 3, and 4, the Gâteaux derivatives of any general dissimilarity measure $\mathcal{I}(R, F^{\mathbf{u}})$ and any (first-order) regularizer $\mathcal{R}(\mathbf{u})$ are given by:

$$d\mathcal{I}(R, F^{\mathbf{u}}; \mathbf{w}) = \int_{\Omega} \langle \mathcal{P}(\mathbf{x}; R, F^{\mathbf{u}}), \mathbf{w} \rangle \, d\mathbf{x} \tag{1.5}$$

and

$$d\mathcal{R}(\mathbf{u};\mathbf{w}) = \int_{\Omega} \langle \mathcal{A}(\mathbf{u}), \mathbf{w} \rangle \, d\mathbf{x} + \int_{\partial\Omega} \langle \mathcal{B}[\mathbf{u}], \mathbf{w} \rangle \, d\sigma, \tag{1.6}$$

respectively. Hence, the Gâteaux derivative of $\mathcal{S}(\mathbf{u})$ is given by:

$$d\mathcal{S}(\mathbf{u};\mathbf{w}) = \int_{\Omega} \langle \mathcal{A}(\mathbf{u}) + \alpha \, \mathcal{P}(\mathbf{x}; R, F^{\mathbf{u}}), \mathbf{w} \rangle \, d\mathbf{x} + \int_{\partial\Omega} \langle \mathcal{B}[\mathbf{u}], \mathbf{w} \rangle \, d\sigma. \tag{1.7}$$

Now, by the Fundamental Theorem of Calculus of Variations [43], we find that the condition $d\mathcal{S}(\mathbf{u};\mathbf{w}) = 0$ implies:

$$\mathcal{A}(\mathbf{u}) = -\alpha \, \mathcal{P}(\mathbf{x}; R, F^{\mathbf{u}}) \qquad \forall \mathbf{x} \in \Omega, \tag{1.8}$$

$$\mathcal{B}[\mathbf{u}] = 0 \qquad \forall \mathbf{x} \in \partial\Omega. \tag{1.9}$$

This system of partial differential equations (1.8)–(1.9) is known as the *Euler-Lagrange* equations, which provide the necessary conditions for a minimum of $\mathcal{E}(\mathbf{u})$. Note that for higher order regularizers, (1.9) is replaced by multiple boundary conditions; i.e.,

$$\mathcal{B}_i[\mathbf{u}] = 0 \qquad \forall \mathbf{x} \in \partial\Omega, \qquad i = 1, \ldots, n, \tag{1.10}$$

where n is the order of the regularizer.

The right hand side of (1.8), which is sometimes referred to as the *force vector*, is nonlinear in $\mathbf{u}$. If the regularizer chosen is linear in the partial derivatives of $\mathbf{u}$, then the left hand side of (1.8) is linear, and the Euler-Lagrange equations are referred to as being *semilinear*. If the regularizer chosen is nonlinear, then the left hand side of (1.8) is nonlinear, and so the Euler-Lagrange equations are nonlinear.

1.3 Limitations of Variational Methods

The variational approach to nonparametric image registration does have some limitations that can make practical implementation tricky or difficult. In this section we describe three

such limitations along with suggestions on how to overcome the limitations in practice.

1.3.1 Small Deformation Assumption

Identifying the Gâteaux derivative of $\mathcal{S}(\mathbf{u})$ relies on Taylor series approximations that are valid only if the displacement field $\mathbf{u}(\mathbf{x})$ is assumed to be small. This assumption limits the applicability of the Euler-Lagrange equations (1.8)–(1.9) to small deformations. In many practical situations, however, large deformations are required to align images.

One strategy that can be useful in handling larger deformations is to perform registration in a multiresolution or multilevel pyramid [59]. This involves first constructing a pyramid of images at coarser and coarser resolutions, and then repeatedly performing registration, starting at the coarsest resolution and moving back up to perform registration at higher resolutions. The key insight that makes this a useful strategy is that a large deformation on a fine grid becomes a smaller deformation on a coarser grid.

Another strategy that can be used either separately or in conjunction with multiresolution pyramids is to apply the Euler-Lagrange equations to the velocity field instead of the displacement field. This turns registration into a flow estimation problem, and it allows for larger deformations to be recovered because the velocity field can remain small as the displacement field grows. This strategy is described in detail in Section 5.1.3 of this thesis.

1.3.2 Multiple Local Minima

Another limitation that arises in variational image registration is that multiple local minima of (1.4) typically exist due to the nonconvexity of $\mathcal{I}(R, F^{\mathbf{u}})$. This problem is particularly acute when the recovery of large deformations are required, because a quadratic (and hence, convex) approximation of $\mathcal{I}(R, F^{\mathbf{u}})$ is only valid for small deformations.

Two conceptual approaches for reducing the impact of this problem are variations on the theme of ensuring that the initial deformation estimate is "close enough" to the true deformation: (i) pre-register the images according to some simpler transformation in order to ensure that the initial deformation estimate is as close as possible to the true deformation, and (ii) pre-process the images in a way that will tend to reduce the number of local minima

of (1.4), so that the basin of attraction around the true deformation is as large as possible.

A rigid or affine pre-registration step is always recommended prior to variational registration. To reduce the number of local minima of (1.4), registration can be carried out in a multiscale manner, by performing registration first on coarser scale versions of the images and then propagating (and refining) the results to finer scales. At coarse scales, much of the rapidly varying content in an image is attenuated, allowing registration to focus on aligning large coarse features. Multiscale registration can be performed either independently of or in combination with multiresolution registration as described in [59] and the references therein.

1.3.3 Non-Smooth Deformations

The other major limitation of variational techniques is that the inclusion of the regularizer $\mathcal{R}(\mathbf{u})$ in (1.4) tends to bias the result of registration towards smooth deformations. In many practical situations in medical imaging, discontinuities exist in the true deformation field. For example, the true deformation field relating two 3-D chest CT images at different points in the breathing cycle should be discontinuous where the diaphragm slides along the chest wall. In another example, in the registration of pre-surgical and post-surgical images of the same patient, a tumor or organ may be present in one image and not the other, yielding a true deformation field that should have a singularity or hole depending on the order in which images are considered. Such discontinuities are not explicitly modeled by variational techniques, although small amounts of sliding can be handled with nonhomogeneous regularizers, as is described in Sections 4.2–4.3 of this thesis.

1.4 Summary

The focus of variational image registration is to determine the Euler-Lagrange equations that arise for a specific choice of dissimilarity measure and regularizer, and then to discretize and solve the Euler-Lagrange equations on a grid of image pixels/voxels. The remaining chapters in this thesis address various components of the variational image registration problem: Chapters 2 and 3 investigate how dissimilarity measures are constructed and how

they behave over the course of registration; Chapter 4 discusses how regularizers can be constructed to exhibit various properties; and, Chapter 5 describes how to discretize and rapidly solve the Euler-Lagrange equations.

Chapter 2

Dissimilarity Measures

Any dissimilarity measure relies on some underlying assumption about the relationship between the images being registered. In optical flow, the assumption is one of intensity constancy [51]; namely, that brightness (or brightness gradient or another related feature [95]) is preserved over time. In medical image registration, constancy assumptions are useful when the floating and reference images are of the same patient, from the same modality, and from similar viewpoints; however, they can be too restrictive when images from multiple patients, multiple modalities, and/or multiple views are aligned. In order to accommodate these situations, more flexible assumptions have been made about the relationship between image data, including linearity in brightness [41, 56, 58], general nonlinearity in brightness [74, 76], and commonality of information content [34, 84, 92, 93].

This chapter details the image data relationship assumptions based on constancy, linearity, general nonlinearity, and commonality of information content, and shows how dissimilarity measures can be constructed by penalizing deviations from these assumptions. Within this framework, some new dissimilarity measures are proposed: the squared orthogonal correlation coefficient (SOCC), the normalized cross cumulative residual entropy (NCCRE), and the cumulative residual entropy correlation coefficient (CRECC). This chapter then discusses the global versus local nature of the assumptions, describing how to extend the work of Hermosillo [50] on local versions of squared correlation coefficient (SCC), correlation ratio (CR), and mutual information (MI) to construct local versions of *any* dissimilarity measure. Next, this chapter presents the Gâteaux derivatives of the global and local versions of all

of the dissimilarity measures, as these are required for variational image registration. In a final note, this chapter mentions the popular "Demons Algorithm" and how Demons force vectors relate to the Gâteaux derivatives of dissimilarity measures.

2.1 Image Data Relationship Assumptions

2.1.1 Constancy Assumptions

The simplest assumption that can be made about the relationship between a reference and floating image is that brightness is preserved; that is:

$$F^{\mathbf{u}}(\mathbf{x}) = R(\mathbf{x}) \qquad \forall x \in \Omega. \tag{2.1}$$

This is known as the *brightness constancy constraint* [51], and it is applicable when images are captured from the same sensor under the same conditions.

In many practical situations, the brightness constancy assumption is too restrictive, and it is useful to generalize to other types of assumptions. One such assumption is that the spatial gradient of brightness is preserved [90], yielding the *brightness gradient constancy constraint*:

$$\nabla F^{\mathbf{u}}(\mathbf{x}) = \nabla R(\mathbf{x}) \qquad \forall x \in \Omega. \tag{2.2}$$

Another possible assumption is that the spatial Hessian (matrix of second derivatives) of brightness is preserved, that is:

$$\nabla\nabla^{\mathrm{T}} F^{\mathbf{u}}(\mathbf{x}) = \nabla\nabla^{\mathrm{T}} R(\mathbf{x}) \qquad \forall x \in \Omega. \tag{2.3}$$

In order to overcome the implicit orientation constancy assumption in (2.2) and (2.3), Weickert [95] suggests assuming constancy of the magnitudes of the gradient/Hessian terms, yielding the *brightness gradient magnitude constancy constraint*:

$$\|\nabla F^{\mathbf{u}}(\mathbf{x})\| = \|\nabla R(\mathbf{x})\| \qquad \forall x \in \Omega, \tag{2.4}$$

the *brightness Hessian determinant constancy constraint*:

$$\left|\nabla\nabla^{\mathrm{T}} F^{\mathbf{u}}(\mathbf{x})\right| = \left|\nabla\nabla^{\mathrm{T}} R(\mathbf{x})\right| \qquad \forall x \in \Omega. \tag{2.5}$$

and the *brightness Laplacian constancy constraint*:

$$\Delta F^{\mathbf{u}}(\mathbf{x}) = \Delta R(\mathbf{x}) \qquad \forall x \in \Omega. \tag{2.6}$$

The constancy constraints (2.5) and (2.6) provide alternatives for imposing constancy on the Hessian magnitude without imposing constancy on Hessian orientation.

Finally, Droske and Rumpf [38] and Haber and Modersitzki [47] make the assumption that the brightness morphology, as measured by the normalized brightness gradient, should be constant across images, even when the images come from different modalities. This so-called *brightness morphology constancy constraint* is given by:

$$\frac{\nabla F^{\mathbf{u}}(\mathbf{x})}{\|\nabla F^{\mathbf{u}}(\mathbf{x})\|} = \frac{\nabla R(\mathbf{x})}{\|\nabla R(\mathbf{x})\|} \qquad \forall x \in \Omega. \tag{2.7}$$

As opposed to the brightness gradient magnitude constancy constraint (2.4), (2.7) can be thought of as a constancy constraint on the *direction* of the brightness gradient.

2.1.2 Functional/Linearity Assumptions

The constancy assumptions, although useful in many circumstances, can still be too restrictive for use in some practical applications. In some cases, it is more useful to assume that there is some underlying functional relationship between the image data. In unimodal applications where serial studies of the same patient are registered, constancy assumptions may be invalid due to the use of different imaging devices or slightly different imaging protocols; however, there is still an underlying functional relationship between the images. In some multimodal applications, such as magnetic resonance (MR) to ultrasound (US) registration, constancy assumptions are clearly invalid, although in some cases [76], US values can be somewhat reliably predicted from MR gradient magnitude values.

An underlying functional assumption between images can be expressed by the *brightness functional constraint*:

$$\exists f \text{ such that } F^{\mathbf{u}}(\mathbf{x}) = f(R(\mathbf{x})) \quad \forall x \in \Omega. \tag{2.8}$$

For the remainder of this thesis, we will not only be interested in the general functional constraint, but also in a restriction of it that we will call the *brightness linearity constraint*, in which $f(R)$ is restricted to be a linear function of R.

In some cases, medical practitioners may want to switch the roles of the floating and reference image for a particular registration task. In whole body MR and CT imaging, for example, different abnormalities may be apparent in each modality. For abnormalities that are visible only in the CT image but not the MR image, the practitioner may desire to treat the MR image as the floating image so as not to introduce any deformation into those abnormalities. For this same reason, the practitioner may desire to treat the CT image as the floating image for abnormalities that are visible in the MR image but not the CT image. Because of this, we require a version of (2.8) in which the brightness functional constraint in expressed the converse direction; i.e.,

$$\exists f \text{ such that } R(\mathbf{x}) = f(F^{\mathbf{u}}(\mathbf{x})) \quad \forall x \in \Omega. \tag{2.9}$$

For this thesis, however, we will use the form given by (2.8). We note that any of the subsequent dissimilarity measures that are derived under the constraint (2.8) can also be derived under the constraint (2.9).

2.1.3 Information Content Assumptions

In many multimodal registration applications, there may not exist any functional relationship between data in the floating and reference images. However, it is possible to assume a more general image relationship by analyzing the information content in both images. The *information content assumption* can be roughly stated in the following way: a combined version of two misaligned images should be more complex than the combined version of the same two images when they are aligned.

If each pixel of an image is considered as a realization of a continuous random variable, one way to describe the complexity of the image is by the differential entropy of the random variable. (We assume continuous random variables in order to simplify the gradient computations required in variational image registration.) For a continuous random variable A, the differential entropy $H(A)$ is given by:

$$H(A) = -\int_{-\infty}^{\infty} p_A(a) \log p_A(a)\, da, \tag{2.10}$$

where $p_A(a)$ is the value of the probability density function of A at a. If we define a fused image to be some function of the values of two underlying images, then each pixel of a fused image can be considered a function of a joint realization of two random variables. The complexity of the fused image can therefore be described by the joint entropy of the two underlying random variables. The differential joint entropy of continuous random variables A and B is given by:

$$H(A,B) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{A,B}(a,b)\log p_{A,B}(a,b)\,dadb, \tag{2.11}$$

where $p_{A,B}(a,b)$ is the value of the joint probability density function of (A,B) at (a,b).

Another way in which the complexity of an image can be described is by the *cumulative residual entropy* (CRE) of the underlying random variable. CRE was introduced in [71, 94] and generalized in [37]. The general form is given by:

$$\varepsilon(A) = -\int_{-\infty}^{\infty} P(A>a)\,\log P(A>a)\,da. \tag{2.12}$$

Rao *et al.* [71] argue that CRE retains many of the important properties of Shannon entropy that differential entropy does not. The analogy of differential joint entropy is the *joint cumulative residual entropy* (JCRE), given by:

$$\varepsilon(A,B) = \varepsilon(A) + \mathrm{E}[\varepsilon(B|A)]\,, \tag{2.13}$$

where

$$\varepsilon(B|A) = -\int_{-\infty}^{\infty} P(B>b\,|\,A)\,\log P(B>b\,|\,A)\,db, \tag{2.14}$$

and where E[•] denotes expected value. Note that unlike differential joint entropy, JCRE is not symmetric; i.e., $\varepsilon(A,B) \neq \varepsilon(B,A)$. A symmetrized version can easily be constructed, however; we define the *symmetric joint cumulative residual entropy* (SJCRE) as:

$$\begin{aligned}\varepsilon_S(A,B) &= \frac{1}{2}\Big(\varepsilon(A,B) + \varepsilon(B,A)\Big) \\ &= \frac{1}{2}\Big(\varepsilon(A) + \varepsilon(B) + \mathrm{E}[\varepsilon(A|B)] + \mathrm{E}[\varepsilon(B|A)]\Big).\end{aligned} \tag{2.15}$$

Rather than explicitly state the information content assumption as a constraint in the form of (2.1)–(2.8), we will consider only the qualitative definition stated at the beginning

of this section; namely, that the combined version of two images should be simplest when the images are in alignment. Later, in Section 2.2.4, we will show how various dissimilarity measures based on (2.10)–(2.15) can be thought of as penalizing deviations from the information content assumption.

2.2 Dissimilarity Measures as Penalties

When two aligned images are assumed to obey one of the constancy constraints, any misalignment should cause a deviation from the assumed constraint. Image dissimilarity measures can be constructed by penalizing this deviation. For the constancy assumptions, this can typically be done by defining a penalty function on the difference between the image intensities/gradients/Hessians. For the functional assumptions, an additional step is required: assuming some functional form and applying it to one of the images. Finally, for the information content assumptions, dissimilarity measures can be constructed by relating the entropies of each image and the joint entropy of both images in a way that gives an optimal value when the images are aligned.

2.2.1 Penalties on Deviation from Constancy Assumptions

We begin by defining an image of residuals r that quantifies the deviation of two images from the assumed constancy constraint. For the brightness constancy constraint, this image is given by:

$$r_{BC}(\mathbf{x}; R, F^{\mathbf{u}}) = F^{\mathbf{u}}(\mathbf{x}) - R(\mathbf{x})\,. \tag{2.16}$$

Residual images for the other constancy constraints are given in Table 2.1.

All residual images in Table 2.1 except for the brightness gradient magnitude and brightness morphology are based on the difference between the left hand side and right hand side of the corresponding constancy constraint. For the brightness gradient magnitude, the square is used so that the residual image is differentiable when $\nabla F^{\mathbf{u}}(\mathbf{x}) = 0$. For the brightness morphology constancy constraint, the vector $\mathbf{n}$ is given by:

$$\mathbf{n}(I(\mathbf{x})) = \begin{cases} \frac{\nabla I(\mathbf{x})}{\|\nabla I(\mathbf{x})\|}, & \nabla I(\mathbf{x}) \neq 0 \\ 0, & \nabla I(\mathbf{x}) = 0 \end{cases}, \tag{2.17}$$

Constancy Constraint	**Symbol**	**Residual Image Value** $r(\mathbf{x})$
Brightness	r_{BC}	$F^{\mathbf{u}}(\mathbf{x}) - R(\mathbf{x})$
Brightness Gradient	r_{BGC}	$\nabla F^{\mathbf{u}}(\mathbf{x}) - \nabla R(\mathbf{x})$
Brightness Hessian	r_{BHC}	$\nabla\nabla^{\mathrm{T}} F^{\mathbf{u}}(\mathbf{x}) - \nabla\nabla^{\mathrm{T}} R(\mathbf{x})$
Brightness Gradient Magnitude	r_{BGMC}	$\|\nabla F^{\mathbf{u}}(\mathbf{x})\|^2 - \|\nabla R(\mathbf{x})\|^2$
Brightness Hessian Determinant	r_{BHDC}	$\det\left(\nabla\nabla^{\mathrm{T}} F^{\mathbf{u}}(\mathbf{x})\right) - \det\left(\nabla\nabla^{\mathrm{T}} R(\mathbf{x})\right)$
Brightness Laplacian	r_{BLC}	$\Delta F^{\mathbf{u}}(\mathbf{x}) - \Delta R(\mathbf{x})$
Brightness Morphology	r_{BMC}	$\|\mathbf{n}(F^{\mathbf{u}}(\mathbf{x})) \times \mathbf{n}(R(\mathbf{x}))\|^2$

Table 2.1: Residual images for constancy constraints.

where the image I can represent either the reference image R or the warped floating image $F^{\mathbf{u}}$. In this case, the residual image r_{BMC} is equivalent to the square of the sine of the angular difference between the left and right hand sides of (2.7) when neither image gradient is zero. In [47], the authors note that this definition of $\mathbf{n}$ suffers from two drawbacks: it is not differentiable at the boundaries of constant regions and it is very sensitive to small gradient values. They propose a regularized version based on replacing $\mathbf{n}$ with $\mathbf{n}_\epsilon$:

$$\mathbf{n}_\epsilon(I(\mathbf{x})) = \frac{\nabla I(\mathbf{x})}{\sqrt{\|\nabla I(\mathbf{x})\|^2 + \epsilon^2}}, \tag{2.18}$$

where ϵ is a nonzero constant with small magnitude that can be tuned according to the observed noise level.

Global image dissimilarity measures can be constructed by applying a nonnegative function Ψ to the residual image r and computing its average over the region Ω:

$$\mathcal{J}^g_{\Psi,r}(R, F^{\mathbf{u}}) := \frac{1}{|\Omega|} \int_\Omega \Psi(r(\mathbf{x}; R, F^{\mathbf{u}}))\, d\mathbf{x} \tag{2.19}$$

The "g" superscript indicates that (2.19) is a global dissimilarity measure. A variety of choices can be made for Ψ. Among the most common are the functions:

$$\Psi_p(r) := |r|^p, \qquad p \in \mathbb{Z}^+, \tag{2.20}$$

which turn (2.19) into measures of the p^{th} power of the L^p-norm of the scalar valued residual image r. For vector and matrix valued residual images $\mathbf{r}$, (2.20) can be written in terms of the corresponding vector or matrix norm:

$$\Psi_p(\mathbf{r}) := \|\mathbf{r}\|_p^p, \qquad p \in \mathbb{Z}^+. \tag{2.21}$$

We will consider matrix norms to be entry-wise norms as opposed to induced norms. In Section 2.4, this will simplify the process of computing the derivatives of penalty functions for nonrigid registration.

Some well known dissimilarity measures are based on specific choices of p and r in (2.20). The choice $\Psi_2(r_{BC})$ yields the well known *mean of squared differences* (MSD) dissimilarity measure, stated explicitly as:

$$MSD(R, F^{\mathbf{u}}) := \frac{1}{|\Omega|} \int_\Omega \left(F^{\mathbf{u}}(\mathbf{x}) - R(\mathbf{x})\right)^2 d\mathbf{x}, \tag{2.22}$$

where $|\Omega|$ is the size of the region Ω over which the images are defined. The MSD and its unnormalized version (the *sum of squared differences*, or SSD) have been used, for instance, in the registration of brain images [7, 48].

The choice $\Psi_1(r_{BC})$ yields the well known *mean of absolute differences* (MAD) dissimilarity measure:

$$MAD(R, F^{\mathbf{u}}) := \frac{1}{|\Omega|} \int_\Omega \left|F^{\mathbf{u}}(\mathbf{x}) - R(\mathbf{x})\right| d\mathbf{x}. \tag{2.23}$$

The MAD and its unnormalized version (the *sum of absolute differences*, or SAD) have been used in template matching [9, 23] and optic flow [69].

Other possibilities for Ψ have been used in order to combat some of the problems inherent with L^p minimization. For example, Chan *et al.* [24] and Brox *et al.* [15] suggest a modified version of Ψ_1:

$$\Psi_{1,\epsilon}(r) := \sqrt{r^2 + \epsilon^2}, \tag{2.24}$$

which, for small ϵ behaves like Ψ_1 but is differentiable at the origin. This modified version of Ψ_1 can be generalized to operate on vector or matrix $\mathbf{r}$ in the following way:

$$\Psi_{1,\epsilon}(\mathbf{r}) := \sqrt{\|\mathbf{r}\|_F^2 + \epsilon^2}, \tag{2.25}$$

where $\|\bullet\|_F^2$ indicates the square of the Frobenius norm.

For robustness against outliers in the residual images, M-estimators can be used to construct dissimilarity measures. Arya *et al.* [5] show that the Huber M-estimator [52] and the Tukey Bisquare M-estimator [89] exhibit good performance for registering pictorial images with large occlusions. The Huber M-estimator is given by:

$$\Psi_{H,k}(r) := \begin{cases} \frac{r^2}{2}, & |r| < k \\ k\left(|r| - \frac{k}{2}\right), & |r| \geq k \end{cases}, \tag{2.26}$$

where k is a tuning parameter that should be chosen to be 1.345σ for 95% asymptotic efficiency on the standard normal distribution, where σ is the estimated standard deviation of the residuals. The Tukey Bisquare M-estimator is given by:

$$\Psi_{T,c}(r) := \begin{cases} \frac{c^2}{6}\left(1 - \left[1 - \left(\frac{r}{c}\right)^2\right]^3\right), & |r| < c \\ \frac{c^2}{6}, & |r| \geq c \end{cases}, \tag{2.27}$$

where c is a tuning parameter that should be chosen to be 4.6851σ for 95% asymptotic efficiency on the standard normal distribution. Other M-estimators that have been used in registration and matching include the Geman-McClure [44] and the Lorentizian M-estimators. The Geman-McClure M-estimator is given by:

$$\Psi_{G,\sigma}(r) := \frac{r^2}{r^2 + \sigma^2}, \tag{2.28}$$

and has been used in registration of 3D MR and ultrasound images in [76]. The Lorentizian M-estimator is given by:

$$\Psi_{L,\sigma}(r) := \log\left(1 + \frac{r^2}{2\sigma^2}\right), \tag{2.29}$$

and has been used in comparison with other M-estimators for template matching in [27].

Table 2.2 lists the penalty functions (2.20) and (2.24)–(2.29) for easy reference. Examples of these penalty functions are illustrated in Fig. 2.1. The upper left figure shows the L^2 penalty, which squares errors and can therefore be problematic in the presence of outliers. The upper right figure shows the L^1 penalty and its modified version that is differentiable at zero; as can be seen, these functions still penalize outliers, but not to the same degree as the L^2 penalty. The middle right, middle left, lower right, and lower left figures illustrate the Huber, Tukey, Geman-McClure, and Lorentizian penalties, respectively. These penalty

Name	**Penalty Function $\Psi(r)$**	**Example References**
L^p-norm	$\|r\|^p$	[7, 9, 23, 48, 69]
Modified L^1-norm	$\sqrt{r^2+\epsilon^2}$	[15, 24]
Huber	$\begin{cases} \frac{r^2}{2}, & \|r\| < k \\ k\left(\|r\| - \frac{k}{2}\right), & \|r\| \geq k \end{cases}$	[5]
Tukey	$\begin{cases} \frac{c^2}{6}\left(1 - \left[1 - \left(\frac{r}{c}\right)^2\right]^3\right), & \|r\| < c \\ \frac{c^2}{6}, & \|r\| \geq c \end{cases}$	[5]
Geman-McClure	$\frac{r^2}{r^2+\sigma^2}$	[76]
Lorentizian	$\log\left(1 + \frac{r^2}{2\sigma^2}\right)$	[27]

Table 2.2: Penalty functions of residuals for image registration.

functions can be compared based on their behavior towards outliers; the Huber functions exhibit linear behavior with respect to outlier magnitude; Lorentizian functions exhibit logarithmic behavior, and the Tukey and Geman-McClure exhibit (nearly) constant behavior for outliers.

One note of caution, however: behavior that is desirable in dealing with outliers might be undesirable in dealing with lack of fit (i.e., misregistration). While the Tukey and Geman-McClure penalties may work well for noisy images with small initial misregistrations, the lack of fit in larger misregistrations may be penalized as nearly constant, making it more difficult to converge to a correct registration result. There is not necessarily a "one-size-fits-all" answer to the question of which penalty function might be the most appropriate; for a particular registration task, it is recommended to experiment with various penalty functions to see which heuristically provides the best trade-off between robustness to outliers and ability to converge from larger initial misregistrations.

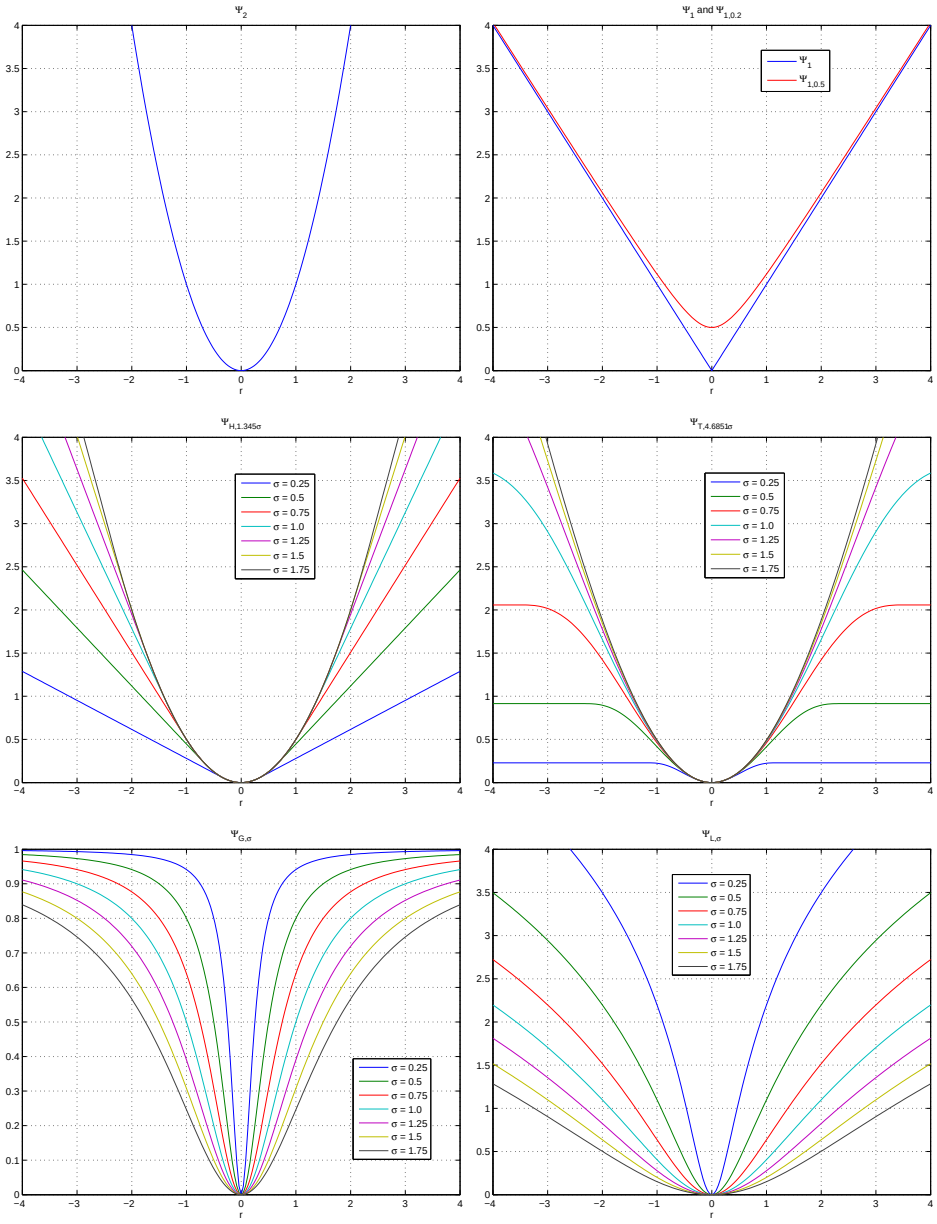


Figure 2.1: Various penalty functions displayed as functions of residuals; Upper left: Ψ_2, Upper right: Ψ_1 and $\Psi_{1,0.5}$, Middle left: $\Psi_{H,1.345\sigma}$ for various values of σ, Middle right: $\Psi_{T,4.6851\sigma}$ for various values of σ, Lower left: $\Psi_{G,\sigma}$ for various values of σ, and Lower right: $\Psi_{L,\sigma}$ for various values of σ.

2.2.2 Penalties on Deviation from Linearity Assumptions

In order to penalize deviations from the linearity assumption, we explicitly represent the unknown linear relationship between the images. If we define the parameter vector $\boldsymbol{\beta} = (\beta_1, \beta_2)^{\mathrm{T}}$, then the brightness linearity constraint can be written as:

$$\exists \boldsymbol{\beta} \text{ such that } F^{\mathbf{u}}(\mathbf{x}) = \beta_1 R(\mathbf{x}) + \beta_2 \quad \forall \mathbf{x} \in \Omega. \tag{2.30}$$

This suggests that one way to define residuals is by subtracting the left hand side of (2.30) from the right hand side:

$$r_{VDL}(\mathbf{x}; R, F^{\mathbf{u}}, \boldsymbol{\beta}) = \beta_1 R(\mathbf{x}) + \beta_2 - F^{\mathbf{u}}(\mathbf{x})\,. \tag{2.31}$$

The subscript VDL denotes *vertical difference from linearity*, indicating that the residuals are measured in terms of vertical distances to the line represented by β (see Fig. 2.2).

Another way in which residuals can be measured is in terms of the signed *orthogonal* distance to the line; i.e.,

$$r_{ODL}(\mathbf{x}; R, F^{\mathbf{u}}, \boldsymbol{\beta}) = \operatorname{sgn}\left(r_{VDL}(\mathbf{x}; R, F^{\mathbf{u}}, \boldsymbol{\beta})\right) \frac{|\beta_1 F^{\mathbf{u}}(\mathbf{x}) + \beta_2 - R(\mathbf{x})|}{\sqrt{1+\beta_1^2}}. \tag{2.32}$$

Here, the subscript ODL denotes *orthogonal difference from linearity*. The relationship between r_{VDL} and r_{ODL} is shown in Fig. 2.2; the form of r_{ODL} in (2.32) can be derived using similar triangles.

Any desired penalty function can be applied to r_{VDL} or r_{ODL} to construct a dissimilarity measure. Richard *et al.* [72] apply the quadratic penalty Ψ_2 to r_{VDL} to construct a dissimilarity measure which we refer to as the *mean squared vertical difference from linearity* (MSVDL):

$$\begin{aligned} MSVDL(R, F^{\mathbf{u}}, \boldsymbol{\beta}) &:= \frac{1}{|\Omega|} \int_{\Omega} \Psi_2(r_{VDL}(\mathbf{x}; R, F^{\mathbf{u}}, \boldsymbol{\beta}))\, d\mathbf{x} \\ &= \frac{1}{|\Omega|} \int_{\Omega} \left(\beta_1 R(\mathbf{x}) + \beta_2 - F^{\mathbf{u}}(\mathbf{x})\right)^2 d\mathbf{x}. \end{aligned} \tag{2.33}$$

A slightly different dissimilarity measure, which we refer to as the *mean squared orthogonal difference from linearity* (MSODL), can be constructed by applying the quadratic penalty

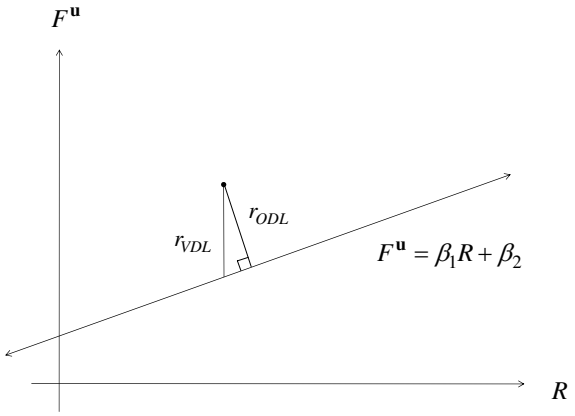


Figure 2.2: Relationship between r_{VDL} and r_{ODL}.

Ψ_2 to r_{ODL}:

$$\begin{aligned} MSODL(R, F^{\mathbf{u}}, \boldsymbol{\beta}) &:= \frac{1}{|\Omega|} \int_{\Omega} \Psi_2(r_{ODL}(\mathbf{x}; R, F^{\mathbf{u}}, \boldsymbol{\beta}))\, d\mathbf{x} \\ &= \frac{1}{|\Omega|} \int_{\Omega} \frac{(\beta_1 R(\mathbf{x}) + \beta_2 - F^{\mathbf{u}}(\mathbf{x}))^2}{1 + \beta_1^2} d\mathbf{x}. \end{aligned} \tag{2.34}$$

Another related dissimilarity measure, the *squared correlation coefficient* (SCC), is often used in registration [41, 19, 66, 35]:

$$SCC(R, F^{\mathbf{u}}) := \rho\left(R, F^{\mathbf{u}}\right)^2, \tag{2.35}$$

where

$$\rho\left(I_1, I_2\right) = \frac{\mathrm{Cov}\left(I_1, I_2\right)}{\sqrt{\mathrm{Var}\left(I_1\right)\mathrm{Var}\left(I_2\right)}}, \tag{2.36}$$

$$\mathrm{Var}\left(I\right) = \frac{1}{|\Omega|} \int_{\Omega} \left(I(\mathbf{x}) - \bar{I}\right)^2 d\mathbf{x}, \tag{2.37}$$

$$\mathrm{Cov}\left(I_1, I_2\right) = \frac{1}{|\Omega|} \int_{\Omega} \left(I_1(\mathbf{x}) - \bar{I}_1\right)\left(I_2(\mathbf{x}) - \bar{I}_2\right) d\mathbf{x}, \quad \text{and} \tag{2.38}$$

$$\bar{I} = \frac{1}{|\Omega|} \int_{\Omega} I(\mathbf{x})\, d\mathbf{x}. \tag{2.39}$$

Under the assumption that Ω does not change over the course of registration, minimizing a normalized version of the MSVDL is equivalent to maximizing the SCC. Furthermore, under

this same assumption, minimizing the MSODL is equivalent to maximizing the following dissimilarity measure, which we call the *squared orthogonal correlation coefficient* (SOCC):

$$SOCC(R, F^{\mathbf{u}}) := \frac{(\mathrm{Var}\,(R) - \mathrm{Var}\,(F^{\mathbf{u}}))^2 + 4\mathrm{Cov}\,(R, F^{\mathbf{u}})^2}{(\mathrm{Var}\,(R) + \mathrm{Var}\,(F^{\mathbf{u}}))^2}. \tag{2.40}$$

See Appendix A for proofs of these assertions.

The SCC and the SOCC are more compact dissimilarity measures than the MSVDL or MSODL, because they do not explicitly depend on $\boldsymbol{\beta}$. This is advantageous because registration can be carried out without the need to estimate the parameters of the linear model.

Any of the other penalty functions mentioned in Section 2.2.1 can be applied to r_{VDL} (2.31) or r_{ODL} (2.32) in order to construct robust dissimilarity measures for use when the linearity assumption holds. In general, unless a corresponding robust version of the correlation coefficient can be found, registration using these dissimilarity measures must either combine estimation of the parameters of $\boldsymbol{\beta}$ with estimation of the alignment transformation $\mathbf{u}$. Kim and Fessler [56] illustrate how to compute robust correlation coefficients for certain assumed error distributions without the need to estimate the linearity parameters. However, a more general approach can be designed for any differentiable penalty function by partitioning the alignment and β parameters in the minimization process.

Suppose the general registration problem can be stated for the linearity assumption as:

$$\min_{\mathbf{u},\beta} \mathcal{I}(R, F^{\mathbf{u}}; \beta) := \frac{1}{|\Omega|} \int_{\Omega} \Psi(r(\mathbf{x}; R, F^{\mathbf{u}}, \beta))\, d\mathbf{x}. \tag{2.41}$$

Then, Algorithm 2.2.2 describes a general approach to solving (2.41) by partitioning the minimization.

To illustrate how Algorithm 2.2.2 can be used in practice in nonrigid registration, we choose a particular penalty function ($\Psi_{1,\epsilon}$) and residual (r_{VDL}), so that the registration process can be stated as the minimization problem:

$$\min_{\mathbf{u},\beta} \mathcal{I}_{\Psi_{1,\epsilon}, r_{VDL}}(R, F^{\mathbf{u}}; \beta) := \frac{1}{|\Omega|} \int_{\Omega} \sqrt{(\beta_1 R(\mathbf{x}) + \beta_2 - F^{\mathbf{u}}(\mathbf{x}))^2 + \epsilon^2}\, d\mathbf{x}. \tag{2.42}$$

To perform the second step in Algorithm 2.2.2 for solving (2.42), we can determine analytically that if we fix $\mathbf{u} = \hat{\mathbf{u}}^{(k)}$, the first order conditions for minimizing (2.42) with respect

Algorithm 2.2.2: Partitioned Solution of (2.41)

1. Select an initial alignment estimate $\hat{\mathbf{u}}^{(0)}$. Initialize $k = 0$.
2. Solve the minimization problem $\hat{\beta}^{(k)} = \arg\min_{\beta} \mathcal{I}\Big(R, F^{\hat{\mathbf{u}}^{(k)}}; \beta\Big)$.
3. Solve the minimization problem $\hat{\mathbf{u}}^{(k+1)} = \arg\min_{\mathbf{u}} \mathcal{I}\Big(R, F^{\mathbf{u}}, \hat{\beta}^{(k)}\Big)$.
4. If $\hat{\mathbf{u}}^{(k+1)}$ is judged to be close enough to $\hat{\mathbf{u}}^{(k)}$, terminate. If not, increment k and return to step 2.

to β yield the solution:

$$\hat{\beta}_1^{(k)} = \frac{\text{Cov}\left(R, F^{\hat{\mathbf{u}}^{(k)}}\right)}{\text{Var}\,(R)}, \tag{2.43}$$

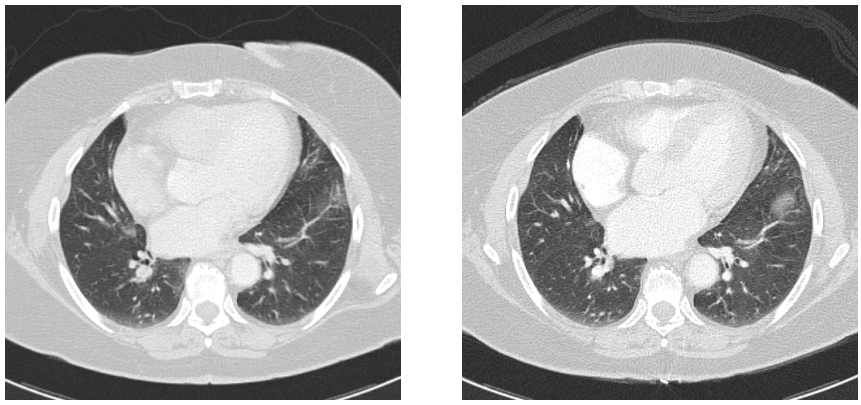
$$\hat{\beta}_2^{(k)} = \bar{F}^{\hat{\mathbf{u}}^{(k)}} - \bar{R}\frac{\text{Cov}\left(R, F^{\hat{\mathbf{u}}^{(k)}}\right)}{\text{Var}\,(R)}. \tag{2.44}$$

Once $\hat{\beta}_1^{(k)}$ and $\hat{\beta}_2^{(k)}$ have been estimated, the next step in Algorithm 2.2.2 can be performed by computing the corresponding force vector (described in Section 2.4) and then solving for $\hat{\mathbf{u}}^{(k+1)}$ using one of the techniques described in Chapter 5.

The images shown in Fig. 2.3 are axial slices from chest CT scans of the same patient taken at different times. After normalizing the values in both images to range from zero to one, we performed nonrigid registration by using Algorithm 2.2.2 to solve (2.42). We chose the value of ϵ^2 to be 0.01, and we used the fluid registration approach (described later in Section 5.1.3) to perform the third step of Algorithm 2.2.2.

Registration terminated after 133 iterations due to the maximum difference in successive estimates of $\hat{\mathbf{u}}^{(k)}$ being less than a prescribed tolerance of 0.1 pixels. The resulting aligned prior image is shown in Fig. 2.2.2, and difference images before and after registration are illustrated in Fig. 2.5.

As shown in Fig. 2.6, the values of β started with $\hat{\beta}_1^{(0)} \approx 0.8444$ and $\hat{\beta}_2^{(0)} \approx 0.0979$ and ended with $\hat{\beta}_1^{(133)} \approx 0.9639$ and $\hat{\beta}_2^{(133)} \approx 0.0088$. Furthermore, the dissimilarity measure had initial value of approximately 0.15644 and decreased to a final value of approximately 0.1211. The slight and sudden decrease in values at iteration 90 corresponds to a regridding step of the fluid registration approach.



(a) Prior (b) Current

Figure 2.3: Example axial slices of serial 3D CT scans of the same patient.

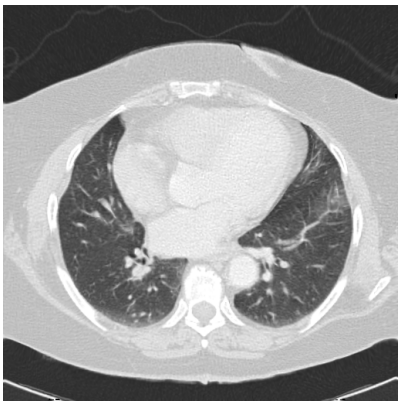
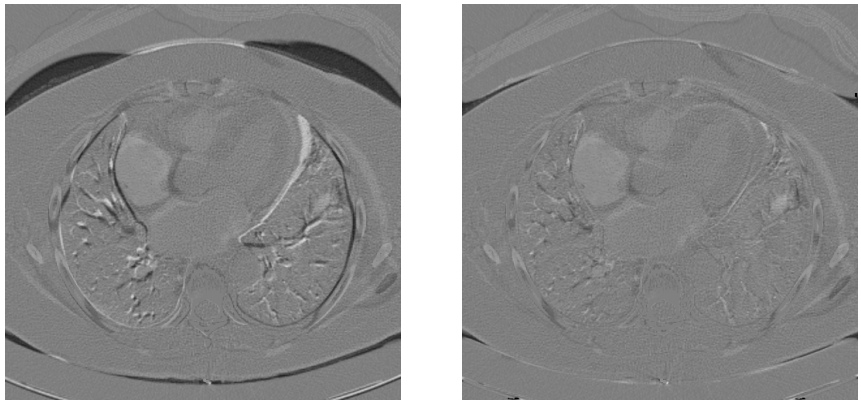
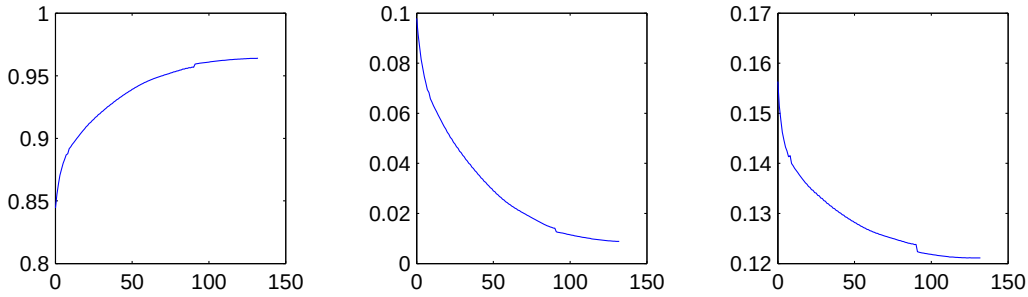


Figure 2.4: Prior CT slice after registration.



(a) Differences before registration (b) Differences after registration

Figure 2.5: Differences between current and prior slices, before (left) and after (right) registration.



(a) $\hat{\beta}_1$ (b) $\hat{\beta}_2$ (c) $\mathcal{I}_{\Psi_{1,\epsilon},r_{VDL}}$

Figure 2.6: Values of $\hat{\beta}_1$, $\hat{\beta}_2$, and dissimilarity measure $\mathcal{I}_{\Psi_{1,\epsilon},r_{VDL}}$ as a function of iteration over the course of registration.

A similar type of behavior should be expected of Algorithm 2.2.2 whenever the general dissimilarity measure is convex in β for a fixed displacement field. However, in situations where convexity is lost, one must take great care in employing a partitioned algorithm. As soon as a local minimum that is different than the global minimum is chosen for β in the second step, this will likely cause Algorithm 2.2.2 to diverge (or to converge to the wrong minimum).

2.2.3 Penalties on Deviation from Functional Assumptions

For functional assumptions that are more general than linearity, the functional form relating the images is rarely known. If it *is* known (e.g., in [98]) and can be explicitly parameterized, any of the penalty functions mentioned in 2.2.1–2.2.2 can be applied, and registration can be carried out according to a partitioned algorithm like Algorithm 2.2.2.

In more common scenarios, the existence but not the form of a functional relationship between images is known. Woods *et al.* [99] investigate the case of MR/PET registration, and they implicitly assume that such a functional relationship exists. By partitioning the MR image into a number of components according to pixel values, Woods develops a criterion that maximizes the uniformity of the PET pixel values within each partition. This is achieved by minimizing a weighted average of the standard deviation of the PET values in each partition. The *Woods criterion* is given by:

$$W(R, F^{\mathbf{u}}) := \frac{1}{N} \sum_i N_i \frac{\sigma_i(F^{\mathbf{u}})}{\mu_i(F^{\mathbf{u}})}, \tag{2.45}$$

where

$$\mu_i(F^{\mathbf{u}}) = \frac{1}{N_i} \sum_{\mathbf{x}\in\Omega_i} F^{\mathbf{u}}(\mathbf{x}), \tag{2.46}$$

$$\sigma_i(F^{\mathbf{u}}) = \sqrt{\frac{1}{N_i} \sum_{\mathbf{x}\in\Omega_i} \left(F^{\mathbf{u}}(\mathbf{x})^2 - \mu_i(F^{\mathbf{u}})^2 \right)}, \tag{2.47}$$

$\Omega = \{\bigcup_{i=1}^{n} \Omega_i\}$ is a partitioning of the image domain, N_i is the number of pixels in Ω_i, and N is the total number of pixels.

A few years after the Woods criterion was introduced, Roche *et al.* [74, 75] showed that if a function $f^*(R)$ is defined as the conditional expectation $E[F^{\mathbf{u}}|R]$, then f^* is the best possible fit of $F^{\mathbf{u}}$ to $f(R)$ for *any* function f, as measured by the 2-norm of the residuals:

$$f^* = \min_f \left\{ \frac{1}{|\Omega|} \int_\Omega \Psi_2 \left(F^{\mathbf{u}}(\mathbf{x}) - f\left(R(\mathbf{x})\right) \right) d\mathbf{x} \right\}. \tag{2.48}$$

Using this result, Roche [75] relates (2.48) to the variance of the residuals, and then uses an application of the total variance theorem to propose the *correlation ratio* as a dissimilarity measure:

$$CR(R, F^{\mathbf{u}}) := \eta(F^{\mathbf{u}}|R) = \frac{\mathrm{Var}\left(E\left[F^{\mathbf{u}}|R\right]\right)}{\mathrm{Var}\left(F^{\mathbf{u}}\right)}. \tag{2.49}$$

The correlation ratio is analogous to the squared correlation coefficient; both range from zero (no functional/linear dependence) to one (total functional/linear dependence). Roche [74] argues that the development of the correlation ratio provides theoretical justification for the Woods criterion; furthermore, he illustrates experimentally that the correlation ratio and the Woods criterion behave similarly in practice, with the correlation ratio exhibiting some small practical improvements.

It is possible to apply other penalty functions when an unknown functional relationship exists between images. However, to this author's knowledge, the only way this has been done is by restricting the unknown functional relationship to be part of a known class of functions. In an application of MR/ultrasound registration, Roche *et al.* [76] assume the functional relationship is a multivariate polynomial, based on the intensities *and* intensity gradient magnitudes of the MR images. They pose the objective function as a Geman-McClure penalty of the differences between the predicted polynomial value and ultrasound

intensity. They estimate the polynomial coefficients and optimal rigid transformation in a partitioned approach, which can be generalized by Algorithm 2.2.3.

Suppose the general registration problem can be stated for the functional assumption as:

$$\min_{\mathbf{u},\theta} \mathcal{I}(R, F^{\mathbf{u}}; \boldsymbol{\theta}) := \frac{1}{|\Omega|} \int_{\Omega} \Psi(r(\mathbf{x}; f(R; \boldsymbol{\theta}), F^{\mathbf{u}}))\, d\mathbf{x}, \tag{2.50}$$

where f is a function parameterized by the vector $\boldsymbol{\theta}$. Then, Algorithm 2.2.3 describes a general approach to solving (2.50) by partitioning the minimization. As with Algorithm 2.2.2, one must be cautious if the objective function $\mathcal{I}\Big(R, F^{\hat{\mathbf{u}}^{(k)}}; \theta\Big)$ is nonconvex in θ.

Algorithm 2.2.3: Partitioned Solution of (2.50)

1. Select an initial alignment estimate $\hat{\mathbf{u}}^{(0)}$. Initialize $k = 0$.
2. Solve the minimization problem $\hat{\theta}^{(k)} = \arg\min_{\theta} \mathcal{I}\Big(R, F^{\hat{\mathbf{u}}^{(k)}}; \theta\Big)$.
3. Solve the minimization problem $\hat{\mathbf{u}}^{(k+1)} = \arg\min_{\mathbf{u}} \mathcal{I}\Big(R, F^{\mathbf{u}}; \hat{\theta}^{(k)}\Big)$.
4. If $\hat{\mathbf{u}}^{(k+1)}$ is judged to be close enough to $\hat{\mathbf{u}}^{(k)}$, terminate. If not, increment k and return to step 2.

2.2.4 Penalties on Deviation from Information Content Assumptions

The simplest way to construct a dissimilarity measure that would penalize deviations from the information content assumption is to use the joint entropy itself, given by (2.11). Joint entropy has been investigated for multimodal image alignment, for instance in [34] and [83].

One problem with the use of joint entropy for image registration is that a reduction in joint entropy can be effected by a reduction in marginal entropy. This could yield an "optimal" alignment having the least amount of information rather than the most *corresponding* information.

Viola and Wells [92] and Collignon *et al.* [34] independently proposed the use of mutual information to combat this problem. Mutual information is given by the difference between the sum of the marginal entropies and the joint entropy:

$$MI(R, F^{\mathbf{u}}) := H(R) + H(F^{\mathbf{u}}) - H(R, F^{\mathbf{u}}). \tag{2.51}$$

The negative mutual information can be thought of as a penalty on deviations from the information content assumption. Maximizing the mutual information (or minimizing the negative mutual information) will tend to simultaneously reduce joint entropy while maintaining marginal entropy.

Studholme *et al.* [84] argued that any image dissimilarity measure should be invariant to changes in the overlap region through the course of registration. They showed that mutual information does in fact vary with changing overlap, and they proposed the *normalized mutual information* as an alternative:

$$NMI(R, F^{\mathbf{u}}) := \frac{H(R) + H(F^{\mathbf{u}})}{H(R, F^{\mathbf{u}})}. \tag{2.52}$$

Studholme *et al.* validated the invariance of NMI to changing overlap for some simple examples; however, NMI still does not guarantee overlap invariance in every situation. (Chapter 3 of this thesis will further explore the notion of overlap invariance and propose overlap invariant modifications to *all* of the standard dissimilarity measures.)

NMI is closely related to Astola's *entropy correlation coefficient* [8], which was explored for multimodal image registration in [33, 61]. The entropy correlation coefficient is given by:

$$ECC(R, F^{\mathbf{u}}) := \sqrt{2 - \frac{2H(R, F^{\mathbf{u}})}{H(R) + H(F^{\mathbf{u}})}}. \tag{2.53}$$

Dissimilarity measures can also be constructed using the cumulative residual entropy in place of the differential entropy. Rao *et al.* [71] use the JCRE itself as a dissimilarity measure. However, similarly to joint entropy, JCRE suffers from the problem that a reduction in JCRE can be effected by a reduction in CRE. A solution to this problem is to use the *cross cumulative residual entropy* (CCRE), which is analogous to mutual information:

$$\begin{aligned} CCRE(R, F^{\mathbf{u}}) &:= \varepsilon(R) + \varepsilon(F^{\mathbf{u}}) - \varepsilon(F^{\mathbf{u}}, R) \\ &= \varepsilon(R) - \mathrm{E}[\varepsilon(R|F^{\mathbf{u}})]\,, \end{aligned} \tag{2.54}$$

with ε defined according to (2.12)–(2.14). CCRE was explored by Wang *et al.* [93] for nonrigid registration; it was also investigated by [37] under the alternate name of cumulative mutual information (CMI).

Since JCRE is not symmetric, neither is CCRE. However, CCRE can easily be symmetrized; the *symmetric cross cumulative residual entropy* (SCCRE) is given by:

$$\begin{aligned} SCCRE(R, F^{\mathbf{u}}) &:= \frac{1}{2}\Big(CCRE(R, F^{\mathbf{u}}) + CCRE(F^{\mathbf{u}}, R)\Big) \\ &= \varepsilon(R) + \varepsilon(F^{\mathbf{u}}) - \varepsilon_S(R, F^{\mathbf{u}})\,. \end{aligned} \tag{2.55}$$

As with mutual information, CCRE and SCCRE vary with changing overlap regions. We can attempt to remedy this problem by using the CRE to form analogous versions of NMI and ECC. We will call these dissimilarity measures the *normalized cross cumulative residual entropy* (NCCRE) and the *cumulative residual entropy correlation coefficient* (CRECC):

$$\begin{aligned} NCCRE(R, F^{\mathbf{u}}) &:= \frac{\varepsilon(R) + \varepsilon(F^{\mathbf{u}})}{\varepsilon_S(R, F^{\mathbf{u}})} \\ &= \frac{2\varepsilon(R) + 2\varepsilon(F^{\mathbf{u}})}{\varepsilon(R) + \varepsilon(F^{\mathbf{u}}) + \mathrm{E}\left[\varepsilon(R|F^{\mathbf{u}})\right] + \mathrm{E}\left[\varepsilon(F^{\mathbf{u}}|R)\right]}, \end{aligned} \tag{2.56}$$

$$\begin{aligned} CRECC(R, F^{\mathbf{u}}) &:= \sqrt{2 - \frac{2\varepsilon_S(R, F^{\mathbf{u}})}{\varepsilon(R) + \varepsilon(F^{\mathbf{u}})}} \\ &= \sqrt{1 - \frac{\mathrm{E}\left[\varepsilon(R|F^{\mathbf{u}})\right] + \mathrm{E}\left[\varepsilon(F^{\mathbf{u}}|R)\right]}{\varepsilon(R) + \varepsilon(F^{\mathbf{u}})}}. \end{aligned} \tag{2.57}$$

In Chapter 3 of this thesis, we will show that the NCCRE and CRECC, like the NMI and ECC, are not overlap invariant in all cases, and we will propose overlap invariant modifications to address this concern.

2.3 From Global to Local

All of the dissimilarity measures presented in Section 2.2 are global in nature. If each pixel/voxel of an image is considered a realization of a random variable, then the entire image can be considered a realization of a stochastic process. The use of global dissimilarity measures for registration implicitly treats images as *stationary* stochastic processes. In most applications, however, images are more adequately represented by *non-stationary* stochastic processes. Non-stationarities can be accounted for in image registration by using *local* dissimilarity measures. This idea was introduced by Hermosillo [50]; he used estimates of

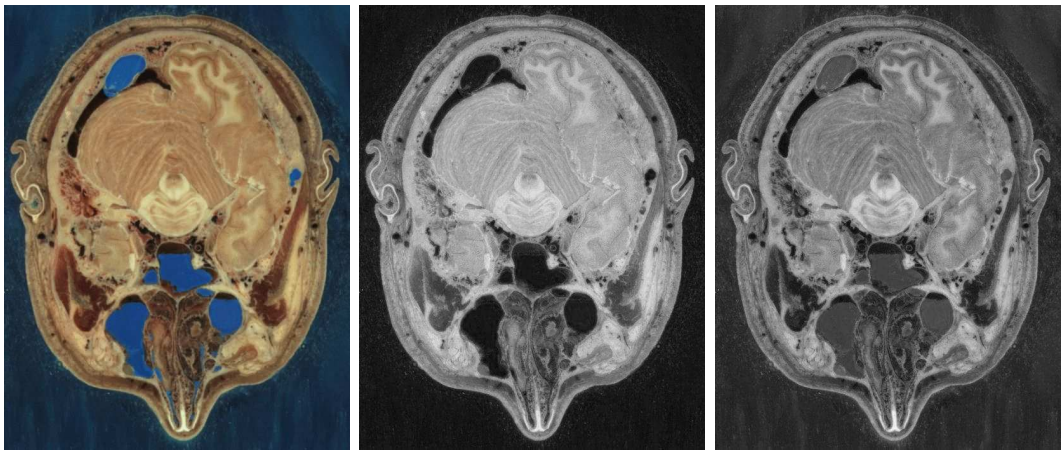


Figure 2.7: Left: color cryosection of a human head, Middle: reference image (red channel), Right: floating image (green channel)

local joint distributions to construct local versions of the SCC, CR, and MI dissimilarity measures.

In this section, we show how to extend Hermosillo's idea to construct local versions of all of the dissimilarity measures in Section 2.2. For the dissimilarity measures based on the constancy and linearity assumptions, this can be done with local moments or with local distributions. For the dissimilarity measures based on the functional and information content assumptions, this is done by a direct application of Hermosillo's estimates of local joint distributions.

To illustrate the local dissimilarity measures on real image data, we use a color cryosection of a human head from the Visible Human Project® [11]. The cryosection contains 700×545 pixels, with each pixel $0.33 \times 0.33\,\mathrm{mm}^2$. The red and green channels are extracted and converted into single-channel intensity images, as illustrated in Fig. 2.7. These images can be considered reference and floating images from a multimodal data set. In the next subsections, we will use these images to show the behavior of various local dissimilarity measures.

2.3.1 Dissimilarity Measures Based on Local Moments

Recall from (2.39) that the expected value of an image can be computed by integrating the image over its support and normalizing by the area (for a 2-D image) or volume (for a 3-D

image). (2.39) represents the *global* mean of an image. The *local* mean of an image at a point $\mathbf{x}_0$ can be computed by weighting the image prior to integration; i.e.,

$$\mathrm{E}_\gamma[I;\mathbf{x}_0] := \frac{\int_\Omega w_\gamma(\mathbf{x}-\mathbf{x}_0)\, I(\mathbf{x})\, d\mathbf{x}}{\int_\Omega w_\gamma(\mathbf{x}-\mathbf{x}_0)\, d\mathbf{x}}. \tag{2.58}$$

Ideally, the weighting function $w_\gamma(\mathbf{x})$ is a radially symmetric function parameterized by scale factor γ. For this thesis, we assume $w_\gamma(\mathbf{x})$ to be Gaussian:

$$w_\gamma(\mathbf{x}) = \frac{1}{\sqrt{(2\pi\gamma^2)^N}} \exp\left(-\frac{\|\mathbf{x}\|^2}{2\gamma^2}\right), \tag{2.59}$$

where N is the dimension of $\mathbf{x}$.

In order to use a Gaussian weighting function, (2.59) requires that $\gamma \in (0,\infty)$. However, we can allow zero and infinite scale by defining:

$$\mathrm{E}_0[I;\mathbf{x}_0] := \lim_{\gamma\to 0^+} \mathrm{E}_\gamma[I;\mathbf{x}_0] = I(\mathbf{x}_0)\,, \tag{2.60}$$

$$\mathrm{E}_\infty[I;\mathbf{x}_0] := \lim_{\gamma\to \infty} \mathrm{E}_\gamma[I;\mathbf{x}_0] = \bar{I}. \tag{2.61}$$

From (2.61), we see that the local mean with infinite scale is equivalent to the global mean.

All of the dissimilarity measures based on constancy or linearity assumptions can be written as expected values; hence, local versions of these dissimilarity measures can be constructed by using local means.

2.3.1.1 Constancy Assumptions

The general constancy dissimilarity measure (2.19) can be written in terms of expected values as:

$$\mathcal{I}^g_{\Psi,r}(R,F^{\mathbf{u}}) = \mathrm{E}[\Psi(r(\mathbf{x};R,F^{\mathbf{u}}))]\,. \tag{2.62}$$

Hence, the local dissimilarity measure $\mathcal{I}^{\ell,\gamma}_{\Psi,r}$ can be computed at $\mathbf{x}_0$ by:

$$\mathcal{I}^{\ell,\gamma}_{\Psi,r}(R,F^{\mathbf{u}},\mathbf{x}_0) = \mathrm{E}_\gamma[\Psi(r(\mathbf{x};R,F^{\mathbf{u}}))\,;\mathbf{x}_0]\,. \tag{2.63}$$

Recall that MSD and MAD are specific versions of $\mathcal{I}^g_{\Psi,r}$, with $MSD = \mathcal{I}^g_{\Psi_2,r_{BC}}$ and $MAD = \mathcal{I}^g_{\Psi_1,r_{BC}}$. Therefore, local versions of MSD and MAD, computed at $\mathbf{x}_0$, are given

by:

$$MSD_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0) = \mathrm{E}_\gamma\left[(F^{\mathbf{u}} - R)^2 ; \mathbf{x}_0\right], \tag{2.64}$$

$$MAD_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0) = \mathrm{E}_\gamma[|F^{\mathbf{u}} - R| ; \mathbf{x}_0]. \tag{2.65}$$

Aggregate local dissimilarity measures are defined by averaging the local dissimilarity measures over all points in the domain; i.e.,

$$\mathcal{I}^{\ell,\gamma}_{\Psi,r}(R, F^{\mathbf{u}}) = \frac{1}{|\Omega|}\int_\Omega \mathcal{I}^{\ell,\gamma}_{\Psi,r}(R, F^{\mathbf{u}}; \mathbf{x}_0)\, d\mathbf{x}_0, \tag{2.66}$$

$$MSD_\gamma(R, F^{\mathbf{u}}) = \frac{1}{|\Omega|}\int_\Omega MSD_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)\, d\mathbf{x}_0, \tag{2.67}$$

$$MAD_\gamma(R, F^{\mathbf{u}}) = \frac{1}{|\Omega|}\int_\Omega MAD_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)\, d\mathbf{x}_0. \tag{2.68}$$

Note that this construction of (2.67) serves to unify Hermosillo's idea with the combined local-global (CLG) optic flow method of Bruhn, Weickert, and Schnörr [17]. The Lucas-Kanade term in the CLG optic flow method is a linearized version of (2.67).

Figures 2.8–2.10 illustrate some examples of the local dissimilarity measures (2.63)–(2.65) when applied to the reference and floating images in Fig. 2.7. Note the blurring effect of the local dissimilarity measures as the scale is increased. As indicated by (2.61), we would expect that as the scale parameter γ approaches infinity, these images would tend towards the constant value taken on by the corresponding global dissimilarity measure. As γ approaches zero, the local dissimilarity measure approaches the point-based measure that can be found by applying the penalty function directly to the residual computed at each position.

One important observation to take away from Fig.'s 2.8–2.10 is that as γ increases, the local neighborhood increases its width independently of spatial position and image content. If nonhomogeneous and/or anisotropic behavior is desired, local neighborhood windows could be constructed with locally varying width and directionality; although it is not clear that such behavior would be appropriate for the task of computing local moments.

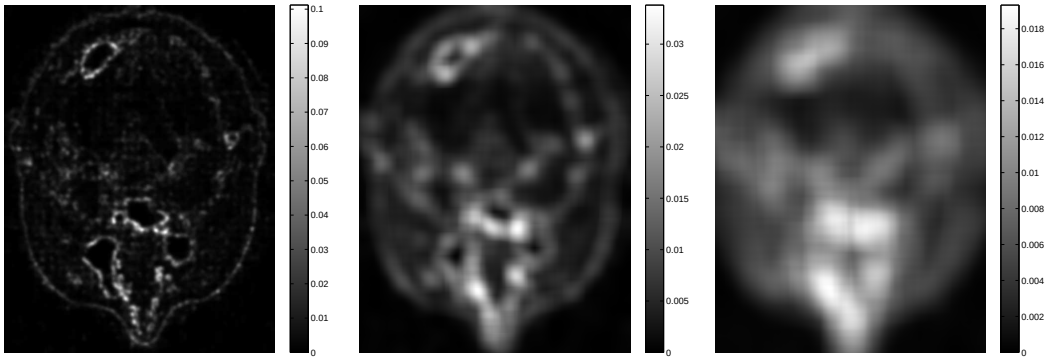


Figure 2.8: Examples of $\mathfrak{I}^{\ell,\gamma}_{\Psi_2,r_{BMC}}(R, F^{\mathbf{u}}; \mathbf{x}_0)$ (the L^2 norm of the brightness morphology residual image) for $\gamma = 1\,\text{mm}$ (left), $\gamma = 5\,\text{mm}$ (middle), and $\gamma = 10\,\text{mm}$ (right).

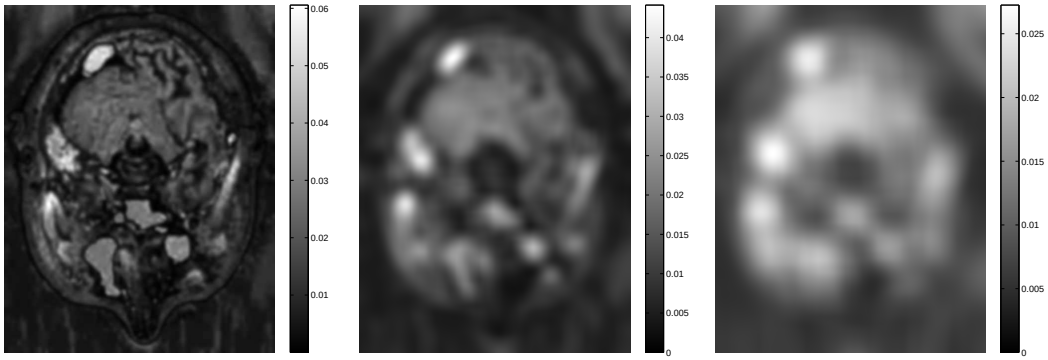


Figure 2.9: Examples of $MSD_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)$ for $\gamma = 1\,\text{mm}$ (left), $\gamma = 5\,\text{mm}$ (middle), and $\gamma = 10\,\text{mm}$ (right).

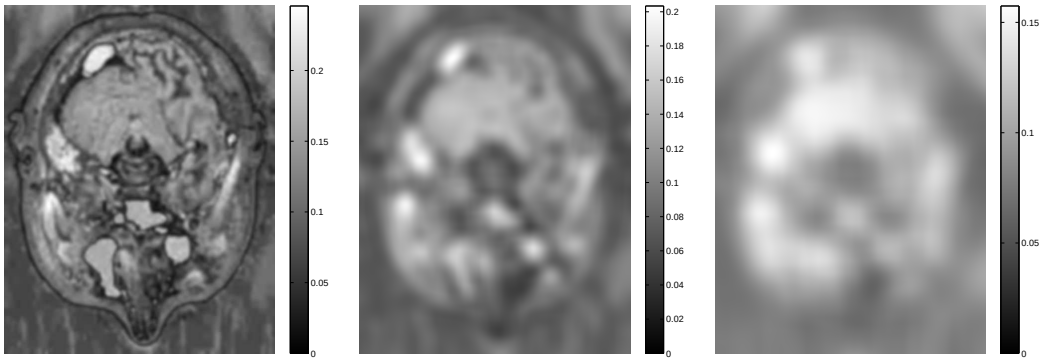


Figure 2.10: Examples of $MAD_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)$ for $\gamma = 1\,\text{mm}$ (left), $\gamma = 5\,\text{mm}$ (middle), and $\gamma = 10\,\text{mm}$ (right).

2.3.1.2 Linearity Assumptions

The global MSVDL and global MSODL can be written in terms of expected values as:

$$MSVDL(R, F^{\mathbf{u}}; \beta) = \mathrm{E}\Big[(\beta_1 R + \beta_2 - F^{\mathbf{u}})^2\Big], \tag{2.69}$$

$$MSODL(R, F^{\mathbf{u}}; \beta) = \frac{MSVDL(R, F^{\mathbf{u}}; \beta)}{1 + \beta_1^2}. \tag{2.70}$$

Hence, local versions of MSVDL and MSODL can be computed at $\mathbf{x}_0$ by:

$$MSVDL_\gamma(R, F^{\mathbf{u}}; \beta, \mathbf{x}_0) = \mathrm{E}_\gamma\Big[(\beta_1(\mathbf{x}_0)\, R + \beta_2(\mathbf{x}_0) - F^{\mathbf{u}})^2 ; \mathbf{x}_0\Big], \tag{2.71}$$

$$MSODL_\gamma(R, F^{\mathbf{u}}; \beta, \mathbf{x}_0) = \frac{MSVDL_\gamma(R, F^{\mathbf{u}}; \beta, \mathbf{x}_0)}{1 + \beta_1(\mathbf{x}_0)^2}. \tag{2.72}$$

Integrating (2.71) and (2.72) over all points in the domain yields the aggregate local dissimilarity measures:

$$MSVDL_\gamma(R, F^{\mathbf{u}}, \beta) = \frac{1}{|\Omega|} \int_\Omega MSVDL_\gamma(R, F^{\mathbf{u}}; \beta, \mathbf{x}_0)\, d\mathbf{x}_0, \tag{2.73}$$

$$MSODL_\gamma(R, F^{\mathbf{u}}, \beta) = \frac{1}{|\Omega|} \int_\Omega MSODL_\gamma(R, F^{\mathbf{u}}; \beta, \mathbf{x}_0)\, d\mathbf{x}_0. \tag{2.74}$$

Note that for (2.71) and (2.72), β depends on $\mathbf{x}_0$. Therefore, on the left hand sides of (2.73) and (2.74), β can be thought of as a vector-valued image of linearity coefficients.

Figure 2.11 illustrates images of the components of β that minimize MSVDL for the reference and floating images shown in Fig. 2.7 under varying scale. Two interesting observations can be made based on these figures. First, at the finer scale (γ = 1mm), there appears to be a significant amount of variation in the values of β_1, especially in the background region. This fluctuation is to be expected, since we are in effect fitting a model to a small amount of data with a low signal-to-noise ratio. Second, at the larger scales (γ = 5mm, 10mm), both β_1 and β_2 seem to have a more piecewise constant appearance to them than the quantities in the previous figures, suggesting that even with larger neighborhoods, estimates of β may not necessarily be insensitive to small neighborhood translations.

In order to present local versions of SCC and SOCC, we first introduce the local variance and local covariance. Variance and covariance can be written in terms of expected values

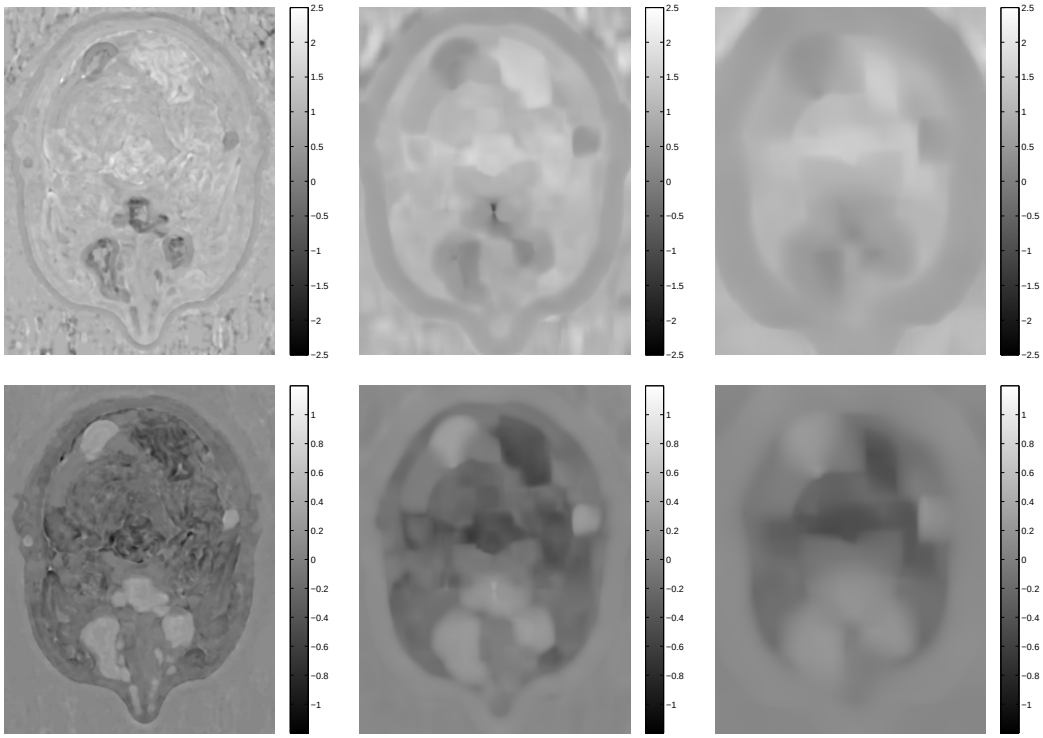


Figure 2.11: Examples of $\beta(\mathbf{x}_0)$ for β_1 (top) and β_2 (bottom), with $\gamma = 1\,\text{mm}$ (left), $\gamma = 5\,\text{mm}$ (middle), and $\gamma = 10\,\text{mm}$ (right).

in the following way:

$$\text{Var}(A) = \text{E}\big[A^2\big] - \text{E}[A]^2\,, \tag{2.75}$$

$$\text{Cov}(A,B) = \text{E}[AB] - \text{E}[A]\cdot\text{E}[B]\,. \tag{2.76}$$

Therefore, local variance and local covariance can be defined as:

$$\text{Var}_\gamma(A;\mathbf{x}_0) := \text{E}_\gamma\big[A^2;\mathbf{x}_0\big] - \text{E}_\gamma[A;\mathbf{x}_0]^2\,, \tag{2.77}$$

$$\text{Cov}_\gamma(A,B;\mathbf{x}_0) := \text{E}_\gamma[AB;\mathbf{x}_0] - \text{E}_\gamma[A;\mathbf{x}_0]\cdot\text{E}_\gamma[B;\mathbf{x}_0]\,. \tag{2.78}$$

Since the global SCC (2.35) can be written in terms of (2.75) and (2.76), the local SCC at $\mathbf{x}_0$ can be written in terms of (2.77) and (2.78):

$$SCC_\gamma(R,F^{\mathbf{u}};\mathbf{x}_0) = \frac{\text{Cov}_\gamma\left(R,F^{\mathbf{u}};\mathbf{x}_0\right)^2}{\text{Var}_\gamma\left(R;\mathbf{x}_0\right)\text{Var}_\gamma\left(F^{\mathbf{u}};\mathbf{x}_0\right)}. \tag{2.79}$$

Like the local SCC, the local SOCC at $\mathbf{x}_0$ can be written in terms of (2.77) and (2.78) in

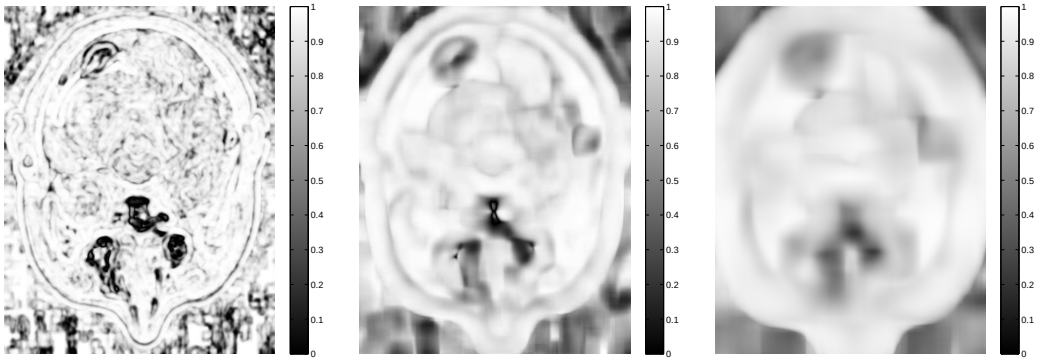


Figure 2.12: Examples of $SCC_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)$ for $\gamma = 1\,\text{mm}$ (left), $\gamma = 5\,\text{mm}$ (middle), and $\gamma = 10\,\text{mm}$ (right).

the following way:

$$SOCC_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0) = \frac{\left(\text{Var}_\gamma\left(R, \mathbf{x}_0\right) - \text{Var}_\gamma\left(F^{\mathbf{u}}, \mathbf{x}_0\right)\right)^2 + 4\text{Cov}_\gamma\left(R, F^{\mathbf{u}}, \mathbf{x}_0\right)^2}{\left(\text{Var}_\gamma\left(R, \mathbf{x}_0\right) + \text{Var}_\gamma\left(F^{\mathbf{u}}, \mathbf{x}_0\right)\right)^2}. \tag{2.80}$$

Integrating (2.79) and (2.80) over the entire domain yields the aggregate measures:

$$SCC_\gamma(R, F^{\mathbf{u}}) = \int_\Omega SCC_\gamma(R, F^{\mathbf{u}}, \mathbf{x}_0)\, d\mathbf{x}_0, \tag{2.81}$$

$$SOCC_\gamma(R, F^{\mathbf{u}}) = \int_\Omega SOCC_\gamma(R, F^{\mathbf{u}}, \mathbf{x}_0)\, d\mathbf{x}_0. \tag{2.82}$$

Figures 2.12 and 2.13 illustrate examples of local SCC and local SOCC for different values of γ. Similarly to what was observed in Fig. 2.11, we see that the local SCC and local SOCC at the smallest scale appear to exhibit much more fluctuation in the background values than do local MSD or local MAD, which could potentially lead to more difficulty in robustly converging to a correct registration result. Furthermore, at the medium and larger scales, local SCC and local SOCC also have a blocky appearance, suggesting a sensitivity to small neighborhood translations.

2.3.2 Dissimilarity Measures Based on Local Distributions

Hermosillo's method [50] for constructing local dissimilarity measures relies on the notion that images are realizations of stochastic processes. Global dissimilarity measures implicitly assume stationarity of the random process, whereas local dissimilarity measures allow a more general nonstationarity assumption.

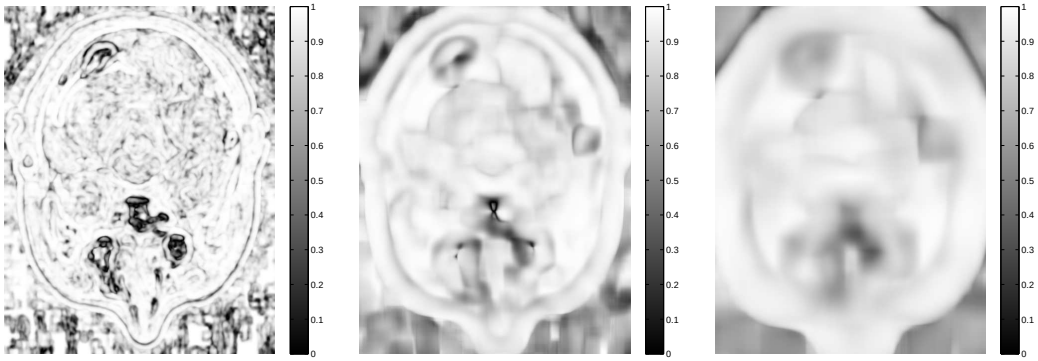


Figure 2.13: Examples of $SOCC_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)$ for $\gamma = 1\,\text{mm}$ (left), $\gamma = 5\,\text{mm}$ (middle), and $\gamma = 10\,\text{mm}$ (right).

The actual distribution of an image's underlying random variable is typically unknown; however, it can be approximated from the image data itself. For an image A, the probability density function p_A can be defined in terms of a Parzen window w by:

$$p_A(a) = \frac{1}{|\Omega|} \int_\Omega w_\sigma(A(\mathbf{x}) - a)\, d\mathbf{x}. \tag{2.83}$$

In the Parzen window method [67], w must integrate to unity; we choose w to be a Gaussian of the form (2.59) to be consistent with [50]. The joint distribution of two images can be similarly approximated:

$$p_{A_1,A_2}(a_1, a_2) = \frac{1}{|\Omega|} \int_\Omega w_\sigma(\mathbf{A}(\mathbf{x}) - \mathbf{a})\, d\mathbf{x}, \tag{2.84}$$

where $\mathbf{A}(\mathbf{x}) = (A_1(\mathbf{x}), A_2(\mathbf{x}))^{\mathrm{T}}$ and $\mathbf{a} = (a_1, a_2)^{\mathrm{T}}$.

Hermosillo [50] shows that nonstationarity can be realized by making the distribution approximation local. The local joint distribution with scale γ at $\mathbf{x}_0$ can be approximated by:

$$p^\gamma_{A_1,A_2}(a_1, a_2; \mathbf{x}_0) = \frac{\int_\Omega w_\sigma(\mathbf{A}(\mathbf{x}) - \mathbf{a})\, w_\gamma(\mathbf{x} - \mathbf{x}_0)\, d\mathbf{x}}{\int_\Omega w_\gamma(\mathbf{x} - \mathbf{x}_0)\, d\mathbf{x}}. \tag{2.85}$$

The marginal distribution of (2.85) then yields an estimate of the local distribution $p^\gamma_{A_1}$; i.e.,

$$p^\gamma_{A_1}(a_1; \mathbf{x}_0) = \int_{-\infty}^{\infty} p_{A_1,A_2}(a_1, a_2; \mathbf{x}_0)\, da_2. \tag{2.86}$$

Since the Parzen window (2.59) is separable, an alternative local estimate of $p^\gamma_{A_1}$ could be

computed directly without estimating the local joint distribution; i.e.:

$$p^{\gamma}_{A_1}(a;\mathbf{x}_0) = \frac{\int_\Omega w_\sigma(A_1(\mathbf{x}) - a)\, w_\gamma(\mathbf{x}-\mathbf{x}_0)\, d\mathbf{x}}{\int_\Omega w_\gamma(\mathbf{x}-\mathbf{x}_0)\, d\mathbf{x}}. \tag{2.87}$$

Local versions of dissimilarity measures based on the constancy assumptions can be constructed from (2.87). Local versions of dissimilarity measures based on the linearity, functional, and information content assumptions can be constructed from (2.85) and (2.87).

One final note: when dissimilarity measures are constructed from estimates of local distributions instead of local moments, a practical problem arises: having enough samples to adequately estimate the local distribution. If the image resolution is good enough and the local neighborhood size is large enough, this is not a problem. For example, sampling pixels within a radius of 3γ on the test images in Fig. 2.7 will utilize approximately 1600 samples when $\gamma = 5$mm. For the local neighborhoods of width $\gamma = 1$mm, approximately 65 samples are drawn, and this number of samples still appears to yield local distributions that exhibit intuitive behavior as the neighborhoods are translated. Caution should be exercised when the desired local neighborhood width approaches the same order as the width of a pixel, to make sure that the resulting local distribution estimates are meaningful.

2.3.2.1 Constancy Assumptions

We have already seen that one way to compute the local mean of an image I is by (2.58). Alternatively, if the local probability distribution $p_I(i;\mathbf{x}_0)$ is available, the local mean can be estimated by:

$$\mathrm{E}_\gamma[I;\mathbf{x}_0] = \int_{-\infty}^{\infty} i\, p^{\gamma}_I(i;\mathbf{x}_0)\, di. \tag{2.88}$$

This suggests that (2.63)–(2.65) can be alternatively expressed as:

$$\mathcal{I}^{\ell,\gamma}_{\Psi,r}(R,F^{\mathbf{u}};\mathbf{x}_0) = \int_{-\infty}^{\infty} i\, p^{\gamma}_{\Psi(r(\mathbf{x};R,F^{\mathbf{u}}))}(i;\mathbf{x}_0)\, di, \tag{2.89}$$

$$MSD_\gamma(R,F^{\mathbf{u}};\mathbf{x}_0) = \int_{-\infty}^{\infty} i\, p^{\gamma}_{(F^{\mathbf{u}}-R)^2}(i;\mathbf{x}_0)\, di, \tag{2.90}$$

$$MAD_\gamma(R,F^{\mathbf{u}};\mathbf{x}_0) = \int_{-\infty}^{\infty} i\, p^{\gamma}_{|F^{\mathbf{u}}-R|}(i;\mathbf{x}_0)\, di, \tag{2.91}$$

and these can be substituted in to (2.66)–(2.68) to form the aggregate local dissimilarity measures.

2.3.2.2 Linearity Assumptions

Local versions of MSVDL and MSODL, in addition to being described by (2.71) and (2.72), can be expressed in terms of local probability distributions by:

$$MSVDL_\gamma(R, F^{\mathbf{u}}; \beta, \mathbf{x}_0) = \int_{-\infty}^{\infty} i\, p^\gamma_{(\beta_1(\mathbf{x}_0)R+\beta_2(\mathbf{x}_0)-F^{\mathbf{u}})^2}(i; \mathbf{x}_0)\, di, \tag{2.92}$$

$$MSODL_\gamma(R, F^{\mathbf{u}}; \beta, \mathbf{x}_0) = \int_{-\infty}^{\infty} i\, p^\gamma_{\frac{(\beta_1(\mathbf{x}_0)R+\beta_2(\mathbf{x}_0)-F^{\mathbf{u}})^2}{(1+\beta_1(\mathbf{x}_0)^2)}}(i; \mathbf{x}_0)\, di. \tag{2.93}$$

To construct local versions of SCC and SOCC in a similar manner, estimates of the local joint distribution of two images is required. The local covariance between two images can be computed from the local joint distribution by:

$$\mathrm{Cov}_\gamma(I, J; \mathbf{x}_0) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} ij\, p^\gamma_{I,J}(i, j; \mathbf{x}_0)\, didj - \mathrm{E}_\gamma[I; \mathbf{x}_0] \cdot \mathrm{E}_\gamma[J; \mathbf{x}_0]\,. \tag{2.94}$$

Local variance can be computed from the local distribution by:

$$\mathrm{Var}_\gamma(I; \mathbf{x}_0) = \int_{-\infty}^{\infty} ij\, p^\gamma_I(i; \mathbf{x}_0)\, di - \mathrm{E}_\gamma[I; \mathbf{x}_0]^2\,. \tag{2.95}$$

(2.94) and (2.95) can be substituted into (2.79)–(2.82) in order to form local versions of SCC and SOCC. Note that the local version of SCC constructed using (2.94), (2.95), (2.79) and (2.81) is equivalent to the measure Hermosillo refers to as local cross correlation (CC) [50].

2.3.2.3 Functional and Information Content Assumptions

Hermosillo presented local versions of CR and MI in [50]. In the notation used here, the local CR evaluated at $\mathbf{x}_0$ would be given by:

$$CR_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0) = \frac{\mathrm{Var}_\gamma\left(\mathrm{E}_\gamma[(F^{\mathbf{u}}|R)\,; \mathbf{x}_0]\,; \mathbf{x}_0\right)}{\mathrm{Var}_\gamma(F^{\mathbf{u}}; \mathbf{x}_0)}, \tag{2.96}$$

where

$$\mathrm{E}_\gamma[(J|I=i)\,; \mathbf{x}_0] = \int_{-\infty}^{\infty} j\, p^\gamma_{J|I}((j|i)\,; \mathbf{x}_0)\, dj \quad \text{and} \tag{2.97}$$

$$p^\gamma_{J|I}((j|i)\,; \mathbf{x}_0) = \frac{p^\gamma_{I,J}(i, j; \mathbf{x}_0)}{p^\gamma_I(i; \mathbf{x}_0)}. \tag{2.98}$$

Figure 2.14 illustrates examples of local CR for different values of γ. An important observation that can be made from this figure is that unlike the local versions of SCC and SOCC,

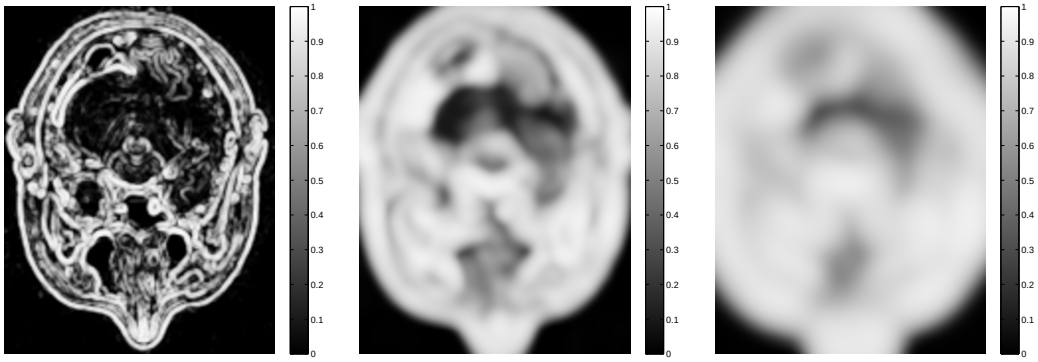


Figure 2.14: Examples of $CR_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)$ for $\gamma = 1\,\text{mm}$ (left), $\gamma = 5\,\text{mm}$ (middle), and $\gamma = 10\,\text{mm}$ (right).

local CR seems to have much smoother behavior in the background region at the smaller scale as well as smoother transitions in the medium and larger scales. This suggests that the use of local CR may provide a more robust approach to registration than local SCC or SOCC.

The local MI evaluated at $\mathbf{x}_0$ would be given by:

$$MI_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p^\gamma_{R,F^{\mathbf{u}}}(i, j; \mathbf{x}_0) \log \frac{p^\gamma_{R,F^{\mathbf{u}}}(i, j; \mathbf{x}_0)}{p^\gamma_R(i; \mathbf{x}_0)\, p^\gamma_{F^{\mathbf{u}}}(j; \mathbf{x}_0)} didj, \tag{2.99}$$

although it could be expressed alternatively using local entropy:

$$MI_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0) = H_\gamma(R; \mathbf{x}_0) + H_\gamma(F^{\mathbf{u}}; \mathbf{x}_0) - H_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)\,, \tag{2.100}$$

where

$$H_\gamma(A; \mathbf{x}_0) = -\int_{-\infty}^{\infty} p^\gamma_A(a; \mathbf{x}_0) \log p^\gamma_A(a; \mathbf{x}_0)\, da \tag{2.101}$$

$$H_\gamma(A, B; \mathbf{x}_0) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p^\gamma_{A,B}(a, b; \mathbf{x}_0) \log p^\gamma_{A,B}(a, b; \mathbf{x}_0)\, dadb. \tag{2.102}$$

Following (2.52) and (2.53), local versions of NMI and ECC about a point $\mathbf{x}_0$ are given by:

$$NMI_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0) := \frac{H_\gamma(R; \mathbf{x}_0) + H_\gamma(F^{\mathbf{u}}; \mathbf{x}_0)}{H_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)}, \tag{2.103}$$

$$ECC_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0) := \sqrt{2 - \frac{2H_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)}{H_\gamma(R; \mathbf{x}_0) + H_\gamma(F^{\mathbf{u}}; \mathbf{x}_0)}}. \tag{2.104}$$

Local versions of the CRE based dissimilarity measures can be defined in a similar manner:

$$CCRE_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0) := \varepsilon_\gamma(R; \mathbf{x}_0) - \mathrm{E}_\gamma[\varepsilon_\gamma((R|F^{\mathbf{u}})\,; \mathbf{x}_0)\,; \mathbf{x}_0]\,, \tag{2.105}$$

$$SCCRE_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0) := \frac{1}{2}\Big(CCRE_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0) + CCRE_\gamma(F^{\mathbf{u}}, R; \mathbf{x}_0)\Big), \tag{2.106}$$

$$NCCRE_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0) := \frac{2}{2 - CRECC_\gamma(R, F^{\mathbf{u}}, \mathbf{x}_0)^2}, \tag{2.107}$$

$$CRECC_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0) := \sqrt{\frac{2SCCRE_\gamma(R, F^{\mathbf{u}}, \mathbf{x}_0)}{\varepsilon_\gamma(R; \mathbf{x}_0) + \varepsilon_\gamma(F^{\mathbf{u}}; \mathbf{x}_0)}}, \tag{2.108}$$

where

$$\varepsilon_\gamma(A; \mathbf{x}_0) = -\int_{-\infty}^{\infty} P^\gamma(A > a; \mathbf{x}_0)\, \log P^\gamma(A > a; \mathbf{x}_0)\, da, \tag{2.109}$$

$$\varepsilon_\gamma((B|A)\,; \mathbf{x}_0) = -\int_{-\infty}^{\infty} \begin{pmatrix} P^\gamma((B > b \,|\, A)\,; \mathbf{x}_0) \\ \cdot \log P^\gamma((B > b \,|\, A)\,; \mathbf{x}_0) \end{pmatrix} db, \tag{2.110}$$

$$P^\gamma(A > a; \mathbf{x}_0) = \int_a^{\infty}\int_{-\infty}^{\infty} p^\gamma_{A,B}(t, b; \mathbf{x}_0)\, dbdt, \tag{2.111}$$

$$P^\gamma((B > b|A = a)\,; \mathbf{x}_0) = \frac{\int_b^{\infty} p^\gamma_{A,B}(a, t; \mathbf{x}_0)\, dt}{p^\gamma_A(a; \mathbf{x}_0)}. \tag{2.112}$$

Figures 2.15–2.20 illustrate examples of local MI, NMI, ECC, SCCRE, NCCRE, and CRECC for different values of γ. Like the local CR, the local versions of these similarity measures all have much smoother behavior in the background regions than local SCC or local SOCC. Moving from left to right in each figure shows a significant amount of smoothing, indicating relatively smooth behavior as small perturbations are made to the locations of neighborhoods. This would indicate that these local similarity measures would be useful in a multiscale strategy for registration in order to improve the ability to converge from larger initial misregistrations. It is difficult to make any observations contrasting the local MI from local versions of NMI, SCCRE, and NCCRE. However, it can be observed that local ECC and local CRECC contain more non-edge information than local MI, NMI, SCCRE, and NCCRE. This is due to the square root factor in the formulation of ECC and CRECC that causes edges to be amplified to a lesser degree than in the other information content based dissimilarity measures.

An aggregate version of the local MI (2.99) can be formed by integrating over the entire

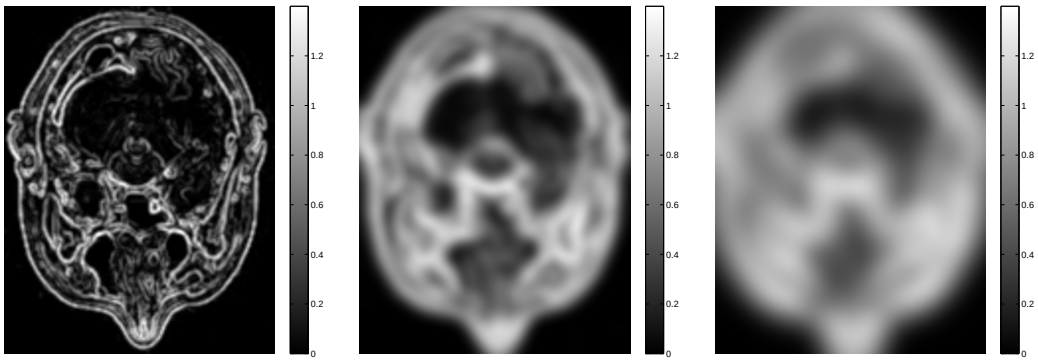


Figure 2.15: Examples of $MI_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)$ for $\gamma = 1\,\text{mm}$ (left), $\gamma = 5\,\text{mm}$ (middle), and $\gamma = 10\,\text{mm}$ (right).

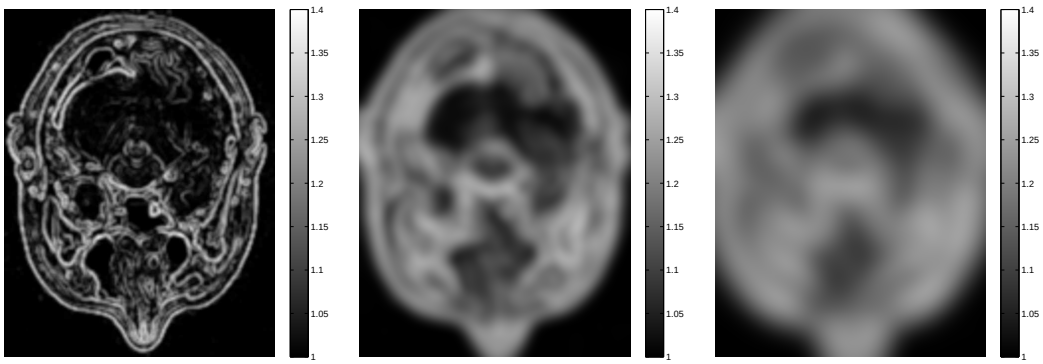


Figure 2.16: Examples of $NMI_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)$ for $\gamma = 1\,\text{mm}$ (left), $\gamma = 5\,\text{mm}$ (middle), and $\gamma = 10\,\text{mm}$ (right).

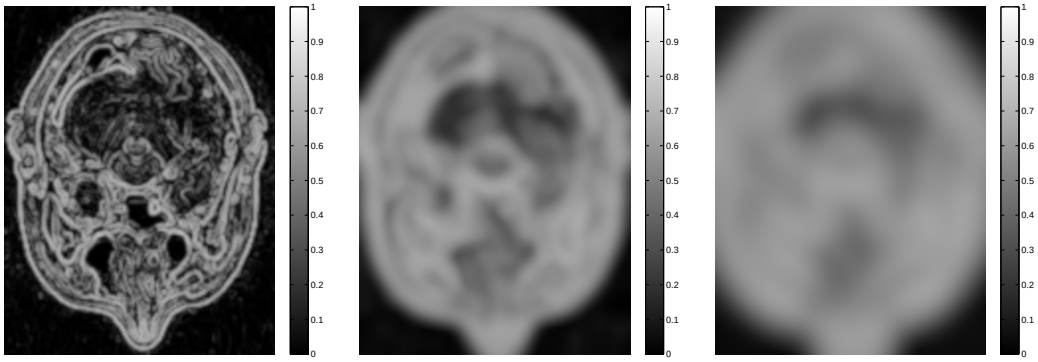


Figure 2.17: Examples of $ECC_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)$ for $\gamma = 1\,\text{mm}$ (left), $\gamma = 5\,\text{mm}$ (middle), and $\gamma = 10\,\text{mm}$ (right).

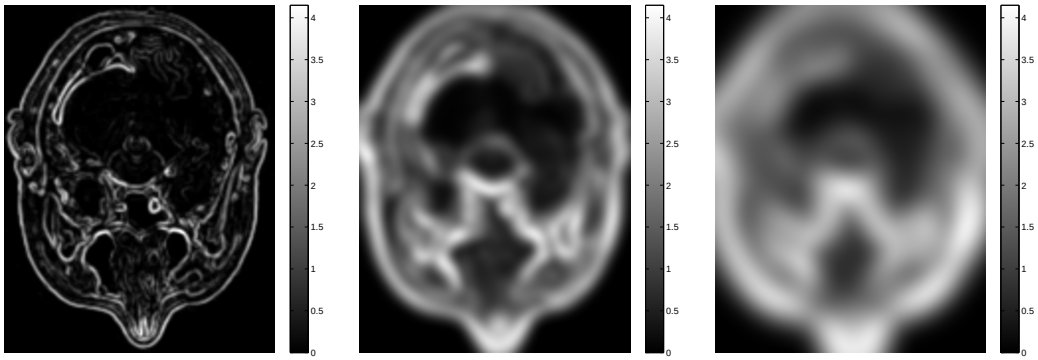


Figure 2.18: Examples of $SCCRE_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)$ for $\gamma = 1\,\text{mm}$ (left), $\gamma = 5\,\text{mm}$ (middle), and $\gamma = 10\,\text{mm}$ (right).

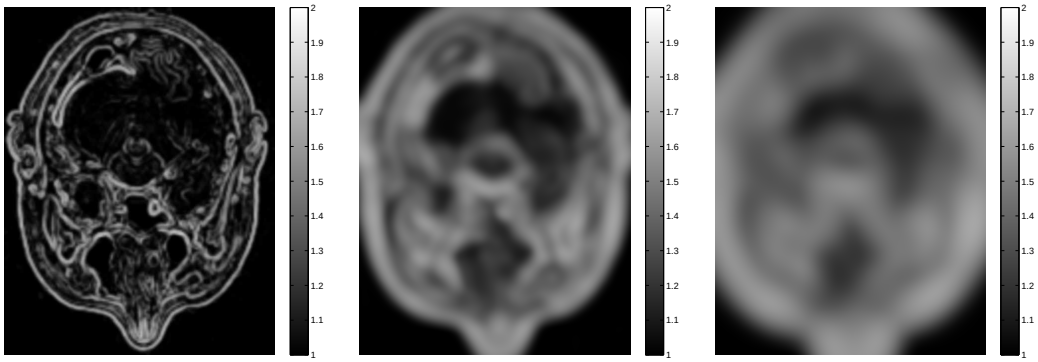


Figure 2.19: Examples of $NCCRE_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)$ for $\gamma = 1\,\text{mm}$ (left), $\gamma = 5\,\text{mm}$ (middle), and $\gamma = 10\,\text{mm}$ (right).

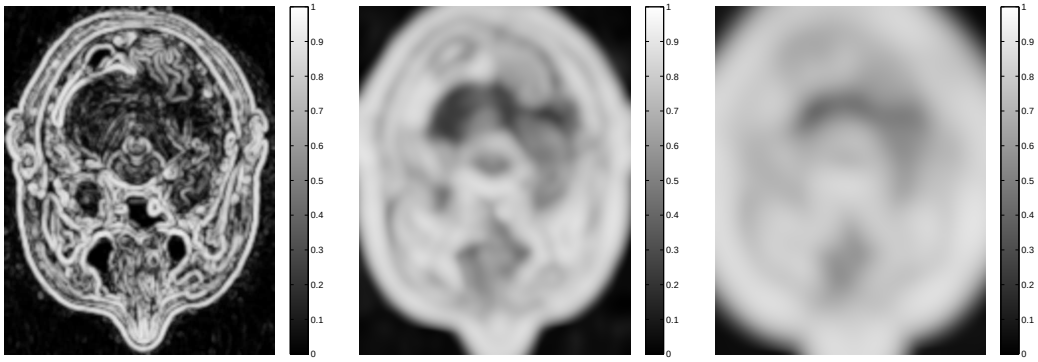


Figure 2.20: Examples of $CRECC_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)$ for $\gamma = 1\,\text{mm}$ (left), $\gamma = 5\,\text{mm}$ (middle), and $\gamma = 10\,\text{mm}$ (right).

image domain:

$$MI_\gamma(R, F^{\mathbf{u}}) = \int_\Omega MI_\gamma(R, F^{\mathbf{u}}, \mathbf{x}_0)\, d\mathbf{x}_0. \tag{2.113}$$

Aggregate versions of the local NMI, ECC, CCRE, SCCRE, NCCRE, and CRECC can be formed in a similar manner.

2.4 Gâteaux Derivatives

To find the optimal deformation field relating a pair of images, it is necessary to determine the variational (or Gâteaux) derivative of the dissimilarity measure with respect to variations in the displacement field. For a general dissimilarity measure of the form $\mathcal{I}(R, F^{\mathbf{u}})$, the Gâteaux derivative is defined in the direction $\mathbf{w}$ by:

$$d\mathcal{I}(R, F^{\mathbf{u}}; \mathbf{w}) = \lim_{h \to 0} \frac{1}{h} \left[\mathcal{I}\Big(R, F^{\mathbf{u}+h\mathbf{w}}\Big) - \mathcal{I}(R, F^{\mathbf{u}}) \right]. \tag{2.114}$$

From (2.114), it is possible to derive the following alternate form of the Gâteaux derivative (see Appendix B for a proof):

$$d\mathcal{I}(R, F^{\mathbf{u}}; \mathbf{w}) = \int_\Omega \left\langle \mathcal{P}(\mathbf{x}; R, F^{\mathbf{u}}), \mathbf{w}(\mathbf{x}) \right\rangle d\mathbf{x}, \tag{2.115}$$

where $\langle \cdot, \cdot \rangle$ indicates the inner product.

For local dissimilarity measures of the form:

$$\mathcal{I}^{\ell,\gamma}(R, F^{\mathbf{u}}) = \int_\Omega \mathcal{I}^{\ell,\gamma}(R, F^{\mathbf{u}}, \mathbf{x}_0)\, d\mathbf{x}, \tag{2.116}$$

the Gâteaux derivative can be expressed by:

$$d\mathcal{I}^{\ell,\gamma}(R, F^{\mathbf{u}}; \mathbf{w}) = \int_\Omega \left\langle \mathcal{P}^{\ell,\gamma}(\mathbf{x}; R, F^{\mathbf{u}}), \mathbf{w}(\mathbf{x}) \right\rangle d\mathbf{x}, \tag{2.117}$$

where $\mathcal{P}^{\ell,\gamma}(\mathbf{x}; R, F^{\mathbf{u}})$ is given by:

$$\mathcal{P}^{\ell,\gamma}(\mathbf{x}; R, F^{\mathbf{u}}) = \int_\Omega \mathcal{P}^{\ell,\gamma}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0)\, d\mathbf{x}_0. \tag{2.118}$$

In this section, we will present the quantities $\mathcal{P}^{g}(\mathbf{x}; R, F^{\mathbf{u}})$ and $\mathcal{P}^{\ell,\gamma}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0)$ that are needed to define the Gateaux derivatives of the global and local versions of each dissimilarity measure according to (2.115) and (2.117)–(2.118).

2.4.1 Constancy Assumptions

For the general dissimilarity measure $\mathcal{I}_{\Psi,r}$, it can be shown that:

$$\mathcal{P}_{\Psi,r}(\mathbf{x}; R, F^{\mathbf{u}}) = -\frac{1}{|\Omega|}\left[\frac{\partial}{\partial F}\Psi(r(\mathbf{x}; R, F^{\mathbf{u}}))\right]\nabla F^{\mathbf{u}}(\mathbf{x}) \tag{2.119}$$

(see (B.3) in Appendix B for details). In general, r can be a scalar, vector, or matrix function of R and $F^{\mathbf{u}}$. If we assume that r is sampled on a discrete grid in $\mathbb{R}^2$, then the chain rule implies:

$$\mathcal{P}_{\Psi,r}(\mathbf{x}; R, F^{\mathbf{u}}) = -\frac{1}{M}\sum_{i,j}\frac{d}{dr_{i,j}}\Psi(r(\mathbf{x}; R, F^{\mathbf{u}}))\,\frac{\partial}{\partial F}r_{i,j}(\mathbf{x}; R, F^{\mathbf{u}})\,\nabla F^{\mathbf{u}}(\mathbf{x}), \tag{2.120}$$

where M is the total number of grid points. Similar expressions can be derived for higher dimensions. All of the residual images in Table 2.1 except for the one based on brightness morphology can be written in terms of the difference of two functions; i.e.,

$$r(\mathbf{x}; R, F^{\mathbf{u}}) = \hat{r}(F^{\mathbf{u}}(\mathbf{x})) - \hat{r}(R(\mathbf{x}))\,. \tag{2.121}$$

Therefore, for these cases, we can write

$$\mathcal{P}_{\Psi,r}(\mathbf{x}; R, F^{\mathbf{u}}) = -\frac{1}{M}\sum_{i,j}\frac{d}{dr_{i,j}}\Psi(r(\mathbf{x}; R, F^{\mathbf{u}}))\,\frac{\partial}{\partial F}\hat{r}_{i,j}(F^{\mathbf{u}}(\mathbf{x}))\,\nabla F^{\mathbf{u}}(\mathbf{x}), \tag{2.122}$$

which reduces via the chain rule to:

$$\mathcal{P}_{\Psi,r}(\mathbf{x}; R, F^{\mathbf{u}}) = -\frac{1}{M}\sum_{i,j}\frac{d}{dr_{i,j}}\Psi(r(\mathbf{x}; R, F^{\mathbf{u}}))\,\nabla\hat{r}_{i,j}(F^{\mathbf{u}}(\mathbf{x})). \tag{2.123}$$

For the brightness morphology constancy constraint, the residual image r_{BM} cannot be expressed in terms of (2.121), but rather in terms of norms and inner products:

$$r_{BM}(\mathbf{x}; R, F^{\mathbf{u}}) = \|\mathbf{n}(F^{\mathbf{u}}(\mathbf{x}))\|^2\,\|\mathbf{n}(R(\mathbf{x}))\|^2 - \langle\mathbf{n}(F^{\mathbf{u}}(\mathbf{x}))\,,\mathbf{n}(R(\mathbf{x}))\rangle^2\,, \tag{2.124}$$

where $\mathbf{n} := \mathbf{n}_\epsilon$ is defined by (2.18). For this specific case, we can apply the chain rule to find:

$$\begin{aligned}\mathcal{P}_{\Psi,r_{BM}}(\mathbf{x}; R, F^{\mathbf{u}}) \;=\; & -\frac{2}{|\Omega|}\frac{d}{dr}\Psi(r_{BM}(\mathbf{x}; R, F^{\mathbf{u}}))\,[\nabla\mathbf{n}(F^{\mathbf{u}}(\mathbf{x}))] \\ & \cdot\Big(\,\|\mathbf{n}(R(\mathbf{x}))\|^2\,\mathbf{n}(F^{\mathbf{u}}(\mathbf{x})) \\ & \quad - \langle\mathbf{n}(F^{\mathbf{u}}(\mathbf{x}))\,,\mathbf{n}(R(\mathbf{x}))\rangle\,\mathbf{n}(R(\mathbf{x}))\,\Big),\end{aligned} \tag{2.125}$$

Constancy Constraint	$\mathcal{P}^g_{\Psi,r}(\mathbf{x}; R, F^{\mathbf{u}})$
Brightness	$-\frac{1}{\lvert\Omega\rvert}\left[\nabla F^{\mathbf{u}}\right]\frac{d}{dr}\Psi(r(\mathbf{x}; R, F, \mathbf{u}))$
Brightness Gradient	$-\frac{1}{\lvert\Omega\rvert}\left[\nabla\nabla^{\mathrm{T}} F^{\mathbf{u}}\right]\frac{d}{dr}\Psi(r(\mathbf{x}; R, F, \mathbf{u}))$
Brightness Hessian	$-\frac{1}{M}\sum_{i,j}\left[\frac{\partial}{\partial x_{i,j}}\nabla\nabla^{\mathrm{T}} F^{\mathbf{u}}\right]\frac{\partial}{\partial r_{i,j}}\Psi(r(\mathbf{x}; R, F, \mathbf{u}))$
Brightness Gradient Magnitude	$-\frac{2}{\lvert\Omega\rvert}\left[\nabla\nabla^{\mathrm{T}} F^{\mathbf{u}}\right]\left[\nabla F^{\mathbf{u}}\right]\frac{d}{dr}\Psi(r(\mathbf{x}; R, F, \mathbf{u}))$
Brightness Hessian Determinant	$-\frac{1}{\lvert\Omega\rvert}\left[\nabla\left\lvert\nabla\nabla^{\mathrm{T}} F^{\mathbf{u}}\right\rvert\right]\frac{d}{dr}\Psi(r(\mathbf{x}; R, F, \mathbf{u}))$
Brightness Laplacian	$-\frac{1}{\lvert\Omega\rvert}\left[\nabla\Delta F^{\mathbf{u}}\right]\frac{d}{dr}\Psi(r(\mathbf{x}; R, F, \mathbf{u}))$
Brightness Morphology	$-\frac{2}{\lvert\Omega\rvert}\frac{d}{dr}\Psi(r_{BM}(\mathbf{x}; R, F^{\mathbf{u}}))\left[\nabla\mathbf{n}(F^{\mathbf{u}}(\mathbf{x}))\right]$ $\cdot\Big(\left\lVert\mathbf{n}(R(\mathbf{x}))\right\rVert^2\mathbf{n}(F^{\mathbf{u}}(\mathbf{x}))$ $-\left\langle\mathbf{n}(F^{\mathbf{u}}(\mathbf{x})), \mathbf{n}(R(\mathbf{x}))\right\rangle\mathbf{n}(R(\mathbf{x}))\Big)$

Table 2.3: Values of $\mathcal{P}^g_{\Psi,r}$ for specific residual images.

where

$$\nabla\mathbf{n}(F^{\mathbf{u}}(\mathbf{x})) = \nabla\nabla^{\mathrm{T}} F^{\mathbf{u}}(\mathbf{x}) \cdot \frac{\left(\left\lVert\nabla F^{\mathbf{u}}(\mathbf{x})\right\rVert^2 + \epsilon^2\right)\mathbf{I} - \left(\nabla F^{\mathbf{u}}(\mathbf{x})\right)\left(\nabla^{\mathrm{T}} F^{\mathbf{u}}(\mathbf{x})\right)}{\left(\left\lVert\nabla F^{\mathbf{u}}(\mathbf{x})\right\rVert^2 + \epsilon^2\right)^{\frac{3}{2}}}, \tag{2.126}$$

and where $\mathbf{I}$ is the identity matrix.

Given (2.123) and (2.125), we can state $\mathcal{P}^g$ for any combination of constancy constraint and penalty function. Table 2.3 shows $\mathcal{P}^g$ for each type of constancy constraint. Note that for the brightness morphology constancy constraint, we have assumed the use of the regularized vector (2.18), so that $\mathcal{P}$ is differentiable when ∇F or ∇R are equal to zero. For each $\mathcal{P}^g$ in Table 2.3, the derivative of Ψ with respect to r is required. Table 2.4 lists these values for each of the penalty functions listed in Table 2.2.

Tables 2.3 and 2.4 allow us to state the Gâteaux derivatives of some specific global dissimilarity measures. For the MSD, we have:

$$\mathcal{P}_{MSD}(\mathbf{x}; R, F^{\mathbf{u}}) = \frac{2}{\lvert\Omega\rvert}\left(R(\mathbf{x}) - F^{\mathbf{u}}(\mathbf{x})\right)\nabla F^{\mathbf{u}}(\mathbf{x})\,. \tag{2.127}$$

For the MAD, $\Psi_1(r)$ is not differentiable at the origin. The modified L^1-norm, which is

Penalty Function	$\frac{d}{dr}\Psi(r)$
L^p-norm (scalar r)	$p\, r^{p-1}\,\mathrm{sign}\,(r)$
L^p-norm (vector r)	$\begin{pmatrix} p\, r_1^{p-1}\,\mathrm{sign}\,(r_1) \\ \vdots \\ p\, r_n^{p-1}\,\mathrm{sign}\,(r_n) \end{pmatrix}$
L^p-norm (matrix r)	$\begin{pmatrix} p\, r_{1,1}^{p-1}\,\mathrm{sign}\,(r_{1,1}) & \cdots & p\, r_{1,n}^{p-1}\,\mathrm{sign}\,(r_{1,n}) \\ \vdots & \ddots & \vdots \\ p\, r_{n,1}^{p-1}\,\mathrm{sign}\,(r_{n,1}) & \cdots & p\, r_{n,n}^{p-1}\,\mathrm{sign}\,(r_{n,n}) \end{pmatrix}$
Modified L^1-norm (scalar r)	$r\left(r^2+\epsilon^2\right)^{-\frac{1}{2}}$
Modified L^1-norm (vector/matrix r)	$r\left(\|r\|_2^2+\epsilon^2\right)^{-\frac{1}{2}}$
Huber	$\begin{cases} -k, & r < -k \\ r, & -k \le r < k \\ k, & r \ge k \end{cases}$
Tukey	$\begin{cases} r\left[1-\left(\frac{r}{c}\right)^2\right]^2, & \|r\| < c \\ 0, & \|r\| \ge c \end{cases}$
Geman-McClure	$2r\sigma^2\left(r^2+\sigma^2\right)^{-2}$
Lorentizian	$2r\left(r^2+2\sigma^2\right)^{-1}$

Table 2.4: Derivatives of penalty functions.

differentiable everywhere, can be applied to r_{BC} in order to approximate the MAD when variational methods are to be used for nonrigid registration.

Gâteaux derivatives of local dissimilarity measures are simple extensions of the Gâteaux derivatives of the global dissimilarity measures. In light of (2.58) and (2.63), we can state:

$$\mathcal{P}^{\ell,\gamma}_{\Psi,r}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0) = w_\gamma(\mathbf{x} - \mathbf{x}_0) \left[\mathcal{G}_\gamma(\mathbf{x}_0)\right]^{-1} \mathcal{P}^{g}_{\Psi,r}(\mathbf{x}; R, F^{\mathbf{u}}) \,, \tag{2.128}$$

where

$$\mathcal{G}_\gamma(\mathbf{x}_0) = \int_\Omega w_\gamma(\mathbf{t} - \mathbf{x}_0)\, d\mathbf{t}. \tag{2.129}$$

This result can then be substituted into (2.118) to form $\mathcal{P}^{\ell,\gamma}_{\Psi,r}(\mathbf{x}; R, F^{\mathbf{u}})$.

2.4.2 Linearity Assumptions

In order to compute the Gâteaux derivatives of the MSVDL and MSODL dissimilarity measures with respect to the displacement, it is helpful to think of the linearity parameters β_1 and β_2 as fixed. (This is valid during Step 3 of Algorithm 2.2.2, for example.) For fixed β, MSVDL and MSODL can actually be thought of as dissimilarity measures that are based on constancy assumptions. Therefore, $\mathcal{P}^{g}_{MSVDL}$ and $\mathcal{P}^{g}_{MSODL}$ can be derived in the same manner as $\mathcal{P}^{g}_{\Psi,r}$ in the previous section. We have:

$$\mathcal{P}^{g}_{MSVDL}(\mathbf{x}; R, F^{\mathbf{u}}) = \frac{2}{|\Omega|} \left(\beta_1 R(\mathbf{x}) + \beta_2 - F^{\mathbf{u}}(\mathbf{x})\right) \nabla F^{\mathbf{u}}(\mathbf{x}) \,, \tag{2.130}$$

$$\mathcal{P}^{g}_{MSODL}(\mathbf{x}; R, F^{\mathbf{u}}) = \frac{2\left(\beta_1 R(\mathbf{x}) + \beta_2 - F^{\mathbf{u}}(\mathbf{x})\right)}{|\Omega|\left(1 + \beta_1^2\right)} \nabla F^{\mathbf{u}}(\mathbf{x}) \,. \tag{2.131}$$

Using (2.128), we see that for local versions of MSVDL and MSODL,

$$\mathcal{P}^{\ell,\gamma}_{MSVDL}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0) = \frac{2}{|\Omega|} w_\gamma(\mathbf{x} - \mathbf{x}_0) \left[\mathcal{G}_\gamma(\mathbf{x}_0)\right]^{-1} \left(\beta_1 R(\mathbf{x}) + \beta_2 - F^{\mathbf{u}}(\mathbf{x})\right) \nabla F^{\mathbf{u}}(\mathbf{x}) \tag{2.132}$$

and

$$\mathcal{P}^{\ell,\gamma}_{MSODL}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0) = \frac{2 w_\gamma(\mathbf{x} - \mathbf{x}_0) \left[\mathcal{G}_\gamma(\mathbf{x}_0)\right]^{-1} \left(\beta_1 R(\mathbf{x}) + \beta_2 - F^{\mathbf{u}}(\mathbf{x})\right)}{|\Omega|\left(1 + \beta_1^2\right)} \nabla F^{\mathbf{u}}(\mathbf{x}) \,. \tag{2.133}$$

The SCC is equivalent to the square of Hermosillo's CC dissimilarity measure. In [50], Hermosillo derives the Gâteaux derivatives for the global and local versions of this measure.

However, he assumes that $\mathcal{I}^g_{SCC}$ is computed using an estimate of the joint probability density function. If, instead, (2.37)–(2.39) are used to compute $\mathcal{I}^g_{SCC}$, then the term $\mathcal{P}^g_{SCC}$ describing the Gâteaux derivative can be written as:

$$\mathcal{P}^g_{SCC}(\mathbf{x}; R, F^{\mathbf{u}}) = -\frac{2}{|\Omega|}\mathcal{K}^g_{SCC}(\mathbf{x}; R, F^{\mathbf{u}})\,\nabla F^{\mathbf{u}}(\mathbf{x})\,, \tag{2.134}$$

where

$$\begin{aligned}\mathcal{K}^g_{SCC}(\mathbf{x}; R, F^{\mathbf{u}}) &= \frac{\mathrm{Cov}\,(R, F^{\mathbf{u}})}{\mathrm{Var}\,(R)\,\mathrm{Var}\,(F^{\mathbf{u}})}\left(R(\mathbf{x}) - \bar{R}\right)\\ &- \frac{SCC(R, F^{\mathbf{u}})}{\mathrm{Var}\,(F^{\mathbf{u}})}\left(F^{\mathbf{u}}(\mathbf{x}) - \bar{F^{\mathbf{u}}}\right).\end{aligned} \tag{2.135}$$

The Gâteaux derivative of SOCC can be determined in a similar manner. We have:

$$\mathcal{P}^g_{SOCC}(\mathbf{x}; R, F^{\mathbf{u}}) = -\frac{2}{|\Omega|}\mathcal{K}^g_{SOCC}(\mathbf{x}; R, F^{\mathbf{u}})\,\nabla F^{\mathbf{u}}(\mathbf{x})\,, \tag{2.136}$$

where

$$\begin{aligned}\mathcal{K}^g_{SOCC}(\mathbf{x}; R, F^{\mathbf{u}}) &= \frac{4\mathrm{Cov}\,(R, F^{\mathbf{u}})\left(R(\mathbf{x}) - \bar{R}\right)}{\left(\mathrm{Var}\,(R) + \mathrm{Var}\,(F^{\mathbf{u}})\right)^2}\\ &- \frac{2\left(\mathrm{Var}\,(R) - \mathrm{Var}\,(F^{\mathbf{u}})\right)\left(F^{\mathbf{u}}(\mathbf{x}) - \bar{F^{\mathbf{u}}}\right)}{\left(\mathrm{Var}\,(R) + \mathrm{Var}\,(F^{\mathbf{u}})\right)^2}\\ &- \frac{2\,SOCC(R, F^{\mathbf{u}})}{\mathrm{Var}\,(R) + \mathrm{Var}\,(F^{\mathbf{u}})}\left(F^{\mathbf{u}}(\mathbf{x}) - \bar{F^{\mathbf{u}}}\right).\end{aligned} \tag{2.137}$$

For local versions of SCC and SOCC, we have:

$$\mathcal{P}^{\ell,\gamma}_{SCC}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0) = -\frac{2}{|\Omega|}\mathcal{K}^{\ell,\gamma}_{SCC}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0)\,\nabla F^{\mathbf{u}}(\mathbf{x})\,, \tag{2.138}$$

$$\mathcal{P}^{\ell,\gamma}_{SOCC}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0) = -\frac{2}{|\Omega|}\mathcal{K}^{\ell,\gamma}_{SOCC}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0)\,\nabla F^{\mathbf{u}}(\mathbf{x})\,, \tag{2.139}$$

where

$$\begin{aligned}\mathcal{K}^{\ell,\gamma}_{SCC}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0) &= \frac{\mathrm{Cov}_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)}{\mathrm{Var}_\gamma(R; \mathbf{x}_0)\,\mathrm{Var}_\gamma(F^{\mathbf{u}}; \mathbf{x}_0)}\left(R(\mathbf{x}) - \mathrm{E}_\gamma[R; \mathbf{x}_0]\right)\\ &- \frac{SCC_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)}{\mathrm{Var}_\gamma(F^{\mathbf{u}}; \mathbf{x}_0)}\left(F^{\mathbf{u}}(\mathbf{x}) - \mathrm{E}_\gamma[F^{\mathbf{u}}; \mathbf{x}_0]\right)\end{aligned} \tag{2.140}$$

and

$$\begin{aligned}\mathcal{K}^{\ell,\gamma}_{SOCC}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0) &= \frac{4\mathrm{Cov}_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)\left(R(\mathbf{x}) - \mathrm{E}_\gamma[R; \mathbf{x}_0]\right)}{\left(\mathrm{Var}_\gamma(R; \mathbf{x}_0) + \mathrm{Var}_\gamma(F^{\mathbf{u}}; \mathbf{x}_0)\right)^2}\\ &- \frac{2\left(\mathrm{Var}_\gamma(R; \mathbf{x}_0) - \mathrm{Var}_\gamma(F^{\mathbf{u}}; \mathbf{x}_0)\right)\left(F^{\mathbf{u}}(\mathbf{x}) - \mathrm{E}_\gamma[F^{\mathbf{u}}; \mathbf{x}_0]\right)}{\left(\mathrm{Var}_\gamma(R; \mathbf{x}_0) + \mathrm{Var}_\gamma(F^{\mathbf{u}}; \mathbf{x}_0)\right)^2}\\ &- \frac{2\,SOCC_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)\left(F^{\mathbf{u}}(\mathbf{x}) - \mathrm{E}_\gamma[F^{\mathbf{u}}; \mathbf{x}_0]\right)}{\mathrm{Var}_\gamma(R; \mathbf{x}_0) + \mathrm{Var}_\gamma(F^{\mathbf{u}}; \mathbf{x}_0)}.\end{aligned} \tag{2.141}$$

$\mathcal{P}^{\ell,\gamma}_{SCC}(\mathbf{x}; R, F^{\mathbf{u}})$ and $\mathcal{P}^{\ell,\gamma}_{SOCC}(\mathbf{x}; R, F^{\mathbf{u}})$ can then be formed by integrating (2.138) and (2.139), respectively, according to (2.118). Note that the use of Gâteaux derivatives that are constructed using local moments enables the use of much smaller neighborhoods than when using local distributions. This is because much fewer samples are required to estimate a mean or variance than to estimate an entire distribution. This provides a practical improvement over Hermosillo's formulation of local dissimilarity measures based on distributions.

2.4.3 Functional Assumptions

The Gâteaux derivatives of the global and local versions of CR are derived in [50]. In the notation of this thesis, $\mathcal{P}^{g}_{CR}(\mathbf{x}; R, F^{\mathbf{u}})$ is given by:

$$\mathcal{P}^{g}_{CR}(\mathbf{x}; R, F^{\mathbf{u}}) = -\frac{2}{|\Omega|}\mathcal{K}^{g}_{CR}(\mathbf{x}; R, F^{\mathbf{u}})\,\nabla F^{\mathbf{u}}(\mathbf{x})\,, \tag{2.142}$$

where

$$\mathcal{K}^{g}_{CR}(\mathbf{x}; R, F^{\mathbf{u}}) = \frac{\mathrm{E}[F^{\mathbf{u}}|R\,(\mathbf{x})] - \mathrm{E}[F^{\mathbf{u}}]}{\mathrm{Var}(F^{\mathbf{u}})} - \frac{CR(R, F^{\mathbf{u}})}{\mathrm{Var}(F^{\mathbf{u}})}\,(F^{\mathbf{u}}(\mathbf{x}) - \mathrm{E}[F^{\mathbf{u}}])\,. \tag{2.143}$$

For the local CR, $\mathcal{P}^{\ell,\gamma}_{CR}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0)$ is given by:

$$\mathcal{P}^{\ell,\gamma}_{CR}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0) = -\frac{2}{|\Omega|}\mathcal{K}^{\ell,\gamma}_{CR}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0)\,\nabla F^{\mathbf{u}}(\mathbf{x})\,, \tag{2.144}$$

where

$$\begin{aligned}\mathcal{K}^{\ell,\gamma}_{CR}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0) \;=\; & \frac{\mathrm{E}_\gamma[(F^{\mathbf{u}}|R\,(\mathbf{x}))\,; \mathbf{x}_0] - \mathrm{E}_\gamma[F^{\mathbf{u}}; \mathbf{x}_0]}{\mathrm{Var}_\gamma(F^{\mathbf{u}}; \mathbf{x}_0)} \\ & - \frac{CR_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)}{\mathrm{Var}_\gamma(F^{\mathbf{u}}; \mathbf{x}_0)}\,(F^{\mathbf{u}}(\mathbf{x}) - \mathrm{E}_\gamma[F^{\mathbf{u}}; \mathbf{x}_0])\,.\end{aligned} \tag{2.145}$$

$\mathcal{P}^{\ell,\gamma}_{CR}(\mathbf{x}; R, F^{\mathbf{u}})$ can then be formed by integrating (2.144) according to (2.118).

2.4.4 Information Content Assumptions

As shown in Sections 2.2.4 and 2.3.2.3, all of the dissimilarity measures based on information content assumptions can be expressed as functions of entropy, joint entropy, CRE, and

JCRE. The Gâteaux derivatives of these quantities can be written in terms of (2.115), with:

$$\mathcal{P}_H^g(\mathbf{x}; F^{\mathbf{u}}) = \frac{1}{|\Omega|}\left(\frac{p'_{F^{\mathbf{u}}}(F^{\mathbf{u}}(\mathbf{x}))}{p_{F^{\mathbf{u}}}(F^{\mathbf{u}}(\mathbf{x}))}\right)\nabla F^{\mathbf{u}}(\mathbf{x}), \tag{2.146}$$

$$\mathcal{P}_H^g(\mathbf{x}; R, F^{\mathbf{u}}) = \frac{1}{|\Omega|}\left(\frac{\frac{\partial}{\partial f}\left[p_{R,F^{\mathbf{u}}}(R(\mathbf{x}), f)\right]_{f=F^{\mathbf{u}}(\mathbf{x})}}{p_{R,F^{\mathbf{u}}}(R(\mathbf{x}), F^{\mathbf{u}}(\mathbf{x}))}\right)\nabla F^{\mathbf{u}}(\mathbf{x}), \tag{2.147}$$

$$\mathcal{P}_\varepsilon^g(\mathbf{x}; F^{\mathbf{u}}) = \frac{1}{|\Omega|}\left(1 + \log P(F^{\mathbf{u}} > F^{\mathbf{u}}(\mathbf{x}))\right)\nabla F^{\mathbf{u}}(\mathbf{x}), \tag{2.148}$$

$$\mathcal{P}_\varepsilon^g(\mathbf{x}; R, F^{\mathbf{u}}) = \frac{1}{|\Omega|}\left(1 + \log P(F^{\mathbf{u}} > F^{\mathbf{u}}(\mathbf{x}) \,|R = R(\mathbf{x}))\right)\nabla F^{\mathbf{u}}(\mathbf{x}), \tag{2.149}$$

$$\mathcal{P}_\varepsilon^g(\mathbf{x}; F^{\mathbf{u}}, R) = \mathcal{P}_\varepsilon^g(\mathbf{x}; F^{\mathbf{u}}) + \frac{1}{|\Omega|}\mathcal{K}_\varepsilon^g(\mathbf{x}; F^{\mathbf{u}}, R)\nabla F^{\mathbf{u}}(\mathbf{x}), \tag{2.150}$$

where

$$\begin{aligned}\mathcal{K}_\varepsilon^g(\mathbf{x}; F^{\mathbf{u}}, R) = &\int_{-\infty}^{R(\mathbf{x})} \frac{\partial}{\partial f}\left[\log P(R > r|F^{\mathbf{u}} = f)\right]_{f=F^{\mathbf{u}}(\mathbf{x})} dr \\ &+ \int_{-\infty}^{R(\mathbf{x})} \frac{\partial}{\partial f}\left[P(R < r|F^{\mathbf{u}} = f)\right]_{f=F^{\mathbf{u}}(\mathbf{x})} dr \\ &- \int_{R(\mathbf{x})}^{\infty} \frac{\partial}{\partial f}\left[P(R > r|F^{\mathbf{u}} = f)\right]_{f=F^{\mathbf{u}}(\mathbf{x})} dr.\end{aligned} \tag{2.151}$$

Derivations of (2.146)–(2.150) are provided in Appendix B.

The Gâteaux derivatives of the dissimilarity measures based on information content assumptions can be expressed in terms of (2.146)–(2.150). Table 2.5 presents these relationships, which can all be derived by application of the chain rule.

For local versions of the dissimilarity measures, the Gâteaux derivatives are simple extensions of the Gâteaux derivatives of the global dissimilarity measures. Derivations similar to those in Appendix B allow us to state:

$$\mathcal{P}_H^{\ell,\gamma}(\mathbf{x}; F^{\mathbf{u}}, \mathbf{x}_0) = w_\gamma(\mathbf{x} - \mathbf{x}_0)\left[\mathcal{G}_\gamma(\mathbf{x}_0)\right]^{-1}\mathcal{K}_H^{\ell,\gamma}(\mathbf{x}; F^{\mathbf{u}}, \mathbf{x}_0)\nabla F^{\mathbf{u}}(\mathbf{x}), \tag{2.152}$$

$$\mathcal{P}_H^{\ell,\gamma}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0) = w_\gamma(\mathbf{x} - \mathbf{x}_0)\left[\mathcal{G}_\gamma(\mathbf{x}_0)\right]^{-1}\mathcal{K}_H^{\ell,\gamma}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0)\nabla F^{\mathbf{u}}(\mathbf{x}), \tag{2.153}$$

$$\mathcal{P}_\varepsilon^{\ell,\gamma}(\mathbf{x}; F^{\mathbf{u}}, \mathbf{x}_0) = w_\gamma(\mathbf{x} - \mathbf{x}_0)\left[\mathcal{G}_\gamma(\mathbf{x}_0)\right]^{-1}\mathcal{K}_\varepsilon^{\ell,\gamma}(\mathbf{x}; F^{\mathbf{u}}, \mathbf{x}_0)\nabla F^{\mathbf{u}}(\mathbf{x}), \tag{2.154}$$

$$\mathcal{P}_\varepsilon^{\ell,\gamma}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0) = w_\gamma(\mathbf{x} - \mathbf{x}_0)\left[\mathcal{G}_\gamma(\mathbf{x}_0)\right]^{-1}\mathcal{K}_\varepsilon^{\ell,\gamma}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0)\nabla F^{\mathbf{u}}(\mathbf{x}), \tag{2.155}$$

$$\begin{aligned}\mathcal{P}_\varepsilon^{\ell,\gamma}(\mathbf{x}; F^{\mathbf{u}}, R, \mathbf{x}_0) = &\, w_\gamma(\mathbf{x} - \mathbf{x}_0)\left[\mathcal{G}_\gamma(\mathbf{x}_0)\right]^{-1}\mathcal{K}_\varepsilon^{\ell,\gamma}(\mathbf{x}; F^{\mathbf{u}}, R, \mathbf{x}_0)\nabla F^{\mathbf{u}}(\mathbf{x}) \\ &+ \mathcal{P}_\varepsilon^{\ell,\gamma}(\mathbf{x}; F^{\mathbf{u}}, \mathbf{x}_0)\end{aligned} \tag{2.156}$$

Disimilarity Measure	**Gâteaux Derivative Component** $\mathcal{P}^g(\mathbf{x}; R, F^{\mathbf{u}})$
MI	$\mathcal{P}^g_H(\mathbf{x}; F^{\mathbf{u}}) - \mathcal{P}^g_H(\mathbf{x}; R, F^{\mathbf{u}})$
NMI	$(H(R, F^{\mathbf{u}}))^{-1} \left[\mathcal{P}^g_H(\mathbf{x}; F^{\mathbf{u}}) - NMI(R, F^{\mathbf{u}})\, \mathcal{P}^g_H(\mathbf{x}; R, F^{\mathbf{u}})\right]$
ECC	$-\left[H(R) + H(F^{\mathbf{u}})\right]^{-1} \Big\{ \frac{1}{2} ECC(R, F^{\mathbf{u}})\, \mathcal{P}^g_H(\mathbf{x}; F^{\mathbf{u}}) - (ECC(R, F^{\mathbf{u}}))^{-1} \left[\mathcal{P}^g_H(\mathbf{x}; F^{\mathbf{u}}) - \mathcal{P}^g_H(\mathbf{x}; R, F^{\mathbf{u}})\right] \Big\}$
$CCRE(R, F^{\mathbf{u}})$	$\mathcal{P}^g_\varepsilon(\mathbf{x}; F^{\mathbf{u}}) - \mathcal{P}^g_\varepsilon(\mathbf{x}; F^{\mathbf{u}}, R)$
$CCRE(F^{\mathbf{u}}, R)$	$\mathcal{P}^g_\varepsilon(\mathbf{x}; F^{\mathbf{u}}) - \mathcal{P}^g_\varepsilon(\mathbf{x}; R, F^{\mathbf{u}})$
$SCCRE$	$\mathcal{P}^g_\varepsilon(\mathbf{x}; F^{\mathbf{u}}) - \frac{1}{2}\mathcal{P}^g_\varepsilon(\mathbf{x}; F^{\mathbf{u}}, R) - \frac{1}{2}\mathcal{P}^g_\varepsilon(\mathbf{x}; R, F^{\mathbf{u}})$
$NCCRE$	$\left[\varepsilon(R, F^{\mathbf{u}}) + \varepsilon(F^{\mathbf{u}}, R)\right]^{-1} \cdot \Big\{ 2\, \mathcal{P}^g_\varepsilon(\mathbf{x}; F^{\mathbf{u}}) - NCCRE(R, F^{\mathbf{u}}) \left[\mathcal{P}^g_\varepsilon(\mathbf{x}; R, F^{\mathbf{u}}) + \mathcal{P}^g_\varepsilon(\mathbf{x}; F^{\mathbf{u}}, R)\right] \Big\}$
$CRECC$	$-\left[\varepsilon(R) + \varepsilon(F^{\mathbf{u}})\right]^{-1} \Big\{ \frac{1}{2} CRECC(R, F^{\mathbf{u}})\, \mathcal{P}^g_\varepsilon(\mathbf{x}; F^{\mathbf{u}}) - (CRECC(R, F^{\mathbf{u}}))^{-1} \cdot \left[\mathcal{P}^g_\varepsilon(\mathbf{x}; F^{\mathbf{u}}) - \frac{1}{2}\mathcal{P}^g_\varepsilon(\mathbf{x}; R, F^{\mathbf{u}}) - \frac{1}{2}\mathcal{P}^g_\varepsilon(\mathbf{x}; F^{\mathbf{u}}, R)\right] \Big\}$

Table 2.5: $\mathcal{P}^g$ for global dissimilarity measures based on information content assumptions, expressed as functions of $\mathcal{P}^g_H$ and $\mathcal{P}^g_\varepsilon$.

where

$$\mathcal{K}_H^{\ell,\gamma}(\mathbf{x}; F^{\mathbf{u}}, \mathbf{x}_0) = \frac{p_{F^{\mathbf{u}}}^{\gamma}{}'(F^{\mathbf{u}}(\mathbf{x}); \mathbf{x}_0)}{p_{F^{\mathbf{u}}}^{\gamma}(F^{\mathbf{u}}(\mathbf{x}); \mathbf{x}_0)}, \tag{2.157}$$

$$\mathcal{K}_H^{\ell,\gamma}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0) = \frac{\frac{\partial}{\partial f}\left[p_{R,F^{\mathbf{u}}}^{\gamma}(R(\mathbf{x}), f; \mathbf{x}_0)\right]_{f=F^{\mathbf{u}}(\mathbf{x})}}{p_{R,F^{\mathbf{u}}}^{\gamma}(R(\mathbf{x}), F^{\mathbf{u}}(\mathbf{x}); \mathbf{x}_0)}, \tag{2.158}$$

$$\mathcal{K}_{\varepsilon}^{\ell,\gamma}(\mathbf{x}; F^{\mathbf{u}}, \mathbf{x}_0) = 1 + \log P^{\gamma}(F^{\mathbf{u}} > F^{\mathbf{u}}(\mathbf{x}); \mathbf{x}_0), \tag{2.159}$$

$$\mathcal{K}_{\varepsilon}^{\ell,\gamma}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0) = 1 + \log P^{\gamma}(F^{\mathbf{u}} > F^{\mathbf{u}}(\mathbf{x}) | R = R(\mathbf{x}); \mathbf{x}_0), \tag{2.160}$$

$$\begin{aligned}\mathcal{K}_{\varepsilon}^{\ell,\gamma}(\mathbf{x}; F^{\mathbf{u}}, R, \mathbf{x}_0) = &\int_{-\infty}^{R(\mathbf{x})} \frac{\partial}{\partial f}\left[\log P^{\gamma}(R > r | F^{\mathbf{u}} = f; \mathbf{x}_0)\right]_{f=F^{\mathbf{u}}(\mathbf{x})} dr \\ &+ \int_{-\infty}^{R(\mathbf{x})} \frac{\partial}{\partial f}\left[P^{\gamma}(R < r | F^{\mathbf{u}} = f; \mathbf{x}_0)\right]_{f=F^{\mathbf{u}}(\mathbf{x})} dr \\ &- \int_{R(\mathbf{x})}^{\infty} \frac{\partial}{\partial f}\left[P^{\gamma}(R > r | F^{\mathbf{u}} = f; \mathbf{x}_0)\right]_{f=F^{\mathbf{u}}(\mathbf{x})} dr,\end{aligned} \tag{2.161}$$

and where $\mathcal{G}_{\gamma}(\mathbf{x}_0)$ is given by (2.129).

The Gâteaux derivatives of the local dissimilarity measures based on information content assumptions can be expressed in terms of (2.152)–(2.156). Table 2.6 presents these local relationships, evaluated at $\mathbf{x}_0$. Aggregate versions of the local dissimilarity measures can be found by integrating these terms according to (2.118). Like the relationships described in Table 2.5, those in Table 2.6 can be derived by application of the chain rule.

2.5 A Note on the Demons Algorithm

Thirion's Demons algorithm [87] is a popular choice for nonrigid registration because of its linear complexity and simple implementation. It involves iteratively computing a force vector field from the images and then using the the force vector field to locally drive the deformation field. In the notation of this thesis, the *Demons force* vector field can be defined as:

$$\mathcal{P}_{Demons}(\mathbf{x}; R, F^{\mathbf{u}}) = \frac{(R(\mathbf{x}) - F^{\mathbf{u}}(\mathbf{x}))\,\nabla F^{\mathbf{u}}(\mathbf{x})}{\|\nabla F^{\mathbf{u}}(\mathbf{x})\|^2 + (R(\mathbf{x}) - F^{\mathbf{u}}(\mathbf{x}))^2}. \tag{2.162}$$

Thirion motivated this force vector as a way to guarantee a smooth and stable solution to the optical flow equation. Although no direct connection to a particular dissimilarity measure was presented in Thirion's original development, it is readily apparent that (2.162)

Disimilarity Measure	**Gâteaux Derivative Component** $\mathcal{P}^{\ell,\gamma}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0)$
$MI_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)$	$\mathcal{P}_H^{\ell,\gamma}(\mathbf{x}; F^{\mathbf{u}}, \mathbf{x}_0) - \mathcal{P}_H^{\ell,\gamma}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0)$
$NMI_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)$	$(H_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0))^{-1} \left[\mathcal{P}_H^{\ell,\gamma}(\mathbf{x}; F^{\mathbf{u}}, \mathbf{x}_0) - NMI_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)\, \mathcal{P}_H^{\ell,\gamma}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0)\right]$
$ECC_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)$	$-\left[H_\gamma(R) + H_\gamma(F^{\mathbf{u}})\right]^{-1} \Big\{ \frac{1}{2} ECC_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)\, \mathcal{P}_H^{\ell,\gamma}(\mathbf{x}; F^{\mathbf{u}}, \mathbf{x}_0)$ $- \left(ECC_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)\right)^{-1} \left[\mathcal{P}_H^{\ell,\gamma}(\mathbf{x}; F^{\mathbf{u}}, \mathbf{x}_0) - \mathcal{P}_H^{\ell,\gamma}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0)\right] \Big\}$
$CCRE_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)$	$\mathcal{P}_\varepsilon^{\ell,\gamma}(\mathbf{x}; F^{\mathbf{u}}, \mathbf{x}_0) - \mathcal{P}_\varepsilon^{\ell,\gamma}(\mathbf{x}; F^{\mathbf{u}}, R, \mathbf{x}_0)$
$CCRE_\gamma(F^{\mathbf{u}}, R; \mathbf{x}_0)$	$\mathcal{P}_\varepsilon^{\ell,\gamma}(\mathbf{x}; F^{\mathbf{u}}, \mathbf{x}_0) - \mathcal{P}_\varepsilon^{\ell,\gamma}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0)$
$SCCRE_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)$	$\mathcal{P}_\varepsilon^{\ell,\gamma}(\mathbf{x}; F^{\mathbf{u}}, \mathbf{x}_0) - \frac{1}{2}\mathcal{P}_\varepsilon^{\ell,\gamma}(\mathbf{x}; F^{\mathbf{u}}, R, \mathbf{x}_0) - \frac{1}{2}\mathcal{P}_\varepsilon^{\ell,\gamma}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0)$
$NCCRE_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)$	$\left[\varepsilon_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0) + \varepsilon_\gamma(F^{\mathbf{u}}, R; \mathbf{x}_0)\right]^{-1} \cdot \Big\{ 2\, \mathcal{P}_\varepsilon^{\ell,\gamma}(\mathbf{x}; F^{\mathbf{u}}, \mathbf{x}_0)$ $- \, NCCRE_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0) \left[\mathcal{P}_\varepsilon^{\ell,\gamma}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0) + \mathcal{P}_\varepsilon^{\ell,\gamma}(\mathbf{x}; F^{\mathbf{u}}, R, \mathbf{x}_0)\right] \Big\}$
$CRECC_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)$	$-\left[\varepsilon_\gamma(R; \mathbf{x}_0) + \varepsilon_\gamma(F^{\mathbf{u}}; \mathbf{x}_0)\right]^{-1} \Big\{ \frac{1}{2} CRECC_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)\, \mathcal{P}_\varepsilon^{\ell,\gamma}(\mathbf{x}; F^{\mathbf{u}}, \mathbf{x}_0)$ $- \left(CRECC_\gamma(R, F^{\mathbf{u}}; \mathbf{x}_0)\right)^{-1} \left[\mathcal{P}_\varepsilon^{\ell,\gamma}(\mathbf{x}; F^{\mathbf{u}}, \mathbf{x}_0) - \frac{1}{2}\mathcal{P}_\varepsilon^{\ell,\gamma}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0) - \frac{1}{2}\mathcal{P}_\varepsilon^{\ell,\gamma}(\mathbf{x}; F^{\mathbf{u}}, R, \mathbf{x}_0)\right] \Big\}$

Table 2.6: $\mathcal{P}^{\ell,\gamma}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0)$ for local dissimilarity measures based on information content assumptions, expressed as functions of $\mathcal{P}_H^{\ell,\gamma}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0)$ and $\mathcal{P}_\varepsilon^{\ell,\gamma}(\mathbf{x}; R, F^{\mathbf{u}}, \mathbf{x}_0)$.

is a locally weighted version of (2.127), so it would seem that the Demons force vector field is in some way related to the MSD dissimilarity measure.

Pennec *et al.* [68] illustrated this relationship by performing a Quasi-Newton minimization on the MSD dissimilarity measure and showing that the resulting incremental update can be expressed as:

$$\mathcal{P}(\mathbf{x}; R, F^{\mathbf{u}}) = \frac{\kappa \left(R(\mathbf{x}) - F^{\mathbf{u}}(\mathbf{x})\right) \nabla F^{\mathbf{u}}(\mathbf{x})}{\|\nabla F^{\mathbf{u}}(\mathbf{x})\|^2 + \left(R(\mathbf{x}) - F^{\mathbf{u}}(\mathbf{x})\right) \left(\Delta F^{\mathbf{u}}(\mathbf{x})\right)}, \tag{2.163}$$

where κ is a constant. Pennec notes that a Levenberg-Marquardt algorithm could be used to derive an alternative incremental update in order to avoid problems with the denominator becoming zero in regions where the Hessian matrix of $F^{\mathbf{u}}$ is not positive definite.

Although the force vector field defined by (2.163) is not exactly the same as the Demons force vector field (2.162), there appears to be great potential in generalizing Pennec's version. The strategy of performing a Quasi-Newton (or Levenberg-Marquardt) minimization could theoretically be applied to *any* of the dissimilarity measures presented in this chapter, yielding a whole family of Demons-like force vector fields for use in image registration.

2.6 Summary

This chapter focused on dissimilarity measures used in medical image registration. It showed how many standard dissimilarity measures arise from penalizing deviations from various image data relationship assumptions, including constancy, linearity, general nonlinearity, and commonality of information content. Within this framework, the chapter proposed three new dissimilarity measures: the squared orthogonal correlation coefficient (SOCC), the normalized cross cumulative residual entropy (NCCRE), and the cumulative residual entropy correlation coefficient (CRECC). It then built upon Hermosillo's work [50] on defining local versions of specific dissimilarity measures and illustrated how to construct local versions of *all* dissimilarity measures. Next, the chapter illustrated how to compute the Gâteaux derivatives necessary for using these dissimilarity measures in variational image registration, and it presented the Gâteaux derivatives for the global and local versions of all of the dissimilarity measures. Finally, the chapter mentioned how the Demons algorithm can be

thought of in terms of the Gâteaux derivatives of the dissimilarity measures described in this chapter.

Chapter 3

Changing Overlap

All of the dissimilarity measures presented in Chapter 2 depend in some way on the region Ω describing the overlap between the floating and reference images. The Gâteaux derivatives presented in Section 2.4 all implicitly assume that Ω remains constant over the course of registration. This assumption is valid if homogeneous Dirichlet boundary conditions are chosen; however, it is invalid if any other type of boundary conditions are chosen, or if the dissimilarity measure is computed over some masked portion of the overlap region (which is often the case in practice). To handle these more general situations, we must explicitly account for the changing overlap region in the dissimilarity measures and their Gâteaux derivatives.

Even when a changing overlap region is appropriately handled, a new problem emerges: sensitivity of the dissimilarity measures to the movement of background content into and out of the overlap region over the course of registration. Studholme [84] investigated this problem in the context of registering whole head MR images with small band CT images. He recognized that MI exhibits such an *overlap sensitivity* problem and showed empirically that it could be mitigated with NMI. It turns out, however, that *all* of the dissimilarity measures presented in Chapter 2, including NMI, exhibit this overlap sensitivity problem, although the problem is less apparent in some dissimilarity measures than others.

This chapter explores the effect of changing overlap on the dissimilarity measures presented in Chapter 2. It first shows how to explicitly account for changing overlap in the dissimilarity measures and their Gâteaux derivatives. Next, it uses an academic example

presented by Studholme [84] to illustrate the overlap sensitivity problem for many of the dissimilarity measures. It then proposes a framework for modifying *all* of the dissimilarity measures in a way that is guaranteed to be insensitive to overlap changes under certain assumptions on background statistics. Finally, it illustrates the performance of the modified dissimilarity measures versus the standard dissimilarity measures on both rigid and nonrigid registration tasks.

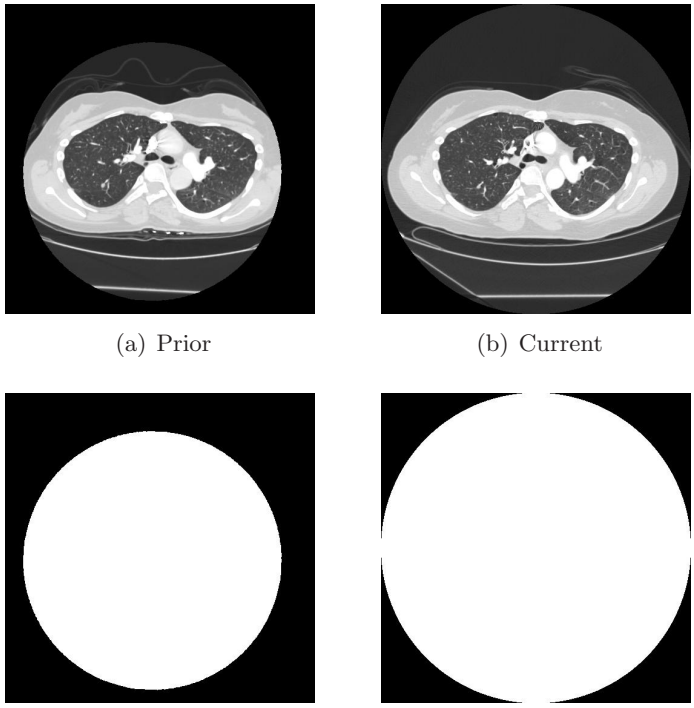
3.1 Accounting for Changing Overlap in Registration

As noted above, in many practical situations, it is desirable to consider only a portion of the floating or reference image when aligning images. Consider, for example, the axial slices from serial 3D chest CT scans of the same patient shown in Fig.'s 3.1(a) and 3.1(b). The particular reconstruction algorithm used to generate these CT scans only reconstructs data within a cylinder. In computing any image dissimilarity measure, we thus only want to use the valid data. One way to ensure this is to construct masks that indicate valid sampling locations (as shown in Fig. 3.1(c) and Fig. 3.1(d)) and deform the masks over the course of registration. As the masks deform over the course of registration, their region of intersection deforms as well. This region of intersection must then be the overlap region Ω, as it is the only region where valid data can be sampled from both images simultaneously.

In other situations, all of the data may be valid in both the reference and (original) floating images; however, one may want to experiment with different boundary conditions in registration. Unless homogeneous Dirichlet boundary conditions are used, the image boundaries will move over the course of registration, causing the overlap region Ω to change as well.

A problem now arises: in computing the Gâteaux derivatives of all of the dissimilarity measures presented in Chapter 2, the implicit assumption was made that Ω remains constant over the course of registration. In order to perform nonrigid registration in situations where Ω *changes*, we must revisit the dissimilarity measures and recompute their Gâteaux derivatives under the assumption that Ω depends on the displacement field $\mathbf{u}(\mathbf{x})$.

To properly account for changing overlap in nonrigid registration, we first investigate



(a) Prior (b) Current

(c) Prior Mask (d) Current Mask

Figure 3.1: Example axial slices of serial 3D CT scans of the same patient and corresponding masks indicating the positions of valid data.

the changing behavior of the region content $|\Omega|$. Then, we describe the impact this has on the Gâteaux derivatives of the dissimilarity measures.

3.1.1 Gâteaux Derivative of Region Content

For either the situation where masks exist indicating valid data, or the situation where flexible boundary conditions are desired, we can describe the overlap region conceptually by Fig. 3.2. Consider the reference image to be defined over the region $\Omega_R \subset \mathbb{R}^n$, and consider the deformed floating image to be defined over the region $\Omega_{F^{\mathbf{u}}} \subset \mathbb{R}^n$. The overlap region Ω over which both images are defined is given by the intersection $\Omega_R \cap \Omega_{F^{\mathbf{u}}}$. (The situation where no masks exist and homogeneous Dirichlet boundary conditions are chosen is still covered; this case simply yields $\Omega = \Omega_R = \Omega_{F^{\mathbf{u}}}$.)

Now, the content $\mathrm{C}(\mathbf{u}) := |\Omega|$ of the overlap region is given by:

$$\mathrm{C}(\mathbf{u}) = \int_{\mathbb{R}^n} \mathrm{H}(\Theta_R(\mathbf{x})) \cdot \mathrm{H}(\Theta_{F^{\mathbf{u}}}(\mathbf{x}))\, d\mathbf{x}, \tag{3.1}$$

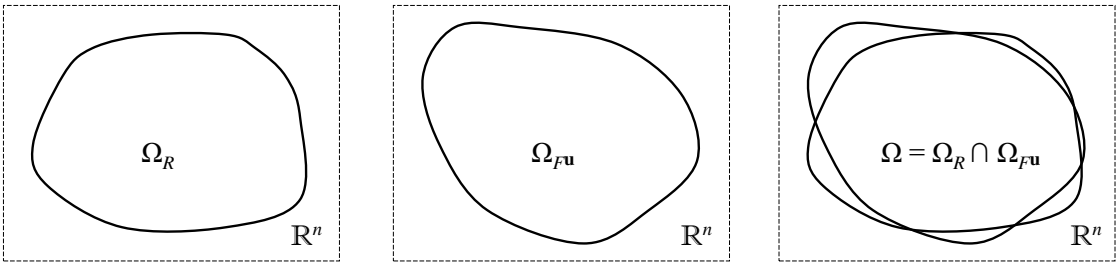


Figure 3.2: Regions over which reference image (left) and floating image (middle) are defined; overlap region Ω as the intersection of these regions (right).

where $\Theta_A(\mathbf{x}) : \mathbb{R}^n \mapsto \mathbb{R}$ is a function exhibiting the following properties:

$$\begin{aligned} \Theta_A(\mathbf{x}) &> 0, \qquad \mathbf{x} \in \Omega_A, \\ \Theta_A(\mathbf{x}) &= 0, \qquad \mathbf{x} \in \partial\Omega_A, \\ \Theta_A(\mathbf{x}) &< 0, \qquad \text{otherwise}, \end{aligned} \tag{3.2}$$

and where $\mathrm{H}(\Theta)$ is the Heaviside step function, defined by:

$$\mathrm{H}(\Theta) = \begin{cases} 1, & \Theta \geq 0 \\ 0, & \Theta < 0 \end{cases} . \tag{3.3}$$

A natural choice for Θ is the signed Euclidean distance transform [62], which can be computed in linear time.

If we approximate H in (3.1) with the regularized Heaviside function H_ϵ defined in [25] by:

$$\mathrm{H}_\epsilon(\Theta) = \frac{1}{2}\left(1 + \frac{2}{\pi}\arctan\left(\frac{\Theta}{\epsilon}\right)\right), \tag{3.4}$$

we find that the Gâteaux derivative $d\mathrm{C}(\mathbf{u};\mathbf{w})$ is given by:

$$d\mathrm{C}(\mathbf{u};\mathbf{w}) = \int_{\mathbb{R}^n} \left\langle \mathcal{P}_\mathrm{C}(\mathbf{x}; R, F^{\mathbf{u}}), \mathbf{w}(\mathbf{x}) \right\rangle d\mathbf{x}, \tag{3.5}$$

where

$$\mathcal{P}_\mathrm{C}(\mathbf{x}; R, F^{\mathbf{u}}) \approx -\mathrm{H}_\epsilon(\Theta_R(\mathbf{x})) \cdot \mathrm{H}_\epsilon^{'}(\Theta_{F^{\mathbf{u}}}(\mathbf{x})) \cdot \nabla\Theta_{F^{\mathbf{u}}}(\mathbf{x}), \tag{3.6}$$

and where

$$\mathrm{H}_\epsilon^{'}(\Theta) = \frac{\epsilon}{\pi\left(\Theta^2 + \epsilon^2\right)} \tag{3.7}$$

is a regularized version of the Dirac delta function. Note that the choice $\epsilon = \pi^{-1}$ yields $\mathrm{H}'_{\pi^{-1}}(0) = 1$. A proof of (3.5) is provided in Appendix C.

To get a visual sense of the Gâteaux derivative of the content of the overlap region, we observe in Fig. 3.3 the various components that are used to form the vector field $\mathcal{P}_{\mathrm{C}}(\mathbf{x}; R, F^{\mathbf{u}})$ for the regions shown in Fig. 3.2. Figures 3.3(a)–3.3(b) show the regions Ω_R and $\Omega_{F^{\mathbf{u}}}$ from Fig. 3.2 as binary masks, and Fig.'s 3.3(c)–3.3(d) show the corresponding signed Euclidean distance transforms of these regions. Figures 3.3(e) and 3.3(f) show the regularized Heaviside function for the reference region and the derivative of the regularized Heaviside function for the floating region, respectively, assuming that the images are 300 pixels × 400 pixels with isotropic pixel spacing of 0.1 units, and that $\epsilon = \pi^{-1}$. Figure 3.3(e) shows the product of $\mathrm{H}_\epsilon(\Theta_R(\mathbf{x}))$ and $\mathrm{H}'_\epsilon(\Theta_{F^{\mathbf{u}}}(\mathbf{x}))$, and Fig. 3.3(h) illustrates the resulting vector field $\mathcal{P}_{\mathrm{C}}(\mathbf{x}; R, F^{\mathbf{u}})$ overlayed on the mask of the floating region. As can be seen in Fig. 3.3(h), the vector field $\mathcal{P}_{\mathrm{C}}(\mathbf{x}; R, F^{\mathbf{u}})$ has nontrivial vectors only near the edges of $\Omega_{F^{\mathbf{u}}}$ that are inside Ω_R, and these vectors point outward from $\Omega_{F^{\mathbf{u}}}$.

3.1.2 Revisiting Gâteaux Derivatives of Dissimilarity Measures

The Gâteaux derivatives presented in Section 2.4 implicitly assume that Ω does not change as a function of the displacement field $\mathbf{u}$. To relax this assumption, we can reformulate the dissimilarity measures to explicitly account for masked regions and then derive the resulting new forms of the Gâteaux derivatives.

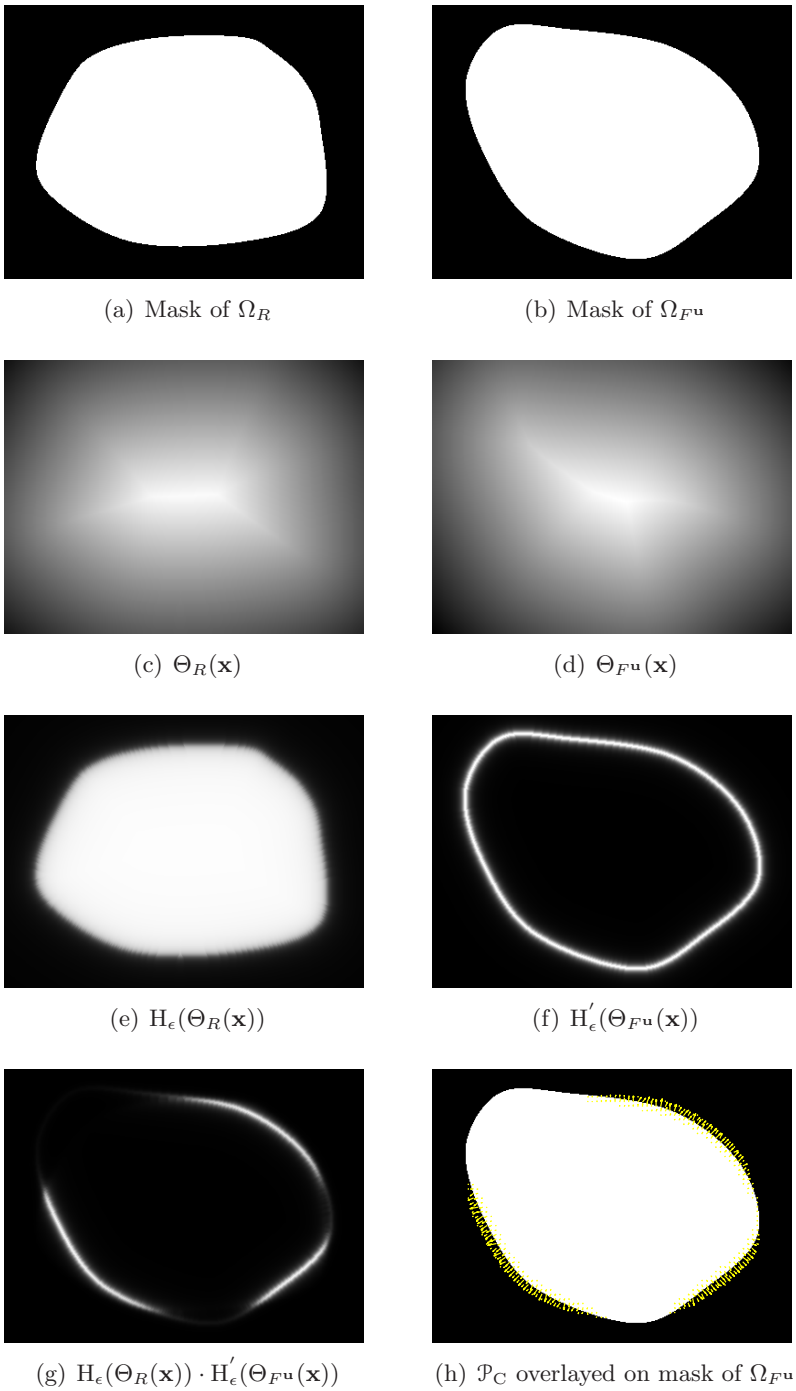
3.1.2.1 General Constancy / MSD

In order to explicitly account for the regions Ω_R and $\Omega_{F^{\mathbf{u}}}$ over which the reference and floating images are defined, we rewrite the general constancy dissimilarity measure (2.19) as:

$$\mathcal{J}^g_{\Psi,r}(R, F^{\mathbf{u}}) := \frac{\mathrm{M}_{\Psi,r}(R, F^{\mathbf{u}})}{\mathrm{C}(\mathbf{u})}, \tag{3.8}$$

with

$$\mathrm{M}_{\Psi,r}(R, F^{\mathbf{u}}) := \int_{\mathbb{R}^n} \mathrm{H}(\Theta_R(\mathbf{x})) \cdot \mathrm{H}(\Theta_{F^{\mathbf{u}}}(\mathbf{x})) \cdot \Psi(r(\mathbf{x}; R, F^{\mathbf{u}}))\, d\mathbf{x} \tag{3.9}$$



(a) Mask of Ω_R (b) Mask of $\Omega_{F^{\mathbf{u}}}$

(c) $\Theta_R(\mathbf{x})$ (d) $\Theta_{F^{\mathbf{u}}}(\mathbf{x})$

(e) $\mathrm{H}_\epsilon(\Theta_R(\mathbf{x}))$ (f) $\mathrm{H}_\epsilon^{'}(\Theta_{F^{\mathbf{u}}}(\mathbf{x}))$

(g) $\mathrm{H}_\epsilon(\Theta_R(\mathbf{x})) \cdot \mathrm{H}_\epsilon^{'}(\Theta_{F^{\mathbf{u}}}(\mathbf{x}))$ (h) $\mathcal{P}_\mathrm{C}$ overlayed on mask of $\Omega_{F^{\mathbf{u}}}$

Figure 3.3: Visualizing the Gâteaux derivative of overlap region content.

and C($\mathbf{u}$) defined as in (3.1). As shown in Appendix C, the Gâteaux derivative of (3.8) is given by:

$$d\mathcal{J}^{g}_{\Psi,r}(R, F^{\mathbf{u}}; \mathbf{w}) = \int_{\mathbb{R}^n} \langle \mathcal{P}_{\Psi,r}(\mathbf{x}; R, F^{\mathbf{u}}) \, , \mathbf{w}(\mathbf{x}) \rangle \, d\mathbf{x}, \tag{3.10}$$

where

$$\begin{aligned} \mathcal{P}_{\Psi,r}(\mathbf{x}; R, F^{\mathbf{u}}) \approx & - \mathrm{H}_{\epsilon}(\Theta_R(\mathbf{x})) \cdot \mathrm{H}^{'}_{\epsilon}(\Theta_{F^{\mathbf{u}}}(\mathbf{x})) \cdot \left(\frac{\Psi(r(\mathbf{x}; R, F^{\mathbf{u}})) - \mathcal{J}^{g}_{\Psi,r}(R, F^{\mathbf{u}})}{\mathrm{C}(\mathbf{u})} \right) \cdot \nabla \Theta_{F^{\mathbf{u}}}(\mathbf{x}) \\ & - \mathrm{H}_{\epsilon}(\Theta_R(\mathbf{x})) \cdot \mathrm{H}_{\epsilon}(\Theta_{F^{\mathbf{u}}}(\mathbf{x})) \cdot \left(\frac{\frac{\partial}{\partial F} \Psi(r(\mathbf{x}; R, F^{\mathbf{u}}))}{\mathrm{C}(\mathbf{u})} \right) \cdot \nabla F^{\mathbf{u}}(\mathbf{x}) \, . \end{aligned} \tag{3.11}$$

For specific residuals and penalty functions, (3.11) can be constructed by referring to Tables 2.3 and 2.4. The MSD, for example, has $\mathcal{P}_{MSD} := \mathcal{P}_{\Psi_2, r_{BC}}$ given by:

$$\begin{aligned} & \mathcal{P}_{MSD}(\mathbf{x}; R, F^{\mathbf{u}}) \approx \\ & \quad - \mathrm{H}_{\epsilon}(\Theta_R(\mathbf{x})) \cdot \mathrm{H}^{'}_{\epsilon}(\Theta_{F^{\mathbf{u}}}(\mathbf{x})) \cdot \left(\frac{(F^{\mathbf{u}}(\mathbf{x}) - R(\mathbf{x}))^2 - MSD(R, F^{\mathbf{u}})}{\mathrm{C}(\mathbf{u})} \right) \cdot \nabla \Theta_{F^{\mathbf{u}}}(\mathbf{x}) \\ & \quad + \mathrm{H}_{\epsilon}(\Theta_R(\mathbf{x})) \cdot \mathrm{H}_{\epsilon}(\Theta_{F^{\mathbf{u}}}(\mathbf{x})) \cdot \left(\frac{2\,(R(\mathbf{x}) - F^{\mathbf{u}}(\mathbf{x}))}{\mathrm{C}(\mathbf{u})} \right) \cdot \nabla F^{\mathbf{u}}(\mathbf{x}). \end{aligned} \tag{3.12}$$

3.1.2.2 Mutual Information

Rather than spending many pages recomputing Gâteaux derivatives for all of the other dissimilarity measures under the assumption that the overlap region changes over the course of registration, we focus only on mutual information (MI). By illustrating how to recompute the Gâteaux derivative of MI, we provide the necessary tools for the interested reader to recompute Gâteaux derivatives for the other dissimilarity measures.

Mutual information relies on estimates of the underlying marginal and joint density functions of the reference and floating images. If the overlap region is assumed to remain constant over the course of registration, the Gâteaux derivative of the marginal density of the reference image is trivial, while the derivatives of the marginal density of the floating image and the joint density of both images are given in Lemmas 3 and 4 of Appendix B. If the overlap region is allowed to deform according to the displacement field $\mathbf{u}$, however, the Gâteaux derivatives of the marginal and joint densities become more complicated.

For the marginal density of the reference image, the Gâteaux derivative is no longer trivial because of the changing region over which the reference image is sampled. It is given

by:

$$dp_R(r;\mathbf{u},\mathbf{w}) = \int_{\mathbb{R}^n} \langle \mathcal{P}_{p_R}(\mathbf{x},r;R,F^{\mathbf{u}})\,,\mathbf{w}(\mathbf{x})\rangle\, d\mathbf{x}, \tag{3.13}$$

where

$$\mathcal{P}_{p_R}(\mathbf{x},r;R,F^{\mathbf{u}}) \approx -\mathrm{H}_\epsilon(\Theta_R(\mathbf{x}))\cdot \mathrm{H}_\epsilon'(\Theta_{F^{\mathbf{u}}}(\mathbf{x}))\cdot\left(\frac{\omega_\sigma(R(\mathbf{x})-r)-p_R(r)}{\mathrm{C}(\mathbf{u})}\right)\cdot\nabla\Theta_{F^{\mathbf{u}}}(\mathbf{x})\,. \tag{3.14}$$

For the marginal density of the floating image and the joint density of both images, we have:

$$dp_{F^{\mathbf{u}}}(f;\mathbf{w}) = \int_{\mathbb{R}^n} \langle \mathcal{P}_{p_{F^{\mathbf{u}}}}(\mathbf{x},f;R,F^{\mathbf{u}})\,,\mathbf{w}(\mathbf{x})\rangle\, d\mathbf{x}, \tag{3.15}$$

$$dp_{R,F^{\mathbf{u}}}(r,f;\mathbf{w}) = \int_{\mathbb{R}^n} \left\langle \mathcal{P}_{p_{R,F^{\mathbf{u}}}}(\mathbf{x},r,f;R,F^{\mathbf{u}})\,,\mathbf{w}(\mathbf{x})\right\rangle d\mathbf{x}, \tag{3.16}$$

where

$$\begin{aligned}\mathcal{P}_{p_{F^{\mathbf{u}}}}(\mathbf{x},f;R,F^{\mathbf{u}}) \approx\ & -\mathrm{H}_\epsilon(\Theta_R(\mathbf{x}))\cdot\mathrm{H}_\epsilon'(\Theta_{F^{\mathbf{u}}}(\mathbf{x}))\cdot\left(\frac{\omega_\sigma(F^{\mathbf{u}}(\mathbf{x})-f)-p_{F^{\mathbf{u}}}(f)}{\mathrm{C}(\mathbf{u})}\right)\cdot\nabla\Theta_{F^{\mathbf{u}}}(\mathbf{x})\\ & -\mathrm{H}_\epsilon(\Theta_R(\mathbf{x}))\cdot\mathrm{H}_\epsilon(\Theta_{F^{\mathbf{u}}}(\mathbf{x}))\cdot\left(\frac{\omega_\sigma'(F^{\mathbf{u}}(\mathbf{x})-f)}{\mathrm{C}(\mathbf{u})}\right)\cdot\nabla F^{\mathbf{u}}(\mathbf{x})\end{aligned} \tag{3.17}$$

and

$$\begin{aligned}&\mathcal{P}_{p_{R,F^{\mathbf{u}}}}(\mathbf{x},r,f;R,F^{\mathbf{u}}) \approx\\ &\quad -\mathrm{H}_\epsilon(\Theta_R(\mathbf{x}))\cdot\mathrm{H}_\epsilon'(\Theta_{F^{\mathbf{u}}}(\mathbf{x}))\cdot\left(\frac{\omega_\sigma(R(\mathbf{x})-r)\,\omega_\sigma(F^{\mathbf{u}}(\mathbf{x})-f)-p_{R,F^{\mathbf{u}}}(r,f)}{\mathrm{C}(\mathbf{u})}\right)\cdot\nabla\Theta_{F^{\mathbf{u}}}(\mathbf{x})\\ &\quad -\mathrm{H}_\epsilon(\Theta_R(\mathbf{x}))\cdot\mathrm{H}_\epsilon(\Theta_{F^{\mathbf{u}}}(\mathbf{x}))\cdot\left(\frac{\omega_\sigma(R(\mathbf{x})-r)\,\omega_\sigma'(F^{\mathbf{u}}(\mathbf{x})-f)}{\mathrm{C}(\mathbf{u})}\right)\cdot\nabla F^{\mathbf{u}}(\mathbf{x}).\end{aligned} \tag{3.18}$$

Proofs of (3.13)–(3.18) are omitted, as they are very similar to the proof of Proposition 10 in Appendix C.

Using derivations similar to those of Propositions 4 and 5 in Appendix B, we can state new forms of the Gâteaux derivatives of the marginal and joint entropies. These new forms are given by:

$$dH_R(\mathbf{u},\mathbf{w}) = \int_{\mathbb{R}^n} \langle \mathcal{P}_H(\mathbf{x};\mathbf{u},R)\,,\mathbf{w}(\mathbf{x})\rangle\, d\mathbf{x}, \tag{3.19}$$

$$dH_{F^{\mathbf{u}}}(\mathbf{w}) = \int_{\mathbb{R}^n} \langle \mathcal{P}_H(\mathbf{x};F^{\mathbf{u}})\,,\mathbf{w}(\mathbf{x})\rangle\, d\mathbf{x}, \tag{3.20}$$

and

$$dH_{R,F^{\mathbf{u}}}(\mathbf{w}) = \int_{\mathbb{R}^n} \langle \mathcal{P}_H(\mathbf{x}; R, F^{\mathbf{u}}), \mathbf{w}(\mathbf{x}) \rangle \, d\mathbf{x}, \tag{3.21}$$

where

$$\mathcal{P}_H(\mathbf{x}; \mathbf{u}, R) \approx \mathrm{H}_\epsilon(\Theta_R(\mathbf{x})) \cdot \mathrm{H}'_\epsilon(\Theta_{F^{\mathbf{u}}}(\mathbf{x})) \cdot \left(\frac{\log p_R(R(\mathbf{x})) + H(R)}{\mathrm{C}(\mathbf{u})} \right) \cdot \nabla \Theta_{F^{\mathbf{u}}}(\mathbf{x}), \tag{3.22}$$

$$\begin{aligned} \mathcal{P}_H(\mathbf{x}; F^{\mathbf{u}}) \approx{}& \mathrm{H}_\epsilon(\Theta_R(\mathbf{x})) \cdot \mathrm{H}'_\epsilon(\Theta_{F^{\mathbf{u}}}(\mathbf{x})) \cdot \left(\frac{\log p_{F^{\mathbf{u}}}(F^{\mathbf{u}}(\mathbf{x})) + H(F^{\mathbf{u}})}{\mathrm{C}(\mathbf{u})} \right) \cdot \nabla \Theta_{F^{\mathbf{u}}}(\mathbf{x}) \\ &+ \mathrm{H}_\epsilon(\Theta_R(\mathbf{x})) \cdot \mathrm{H}_\epsilon(\Theta_{F^{\mathbf{u}}}(\mathbf{x})) \cdot \left(\frac{p'_{F^{\mathbf{u}}}(F^{\mathbf{u}}(\mathbf{x}))}{p_{F^{\mathbf{u}}}(F^{\mathbf{u}}(\mathbf{x})) \, \mathrm{C}(\mathbf{u})} \right) \cdot \nabla F^{\mathbf{u}}(\mathbf{x}), \end{aligned} \tag{3.23}$$

and

$$\begin{aligned} &\mathcal{P}_H(\mathbf{x}; R, F^{\mathbf{u}}) \approx \\ &\quad \mathrm{H}_\epsilon(\Theta_R(\mathbf{x})) \cdot \mathrm{H}'_\epsilon(\Theta_{F^{\mathbf{u}}}(\mathbf{x})) \cdot \left(\frac{\log p_{R,F^{\mathbf{u}}}(R(\mathbf{x}), F^{\mathbf{u}}(\mathbf{x})) + H(R, F^{\mathbf{u}})}{\mathrm{C}(\mathbf{u})} \right) \cdot \nabla \Theta_{F^{\mathbf{u}}}(\mathbf{x}) \\ &\quad + \mathrm{H}_\epsilon(\Theta_R(\mathbf{x})) \cdot \mathrm{H}_\epsilon(\Theta_{F^{\mathbf{u}}}(\mathbf{x})) \cdot \left(\frac{\frac{\partial}{\partial f} \left[p_{R,F^{\mathbf{u}}}(R(\mathbf{x}), f) \right]_{f=F^{\mathbf{u}}(\mathbf{x})}}{p_{R,F^{\mathbf{u}}}(R(\mathbf{x}), F^{\mathbf{u}}(\mathbf{x})) \, \mathrm{C}(\mathbf{u})} \right) \cdot \nabla F^{\mathbf{u}}(\mathbf{x}). \end{aligned} \tag{3.24}$$

Note the difference in the notation H (entropy) and H (Heaviside function).

The new form of the Gâteaux derivative of mutual information is then given by:

$$\begin{aligned} dMI(R, F^{\mathbf{u}}; \mathbf{w}) &= dH_R(\mathbf{u}, \mathbf{w}) + dH_{F^{\mathbf{u}}}(\mathbf{w}) - dH_{R,F^{\mathbf{u}}}(\mathbf{w}) \\ &= \int_{\mathbb{R}^n} \langle \mathcal{P}_{MI}(\mathbf{x}; R, F^{\mathbf{u}}), \mathbf{w}(\mathbf{x}) \rangle \, d\mathbf{x}, \end{aligned} \tag{3.25}$$

where

$$\begin{aligned} &\mathcal{P}_{MI}(\mathbf{x}; R, F^{\mathbf{u}}) \approx \\ &\left(\frac{\mathrm{H}_\epsilon(\Theta_R(\mathbf{x})) \cdot \mathrm{H}'_\epsilon(\Theta_{F^{\mathbf{u}}}(\mathbf{x}))}{\mathrm{C}(\mathbf{u})} \right) \left(\log \left(\frac{p_R(R(\mathbf{x})) \, p_{F^{\mathbf{u}}}(F^{\mathbf{u}}(\mathbf{x}))}{p_{R,F^{\mathbf{u}}}(R(\mathbf{x}), F^{\mathbf{u}}(\mathbf{x}))} \right) + MI(R, F^{\mathbf{u}}) \right) \nabla \Theta_{F^{\mathbf{u}}}(\mathbf{x}) \\ &+ \left(\frac{\mathrm{H}_\epsilon(\Theta_R(\mathbf{x})) \cdot \mathrm{H}_\epsilon(\Theta_{F^{\mathbf{u}}}(\mathbf{x}))}{\mathrm{C}(\mathbf{u})} \right) \left(\frac{p'_{F^{\mathbf{u}}}(F^{\mathbf{u}}(\mathbf{x}))}{p_{F^{\mathbf{u}}}(F^{\mathbf{u}}(\mathbf{x}))} - \frac{\frac{\partial}{\partial f} \left[p_{R,F^{\mathbf{u}}}(R(\mathbf{x}), f) \right]_{f=F^{\mathbf{u}}(\mathbf{x})}}{p_{R,F^{\mathbf{u}}}(R(\mathbf{x}), F^{\mathbf{u}}(\mathbf{x}))} \right) \nabla F^{\mathbf{u}}(\mathbf{x}). \end{aligned} \tag{3.26}$$

3.1.2.3 Visualizing the Revised Gâteaux Derivatives

To investigate the difference between the revised Gâteaux derivatives presented in this section that account for overlap and the original Gâteaux derivatives presented in Chapter

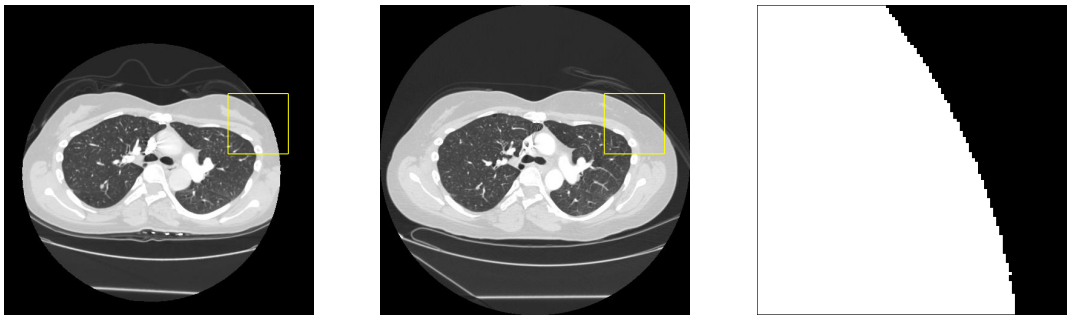
2 that do not account for overlap, we visualize the Gâteaux components of the MSD and MI dissimilarity measures for the chest CT slices shown in Fig. 3.1. Figures 3.4(a) and 3.4(b) show the prior and current slices with yellow boxes outlining the subregion over which $\mathcal{P}_{MSD}$ and $\mathcal{P}_{MI}$ are shown. As illustrated by the subregion mask in Fig. 3.4(c), the subregion is chosen to encompass regions both inside and outside the original prior mask shown in Fig. 3.1(c).

In Fig.'s 3.4(d) and 3.4(g), the subregion from the prior slice is superimposed with the vector fields $\mathcal{P}_{MSD}$ and $\mathcal{P}_{MI}$, respectively, as computed according to the definitions in Chapter 2. These definitions assume that all image information is valid and that the overlap region between the images remains constant over the course of registration. Only vectors having significant nontrivial magnitude are shown. Figures 3.4(e) and 3.4(h) show the vector fields $\mathcal{P}_{MSD}$ and $\mathcal{P}_{MI}$ as computed according to the revised definitions (3.12) and (3.26).

Differences between the revised and original versions of the Gâteaux derivative of the MSD are readily apparent. In Fig. 3.4(d), the vectors along the edge of the mask have greater magnitude than the corresponding vectors in Fig. 3.4(e). Since the original definition assumes all data is valid, it misinterprets the mask edge as a real (physical) edge. The difference between the original and revised vector fields for the MSD is illustrated in Fig. 3.4(f); it is clear from this difference field that the revised definition is identical to the original definition everywhere *except* at the mask edge.

For the revised and original versions of the Gâteaux derivative of the MI, differences are not visually apparent in the vector fields displayed in Fig.'s 3.4(g) and 3.4(h). Slight differences do exist at the mask edge, however. Figure 3.4(i) shows a magnified version of the difference field between the vector fields in Fig. 3.4(h) and 3.4(g).

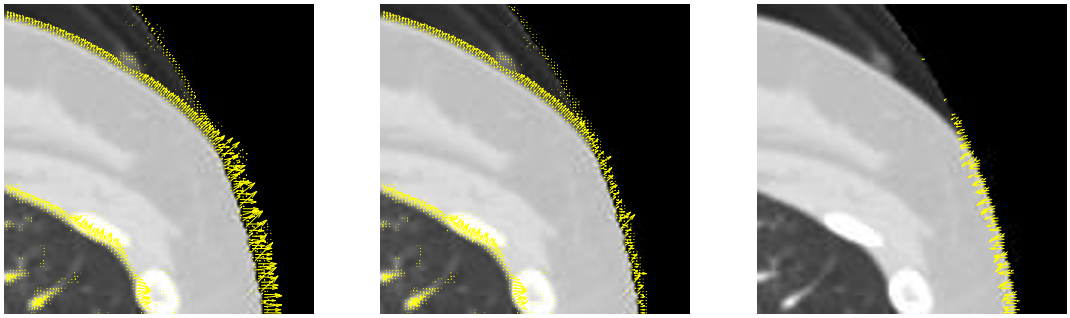
When registration is carried out using the revised and original versions of the Gâteaux derivative of the MSD dissimilarity measure, differences in the final registration result are also readily apparent. To compare the use of the revised and original Gâteaux derivatives, we performed fluid registration (described in Section 5.1.3) twice; first with force vectors corresponding to the original version of the MSD Gâteaux derivative, and next with



(a) Prior slice with subregion

(b) Current slice with subregion

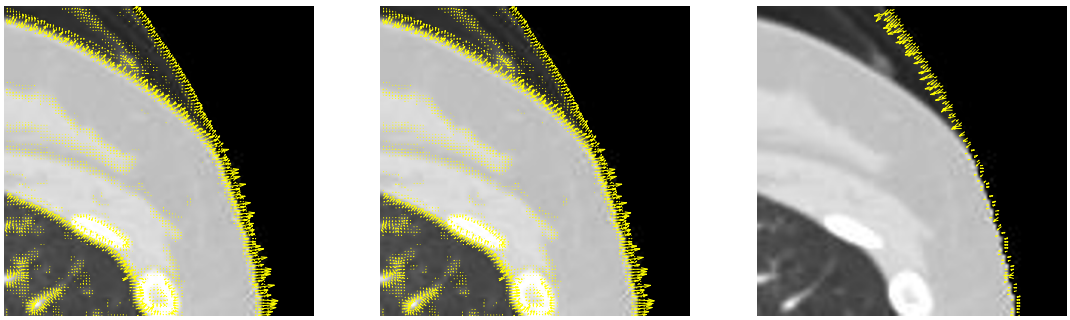
(c) Mask of prior subregion



(d) $\mathcal{P}_{MSD}$ via (2.119)

(e) $\mathcal{P}_{MSD}$ via (3.11)

(f) Difference (e) − (d)



(g) $\mathcal{P}_{MI}$ via Table 2.5

(h) $\mathcal{P}_{MI}$ via (3.26)

(i) 20 × Difference (h) − (g)

Figure 3.4: CT slices from Fig. 3.1 with superimposed vector fields from the MSD and MI dissimilarity measures, computed with and without the implicit assumption of the constancy of Ω. Only vectors with significant magnitudes are displayed.

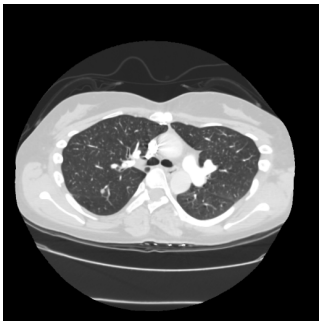
force vectors corresponding to the revised version. Using the revised force vectors, fluid registration terminated after 150 iterations based on the displacement field update having maximum magnitude below a predefined tolerance of 0.1 pixel. Using the original force vectors, fluid registration appeared to continue indefinitely, so we manually terminated it after 150 iterations in order to make a direct comparison.

In the top row of Fig. 3.5, we see the results of fluid registration with the original force vectors. The top left image is of the prior slice aligned and warped into the frame of reference of the current slice. The top middle image shows the difference between the current slice and the warped prior slice, and the top right image shows the magnitude of the resulting displacement field (with black/white indicating 0cm/1.5cm magnitudes, respectively). In the bottom row, we see similar figures that are based on the results of fluid registration with the revised force vectors.

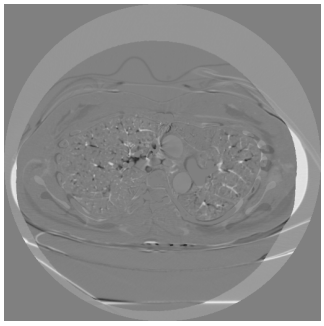
The most important observation to take away from Fig. 3.5 is apparent from comparing Fig.'s 3.5(c) and 3.5(f). Clearly, in Fig. 3.5(c), the regions that undergo the largest displacements are those near the boundary of the mask, even though this type of displacement does not represent physically valid behavior. In Fig. 3.5(f), when the revised Gâteaux derivatives are used, the large displacement magnitudes near the mask boundaries are no longer seen, suggesting a more plausible deformation.

3.2 Sensitivity of Standard Dissimilarity Measures to Changing Overlap in Background Regions

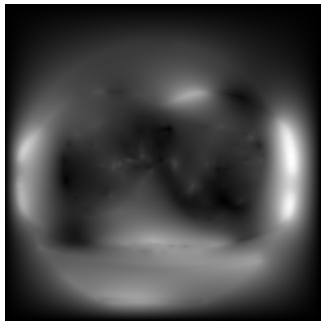
Even if the Gâteaux derivatives of dissimilarity measures are properly accounted for when the overlap region is changing, another problem emerges with the dissimilarity measures themselves. As registration progresses, image content moves into and out of the changing overlap region. Ideally, we would like dissimilarity measures to be immune to fluctuation if the only content moving into and out of the changing overlap region is *background*. But dissimilarity measures depend both on foreground and background information, so they will fluctuate if the amount of background information changes. We refer to this problem as the *overlap sensitivity problem*.



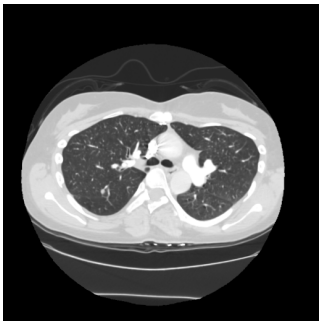
(a) Aligned prior (original force)



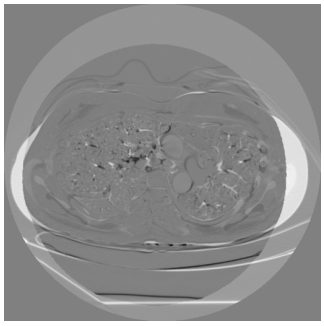
(b) Current − aligned prior (original force)



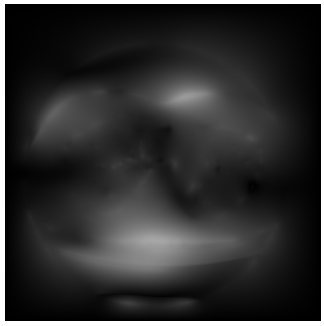
(c) Displacement Magnitude (original force)



(d) Aligned prior (revised force)



(e) Current − aligned prior (revised force)



(f) Displacement Magnitude (revised force)

Figure 3.5: Results of registering CT slices using original (top) and revised (bottom) versions of the Gâteaux derivatives of the MSD dissimilarity measure. Left images are the prior slices warped into the frame of reference of the current slice; Middle images are the differences between current slice and warped prior slice; Right images are the magnitude of the displacement field, with black/white indicating 0/1.5 cm, respectively.

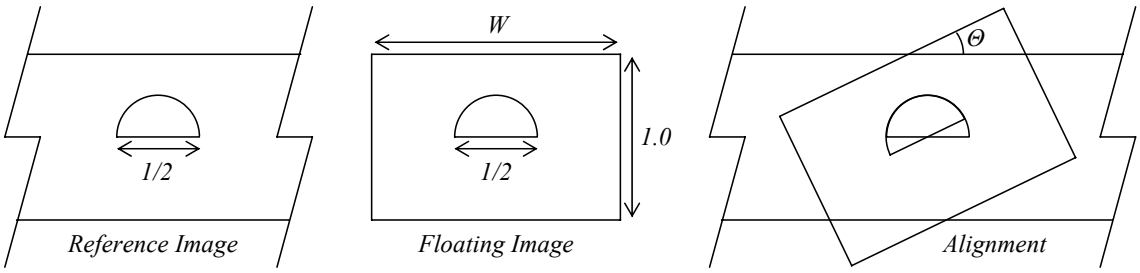


Figure 3.6: Overlap sensitivity example: reference image (left) is held in position as the floating image (middle) is overlayed and rotated (right) about its center by an amount θ.

Studholme [84] originally noticed this problem in the context of aligning MR volumes of the entire head with CT volumes that represented only a small portion of the head. CT scans are frequently limited in extent to reduce radiation exposure to the patient; however, this makes the resulting MR/CT registration problem more difficult because in addition to the multimodal nature of the images, there are anatomical structures imaged in the MR volume that are not covered by the extent of the CT volume. As registration progresses, these structures may move into or out of the overlap region, obfuscating the dissimilarity measure. Furthermore, the changing size of the overlap region itself may also obfuscate the dissimilarity measure, independently of whether or not any "interesting" structures are moving in or out.

To analyze this overlap sensitivity problem, we consider the example shown in Fig. 3.6, which is the same simplification adopted by Studholme in Fig. 6 of [84]. The reference and floating images are images of a half disc centered on a dark background; the reference image has unit height and infinite width, and the floating image has unit height and width W. The floating image is rotated about its center by an amount θ. Even though it can be argued that no "interesting" information is moving into or out of the overlap region, this simple example illustrates a problem with dissimilarity measures. Ideally, if W is varied while θ remains fixed, we would like any useful image dissimilarity measure to remain constant. However, this will not happen because dissimilarity measures are computed from the information contained in the overlap region, and the size of this overlap region changes with respect to W.

If we assign intensity values to the regions of the reference and floating images, we can observe the overlap sensitivity of various dissimilarity measures. We assign the values of 0 and 1 to the backgrounds of the reference and floating images, respectively. To the half discs, we assign the values of 2 and 4 for the reference and floating images, respectively. Using these values, we can analytically compute various dissimilarity measures as functions of W and θ. Figure 3.7 illustrates the MSD, MAD, SCC, SOCC, CR, MI, NMI, ECC, SCCRE, NCCRE, and CRECC for the example in Fig. 3.6.

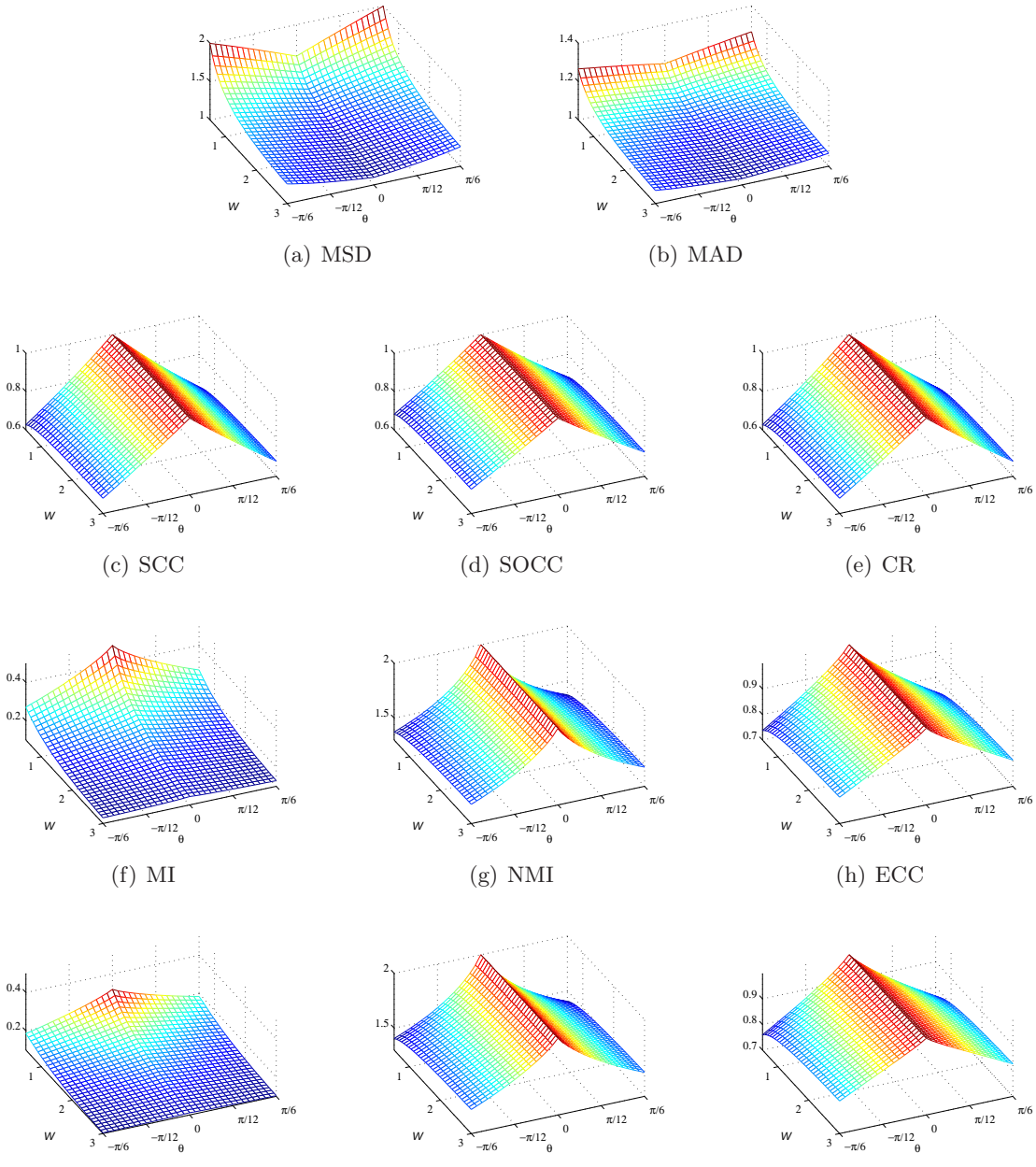
As can be seen in Fig. 3.7, if θ is fixed, the MSD, MAD, MI, and SCCRE all vary with W. Studholme [84] recognized this behavior in MI, and proposed NMI, which appears from Fig. 3.7(g) to exhibit nearly constant behavior as a function of W for fixed θ. As seen in Fig. 3.7, the SCC, SOCC, CR, ECC, NCCRE, and CRECC all appear to exhibit a similar type of invariance to W for fixed θ.

If we look more closely at the SCC, SOCC, CR, NMI, ECC, NCCRE, CRECC, however, we find that these dissimilarity measures exhibit a subtle overlap sensitivity. Figure 3.8 shows a slightly different view of all of the dissimilarity measures than shown in Fig. 3.7. For fixed values of θ (excepting $\theta = 0$), we see that *all* of the dissimilarity measures vary with respect to W.

3.3 Image Statistics and Changing Overlap in Background Regions

In the example shown in Fig. 3.6, it would seem appropriate that if θ is fixed, any dissimilarity measure that is used to compare the images be invariant to the image width W. This is because there is no image information that tells us whether one image width should be superior to another. However, none of the standard dissimilarity measures satisfy this notion of invariance, except for the special case when $\theta = 0$. This is due to the fact that as the width of the floating image changes, the proportion of background to foreground area varies, skewing the dissimilarity measures.

In order to modify dissimilarity measures to account for changes in overlap size, we first consider the relationship between image statistics and changing overlap. Recalling the



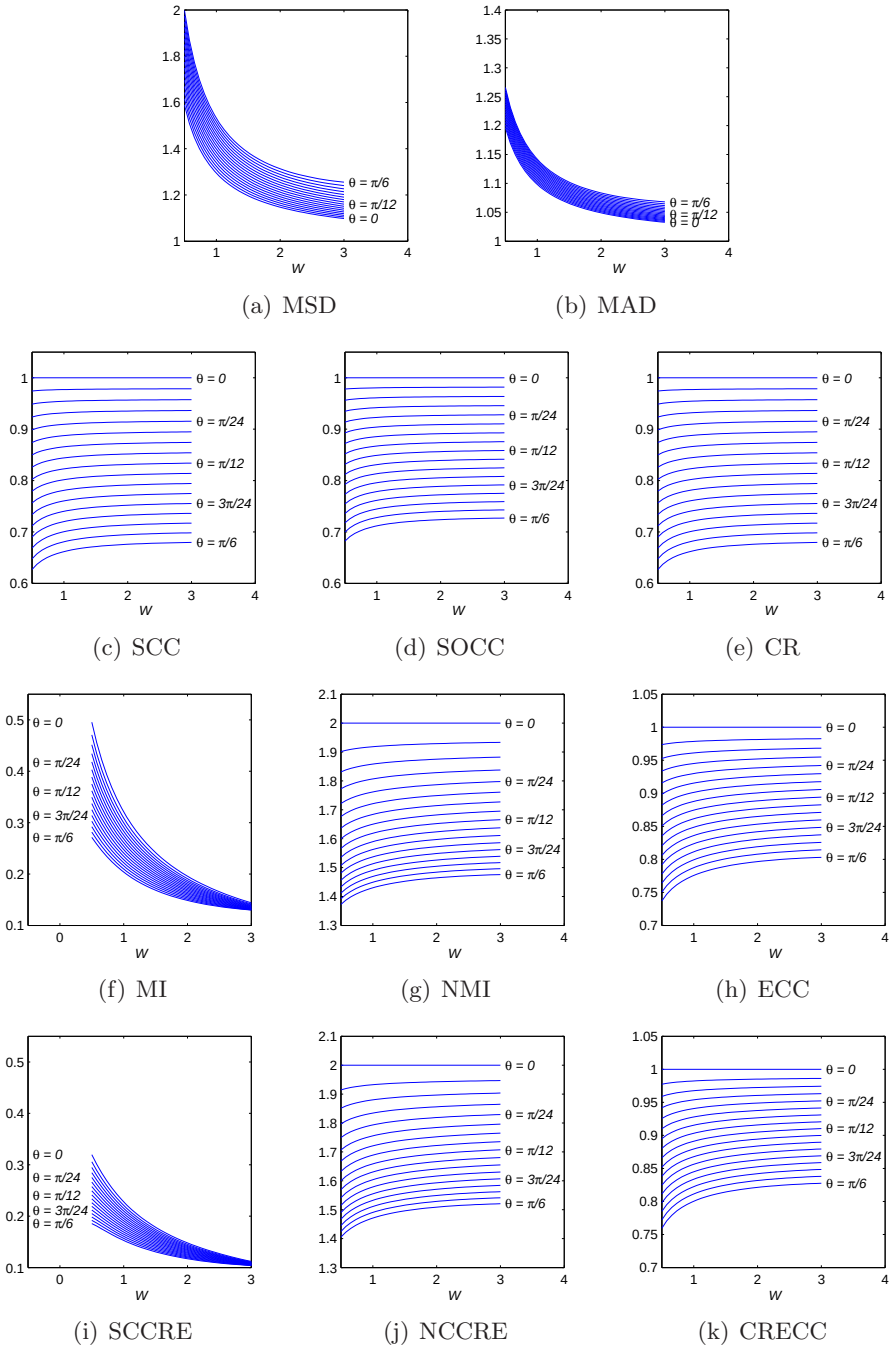
(a) MSD (b) MAD

(c) SCC (d) SOCC (e) CR

(f) MI (g) NMI (h) ECC

(i) SCCRE (j) NCCRE (k) CRECC

Figure 3.7: Values of various dissimilarity measures as functions of image width W and rotation angle θ for the example in Fig. 3.6. (Note the differing scales given to the z-axis for individual cases.)



(a) MSD (b) MAD

(c) SCC (d) SOCC (e) CR

(f) MI (g) NMI (h) ECC

(i) SCCRE (j) NCCRE (k) CRECC

Figure 3.8: Values of various dissimilarity measures as functions of image width W for fixed rotation angles θ for the example in Fig. 3.6.

diagram in Fig. 3.2, we introduce some new notation. If we allow the displacement field $\mathbf{u}$ to change as a function of time t, so that $\mathbf{u} := \mathbf{u}(\mathbf{x}, t)$, then we can denote the overlap region $\Omega = \Omega_R \cap \Omega_{F^{\mathbf{u}}}$ as Ω_t to indicate its time dependence. Then Ω_0 is the original overlap region between images R and F. We define ω_t to be the ratio of the change in overlap size to the original overlap size:

$$\omega_t = \frac{|\Omega_t| - |\Omega_0|}{|\Omega_0|} \ . \tag{3.27}$$

Let us further define $\Omega'_t := \Omega_0 \cap \Omega_t^{\mathrm{c}}$ to be the region lost and $\hat{\Omega}_t := \Omega_0^{\mathrm{c}} \cap \Omega_t$ to be the region gained at time t. Now, we make the crucial assumption that Ω'_t and $\hat{\Omega}_t$ share common statistics that are independent of t. This assumption holds, for example, when Ω'_t and $\hat{\Omega}_t$ cover only background regions in both images. We will refer to this assumption as the *background statistics assumption*.

3.3.1 Probability Distributions Over Regions

In order to describe the statistics of images over a region, we must modify our previous notation in order to explicitly account for the region. The global and local probability density functions (2.83) and (2.87) of an image A over region Ω are rewritten as $p_A^{\Omega}(a)$ and $p_A^{\Omega,\gamma}(a; \mathbf{x}_0)$, respectively. The global and local joint probability density functions of images A and B over region Ω are rewritten as $p_{A,B}^{\Omega}(a, b)$ and $p_{A,B}^{\Omega,\gamma}(a, b; \mathbf{x}_0)$, respectively.

In some situations, such as in the example of Fig. 3.6, the underlying random variables are actually discrete. In these situations, we will consider p^{Ω} and $p^{\Omega,\gamma}$ to refer to global and local probability *mass* functions. Where possible, we will use Stieltjes integrals [55] in order to derive equations and present results for both discrete and continuous random variables in a common framework. This requires defining the cumulative distribution functions:

$$P_A^{\Omega}(a) := \int_{-\infty}^{a} p_A^{\Omega}(x)\, dx \ , \tag{3.28}$$

$$P_A^{\Omega,\gamma}(a; \mathbf{x}_0) := \int_{-\infty}^{a} p_A^{\Omega,\gamma}(x; \mathbf{x}_0)\, dx \ , \tag{3.29}$$

$$P_{A,B}^{\Omega}(a, b) := \int_{-\infty}^{b} \int_{-\infty}^{a} p_{A,B}^{\Omega}(x, y)\, dxdy \ , \tag{3.30}$$

$$P_{A,B}^{\Omega,\gamma}(a, b; \mathbf{x}_0) := \int_{-\infty}^{b} \int_{-\infty}^{a} p_{A,B}^{\Omega,\gamma}(x, y; \mathbf{x}_0)\, dxdy \ . \tag{3.31}$$

Using this notation, we can now introduce the *background distributions*, which are described by their cumulative distribution functions. The global background distributions have c.d.f.'s given by:

$$\tilde{P}_A(a) := P_A^{\Omega'_t}(a) = P_A^{\hat{\Omega}_t}(a) \ , \tag{3.32}$$

$$\tilde{P}_B(b) := P_B^{\Omega'_t}(b) = P_B^{\hat{\Omega}_t}(b) \ , \tag{3.33}$$

$$\tilde{P}_{A,B}(a,b) := P_{A,B}^{\Omega'_t}(a,b) = P_{A,B}^{\hat{\Omega}_t}(a,b) \ . \tag{3.34}$$

For convenience, various global first and second order statistics will be denoted by:

$$\tilde{\mu}_A := \int_{-\infty}^{\infty} x d\tilde{P}_A(x) \ , \tag{3.35}$$

$$\tilde{\mu}_B := \int_{-\infty}^{\infty} y d\tilde{P}_B(y) \ , \tag{3.36}$$

$$\tilde{\sigma}_A^2 := \int_{-\infty}^{\infty} (x - \tilde{\mu}_A)^2 \, d\tilde{P}_A(x) \ , \tag{3.37}$$

$$\tilde{\sigma}_B^2 := \int_{-\infty}^{\infty} (y - \tilde{\mu}_B)^2 \, d\tilde{P}_B(y) \ , \tag{3.38}$$

$$\tilde{\sigma}_{A,B} := \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \tilde{\mu}_A)(y - \tilde{\mu}_B) \, d_y d_x \tilde{P}_{A,B}(x,y), \tag{3.39}$$

We will refer collectively to (3.32)–(3.39) as *global background statistics*. Local background statistics can be formed in a similar manner, but with the local distributions (3.29) and (3.31) in place of (3.28) and (3.30).

Under the background statistics assumption, we can relate distributions over Ω_t to those over Ω_0. For example, using conditioning arguments, we can establish:

$$P_A^{\Omega_0}(a) = \frac{|\Omega'_t|}{|\Omega_0|} P_A^{\Omega'_t}(a) + \frac{|\Omega_0 \cap \Omega_t|}{|\Omega_0|} P_A^{\Omega_0 \cap \Omega_t}(a) \tag{3.40}$$

and

$$P_A^{\Omega_t}(a) = \frac{\left|\hat{\Omega}_t\right|}{|\Omega_t|} P_A^{\hat{\Omega}_t}(a) + \frac{|\Omega_0 \cap \Omega_t|}{|\Omega_t|} P_A^{\Omega_0 \cap \Omega_t}(a) \ . \tag{3.41}$$

A small amount of algebraic manipulation, along with the substitution of (3.27) and (3.32) into (3.40) and (3.41), yields:

$$P_A^{\Omega_0}(a) = (1 + \omega_t) \, P_A^{\Omega_t}(a) - \omega_t \tilde{P}_A(a) \ . \tag{3.42}$$

Using similar conditioning arguments, the following relationships can also be established:

$$P_B^{\Omega_0}(b) = (1+\omega_t)\, P_B^{\Omega_t}(b) - \omega_t \tilde{P}_B(b) \quad , \tag{3.43}$$

$$P_{A,B}^{\Omega_0}(a,b) = (1+\omega_t)\, P_{A,B}^{\Omega_t}(a,b) - \omega_t \tilde{P}_{A,B}(a,b) \quad . \tag{3.44}$$

3.3.2 From Probability Distributions to Image Statistics

Now that we know how probability distributions over changing regions are related, we can use (3.42)–(3.44) in conjunction with (3.35)–(3.39) to compute the relationships between various statistics computed over Ω_0 and Ω_t. The first order statistics (expected values) of A and B are related by:

$$\mathrm{E}_{\Omega_0}[A] = (1+\omega_t)\, \mathrm{E}_{\Omega_t}[A] - \omega_t \tilde{\mu}_A \quad , \tag{3.45}$$

$$\mathrm{E}_{\Omega_0}[B] = (1+\omega_t)\, \mathrm{E}_{\Omega_t}[B] - \omega_t \tilde{\mu}_B \quad . \tag{3.46}$$

The second order statistics are related by:

$$\mathrm{Var}_{\Omega_0}(A) = (1+\omega_t)\, \mathrm{Var}_{\Omega_t}(A) - \omega_t \tilde{\sigma}_A^2 - \omega_t(1+\omega_t)(\mathrm{E}_{\Omega_t}[A] - \tilde{\mu}_A)^2 \quad , \tag{3.47}$$

$$\mathrm{Var}_{\Omega_0}(B) = (1+\omega_t)\, \mathrm{Var}_{\Omega_t}(B) - \omega_t \tilde{\sigma}_B^2 - \omega_t(1+\omega_t)(\mathrm{E}_{\Omega_t}[B] - \tilde{\mu}_B)^2 \quad , \tag{3.48}$$

$$\begin{aligned} \mathrm{Cov}_{\Omega_0}(A,B) &= (1+\omega_t)\, \mathrm{Cov}_{\Omega_t}(A,B) - \omega_t \tilde{\sigma}_{A,B} \\ &\quad -\omega_t(1+\omega_t)(\mathrm{E}_{\Omega_t}[A] - \tilde{\mu}_A)\,(\mathrm{E}_{\Omega_t}[B] - \tilde{\mu}_B) \quad . \end{aligned} \tag{3.49}$$

Similar relationships can be derived for first and second order *local* statistics.

What is important to note in (3.42) – (3.49) is that the left hand side of each equation is independent of t. This means that as t increases and the overlap region changes, the right hand side of each of these equations *remains constant*. It is this property that we will exploit in the next section to form overlap invariant versions of the similarity measures.

3.4 Modified Dissimilarity Measures

If images A and B remain in a fixed position while the overlap region is allowed to change, the right hand sides of (3.42) – (3.49) remain constant as t increases. We can use this knowledge to construct new image dissimilarity measures that are invariant to changes in overlap size.

3.4.1 Constancy Assumptions

As shown in (2.62), the general constancy dissimilarity measure (2.19) can be expressed in terms of first order moments. Rewriting (2.62) to explicitly account for the region, we have:

$$\mathcal{I}^{\Omega}_{\Psi,r}(R, F^{\mathbf{u}}) = \mathrm{E}_{\Omega}[\Psi(r(\mathbf{x}; R, F^{\mathbf{u}}))]\,. \tag{3.50}$$

From (3.45), we can see that as the region changes from Ω_0 to Ω_t, the general constancy dissimilarity measure changes according to:

$$\begin{aligned}
\mathcal{I}^{\Omega_0}_{\Psi,r}(R, F^{\mathbf{u}}) &= \mathrm{E}_{\Omega_0}[\Psi(r(\mathbf{x}; R, F^{\mathbf{u}}))] \\
&= (1+\omega_t)\,\mathrm{E}_{\Omega_t}[\Psi(r(\mathbf{x}; R, F^{\mathbf{u}}))] - \omega_t \tilde{\mu}_{\Psi(r(\mathbf{x};R,F^{\mathbf{u}}))} \\
&= (1+\omega_t)\,\mathcal{I}^{\Omega_t}_{\Psi,r}(R, F^{\mathbf{u}}) - \omega_t \tilde{\mu}_{\Psi(r(\mathbf{x};R,F^{\mathbf{u}}))}\ .
\end{aligned} \tag{3.51}$$

Since the left hand side of (3.51) remains constant as a function of t, the right hand side must also remain constant as a function of t. Therefore, we can use the right hand side of (3.51) as a modified version of the general constancy dissimilarity measure:

$$M\text{-}\mathcal{I}^{\Omega}_{\Psi,r}(R, F^{\mathbf{u}}) := (1+\omega)\,\mathcal{I}^{\Omega}_{\Psi,r}(R, F^{\mathbf{u}}) - \omega \tilde{\mu}_{\Psi(r(\mathbf{x};R,F^{\mathbf{u}}))}\ . \tag{3.52}$$

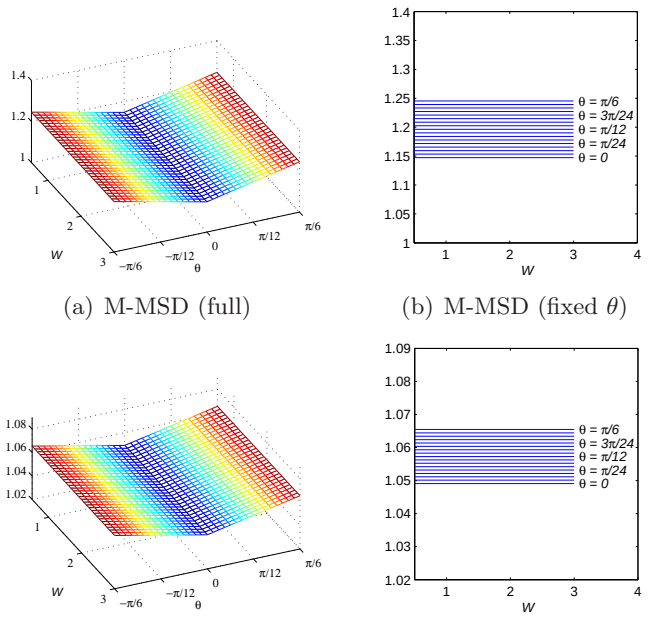
The "M-" before the dissimilarity measure indicates "modified." Note that we have dropped the subscript t in order to simplify notation. This *modified general constancy* dissimilarity measure, evaluated over region Ω, can be interpreted as a weighted sum of two terms: a dissimilarity term based on the general constancy dissimilarity measure evaluated over Ω, and a correction term based on the background statistics.

Modified versions of MSD and MAD can be constructed directly from (3.52):

$$M\text{-}MSD(R, F^{\mathbf{u}}) := (1+\omega)\,MSD(R, F^{\mathbf{u}}) - \omega \tilde{\mu}_{(F^{\mathbf{u}}(\mathbf{x})-R(\mathbf{x}))^2}\ , \tag{3.53}$$

$$M\text{-}MAD(R, F^{\mathbf{u}}) := (1+\omega)\,MAD(R, F^{\mathbf{u}}) - \omega \tilde{\mu}_{|F^{\mathbf{u}}(\mathbf{x})-R(\mathbf{x})|}\ . \tag{3.54}$$

Figure 3.9 illustrates the behavior of the modified MSD and MAD on the example shown in Fig 3.6. The reference region Ω_0 is chosen to be the overlap region corresponding to $W = 2$. As opposed to the MSD and MAD behavior shown in Fig.'s 3.7–3.7, the M-MSD and M-MAD are constant with respect to the changing overlap region.



(a) M-MSD (full) (b) M-MSD (fixed θ)

(c) M-MAD (full) (d) M-MAD (fixed θ)

Figure 3.9: Values of M-MSD and M-MAD as functions of rotation angle θ and image width W for the example in Fig. 3.6.

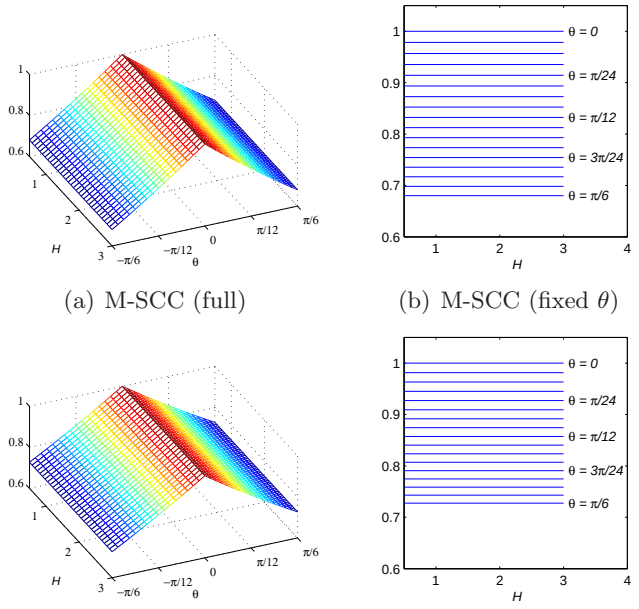
3.4.2 Linearity Assumptions

Modified versions of SSVDL and SSODL could be constructed by considering the behavior of the vector $\boldsymbol{\beta}$ of linearity parameters as the region of overlap changes. This can be done by substituting (3.45)–(3.49) into the expressions for $\hat{\beta}_1$ and $\hat{\beta}_2$ in Appendix A to construct a modified $\boldsymbol{\beta}$ vector. However, the resulting modified versions of SSVDL and SSODL are extremely complicated. Therefore, we skip the formation of M-SSVDL and M-SSODL, and jump to describing the much simpler construction of modified versions of SCC and SOCC.

M-SCC and M-SOCC can be constructed using second order statistics (3.47)–(3.49). First, we define the *modified variance* according to (3.47):

$$\mathrm{MVar}(A) := (1+\omega)\,\mathrm{Var}(A) - \omega\tilde{\sigma}_A^2 - \omega\,(1+\omega)(\mathrm{E}[A] - \tilde{\mu}_A)^2 \quad . \tag{3.55}$$

Note that as in (3.52), we have dropped the subscript t from ω. We have also dropped the subscripts from the expected values and covariance terms. This is done to simplify notation; it should be noted that $\mathrm{E}[A]$ and $\mathrm{Var}(A)$ are to be computed over the current overlap region, and that an original overlap region Ω_0 must be assumed in order to compute ω.



(a) M-SCC (full) (b) M-SCC (fixed θ)

(c) M-SOCC (full) (d) M-SOCC (fixed θ)

Figure 3.10: Values of M-SCC and M-SOCC as functions of rotation angle θ and image width H for the example in Fig. 3.6.

In a manner similar to (3.55), we can define the *modified covariance*:

$$\mathrm{MCov}(A,B) := (1+\omega)\,\mathrm{Cov}(A,B) - \omega\tilde{\sigma}_{A,B} - \omega(1+\omega)(\mathrm{E}[A]-\tilde{\mu}_A)\,(\mathrm{E}[B]-\tilde{\mu}_B) \quad . \tag{3.56}$$

The modified variance and modified covariance can now be used to define modified versions of SCC and SOCC:

$$M\text{-}SCC(R,F^{\mathbf{u}}) \;:=\; \frac{\mathrm{MCov}\,(R,F^{\mathbf{u}})^2}{\mathrm{MVar}\,(R)\,\mathrm{MVar}\,(F^{\mathbf{u}})} \quad , \tag{3.57}$$

$$M\text{-}SOCC(R,F^{\mathbf{u}}) \;:=\; \frac{(\mathrm{MVar}\,(R)-\mathrm{MVar}\,(F^{\mathbf{u}}))^2+4\mathrm{MCov}\,(R,F^{\mathbf{u}})^2}{(\mathrm{MVar}\,(R)+\mathrm{MVar}\,(F^{\mathbf{u}}))^2} \quad . \tag{3.58}$$

Figure 3.10 illustrates the behavior of the modified SCC and SOCC on the half disc example. As with the M-SSD and M-SAD, the reference region Ω_0 is chosen to be the overlap region corresponding to $\theta = 0$. The M-SCC and M-SOCC are clearly constant with respect to the changing overlap region.

3.4.3 Functional Assumptions

Modifying the correlation ratio for overlap invariance requires describing the behavior of conditional random variables over changing regions. The proper way to construct the M-

CR starts with establishing how the conditional distribution behaves over changing regions. There appears to be no easy way to do this using cumulative distribution functions. However, if we assume the underlying random variables are continuous, we can take the derivatives of (3.43)–(3.44), yielding:

$$p_A^{\Omega_0}(a) = (1+\omega_t)\, p_A^{\Omega_t}(a) - \omega_t \tilde{p}_A(a) \quad , \tag{3.59}$$

$$p_{A,B}^{\Omega_0}(a,b) = (1+\omega_t)\, p_{A,B}^{\Omega_t}(a,b) - \omega_t \tilde{p}_{A,B}(a,b) \quad . \tag{3.60}$$

We can now apply Bayes' rule to (3.60) in order to establish the following relationship:

$$p_A^{\Omega_0}(a)\, p_{B|A=a}^{\Omega_0}(b) = (1+\omega_t)\, p_A^{\Omega_t}(a)\, p_{B|A=a}^{\Omega_t}(b) - \omega_t \tilde{p}_A(a)\, \tilde{p}_{B|A=a}(b) \quad . \tag{3.61}$$

Substituting (3.59) into (3.61) and simplifying yields:

$$p_{B|A=a}^{\Omega_0}(b) = \alpha_t(a)\, p_{B|A=a}^{\Omega_t}(b) + (1-\alpha_t(a))\, \tilde{p}_{B|A=a}(b) \quad , \tag{3.62}$$

where

$$\alpha_t(a) = \frac{|\Omega_t|\, p_A^{\Omega_t}(a)}{|\Omega_t|\, p_A^{\Omega_t}(a) + (|\Omega_0| - |\Omega_t|)\, \tilde{p}_A(a)} \quad . \tag{3.63}$$

In order to determine how the correlation ratio behaves over a changing region, we first use (3.62) to analyze the behavior of $\mathrm{Var}\left(\mathrm{E}\left[B|A\right]\right)$ over a changing region. By the definition of variance, we can state:

$$\begin{aligned} \mathrm{Var}_{\Omega_0}(\mathrm{E}_{\Omega_0}[B|A]) &= \mathrm{E}_{\Omega_0}\left[\mathrm{E}_{\Omega_0}[B|A]^2\right] - \mathrm{E}_{\Omega_0}[\mathrm{E}_{\Omega_0}[B|A]]^2 \\ &= \mathrm{E}_{\Omega_0}\left[\mathrm{E}_{\Omega_0}[B|A]^2\right] - \mathrm{E}_{\Omega_0}[B]^2 \quad . \end{aligned} \tag{3.64}$$

From (3.46), we see that the term $\mathrm{E}_{\Omega_0}[B]^2$ is given by:

$$\mathrm{E}_{\Omega_0}[B]^2 = (1+\omega_t)^2\, \mathrm{E}_{\Omega_t}[B]^2 - \omega_t\, (1+\omega_t)\, \tilde{\mu}_B \mathrm{E}_{\Omega_t}[B] + \omega_t^2 \tilde{\mu}_B^2 \quad . \tag{3.65}$$

By the definition of expectation, the term $\mathrm{E}_{\Omega_0}\left[\mathrm{E}_{\Omega_0}[B|A]^2\right]$ is given by:

$$\mathrm{E}_{\Omega_0}\left[\mathrm{E}_{\Omega_0}[B|A]^2\right] = \int_{-\infty}^{\infty} p_A^{\Omega_0}(a)\, \mathrm{E}_{\Omega_0}[B|A=a]^2\, da \quad . \tag{3.66}$$

As a direct consequence of (3.62), $\mathrm{E}_{\Omega_0}[B|A=a]$ is given by:

$$\mathrm{E}_{\Omega_0}[B|A=a] = \alpha_t(a)\, \mathrm{E}_{\Omega_t}[B|A=a] + (1-\alpha_t(a))\, \tilde{\mu}_{B|A=a} \quad . \tag{3.67}$$

Substituting (3.67) and (3.59) into (3.66) and simplifying yields:

$$\begin{aligned} \mathrm{E}_{\Omega_0}\left[\mathrm{E}_{\Omega_0}[B|A]^2\right] &= (1+\omega_t)\int_{-\infty}^{\infty}\alpha_t(a)\,p_A^{\Omega_t}(a)\left(\mathrm{E}_{\Omega_t}[B|A=a]-\tilde{\mu}_{B|A=a}\right)^2 da \\ &\quad + 2\,(1+\omega_t)\,\tilde{\mu}_{B|A=a}\mathrm{E}_{\Omega_t}[B] - (1+2\omega_t)\,\tilde{\mu}^2_{B|A=a}\ . \end{aligned} \tag{3.68}$$

Substituting (3.65) and (3.68) into (3.64) and making the assumption that A and B are independent in the background region yields:

$$\begin{aligned} \mathrm{Var}_{\Omega_0}(\mathrm{E}_{\Omega_0}[B|A]) &= (1+\omega_t)\int_{-\infty}^{\infty}\alpha_t(a)\,p_A^{\Omega_t}(a)\left(\mathrm{E}_{\Omega_t}[B|A=a]-\tilde{\mu}_B\right)^2 da \\ &\quad - (1+\omega_t)^2\left(\mathrm{E}_{\Omega_t}[B]-\tilde{\mu}_B\right)^2\ , \end{aligned} \tag{3.69}$$

which describes the behavior of $\mathrm{Var}\left(\mathrm{E}\left[B|A\right]\right)$ over a changing region.

Now, we can exploit the fact that the left hand side of (3.69) remains constant as t increases to form the *modified correlation ratio*:

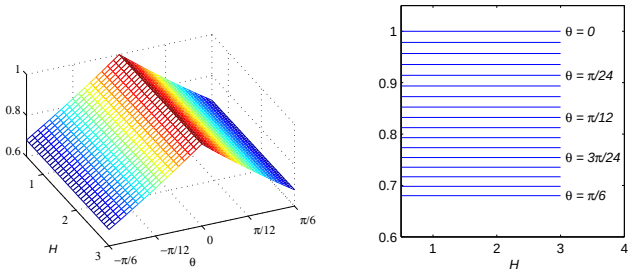
$$\begin{aligned} \textit{M-CR}(R,F^{\mathbf{u}}) &= \frac{(1+\omega)\int_{-\infty}^{\infty}\alpha(r)\,p_R(r)\left(\mathrm{E}[F^{\mathbf{u}}|R=r]-\tilde{\mu}_{F^{\mathbf{u}}}\right)^2 dr}{\mathrm{MVar}(F^{\mathbf{u}})} \\ &\quad - \frac{(1+\omega)^2\left(\mathrm{E}[F^{\mathbf{u}}]-\tilde{\mu}_{F^{\mathbf{u}}}\right)^2}{\mathrm{MVar}(F^{\mathbf{u}})}\ , \end{aligned} \tag{3.70}$$

with MVar given by (3.55). As with the other modified similarity measures, we have dropped the subscript t from ω and α, and we have dropped the subscript from the expected values and the superscript from the probability density function.

Figure 3.11 illustrates the behavior of the M-CR on the half disc example. As with the other modified similarity measures, the reference region Ω_0 is chosen to be the overlap region corresponding to $\theta = 0$. The M-CR is clearly constant with respect to the changing overlap region.

3.4.4 Information Content Assumptions

To construct modified versions of the dissimilarity measures that are based on information content assumptions, we start by defining modified versions of entropy and cumulative residual entropy, and then use these modified versions to modify MI, NMI, ECC, CCRE, SCCRE, NCCRE, and CRECC for overlap invariance.



(a) M-CR (full) (b) M-CR (fixed θ)

Figure 3.11: Values of M-CR as a function of rotation angle θ and image width H for the example in Fig. 3.6.

3.4.4.1 Modified Entropy and CRE

In modifying entropy to be overlap invariant, we must consider the discrete and continuous cases separately, as a Stieltjes integral representation of entropy does not exist for discrete random variables. (This is because $p(x)$ and $\log p(x)$ have the same points of discontinuity.)

For continuous random variables A and B, the behavior of the differential entropy and the joint differential entropy over changing regions can be found by substituting (3.59) and (3.60) into (2.10)–(2.11):

$$\begin{aligned} H_{\Omega_0}(A) &= -(1+\omega_t)\int_{-\infty}^{\infty} p_A^{\Omega_t}(a)\log\Big((1+\omega_t)\,p_A^{\Omega_t}(a)-\omega_t\tilde{p}_A(a)\Big)\,da \\ &\quad +\omega_t\int_{-\infty}^{\infty}\tilde{p}_A(a)\log\Big((1+\omega_t)\,p_A^{\Omega_t}(a)-\omega_t\tilde{p}_A(a)\Big)\,da\ , \qquad (3.71) \\ H_{\Omega_0}(A,B) &= -(1+\omega_t)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{A,B}^{\Omega_t}(a,b)\log\Big((1+\omega_t)\,p_{A,B}^{\Omega_t}(a,b)-\omega_t\tilde{p}_{A,B}(a,b)\Big)\,dbda \\ &\quad +\omega_t\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tilde{p}_{A,B}(a,b)\log\Big((1+\omega_t)\,p_{A,B}^{\Omega_t}(a)-\omega_t\tilde{p}_{A,B}(a)\Big)\,dbda\ . \qquad (3.72) \end{aligned}$$

We can exploit the fact that (3.71)–(3.72) remain constant in t and construct the *modified differential entropy* and *modified joint differential entropy*:

$$\begin{aligned} MH(A) &= -\int_{-\infty}^{\infty} q_A(a)\log q_A(a)\,da\ , \qquad (3.73) \\ MH(A,B) &= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} q_{A,B}(a,b)\log q_{A,B}(a,b)\,dbda\ , \qquad (3.74) \end{aligned}$$

where

$$\begin{aligned} q_A(a) &= (1+\omega)\,p_A(a)-\omega\tilde{p}_A(a)\ , \qquad (3.75) \\ q_{A,B}(a,b) &= (1+\omega)\,p_{A,B}(a,b)-\omega\tilde{p}_{A,B}(a,b)\ . \qquad (3.76) \end{aligned}$$

If the random variables A and B are discrete, (3.73) and (3.74) can be replaced with the *modified Shannon entropy* and *modified joint Shannon entropy*, given by:

$$MH(A) = -\sum_a q_A(a) \log q_A(a) \ , \tag{3.77}$$

$$MH(A,B) = -\sum_a \sum_b q_{A,B}(a,b) \log q_{A,B}(a,b) \ . \tag{3.78}$$

In order to define modified versions of cumulative residual entropy and joint cumulative residual entropy, we examine their behavior over changing regions. For the CRE, we see that substituting (3.42) into (2.12) yields:

$$\varepsilon_{\Omega_0}(A) = -(1+\omega_t) \int_{-\infty}^{\infty} S_A^{\Omega_t}(a) \, \log S_A^{\Omega_0}(a) \, da + \omega_t \int_{-\infty}^{\infty} \tilde{S}_A(a) \, \log \tilde{S}_A(a) \, da \ , \tag{3.79}$$

where S_A is the survival function, given by:

$$S_A(a) := 1 - P_A(a) \ . \tag{3.80}$$

Therefore, we define the *modified cumulative residual entropy* by:

$$MCRE(A) = -\int_{-\infty}^{\infty} Q_A(a) \log Q_A(a) \, da \ , \tag{3.81}$$

where

$$Q_A(a) = (1+\omega) \, S_A(a) - \omega \tilde{S}_A(a) \ . \tag{3.82}$$

Note that this definition holds independently of whether A is a discrete or continuous random variable.

To modify the JCRE, we first concentrate on the term $\mathrm{E}[\varepsilon(B|A)]$, which is defined by:

$$\mathrm{E}[\varepsilon(B|A)] = -\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_A(a) \, S_{B|A}(b|a) \, \log S_{B|A}(b|a) \, db \, da \ , \tag{3.83}$$

where $S_{B|A}$ is the conditional survival function, given by:

$$S_{B|A}(b|a) := 1 - \Pr\{B \leq b | A = a\} \ . \tag{3.84}$$

If we integrate the conditional p.d.f.'s in (3.61) and assume that the backgrounds of A and B are independent, we have:

$$p_A^{\Omega_0}(a) \, S_{B|A}^{\Omega_0}(b|a) = (1+\omega_t) \, p_A^{\Omega_t}(a) \, S_{B|A}^{\Omega_t}(b|a) - \omega_t \tilde{p}_A(a) \, \tilde{S}_B(b) \, . \tag{3.85}$$

This suggests that the term $p_A(a)\, S_{B|A}(b|a)$ in (3.83) can be replaced with:

$$Q_{A,B}(a,b) = (1+\omega)\, p_A(a)\, S_{B|A}(b|a) - \omega \tilde{p}_A(a)\, \tilde{S}_B(b)\,. \tag{3.86}$$

To determine how to modify the $\log S_{B|A}(b|a)$ term in (3.83), we divide both sides of (3.85) by $p_A^{\Omega_0}(a)$ and plug in (3.59) to yield:

$$S_{B|A}^{\Omega_0}(b|a) = \frac{(1+\omega_t)\, p_A^{\Omega_t}(a)\, S_{B|A}^{\Omega_t}(b|a) - \omega_t \tilde{p}_A(a)\, \tilde{S}_B(b)}{(1+\omega_t)\, p_A^{\Omega_t}(a) - \omega_t \tilde{p}_A(a)}. \tag{3.87}$$

This suggests that the term $\log S_{B|A}(b|a)$ in (3.83) can be replaced with $\log\left(Q_{A,B}(a,b)\ /\ q_A(a)\right)$, where $q_A(a)$ is given by (3.75). We can form a modified version of $\mathrm{E}[\varepsilon(B|A)]$ by combining these results:

$$M\mathrm{E}[\varepsilon(B|A)] := -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Q_{A,B}(a,b)\, \log\left(\frac{Q_{A,B}(a,b)}{q_A(a)}\right)\, db\, da. \tag{3.88}$$

This allows us to define modified versions of the JCRE and symmetric JCRE by combining (3.81) and (3.88):

$$MJCRE(A,B) \;:=\; MCRE(A) + M\mathrm{E}[\varepsilon(B|A)]\,, \tag{3.89}$$

$$MJCRE_S(A,B) \;:=\; \frac{1}{2}\Big(MJCRE(A,B) + MJCRE(B,A)\Big). \tag{3.90}$$

3.4.4.2 Dissimilarity Measures Based on Modified Entropy and CRE

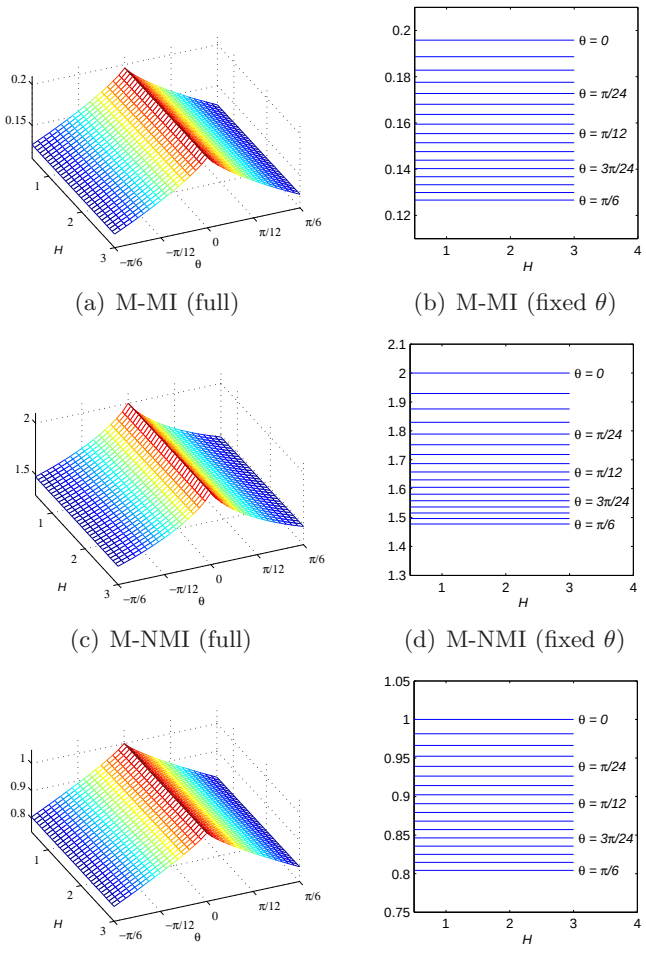
Using the modified entropy (3.73) and modified joint entropy (3.74), we can construct modified versions of MI, NMI, and ECC. The *modified mutual information* (M-MI), *modified normalized mutual information* (M-NMI), and *modified entropy correlation coefficient* (M-ECC) are constructed by replacing (2.10) and (2.11) with (3.73) and (3.74):

$$M\text{-}MI(R,F^{\mathbf{u}}) \;:=\; MH(R) + MH(F^{\mathbf{u}}) - MH(R,F^{\mathbf{u}})\ \ , \tag{3.91}$$

$$M\text{-}NMI(R,F^{\mathbf{u}}) \;:=\; \frac{MH(R) + MH(F^{\mathbf{u}})}{MH(R,F^{\mathbf{u}})}\ \ , \tag{3.92}$$

$$M\text{-}ECC(R,F^{\mathbf{u}}) \;:=\; \sqrt{2 - \frac{2MH(R,F^{\mathbf{u}})}{MH(R) + MH(F^{\mathbf{u}})}}\ \ . \tag{3.93}$$

As can be seen in Fig. 3.12, each of these modified similarity measures is constant with respect to changing overlap size when applied to the half disc example of Fig. 3.6 for a fixed value of θ.



(a) M-MI (full) (b) M-MI (fixed θ)

(c) M-NMI (full) (d) M-NMI (fixed θ)

(e) M-ECC (full) (f) M-ECC (fixed θ)

Figure 3.12: Values of entropy-based dissimilarity measures as functions of rotation angle θ and image width H for the example in Fig. 3.6.

Modified versions of CCRE, SCCRE, NCCRE, and CRECC can be constructed in a similar manner to (3.91)–(3.93). Substituting MCRE and MJCRE for CRE and JCRE in (2.54)–(2.57) yield the *modified cross cumulative residual entropy* (M-CCRE), *modified symmetric cross cumulative residual entropy* (M-SCCRE), *modified normalized cross cumulative residual entropy* (M-NCCRE), and *modified cumulative residual entropy correlation coefficient* (M-CRECC), given by:

$$M\text{-}CCRE(R, F^{\mathbf{u}}) := MCRE(R) + MCRE\varepsilon(F^{\mathbf{u}}) - MJCRE(F^{\mathbf{u}}, R) \quad , \quad (3.94)$$

$$M\text{-}SCCRE(R, F^{\mathbf{u}}) := MCRE(R) + MCRE\varepsilon(F^{\mathbf{u}}) - MJCRE_S(F^{\mathbf{u}}, R) \quad , \quad (3.95)$$

$$M\text{-}NCCRE(R, F^{\mathbf{u}}) := \frac{MCRE(R) + MCRE(F^{\mathbf{u}})}{MJCRE_S(R, F^{\mathbf{u}})} \quad , \quad (3.96)$$

$$M\text{-}CRECC(R, F^{\mathbf{u}}) := \sqrt{2 - \frac{2MCRE_S(R, F^{\mathbf{u}})}{MCRE(R) + MCRE(F^{\mathbf{u}})}} \quad . \quad (3.97)$$

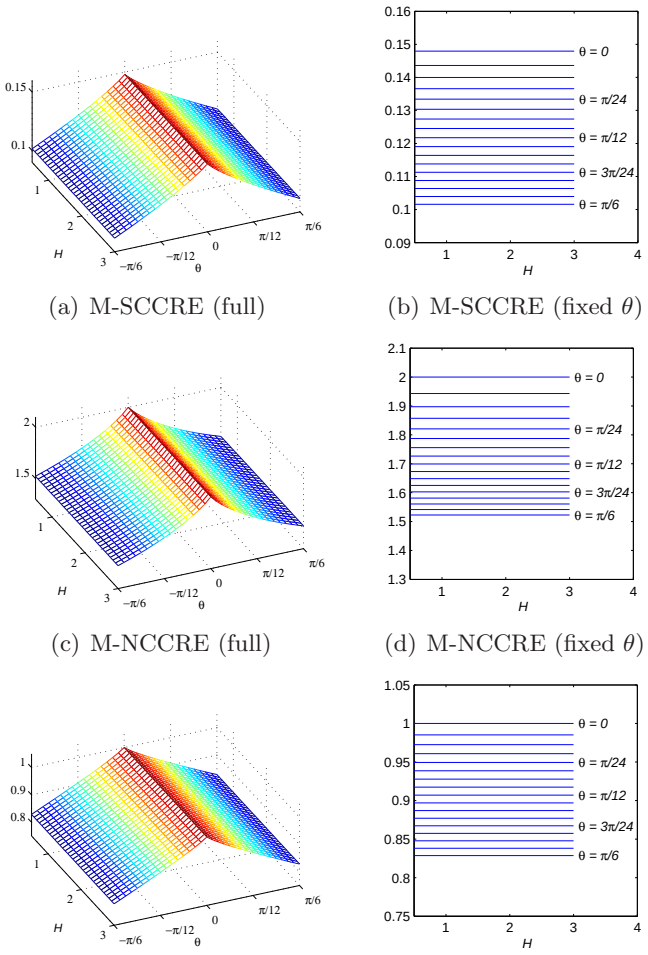
Figure 3.13 shows each of the symmetric modified CRE-based dissimilarity measures as applied to the half disc example of Fig. 3.6. Note that each is constant with respect to changing overlap size for fixed values of θ.

3.5 Rigid Registration Experiment

In order to illustrate the behavior of the modified dissimilarity measures on real-world data, we focus on the rigid registration case, and we use images from the Retrospective Image Registration Evaluation project[1]. The RIRE project database contains CT, MR, and PET images for a variety of patients, and has a sequestered set of ground truth rigid body transformations that were computed from fiducial markers implanted in the skull. (The fiducial markers were removed from the images prior to retrospectively evaluating registration algorithms.) Results of the original RIRE study are provided by West *et al.*[97].

In this thesis, we used the images from nine patient datasets. Each patient dataset contains MR images from some or all of the following protocols: T1-weighted, T2-weighted, PD-weighted, and rectified versions of the T1, T2, and PD-weighted images. Five of the

[1]The images and standard transformations were provided as part of the project, "Retrospective Image Registration Evaluation," National Institutes of Health, Project Number 8R01EB002124-03, Principal Investigator, J. Michael Fitzpatrick, Vanderbilt University, Nashville, TN.



(a) M-SCCRE (full) (b) M-SCCRE (fixed θ)

(c) M-NCCRE (full) (d) M-NCCRE (fixed θ)

(e) M-CRECC (full) (f) M-CRECC (fixed θ)

Figure 3.13: Values of CRE-based dissimilarity measures as functions of rotation angle θ and image width H for the example in Fig. 3.6.

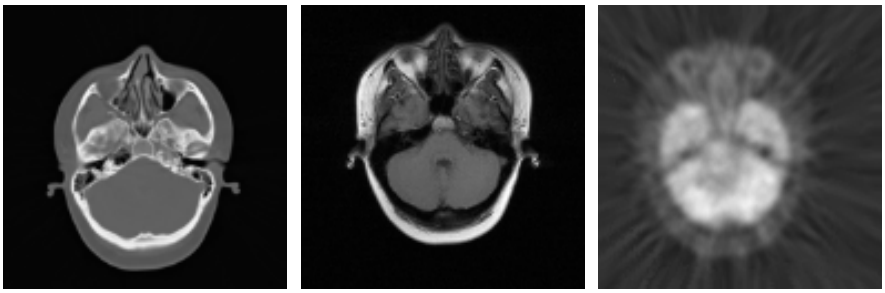


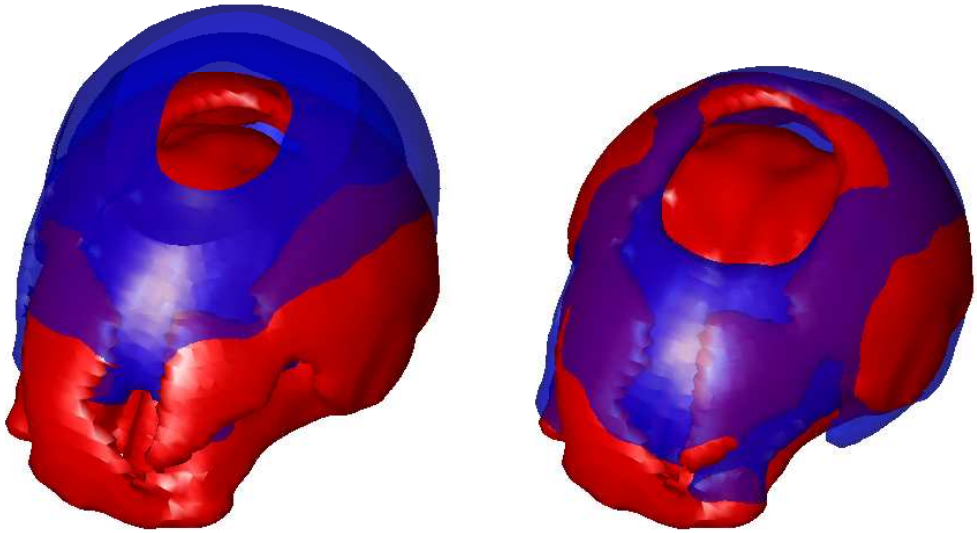
Figure 3.14: Axial slices of CT, MR (T1-weighted) and PET images from the patient 5 dataset.

nine datasets contain both CT and PET images in addition to the MR images. Two of the datasets contain CT but not PET images, and the remaining two datasets contain PET but not CT images. The CT images have resolution $0.65 \times 0.65 \times 4.0$ mm^3, the MR images have approximate resolution $1.25 \times 1.25 \times 4.0$ mm^3, and the PET images have resolution $2.59 \times 2.59 \times 8.0$ mm^3. We performed registration at two isotropic resolutions: $6.0 \times 6.0 \times 6.0$ mm^3 and $3.0 \times 3.0 \times 3.0$ mm^3.

Examples of some of the RIRE images are shown in Fig.'s 3.14 and 3.15. Figure 3.14 shows axial views of the CT, MR-T1, and PET images from patient 5. Figure 3.15 illustrates overlayed isosurfaces of the CT (blue) and MR-T1 images from patient 4, both before (left) and after (right) rigid registration.

All probability densities (and joint densities) were estimated via histograms (and joint histograms) that were constructed with 32 (or 32×32) equally spaced bins. Linear (or bilinear) interpolation was used to accumulate partial weights in neighboring bins. For the modified dissimilarity measures, background distributions for each volume were estimated via histograms that were constructed from the data in manually selected background regions.

To carry out the rigid registration process, we first performed a pre-alignment step by equating the centroids of the floating and reference volumes. We parameterized rigid transformations in terms of Euler angles and translation vectors, and we used the translation from the pre-alignment step as an initial estimate. To constrain the parameters, we employed bounds of $\pm\pi/6$ radians on each Euler angle, and $\pm 1/5$ of the width of the corresponding reference image dimension on each translation component. (We visually verified



(a) Before registration (b) After registration

Figure 3.15: Isosurfaces of CT (blue) and MR-T1 (red) images of patient 4, before and after rigid registration.

for each case that after the prealignment step is performed, the true rigid transformation parameters fall within these bounds.) For each dissimilarity measure, we carried out a bound-constrained optimization procedure using the active set algorithm employed by the `fmincon` function of MATLAB's Optimization Toolbox. All gradient vectors and Hessian matrices were estimated numerically via finite differences. The optimization was terminated when the maximum change in magnitude in any parameter was less than 10^{-4} or after 500 iterations, whichever occured first.

3.5.1 Estimating Background Statistics

Each of the modified similarity measures requires knowledge of the background statistics of the reference and floating images. Probability distribution functions, means, and standard deviations of the background regions were collected by manually selecting background regions from each volume. The estimated background mean and standard deviation values for patient 5 are displayed in Table 3.1.

Computing estimates of the joint p.d.f., covariance, and variance of the conditional mean of the background regions seems problematic, as it would require registered background regions for believable estimates. However, we make the assumption for CT/MR and PET/MR

registration that the background values of the reference and floating images are independent. This allows us to estimate the joint p.d.f. as the outer product of the background p.d.f.'s of the reference and floating image. Furthermore, this yields trivial values for the covariance and variance of the conditional mean of the background regions.

3.5.2 Results

The performance of dissimilarity measures on the various registration tasks is measured via Target Registration Error (TRE). For the RIRE project [97], a number of anatomically meaningful volumes of interest (VOI) were annotated. TRE is computed as the average Euclidean distance (in mm) between VOI centroids in the reference image and their predicted positions after registration.

Tables 3.2–3.3 report the mean, median, and standard deviation of the TRE values measured across every pair of images for every patient for each dissimilarity measure. The three sections of columns refer to registration based on: (a) low resolution ($6.0 \times 6.0 \times 6.0$ mm^3) images with initial estimates provided by aligning image centroids; (b) high resolution ($3.0 \times 3.0 \times 3.0$ mm^3) images with initial estimates provided by aligning image centroids; and, (c) high resolution images with initial estimates provided by the resulting transformations from (a).

When compared to the results reported in the original RIRE paper [97] for techniques that perform registration by optimizing some dissimilarity measure, the results achieved here

Image	$\tilde{\mu}$	$\tilde{\sigma}$
CT	-998.66	4.91
PET	3.95	20.11
MR T1	25.80	13.77
MR T1-Rectified	22.49	10.31
MR T2	27.76	14.81
MR T2-Rectified	23.78	10.92
MR PD	29.66	15.80
MR PD-Rectified	25.54	11.74

Table 3.1: Estimated mean and standard deviation of background regions for the images from the patient 5 dataset. Values reported for the CT volumes are in Hounsfield units; other values are in units of PET or MR code value.

	$6 \times 6 \times 6\,\mathrm{mm}^3$ Resolution			$3 \times 3 \times 3\,\mathrm{mm}^3$ Resolution			Both Resolutions		
Measure	Mean TRE	Median TRE	Std. Dev. TRE	Mean TRE	Median TRE	Std. Dev. TRE	Mean TRE	Median TRE	Std. Dev. TRE
MSD	12.6	12.0	6.3	15.1	14.0	7.0	26.6	24.9	17.5
MMSD	13.0	12.0	5.1	14.2	13.5	5.8	12.9	11.6	4.7
SCC	9.5	8.2	4.8	13.0	12.9	5.8	14.3	11.2	8.1
MSCC	11.3	10.8	3.5	11.3	10.8	3.5	10.9	11.5	3.0
SOCC	8.3	7.4	4.2	11.3	11.0	4.1	12.5	10.9	7.0
MSOCC	14.7	10.9	11.9	23.0	13.5	20.6	20.6	13.7	16.5
CR	6.6	6.3	3.1	6.1	5.3	3.2	5.8	5.1	4.0
MCR	16.3	16.0	6.8	15.5	15.1	6.7	16.9	16.9	7.0
MI	3.5	3.1	2.0	2.0	1.8	1.0	1.8	1.8	0.8
MMI	3.1	2.8	1.7	2.0	1.8	1.0	1.9	1.6	1.0
NMI	4.7	3.7	3.0	2.3	2.4	1.4	2.3	2.1	1.2
MNMI	4.7	3.6	3.3	2.4	2.3	1.5	2.2	2.1	1.3
ECC	4.6	3.6	3.1	2.3	2.0	1.3	2.3	2.1	1.1
MECC	4.6	3.6	3.3	2.3	2.2	1.2	2.3	2.3	1.1
SCCRE	6.0	4.8	3.7	6.4	4.6	5.9	4.7	4.2	3.2
MSCCRE	4.3	3.4	3.3	2.6	2.5	1.4	2.6	2.6	1.3
NCCRE	4.6	3.7	2.8	3.2	3.0	1.7	2.8	2.7	1.6
MNCCRE	4.3	3.5	2.7	2.9	2.8	1.6	2.7	2.5	1.3
CRECC	4.5	3.6	2.6	3.4	3.0	1.9	2.6	2.6	1.3
MCRECC	4.1	3.3	2.7	3.2	2.8	2.0	2.7	2.6	1.4

Table 3.2: Statistics of TRE (in mm) for CT/MR registration, aggregated across all patients.

	$6 \times 6 \times 6\,\mathrm{mm}^3$ Resolution			$3 \times 3 \times 3\,\mathrm{mm}^3$ Resolution			Both Resolutions		
Measure	Mean TRE	Median TRE	Std. Dev. TRE	Mean TRE	Median TRE	Std. Dev. TRE	Mean TRE	Median TRE	Std. Dev. TRE
MSD	5.8	3.7	4.7	5.3	3.1	5.8	6.4	3.8	7.8
MMSD	10.1	7.3	8.6	9.7	6.1	8.9	11.6	7.3	13.1
SCC	10.0	8.8	6.2	8.2	4.6	7.8	9.6	6.0	9.8
MSCC	7.4	7.0	4.8	5.6	3.2	5.9	6.8	5.1	5.0
SOCC	7.8	6.5	4.8	7.0	4.0	7.6	7.4	4.3	7.2
MSOCC	5.8	4.2	3.7	5.6	3.0	5.4	5.5	3.9	5.2
CR	10.9	7.9	9.1	14.3	11.2	9.3	13.4	9.3	13.1
MCR	11.1	9.9	6.6	10.2	10.5	4.8	11.9	10.7	6.8
MI	5.7	4.7	4.8	6.6	3.1	8.3	4.5	3.3	4.9
MMI	4.2	3.1	2.5	3.3	2.8	1.9	3.1	2.6	1.9
NMI	4.1	3.5	2.2	3.2	2.8	1.7	3.3	2.9	1.8
MNMI	3.9	3.0	2.4	3.5	2.8	2.2	3.1	2.5	2.0
ECC	4.1	3.1	2.3	3.1	2.8	1.6	3.2	2.9	1.8
MECC	4.1	3.1	2.3	3.4	2.6	2.0	3.1	2.7	2.1
SCCRE	5.2	4.7	2.6	6.3	4.3	4.7	5.0	4.4	2.9
MSCCRE	4.7	3.8	2.5	3.6	3.0	2.3	3.6	3.2	1.9
NCCRE	4.4	3.9	2.6	3.6	3.4	2.2	3.9	3.2	2.6
MNCCRE	4.3	3.3	2.4	3.4	2.8	2.4	3.3	2.6	2.0
CRECC	4.4	3.8	2.5	3.6	3.0	2.5	3.8	3.2	2.3
MCRECC	4.3	3.3	2.6	3.5	3.0	2.4	3.3	2.8	2.0

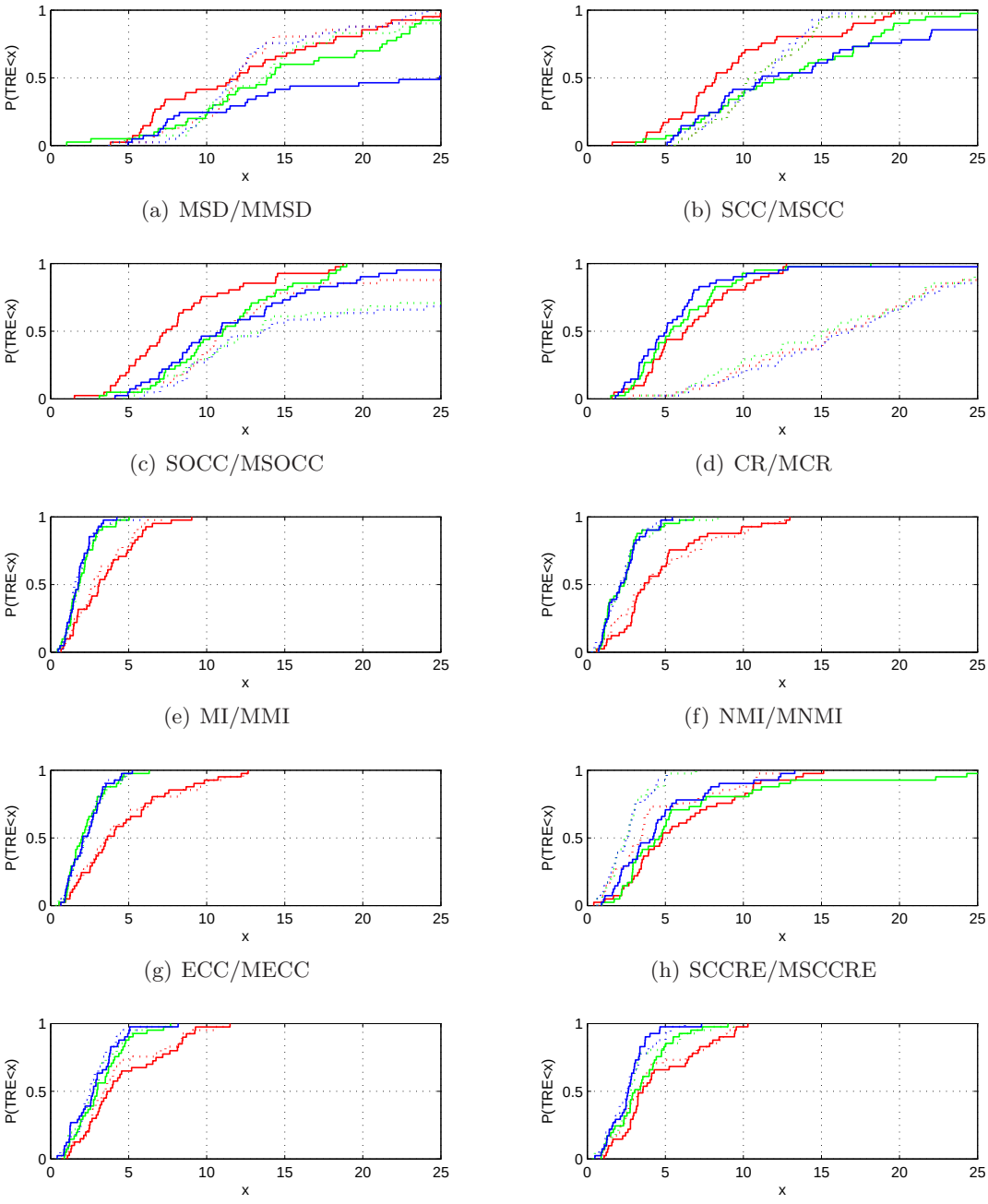
Table 3.3: Statistics of TRE (in mm) for PET/MR registration, aggregated across all patients.

by performing registration on the high resolution images using dissimilarity measures based on the the information content assumptions (namely, MI, NMI, ECC, SCCRE, NCCRE, CRECC, and their modified versions) seem marginally worse. However, we note that our median TRE values with these dissimilarity measures are in general less than voxel width. If we further performed registration at the original image resolution, it is likely that we would achieve results competitive with those in the original paper. As a side note, as of the date of this thesis, the RIRE images and data as well as up-to-date results are still available online at `http://www.insight-journal.org/rire/`.

For a visual interpretation of these results, we compute and plot empirical estimates of the TRE distribution for each dissimilarity measure. Figures 3.16–3.17 illustrate these distributions for CT/MR and PET/MR registration, respectively. To interpret these figures, note that dissimilarity measures with better TRE performance will have TRE distributions that rise more rapidly than those with worse TRE performance. In both the CT/MR and PET/MR registration cases, it is clear that registration with dissimilarity measures that are based on the information content assumptions yield more rapidly increasing TRE distributions than registration with the other dissimilarity measures. This is to be expected, as the information content assumption is the most readily applicable assumption for the multimodal registration tasks in question.

It is clear from Fig. 3.16(h) that the use of MSCCRE for CT/MR provides a significant improvement in the TRE distribution over the use of SCCRE. For MR/PET registration, Fig.'s 3.17(e) and (h) appear to indicate a slight improvement in the use of MMI/MSCCRE for PET/MR registration over MI/SCCRE, respectively. It is also clear that in CT/MR registration, the use of higher resolution volumes provides a significant improvement over registration with only low resolution volumes for all of the information content based dissimilarity measures (except possibly SCCRE).

The benefits of the modified versions of the other information content based dissimilarity measures are not as apparent from Fig.'s 3.16 and 3.17; however, we can establish what differences are statistically significant via hypothesis testing. Given two sets of TRE's, S_1 and S_2, we define the null hypothesis H_0 to state that the TRE's from S_1 and S_2 arise from



(a) MSD/MMSD

(b) SCC/MSCC

(c) SOCC/MSOCC

(d) CR/MCR

(e) MI/MMI

(f) NMI/MNMI

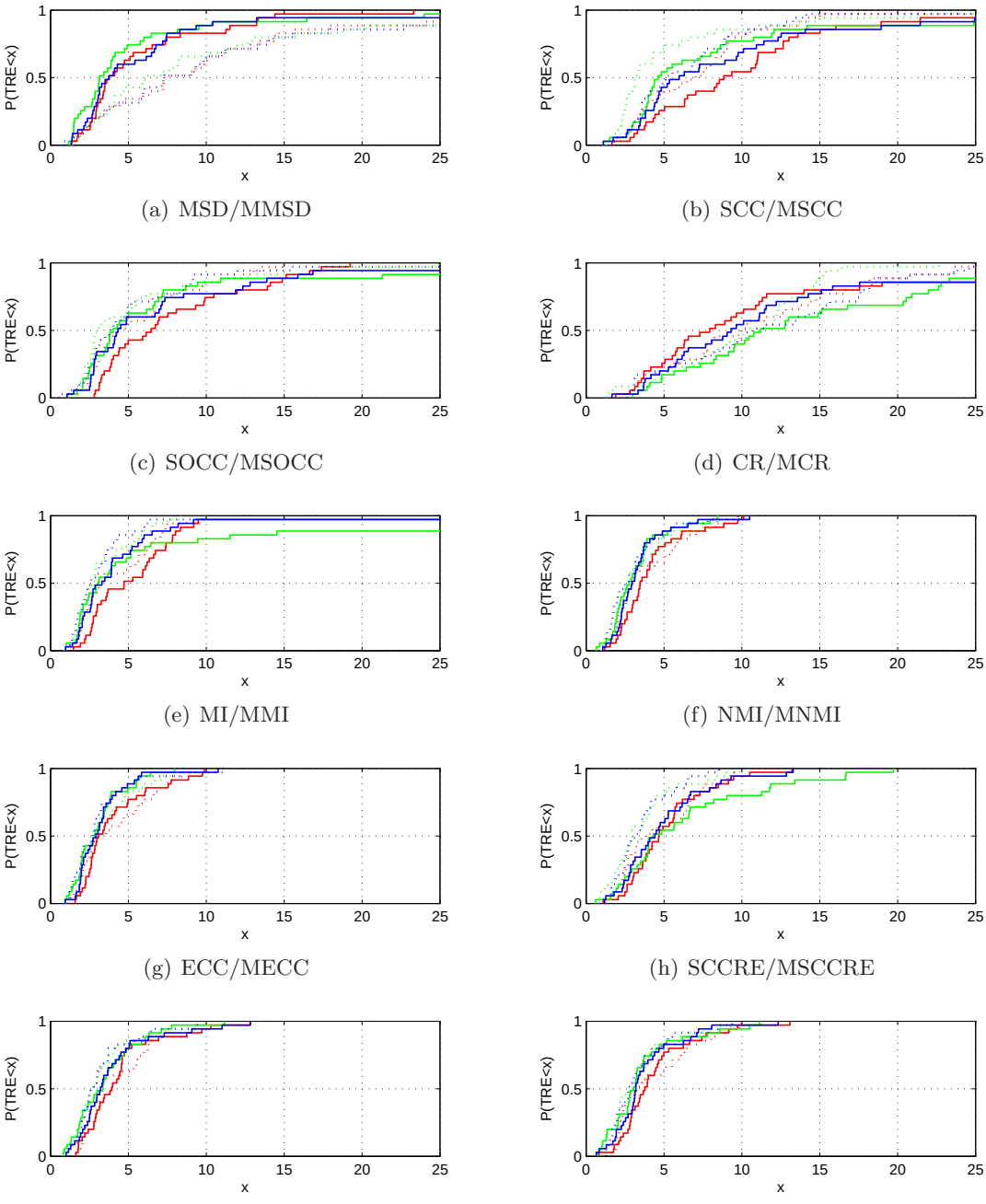
(g) ECC/MECC

(h) SCCRE/MSCCRE

(i) NCCRE/MNCCRE

(j) CRECC/MCRECC

Figure 3.16: Empirical cumulative distribution functions of TRE (in mm) for each dissimilarity measure (solid lines) and each modified dissimilarity measure (dashed lines) for CT/MR registration of RIRE images. Colors indicate use of images of low resolution (red), high resolution (green), or both resolutions (blue) for registration.



(a) MSD/MMSD

(b) SCC/MSCC

(c) SOCC/MSOCC

(d) CR/MCR

(e) MI/MMI

(f) NMI/MNMI

(g) ECC/MECC

(h) SCCRE/MSCCRE

(i) NCCRE/MNCCRE

(j) CRECC/MCRECC

Figure 3.17: Empirical cumulative distribution functions of TRE (in mm) for each dissimilarity measure (solid lines) and each modified dissimilarity measure (dashed lines) for PET/MR registration of RIRE images. Colors indicate use of images of low resolution (red), high resolution (green), or both resolutions (blue) for registration.

the same underlying distribution. We define the alternative hypothesis H_A to state that the TRE's from S_1 are "worse" than the TRE's from S_2, in the sense that the cumulative distribution values are smaller everywhere. The two-sample Kolmogorov-Smirnov test can be used to determine whether or not H_0 can be rejected in favor of H_A. Comparisons of dissimilarity measures for CT/MR and PET/MR registration are illustrated in Tables 3.4–3.5.

3.6 Gâteaux Derivatives of Modified Similarity Measures

In order to use modified dissimilarity measures in deformable registration, we need to determine their Gâteaux derivatives. This can be done in a straightforward task by extending the analysis of Section 3.1.2. As in Section 3.1.2, we present the resulting Gâteaux derivatives for the modified general constancy dissimilarity measure and for the M-MI. The explanations herein should serve as a framework for the interested reader to compute Gâteaux derivatives for the other modified dissimilarity measures.

3.6.1 Modified General Constancy / M-MSD

The Gâteaux derivative of the modified general constancy dissimilarity measure (3.52) can be found by applying the product rule and simplifying:

$$dM\text{-}\mathcal{I}^{\Omega}_{\Psi,r}(R, F^{\mathbf{u}}; \mathbf{w}) = (1+\omega)\, d\mathcal{I}^{\Omega}_{\Psi,r}(R, F^{\mathbf{u}}; \mathbf{w}) + \left(\mathcal{I}^{\Omega}_{\Psi,r}(R, F^{\mathbf{u}}) - \tilde{\mu}_{\Psi(r(\mathbf{x};R,F^{\mathbf{u}}))}\right) d\omega(\mathbf{u}, \mathbf{w})\,. \tag{3.98}$$

Now, ω_t can be written in terms of (3.1) by $\omega_t = \left(\mathrm{C}(\mathbf{u}) - \mathrm{C}(\mathbf{0})\right)/\mathrm{C}(\mathbf{0})$, so it is clear that:

$$d\omega(\mathbf{u}, \mathbf{w}) = \frac{d\mathrm{C}(\mathbf{u}; \mathbf{w})}{\mathrm{C}(\mathbf{0})}. \tag{3.99}$$

Hence, (3.98) can be written as:

$$dM\text{-}\mathcal{I}^{\Omega}_{\Psi,r}(R, F^{\mathbf{u}}; \mathbf{w}) = (1+\omega)\, d\mathcal{I}^{\Omega}_{\Psi,r}(R, F^{\mathbf{u}}; \mathbf{w}) + \left(\frac{\mathcal{I}^{\Omega}_{\Psi,r}(R, F^{\mathbf{u}}) - \tilde{\mu}_{\Psi(r(\mathbf{x};R,F^{\mathbf{u}}))}}{\mathrm{C}(\mathbf{0})}\right) d\mathrm{C}(\mathbf{u}, \mathbf{w})\,, \tag{3.100}$$

with $d\mathcal{I}^{\Omega}_{\Psi,r}$ and $d\mathrm{C}(\mathbf{u}, \mathbf{w})$ given by (3.10) and (3.5), respectively.

From (3.100), we can immediately establish that the Gâteaux derivative of the modified

Measure	MSD	MMSD	SCC	MSCC	SOCC	MSOCC	CR	MCR	MI	MMI	NMI	MNMI	ECC	MECC	SCCRE	MSCCRE	NCCRE	MNCCRE	CRECC	MCRECC
MSD	–	>	>	>	>	>	>	>	>	>	>	>	>	>	>	>	>	>	>	>
MMSD	<	–	>	∅	>	<	>	<	>	>	>	>	>	>	>	>	>	>	>	>
SCC	<	<	–	>	∅	∅	>	<	>	>	>	>	>	>	>	>	>	>	>	>
MSCC	<	∅	<	–	∅	<	>	<	>	>	>	>	>	>	>	>	>	>	>	>
SOCC	<	<	∅	∅	–	<	>	<	>	>	>	>	>	>	>	>	>	>	>	>
MSOCC	<	>	∅	>	>	–	>	∅	>	>	>	>	>	>	>	>	>	>	>	>
CR	<	<	<	<	<	<	–	<	>	>	>	>	>	>	∅	>	>	>	>	>
MCR	<	>	>	>	>	∅	>	–	>	>	>	>	>	>	>	>	>	>	>	>
MI	<	<	<	<	<	<	<	<	–	∅	<	∅	<	<	<	<	<	<	<	<
MMI	<	<	<	<	<	<	<	<	∅	–	<	∅	<	∅	<	<	<	<	<	<
NMI	<	<	<	<	<	<	<	<	>	>	–	∅	∅	∅	<	∅	∅	∅	∅	∅
MNMI	<	<	<	<	<	<	<	<	∅	∅	∅	–	∅	∅	<	∅	∅	∅	∅	∅
ECC	<	<	<	<	<	<	<	<	>	>	∅	∅	–	∅	<	∅	∅	∅	∅	∅
MECC	<	<	<	<	<	<	<	<	>	∅	∅	∅	∅	–	<	∅	∅	∅	∅	∅
SCCRE	<	<	<	<	<	<	∅	<	>	>	>	>	>	>	–	>	>	>	>	>
MSCCRE	<	<	<	<	<	<	<	<	>	>	∅	∅	∅	∅	<	–	∅	∅	∅	∅
NCCRE	<	<	<	<	<	<	<	<	>	>	∅	∅	∅	∅	<	∅	–	∅	∅	∅
MNCCRE	<	<	<	<	<	<	<	<	>	>	∅	∅	∅	∅	<	∅	∅	–	∅	∅
CRECC	<	<	<	<	<	<	<	<	>	>	∅	∅	∅	∅	<	∅	∅	∅	–	∅
MCRECC	<	<	<	<	<	<	<	<	>	>	∅	∅	∅	∅	<	∅	∅	∅	∅	–

Table 3.4: Outcome of two-sample Kolmogorov-Smirnov test ($\alpha = 0.05$) comparing pairs of dissimilarity measures for CT/MR registration using both low and high resolution images. '<' / '>' indicates that the dissimilarity measure of the current row has smaller/larger errors than the dissimilarity measure of the current column. '∅' indicates that the null hypothesis cannot be rejected.

Measure	MSD	MMSD	SCC	MSCC	SOCC	MSOCC	CR	MCR	MI	MMI	NMI	MNMI	ECC	MECC	SCCRE	MSCCRE	NCCRE	MNCCRE	CRECC	MCRECC
MSD	–	<	<	∅	∅	∅	<	<	∅	>	>	>	>	>	∅	∅	∅	>	∅	>
MMSD	>	–	∅	>	>	>	∅	∅	>	>	>	>	>	>	>	>	>	>	>	>
SCC	>	∅	–	∅	∅	>	∅	<	>	>	>	>	>	>	>	>	>	>	>	>
MSCC	∅	<	∅	–	∅	∅	<	<	>	>	>	>	>	>	∅	>	>	>	>	>
SOCC	∅	<	∅	∅	–	∅	<	<	>	>	>	>	>	>	∅	>	>	>	>	>
MSOCC	∅	<	<	∅	∅	–	<	<	∅	>	>	>	>	>	∅	∅	∅	>	∅	>
CR	>	∅	∅	>	>	>	–	∅	>	>	>	>	>	>	>	>	>	>	>	>
MCR	>	∅	>	>	>	>	∅	–	>	>	>	>	>	>	>	>	>	>	>	>
MI	∅	<	<	<	<	∅	<	<	–	∅	∅	∅	∅	∅	∅	∅	∅	∅	∅	∅
MMI	<	<	<	<	<	<	<	<	∅	–	∅	∅	∅	∅	<	∅	∅	∅	∅	∅
NMI	<	<	<	<	<	<	<	<	∅	∅	–	>	∅	∅	<	∅	∅	∅	∅	∅
MNMI	<	<	<	<	<	<	<	<	∅	∅	<	–	∅	∅	<	∅	∅	∅	∅	∅
ECC	<	<	<	<	<	<	<	<	∅	∅	∅	∅	–	∅	<	∅	∅	∅	∅	∅
MECC	<	<	<	<	<	<	<	<	∅	∅	∅	∅	∅	–	<	∅	∅	∅	<	∅
SCCRE	∅	<	<	∅	∅	∅	<	<	∅	>	>	>	>	>	–	>	∅	>	>	>
MSCCRE	∅	<	<	<	<	∅	<	<	∅	∅	∅	∅	∅	∅	<	–	∅	∅	∅	∅
NCCRE	∅	<	<	<	<	∅	<	<	∅	∅	∅	∅	∅	∅	∅	∅	–	∅	∅	∅
MNCCRE	<	<	<	<	<	<	<	<	∅	∅	∅	∅	∅	∅	<	∅	∅	–	∅	∅
CRECC	∅	<	<	<	<	∅	<	<	∅	∅	∅	∅	∅	>	<	∅	∅	∅	–	∅
MCRECC	<	<	<	<	<	<	<	<	∅	∅	∅	∅	∅	∅	<	∅	∅	∅	∅	–

Table 3.5: Outcome of two-sample Kolmogorov-Smirnov test ($\alpha = 0.05$) comparing pairs of dissimilarity measures for PET/MR registration using both low and high resolution images. '<' / '>' indicates that the dissimilarity measure of the current row has smaller/larger errors than the dissimilarity measure of the current column. '∅' indicates that the null hypothesis cannot be rejected.

MSD is given by:

$$dM\text{-}MSD(R, F^{\mathbf{u}}; \mathbf{w}) = (1+\omega)\, dMSD(R, F^{\mathbf{u}}; \mathbf{w}) + \left(\frac{MSD(R, F^{\mathbf{u}}) - \tilde{\mu}_{(R(\mathbf{x})-F^{\mathbf{u}}(\mathbf{x}))^2}}{\mathrm{C}(\mathbf{0})} \right) d\mathrm{C}(\mathbf{u}, \mathbf{w})\,. \tag{3.101}$$

Hence, we can write:

$$dM\text{-}MSD(R, F^{\mathbf{u}}; \mathbf{w}) = \int_{\mathbb{R}^n} \langle \mathcal{P}_{M\text{-}MSD}(\mathbf{x}; R, F^{\mathbf{u}})\,, \mathbf{w}(\mathbf{x}) \rangle\, d\mathbf{x}, \tag{3.102}$$

where

$$\mathcal{P}_{M\text{-}MSD}(\mathbf{x}; R, F^{\mathbf{u}}) = (1+\omega)\, \mathcal{P}_{MSD}(R, F^{\mathbf{u}}) + \left(\frac{MSD(R, F^{\mathbf{u}}) - \tilde{\mu}_{(R(\mathbf{x})-F^{\mathbf{u}}(\mathbf{x}))^2}}{\mathrm{C}(\mathbf{0})} \right) \mathcal{P}_{\mathrm{C}}(\mathbf{x}; R, F^{\mathbf{u}})\,, \tag{3.103}$$

with $\mathcal{P}_{MSD}$ and $\mathcal{P}_C$ given by (3.12) and (3.6), respectively.

3.6.2 Modified Mutual Information

Using analysis similar to that in Section 3.1.2.2, we find that the Gâteaux derivatives of the modified marginal and joint entropies are given by:

$$dMH_R(\mathbf{u}, \mathbf{w}) = \int_{\mathbb{R}^n} \langle \mathcal{P}_{MH}(\mathbf{x}; \mathbf{u}, R)\,, \mathbf{w}(\mathbf{x}) \rangle\, d\mathbf{x}, \tag{3.104}$$

$$dMH_{F^{\mathbf{u}}}(\mathbf{w}) = \int_{\mathbb{R}^n} \langle \mathcal{P}_{MH}(\mathbf{x}; F^{\mathbf{u}})\,, \mathbf{w}(\mathbf{x}) \rangle\, d\mathbf{x}, \tag{3.105}$$

and

$$dMH_{R,F^{\mathbf{u}}}(\mathbf{w}) = \int_{\mathbb{R}^n} \langle \mathcal{P}_{MH}(\mathbf{x}; R, F^{\mathbf{u}})\,, \mathbf{w}(\mathbf{x}) \rangle\, d\mathbf{x}, \tag{3.106}$$

where

$$\mathcal{P}_{MH}(\mathbf{x}; \mathbf{u}, R) \approx \mathrm{C}(\mathbf{0})^{-1} \cdot \mathrm{H}_\epsilon(\Theta_R(\mathbf{x})) \cdot \mathrm{H}'_\epsilon(\Theta_{F^{\mathbf{u}}}(\mathbf{x})) \cdot \left(\log q_R(R(\mathbf{x})) - \int_{-\infty}^{\infty} \tilde{p}_R(r) \log q_R(r)\, dr \right) \cdot \nabla \Theta_{F^{\mathbf{u}}}(\mathbf{x})\,, \tag{3.107}$$

$$\begin{aligned} \mathcal{P}_{MH}(\mathbf{x}; F^{\mathbf{u}}) \approx\ & \mathrm{C}(\mathbf{0})^{-1} \cdot \mathrm{H}_\epsilon(\Theta_R(\mathbf{x})) \cdot \mathrm{H}'_\epsilon(\Theta_{F^{\mathbf{u}}}(\mathbf{x})) \\ & \cdot \left(\log q_{F^{\mathbf{u}}}(F^{\mathbf{u}}(\mathbf{x})) - \int_{-\infty}^{\infty} \tilde{p}_{F^{\mathbf{u}}}(f) \log q_{F^{\mathbf{u}}}(f)\, df \right) \cdot \nabla \Theta_{F^{\mathbf{u}}}(\mathbf{x}) \\ & + \mathrm{H}_\epsilon(\Theta_R(\mathbf{x})) \cdot \mathrm{H}_\epsilon(\Theta_{F^{\mathbf{u}}}(\mathbf{x})) \cdot \left(\frac{q'_{F^{\mathbf{u}}}(F^{\mathbf{u}}(\mathbf{x}))}{q_{F^{\mathbf{u}}}(F^{\mathbf{u}}(\mathbf{x}))\, \mathrm{C}(\mathbf{0})} \right) \cdot \nabla F^{\mathbf{u}}(\mathbf{x})\,, \end{aligned} \tag{3.108}$$

and

$$\begin{aligned}
&\mathcal{P}_{MH}(\mathbf{x}; R, F^{\mathbf{u}}) \approx \\
&\quad \mathrm{C}(\mathbf{0})^{-1} \cdot \mathrm{H}_\epsilon(\Theta_R(\mathbf{x})) \cdot \mathrm{H}'_\epsilon(\Theta_{F^{\mathbf{u}}}(\mathbf{x})) \\
&\quad \cdot \left(\log q_{R,F^{\mathbf{u}}}(R(\mathbf{x}), F^{\mathbf{u}}(\mathbf{x})) - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{p}_R(r)\, \tilde{p}_{F^{\mathbf{u}}}(f) \log q_{R,F^{\mathbf{u}}}(r,f)\, df dr \right) \cdot \nabla\Theta_{F^{\mathbf{u}}}(\mathbf{x}) \\
&\quad + \mathrm{H}_\epsilon(\Theta_R(\mathbf{x})) \cdot \mathrm{H}_\epsilon(\Theta_{F^{\mathbf{u}}}(\mathbf{x})) \cdot \left(\frac{\frac{\partial}{\partial f}\left[q_{R,F^{\mathbf{u}}}(R(\mathbf{x}), f)\right]_{f=F^{\mathbf{u}}(\mathbf{x})}}{q_{R,F^{\mathbf{u}}}(R(\mathbf{x}), F^{\mathbf{u}}(\mathbf{x}))\, \mathrm{C}(\mathbf{0})} \right) \cdot \nabla F^{\mathbf{u}}(\mathbf{x}).
\end{aligned} \tag{3.109}$$

Combining these terms yields the Gâteaux derivative of the modified mutual information:

$$\begin{aligned}
dM\text{-}MI(R, F^{\mathbf{u}}; \mathbf{w}) &= dMH_R(\mathbf{u}, \mathbf{w}) + dMH_{F^{\mathbf{u}}}(\mathbf{w}) - dMH_{R,F^{\mathbf{u}}}(\mathbf{w}) \\
&= \int_{\mathbb{R}^n} \left\langle \mathcal{P}_{M\text{-}MI}(\mathbf{x}; R, F^{\mathbf{u}}), \mathbf{w}(\mathbf{x}) \right\rangle d\mathbf{x},
\end{aligned} \tag{3.110}$$

where

$$\begin{aligned}
&\mathcal{P}_{M\text{-}MI}(\mathbf{x}; R, F^{\mathbf{u}}) \approx \\
&\left(\frac{\mathrm{H}_\epsilon(\Theta_R(\mathbf{x})) \cdot \mathrm{H}'_\epsilon(\Theta_{F^{\mathbf{u}}}(\mathbf{x}))}{\mathrm{C}(\mathbf{0})} \right) \left(\log \left(\frac{q_R(R(\mathbf{x}))\, q_{F^{\mathbf{u}}}(F^{\mathbf{u}}(\mathbf{x}))}{q_{R,F^{\mathbf{u}}}(R(\mathbf{x}), F^{\mathbf{u}}(\mathbf{x}))} \right) + Y(R, F^{\mathbf{u}}) \right) \nabla\Theta_{F^{\mathbf{u}}}(\mathbf{x}) \\
&+ \left(\frac{\mathrm{H}_\epsilon(\Theta_R(\mathbf{x})) \cdot \mathrm{H}_\epsilon(\Theta_{F^{\mathbf{u}}}(\mathbf{x}))}{\mathrm{C}(\mathbf{0})} \right) \left(\frac{q'_{F^{\mathbf{u}}}(F^{\mathbf{u}}(\mathbf{x}))}{q_{F^{\mathbf{u}}}(F^{\mathbf{u}}(\mathbf{x}))} - \frac{\frac{\partial}{\partial f}\left[q_{R,F^{\mathbf{u}}}(R(\mathbf{x}), f)\right]_{f=F^{\mathbf{u}}(\mathbf{x})}}{q_{R,F^{\mathbf{u}}}(R(\mathbf{x}), F^{\mathbf{u}}(\mathbf{x}))} \right) \nabla F^{\mathbf{u}}(\mathbf{x})
\end{aligned} \tag{3.111}$$

and where

$$\begin{aligned}
Y(R, F^{\mathbf{u}}) = &- \int_{-\infty}^{\infty} \tilde{p}_R(r) \log q_R(r)\, dr - \int_{-\infty}^{\infty} \tilde{p}_{F^{\mathbf{u}}}(f) \log q_{F^{\mathbf{u}}}(f)\, df \\
&+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{p}_R(r)\, \tilde{p}_{F^{\mathbf{u}}}(f) \log q_{R,F^{\mathbf{u}}}(r,f)\, df dr.
\end{aligned} \tag{3.112}$$

3.7 Summary

This chapter focused on the effect of changing overlap on the dissimilarity measures presented in Chapter 2. It first proposed a technique to account for changing overlap in the dissimilarity measures and their Gâteaux derivatives by using regularized Heaviside functions to describe regions of valid data. It then illustrated the overlap sensitivity problem that arises due to information moving in and out of the overlap region, and it proposed a

framework for correcting this problem by modifying the dissimilarity measures according to the image statistics in background regions. Finally, it presented an experiment on registering multimodal brain images in order to compare and contrast the behavior of different dissimilarity measures and their modified versions.

Chapter 4

Regularizers

If any of the dissimilarity measures defined in Chapters 2 and 3 are optimized with respect to a nonparametric displacement field, there is no guarantee of a unique solution. In fact, there may be infinitely many solutions, most of which exhibit no reasonable amount of continuity or smoothness. In order to guarantee that the nonparametric registration problem is well-posed and has a smooth solution, a regularization term must be added to the dissimilarity measure.

In medical image registration, various regularizers have been proposed. Modersitzki [63] provides a taxonomy of quadratic regularizers, referring to them as elastic, fluid, diffusion, and curvature regularizers. Each of these regularizers are homogeneous and isotropic, meaning that they ensure smoothness independently of location or direction. In the computer vision community, the regularizer used by Horn & Schunck [51] (which is equivalent to the diffusion regularizer) serves as the basis for dense optical flow estimation, and has been generalized to the nonhomogeneous and anisotropic regularizers of the Weickert taxonomy [95]. Although both the quadratic and Weickert taxonomies can be thought of as being based on generalizations of the diffusion regularizer, the taxonomies follow orthogonal paths. Aside from the nonhomogeneous elastic regularizer presented by Kabus [53], there appears to have been little work done in extending the elastic, fluid, and curvature regularizers in nonhomogeneous and/or anisotropic ways.

This chapter details the various types of regularizers in both the quadratic and Weickert taxonomies. It extends the quadratic taxonomy by adding the *second order elastic* regular-

izer in order to provide a connection with the family of second-order rotationally invariant regularizers. It also extends the Weickert taxonomy in the manner of Kabus [53] to allow the weighting functions in image-driven regularizers to deform according to the displacement field. It then illustrates how the ideas behind the nonhomogeneous regularizers in Weickert's taxonomy can be applied to all of the quadratic regularizers. Finally, it presents the Gâteaux derivatives of all of the regularizers, yielding differential operators that can be considered generalizations of the operators that have emerged out of the work of Grenander and Miller [46] and Christensen [28] in the field of computational anatomy.

4.1 Homogeneous Regularizers

The regularizers commonly used in medical image registration are typically homgeneous and isotropic. In this section, we describe the quadratic taxonomy of regularizers, and then we show how the addition of what we call the *second-order elastic* regularizer provides a basis for a family of regularizers.

Every regularizer in this chapter will be represented by the notation $\mathcal{R}(\mathbf{u})$ to indicate that it is applied to the displacement field $\mathbf{u}$ of the deformation.

4.1.1 Quadratic Taxonomy

In [63], the diffusion, elastic, fluid, and curvature regularizers are theoretically motivated and derived. In this subsection, we briefly present each of these regularizers and mention their advantages.

4.1.1.1 Diffusion

The *diffusion* regularizer is designed to minimize components of the displacement field:

$$\mathcal{R}_{diffusion}(\mathbf{u}) = \frac{1}{2}\int_{\Omega}\left(\sum_{j=1}^{n}\|\nabla u_j\|^2\right)d\mathbf{x}. \tag{4.1}$$

It gets its name because of the diffusion equation that arises when the Gâteaux derivatives of the dissimilarity measure and regularizer are equated. One of the earliest references where the diffusion regularizer formally appears for use in medical image registration is [39], although Modersitzki shows in [63] that the earlier work done by Thirion [86] can be

thought of as a form of diffusion-based registration. The diffusion regularizer is equivalent to the much earlier regularizer proposed by Horn and Schunck [51] that is widely used in the field of optical flow.

The advantage of the diffusion regularizer over the other homogeneous regularizers is mainly a computational one. A fast algorithm based on Additive Operator Splitting [63] can be used to solve the diffusion equation but cannot be used to solve the partial differential equations arising from the other regularizers. The disadvantages of using the diffusion regularizer for registration are that it does not couple the spatial directions of the deformation field (which would be physically appropriate), and that it is not immune to initial affine misregistrations.

4.1.1.2 Elastic/Fluid

The *elastic* regularizer measures the linearized elastic potential of the displacement field:

$$\mathcal{R}_{elastic}(\mathbf{u}) = \frac{1}{2}\int_{\Omega}\left(\frac{\mu}{2}\sum_{j=1}^{n}\sum_{k=1}^{n}\left(\partial_{x_j}u_k + \partial_{x_k}u_j\right)^2 + \lambda\left(\operatorname{div}\mathbf{u}\right)^2\right)d\mathbf{x}, \tag{4.2}$$

where μ and λ are the Lamé constants which are based on properties of an elastic body.

The *fluid* regularizer takes the same form as the elastic regularizer, but applied to the velocity field instead of the displacement field; i.e.,

$$\mathcal{R}_{fluid}(\mathbf{u}) = \mathcal{R}_{elastic}(\mathbf{v})\,, \tag{4.3}$$

where the velocity field is related to the displacement field by:

$$\mathbf{v}\left(\mathbf{x},t\right) = \frac{d}{dt}\mathbf{u}\left(\mathbf{x},t\right) = \partial_t\mathbf{u}\left(\mathbf{x},t\right) + \left(\nabla\mathbf{u}\left(\mathbf{x},t\right)\right)^{\mathrm{T}}\mathbf{v}\left(\mathbf{x},t\right). \tag{4.4}$$

Early use of the elastic and fluid regularizers for image registration appeared in [10, 14, 30, 31, 36]. Other examples of these regularizers in the research literature can be found in [13, 20, 35, 41, 63, 81].

The elastic and fluid regularizers have the advantage over the diffusion and curvature regularizers (the curvature regularizer will be defined in the next section) that they appropriately handle coupling of the spatial directions in the displacement field. Since the

physical deformations of anatomical objects that are observed in medical images almost always contain nontrivial terms based on mixed partial derivatives, this property suggests that the elastic and fluid regularizers might allow nonparametric registration to converge to a more physically meaningful solution. However, elastic and fluid regularizers also suffer the disadvantage of the diffusion regularizer of not being immune to initial affine misregistrations.

For the remainder of this chapter, we will ignore the fluid regularizer as part of the quadratic taxonomy. As will be described in Chapter 5, fluid regularization can be thought of as a generalization of elastic regularization for large-deformation problems. Furthermore, we will see that this type of generalization can be realized by *any* of the other regularizers by applying them to the velocity field instead of the displacement field.

4.1.1.3 Curvature

The *curvature* regularizer is based on the second derivatives of the displacement field:

$$\mathcal{R}_{curvature}(\mathbf{u}) = \frac{1}{2}\int_{\Omega}\left(\sum_{j=1}^{n}(\Delta u_j)^2\right)d\mathbf{x}. \tag{4.5}$$

The curvature regularizer has been used for image registration in [40, 63]. Like the diffusion regularizer, it has the disadvantage of ignoring any coupling of the spatial directions of the displacement field. Its main advantage over the elastic and fluid regularizers is that affine transformations are contained within its kernel (i.e., $\mathbf{u} = \mathbf{C}\mathbf{x} + \mathbf{b} \Rightarrow \Delta\mathbf{u} = 0 \Rightarrow \mathcal{R}_{curvature}(\mathbf{u}) = 0$). Modersitzki shows that this property makes the curvature regularizer more immune to initial affine misregistrations than the elastic, fluid or diffusion regularizers.

4.1.2 Extending the Quadratic Taxonomy

The quadratic taxonomy includes regularizers that have advantages over the basic diffusion regularizer: the elastic regularizer handles coupling of the spatial directions in the displacement field, and the curvature regularizer is immune to initial affine misregistrations. The question naturally arises, then: does there exist a regularizer that handles coupling *and* is immune to affine misregistrations?

Fortunately, the answer is yes: regularizers that incorporate second partial and mixed partial derivatives can have both advantages. One such regularizer is presented by Amodei *et al.* [2] in the context of vector splines, and later appears for use in image reconstruction problems [3, 4] and B-Spline based image registration [81]:

$$\mathcal{R}_{Amodei}(\mathbf{u}) = \int_{\Omega} \left(\alpha_1 \left\| \nabla \mathrm{div}\, \mathbf{u} \right\|^2 + \alpha_2 \left\| \nabla \mathrm{rot}\, \mathbf{u} \right\|^2 \right) d\mathbf{x}, \tag{4.6}$$

where $\mathrm{rot}\, \mathbf{u}$ refers to the rotational component of a vector field, which is defined by:

$$\mathrm{rot}\, \mathbf{u} = \begin{cases} -\partial_{x_2} u_1 + \partial_{x_1} u_2, & n = 2 \\ \mathrm{curl}\, \mathbf{u}, & n = 3 \end{cases}, \tag{4.7}$$

where n is the number of dimensions.

In order to extend the quadratic taxonomy to include a regularizer that shares these advantages, however, we present a slightly different second-order differential quadratic form, which we call the *second-order elastic* regularizer:

$$\mathcal{R}_{elastic\text{-}2}(\mathbf{u}) = \frac{1}{2} \int_{\Omega} \left(\frac{\mu}{3} \sum_{j,k,l=1}^{n} \left(\partial^2_{x_j x_k} u_l + \partial^2_{x_j x_l} u_k + \partial^2_{x_k x_l} u_j \right)^2 + \lambda \left\| \nabla \mathrm{div}\, \mathbf{u} \right\|^2 \right) d\mathbf{x}. \tag{4.8}$$

Even though (4.8) and (4.6) are different regularizers, they have the same Gâteaux derivative up to boundary conditions. This can be established by referring to the results of Cachier and Ayache [18] on second-order differential quadratic forms. The reason, therefore, that we choose (4.8) over (4.6) to extend the quadratic taxonomy is that (4.8) reveals a direct generalization of the elastic regularizer in terms of the Lamé parameters μ and λ.

For the remainder of this thesis, we will consider that the *extended* quadratic taxonomy refers to the diffusion, elastic, curvature, and second-order elastic regularizers, whereas the original quadratic taxonomy refers only to the diffusion, elastic, and curvature regularizers.

4.1.3 Families of Homogeneous Regularizers

The diffusion, elastic/fluid, curvature, and second-order elastic regularizers are not the only homogeneous regularizers that can be used for nonparametric registration. However, we can establish an interesting link between these regularizers and the space of all rotationally invariant regularizers of second order or less. Arigovindan [3, 4] developed a family of

regularizers given by the linear combination of four terms:

$$\mathcal{R}_{\gamma_1,\gamma_2,\gamma_3,\gamma_4}(\mathbf{u}) = \gamma_1 R_1(\mathbf{u}) + \gamma_2 R_2(\mathbf{u}) + \gamma_3 R_3(\mathbf{u}) + \gamma_4 R_4(\mathbf{u}) \,, \tag{4.9}$$

where

$$\mathcal{R}_1(\mathbf{u}) = \frac{1}{2}\int_\Omega (\operatorname{div}\mathbf{u})^2 \, d\mathbf{x}, \tag{4.10}$$

$$\mathcal{R}_2(\mathbf{u}) = \frac{1}{2}\int_\Omega \|\operatorname{rot}\mathbf{u}\|^2 \, d\mathbf{x}, \tag{4.11}$$

$$\mathcal{R}_3(\mathbf{u}) = \frac{1}{2}\int_\Omega \|\nabla \operatorname{div}\mathbf{u}\|^2 \, d\mathbf{x}, \quad \text{and} \tag{4.12}$$

$$\mathcal{R}_4(\mathbf{u}) = \frac{1}{2}\int_\Omega \|\nabla \operatorname{rot}\mathbf{u}\|_F^2 \, d\mathbf{x}. \tag{4.13}$$

Note that if $\gamma_1 = \gamma_2 = 0$, the Arigovindan regularizer reduces to the Amodei regularizer (4.6). As explained in [4], $\mathcal{R}_1$ quantifies the compression rate of the displacement field, $\mathcal{R}_2$ quantifies the squared angular velocity, $\mathcal{R}_3$ quantifies the spatial roughness of the compression rate, and $\mathcal{R}_4$ quantifies the variation in angular velocity. Arigovindan proved that $\mathcal{R}_{\gamma_1,\gamma_2,\gamma_3,\gamma_4}$ generates all possible rotationally invariant regularizers of second order or less.

By referring to the results of Cachier and Ayache [18] on first and second order differential quadratic forms, we can see that the space spanned by the Gâteaux derivatives of the diffusion, elastic, curvature, and second-order elastic regularizers is equivalent, up to boundary conditions, of the space spanned by the Gâteaux derivative of the Arigovindan regularizer. This link to the Arigovindan regularizer family suggests that if boundary conditions are ignored, the Gâteaux derivative of *any* rotationally invariant regularizer up to second order is in the span of the Gâteaux derivatives of the regularizers of the extended quadratic taxonomy.

4.2 Nonhomogeneous Regularizers in Optical Flow

The homogeneous regularizers of the extended quadratic taxonomy penalize displacement field variation in a manner that is independent of location. In medical imaging, however, there are situations where the image content dictates that different penalties should be applied in different places. For example, bones should deform approximately rigidly, whereas

organs or soft tissue are more flexible. This suggests that regularizers should be allowed to adjust the amount of penalty based on the underlying anatomical structure. In other situations, sliding may occur (between an organ and the chest wall, for example), which suggests that regularizers should be allowed to penalize the displacement field differently in transitional areas of the image than in homogeneous regions.

These concerns have been addressed in part in the computer vision community. In the area of optical flow, which deals with small deformations between images, there has been a large amount of research into nonhomogeneous regularizers that are generalizations of the diffusion regularizer. Weickert *et al.* [95, 96] provide a taxonomy of these isotropic and anisotropic regularizers. In this section, we present the nonhomogeneous regularizers of the Weickert taxonomy.

4.2.1 Image-driven Isotropic

The simplest way to allow variable regularization is to incorporate a scalar-valued weighting function into the diffusion regularizer. The weighting function, denoted below as $\beta(\mathbf{x})$, should be strictly positive and differentiable over Ω. Incorporating $\beta(\mathbf{x})$ yields the image-driven isotropic diffusion regularizer:

$$\mathcal{R}^{id\text{-}iso}_{diffusion}(\mathbf{u}) = \frac{1}{2}\int_{\Omega} \beta(\mathbf{x}) \left(\sum_{j=1}^{n} \|\nabla u_j\|^2 \right) d\mathbf{x}. \tag{4.14}$$

Particular choices of β can be tailored to the application at hand. Alvarez [1] proposes using a weighting function that is inversely proportional to the gradient magnitude of the underlying image. Charbonnier [26] and Bruhn [16] show that a function fitting this description is $\beta(\mathbf{x}) = \Psi_\epsilon\Big(\|\nabla F(\mathbf{x})\|^2\Big)$, where

$$\Psi_\epsilon\big(s^2\big) = \frac{\epsilon}{\sqrt{s^2+\epsilon^2}} \ . \tag{4.15}$$

In essence, this type of image-driven regularization allows the resulting deformation to be smooth in areas where the original floating (or reference) image is smooth and allows more variable deformations in areas where the image gradient is large.

One questionable aspect of $\mathcal{R}^{id\text{-}iso}_{diffusion}$ is that $\beta(\mathbf{x})$ does not depend on the displacement field u. If $\beta(\mathbf{x})$ is based on some property of the reference image, then there is no problem.

However, if $\beta(\mathbf{x})$ is based on some property of the floating image (as in the Alvarez regularizer), then it should deform according to the changing displacement field. This suggests that a more appropriate way to write the image-driven isotropic regularizer is by:

$$\mathcal{R}_{diffusion}^{id-iso-nonlin}(\mathbf{u}) = \frac{1}{2}\int_{\Omega} \beta^{\mathbf{u}}(\mathbf{x}) \left(\sum_{j=1}^{n} \|\nabla u_j\|^2 \right) d\mathbf{x}, \tag{4.16}$$

where $\beta^{\mathbf{u}}(\mathbf{x}) := \beta(\mathbf{x} - \mathbf{u}(\mathbf{x}))$. The "nonlin" superscript is added to indicate that the Gâteaux derivative of this regularizer is nonlinear with respect to $\mathbf{u}$.

4.2.2 Image-driven Anisotropic

Replacing the scalar-valued weighting function $\beta(\mathbf{x})$ in the linear image-driven isotropic regularizer with a matrix-valued weighting function $\mathbf{B}(\mathbf{x})$ yields the linear image-driven *anisotropic* regularizer:

$$\mathcal{R}_{diffusion}^{id\text{-}aniso}(\mathbf{u}) = \frac{1}{2}\int_{\Omega} \left(\sum_{j=1}^{n} (\nabla u_j)^{\mathrm{T}} \mathbf{B}(\mathbf{x}) \nabla u_j \right) d\mathbf{x}. \tag{4.17}$$

In general, any $\mathbf{B}(\mathbf{x})$ can be chosen whose range is the set of symmetric positive definite matrices. The purpose of such a weighting function is to locally affect the direction of smoothing. The amount of penalty applied in the direction of an eigenvector of $\mathbf{B}(\mathbf{x})$ is proportional to the corresponding eigenvalue.

Nagel [64, 65] was the first in the computer vision community to suggest this type of anisotropic regularization, which he referred to as *oriented smoothness*. He proposed the following weighting matrix, which inhibits smoothing in the direction of the image gradient:

$$\mathbf{B}(\mathbf{x}) = \frac{\left(\nabla^{\perp} F(\mathbf{x})\right)\left(\nabla^{\perp} F(\mathbf{x})\right)^{\mathrm{T}} + \gamma^2 \mathbf{I}}{\|\nabla F(\mathbf{x})\|^2 + 2\gamma^2}, \tag{4.18}$$

where $\nabla^{\perp}$ indicates the vector normal to $\nabla F(\mathbf{x})$ with the same magnitude as $\nabla F(\mathbf{x})$, and where γ is chosen small enough so that $\gamma^2 << \max_{\mathbf{x}\in\Omega} \|\nabla F(\mathbf{x})\|^2$.

Nagel's weighting matrix can only be used for registration problems with two spatial dimensions because $\nabla^{\perp} F$ is not uniquely defined for three or more dimensions. In order to generalize Nagel's idea of oriented smoothness to registration problems in n dimensions, we first recognize that in two dimensions, the following statement holds true:

$$\left(\nabla^{\perp} F(\mathbf{x})\right)\left(\nabla^{\perp} F(\mathbf{x})\right)^{\mathrm{T}} = \|\nabla F(\mathbf{x})\|^2 \mathbf{I} - \left(\nabla F(\mathbf{x})\right)\left(\nabla F(\mathbf{x})\right)^{\mathrm{T}}. \tag{4.19}$$

This allows us to rewrite (4.18) as:

$$\mathbf{B}(\mathbf{x}) = \frac{\left(\|\nabla F(\mathbf{x})\|^2 + \gamma^2\right)\mathbf{I} - (\nabla F(\mathbf{x}))\,(\nabla F(\mathbf{x}))^{\mathrm{T}}}{\|\nabla F(\mathbf{x})\|^2 + 2\gamma^2}. \tag{4.20}$$

This form of weighting matrix easily generalizes to n dimensions because $\nabla F(\mathbf{x})$ is always uniquely defined (assuming F is differentiable everywhere). The only adjustment to be made for $n > 2$ is in the denominator, which must be rescaled to take into account the trace of the numerator. This yields:

$$\mathbf{B}(\mathbf{x}) = \frac{\left(\|\nabla F(\mathbf{x})\|^2 + \gamma^2\right)\mathbf{I} - (\nabla F(\mathbf{x}))\,(\nabla F(\mathbf{x}))^{\mathrm{T}}}{\|\nabla F(\mathbf{x})\|^2 + n\gamma^2/(n-1)}, \tag{4.21}$$

which is just a scaled version of the matrix presented in [49] that can be considered the generalization of Nagel's oriented smoothness weighting matrix to n dimensions.

As argued in Section 4.2.1, if the weighting function is based on some property of the floating image, it should deform over the course of registration. This suggests that the image-driven anisotropic regularizer should be written as:

$$\mathcal{R}^{\textit{id-aniso-nonlin}}_{\textit{diffusion}}(\mathbf{u}) = \frac{1}{2}\int_{\Omega}\left(\sum_{j=1}^{n}(\nabla u_j)^{\mathrm{T}}\,\mathbf{B}^{\mathbf{u}}(\mathbf{x})\,\nabla u_j\right)d\mathbf{x}. \tag{4.22}$$

where $\mathbf{B}^{\mathbf{u}}(\mathbf{x}) := \mathbf{B}(\mathbf{x} - \mathbf{u}(\mathbf{x}))$.

The generalized oriented smoothness weighting matrix (4.21) can still be used, although it must be modified to vary according to the displacement field:

$$\mathbf{B}^{\mathbf{u}}(\mathbf{x}) = \frac{\left(\|\nabla F^{\mathbf{u}}(\mathbf{x})\|^2 + \gamma^2\right)\mathbf{I} - (\nabla F^{\mathbf{u}}(\mathbf{x}))\,(\nabla F^{\mathbf{u}}(\mathbf{x}))^{\mathrm{T}}}{\|\nabla F^{\mathbf{u}}(\mathbf{x})\|^2 + n\gamma^2/(n-1)}. \tag{4.23}$$

4.2.3 Flow-driven Isotropic

One potential problem with the image-driven isotropic and anisotropic regularizers is that there may be edges in the image data in which smoothness should be enforced (for example, highly textured regions). A solution to this problem is to design regularizers in which smoothness is penalized based on the locations of edges in the *flow* field. Schnörr and Weickert [80, 95, 96] show that this can be done by applying a local weighting function to

the diffusion regularizer in the following manner:

$$\mathcal{R}^{\textit{fd-iso}}_{\textit{diffusion}}(\mathbf{u}) = \frac{1}{2}\int_{\Omega} \beta\left(\sum_{j=1}^{n} \|\nabla u_j\|^2\right) d\mathbf{x}. \tag{4.24}$$

Ideally, the weighting function β should be convex in $\nabla\mathbf{u}$. Brox *et al.* [15] show that a suitable choice is $\beta\left(\sum_{j=1}^{n} \|\nabla u_j\|^2\right) := \Psi_{1,\epsilon}(\nabla\mathbf{u})$, where $\Psi_{1,\epsilon}$ is the modified L^1 norm (2.25). This choice yields a regularizer that is similar in form to the total variation (TV) regularizer [42].

4.2.4 Flow-driven Anisotropic

Weickert and Schnörr [96] show that regularizers can be designed in which penalties are not only based on the *locations* of edges in the flow field, but on the *directions* of those edges as well. This can be accomplished by incorporating a matrix-valued weighting function $\mathbf{B}(\mathbf{M})$ into the diffusion regularizer:

$$\mathcal{R}^{\textit{fd-aniso}}_{\textit{diffusion}}(\mathbf{u}) = \frac{1}{2}\int_{\Omega} \operatorname{tr}\left(\mathbf{B}\left(\nabla\mathbf{u}\nabla\mathbf{u}^{\mathrm{T}}\right)\right) d\mathbf{x}. \tag{4.25}$$

Note that the choice $\mathbf{B} = \mathbf{I}$ reduces (4.25) to the homogeneous diffusion regularizer (4.1).

The weighting function $\mathbf{B}(\mathbf{M})$ can be defined in terms of scalar weighting functions applied to the eigenvalues of $\mathbf{M}$. If we assume $\mathbf{M}$ is a symmetric matrix with eigenvector/eigenvalue pairs $(\mathbf{m}_j, \mu_j)$, then we have:

$$\mathbf{B}(\mathbf{M}) = \sum_{j=1}^{n} \beta(\mu_j)\, \mathbf{m}_j \mathbf{m}_j^{\mathrm{T}}, \tag{4.26}$$

and therefore

$$\operatorname{tr}\left(\mathbf{B}(\mathbf{M})\right) = \sum_{j=1}^{n} \beta(\mu_j). \tag{4.27}$$

As in the flow-driven isotropic case, a suitable choice for $\beta(\mu)$ is $\Psi_{1,\epsilon}(\mu)$.

4.3 Constructing Nonhomogeneous Regularizers for Image Registration

Both the extended quadratic taxonomy of regularizers for image registration and the Weickert taxonomy of regularizers for optical flow start with the diffusion regularizer and

generalize it in different ways. The quadratic taxonomy considers different homogeneous extensions that allow for coupling and/or greater insensitivity to initial affine misregistrations. Weickert's taxonomy considers various nonhomogeneous extensions that allow for varying the regularization penalty based on image or flow characteristics.

In this section, we show how some of the ideas behind both taxonomies can be combined. We construct both image-driven and flow-driven isotropic versions of the elastic, curvature, and second-order elastic regularizers. These generalizations build off of the work of Kabus [53, 54], who presented an image-driven isotropic version of the elastic regularizer. We then provide some notes on why anisotropic versions of the elastic, curvature, and second-order elastic regularizers cannot be directly deduced from the form of the anisotropic diffusion regularizers.

4.3.1 Image-driven Isotropic Regularizers

Image-based isotropic versions of the elastic, curvature, and second-order elastic regularizers can be constructed by incorporating weighting functions in a manner analogous to (4.14). For the elastic regularizer, we have:

$$\mathcal{R}^{\textit{id-iso}}_{\textit{elastic}}(\mathbf{u}) = \frac{1}{2}\int_{\Omega}\left(\frac{\beta_1(\mathbf{x})}{2}\sum_{j,k=1}^{n}\left(\partial_{x_j}u_k + \partial_{x_k}u_j\right)^2 + \beta_2(\mathbf{x})\left(\operatorname{div}\mathbf{u}\right)^2\right)d\mathbf{x}. \tag{4.28}$$

For the curvature regularizer, we have:

$$\mathcal{R}^{\textit{id-iso}}_{\textit{curvature}}(\mathbf{u}) = \frac{1}{2}\int_{\Omega}\beta(\mathbf{x})\left(\sum_{j=1}^{n}\left(\Delta u_j\right)^2\right)d\mathbf{x}. \tag{4.29}$$

Finally, for the second-order elastic regularizer, we have:

$$\begin{aligned}\mathcal{R}^{\textit{id-iso}}_{\textit{elastic-2}}(\mathbf{u}) = \frac{1}{2}\int_{\Omega}\frac{\beta_1(\mathbf{x})}{3}\left(\sum_{j,k,l=1}^{n}\left(\partial^2_{x_jx_k}u_l + \partial^2_{x_jx_l}u_k + \partial^2_{x_kx_l}u_j\right)^2\right)d\mathbf{x} \\ + \quad \frac{1}{2}\int_{\Omega}\beta_2(\mathbf{x})\left\|\nabla\operatorname{div}\mathbf{u}\right\|^2 d\mathbf{x}.\end{aligned} \tag{4.30}$$

A variety of choices can be made for the weighting functions. If $\beta_1(\mathbf{x}) = \beta_2(\mathbf{x}) = \beta(\mathbf{x})$, and $\beta(\mathbf{x})$ is chosen as in (4.15), then the elastic regularizer will be modified so that the resulting displacement field is smoother in areas where the original image is smooth.

For the elastic and second-order elastic regularizers, the weighting functions can be tailored to the Lamé parameters. Defining $\beta_1(\mathbf{x}) := \mu(\mathbf{x})$ and $\beta_1(\mathbf{x}) := \lambda(\mathbf{x})$ allows regularization based on spatially varying Lamé parameters, which may be desirable if the images have been segmented into different tissue types.

Since the floating image deforms as the displacement field changes, any weighting functions based on properties of the floating image should also deform. Analogous to the nonlinear image-driven isotropic diffusion regularizer (4.16), nonlinear versions of (4.28)–(4.30) can be defined by:

$$\mathcal{R}_{elastic}^{id\text{-}iso\text{-}nonlin}(\mathbf{u}) = \frac{1}{2}\int_\Omega \left(\frac{\beta_1^{\mathbf{u}}(\mathbf{x})}{2} \sum_{j,k=1}^{n} \left(\partial_{x_j} u_k + \partial_{x_k} u_j\right)^2 + \beta_2^{\mathbf{u}}(\mathbf{x}) \left(\operatorname{div} \mathbf{u}\right)^2 \right) d\mathbf{x}, \quad (4.31)$$

$$\mathcal{R}_{curvature}^{id\text{-}iso\text{-}nonlin}(\mathbf{u}) = \frac{1}{2}\int_\Omega \beta^{\mathbf{u}}(\mathbf{x}) \left(\sum_{j=1}^{n} \left(\Delta u_j\right)^2 \right) d\mathbf{x}, \quad (4.32)$$

$$\begin{aligned} \mathcal{R}_{elastic\text{-}2}^{id\text{-}iso\text{-}nonlin}(\mathbf{u}) &= \frac{1}{2}\int_\Omega \frac{\beta_1^{\mathbf{u}}(\mathbf{x})}{3} \left(\sum_{j,k,l=1}^{n} \left(\partial_{x_j x_k}^2 u_l + \partial_{x_j x_l}^2 u_k + \partial_{x_k x_l}^2 u_j\right)^2 \right) d\mathbf{x} \\ &+ \frac{1}{2}\int_\Omega \beta_2^{\mathbf{u}}(\mathbf{x}) \left\|\nabla \operatorname{div} \mathbf{u}\right\|^2 d\mathbf{x}. \end{aligned} \quad (4.33)$$

Kabus [53, 54] presents a version of the nonlinear image-driven isotropic elastic regularizer (4.31) in which he uses the weighting functions $\beta_1^{\mathbf{u}}(\mathbf{x}) := \alpha^{\mathbf{u}}(\mathbf{x})\,\mu^{\mathbf{u}}(\mathbf{x})$ and $\beta_2^{\mathbf{u}}(\mathbf{x}) := \alpha^{\mathbf{u}}(\mathbf{x})\,\lambda^{\mathbf{u}}(\mathbf{x})$. The Lamé parameters $\mu^{\mathbf{u}}(\mathbf{x})$ and $\lambda^{\mathbf{u}}(\mathbf{x})$ are considered to be spatially varying, and $\alpha^{\mathbf{u}}(\mathbf{x})$ is a weighting function based on a segmentation of the floating image into foreground/background regions.

4.3.2 Flow-driven Isotropic Regularizers

In order to modify the elastic regularizer based on the *flow* field, we can apply local weighting functions in the following manner:

$$\mathcal{R}_{elastic}^{fd\text{-}iso}(\mathbf{u}) = \frac{1}{2}\int_\Omega \left(\frac{1}{2}\,\beta_1 \left(\sum_{j,k=1}^{n} \left(\partial_{x_j} u_k + \partial_{x_k} u_j\right)^2 \right) + \beta_2\Big(\left(\operatorname{div} \mathbf{u}\right)^2 \Big) \right) d\mathbf{x}. \quad (4.34)$$

The choices $\beta_1\big(s^2\big) := \mu\Psi_{1,\epsilon}(s)$ and $\beta_2\big(s^2\big) := \lambda\Psi_{1,\epsilon}(s)$, where $\Psi_{1,\epsilon}(s)$ is the modified L^1 norm (2.24), allow for a direct generalization of the flow-driven isotropic diffusion regularizer

presented by Brox *et al.* [15]. Note that if the alternative choices $\beta_1\big(s^2\big) := \mu s^2$ and $\beta_2\big(s^2\big) := \lambda s^2$ are made, (4.34) reduces to the homgeneous elastic regularizer (4.2).

For a flow-driven isotropic version of the curvature regularizer, we have:

$$\mathcal{R}^{\textit{fd-iso}}_{\textit{curvature}}(\mathbf{u}) = \frac{1}{2}\int_\Omega \beta\left(\sum_{j=1}^{n}(\Delta u_j)^2\right) d\mathbf{x}, \tag{4.35}$$

with a suitable choice of local weighting function $\beta\Big(\|\Delta\mathbf{u}\|^2\Big) := \Psi_{1,\epsilon}(\Delta\mathbf{u})$.

Finally, for a flow-driven isotropic version of the second-order elastic regularizer, we have:

$$\begin{aligned}\mathcal{R}^{\textit{fd-iso}}_{\textit{elastic-2}}(\mathbf{u}) &= \frac{1}{2}\int_\Omega \frac{1}{3}\,\beta_1\left(\sum_{j,k,l=1}^{n}\left(\partial^2_{x_jx_k}u_l + \partial^2_{x_jx_l}u_k + \partial^2_{x_kx_l}u_j\right)^2\right) d\mathbf{x} \\ &+ \frac{1}{2}\int_\Omega \beta_2\Big(\|\nabla \mathrm{div}\,\mathbf{u}\|^2\Big)\, d\mathbf{x}.\end{aligned} \tag{4.36}$$

As in (4.34), suitable choices for the weighting functions are $\beta_1\big(s^2\big) := \mu\Psi_{1,\epsilon}(s)$ and $\beta_2\big(s^2\big) := \lambda\Psi_{1,\epsilon}(s)$.

4.3.3 Notes on Anisotropic Regularizers

There is a temptation to form image-driven anisotropic versions of the elastic, curvature, and second-order elastic regularizers by premultiplying any occurrence of $\nabla\mathbf{u}$ in the homogeneous regularizer by the generalized oriented smoothness matrices (4.21) or (4.23). After all, it is straightforward to express each of the homogeneous regularizers in terms of $\nabla\mathbf{u}$:

$$\begin{aligned}\mathcal{R}_{\textit{elastic}}(\mathbf{u}) &= \frac{1}{2}\int_\Omega \mu\left(\mathrm{tr}\,\langle\nabla\mathbf{u},\nabla\mathbf{u}\rangle + \mathrm{tr}\left\langle(\nabla\mathbf{u})^{\mathrm{T}},\nabla\mathbf{u}\right\rangle\right) d\mathbf{x} \\ &+ \frac{1}{2}\int_\Omega \lambda \mathrm{tr}^2\left(\nabla\mathbf{u}\right) d\mathbf{x},\end{aligned} \tag{4.37}$$

$$\mathcal{R}_{\textit{curvature}}(\mathbf{u}) = \frac{1}{2}\int_\Omega \|\mathrm{div}\,\nabla\mathbf{u}\|^2\, d\mathbf{x}, \tag{4.38}$$

$$\begin{aligned}\mathcal{R}_{\textit{elastic-2}}(\mathbf{u}) &= \frac{1}{2}\int_\Omega \mu\left(\sum_{j=1}^{n}\mathrm{tr}\left\langle\partial_{x_j}\nabla\mathbf{u},\partial_{x_j}\nabla\mathbf{u}\right\rangle\right) d\mathbf{x} \\ &+ \frac{1}{2}\int_\Omega 2\mu\left(\sum_{j=1}^{n}\mathrm{tr}\left\langle\left(\partial_{x_j}\nabla\mathbf{u}\right)^{\mathrm{T}},\partial_{x_j}\nabla\mathbf{u}\right\rangle\right) d\mathbf{x} \\ &+ \frac{1}{2}\int_\Omega \lambda\left\|\nabla\mathrm{tr}\left(\nabla\mathbf{u}\right)\right\|^2 d\mathbf{x},\end{aligned} \tag{4.39}$$

where $\|\bullet\|_F$ indicates the Frobenius norm.

In fact, we initially followed this path as a possibility for constructing anisotropic generalizations of elastic, curvature, and second-order elastic regularizers. However, we ultimately discovered that this idea is misguided, and we were unable to determine any anisotropic generalizations. The main reason for this is that the substitution $\mathbf{B}(\mathbf{x})\,\nabla\mathbf{u}(\mathbf{x}) \rightarrow \nabla\mathbf{u}(\mathbf{x})$ into (4.37)–(4.39) does not actually guarantee the desired reduction in penalty when the variations in the flow field are predominantly in the direction of the image gradient. This is in opposition to the diffusion regularizer, in which the insertion of the oriented smoothness matrix does guarantee the desired reduction in penalty.

4.4 Gâteaux Derivatives

To construct the system of partial differential equations whose solution yields the optimal deformation field relating a pair of images, it is necessary to determine the Gâteaux derivative of the regularizer with respect to variations in the displacement field. For a general regularizer of the form $\mathcal{R}(\mathbf{u})$, the Gâteaux derivative is defined in the same manner as in (2.114); namely:

$$d\mathcal{R}(\mathbf{u};\mathbf{w}) = \lim_{h\rightarrow 0}\frac{1}{h}[\mathcal{R}(\mathbf{u}+h\mathbf{w}) - \mathcal{R}(\mathbf{u})]\,. \tag{4.40}$$

All of the regularizers presented in this chapter have Gâteaux derivative that can be expressed in terms of a differential operator $\mathcal{A}$ and boundary conditions $\mathcal{B}$:

$$d\mathcal{R}(\mathbf{u};\mathbf{w}) = \int_\Omega \langle \mathcal{A}(\mathbf{u})\,,\mathbf{w}\rangle\, d\mathbf{x} \tag{4.41}$$

with

$$\mathcal{B}[\mathbf{u}] = 0. \tag{4.42}$$

The second-order regularizers (all versions of the curvature and second-order elastic regularizers) have two sets of boundary conditions.

In this section, we will present the Gâteaux derivatives of all of the regularizers, written in terms of their partial differential operators and natural boundary conditions. Derivations of some of these terms will be presented in Appendix D.

4.4.1 Homogeneous Quadratic Regularizers

The partial differential operators and boundary conditions corresponding to the Gâteaux derivatives of the homogeneous versions of the diffusion, elastic, curvature, and second order regularizers are given in Table 4.1. The partial differential operator and boundary conditions for the fluid regularizer are the same as for the elastic regularizer, after replacing the displacement field $\mathbf{u}$ with the velocity field $\mathbf{v}$.

Derivations of these terms for the diffusion, elastic, and curvature regularizers are provided by Modersitzki [63]. For the second-order elastic regularizer, we provide a derivation in Appendix D.

Regularizer	**Differential Operator** $\mathcal{A}(\mathbf{u})$	**Boundary Conditions** $\mathcal{B}[\mathbf{u}]$
Diffusion	$-\Delta\mathbf{u}$	$\langle\nabla\mathbf{u}, \vec{\mathbf{n}}\rangle$
Elastic	$-\mu\Delta\mathbf{u} - (\lambda+\mu)\nabla\mathrm{div}\,\mathbf{u}$	$\lambda(\mathrm{div}\,\mathbf{u})\vec{\mathbf{n}} + \mu\langle\nabla\mathbf{u} + (\nabla\mathbf{u})^{\mathrm{T}}, \vec{\mathbf{n}}\rangle$
Curvature	$\Delta^2\mathbf{u}$	$\langle\nabla\mathbf{u}, \vec{\mathbf{n}}\rangle$; $\langle\Delta\nabla\mathbf{u}, \vec{\mathbf{n}}\rangle$
2^{nd}-order Elastic	$\mu\Delta^2\mathbf{u} + (\lambda+2\mu)\Delta\nabla\mathrm{div}\,\mathbf{u}$	$\lambda(\mathrm{div}\,\mathbf{u})\vec{\mathbf{n}} + \mu\langle\nabla\mathbf{u}, \vec{\mathbf{n}}\rangle$ $+2\mu\langle(\nabla\mathbf{u})^{\mathrm{T}}, \vec{\mathbf{n}}\rangle$; $\lambda(\Delta\mathrm{div}\,\mathbf{u})\vec{\mathbf{n}} + \mu\langle\Delta\nabla\mathbf{u}, \vec{\mathbf{n}}\rangle$ $+2\mu\langle\Delta\,(\nabla\mathbf{u})^{\mathrm{T}}, \vec{\mathbf{n}}\rangle$

Table 4.1: Gâteaux derivative components for homogeneous regularizers.

Note that the partial differential operators that appear in Table 4.1 have also emerged in a different paradigm for registration. Early work in computational anatomy by Grenander, Miller, and Christensen [28, 46] tackled the large deformation registration problem in a Bayesian estimation framework by inducing priors via partial differential equations. The operators in their partial differential equations are identical to the differential operators corresponding to the Gâteaux derivatives of the diffusion, elastic, and curvature regularizers.

4.4.2 Image-driven Isotropic Regularizers

The partial differential operators and boundary conditions corresponding to the Gâteaux derivatives of the image-driven isotropic regularizers are given in Tables 4.2 and 4.3. Table 4.2 corresponds to the linear versions of these regularizers (i.e. (4.14) and (4.28)–(4.30)), and Table 4.3 corresponds to the nonlinear versions (i.e. (4.16) and (4.31)–(4.33)). The boundary conditions are omitted from Table 4.3, as they are the same the boundary conditions listed in Table 4.2, after the substitution of $\beta^{\mathbf{u}}$, $\beta_1^{\mathbf{u}}$ and $\beta_2^{\mathbf{u}}$ for β, β_1 and β_2.

A derivation of these terms for the linear isotropic version of the diffusion regularizer is provided in Appendix D. Terms corresponding to the Gâteaux derivatives of the other image-driven isotropic regularizers can be derived in a similar manner; their derivations are not included in the Appendix due to space considerations.

Regularizer	**Differential Operator** $\mathcal{A}(\mathbf{u})$	**Boundary Conditions** $\mathcal{B}[\mathbf{u}]$
Diffusion	$-\operatorname{div}\left(\beta(\mathbf{x})\,\nabla\mathbf{u}\right)$	$\langle\beta(\mathbf{x})\,\nabla\mathbf{u},\vec{\mathbf{n}}\rangle$
Elastic	$-\operatorname{div}\left(\beta_1(\mathbf{x})\left(\nabla\mathbf{u}+(\nabla\mathbf{u})^{\mathrm{T}}\right)\right)$ $-\nabla\left(\beta_2(\mathbf{x})\operatorname{div}\mathbf{u}\right)$	$\beta_1(\mathbf{x})\,\langle\nabla\mathbf{u}+(\nabla\mathbf{u})^{\mathrm{T}},\vec{\mathbf{n}}\rangle$ $+\beta_2(\mathbf{x})\,(\operatorname{div}\mathbf{u})\,\vec{\mathbf{n}}$
Curvature	$\Delta\left(\beta\left(\mathbf{x}\right)\Delta\mathbf{u}\right)$	$\langle\beta(\mathbf{x})\,\nabla\mathbf{u},\vec{\mathbf{n}}\rangle$; $\langle\nabla(\beta(\mathbf{x})\,\Delta\mathbf{u}),\vec{\mathbf{n}}\rangle$
2^{nd}-order Elastic	$2\sum_{j=1}^{n}\partial_{x_j}\operatorname{div}\left(\beta_1(\mathbf{x})\,\partial_{x_j}\left(\nabla\mathbf{u}\right)^{\mathrm{T}}\right)$ $+\sum_{j=1}^{n}\partial_{x_j}\operatorname{div}\left(\beta_1(\mathbf{x})\,\partial_{x_j}\nabla\mathbf{u}\right)$ $+\nabla\operatorname{div}\left(\beta_2(\mathbf{x})\,\nabla\operatorname{div}\mathbf{u}\right)$	$\beta_2(\mathbf{x})(\operatorname{div}\mathbf{u})\vec{\mathbf{n}}$ $+\beta_1(\mathbf{x})\,\langle\nabla\mathbf{u},\vec{\mathbf{n}}\rangle$ $+2\beta_1(\mathbf{x})\,\langle(\nabla\mathbf{u})^{\mathbf{T}},\vec{\mathbf{n}}\rangle$; $\beta_2(\mathbf{x})(\Delta\operatorname{div}\mathbf{u})\vec{\mathbf{n}}$ $+\beta_1(\mathbf{x})\,\langle\Delta\nabla\mathbf{u},\vec{\mathbf{n}}\rangle$ $+2\beta_1(\mathbf{x})\,\langle\Delta\left(\nabla\mathbf{u}\right)^{\mathbf{T}},\vec{\mathbf{n}}\rangle$

Table 4.2: Gâteaux derivative components for linear image-driven isotropic regularizers.

Regularizer	**Differential Operator** $\mathcal{A}(\mathbf{u})$
Diffusion	$-\operatorname{div}\left(\beta^{\mathbf{u}}(\mathbf{x})\,\nabla\mathbf{u}\right) - \frac{1}{2}\left(\sum_{j=1}^{n}\|\nabla u_j\|^2\right)\nabla\beta^{\mathbf{u}}(\mathbf{x})$
Elastic	$-\operatorname{div}\left(\beta_1^{\mathbf{u}}(\mathbf{x})\left(\nabla\mathbf{u}+(\nabla\mathbf{u})^{\mathrm{T}}\right)\right) - \nabla\left(\beta_2^{\mathbf{u}}(\mathbf{x})\operatorname{div}\mathbf{u}\right)$ $-\frac{1}{4}\left(\sum_{j,k=1}^{n}\left(\partial_{x_j}u_k+\partial_{x_k}u_j\right)^2\right)\nabla\beta_1^{\mathbf{u}}(\mathbf{x}) - \frac{1}{2}(\operatorname{div}\mathbf{u})^2\,\nabla\beta_2^{\mathbf{u}}(\mathbf{x})$
Curvature	$\Delta\left(\beta^{\mathbf{u}}(\mathbf{x})\,\Delta\mathbf{u}\right) - \frac{1}{2}\left(\sum_{j=1}^{n}\left(\Delta u_j\right)^2\right)\nabla\beta^{\mathbf{u}}(\mathbf{x})$
2^{nd}-order Elastic	$\sum_{j=1}^{n}\partial_{x_j}\operatorname{div}\left(\beta_1^{\mathbf{u}}(\mathbf{x})\,\partial_{x_j}\left[\nabla\mathbf{u}+2\,(\nabla\mathbf{u})^{\mathrm{T}}\right]\right)$ $+\nabla\operatorname{div}\left(\beta_2^{\mathbf{u}}(\mathbf{x})\,\nabla\operatorname{div}\mathbf{u}\right)$ $-\frac{1}{6}\left(\sum_{j,k,l=1}^{n}\left(\partial^2_{x_j,x_k}u_l+\partial^2_{x_j,x_l}u_k+\partial^2_{x_k,x_l}u_j\right)^2\right)\nabla\beta_1^{\mathbf{u}}(\mathbf{x})$ $-\frac{1}{2}\|\nabla\operatorname{div}\mathbf{u}\|^2\,\nabla\beta_2^{\mathbf{u}}(\mathbf{x})$

Table 4.3: Partial derivative operators from Gâteaux derivatives of nonlinear image-driven isotropic regularizers. Boundary conditions are the same as those in Table 4.2, after substituting $\beta^{\mathbf{u}}$, $\beta_1^{\mathbf{u}}$ and $\beta_2^{\mathbf{u}}$ for β, β_1 and β_2.

4.4.3 Image-driven Anisotropic Regularizers

The image-driven anisotropic regularizer (4.17) has Gâteaux derivative with partial differential operator given by:

$$\mathcal{A}^{\textit{id-aniso}}_{\textit{diffusion}}(\mathbf{u}) = -\operatorname{div}\left(\mathbf{B}(\mathbf{x})\,\nabla\mathbf{u}\right), \tag{4.43}$$

and boundary conditions given by:

$$\mathcal{B}^{\textit{id-aniso}}_{\textit{diffusion}}[\mathbf{u}] = \langle\mathbf{B}(\mathbf{x})\nabla\mathbf{u},\vec{\mathbf{n}}\rangle. \tag{4.44}$$

The *nonlinear* image-driven anisotropic regularizer (4.22) has Gâteaux derivative with partial differential operator given by:

$$\mathcal{A}^{\textit{id-aniso-nonlin}}_{\textit{diffusion}}(\mathbf{u}) = -\operatorname{div}\left(\mathbf{B}^{\mathbf{u}}(\mathbf{x})\,\nabla\mathbf{u}\right) - \mathbf{c}[\mathbf{u}], \tag{4.45}$$

where $\mathbf{c}[\mathbf{u}]$ is a nonlinear vector in $\mathbf{u}$, with k^{th} element given by:

$$\mathbf{c}_k[\mathbf{u}] = \frac{1}{2}\sum_{j=1}^{n}\left(\nabla u_j\right)^{\mathrm{T}}\left[\frac{\partial}{\partial x_k}\mathbf{B}^{\mathbf{u}}(\mathbf{x})\right]\nabla u_j. \tag{4.46}$$

The corresponding boundary conditions are given by:

$$\mathcal{B}^{\textit{id-aniso-nonlin}}_{\textit{diffusion}}[\mathbf{u}] = \langle\mathbf{B}^{\mathbf{u}}(\mathbf{x})\nabla\mathbf{u},\vec{\mathbf{n}}\rangle. \tag{4.47}$$

The partial derivative of $\mathbf{B}^{\mathbf{u}}(\mathbf{x})$ in (4.46) is given by:

$$\frac{\partial}{\partial x_k}\mathbf{B}^{\mathbf{u}}(\mathbf{x}) = \frac{2\left(\mathbf{I}-\mathbf{B}^{\mathbf{u}}(\mathbf{x})\right)\left\langle \nabla F^{\mathbf{u}}(\mathbf{x}), \frac{\partial}{\partial x_k}\nabla F^{\mathbf{u}}(\mathbf{x})\right\rangle}{\left\|\nabla F^{\mathbf{u}}(\mathbf{x})\right\|^2 + n\gamma^2/(n-1)} - \frac{\left(\nabla F^{\mathbf{u}}(\mathbf{x})\right)\left(\frac{\partial}{\partial x_k}\nabla F^{\mathbf{u}}(\mathbf{x})\right)^{\mathrm{T}} + \left(\frac{\partial}{\partial x_k}\nabla F^{\mathbf{u}}(\mathbf{x})\right)\left(\nabla F^{\mathbf{u}}(\mathbf{x})\right)^{\mathrm{T}}}{\left\|\nabla F^{\mathbf{u}}(\mathbf{x})\right\|^2 + n\gamma^2/(n-1)}. \quad (4.48)$$

4.4.4 Flow-driven Isotropic Regularizers

The partial differential operators and boundary conditions corresponding to the Gâteaux derivatives of the flow-driven isotropic regularizers are given in Tables 4.4 and 4.5.

4.4.5 Flow-driven Anisotropic Regularizers

The Gâteaux derivative of the flow-driven anisotropic regularizer (4.25) has partial differential operator given by:

$$\mathcal{A}^{\textit{fd-iso}}_{\textit{diffusion}}(\mathbf{u}) = -\operatorname{div}\left(\mathbf{B}'\left(\nabla\mathbf{u}\nabla\mathbf{u}^{\mathrm{T}}\right)\nabla\mathbf{u}\right), \quad (4.49)$$

and boundary conditions given by:

$$\mathcal{B}^{\textit{fd-iso}}_{\textit{diffusion}}[\mathbf{u}] = \langle \mathbf{B}'\left(\nabla\mathbf{u}\nabla\mathbf{u}^{\mathrm{T}}\right)\nabla\mathbf{u}, \vec{\mathbf{n}}\rangle. \quad (4.50)$$

4.5 Visual Comparisons

In order to understand how the various regularizers differ, it is useful to visualize their behavior on some simple examples. As described in Chapter 5, the solution to the variational registration problem with Dirichlet boundary conditions can be found by performing a fixed point iteration and successively solving the PDE system:

$$\mathcal{A}\left(\mathbf{u}^{(k+1)}\right) = -\alpha\mathcal{P}\left(\mathbf{x}; R, F^{\mathbf{u}^{(k)}}\right), \quad (4.51)$$

where $\mathcal{A}$ is the partial differential operator corresponding to the regularizer, and $\mathcal{P}$ is the force vector arising from the Gâteaux derivative of the dissimilarity measure.

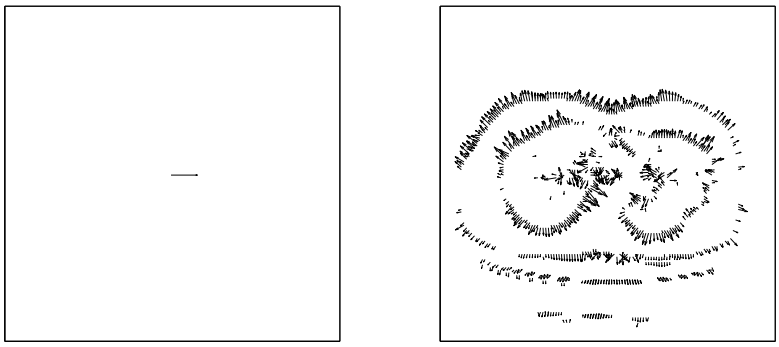
Here, we visualize the solution of a single iteration of (4.51) for two separate examples of $\mathcal{P}$. In the first example, shown in Fig. 4.1(a), $\mathcal{P}$ is an impulse function in the x direction

Regularizer	**Differential Operator** $\mathcal{A}(\mathbf{u})$
Diffusion	$-\operatorname{div}\left(\beta'\Big(\sum_{j=1}^{n}\|\nabla u_j\|^2\Big)\,\nabla\mathbf{u}\right)$
Elastic	$-\operatorname{div}\left(\beta_1'\Big(\sum_{j,k=1}^{n}\left(\partial_{x_j}u_k+\partial_{x_k}u_j\right)^2\Big)\left(\nabla\mathbf{u}+(\nabla\mathbf{u})^{\mathrm{T}}\right)\right)-\nabla\left(\beta_2'\Big((\operatorname{div}\mathbf{u})^2\Big)\operatorname{div}\mathbf{u}\right)$
Curvature	$\Delta\left(\beta'\Big(\sum_{j=1}^{n}(\Delta u_j)^2\Big)\,\Delta\mathbf{u}\right)$
2^{nd}-order Elastic	$\sum_{m=1}^{n}\partial_{x_m}\operatorname{div}\left(\beta_1'\Big(\sum_{j,k,l=1}^{n}\left(\partial^2_{x_j,x_k}u_l+\partial^2_{x_j,x_l}u_k+\partial^2_{x_k,x_l}u_j\right)^2\Big)\,\partial_{x_m}\left[\nabla\mathbf{u}+2\,(\nabla\mathbf{u})^{\mathrm{T}}\right]\right)$ $+\nabla\operatorname{div}\left(\beta_2'\Big(\|\nabla\operatorname{div}\mathbf{u}\|^2\Big)\,\nabla\operatorname{div}\mathbf{u}\right)$

Table 4.4: Partial derivative operators from Gâteaux derivatives of flow-driven isotropic regularizers.

Regularizer	**Boundary Conditions** $\mathcal{B}[\mathbf{u}]$
Diffusion	$\langle \beta'\Big(\sum_{j=1}^{n} \|\nabla u_j\|^2\Big)\, \nabla \mathbf{u}, \vec{\mathbf{n}} \rangle$
Elastic	$\beta_1'\Big(\sum_{j,k=1}^{n} \left(\partial_{x_j} u_k + \partial_{x_k} u_j\right)^2\Big) \langle \nabla \mathbf{u} + (\nabla \mathbf{u})^{\mathrm{T}}, \vec{\mathbf{n}} \rangle + \beta_2'\Big((\operatorname{div} \mathbf{u})^2\Big) (\operatorname{div} \mathbf{u})\, \vec{\mathbf{n}}$
Curvature	$\langle \beta'\Big(\sum_{j=1}^{n} (\Delta u_j)^2\Big)\, \nabla \mathbf{u}, \vec{\mathbf{n}} \rangle;$ $\langle \nabla\Big(\beta'\Big(\sum_{j=1}^{n} (\Delta u_j)^2\Big)\, \Delta \mathbf{u}\Big), \vec{\mathbf{n}} \rangle$
2^{nd}-order Elastic	$\beta_1'\Big(\sum_{j,k,l=1}^{n} \left(\partial^2_{x_j,x_k} u_l + \partial^2_{x_j,x_l} u_k + \partial^2_{x_k,x_l} u_j\right)^2\Big) \langle \nabla \mathbf{u} + 2\,(\nabla \mathbf{u})^{\mathbf{T}}, \vec{\mathbf{n}} \rangle + \beta_2'\Big(\|\nabla \operatorname{div} \mathbf{u}\|^2\Big) (\operatorname{div} \mathbf{u}) \vec{\mathbf{n}}\,;$ $\beta_1'\Big(\sum_{j,k,l=1}^{n} \left(\partial^2_{x_j,x_k} u_l + \partial^2_{x_j,x_l} u_k + \partial^2_{x_k,x_l} u_j\right)^2\Big) \langle \nabla \Delta \mathbf{u} + 2\Delta\,(\nabla \mathbf{u})^{\mathbf{T}}, \vec{\mathbf{n}} \rangle + \beta_2'\Big(\|\nabla \operatorname{div} \mathbf{u}\|^2\Big) (\Delta \operatorname{div} \mathbf{u}) \vec{\mathbf{n}}$

Table 4.5: Boundary conditions from Gâteaux derivatives of flow-driven isotropic regularizers.



(a) Impulse Example (b) CT Slice Example

Figure 4.1: Examples of force vector fields.

and zero in the y direction. In the second example, shown in Fig. 4.1(b), $\mathcal{P}$ is the force field arising from the MSD similarity measure applied to the CT slice images in Fig. 3.1.

In Fig.'s 4.2 – 4.15, we see the magnitudes and angles of the solutions to (4.51) with various regularizers. For every case, α has been chosen so that the maximum magnitude of the solution is scaled to be one.

The top row of Fig. 4.2 shows the magnitudes of the solutions of (4.51) with each homogeneous quadratic regularizer for the impulse example. It is clear that the diffusion and elastic regularizers propagate the information from the impulse to a much lesser extent than the curvature and second-order elastic regularizers. The middle row of Fig. 4.2 shows the solution magnitude with image-driven versions of the quadratic regularizers, and the bottom row shows the solution magnitude with flow-driven versions of the quadratic regularizers. For all of the image-driven and flow-driven regularizers, the function $\beta(\mathbf{x})$ was chosen as in (4.15). For the elastic and second-order elastic regularizers, the choice of Lamé parameters is $\mu = 1$, $\lambda = 0.1$. It is clear from this figure that that the image-driven and flow-driven regularizers propagate the information from the impulse to a lesser extent than the homogeneous regularizers.

Figure 4.3 shows the angles of the solutions of (4.51) for the impulse example. Since the impulse vector points only in the x direction, the separable nature of the diffusion and curvature regularizers yields solutions that have trivial y components. The elastic and second-order elastic regularizers, on the other hand, couple the x and y components, and

therefore, yield solutions with nontrivial y components.

Figures 4.4–4.13 illustrate the magnitudes and angles of the solutions of (4.51) for the image-driven and flow-driven regularizers when various values of the parameters ε and λ are chosen. In Fig.'s 4.4–4.5, we only observe the solution magnitudes for the diffusion and curvature regularizers, as the angles are all identically 0. It is clear from these figures that the solutions found using the image-driven and flow-driven diffusion regularizers are fairly insensitive to the parameter ε used in the weighting function $\beta(\mathbf{x})$; whereas, when the image-driven and flow-driven curvature regularizers are used, there is a significant narrowing of the extent of the solution magnitude as ε decreases.

Figures 4.6–4.13 show the solution magnitudes and angles for the elastic and second-order elastic versions of the image-driven and flow-driven regularizers for a range of values of ε and λ. (Note that μ was held constant at 1; it is really the ratio λ/μ that matters, as any scale factor common to both μ and λ can be factored out and accomodated by α). It is clear that the behavior of the image-driven elastic regularizer seems independent of λ and μ, and that the behavior of the flow-based elastic regularizer seems nearly independent of λ and varies only as ε approaches 1.

Like the with the curvature regularizer, the solution achieved with the second-order regularizer appears to vary much more with ε than its first-order cousin, the elastic regularizer. In Fig.'s 4.10–4.13, we can see that the solution magnitudes narrow significantly as ε decreases, but exhibit little variability with respect to λ.

Figures 4.14 and 4.15 illustrate the magnitudes and angles of the solutions of (4.51) for the CT slice example. For this example, the vector field was padded on all sides with zeros; Dirichlet boundary conditions were imposed on the padded vector field, and the solutions were cropped back to the original (unpadded) boundaries. As can be seen in Fig. 4.14, the regularizers propagate information from the force vector field in different ways. The diffusion and elastic regularizers propagate information to a lesser extent than the curvature and second-order elastic regularizers, and the image-driven and flow-driven regularizers propagate information to a lesser extent than the homogeneous regularizers.

Although we do not report resulting solutions for the CT slice example over a range

of values for ε and λ, we would anticipate a similar type of behavior as the solutions to the impulse example. The solutions when image-driven and flow-driven versions of the first order regularizers (diffusion and elastic) are used should be relatively invariant to particular values of ε (and λ), whereas solutions when the versions of the second-order regularizers (curvature and second-order elastic) are used should have magnitudes that narrow as ε increases and be relatively invariant to λ.

4.6 Summary

This chapter investigated regularizers, which are a necessary component that guarantees a unique solution to the variational registration problem. It first presented the quadratic taxonomy of homogeneous regularizers, and then added the second-order elastic regularizer in order to illustrate the connection between the regularizers of the quadratic taxonomy and the second-order families of regularizers described by Arigovindan [3] and Cachier and Ayache [18]. Next, it presented the Weickert taxonomy [95] of nonhomogeneous isotropic and anisotropic regularizers that have been used in optical flow, and it generalized these regularizers where necessary to allow nonhomogeneous weighting functions to deform over the course of registration. Then, the chapter illustrated how to combine the ideas of the quadratic and Weickert taxonomies to construct nonhomogeneous versions for all of the homogeneous regularizers in the quadratic taxonomy. Finally, it presented the Gâteaux derivatives of all regularizers, which are necessary for variational registration.

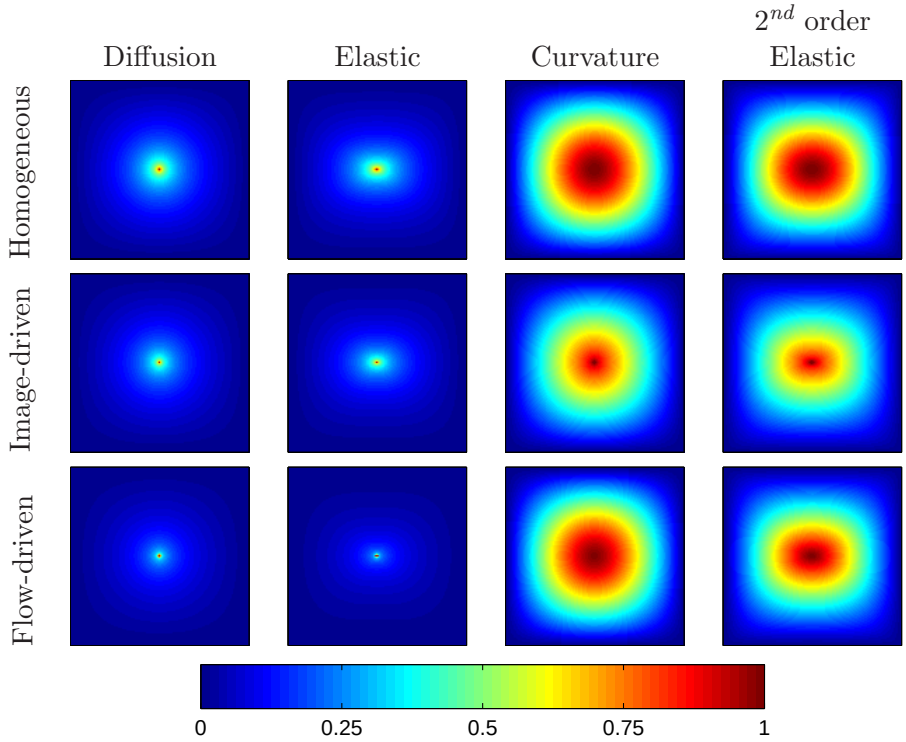


Figure 4.2: Magnitudes of solution fields for impulse example with various regularizers.

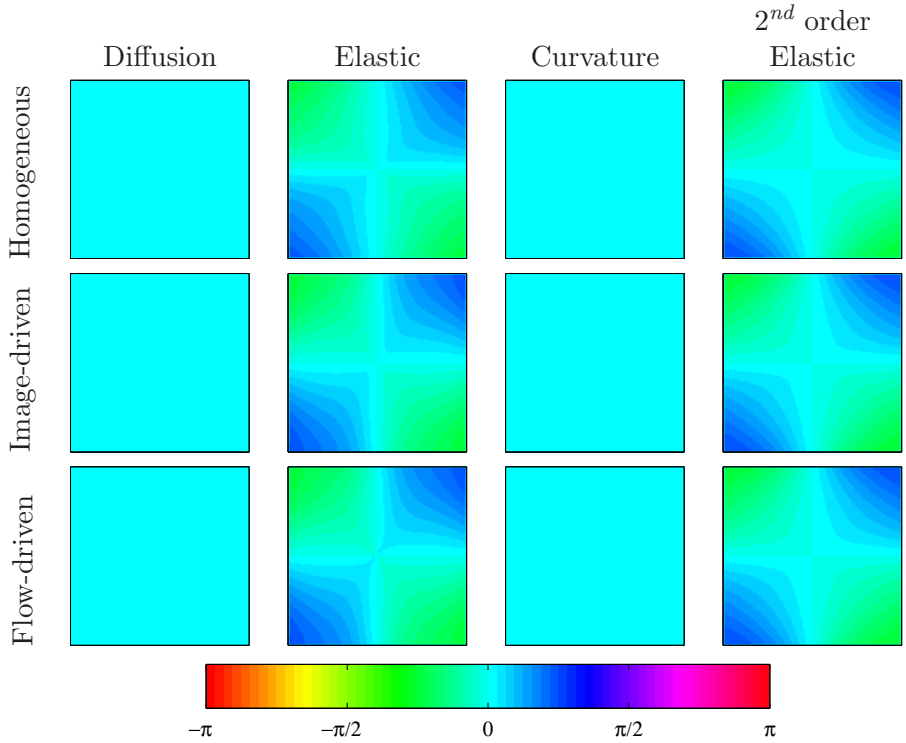


Figure 4.3: Angles of solution fields for impulse example with various regularizers.

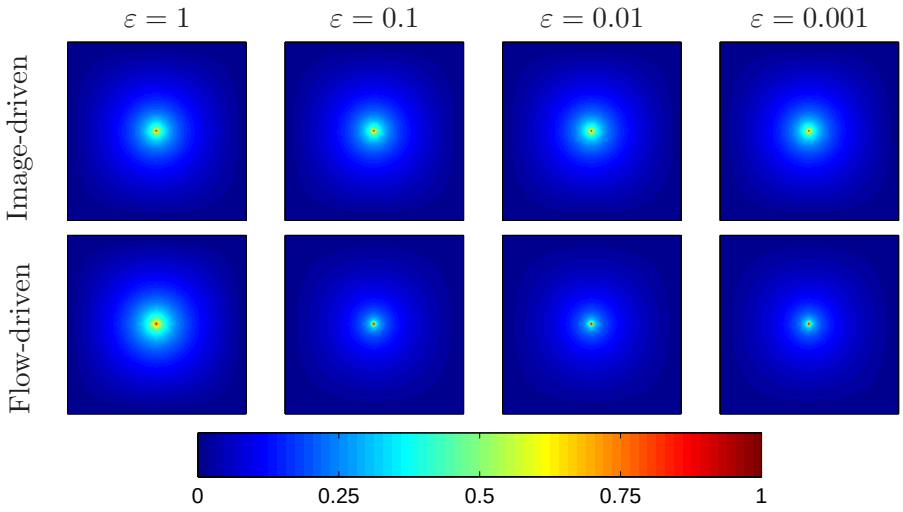


Figure 4.4: Magnitudes of solution fields for impulse example with image-driven and flow-driven diffusion regularizers for various values of ε.

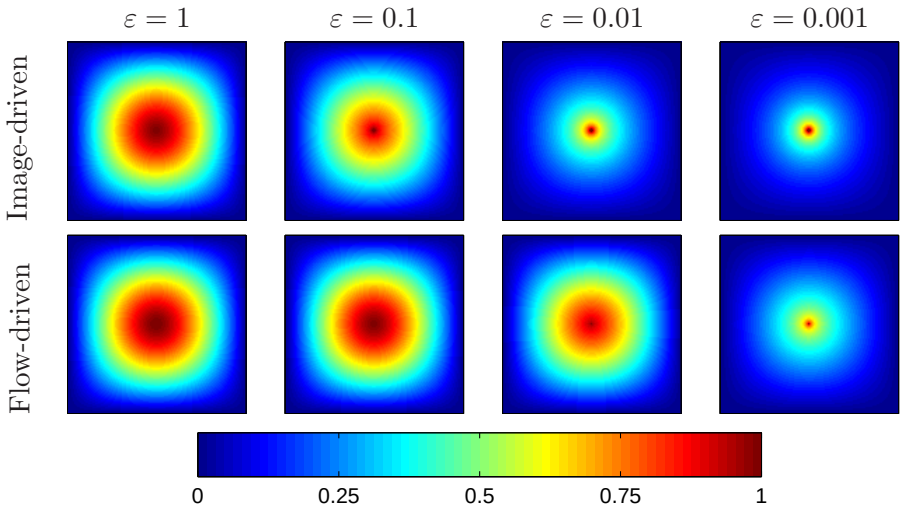


Figure 4.5: Magnitudes of solution fields for impulse example with image-driven and flow-driven curvature regularizers for various values of ε.

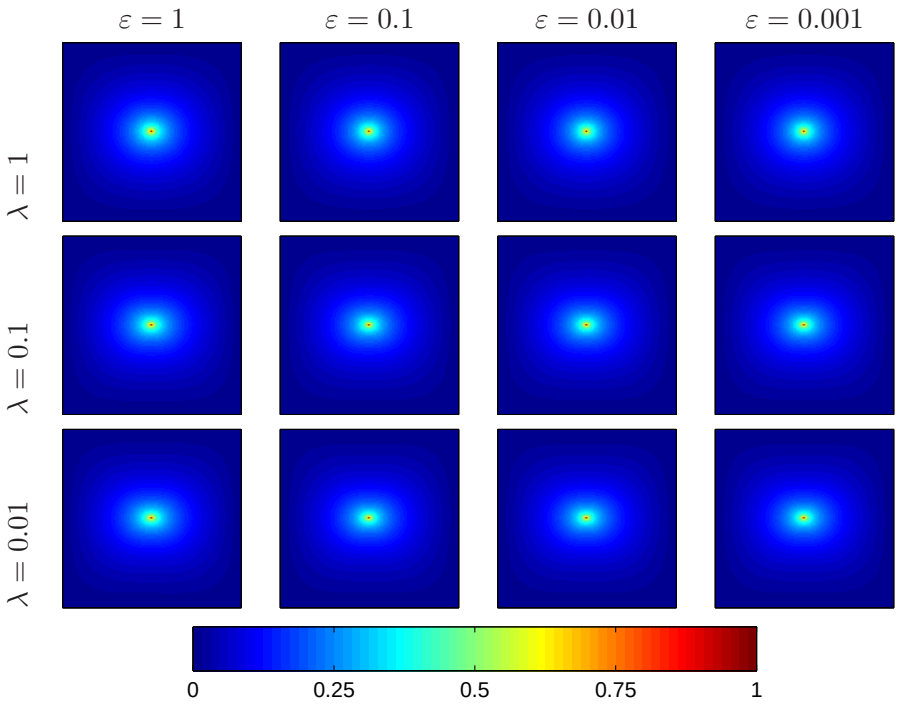


Figure 4.6: Magnitudes of solution fields for impulse example with image-driven elastic regularizers for various values of ε and λ.

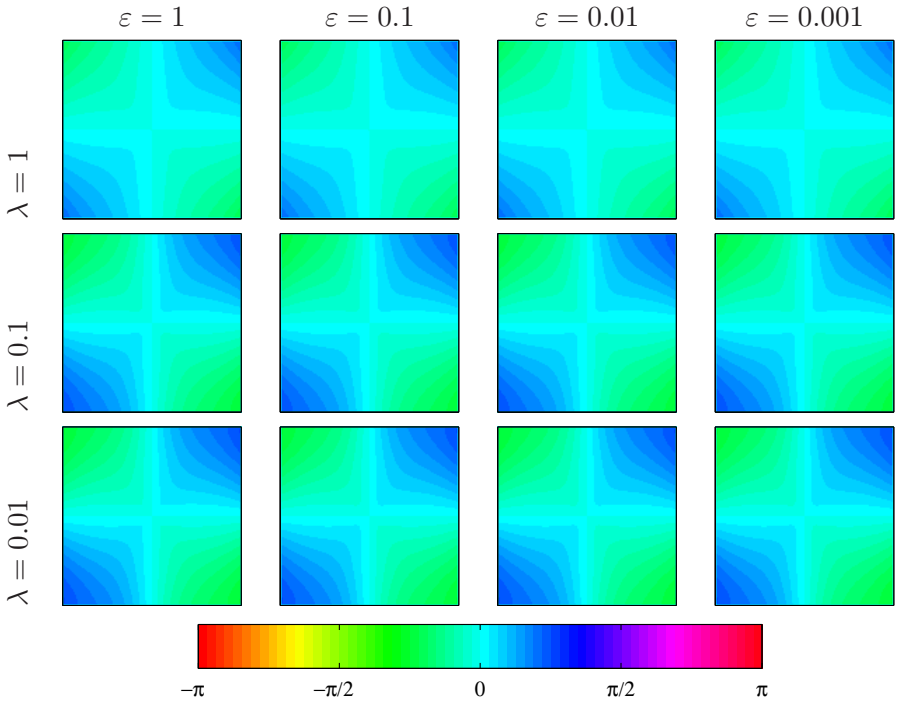


Figure 4.7: Angles of solution fields for impulse example with image-driven elastic regularizers for various values of ε and λ.

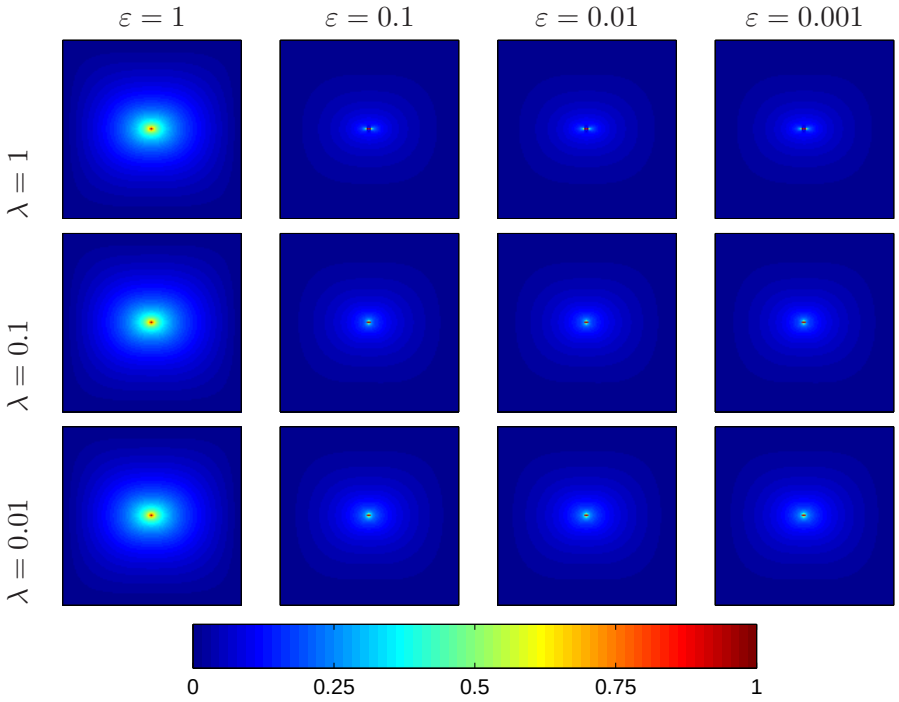


Figure 4.8: Magnitudes of solution fields for impulse example with flow-driven elastic regularizers for various values of ε and λ.

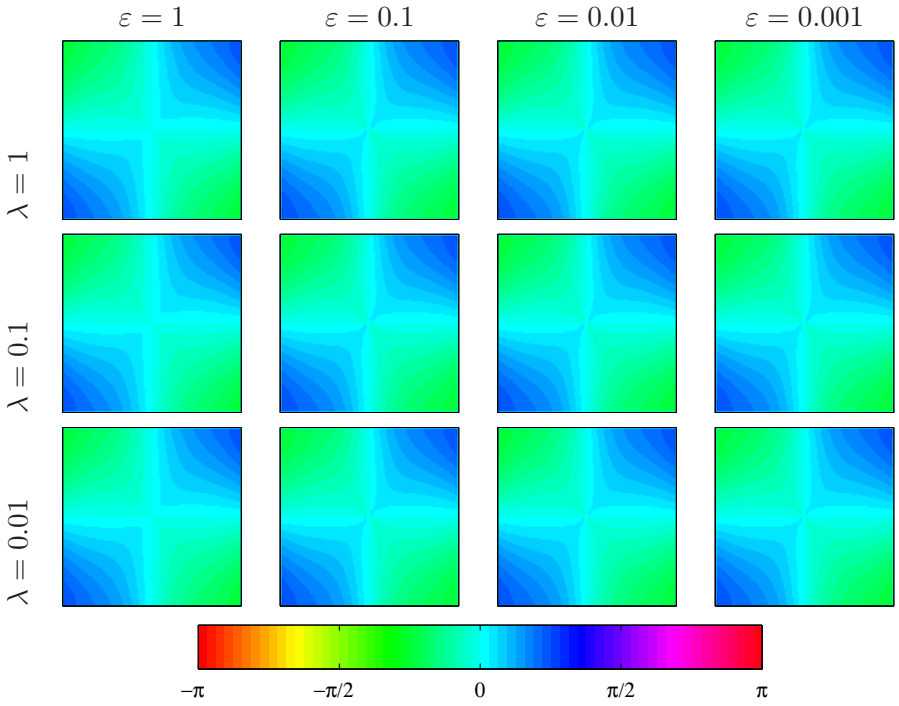


Figure 4.9: Angles of solution fields for impulse example with flow-driven elastic regularizers for various values of ε and λ.

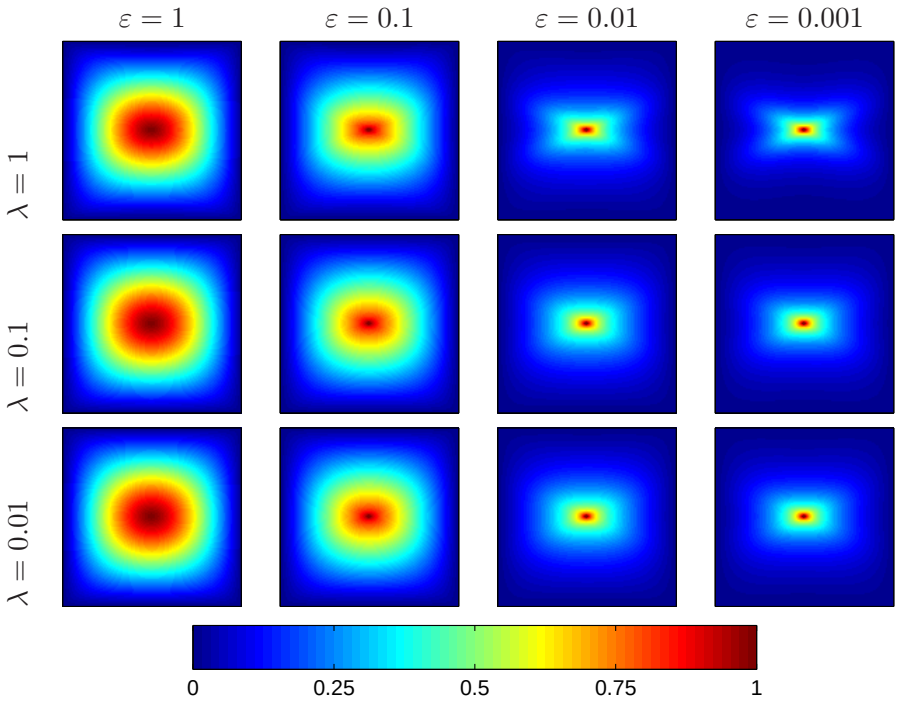


Figure 4.10: Magnitudes of solution fields for impulse example with image-driven second-order elastic regularizers for various values of ε and λ.

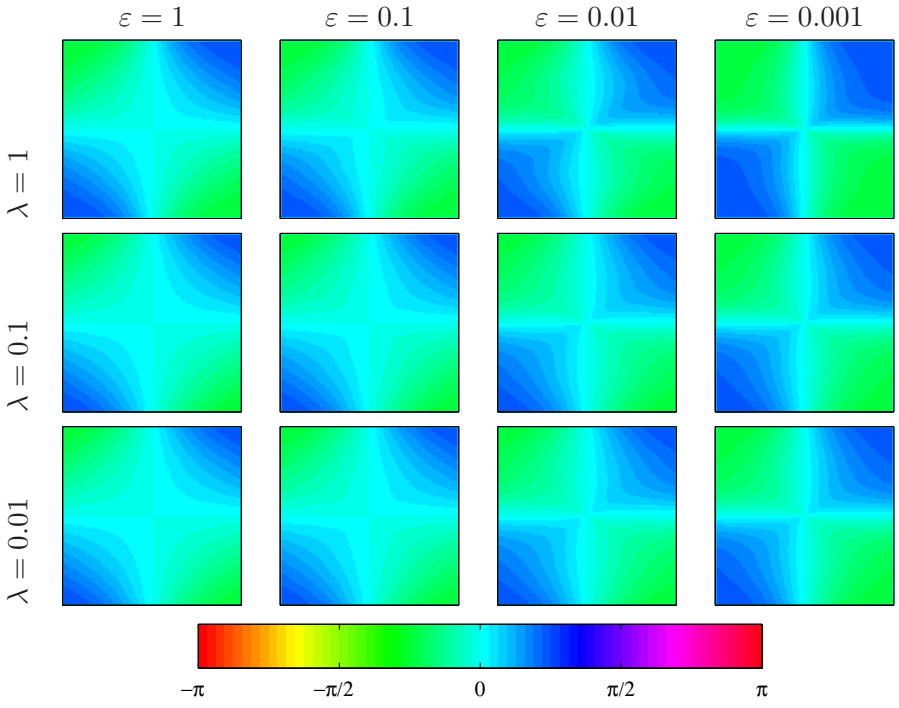


Figure 4.11: Angles of solution fields for impulse example with image-driven second-order elastic regularizers for various values of ε and λ.

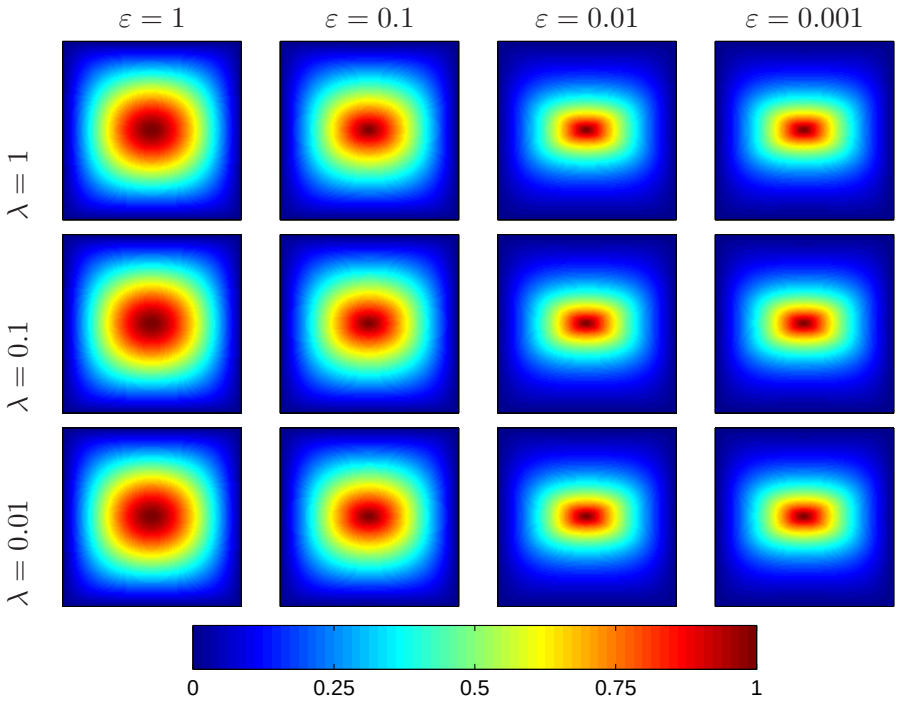


Figure 4.12: Magnitudes of solution fields for impulse example with flow-driven second-order elastic regularizers for various values of ε and λ.

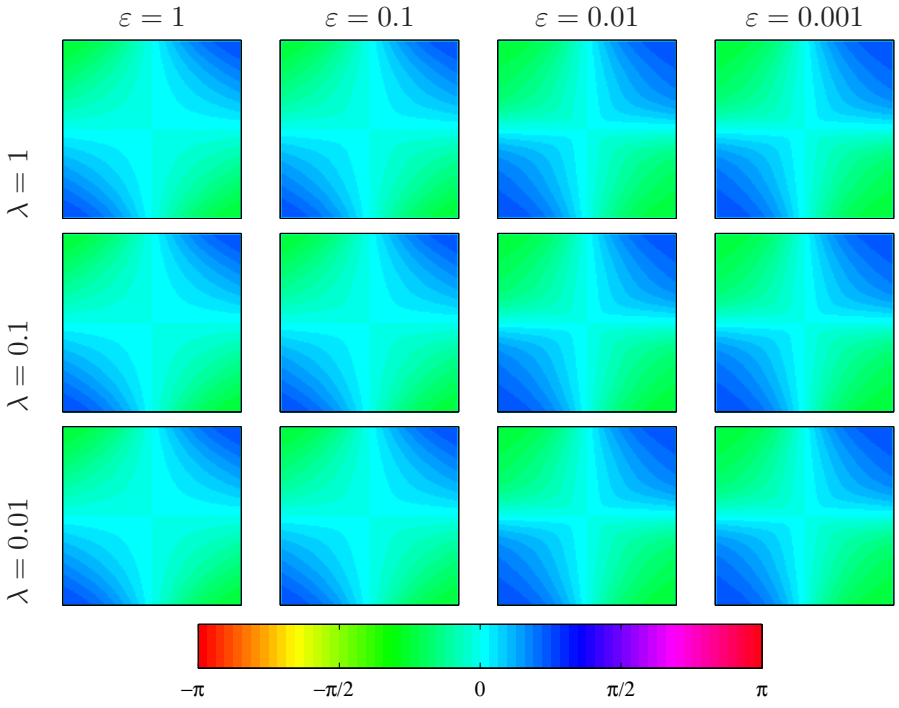


Figure 4.13: Angles of solution fields for impulse example with flow-driven second-order elastic regularizers for various values of ε and λ.

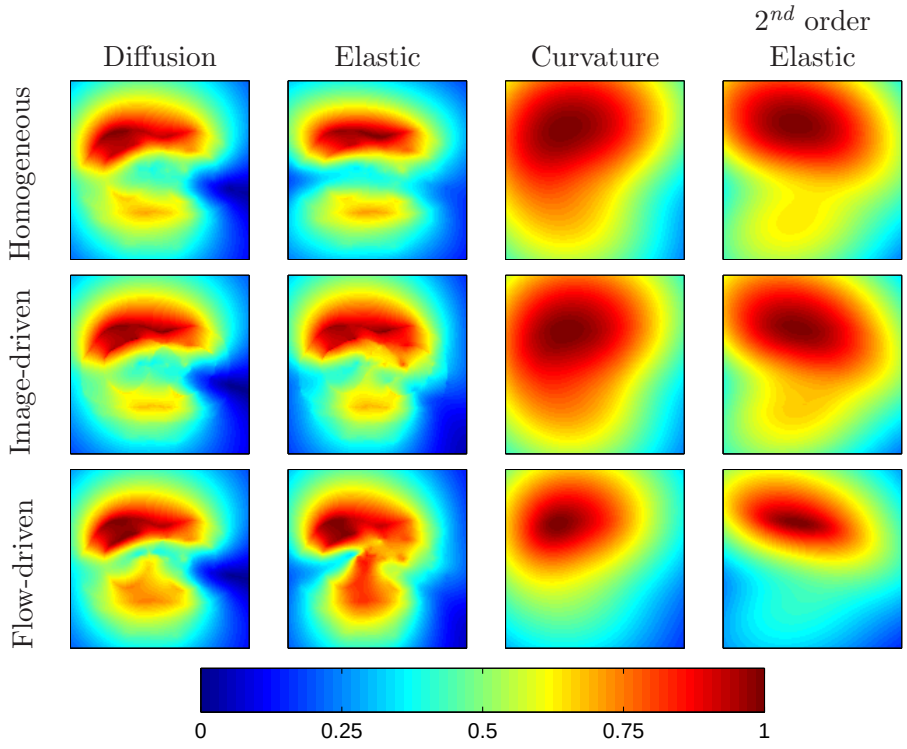


Figure 4.14: Magnitudes of solution fields for CT slice example with various regularizers.

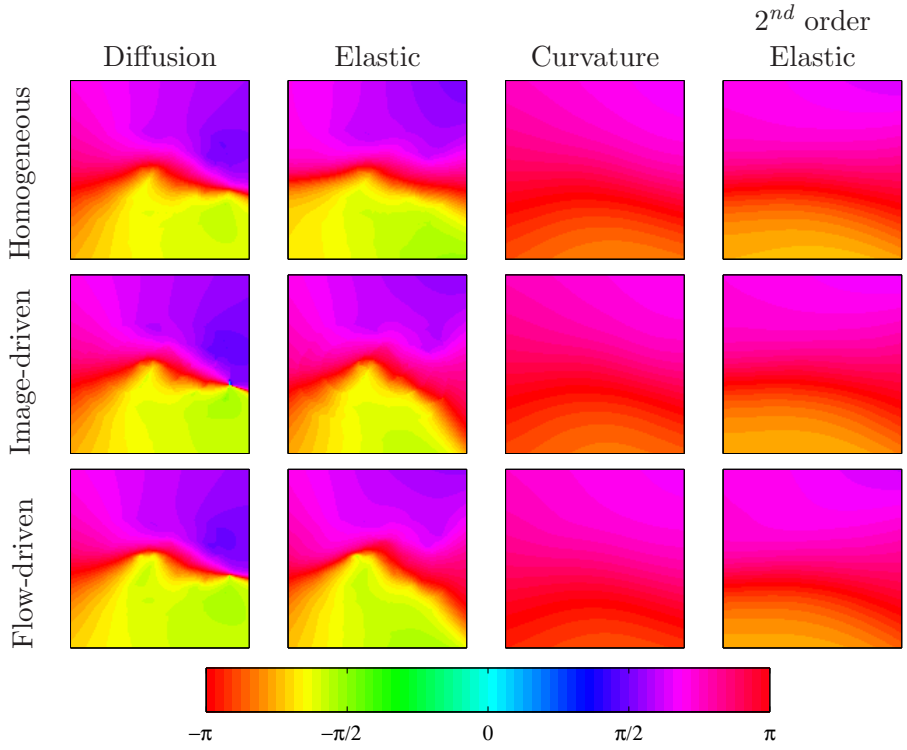


Figure 4.15: Angles of solution fields for CT slice example with various regularizers.

Chapter 5

Solving the Variational Registration Problem

In the previous chapters of this thesis, we have presented a wide variety of dissimilarity measures and regularizers that are useful for registration. We are now interested in methods for finding the deformation field that minimizes a given combination of dissimilarity measure and regularizer, as in (1.4), which we repeat here for reference:

$$\min_{\mathbf{u}} \qquad \mathcal{S}(\mathbf{u}) := \mathcal{R}(\mathbf{u}) + \alpha\, \mathcal{I}(R, F^{\mathbf{u}})\,. \tag{5.1}$$

Using variational techniques, it can be shown that any stationary point of $\mathcal{E}(\mathbf{u})$ satisfies a system of partial differential equations known as the Euler-Lagrange equations. In general, for any regularizer and dissimilarity measure, it is possible to approximate the solution to the Euler-Lagrange equations by solving a succession of linear PDE systems. One important aspect of rapid registration algorithms, therefore, is the use of linear systems solvers that exploit any structure in the Euler-Lagrange equations for computational efficiency.

In this chapter, we first describe the PDE systems that arise from minimizing (5.1). We then illustrate how to construct the resulting succession of linear PDE systems, taking into account both the linear and nonlinear regularizers presented in Chapter 4. Next, we describe how to discretize and approximate the solution to the linear PDE systems. We first focus on general iterative solvers that can be applied for any regularizer, and then we turn our focus to rapid methods (such as the discrete cosine transform [63] and successive Gaussian convolution [87]) that exploit the structure of specific regularizers. Then, we show how Fourier methods and successive Gaussian convolution can be applied to registration with

any of the homogeneous regularizers. Finally, we experiment with some of these techniques on the problem of registering serial chest CT examinations of patients with lung nodules.

5.1 Successive Systems of Linear Partial Differential Equations

In Chapter 1, we showed that the Euler-Lagrange equations corresponding to (5.1) are given by:

$$\mathcal{A}(\mathbf{u}) = -\,\alpha\,\mathcal{P}(\mathbf{x}; R, F^{\mathbf{u}}) \qquad \forall \mathbf{x} \in \Omega, \tag{5.2}$$

$$\mathcal{B}[\mathbf{u}] = 0 \qquad \forall \mathbf{x} \in \partial\Omega. \tag{5.3}$$

In general, it is possible to devise iterative approaches for solving (5.2)–(5.3), either via fixed point iteration or steepest descent. If the partial differential operator $\mathcal{A}$ is linear (if the regularizer is homogeneous or image-driven nonhomogeneous with weighting functions that do not depend on $\mathbf{u}$), then these iterative approaches yield a succession of linear PDE systems. If $\mathcal{A}$ is nonlinear (flow-driven nonhomogeneous regularizers or image-driven nonhomogeneous regularizers with weighting functions that depend on $\mathbf{u}$), then the succession of PDE systems are nonlinear; however they can be made to be linear by employing an inner Newton-Raphson style iteration.

In this section, we illustrate how to construct the successive linear PDE systems arising from both the fixed-point and steepest descent iterations. Along the way, we discuss why it is sometimes necessary (e.g., when Neumann boundary conditions are chosen) to avoid the fixed-point iteration. We then discuss how these iterative schemes can be applied to the velocity field $\mathbf{v}$ instead of the displacement field $\mathbf{u}$ to handle large deformation problems in a fluid-like manner.

5.1.1 Semilinear Euler-Lagrange Equations

We will assume in this section that the partial differential operator $\mathcal{A}$ is linear in $\mathbf{u}$; this assumption yields Euler-Lagrange equations that are semilinear. This allows us to denote $\mathcal{A}(\mathbf{u})$ by $\mathcal{A}\mathbf{u}$. Linear partial differential operators arise from any of the homogeneous regularizers or from any of the nonhomogeneous image-driven regularizers with weighting

functions that do *not* depend on $\mathbf{u}$. The other nonhomogeneous image-driven and flow-driven regularizers yield nonlinear partial differential operators and will be addressed in the next subsection.

If $\mathcal{A}$ is linear in $\mathbf{u}$, one way to approximate the solution of (5.2)–(5.3) is by fixed-point iteration. If an initial guess $\mathbf{u}^{(0)}$ is chosen for the displacement, then updated iterates for the displacement are given implicitly by the iteration:

$$\hat{\mathcal{A}}\mathbf{u}^{(k+1)}(\mathbf{x}) = \mathbf{f}\Big(\mathbf{x}, \mathbf{u}^{(k)}(\mathbf{x})\Big)\,, \tag{5.4}$$

where $\mathbf{f}\big(\mathbf{x}, \mathbf{u}^{(k)}(\mathbf{x})\big) := -\,\alpha^{-1}\,\mathcal{P}\Big(\mathbf{x}; R, F^{\mathbf{u}^{(k)}}\Big)$, and where $\hat{\mathcal{A}}$ is the extension of $\mathcal{A}$ to $\partial\Omega$ such that the boundary conditions are satisfied. For each k, (5.4) is a linear PDE system; hence, the fixed point iteration can be thought of as a way to approximate the solution to the semilinear Euler-Lagrange equations by solving a succession of linear partial differential equations.

If $\hat{\mathcal{A}}$ is nonsingular, each iteration of (5.4) has a unique solution. However, $\hat{\mathcal{A}}$ is not necessarily guaranteed to be nonsingular. Consider, for example, the positive semidefinite Laplacian operator $\mathcal{A} = -\Delta$ with Neumann boundary conditions $\mathcal{B}[\mathbf{u}] = \langle \nabla\mathbf{u}(\mathbf{x})\,, \vec{\mathbf{n}}(\mathbf{x})\rangle$ (which arises from the homogeneous diffusion regularizer). In this case, $\hat{\mathcal{A}}$ has a nontrivial null space comprised of all constant displacement fields. Therefore, if we attempt to apply the fixed point iteration (5.4), there are either infinitely many or zero solutions, depending on whether or not $\mathbf{f}\big(\mathbf{x}, \mathbf{u}^{(k)}(\mathbf{x})\big)$ is in the range of $\hat{\mathcal{A}}$. (Note that if Dirichlet boundary conditions are chosen with the Laplacian operator $\mathcal{A} = -\Delta$, $\hat{\mathcal{A}}$ is nonsingular and the fixed-point iteration (5.4) is guaranteed to have a unique solution for each k. See [45] for a further discussion on existence and uniqueness of solutions to the Poisson equation when Dirichlet or Neumann boundary conditions are employed.)

In order to overcome this problem, we use the standard trick [73] of introducing an artificial time variable t and embedding (5.2) in the parabolic PDE system:

$$\begin{gathered}\partial_t\mathbf{u}(\mathbf{x},t) + \mathcal{A}\mathbf{u}(\mathbf{x},t) = \mathbf{f}(\mathbf{x},t,\mathbf{u}(\mathbf{x},t))\,, \qquad \mathbf{x}\in\Omega,\ t>0, \\ \mathbf{u}(\mathbf{x},0) = \mathbf{u}^{(0)}, \\ \mathbf{B}[\mathbf{u}(\mathbf{x},t)] = 0, \qquad \mathbf{x}\in\partial\Omega,\ t\geq 0.\end{gathered} \tag{5.5}$$

This parabolic system also arises from applying the steepest descent algorithm to solve (5.1). The stationary solution of (5.5) is given by:

$$\mathbf{u}^*(\mathbf{x}) = \lim_{t\to\infty} \mathbf{u}(\mathbf{x}, t). \tag{5.6}$$

This stationary solution satisfies $\partial_t \mathbf{u}^*(\mathbf{x}) = 0$, and, hence, is a solution to (5.2)–(5.3).

Now, in order to approximate the stationary solution of (5.5), we discretize either forwards or backwards in time. A forward time discretization with time step τ yields the explicit iteration:

$$\frac{\mathbf{u}^{(k+1)}(\mathbf{x}) - \mathbf{u}^{(k)}(\mathbf{x})}{\tau} = -\hat{\mathcal{A}}\mathbf{u}^{(k)}(\mathbf{x}) + \mathbf{f}\Big(\mathbf{x}, \mathbf{u}^{(k)}(\mathbf{x})\Big), \tag{5.7}$$

whereas a backwards time discretization yields the implicit iteration:

$$\frac{\mathbf{u}^{(k+1)}(\mathbf{x}) - \mathbf{u}^{(k)}(\mathbf{x})}{\tau} = -\hat{\mathcal{A}}\mathbf{u}^{(k+1)}(\mathbf{x}) + \mathbf{f}\Big(\mathbf{x}, \mathbf{u}^{(k)}(\mathbf{x})\Big). \tag{5.8}$$

The explicit iteration, although simple to implement, requires such a small time step for stability that a huge number of iterations may be required. The implicit iteration does not suffer from the same instability, so larger time steps can be used. If we rearrange the terms in (5.8), we see that the implicit iteration can be written as:

$$\Big[\mathbf{I} + \tau\hat{\mathcal{A}}\Big]\, \mathbf{u}^{(k+1)}(\mathbf{x}) = \mathbf{u}^{(k)}(\mathbf{x}) + \tau\mathbf{f}\Big(\mathbf{x}, \mathbf{u}^{(k)}(\mathbf{x})\Big). \tag{5.9}$$

Therefore, we see that for any of the homogeneous or linear image-driven regularizers, a solution to the registration problem (5.1) can be approximated by the iteration:

$$\mathbf{A}\mathbf{u}^{(k+1)}(\mathbf{x}) = \mathbf{b}\Big(\mathbf{x}, \mathbf{u}^{(k)}(\mathbf{x})\Big), \tag{5.10}$$

with $\mathbf{A} := \Big[\mathbf{I} + \tau\hat{\mathcal{A}}\Big]$ and $\mathbf{b}\big(\mathbf{x}, \mathbf{u}^{(k)}(\mathbf{x})\big) := \mathbf{u}^{(k)}(\mathbf{x}) + \tau\mathbf{f}\big(\mathbf{x}, \mathbf{u}^{(k)}(\mathbf{x})\big)$. We refer to this iteration as the *steepest descent* iteration.

As with the fixed-point iteration (5.4), (5.10) is now a linear PDE system for each k; hence, the iteration (5.10) can also be thought of as a way to approximate the solution to the semilinear Euler-Lagrange equations by solving a succession of linear partial differential equations. The benefit of (5.10) over the fixed-point iteration (5.4) is that (5.10) is guaranteed to have a unique solution at each iteration.

For quick reference, we list the algorithms for solving the Euler-Lagrange equations by fixed-point iteration and steepest descent iteration in Algorithms 1 and 2, respectively. The termination criterion is given as a function $d\big(\mathbf{u}^{(k)}, \mathbf{u}^{(k-1)}\big)$ that measures the distance between successive estimates of the displacement field. This function can be chosen, for example, to be the 2-norm or ∞-norm of the differences.

Algorithm 1 Fixed-Point Iteration for Registration with Linear $\mathcal{A}$

Set $\mathbf{u}^{(0)} = 0$, $k = 0$; Select τ such that $0 < \tau << 1$.
repeat
  Compute $\mathbf{f}^{(k)}(\mathbf{x}) = -\alpha^{-1}\,\mathcal{P}\Big(\mathbf{x}; R, F^{\mathbf{u}^{(k)}}\Big)$.
  Solve the linear PDE system $\hat{\mathcal{A}}\mathbf{u}^{(k+1)} = \mathbf{f}^{(k)}$ for $\mathbf{u}^{(k+1)}$.
  Increment k.
until $d\big(\mathbf{u}^{(k)}, \mathbf{u}^{(k-1)}\big) < \tau$.

Algorithm 2 Steepest Descent Iteration for Registration with Linear $\mathcal{A}$

Set $\mathbf{u}^{(0)} = 0$, $k = 0$; Select τ such that $0 < \tau << 1$.
repeat
  Compute $\mathbf{b}^{(k)}(\mathbf{x}) = \mathbf{u}^{(k)}(\mathbf{x}) - \tau\alpha^{-1}\,\mathcal{P}\Big(\mathbf{x}; R, F^{\mathbf{u}^{(k)}}\Big)$.
  Solve the linear PDE system $\Big[\mathbf{I} + \tau\hat{\mathcal{A}}\Big]\,\mathbf{u}^{(k+1)} = \mathbf{b}^{(k)}$ for $\mathbf{u}^{(k+1)}$.
  Increment k.
until $d\big(\mathbf{u}^{(k)}, \mathbf{u}^{(k-1)}\big) < \tau$.

5.1.2 Nonlinear Euler-Lagrange Equations

If one of the nonlinear image-driven or flow-driven regularizers are selected, then the resulting Euler-Lagrange equations (5.2) are fully nonlinear. A fixed-point iteration can be constructed in the manner of (5.4); i.e.,

$$\hat{\mathcal{A}}\Big(\mathbf{u}^{(k+1)}(\mathbf{x})\Big) = \mathbf{f}\Big(\mathbf{x}, \mathbf{u}^{(k)}(\mathbf{x})\Big)\,. \tag{5.11}$$

Unlike (5.4), (5.11) is a *nonlinear* PDE system for each k. In order to approximate the solutions to these nonlinear PDE systems, we utilize an inner Newton-Raphson style iteration.

Suppose we want to estimate $\mathbf{u}^{(k+1)}$. We introduce an inner iteration given by:

$$\mathbf{u}^{(k,l+1)}(\mathbf{x}) = \mathbf{u}^{(k,l)}(\mathbf{x}) + \mathbf{du}^{(k,l)}(\mathbf{x})\,, \tag{5.12}$$

and equate $\mathbf{u}^{(k+1)}(\mathbf{x}) := \mathbf{u}^{(k+1,0)}(\mathbf{x})$ with the limiting value of $\mathbf{u}^{(k,l)}(\mathbf{x})$ as l becomes large. Therefore, for large enough l, we can rewrite (5.11) as:

$$\hat{\mathcal{A}}\Big(\mathbf{u}^{(k,l+1)}(\mathbf{x})\Big) = \mathbf{f}\Big(\mathbf{x}, \mathbf{u}^{(k,0)}(\mathbf{x})\Big). \tag{5.13}$$

For small $\mathbf{du}^{(k,l)}(\mathbf{x})$, we can make the first-order Taylor series approximation:

$$\hat{\mathcal{A}}\Big(\mathbf{u}^{(k,l+1)}(\mathbf{x})\Big) \approx \hat{\mathcal{A}}\Big(\mathbf{u}^{(k,l)}(\mathbf{x})\Big) + \Big\langle \nabla_{\mathbf{u}}\hat{\mathcal{A}}(\mathbf{u}(\mathbf{x}))\,|_{\mathbf{u}=\mathbf{u}^{(k,l)}}, \mathbf{du}^{(k,l)}(\mathbf{x})\Big\rangle. \tag{5.14}$$

Substituting (5.14) into (5.13) and simplifying yields the iteration:

$$\Big\langle \nabla_{\mathbf{u}}\hat{\mathcal{A}}(\mathbf{u}(\mathbf{x}))\,|_{\mathbf{u}=\mathbf{u}^{(k,l)}}, \mathbf{du}^{(k,l)}(\mathbf{x})\Big\rangle = \mathbf{f}\Big(\mathbf{x}, \mathbf{u}^{(k,0)}(\mathbf{x})\Big) - \hat{\mathcal{A}}\Big(\mathbf{u}^{(k,l)}(\mathbf{x})\Big). \tag{5.15}$$

This is a *linear* system of partial differential equations in the unknown increment $\mathbf{du}^{(k,l)}$. If we nest this iteration inside the fixed-point iteration (5.11), we see that the solution to the nonlinear Euler-Lagrange equations can be approximated by successively solving a set of linear PDE systems. This nested algorithm is summarized by Algorithm 3.

Algorithm 3 Fixed-Point Iteration for Registration with Nonlinear $\mathcal{A}$

Set $\mathbf{u}^{(0,0)} = 0$, $k = 0$; Select τ_1, τ_2 such that $0 < \tau_1, \tau_2 << 1$.
repeat
  Set $l = 0$.
  repeat
    Compute $\mathbf{b}^{(k,l)}(\mathbf{x}) = -\alpha^{-1}\,\mathcal{P}\Big(\mathbf{x}; R, F^{\mathbf{u}^{(k,0)}}\Big) - \hat{\mathcal{A}}\big(\mathbf{u}^{(k,l)}(\mathbf{x})\big)$.
    Solve the linear PDE system $\Big\langle \nabla_{\mathbf{u}}\hat{\mathcal{A}}(\mathbf{u})\,|_{\mathbf{u}=\mathbf{u}^{(k,l)}}, \mathbf{du}^{(k,l)}\Big\rangle = \mathbf{b}^{(k,l)}$ for $\mathbf{du}^{(k,l)}$.
    Set $\mathbf{u}^{(k,l+1)} = \mathbf{u}^{(k,l)} + \mathbf{du}^{(k,l)}$.
    Increment l.
  until $d\big(\mathbf{u}^{(k,l)}, \mathbf{u}^{(k,l-1)}\big) < \tau_2$.
  Set $\mathbf{u}^{(k+1,0)} = \mathbf{u}^{(k,l)}$.
  Increment k.
until $d\big(\mathbf{u}^{(k,0)}, \mathbf{u}^{(k-1,0)}\big) < \tau_1$.

For the semi-linear Euler-Lagrange equations, we established that the fixed-point iteration can run into problems if the linear partial differential operator has a nontrivial null space. In fact, this problem can occur with the *nonlinear* Euler-Lagrange equations as well. If we consider the flow-driven isotropic diffusion regularizer with Neumann boundary conditions, it is easy to see that the partial differential operator $\hat{\mathcal{A}}^{\textit{fb-iso}}_{\textit{diffusion}}$ has constant

displacement fields in its kernel. For this reason, we can use the same trick as in the previous section of introducing and artificial time variable and embedding the nonlinear Euler-Lagrange equations into a parabolic PDE system:

$$\begin{aligned} \partial_t \mathbf{u}(\mathbf{x},t) + \mathcal{A}(\mathbf{u}(\mathbf{x},t)) &= \mathbf{f}(\mathbf{x},t,\mathbf{u}(\mathbf{x},t)), \qquad \mathbf{x}\in\Omega,\ t>0, \\ \mathbf{u}(\mathbf{x},0) &= \mathbf{u}^{(0)}, \\ \mathbf{B}[\mathbf{u}(\mathbf{x},t)] &= 0, \qquad \mathbf{x}\in\partial\Omega,\ t\geq 0. \end{aligned} \tag{5.16}$$

Introducing a backwards time discretization yields the (nonlinear) implicit iteration:

$$\mathbf{u}^{(k+1)}(\mathbf{x}) + \tau\hat{\mathcal{A}}\Big(\mathbf{u}^{(k+1)}(\mathbf{x})\Big) = \mathbf{u}^{(k)}(\mathbf{x}) + \tau\mathbf{f}\Big(\mathbf{x},\mathbf{u}^{(k)}(\mathbf{x})\Big). \tag{5.17}$$

Solutions to the successive nonlinear PDE systems in (5.17) can be approximated by utilizing an inner Newton-Raphson style iteration. The inner iteration can be derived in a similar manner to (5.15), and is given by:

$$\begin{aligned} \mathbf{du}^{(k,l)}(\mathbf{x}) + \tau\left\langle \nabla_{\mathbf{u}}\hat{\mathcal{A}}(\mathbf{u}(\mathbf{x}))\,|_{\mathbf{u}=\mathbf{u}^{(k,l)}}, \mathbf{du}^{(k,l)}(\mathbf{x})\right\rangle \quad &= \\ \mathbf{u}^{(k,0)} - \mathbf{u}^{(k,l)} + \tau\left(\mathbf{f}\big(\mathbf{x},\mathbf{u}^{(k,0)}(\mathbf{x})\big) - \hat{\mathcal{A}}\big(\mathbf{u}^{(k,l)}(\mathbf{x})\big)\right). \end{aligned} \tag{5.18}$$

If we nest this iteration inside the steepest descent iteration (5.17), we see that we can now approximate the solution to the nonlinear Euler-Lagrange equations by successively solving a set of linear PDE systems. The use of this iteration comes with the guarantee that there is a unique solution for each (outer) iteration. The nested steepest descent algorithm is summarized in Algorithm 4.

The only remaining task is to determine the linear partial differential terms (i.e., the left hand sides of (5.15)) for each of the nonlinear image-driven and flow-driven isotropic regularizers. Due to space considerations, we simply present these terms in the following subsections. They can all be derived using principles from multivariate calculus.

Algorithm 4 Steepest Descent Iteration for Registration with Nonlinear $\mathcal{A}$

Set $\mathbf{u}^{(0,0)} = 0$, $k = 0$; Select τ_1, τ_2 such that $0 < \tau_1, \tau_2 << 1$.
repeat
  Set $l = 0$.
  repeat
    Compute $\mathbf{b}^{(k,l)}(\mathbf{x})$ = right hand side of (5.18).
    Solve the linear PDE system $\mathbf{du}^{(k,l)} + \tau \left\langle \nabla_{\mathbf{u}} \hat{\mathcal{A}}(\mathbf{u}) \mid_{\mathbf{u}=\mathbf{u}^{(k,l)}}, \mathbf{du}^{(k,l)} \right\rangle = \mathbf{b}^{(k,l)}$
      for $\mathbf{du}^{(k,l)}$.
    Set $\mathbf{u}^{(k,l+1)} = \mathbf{u}^{(k,l)} + \mathbf{du}^{(k,l)}$.
    Increment l.
  until $d\big(\mathbf{u}^{(k,l)}, \mathbf{u}^{(k,l-1)}\big) < \tau_2$.
  Set $\mathbf{u}^{(k+1,0)} = \mathbf{u}^{(k,l)}$.
  Increment k.
until $d\big(\mathbf{u}^{(k,0)}, \mathbf{u}^{(k-1,0)}\big) < \tau_1$.

5.1.2.1 Linear Terms for the Nonlinear Image-driven Isotropic Regularizers

For the nonlinear image-driven isotropic diffusion regularizer, the left hand side of (5.15) reduces to the following linear partial differential expression in $\mathbf{du}^{(k,l)}$:

$$\begin{aligned} \left\langle \nabla_{\mathbf{u}} \mathcal{A}(\mathbf{u}) \mid_{\mathbf{u}=\mathbf{u}^{(k,l)}}, \mathbf{du}^{(k,l)} \right\rangle \;=\; & -\mathrm{div}\left(\beta^{\mathbf{u}^{(k,l)}}(\mathbf{x})\, \nabla \mathbf{du}^{(k,l)}\right) \\ & - \nabla \beta^{\mathbf{u}^{(k,l)}}(\mathbf{x})\ \mathrm{tr}\left\langle \nabla \mathbf{u}^{(k,l)}, \nabla \mathbf{du}^{(k,l)} \right\rangle \\ & + \mathrm{div}\left(\left\langle \nabla \beta^{\mathbf{u}^{(k,l)}}(\mathbf{x}), \mathbf{du}^{(k,l)} \right\rangle \nabla \mathbf{u}^{(k,l)}\right) \\ & + \frac{1}{2}\left(\sum_{j=1}^{n} \left\| \nabla u_j^{(k,l)} \right\|^2\right) \left\langle \nabla \nabla^{\mathrm{T}} \beta^{\mathbf{u}^{(k,l)}}(\mathbf{x}), \mathbf{du}^{(k,l)} \right\rangle . \end{aligned} \tag{5.19}$$

One potential issue with using (5.19) in practice is that geometric warping has to be performed inside each inner iteration to form $\beta^{\mathbf{u}^{(k,l)}}(\mathbf{x})$. To avoid this heavy computational burden, we choose to update $\beta^{\mathbf{u}}(\mathbf{x})$ only in the outer iteration. In effect, this means that for each k, we can consider $\beta^{\mathbf{u}^{(k,l)}}(\mathbf{x}) = \beta^{\mathbf{u}^{(k,0)}}(\mathbf{x}) = \beta^{\mathbf{u}^{(k)}}(\mathbf{x})$ for all l.

This choice of updating the weighting function only in the outer iteration simplifies the partial differential expression (5.19) to the following:

$$\begin{aligned} \left\langle \nabla_{\mathbf{u}} \mathcal{A}(\mathbf{u}) \mid_{\mathbf{u}=\mathbf{u}^{(k,l)}}, \mathbf{du}^{(k,l)} \right\rangle \;=\; & -\mathrm{div}\left(\beta^{\mathbf{u}^{(k)}}(\mathbf{x})\, \nabla \mathbf{du}^{(k,l)}\right) \\ & - \nabla \beta^{\mathbf{u}^{(k)}}(\mathbf{x})\ \mathrm{tr}\left\langle \nabla \mathbf{u}^{(k,l)}, \nabla \mathbf{du}^{(k,l)} \right\rangle . \end{aligned} \tag{5.20}$$

For the nonlinear image-driven isotropic versions of the elastic, curvature and second-

order elastic regularizers, we also choose to update the weighting functions only in the outer iteration. We present the resulting partial differential terms in Table 5.1.

5.1.2.2 Linear Terms for the Flow-driven Isotropic Regularizers

Any attempt to compute the left hand side of (5.15) for one of the flow-driven isotropic regularizers is a prescription for a rapid descent into madness. Luckily, we can resort to a trick introduced by Brox *et al.* [15] in order to approximate the partial differential operator. Brox suggests introducing a lag in the evaluation of the nonlinear portion of $\mathcal{A}_{diffusion}^{fb\text{-}iso}$. In essence, this allows us to approximate the left hand side of (5.15) for the flow-driven isotropic diffusion regularizer by:

$$\left\langle \nabla_{\mathbf{u}} \mathcal{A}(\mathbf{u}) \mid_{\mathbf{u}=\mathbf{u}^{(k,l)}}, \mathbf{du}^{(k,l)} \right\rangle \approx -\operatorname{div} \left(\beta' \left(\sum_{j=1}^{n} \left\| \nabla u_j^{(k,l)} \right\|^2 \right) \nabla \mathbf{du}^{(k,l)} \right). \tag{5.21}$$

The same trick can be used to approximate the linear partial differential terms for the flow-driven isotropic versions of the elastic, curvature, and second-order elastic regularizers. We present the resulting partial differential terms in Table 5.2.

5.1.3 Strategies for Large Deformations

For regularizers based on the displacement field, Modersitzki [63] notes that the Euler-Lagrange equations (5.2)–(5.3) are only valid for small deformations. To capture larger deformations, Modersitzki suggests generalizing the fluid registration approach to the other regularizers. Fluid registration, as originally proposed by Christensen [31], can be achieved by applying the elastic partial differential operator $\mathcal{A}$ to the velocity field of the deformation. The velocity field $\mathbf{v}$ is related to the displacement field $\mathbf{u}$ by the material derivative:

$$\mathbf{v}(\mathbf{x}, t) = \frac{d}{dt}\mathbf{u}(\mathbf{x}, t) = \partial_t \mathbf{u}(\mathbf{x}, t) + (\nabla \mathbf{u}(\mathbf{x}, t))^{\mathrm{T}} \mathbf{v}(\mathbf{x}, t). \tag{5.22}$$

To generalize the fluid registration approach to other homogeneous regularizers, we can simply applying the corresponding partial differential operator to the velocity field. (Admittedly, this approach is *ad hoc* in the sense that the resulting PDE systems are not Euler-Lagrange equations arising from the minimization of a functional of the velocity field.)

Regularizer	**Partial differential term** $\left\langle \nabla_{\mathbf{u}} \mathcal{A}(\mathbf{u})\,|_{\mathbf{u}=\mathbf{u}^{(k,l)}}, \mathbf{du}^{(k,l)} \right\rangle$
Diffusion	$-\operatorname{div}\left(\beta^{\mathbf{u}^{(k)}}(\mathbf{x})\,\nabla \mathbf{du}^{(k,l)}\right) - \nabla\beta^{\mathbf{u}^{(k)}}(\mathbf{x})\ \operatorname{tr}\left\langle \nabla \mathbf{u}^{(k,l)}, \nabla \mathbf{du}^{(k,l)} \right\rangle$
Elastic	$-\operatorname{div}\left(\beta_1^{\mathbf{u}^{(k)}}(\mathbf{x})\left(\nabla \mathbf{du}^{(k,l)} + \left(\nabla \mathbf{du}^{(k,l)}\right)^{\mathrm{T}}\right)\right) - \nabla\left(\beta_2^{\mathbf{u}^{(k)}}(\mathbf{x})\operatorname{div}\mathbf{du}^{(k,l)}\right) - \frac{1}{2}\left(\operatorname{div}\mathbf{u}^{(k,l)}\right)\left(\operatorname{div}\mathbf{du}^{(k,l)}\right)\nabla\beta_2^{\mathbf{u}^{(k)}}(\mathbf{x})$ $-\frac{1}{4}\left(\sum_{j,m=1}^{n}\left(\partial_{x_j}u_m^{(k,l)} + \partial_{x_m}u_j^{(k,l)}\right)\left(\partial_{x_j}du_m^{(k,l)} + \partial_{x_m}du_j^{(k,l)}\right)\right)\nabla\beta_1^{\mathbf{u}^{(k)}}(\mathbf{x})$
Curvature	$\Delta\left(\beta^{\mathbf{u}^{(k)}}(\mathbf{x})\,\Delta \mathbf{du}^{(k,l)}\right) - \left\langle \Delta \mathbf{u}^{(k,l)}, \Delta \mathbf{du}^{(k,l)} \right\rangle \nabla\beta^{\mathbf{u}^{(k)}}(\mathbf{x})$
2^{nd}-order Elastic	$\sum_{j=1}^{n}\partial_{x_j}\operatorname{div}\left(\beta_1^{\mathbf{u}^{(k)}}(\mathbf{x})\,\partial_{x_j}\left[\nabla \mathbf{du}^{(k,l)} + 2\left(\nabla \mathbf{du}^{(k,l)}\right)^{\mathrm{T}}\right]\right) + \nabla\operatorname{div}\left(\beta_2^{\mathbf{u}^{(k)}}(\mathbf{x})\,\nabla\operatorname{div}\mathbf{du}^{(k,l)}\right)$ $-\frac{1}{3}\left(\sum_{j,m,r=1}^{n}\left(\partial^2_{x_j,x_m}u_r^{(k,l)} + \partial^2_{x_j,x_r}u_m^{(k,l)} + \partial^2_{x_m,x_r}u_j^{(k,l)}\right)\left(\partial^2_{x_j,x_m}du_r^{(k,l)} + \partial^2_{x_j,x_r}du_m^{(k,l)} + \partial^2_{x_m,x_r}du_j^{(k,l)}\right)\right)\nabla\beta_1^{\mathbf{u}^{(k)}}(\mathbf{x})$ $-\left\langle \nabla\operatorname{div}\mathbf{u}^{(k,l)}, \nabla\operatorname{div}\mathbf{du}^{(k,l)} \right\rangle \nabla\beta_2^{\mathbf{u}^{(k)}}(\mathbf{x})$

Table 5.1: Linear partial differential terms for nonlinear image-driven isotropic regularizers.

Regularizer	**Approximation of partial differential term** $\left\langle \nabla_{\mathbf{u}} \mathcal{A}(\mathbf{u}) \,|_{\mathbf{u}=\mathbf{u}^{(k,l)}}, \mathbf{du}^{(k,l)} \right\rangle$
Diffusion	$-\operatorname{div}\left(\beta'\left(\sum_{j=1}^{n}\left\|\nabla u_j^{(k,l)}\right\|^2\right)\nabla \mathbf{du}^{(k,l)}\right)$
Elastic	$-\operatorname{div}\left(\beta_1'\left(\sum_{j,m=1}^{n}\left(\partial_{x_j} u_m^{(k,l)} + \partial_{x_m} u_j^{(k,l)}\right)^2\right)\left(\nabla \mathbf{du}^{(k,l)} + \left(\nabla \mathbf{du}^{(k,l)}\right)^{\mathrm{T}}\right)\right) - \nabla\left(\beta_2'\left(\left(\operatorname{div}\mathbf{u}^{(k,l)}\right)^2\right)\operatorname{div}\mathbf{du}^{(k,l)}\right)$
Curvature	$\Delta\left(\beta'\left(\sum_{j=1}^{n}\left(\Delta u_j^{(k,l)}\right)^2\right)\Delta \mathbf{du}^{(k,l)}\right)$
2^{nd}-order Elastic	$\sum_{m=1}^{n}\partial_{x_m}\operatorname{div}\left(\beta_1'\left(\sum_{j,r,s=1}^{n}\left(\partial^2_{x_j,x_r} u_s^{(k,l)} + \partial^2_{x_j,x_s} u_r^{(k,l)} + \partial^2_{x_r,x_s} u_j^{(k,l)}\right)^2\right)\partial_{x_m}\left[\nabla \mathbf{du}^{(k,l)} + 2\left(\nabla \mathbf{du}^{(k,l)}\right)^{\mathrm{T}}\right]\right)$ $+\nabla\operatorname{div}\left(\beta_2'\left(\left\|\nabla\operatorname{div}\mathbf{u}^{(k,l)}\right\|^2\right)\nabla\operatorname{div}\mathbf{du}^{(k,l)}\right)$

Table 5.2: Approximations to linear partial differential terms for flow-driven isotropic regularizers.

The Euler-Lagrange equations (5.2)–(5.3) written in terms of the velocity field become:

$$\mathcal{A}\mathbf{v} = -\,\alpha\,\mathcal{P}(\mathbf{x}; R, F^{\mathbf{u}}) \qquad \forall \mathbf{x} \in \Omega, \tag{5.23}$$

$$\mathcal{B}[\mathbf{v}] = 0 \qquad \forall \mathbf{x} \in \partial\Omega, \tag{5.24}$$

and the parabolic system (5.5) becomes:

$$\partial_t \mathbf{v}(\mathbf{x}, t) + \mathcal{A}\mathbf{v}(\mathbf{x}, t) = \mathbf{f}(\mathbf{x}, t, \mathbf{u}(\mathbf{x}, t))\,, \qquad \mathbf{x} \in \Omega,\ t > 0, \tag{5.25}$$

$$\mathbf{u}(\mathbf{x}, 0) = \mathbf{u}^{(0)},$$

$$\mathbf{v}(\mathbf{x}, 0) = \mathbf{v}^{(0)},$$

$$\mathbf{B}[\mathbf{v}(\mathbf{x}, t)] = 0, \qquad \mathbf{x} \in \partial\Omega,\ t \geq 0.$$

Some extra care must be taken when posing PDE systems with the nonhomogeneous regularizers in terms of the velocity fields. For the nonlinear image-driven regularizers, the weighting functions $\beta^{\mathbf{u}}(\mathbf{x})$ must still deform according to the displacement field, not the velocity field. Therefore, the appropriate way to express the large-deformation version of the partial differential operator for nonlinear image-driven diffusion regularizer, for example, is by:

$$\mathcal{A}^{\textit{id-iso-nonlin}}_{\textit{diffusion}}(\mathbf{v}) := -\mathrm{div}\left(\beta^{\mathbf{u}}(\mathbf{x})\,\nabla\mathbf{v}\right) - \frac{1}{2}\left(\sum_{j=1}^{n} \|\nabla v_j\|^2\right)\nabla\beta^{\mathbf{u}}(\mathbf{x})\,. \tag{5.26}$$

It is *not* meaningful, however, to define the large-deformation version simply by replacing $\mathbf{u}$ with $\mathbf{v}$ everywhere, as in:

$$\mathcal{A}^{\textit{id-iso-nonlin}}_{\textit{diffusion}}(\mathbf{v}) :\neq -\mathrm{div}\left(\beta^{\mathbf{v}}(\mathbf{x})\,\nabla\mathbf{v}\right) - \frac{1}{2}\left(\sum_{j=1}^{n} \|\nabla v_j\|^2\right)\nabla\beta^{\mathbf{v}}(\mathbf{x})\,. \tag{5.27}$$

Likewise, for the flow-driven regularizers, the weighting terms in the partial differential operators should still operate on $\mathbf{u}$. For example, the appropriate definition of the large-deformation version of the partial differential operator corresponding to the flow-driven diffusion regularizer should be:

$$\mathcal{A}^{\textit{fd-iso}}_{\textit{diffusion}}(\mathbf{v}) := -\mathrm{div}\left(\beta'\left(\sum_{j=1}^{n} \|\nabla u_j\|^2\right)\nabla\mathbf{v}\right), \tag{5.28}$$

and *not*:

$$\mathcal{A}^{\textit{fd-iso}}_{\textit{diffusion}}(\mathbf{v}) :\neq -\text{div}\left(\beta'\left(\sum_{j=1}^{n}\|\nabla v_j\|^2\right)\nabla\mathbf{v}\right). \tag{5.29}$$

When the partial differential operators are applied to $\mathbf{v}$ instead of $\mathbf{u}$, the fixed point and steepest descent iteration algorithms change slightly. First, where it is appropriate, instances of $\mathbf{u}^{(k)}$, $\mathbf{u}^{(k,l)}$, or $\mathbf{du}^{(k,l)}$ should be replaced with $\mathbf{v}^{(k)}$, $\mathbf{v}^{(k,l)}$, or $\mathbf{dv}^{(k,l)}$, respectively. Next, an additional step is required after $\mathbf{v}^{(k+1)}$ is computed to determine $\mathbf{u}^{(k+1)}$. This step can be performed by Euler integration of (5.22); i.e.,

$$\mathbf{u}^{(k+1)}(\mathbf{x}) = \mathbf{u}^{(k)}(\mathbf{x}) + \tau\left(\mathbf{I} - \nabla\mathbf{u}^{(k)}(\mathbf{x})\right)^{\mathrm{T}}\mathbf{v}^{(k+1)}(\mathbf{x}). \tag{5.30}$$

Finally, a regridding step should be employed when necessary to ensure a diffeomorphic solution. As described in [31], the deformation field should be regridded if the Jacobian determinant $|\mathbf{I} - \nabla\mathbf{u}|$ falls below a specified threshold anywhere in the image.

These changes are reflected in the updated algorithmic descriptions of the fixed point and steepest descent iterations given in Algorithms 5–6. Similar changes are required to adapt Algorithms 3–4 to the large deformation setting. Note that the comments in sections 5.1.1–5.1.2 about boundary conditions still apply here; the fixed-point iteration algorithm is fine when Dirichlet boundary conditions are chosen, but the steepest descent iteration algorithm should be used when Neumann boundary conditions are chosen.

Algorithm 5 Fixed-Point Iteration for Large Deformation Registration with Linear $\mathcal{A}$

Set $\mathbf{v}^{(0)} = \mathbf{u}^{(0)} = 0$, $k = 0$; Select τ_1, τ_2 such that $0 < \tau_1, \tau_2 << 1$.
repeat
  Compute $\mathbf{f}^{(k)}(\mathbf{x}) = -\alpha^{-1}\,\mathcal{P}\Big(\mathbf{x}; R, F^{\mathbf{u}^{(k)}}\Big)$.
  Solve the linear PDE system $\hat{\mathcal{A}}\mathbf{v}^{(k+1)} = \mathbf{f}^{(k)}$ for $\mathbf{v}^{(k+1)}$.
  Solve for $\mathbf{u}^{(k+1)}$ by Euler integration.
  If $\min\left|\mathbf{I} - \nabla\mathbf{u}^{(k+1)}\right| < \tau_1$, regrid.
  Increment k.
until $d\big(\mathbf{u}^{(k)}, \mathbf{u}^{(k-1)}\big) < \tau_2$.

5.2 Spatial Discretization

At the heart of Algorithms 1–6 lies the step of solving a succession of linear PDE systems. We focus here on approximating the solutions to these PDE systems by employing a spatial

Algorithm 6 Steepest Descent Iteration for Large Deformation Registration with Linear $\mathcal{A}$

Set $\mathbf{v}^{(0)} = \mathbf{u}^{(0)} = 0$, $k = 0$; Select τ_1, τ_2 such that $0 < \tau_1, \tau_2 << 1$.
repeat
  Compute $\mathbf{b}^{(k)}(\mathbf{x}) = \mathbf{v}^{(k)}(\mathbf{x}) - \tau\alpha^{-1}\,\mathcal{P}\Big(\mathbf{x}; R, F^{\mathbf{u}^{(k)}}\Big)$.
  Solve the linear PDE system $\left[\mathbf{I} + \tau\hat{\mathcal{A}}\right]\mathbf{v}^{(k+1)} = \mathbf{b}^{(k)}$ for $\mathbf{v}^{(k+1)}$.
  Solve for $\mathbf{u}^{(k+1)}$ by Euler integration.
  If $\min\left|\mathbf{I} - \nabla\mathbf{u}^{(k+1)}\right| < \tau_1$, regrid.
  Increment k.
until $d\big(\mathbf{u}^{(k)}, \mathbf{u}^{(k-1)}\big) < \tau_2$.

discretization of the images, vector fields, and differential operators, and by solving the resulting linear system of algebraic equations.

We assume that the images and vector fields are sampled on an n-dimensional lattice that contains N_j samples along the j^{th} dimension; i.e.,

$$\begin{aligned} F := [F_{\mathbf{m}}]\,, \quad R := [R_{\mathbf{m}}]\,, \\ \mathbf{u} := [\mathbf{u}_{\mathbf{m}}] = \left[(u_{1;\mathbf{k}}, u_{2;\mathbf{m}}, \ldots, u_{n;\mathbf{m}})^{\mathrm{T}}\right], \\ \mathbf{m} = (m_1, \ldots, m_n)\,, \quad m_j = 0, \ldots, N_j - 1, \end{aligned} \tag{5.31}$$

and where the images and vector fields can be defined at non-lattice points by some interpolation scheme (such as nearest-neighbor or multilinear interpolation). The spacing of lattice points can differ in each dimension; hence, we define the vector $\mathbf{h} = (h_1, \ldots, h_n)$ so that h_j is the lattice spacing in the j^{th} dimension.

To discretize the various partial differential operators in space, we use centered differences. For any interior lattice point, the first partial derivative of an image F (or of a component of a vector field) can be discretized according to:

$$\left[\frac{\partial}{\partial x_j}F\right]_{\mathbf{m}} \approx \frac{1}{2h_j}\left(F_{\mathbf{m}+\mathbf{e}_j} - F_{\mathbf{m}-\mathbf{e}_j}\right), \tag{5.32}$$

where $\mathbf{e}_j$ is the j^{th} column of the $n \times n$ identity matrix. For lattice points on the boundary, the discretization must be modified to account for the chosen boundary conditions. This discretization is referred to in [6] as *long centered* because the difference stencil is not of minimal size for a first-order differential operator. The corresponding *short centered*

discretization requires a dual lattice containing the midpoints of the original lattice, and is defined by:

$$\left[\frac{\partial}{\partial x_j}F\right]_{\mathbf{m}} \approx \frac{1}{h_j}\left(F_{\mathbf{m}+\frac{1}{2}\mathbf{e}_j} - F_{\mathbf{m}-\frac{1}{2}\mathbf{e}_j}\right). \tag{5.33}$$

Second partial derivatives can be discretized by recursively applying (5.32) or (5.33), yielding:

$$\left[\frac{\partial^2}{\partial x_i \partial x_j}F\right]_{\mathbf{m}} \approx \frac{1}{4h_ih_j}\left(F_{\mathbf{m}+\mathbf{e}_i+\mathbf{e}_j} - F_{\mathbf{m}+\mathbf{e}_i-\mathbf{e}_j} - F_{\mathbf{m}-\mathbf{e}_i+\mathbf{e}_j} + F_{\mathbf{m}-\mathbf{e}_i-\mathbf{e}_j}\right) \tag{5.34}$$

as a long centered discretization, or

$$\left[\frac{\partial^2}{\partial x_i \partial x_j}F\right]_{\mathbf{m}} \approx \frac{1}{h_ih_j}\left(F_{\mathbf{m}+\frac{1}{2}\mathbf{e}_i+\frac{1}{2}\mathbf{e}_j} - F_{\mathbf{m}+\frac{1}{2}\mathbf{e}_i-\frac{1}{2}\mathbf{e}_j} - F_{\mathbf{m}-\frac{1}{2}\mathbf{e}_i+\frac{1}{2}\mathbf{e}_j} + F_{\mathbf{m}-\frac{1}{2}\mathbf{e}_i-\frac{1}{2}\mathbf{e}_j}\right) \tag{5.35}$$

as a short centered discretization. Note that when $i = j$, (5.35) reduces to the compact discretization:

$$\left[\frac{\partial^2}{\partial x_j^2}F\right]_{\mathbf{m}} \approx \frac{1}{h_j^2}\left(F_{\mathbf{m}+\mathbf{e}_j} - 2F_{\mathbf{m}} + F_{\mathbf{m}-\mathbf{e}_j}\right). \tag{5.36}$$

Discretized versions of the gradient vector ∇, the divergence operator $\nabla\cdot$, and the Laplacian operator Δ can be constructed using (5.32)–(5.36), and will be given the notation $\check{\nabla}$, $\check{\nabla}\cdot$, and $\check{\Delta}$, respectively. Discretized versions of higher-order operators can be constructed by composing the discrete gradient, divergence, and Laplacian operators.

For two examples of how these discretized operators are used to transform linear PDE systems into linear systems of algebraic equations, we consider the linear PDE system $\left[\mathbf{I} + \tau\hat{\mathcal{A}}\right]\mathbf{v}^{(k+1)} = \mathbf{b}^{(k)}$ in Algorithm 6, first with the homogeneous diffusion regularizer and then with the image-driven isotropic diffusion regularizer. For the homogeneous diffusion regularizer $\mathcal{A} = -\Delta$ with homogeneous Dirichlet boundary conditions, replacing Δ with $\check{\Delta}$ (under the short centered discretization) and enforcing the boundary conditions yields the linear system of algebraic equations given by:

$$\begin{aligned}\mathbf{v}_{\mathbf{m}}^{(k+1)} - \tau\sum_{j=1}^{n}\frac{1}{h_j^2}\left(\mathbf{v}_{\mathbf{m}+\mathbf{e}_j}^{(k+1)} - 2\mathbf{v}_{\mathbf{m}}^{(k+1)} + \mathbf{v}_{\mathbf{m}-\mathbf{e}_j}^{(k+1)}\right) &= \mathbf{b}_{\mathbf{m}}^{(k)}, \qquad \mathbf{m} \in \text{interior},\\ \mathbf{v}_{\mathbf{m}}^{(k+1)} &= \mathbf{0}, \qquad \mathbf{m} \in \text{boundary}.\end{aligned} \tag{5.37}$$

For the image-driven isotropic diffusion regularizer $\mathcal{A} = -\text{div}\left(\beta(\mathbf{x})\,\nabla\mathbf{u}\right)$, the linear system of algebraic equations is given by:

$$\mathbf{v}_{\mathbf{m}}^{(k+1)} - \tau \sum_{j=1}^{n} \frac{1}{h_j^2} \left[\left(\mathbf{v}_{\mathbf{m}+\mathbf{e}_j}^{(k+1)} - \mathbf{v}_{\mathbf{m}}^{(k+1)} \right) \beta_{\mathbf{m}+\frac{1}{2}\mathbf{e}_j} - \left(\mathbf{v}_{\mathbf{m}}^{(k+1)} - \mathbf{v}_{\mathbf{m}-\mathbf{e}_j}^{(k+1)} \right) \beta_{\mathbf{m}-\frac{1}{2}\mathbf{e}_j} \right] = \mathbf{b}_{\mathbf{m}}^{(k)} \quad (5.38)$$

for $\mathbf{m}$ on the lattice interior, and $\mathbf{v}_{\mathbf{m}}^{(k+1)} = \mathbf{0}$ on the lattice boundary.

5.3 Iterative Solvers for Linear Systems of Algebraic Equations

Once a spatial discretization has been employed, the solutions of the linear PDE systems in Algorithms 1–6 can now be approximated by the solutions of the corresponding linear systems of algebraic equations. A wide variety of algorithms have been developed in the field of numerical linear algebra for approximately solving linear systems; among them, simple relaxation methods such as Jacobi iteration, Gauss-Seidel iteration, and successive overrelaxation, and more complicated multigrid techniques. References describing these techniques include [6, 70, 88].

The simple relaxation methods are easy to implement and can be applied for any choice of regularizer presented in this thesis. However, simple relaxation methods can require many iterations to converge. Even though the high frequency content of the error vectors can be reduced in just a few iterations, it takes many more iterations to reduce the low frequency content in the error vectors. Multigrid techniques solve this problem by traversing a series of grids of varying resolutions, performing only a few iterations at each resolution. However, implementation of multigrid algorithms can be much more difficult than simple relaxation methods.

For some specific combinations of regularizers and boundary conditions, specialized algorithms have been developed that are more efficient and/or easier to implement than the standard iterative solvers. Christensen [28] illustrates how the eigenfunctions of the continuous versions of the Laplacian and Navier-Lamé operators could be used to derive solutions subject to bending and sliding boundary conditions; Grenander and Miller [46] and Christensen and Johnson [29] show how any homogeneous linear differential operators subject

to periodic boundary conditions can be resolved via Fourier series solution. Modersitzki [63] describes rapid algorithms based on additive operator splitting (AOS) and Gaussian filtering for the homogeneous diffusion regularizer, and the discrete cosine transform (DCT) for homogeneous diffusion and homogeneous curvature regularizers with Neumann boundary conditions. Bro-Nielsen [13] describes a filtering algorithm for fluid registration that implicitly assumes no boundary conditions (i.e., images are defined over infinite space). Stefanescu *et al.* [82] illustrate a rapid AOS scheme for use with an image-driven anisotropic diffusion regularizer with implicit boundary conditions.

In the remaining sections of this chapter, we illustrate how specialized algorithms using Fourier methods and successive Gaussian convolution can be constructed for variational registration using *any* of the homogeneous regularizers under Dirichlet, Neumann, and periodic boundary conditions, as well as for *nonhomogeneous* versions of the curvature regularizer.

5.4 Fourier Methods for Homogeneous Regularizers

In [46], Grenanader and Miller illustrate how the DFT can be used to rapidly solve each iteration of registration with a homogeneous diffusion or elastic regularizer subject to periodic boundary conditions. Modersitzki [63] illustrates how to use the DCT for homogeneous diffusion-based or curvature-based registration when Neumann boundary conditions are assumed. In this section, we generalize these ideas to show that Fourier-based methods can be used to construct rapid solution algorithms for registration using any of the homogeneous regularizers in the extended quadratic taxonomy (diffusion, elastic, curvature and second-order elastic) with Dirichlet, Neumann, or periodic boundary conditions.

5.4.1 Diffusion Registration

5.4.1.1 Dirichlet Boundary Conditions

Let us consider the linear PDE system in Algorithm 1, with $\hat{\mathcal{A}}$ chosen to be the negative Laplacian operator $-\Delta$ with homogeneous Dirichlet boundary conditions. As is shown in [21], if the boundary conditions are eliminated by setting $\mathbf{u}_{\mathbf{j}}^{(k+1)} = \mathbf{0}$ for all points on

the lattice boundary, then the discretized version of $\hat{\mathcal{A}}_{diffusion}$ has a set of eigenfunctions $\left\{\Theta^\ell | \ell \in \text{lattice interior}\right\}$, where Θ^ℓ is given at each interior lattice position $\mathbf{j}$ by:

$$\Theta^\ell_{\mathbf{j}} = \prod_{r=1}^{n} \sin \frac{\ell_r j_r \pi}{N_r - 1}, \tag{5.39}$$

and the corresponding eigenvalues are given by:

$$\omega^D_\ell = \sum_{r=1}^{n} \left(2 - 2\cos \frac{\ell_r \pi}{N_r - 1}\right). \tag{5.40}$$

Hence, if we expand $\mathbf{u}^{(k+1)}$ and $\mathbf{f}^{(k)}$ in terms of these eigenfunctions; i.e.,

$$\mathbf{u}^{(k+1)}_{\mathbf{j}} = \sum_\ell \hat{\mathbf{u}}^{(k+1)}_\ell \Theta^\ell_{\mathbf{j}}, \qquad \mathbf{f}^{(k)}_{\mathbf{j}} = \sum_\ell \hat{\mathbf{f}}^{(k)}_\ell \Theta^\ell_{\mathbf{j}}, \qquad \forall \mathbf{j} \in \text{lattice}, \tag{5.41}$$

then the system $-\check{\Delta}\mathbf{u}^{(k+1)} = \mathbf{f}^{(k)}$ can be expressed at each interior lattice point by:

$$\sum_\ell \hat{\mathbf{u}}^{(k+1)}_\ell \omega^D_\ell \Theta^\ell_{\mathbf{j}} = \sum_\ell \hat{\mathbf{f}}^{(k)}_\ell \Theta^\ell_{\mathbf{j}}. \tag{5.42}$$

From (5.42), we see that the coefficients in the expansions of $\mathbf{u}^{(k+1)}$ and $\mathbf{f}^{(k)}$ are related by:

$$\hat{\mathbf{u}}^{(k+1)}_\ell = \hat{\mathbf{f}}^{(k)}_\ell / \omega^D_\ell \qquad \forall \ell \in \text{lattice interior}. \tag{5.43}$$

If an efficient way exists to compute $\hat{\mathbf{f}}$ from $\mathbf{f}$ and $\mathbf{u}$ from $\hat{\mathbf{u}}$, then it is possible to solve $-\check{\Delta}\mathbf{u}^{(k+1)} = \mathbf{f}^{(k)}$ rapidly in three steps: (i) compute $\hat{\mathbf{f}}$ from $\mathbf{f}$; (ii) solve for $\hat{\mathbf{u}}$ using (5.43); and, (iii) compute $\mathbf{u}$ from $\hat{\mathbf{u}}$. With arbitrary eigenfunctions, there is not necessarily any rapid expansion method. However, the structure of the eigenfunctions given by (5.39) is such that the expansions in (5.41) (as well as their inverses) are versions of the discrete sine transform (DST), which can be computed rapidly via the fast Fourier transform (FFT) algorithm [70].

To extend this idea to work in the context of the steepest descent approach in Algorithm 2, we note that $\mathbf{I} - \tau\check{\Delta}$ has the same eigenfunctions as $-\check{\Delta}$, namely (5.39), but with eigenvalues given by $1 + \tau\omega^D_\ell$. Hence, on the lattice interior, we have the following equivalence:

$$\left[\mathbf{I} - \tau\check{\Delta}\right]\mathbf{u} = \mathbf{f} \qquad \Leftrightarrow \qquad \hat{\mathbf{u}}_\ell = \hat{\mathbf{f}}_\ell / \left(1 + \tau\omega^D_\ell\right) \quad \forall \ell, \tag{5.44}$$

and therefore it is possible to solve this linear system using the DST.

This approach can also be used in the large deformation setting to solve the linear systems arising in Algorithms 5 and 6. The only difference is that the velocity field $\mathbf{v}$, instead of the displacement field $\mathbf{u}$, must be expanded in terms of the eigenfunctions.

5.4.1.2 Neumann and Periodic Boundary Conditions

If the negative Laplacian operator is chosen with boundary conditions other than homogeneous Dirichlet, the discretized operator $\check{\mathcal{A}}_{diffusion}$ no longer has eigenfunctions given by (5.39). This means that the DST cannot be exploited to rapidly solve the linear systems $-\check{\Delta}\mathbf{u} = \mathbf{f}$, $-\check{\Delta}\mathbf{v} = \mathbf{f}$, $\left[\mathbf{I} - \tau\check{\Delta}\right]\mathbf{u} = \mathbf{f}$ or $\left[\mathbf{I} - \tau\check{\Delta}\right]\mathbf{v} = \mathbf{f}$.

However, other specific choices of boundary conditions do yield eigenfunctions that allow for rapid solution algorithms. If $-\check{\Delta}$ is chosen to have homogeneous Neumann boundary conditions, for example, the eigenfunctions of $\check{\mathcal{A}}_{diffusion}$ are given by:

$$\Theta_{\mathbf{j}}^{\ell} = \prod_{r=1}^{n} \cos \frac{\ell_r j_r \pi}{N_r - 1}, \tag{5.45}$$

and the corresponding eigenvalues are given by:

$$\omega_{\ell}^{N} = \omega_{\ell}^{D} = \sum_{r=1}^{n} \left(2 - 2\cos \frac{\ell_r \pi}{N_r - 1}\right). \tag{5.46}$$

If periodic boundary conditions are chosen with $-\check{\Delta}$, the eigenfunctions of $\check{\mathcal{A}}_{diffusion}$ are given by:

$$\Theta_{\mathbf{j}}^{\ell} = \prod_{r=1}^{n} \exp \frac{2i\ell_r j_r \pi}{N_r}, \tag{5.47}$$

and the corresponding eigenvalues are given by:

$$\omega_{\ell}^{P} = \sum_{r=1}^{n} \left(2 - 2\cos \frac{2\ell_r \pi}{N_r}\right). \tag{5.48}$$

Proofs of (5.45)–(5.48) are similar to the proof of (5.39)–(5.40) in [21], and are therefore omitted.

Expansions in terms of (5.45) and (5.47) can be computed via the discrete cosine transform (DCT) and discrete Fourier transform (DFT), respectively. Like the DST, the DCT and DFT can be computed rapidly using the FFT algorithm [70]. Therefore, the linear systems $\left[\mathbf{I} - \tau\check{\Delta}\right]\mathbf{u} = \mathbf{f}$ and $\left[\mathbf{I} - \tau\check{\Delta}\right]\mathbf{v} = \mathbf{f}$ can be solved rapidly via the DCT if Neumann

boundary conditions are chosen, or via the DFT if periodic boundary conditions are chosen. (Solving the linear systems $-\check{\Delta}\mathbf{u} = \mathbf{f}$ or $-\check{\Delta}\mathbf{v} = \mathbf{f}$ can be problematic, however. For Neumann or periodic boundary conditions, lattice points on the boundary can no longer be eliminated, and $-\check{\Delta}$ has a trivial eigenvalue corresponding to the lattice point $\ell = \mathbf{0}$. This makes the linear system singular, as mentioned previously in section 5.1.1. For this reason, the steepest descent algorithm is recommended over the fixed point iteration when Neumann or periodic boundary conditions are chosen.)

5.4.2 Curvature Registration

Since the partial differential operator $\mathcal{A}_{curvature} = \Delta^2 = (-\Delta)^2$ is simply the composition of $\mathcal{A}_{diffusion}$ with itself, any eigenfunction of $-\check{\Delta}$ is also an eigenfunction of $\check{\Delta}^2$. The corresponding eigenvalues are simply the squares of the eigenvalues of $-\check{\Delta}$. Hence, for the fixed point iteration with the curvature regularizer, we can establish:

$$\check{\Delta}^2\mathbf{u} = \mathbf{f} \qquad \Leftrightarrow \qquad \hat{\mathbf{u}}_\ell = \hat{\mathbf{f}}_\ell / \left(\omega_\ell\right)^2 \quad \forall \ell, \tag{5.49}$$

where ω_ℓ is defined as in (5.40), (5.46), or (5.48), depending on the chosen boundary conditions. For the steepest descent iteration, we have:

$$\left[\mathbf{I} + \tau\check{\Delta}^2\right]\mathbf{u} = \mathbf{f} \qquad \Leftrightarrow \qquad \hat{\mathbf{u}}_\ell = \hat{\mathbf{f}}_\ell / \left(1 + \tau\left(\omega_\ell\right)^2\right) \quad \forall \ell. \tag{5.50}$$

These results extend also to the large deformation case by replacing $\mathbf{u}$ with $\mathbf{v}$.

Therefore, like diffusion registration, curvature registration can be performed rapidly via DST, DCT, or DFT, depending on the selected boundary conditions. As with diffusion registration, when Neumann or periodic boundary conditions are chosen, the steepest descent algorithm is recommended in order to avoid a singular linear system due to a trivial eigenvalue when $\ell = \mathbf{0}$.

5.4.3 Elastic Registration

The elastic regularizer, when subject to Dirichlet or Neumann boundary conditions, no longer has eigenfunctions of the form (5.39) or (5.45). Hence, elastic or fluid registration cannot be performed directly using the DST or DCT. However, the linear system

$\check{\mathcal{A}}_{elastic}\mathbf{u} = \mathbf{f}$ can be transformed into a linear system of the form $\check{\Delta}^2 = \mathbf{g}$ (as long as $\check{\mathcal{A}}_{elastic}$ is nonsingular), which *can* be solved rapidly as described in the previous subsection on curvature registration.

We define the linear operator $\mathcal{L}_1$ by:

$$\mathcal{L}_1 = -(\lambda + 2\mu)\,\Delta\mathbf{I} + (\lambda + \mu)\,\nabla\mathrm{div}. \tag{5.51}$$

In Appendix E, we show that $\mathcal{L}_1 \circ \mathcal{A}_{elastic}$ is equivalent to a scaled version of $\mathcal{A}_{curvature}$. This establishes an equivalence between the linear systems:

$$-\mu\check{\Delta}\mathbf{u} - (\lambda + \mu)\,\check{\nabla}\check{\mathrm{div}}\,\mathbf{u} = \mathbf{f} \qquad \Leftrightarrow \qquad \mu\,(\lambda + 2\mu)\,\check{\Delta}^2 = \check{\mathcal{L}}_1\mathbf{f}. \tag{5.52}$$

This equivalence, in conjunction with (5.49), allows us to state the following:

$$-\mu\check{\Delta}\mathbf{u} - (\lambda + \mu)\,\check{\nabla}\check{\mathrm{div}}\,\mathbf{u} = \mathbf{f} \qquad \Leftrightarrow \qquad \hat{\mathbf{u}}_\ell = \hat{\mathbf{g}}_\ell / (\omega_\ell)^2 \quad \forall \ell, \tag{5.53}$$

where

$$\mathbf{g} = \frac{\mathcal{L}_1\mathbf{f}}{\mu\,(\lambda + 2\mu)}, \tag{5.54}$$

and where ω_ℓ is defined as in (5.40), (5.46), or (5.48), depending on the chosen boundary conditions. Hence, when the elastic and fluid regularizers are used within a fixed point iterative scheme for registration, the solutions to the linear PDE systems in Algorithms 1 and 5 can be found by transforming them according to (5.52), discretizing and then employing forward and inverse versions of the DST, DCT, or DFT.

For Algorithms 2 and 6 that employ the steepest descent iteration (which are necessary when $\check{\mathcal{A}}_{elastic}$ is singular), a similar trick can be used to transform the linear system $\left[\mathbf{I} + \tau\check{\mathcal{A}}_{elastic}\right]\mathbf{u} = \mathbf{f}$ into one that can be rapidly solved via DST, DCT, or DFT. If we apply the linear operator:

$$\mathcal{L}_2 = \mathbf{I} + \tau\mathcal{L}_1 \tag{5.55}$$

to both sides of the equation $[\mathbf{I} + \tau\mathcal{A}_{elastic}]\,\mathbf{u} = \mathbf{f}$, we find that the resulting linear PDE system is given by:

$$\left[\mathbf{I} - \tau\,(\lambda + 3\mu)\,\Delta + \tau^2\mu\,(\lambda + 2\mu)\,\Delta^2\right]\mathbf{u} = \mathcal{L}_2\mathbf{f}. \tag{5.56}$$

A proof of (5.56) is provided in Appendix E. Now, when the linear operator on the left hand side of (5.56) is discretized, it is a linear combination of the identity, $-\check{\Delta}$, and $\check{\Delta}^2$, and therefore has the same eigenfunctions as $-\check{\Delta}$ and $\check{\Delta}^2$. The eigenvalues arise from the same linear combination of the eigenvalues of $\mathbf{I}$, $-\check{\Delta}$ and $\check{\Delta}^2$. Therefore, we can state the equivalence:

$$\left[\mathbf{I} - \tau\left(\mu\check{\Delta} + (\lambda+\mu)\,\check{\nabla}\check{\mathrm{div}}\right)\right]\mathbf{u} = \mathbf{f} \qquad \Leftrightarrow$$

$$\hat{\mathbf{u}}_\ell = \frac{\hat{\mathbf{q}}_\ell}{1 + \tau\,(\lambda+3\mu)\,\omega_\ell + \tau^2\mu\,(\lambda+2\mu)\,(\omega_\ell)^2} \quad \forall \ell, \tag{5.57}$$

where $\mathbf{q} = \mathcal{L}_2\mathbf{f}$, and where ω_ℓ is defined as in (5.40), (5.46), or (5.48), depending on the chosen boundary conditions. Hence, for elastic and fluid registration, the solutions to the linear PDE systems in Algorithms 2 and 6 can be found by transforming them according to (5.56), discretizing, and then employing forward and inverse versions of the DST, DCT, or DFT.

5.4.4 Second-order Elastic Registration

As with elastic registration, second-order elastic registration requires transformations of the linear PDE systems in Algorithms 1, 2, 5 and 6 in order to exploit the DST, DCT, and DFT for rapid solutions.

We define the linear operator $\mathcal{L}_3$ by:

$$\mathcal{L}_3 = -\,(\lambda+3\mu)\,\Delta\mathbf{I} + (\lambda+2\mu)\,\nabla\mathrm{div}. \tag{5.58}$$

In Appendix E, we show that $\mathcal{L}_3 \circ \mathcal{A}_{\textit{elastic-2}}$ is equivalent to a scaled version of $(-\Delta)^3$. This establishes an equivalence between the linear systems:

$$\mu\check{\Delta}^2\mathbf{u} + (\lambda+2\mu)\,\check{\Delta}\check{\nabla}\check{\mathrm{div}}\,\mathbf{u} = \mathbf{f} \qquad \Leftrightarrow \qquad \mu\,(\lambda+3\mu)\left(-\check{\Delta}\right)^3 = \check{\mathcal{L}}_3\mathbf{f}. \tag{5.59}$$

Since $\left(-\check{\Delta}\right)^3$ has the same eigenfunctions as $-\check{\Delta}$, and since the corresponding eigenvalues are the cubes of the eigenvalues of $-\check{\Delta}$, we can state the following:

$$\mu\check{\Delta}^2\mathbf{u} + (\lambda+2\mu)\,\check{\Delta}\check{\nabla}\check{\mathrm{div}}\,\mathbf{u} \qquad \Leftrightarrow \qquad \hat{\mathbf{u}}_\ell = \hat{\mathbf{p}}_\ell/\,(\omega_\ell)^3 \quad \forall \ell, \tag{5.60}$$

where

$$\mathbf{p} = \frac{\mathcal{L}_3 \mathbf{f}}{\mu\left(\lambda + 3\mu\right)}, \tag{5.61}$$

and where ω_ℓ is defined as in (5.40), (5.46), or (5.48), depending on the chosen boundary conditions. Hence, when the second-order elastic regularizer is used within a fixed point iterative scheme for registration, the solutions to the linear PDE systems in Algorithms 1 and 5 can be found by discretizing, transforming them according to (5.59) and then employing forward and inverse versions of the DST, DCT, or DFT.

To transform the linear PDE system $\left[\mathbf{I} + \tau \mathcal{A}_{\textit{elastic-2}}\right] \mathbf{u} = \mathbf{f}$ into one that can be discretized and rapidly solved via DST, DCT, or DFT, we can apply the linear operator:

$$\mathcal{L}_4 = \mathbf{I} - \tau \Delta \circ \mathcal{L}_3 \tag{5.62}$$

to both sides of the equation. As shown in Appendix E, the resulting linear PDE system is given by:

$$\left[\mathbf{I} + \tau\left(\lambda + 4\mu\right)\Delta^2 + \tau^2 \mu\left(\lambda + 3\mu\right)\Delta^4\right] \mathbf{u} = \mathcal{L}_4 \mathbf{f}. \tag{5.63}$$

Now, when the linear operator on the left hand side of (5.63) is discretized, it is a linear combination of the identity, $\check{\Delta}^2$, and $\check{\Delta}^4$, and therefore has the same eigenfunctions as $\check{\Delta}^2$ and $\check{\Delta}^4$. The eigenvalues arise from the same linear combination of the eigenvalues of $\mathbf{I}$, $\check{\Delta}^2$ and $\check{\Delta}^4$. The eigenvalues of $\check{\Delta}^4$ are simply the squares of the eigenvalues of $\check{\Delta}^2$. Therefore, we can state the equivalence:

$$\begin{aligned} \left[\mathbf{I} - \tau\left(\mu\check{\Delta}^2 + (\lambda + 2\mu)\check{\Delta}\check{\nabla}\check{\mathrm{div}}\right)\right] \mathbf{u} = \mathbf{f} \qquad &\Leftrightarrow \\ \hat{\mathbf{u}}_\ell = \frac{\hat{\mathbf{r}}_\ell}{1 + \tau\left(\lambda + 4\mu\right)\left(\omega_\ell\right)^2 + \tau^2 \mu\left(\lambda + 3\mu\right)\left(\omega_\ell\right)^4} \quad &\forall \ell, \end{aligned} \tag{5.64}$$

where $\mathbf{r} = \mathcal{L}_4 \mathbf{f}$, and where ω_ℓ is defined as in (5.40), (5.46), or (5.48), depending on the chosen boundary conditions. Hence, for second-order elastic registration, the solutions to the linear PDE systems in Algorithms 2 and 6 can be found by transforming them according to (5.63), discretizing, and then employing forward and inverse versions of the DST, DCT, or DFT.

5.5 Successive Gaussian Convolution for Homogeneous Regularizers

In addition to presenting the Demon's force vector described in Chapter 2, Thirion [87] also provided a fast and simple to implement algorithm for registration that can be applied with any force vector. The algorithm involves iteratively computing the force vector and convolving it with a Gaussian kernel to form the next iterate of the displacement/velocity field. Modersitzki [63] illustrates that the Gaussian kernel is the Green's function for the linear diffusion equation over infinite space; hence, he argues that Thirion's successive Gaussian convolution algorithm provides an approximate solution to the diffusion registration problem.

A natural question therefore arises: can a similar algorithm with linear complexity and simple implementation be designed for use with other regularizers? Vercauteren [91] hints at this possibility by noting that the advanced vectorial filters of Cachier and Ayache [18] could be employed for this purpose. This is true; however, the advanced vectorial filters are applied in the Fourier domain, and therefore, when they are discretized and appropriate boundary conditions are chosen, the resulting registration algorithm would bear more resemblance to the Fourier approaches mentioned in Section 5.4.

It is possible, though, to devise algorithms that utilize *Gaussian* convolution for more regularizers than the homogeneous diffusion regularizer. In fact, *all* of the homogeneous regularizers admit such algorithms. This is done by constructing coupled systems of diffusion equations whose stationary solution is identical to the steepest descent solution for the homogeneous regularizer at hand.

To describe this process, we first revisit Modersitzki's notion [63] that Thirion's algorithm is related to diffusion registration. By rigorously examining the inhomogeneous diffusion equation, we can see how Gaussian convolution can be used to approximate the solution to diffusion registration.

5.5.1 Inhomogeneous Diffusion Equation

Consider the following inhomogeneous diffusion equation:

$$\begin{aligned} \partial_t \mathbf{u}(\mathbf{x},t) - \Delta \mathbf{u}(\mathbf{x},t) &= \mathbf{g}(\mathbf{x},t) \\ \mathbf{u}(\mathbf{x},0) &= 0 \ , \end{aligned} \tag{5.65}$$

where $\mathbf{x} \in \mathbb{R}^n$ and $t \geq 0$. The solution to (5.65) is given by:

$$\mathbf{u}(\mathbf{x},t) = \int_0^t \int_{\mathbb{R}^n} \Psi(\mathbf{x}-\mathbf{y}, t-s)\, \mathbf{g}(\mathbf{y},t)\, d\mathbf{y} ds \ , \tag{5.66}$$

where the Green's function $\Psi(\mathbf{x}-\mathbf{y},t)$ is given by the Gaussian kernel $\mathcal{K}\big(\mathbf{x}-\mathbf{y}, \sqrt{2t}\big)$, where

$$\mathcal{K}(\mathbf{w},\sigma) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} \exp\left(\frac{-\mathbf{w}^{\mathrm{T}}\mathbf{w}}{2\sigma^2}\right) \ . \tag{5.67}$$

If we consider a sequence of time points $t_j = j\tau$, $j = 0, \ldots, m+1$, with $\tau = t/(m+1)$, then (5.66) can be approximated by left Riemann sums:

$$\mathbf{u}(\mathbf{x}, t_{m+1}) \approx \sum_{j=0}^{m} \int_{\mathbb{R}^n} \Psi(\mathbf{x}-\mathbf{y}, (m+1-j)\,\tau)\, \mathbf{g}(\mathbf{y}, t_j)\, d\mathbf{y} \ . \tag{5.68}$$

This allows us to establish the following recursive relationship:

$$\mathbf{u}(\mathbf{x}, t_{m+1}) \approx \mathcal{K}\Big(\mathbf{x}, \sqrt{2\tau}\Big) * \Big[\mathbf{g}(\mathbf{x}, t_m) + \mathbf{u}(\mathbf{x}, t_m)\Big] \ , \tag{5.69}$$

where $*$ denotes convolution.

5.5.2 Diffusion Registration

The first variant of Thirion's algorithm [87] essentially poses the diffusion registration problem in terms of the inhomogeneous diffusion equation (5.65) and then follows the recursion (5.69) until the stationary solution is found. This procedure is summarized by Algorithm 7. Note that Thirion [87] uses the Demons force vector; however, we choose here to generalize the algorithm for use with any dissimilarity measure. An interesting observation about this interpretation of Thirion's algorithm is that it provides a value for the width of the Gaussian kernel used in step 4, basing it on the time step τ. Thirion's original paper did not provide such a result.

Algorithm 7 Gaussian Convolution Algorithm for Diffusion Registration

Select time step τ; define $t_j = j\tau$, $j = 0, 1, \ldots$.
Set $\mathbf{u}(\mathbf{x}, t_0) = 0$, $m = 0$.
repeat
  $\mathbf{u}(\mathbf{x}, t_{m+1}) \leftarrow \mathcal{K}\big(\mathbf{x}, \sqrt{2\tau}\big) * \big[\alpha \mathbf{f}(\mathbf{x}, R, F, \mathbf{u}(\mathbf{x}, t_m)) + \mathbf{u}(\mathbf{x}, t_m)\big]$.
  $m \leftarrow m + 1$.
until convergence.

Algorithm 8 Gaussian Convolution Algorithm for Large Deformation Diffusion Registration

Select time step τ; define $t_j = j\tau$, $j = 0, 1, \ldots$.
Set $\mathbf{v}(\mathbf{x}, t_0) = \mathbf{u}(\mathbf{x}, t_0) = 0$, $m = 0$.
repeat
  $\mathbf{v}(\mathbf{x}, t_{m+1}) \leftarrow \mathcal{K}\big(\mathbf{x}, \sqrt{2\tau}\big) * \big[\alpha \mathbf{f}(\mathbf{x}, R, F, \mathbf{u}(\mathbf{x}, t_m)) + \mathbf{v}(\mathbf{x}, t_m)\big]$.
  $\mathbf{u}(\mathbf{x}, t_{m+1}) \leftarrow \mathbf{u}(\mathbf{x}, t_m) + \tau \left[\mathbf{I} - \nabla \mathbf{u}(\mathbf{x}, t_m)\right]^{\mathrm{T}} \mathbf{v}(\mathbf{x}, t_{m+1})$.
  $m \leftarrow m + 1$.
until convergence.

Large deformations can also be handled in the context of the Thirion's Algorithm. If we note that the development of (5.65)–(5.69) is valid if we replace $\mathbf{u}(\mathbf{x}, t)$ with $\mathbf{v}(\mathbf{x}, t)$, then we can define a successive Gaussian convolution to operate on the velocity field, as long as we add an Euler integration step to compute the displacement field from the velocity field at each iteration. This procedure is summarized by Algorithm 8.

Two final notes: first, it is important to ensure that the displacement field $\mathbf{u}(\mathbf{x}, t)$ remains diffeomorphic as t increases. This can be done by employing the regridding strategy of [31]. Second, the Gaussian kernel must be discretized and truncated, and images should be padded or mirrored to mimic an infinite domain.

5.5.3 Generalizing Successive Gaussian Convolution for Other Homogeneous Regularizers

The structure of the elastic, curvature, and second-order elastic operators allows the Euler-Lagrange equations to be represented by coupled diffusion equations. For the curvature operator, the equations are described in the following Lemma:

Lemma 1. *The stationary solution of the coupled diffusion equations:*

$$\partial_t \mathbf{w}(\mathbf{x},t) - \Delta \mathbf{w}(\mathbf{x},t) = \sqrt{\alpha}\,\mathbf{f}(\mathbf{x},R,F,\mathbf{u}(\mathbf{x},t_m)) \tag{5.70}$$

$$\partial_t \mathbf{u}(\mathbf{x},t) - \Delta \mathbf{u}(\mathbf{x},t) = \sqrt{\alpha}\,\mathbf{w}(\mathbf{x},t) \tag{5.71}$$

$$\mathbf{w}(\mathbf{x},0) = \mathbf{u}(\mathbf{x},0) = 0\ , \tag{5.72}$$

is equivalent to the solution of the biharmonic equation $\Delta^2 \mathbf{u}(\mathbf{x}) = \alpha\,\mathbf{f}(\mathbf{x},R,F,\mathbf{u}(\mathbf{x},t_m))$.

Proof. At the stationary solution of the coupled diffusion equations, $\partial_t \mathbf{w}(\mathbf{x},t) = \partial_t \mathbf{u}(\mathbf{x},t) = 0$. Therefore, we have $-\Delta \mathbf{w}(\mathbf{x},t) = \sqrt{\alpha}\,\mathbf{f}(\mathbf{x},R,F,\mathbf{u}(\mathbf{x},t_m))$ and $-\Delta \mathbf{u}(\mathbf{x},t) = \sqrt{\alpha}\,\mathbf{w}(\mathbf{x},t)$. Solving the second equation for $\mathbf{w}(\mathbf{x},t)$ and substituting into the first equation yields the desired result. □

5.5.3.1 Curvature Registration

Using Lemma 1, we can design an algorithm to approximate the solutions of the coupled diffusion equations by successive convolution with Gaussian kernels. With a semi-implicit time discretization scheme, we evaluate the left hand sides of (5.70) and (5.71) at time t_{m+1} while computing the right hand sides with values taken at time t_m. The vector fields $\mathbf{u}$ and $\mathbf{w}$ can either be computed simultaneously or successively. We choose to compute them successively, so $\mathbf{u}$ is formed using the newest possible values of $\mathbf{w}$. With these conventions, we define the Gaussian convolution algorithm for curvature registration by Algorithm 9. A large deformation version of this algorithm can also be constructed for use with the curvature regularizer, by operating on velocities and adding an Euler integration step, as illustrated in Algorithm 10.

Algorithm 9 Gaussian Convolution Algorithm for Curvature Registration

Select time step τ; define $t_j = j\tau$, $j = 0, 1, \ldots$.
Set $\mathbf{w}(\mathbf{x},t_0) = \mathbf{u}(\mathbf{x},t_0) = 0$, $m = 0$.
repeat
  $\mathbf{w}(\mathbf{x},t_{m+1}) \leftarrow \mathcal{K}\big(\mathbf{x},\sqrt{2\tau}\big) * \big[\sqrt{\alpha}\ \mathbf{f}(\mathbf{x},R,F,\mathbf{u}(\mathbf{x},t_m)) + \mathbf{w}(\mathbf{x},t_m)\big]$.
  $\mathbf{u}(\mathbf{x},t_{m+1}) \leftarrow \mathcal{K}\big(\mathbf{x},\sqrt{2\tau}\big) * \big[\sqrt{\alpha}\ \mathbf{w}(\mathbf{x},t_{m+1}) + \mathbf{u}(\mathbf{x},t_m)\big]$.
  $m \leftarrow m + 1$.
until convergence.

Algorithm 10 Gaussian Convolution Algorithm for Large Deformation Curvature Registration

Select time step τ; define $t_j = j\tau$, $j = 0, 1, \ldots$.
Set $\mathbf{v}(\mathbf{x}, t_0) = \mathbf{w}(\mathbf{x}, t_0) = \mathbf{u}(\mathbf{x}, t_0) = 0$, $m = 0$.
repeat
  $\mathbf{w}(\mathbf{x}, t_{m+1}) \leftarrow \mathcal{K}\big(\mathbf{x}, \sqrt{2\tau}\big) * \big[\sqrt{\alpha}\ \mathbf{f}(\mathbf{x}, R, F, \mathbf{u}(\mathbf{x}, t_m)) + \mathbf{w}(\mathbf{x}, t_m)\big]$.
  $\mathbf{v}(\mathbf{x}, t_{m+1}) \leftarrow \mathcal{K}\big(\mathbf{x}, \sqrt{2\tau}\big) * \big[\sqrt{\alpha}\ \mathbf{w}(\mathbf{x}, t_{m+1}) + \mathbf{v}(\mathbf{x}, t_m)\big]$.
  $\mathbf{u}(\mathbf{x}, t_{m+1}) \leftarrow \mathbf{u}(\mathbf{x}, t_m) + \tau\,[\mathbf{I} - \nabla\mathbf{u}(\mathbf{x}, t_m)]^{\mathrm{T}}\mathbf{v}(\mathbf{x}, t_{m+1})$.
  $m \leftarrow m + 1$.
until convergence.

Algorithm 11 Gaussian Convolution Algorithm for Elastic Registration

Select time step τ; define $t_j = j\tau$, $j = 0, 1, \ldots$.
Set $\mathbf{w}(\mathbf{x}, t_0) = \mathbf{u}(\mathbf{x}, t_0) = 0$, $m = 0$.
repeat
  $\mathbf{h}(\mathbf{x}, t_m) \leftarrow \mathcal{L}_1\mathbf{f}(\mathbf{x}, R, F, \mathbf{u}(\mathbf{x}, t_m))\,/(\mu(\lambda + 2\mu))$
  $\mathbf{w}(\mathbf{x}, t_{m+1}) \leftarrow \mathcal{K}\big(\mathbf{x}, \sqrt{2\tau}\big) * \big[\sqrt{\alpha}\ \mathbf{h}(\mathbf{x}, t_m) + \mathbf{w}(\mathbf{x}, t_m)\big]$.
  $\mathbf{u}(\mathbf{x}, t_{m+1}) \leftarrow \mathcal{K}\big(\mathbf{x}, \sqrt{2\tau}\big) * \big[\sqrt{\alpha}\ \mathbf{w}(\mathbf{x}, t_{m+1}) + \mathbf{u}(\mathbf{x}, t_m)\big]$.
  $m \leftarrow m + 1$.
until convergence.

5.5.3.2 Elastic Registration

To construct a successive Gaussian convolution algorithm for elastic registration, we use the same trick as in Section 5.4.3 of applying a scaled version of the linear operator $\mathcal{L}_1$ from (5.51) to the Navier-Lamé operator in order to establish a relationship with the biharmonic equation; i.e.,

$$-\mu\Delta\mathbf{u} - (\lambda + \mu)\,\nabla\mathrm{div}\,\mathbf{u} = \mathbf{f} \qquad \Longrightarrow \qquad \Delta^2\mathbf{u} = \frac{\mathcal{L}_1\mathbf{f}}{\mu\,(\lambda + 2\mu)}\ . \tag{5.73}$$

Note that this relationship is written in terms of the differential operators, not their discretized versions, and therefore the inverse is not necessarily implied. However, we ignore this problem and implicitly assume an equivalence. This allows us to exploit the result of Lemma 1 to define a successive Gaussian convolution algorithm for elastic registration, which is described in Algorithm 11. A fluid (or large deformation elastic) version of this algorithm can be constructed by operating on velocities and adding an Euler integration step, as illustrated in Algorithm 12.

Algorithm 12 Gaussian Convolution Algorithm for Fluid Registration

Select time step τ; define $t_j = j\tau$, $j = 0, 1, \ldots$.
Set $\mathbf{v}(\mathbf{x}, t_0) = \mathbf{w}(\mathbf{x}, t_0) = \mathbf{u}(\mathbf{x}, t_0) = 0$, $m = 0$.
repeat
  $\mathbf{h}(\mathbf{x}, t_m) \leftarrow \mathcal{L}_1 \mathbf{f}(\mathbf{x}, R, F, \mathbf{u}(\mathbf{x}, t_m)) / (\mu(\lambda + 2\mu))$
  $\mathbf{w}(\mathbf{x}, t_{m+1}) \leftarrow \mathcal{K}\big(\mathbf{x}, \sqrt{2\tau}\big) * \big[\sqrt{\alpha}\ \mathbf{h}(\mathbf{x}, t_m) + \mathbf{w}(\mathbf{x}, t_m)\big]$.
  $\mathbf{v}(\mathbf{x}, t_{m+1}) \leftarrow \mathcal{K}\big(\mathbf{x}, \sqrt{2\tau}\big) * \big[\sqrt{\alpha}\ \mathbf{w}(\mathbf{x}, t_{m+1}) + \mathbf{v}(\mathbf{x}, t_m)\big]$.
  $\mathbf{u}(\mathbf{x}, t_{m+1}) \leftarrow \mathbf{u}(\mathbf{x}, t_m) + \tau\, [\mathbf{I} - \nabla \mathbf{u}(\mathbf{x}, t_m)]^{\mathrm{T}} \mathbf{v}(\mathbf{x}, t_{m+1})$.
  $m \leftarrow m + 1$.
until convergence.

5.5.3.3 Second-order Elastic Registration

In order to construct a successive Gaussian convolution algorithm that uses the second-order elastic regularizer, we can no longer rely on Lemma 1, because we cannot establish an equivalence between the Euler-Lagrange equations for the second-order elastic regularizer and the biharmonic equation. However, we *can* exploit the relationship with the triple Laplacian equation by using the linear operator $\mathcal{L}_3$ from (5.58); i.e.,

$$\mu \Delta^2 \mathbf{u} + (\lambda + 2\mu)\, \Delta \nabla \mathrm{div}\, \mathbf{u} = \mathbf{f} \qquad \Longrightarrow \qquad (-\Delta)^3\, \mathbf{u} = \frac{\mathcal{L}_3 \mathbf{f}}{\mu\, (\lambda + 3\mu)}\ . \tag{5.74}$$

As in the case of elastic registration, this relationship does not necessarily hold in the opposite direction, although we will ignore this problem for the sake of constructing a successive Gaussian convolution algorithm.

The triple Laplacian equation can be related to a system of three partial differential equations, as described by Lemma 2:

Lemma 2. *The stationary solution of the coupled diffusion equations:*

$$\partial_t \mathbf{w}(\mathbf{x}, t) - \Delta \mathbf{w}(\mathbf{x}, t) = \sqrt[3]{\alpha}\, \mathbf{f}(\mathbf{x}, R, F, \mathbf{u}(\mathbf{x}, t_m)) \tag{5.75}$$

$$\partial_t \mathbf{z}(\mathbf{x}, t) - \Delta \mathbf{z}(\mathbf{x}, t) = \sqrt[3]{\alpha}\, \mathbf{w}(\mathbf{x}, t) \tag{5.76}$$

$$\partial_t \mathbf{u}(\mathbf{x}, t) - \Delta \mathbf{u}(\mathbf{x}, t) = \sqrt[3]{\alpha}\, \mathbf{z}(\mathbf{x}, t) \tag{5.77}$$

$$\mathbf{w}(\mathbf{x}, 0) = \mathbf{w}(\mathbf{x}, 0) = \mathbf{u}(\mathbf{x}, 0) = 0\ , \tag{5.78}$$

is equivalent to the solution of the equation $(-\Delta)^3\, \mathbf{u}(\mathbf{x}) = \alpha\, \mathbf{f}(\mathbf{x}, R, F, \mathbf{u}(\mathbf{x}, t_m))$.

Algorithm 13 Gaussian Convolution Algorithm for Second-order Elastic Registration

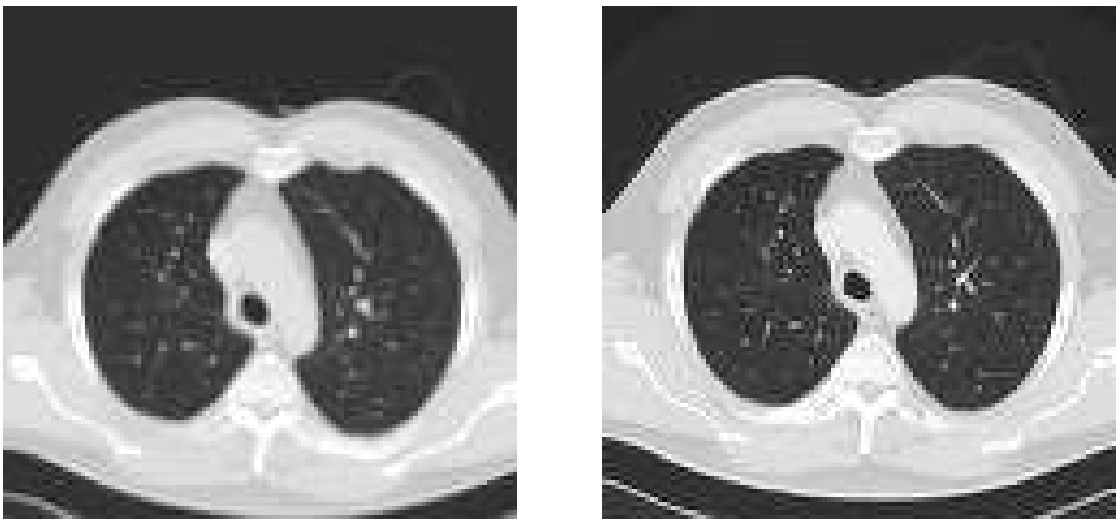
Select time step τ; define $t_j = j\tau$, $j = 0, 1, \ldots$.
Set $\mathbf{w}(\mathbf{x}, t_0) = \mathbf{z}(\mathbf{x}, t_0) = \mathbf{u}(\mathbf{x}, t_0) = 0$, $m = 0$.
repeat
  $\mathbf{h}(\mathbf{x}, t_m) \leftarrow \mathcal{L}_3\mathbf{f}(\mathbf{x}, R, F, \mathbf{u}(\mathbf{x}, t_m))\,/(\mu(\lambda + 3\mu))$
  $\mathbf{w}(\mathbf{x}, t_{m+1}) \leftarrow \mathcal{K}\big(\mathbf{x}, \sqrt{2\tau}\big) * \big[\sqrt[3]{\alpha}\ \mathbf{h}(\mathbf{x}, t_m) + \mathbf{w}(\mathbf{x}, t_m)\big]$.
  $\mathbf{z}(\mathbf{x}, t_{m+1}) \leftarrow \mathcal{K}\big(\mathbf{x}, \sqrt{2\tau}\big) * \big[\sqrt[3]{\alpha}\ \mathbf{w}(\mathbf{x}, t_{m+1}) + \mathbf{z}(\mathbf{x}, t_m)\big]$.
  $\mathbf{u}(\mathbf{x}, t_{m+1}) \leftarrow \mathcal{K}\big(\mathbf{x}, \sqrt{2\tau}\big) * \big[\sqrt[3]{\alpha}\ \mathbf{z}(\mathbf{x}, t_{m+1}) + \mathbf{u}(\mathbf{x}, t_m)\big]$.
  $m \leftarrow m + 1$.
until convergence.

Algorithm 14 Gaussian Convolution Algorithm for Large Deformation Second-order Elastic Registration

Select time step τ; define $t_j = j\tau$, $j = 0, 1, \ldots$.
Set $\mathbf{v}(\mathbf{x}, t_0) = \mathbf{w}(\mathbf{x}, t_0) = \mathbf{z}(\mathbf{x}, t_0) = \mathbf{u}(\mathbf{x}, t_0) = 0$, $m = 0$.
repeat
  $\mathbf{h}(\mathbf{x}, t_m) \leftarrow \mathcal{L}_3\mathbf{f}(\mathbf{x}, R, F, \mathbf{u}(\mathbf{x}, t_m))\,/(\mu(\lambda + 3\mu))$
  $\mathbf{w}(\mathbf{x}, t_{m+1}) \leftarrow \mathcal{K}\big(\mathbf{x}, \sqrt{2\tau}\big) * \big[\sqrt[3]{\alpha}\ \mathbf{h}(\mathbf{x}, t_m) + \mathbf{w}(\mathbf{x}, t_m)\big]$.
  $\mathbf{z}(\mathbf{x}, t_{m+1}) \leftarrow \mathcal{K}\big(\mathbf{x}, \sqrt{2\tau}\big) * \big[\sqrt[3]{\alpha}\ \mathbf{w}(\mathbf{x}, t_{m+1}) + \mathbf{z}(\mathbf{x}, t_m)\big]$.
  $\mathbf{v}(\mathbf{x}, t_{m+1}) \leftarrow \mathcal{K}\big(\mathbf{x}, \sqrt{2\tau}\big) * \big[\sqrt[3]{\alpha}\ \mathbf{z}(\mathbf{x}, t_{m+1}) + \mathbf{v}(\mathbf{x}, t_m)\big]$.
  $\mathbf{u}(\mathbf{x}, t_{m+1}) \leftarrow \mathbf{u}(\mathbf{x}, t_m) + \tau\,[\mathbf{I} - \nabla\mathbf{u}(\mathbf{x}, t_m)]^{\mathrm{T}}\mathbf{v}(\mathbf{x}, t_{m+1})$.
  $m \leftarrow m + 1$.
until convergence.

Proof. At the stationary solution of the coupled diffusion equations, $\partial_t\mathbf{w}(\mathbf{x}, t) = \partial_t\mathbf{z}(\mathbf{x}, t) = \partial_t\mathbf{u}(\mathbf{x}, t) = 0$. Therefore, we have $-\Delta\mathbf{w}(\mathbf{x}, t) = \sqrt[3]{\alpha}\,\mathbf{f}(\mathbf{x}, R, F, \mathbf{u}(\mathbf{x}, t_m))$, $-\Delta\mathbf{z}(\mathbf{x}, t) = \sqrt[3]{\alpha}\,\mathbf{w}(\mathbf{x}, t)$, and $-\Delta\mathbf{u}(\mathbf{x}, t) = \sqrt[3]{\alpha}\,\mathbf{z}(\mathbf{x}, t)$. Solving the third equation for $\mathbf{z}(\mathbf{x}, t)$, substituting into the second equation, solving for $\mathbf{w}(\mathbf{x}, t)$ and substituting into the first equation yields the desired result. □

Hence, we can design an algorithm to perform second-order elastic registration by modifying the force vector $\mathbf{f}$ and then solving the diffusion equations in Lemma 2. This process is described in Algorithm 13. A large deformation version of this algorithm can be constructed by operating on velocities and adding an Euler integration step, as illustrated in Algorithm 14.



(a) Reference Slice (b) Floating Slice

Figure 5.1: Axial slices from 3D CT chest scans.

5.6 Registration Experiment

In this section, we perform an experiment that focuses on the practical problem of registering serial chest CT studies. In this experiment, Fourier methods and successive Gaussian convolution algorithms using all of the homogeneous regularizers are tested and compared according to how well they predict the positions of manually identified nodules as well as the amount of computation required.

5.6.1 Serial Chest CT Images

The data we use comprises a set of serial chest CT examinations of 18 patients with lung nodules. Each serial examination included a prior image and a current image. Axial slices of the one of the patient examinations are shown in Fig. 5.1.

5.6.2 Experimental Setup

For each patient, we resampled the images to approximately $3 \times 3 \times 3\text{mm}^3$ isotropic resolution and performed a rigid preregistration step. We then performed nonrigid registration using the large deformation versions of the successive Gaussian convolution algorithm with the diffusion, curvature, elastic, and second-order elastic regularizers, and with the SCC

(squared correlation coefficient) dissimilarity measure. Values for α were chosen from the set $\{0.001, 0.01, 0.1\}$. (Other values of α were tried; however, larger values tended to cause the registration to diverge, whereas smaller values caused a very large number of iterations.) The time step τ was chosen to be identical to one voxel, and the Gaussian kernels were discretized and truncated to $5 \times 5 \times 5$ voxels. Registration was performed in a multiresolution pyramid at four resolution levels; at each level, the registration process was terminated if there was no significant change in SCC or if a maximum number of 50 iterations have been performed.

To establish a set of ground truth points that can be used to measure target registration error (TRE), we manually identified the centers of lung nodules less than 6mm in diameter that are observable in both the prior and current images of a patient. This yielded from 4-20 ground truth points for each patient. After each nonrigid registration algorithm was performed, the manually identified locations of nodules in the prior image were mapped through the resulting deformation field to predict their positions in the current image. The TRE is then defined as the Euclidean distance between the predicted positions and the manually identified positions of the nodules in the current image.

Computational requirements of each algorithm are measured in terms of *effective convolution steps* (ECS). A unit of ECS is defined as the amount of computation required to convolve a vector field at the finest resolution level with a Gaussian kernel. Hence, 50 iterations of successive Gaussian convolution (with the diffusion regularizer) at each of the two finest resolution levels would require $50 + (1/8) * 50 = 56.25$ ECS. Curvature and elastic-based regularizers would require twice as many Gaussian convolution steps, yielding 112.5 ECS, and second-order elastic regularizers would require three times as many, yielding 168.75 ECS. For the Fourier-based methods, we note that in our implementation, solution of one linear system at the finest resolution level requires approximately twice the computation as convolution with a $5 \times 5 \times 5$ kernel; hence, we compute ECS for Fourier-based methods by multiplying the number of iterations at each resolution level by 2 and weighting according to the resolution level. Therefore, 50 iterations of any Fourier-based registration algorithm at each of the two finest resolution levels would require 112.5 ECS.

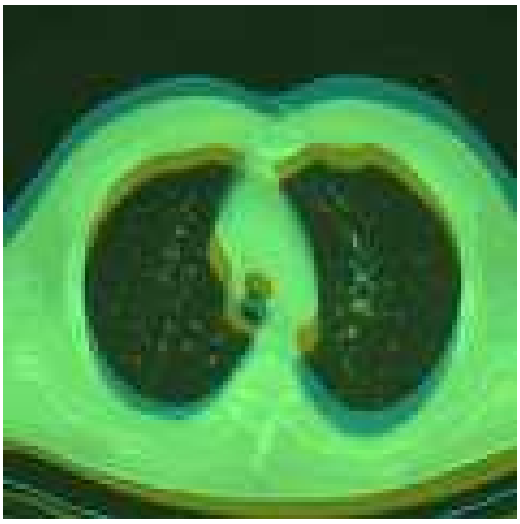
5.6.3 Results

To visualize how nonrigid registration brings images into alignment, we use the example axial slices from Fig. 5.1. Figure 5.2 shows color-coded, fused versions of the slices, with no registration, and then after the floating image has been warped according to various registration algorithms.

Tables 5.3 and 5.4 list various statistics that allow a direct comparison of the speed and accuracy of the various registration procedures. Target registration error (TRE) is computed for each nodule as the distance between predicted location and actual location of the nodule in the current image. Aggregate TRE (ATRE) is computed for each case as the median TRE of all the nodules for a case; the ATRE allows a TRE-based measure that is normalized for the different numbers of nodules in each case. Finally, effective convolution steps (ECS) are computed for each algorithm and roughly allow algorithms to be compared in terms of their relative efficiency.

Based on the results reported in Tables 5.3 and 5.4, we see that all of the variations of nonrigid registration appear to predict nodule locations to within voxel accuracy. In terms of TRE and ATRE, there does not appear to be one variation that stands out as significantly better than the rest. However, when ECS is taken into account, we see that with one exception, the second-order regularizers (curvature and second-order elastic) outperform their first order counterparts (diffusion and elastic, respectively). Overall, the choice of the curvature regularizer with $\alpha = 0.1$, using either the successive Gaussian convolution algorithm or the Fourier-based approach appears to yield the best performance in terms of ECS.

To further investigate the amount of computational effort required by both algorithms with various regularizers, we provide box plots to illustrate the distributions of the number of iterations required at each level of the multilevel pyramid. Figure 5.3 shows these distributions for the case of registration via successive Gaussian convolution, and Fig. 5.4 shows the distributions for the case of registration via Fourier methods. In each of these figures, box plots illustrate the number of iterations required for each particular regularizer at the



(c) No Registration

(d) Diffusion Registration (Fourier)

(e) Curvature Registration (Gaussian Conv.)

(f) Curvature Registration (Fourier)

Figure 5.2: Axial slices from 3D CT chest scans, colored cyan and yellow and fused.

Registration Type	Mean TRE (mm)	Median TRE (mm)	Std. dev. TRE (mm)	Mean ATRE (mm)	Median ATRE (mm)	Std. dev. ATRE (mm)	25$^{\text{th}}$ Percentile ECS	50$^{\text{th}}$ Percentile ECS	75$^{\text{th}}$ Percentile ECS
Rigid Only	7.2	5.8	4.9	7.7	7.0	3.3	–	–	–
Diffusion									
$\alpha = 0.001$	6.8	5.7	5.0	7.2	6.1	3.8	20.1	28.2	52.3
$\alpha = 0.01$	6.1	4.7	4.8	6.3	5.6	4.0	53.7	54.8	56.6
$\alpha = 0.1$	6.3	4.3	6.9	6.4	5.1	4.2	22.6	27.8	41.3
$\alpha = 1.0$	6.8	5.1	5.7	6.7	6.0	4.1	8.8	14.4	27.0
Elastic									
$\alpha = 0.001$	6.9	5.7	5.0	7.2	5.9	3.8	15.5	21.5	36.8
$\alpha = 0.01$	6.6	5.4	5.0	6.8	5.6	4.0	101.8	107.1	113.0
$\alpha = 0.1$	6.4	5.1	4.9	6.1	4.9	4.0	76.2	87.8	90.9
$\alpha = 1.0$	7.4	5.8	5.4	7.0	5.8	4.1	22.0	26.3	65.4
Curvature									
$\alpha = 0.001$	6.1	4.7	4.9	6.6	5.7	4.2	98.3	104.2	109.5
$\alpha = 0.01$	5.9	4.4	4.9	6.2	5.2	4.3	48.6	50.2	53.7
$\alpha = 0.1$	6.2	4.5	5.4	6.3	5.0	4.6	19.8	20.0	21.8
$\alpha = 1.0$	7.1	4.8	7.3	6.4	4.7	4.7	11.5	13.5	17.6
2^{nd}-order Elastic									
$\alpha = 0.001$	6.9	5.7	5.0	7.1	5.9	3.8	17.6	23.3	57.4
$\alpha = 0.01$	6.6	5.4	5.1	6.6	5.9	4.1	71.5	77.8	85.4
$\alpha = 0.1$	7.0	5.5	5.6	6.8	5.5	4.2	36.0	42.0	48.2
$\alpha = 1.0$	6.9	5.5	5.0	6.9	5.8	3.6	26.2	26.2	29.6

Table 5.3: Statistics of TRE across all nodules, ATRE across all cases, and ECS for rigid registration and nonrigid registration via large deformation successive Gaussian convolution algorithm with each homogeneous regularizer and various values of α.

Registration Type	Mean TRE (mm)	Median TRE (mm)	Std. dev. TRE (mm)	Mean ATRE (mm)	Median ATRE (mm)	Std. dev. ATRE (mm)	25^{th} Percentile ECS	50^{th} Percentile ECS	75^{th} Percentile ECS
Rigid Only	7.2	5.8	4.9	7.7	7.0	3.3	–	–	–
Diffusion									
$\alpha = 0.001$	6.7	5.5	4.9	6.9	5.9	3.8	114.2	114.2	114.2
$\alpha = 0.01$	5.8	4.8	4.6	5.9	5.2	3.7	114.2	114.2	114.2
$\alpha = 0.1$	5.3	4.2	4.1	5.3	5.0	2.1	93.4	108.1	111.4
$\alpha = 1.0$	6.6	5.5	4.2	6.4	5.9	2.7	24.8	35.8	75.1
Elastic									
$\alpha = 0.001$	6.8	5.7	4.9	7.1	6.0	3.7	114.2	114.2	114.2
$\alpha = 0.01$	6.5	5.6	4.5	6.5	5.9	2.9	104.4	110.7	113.5
$\alpha = 0.1$	6.6	5.7	4.6	6.4	5.9	3.0	18.2	20.6	25.8
$\alpha = 1.0$	7.2	6.6	4.7	6.7	5.8	2.9	11.8	15.5	17.7
Curvature									
$\alpha = 0.001$	6.6	5.3	4.9	6.7	5.8	3.8	114.2	114.2	114.2
$\alpha = 0.01$	5.7	4.3	4.7	5.9	4.8	4.0	113.5	113.9	114.2
$\alpha = 0.1$	6.1	4.5	5.9	6.0	5.8	3.0	40.8	50.3	81.9
$\alpha = 1.0$	7.2	5.8	4.5	7.3	6.9	3.0	19.8	27.8	56.1
2^{nd}-order Elastic									
$\alpha = 0.001$	6.8	5.7	4.9	7.0	6.0	3.7	114.2	114.2	114.2
$\alpha = 0.01$	6.6	5.9	4.7	6.5	5.7	3.1	105.7	110.9	113.4
$\alpha = 0.1$	6.8	6.0	4.8	6.7	5.7	3.3	21.6	27.3	33.1
$\alpha = 1.0$	7.6	6.6	4.8	6.8	5.9	2.8	11.5	11.5	11.5

Table 5.4: Statistics of TRE across all nodules, ATRE across all cases, and ECS for rigid registration and nonrigid registration via large deformation Fourier solvers with each homogeneous regularizer, homogeneous Dirichlet boundary conditions on the displacement and velocity fields, and various values of α.

low resolution, mid-level resolution, and high resolution levels of a multiresolution pyramid. For consistency, α is chosen to be 0.1, as this value seems to provide the best general TRE values across all regularizers.

An interesting observation that can be made from Fig.'s 5.3–5.4 is that in registration via successive Gaussian convolution, the first order regularizers (diffusion and elastic) require many more iterations at all levels than the second order regularizers (curvature and second-order elastic). However, when registration is performed via Fourier methods, it is the diffusion and curvature regularizers that require more iterations at all levels than the elastic and second-order elastic regularizers.

5.7 Summary

This chapter presented strategies for approximating the solutions to the PDE systems that arise in variational registration. It illustrated fixed-point and steepest descent iterative approaches for transforming the semilinear or nonlinear PDE systems that arise from various regularizers into successions of linear PDE systems. It then showed how the resulting linear PDE systems can be discretized and solved via general iterative solvers. The chapter then described how the structure of the linear systems arising from any of the homogeneous regularizers can be exploited, and how Fourier methods and successive Gaussian convolution algorithms can be constructed to rapidly solve the linear systems in these situations. Finally, the chapter illustrated the performance of the proposed algorithms on the problem of registering serial 3D chest CT images of patients with lung nodules.

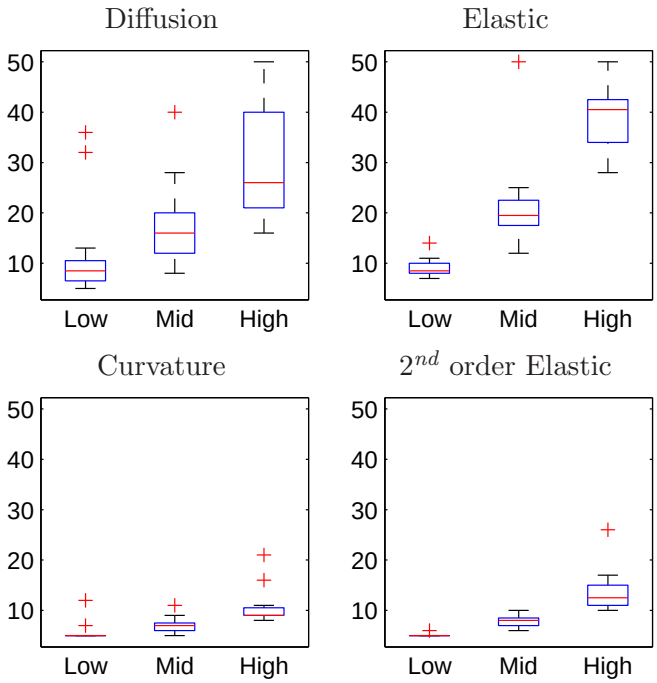


Figure 5.3: Distributions of the number of iterations required at the low resolution, medium resolution, and high resolution levels in multilevel registration using successive Gaussian convolution.

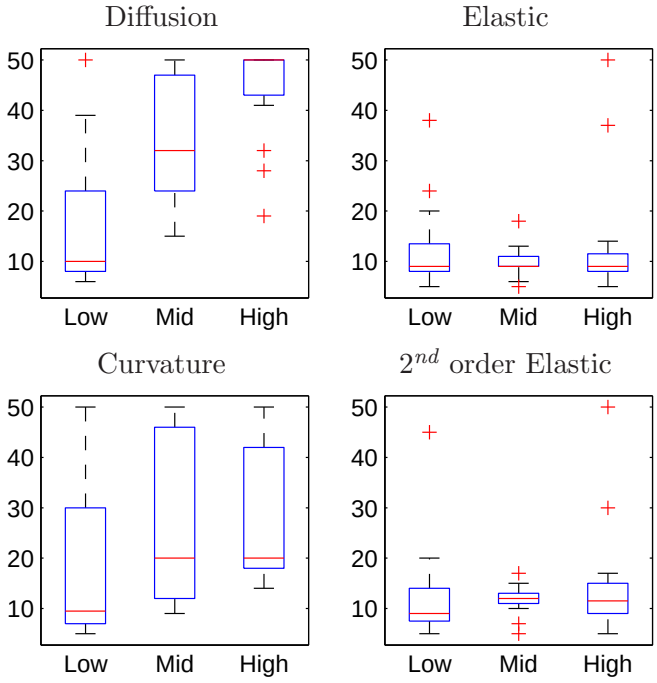


Figure 5.4: Distributions of the number of iterations required at the low resolution, medium resolution, and high resolution levels in multilevel registration using Fourier methods.

Chapter 6

Conclusions

This thesis has explored key components of variational image registration problem; namely, dissimilarity measures, changing overlap regions, regularizers, and computational solution techniques. In this chapter, we summarize the key contributions of this thesis, and then we provide some ideas for future investigation.

6.1 Contributions of this Thesis

6.1.1 Dissimilarity Measures

In Chapter 2, we described a variety of assumptions on the relationship between images being registered, and we classified standard dissimilarity measures according to these assumptions. In the process, we introduced three new dissimilarity measures: the squared orthogonal correlation coefficient (SOCC), normalized cross cumulative residual entropy (NCCRE), and cumulative residual entropy correlation coefficient (CRECC). In addition, we showed how to construct local versions of any dissimilarity measure, using local moments or local distributions. Finally, we presented the Gâteaux derivatives for the global and local versions of all dissimilarity measures.

6.1.2 Changing Overlap

In Chapter 3, we investigated problems that emerge when considering the changing nature of the overlap region over the course of registration. We showed how the Gâteaux derivatives presented in the previous chapter can be computed under the general assumption that the overlap region varies according to the deformation field. We then showed how many

dissimilarity measures suffer from a sensitivity to changing overlap when background content moves in and out of the overlap region. We proposed a method for modifying the dissimilarity measures to remove this sensitivity. Finally, we showed how the modifications affect the Gâteaux derivatives presented in the previous chapter.

6.1.3 Regularizers

In Chapter 4, we described the quadratic taxonomy of homogeneous regularizers used in medical image registration as well as the Weickert taxonomy of nonhomogeneous regularizers used in optical flow. We extended the quadratic taxonomy to include a second-order elastic regularizer, and then we illustrated how the two taxonomies can be combined in order to construct nonhomogeneous versions of any of the quadratic regularizers. Finally, we presented the Gâteaux derivatives of every regularizer in the combined taxonomy.

6.1.4 Solving the Variational Registration Problem

In Chapter 5, we presented strategies for solving the PDE systems arising from the Gâteaux derivatives of the dissimilarity measures and regularizers. By employing semi-implicit discretizations in conjunction with fixed-point or steepest descent algorithms, we showed that any PDE system can be transformed into a sequence of linear algebraic equations. We illustrated that iterative techniques can be applied to solve the successive linear algebraic equations for any regularizer, and then we focused on how Fourier methods and successive Gaussian convolution can be used to rapidly solve the equations arising from any of the homogeneous regularizers.

6.2 Future Work

The field of variational image registration is ripe with problems for future research. In this section, we describe a few of these problems, along with some suggestions of ideas that might be applicable in addressing these problems.

6.2.1 Optimal Dissimilarity Measures

The dissimilarity measures presented in Chapter 2 are all described in terms of assumptions on the underlying relationship between the images. A natural question arises: is there a way to identify an *optimal* dissimilarity measure for a particular registration problem? One way to approach an answer is to learn the joint intensity distribution from exemplar aligned images, and then to define a dissimilarity measure based on how different an observed joint distribution is from the learned joint distribution. Some examples of this type of approach are found in [32, 60, 78]. Another approach involves learning an optimal dissimilarity measure directly from exemplar aligned and misaligned image patches [57].

The difficulty with generalizing any of these approaches to variational registration lies in the practical problem of acquiring suitable examples of aligned images with which to train the dissimilarity measure. This may be possible by capturing images from different modalities simultaneously, as with PET/CT or PET/MR devices. However, a dissimilarity measure learned from PET and CT images captured by a PET/CT device may not be optimal if applied to arbitrary PET and CT images that are captured from separate devices with different protocols. But, if it can be ensured that adequate aligned training images are available, then it should be possible to generalize these approaches to variational registration by deriving and employing the force vectors corresponding to the learned dissimilarity measure.

6.2.2 Rapid Computation of Local Dissimilarity Measures

A major difficulty with the use of local dissimilarity measures in variational image registration is that they can be prohibitively expensive to compute, especially if the local neighborhoods are relatively wide. A variety of approaches [12, 77, 79, 85] have been proposed to approximate mutual information (or other global dissimilarity measures) by subsampling the image data. These approaches aim to come up with rapid, robust approximations to *global* dissimilarity measures. Since the computation of *local* dissimilarity measures can require orders of magnitude more computations than the computation of global dissimilarity measures, it is critical to develop rapid ways to compute or approximate the local

dissimilarity measures and their corresponding force vectors.

6.2.3 Handling Truncated Data

The modified dissimilarity measures presented in Section 3.4 all rely on the assumption that only background (non-interesting) content is moving in and out of the overlap region over the course of registration. Section 3.3 describes the behavior of probability distributions and image statistics over the changing overlap regions; this behavior is only valid when the background statistics assumption holds. In practice, however, there are many situations where the background statistics assumption is violated. One common example is when images contain truncated data, for example, at the top and bottom slices of a chest CT exam. These situations can cause undesirable behavior, generating modified probability distributions with values outside of the range $[0, 1]$, or modified variance values that are negative. An open question is whether or not dissimilarity measures can be modified in a suitable manner in situations where the background statistics assumption is violated.

6.2.4 Anisotropic Generalizations of Homogeneous Regularizers

As alluded to in Section 4.3.3, it is tempting to substitute $\mathbf{B}(\mathbf{x})\,\nabla\mathbf{u}(\mathbf{x}) \rightarrow \nabla\mathbf{u}(\mathbf{x})$, where $\mathbf{B}(\mathbf{x})$ is the generalization of Nagel's oriented smoothness matrix defined in (4.21), into (4.37)–(4.39) in order to construct anisotropic versions of the elastic, curvature, and second-order elastic regularizers. Even though this substitution works in generalizing the diffusion regularizer to its anisotropic version, this substitution does *not* enable the desired anisotropic behavior for the elastic, curvature, and second-order elastic regularizers. It appears that it is an open question as to whether or not it is possible to generalize the non-diffusion regularizers to anisotropic versions.

6.2.5 Rapid Registration with Nonhomogeneous Regularizers

Certainly, registration with any of the nonhomogeneous regularizers presented in Section 4.3 can be rapidly performed using a multigrid algorithm. In fact, Bruhn et al. [16] describe how the full multigrid (FMG) and full approximation scheme (FAS) can be employed for performing optical flow with any of the regularizers in the Weickert taxonomy. The

FMG and FAS approaches should be able to be generalized to the other nonhomogeneous regularizers in the same manner.

It would be of interest, however, if the structure of the nonhomogeneous regularizers can be exploited to design rapid solution algorithms that would be much simpler to implement than multigrid algorithms. We have demonstrated in [22] that this is possible at least for the nonhomogeneous version of the curvature regularizer, by representing the solution to the Euler-Lagrange equations as the stationary solution of a pair of coupled PDE systems and employing a Demons style algorithm, as in Section 5.5. It remains an open question as to whether or not any simple-to-implement, rapid solution algorithms exist for nonhomogeneous elastic or nonhomogeneous second-order elastic registration.

Appendix A

Equivalence of Dissimilarity Measures Under the Linearity Assumption

This appendix relates to Section 2.2.2 of Chapter 2, in which we state the equivalence of optimizing the MSVDL and SCC dissimilarity measures, and of optimizing the MSODL and SOCC dissimilarity measures.

Consider the MSVDL, MSODL, SCC, and SOCC dissimilarity measures, repeated here for reference:

$$MSVDL(R, F^{\mathbf{u}}, \boldsymbol{\beta}) = \frac{1}{|\Omega|} \int_{\Omega} \left(\beta_1 R(\mathbf{x}) + \beta_2 - F^{\mathbf{u}}(\mathbf{x})\right)^2 d\mathbf{x}, \tag{A.1}$$

$$MSODL(R, F^{\mathbf{u}}, \boldsymbol{\beta}) = \frac{1}{|\Omega|} \int_{\Omega} \frac{\left(\beta_1 R(\mathbf{x}) + \beta_2 - F^{\mathbf{u}}(\mathbf{x})\right)^2}{1 + \beta_1^2} d\mathbf{x}, \tag{A.2}$$

$$SCC(R, F^{\mathbf{u}}) = \frac{\mathrm{Cov}\left(R, F^{\mathbf{u}}\right)^2}{\mathrm{Var}\left(R\right)\mathrm{Var}\left(F^{\mathbf{u}}\right)}, \tag{A.3}$$

$$SOCC(R, F^{\mathbf{u}}) = \frac{\left(\mathrm{Var}\left(R\right) - \mathrm{Var}\left(F^{\mathbf{u}}\right)\right)^2 + 4\mathrm{Cov}\left(R, F^{\mathbf{u}}\right)^2}{\left(\mathrm{Var}\left(R\right) + \mathrm{Var}\left(F^{\mathbf{u}}\right)\right)^2}. \tag{A.4}$$

Under the assumption that Ω remains the same over the course of registration, we prove the following propositions:

Proposition 1. *Minimizing a normalized version of the MSVDL with respect to β and $\mathbf{u}$ is equivalent to maximizing the SCC with respect to $\mathbf{u}$.*

Proposition 2. *Minimizing a normalized version of the MSODL with respect to β and $\mathbf{u}$ is equivalent to maximizing the SOCC with respect to $\mathbf{u}$, so long as the covariance between the images is nonzero.*

Proposition 1 is not a new result, as it describes the well-known relationship between linear least squares data fitting and the correlation coefficient; however, we include its proof in order to establish a framework with which to prove Proposition 2.

Proof of Proposition 1. For a particular value of $\mathbf{u}$, an estimate $\hat{\boldsymbol{\beta}}(\mathbf{u})$ of the linear parameters relating R and $F^{\mathbf{u}}$ can be computed by solving the minimization problem:

$$\hat{\boldsymbol{\beta}}(\mathbf{u}) = \arg\min_{\boldsymbol{\beta}} MSVDL(R, F^{\mathbf{u}}, \beta)\,. \tag{A.5}$$

In order to solve (A.5), we introduce the notation $M_V(\boldsymbol{\beta}) := MSVDL(R, F^{\mathbf{u}}, \boldsymbol{\beta})$, and we rewrite $MSVDL$ using (2.37)–(2.39):

$$M_V(\boldsymbol{\beta}) = \boldsymbol{\beta}^{\mathrm{T}} \begin{pmatrix} \mathrm{Var}\,(R) + \bar{R}^2 & \bar{R} \\ \bar{R} & 1 \end{pmatrix} \boldsymbol{\beta} - 2\boldsymbol{\beta}^{\mathrm{T}} \begin{pmatrix} \mathrm{Cov}\,(R, F^{\mathbf{u}}) + \bar{R}\bar{F^{\mathbf{u}}} \\ \bar{F^{\mathbf{u}}} \end{pmatrix} + \mathrm{Var}\,(F^{\mathbf{u}}) + \bar{F^{\mathbf{u}}}^2. \tag{A.6}$$

The gradient of M_V with respect to β is given by:

$$\nabla M_V(\boldsymbol{\beta}) = 2 \begin{pmatrix} \mathrm{Var}\,(R) + \bar{R}^2 & \bar{R} \\ \bar{R} & 1 \end{pmatrix} \boldsymbol{\beta} - 2 \begin{pmatrix} \mathrm{Cov}\,(R, F^{\mathbf{u}}) + \bar{R}\bar{F^{\mathbf{u}}} \\ \bar{F^{\mathbf{u}}} \end{pmatrix}. \tag{A.7}$$

Solving the system $\nabla M_V\left(\hat{\boldsymbol{\beta}}\right) = 0$ for $\hat{\boldsymbol{\beta}}$ yields:

$$\hat{\beta}_1 = \frac{\mathrm{Cov}\,(R, F^{\mathbf{u}})}{\mathrm{Var}\,(R)}, \tag{A.8}$$

$$\hat{\beta}_2 = \bar{F^{\mathbf{u}}} - \bar{R}\frac{\mathrm{Cov}\,(R, F^{\mathbf{u}})}{\mathrm{Var}\,(R)}. \tag{A.9}$$

Since M_V is convex in $\boldsymbol{\beta}$, the solution to (A.5) is

$$\hat{\beta}_1(\mathbf{u}) = \frac{\mathrm{Cov}\,(R, F^{\mathbf{u}})}{\mathrm{Var}\,(R)}, \tag{A.10}$$

$$\hat{\beta}_2(\mathbf{u}) = \bar{F^{\mathbf{u}}} - \bar{R}\frac{\mathrm{Cov}\,(R, F^{\mathbf{u}})}{\mathrm{Var}\,(R)}. \tag{A.11}$$

Substituting (A.10) and (A.11) into (A.1) for $\boldsymbol{\beta}$ and simplifying yields:

$$MSVDL\left(R, F^{\mathbf{u}}, \hat{\boldsymbol{\beta}}\right) = \mathrm{Var}\,(F^{\mathbf{u}})\,(1 - SCC(R, F^{\mathbf{u}}))\,. \tag{A.12}$$

Since Ω remains the same over the course of registration, $MSVDL$, when normalized by $\mathrm{Var}\,(F^{\mathbf{u}})$, is a negative linear function of SCC. Hence, minimizing the ratio $MSVDL/\mathrm{Var}\,(F^{\mathbf{u}})$ with respect to $\boldsymbol{\beta}$ and $\mathbf{u}$ is equivalent to maximizing SCC with respect to $\mathbf{u}$. □

Proof of Proposition 2. For a particular value of $\mathbf{u}$, an estimate $\hat{\boldsymbol{\beta}}(\mathbf{u})$ of the linear parameters relating the images can be computed by solving the minimization problem:

$$\hat{\boldsymbol{\beta}}(\mathbf{u}) = \arg\min_{\boldsymbol{\beta}} MSODL(R, F^{\mathbf{u}}, \boldsymbol{\beta})\,. \tag{A.13}$$

Unlike $MSVDL$, $MSODL$ is not convex in $\boldsymbol{\beta}$, so we are not necessarily guaranteed that a solution to (A.13) exists. We will proceed by investigating the existence of any stationary points of $MSODL$.

Defining $M_O(\boldsymbol{\beta}) := MSODL(R, F^{\mathbf{u}}, \boldsymbol{\beta})$, we find that $MSODL$ can be rewritten as:

$$M_O(\boldsymbol{\beta}) = \frac{\boldsymbol{\beta}^{\mathrm{T}}\begin{pmatrix} \mathrm{Var}\,(R) + \bar{R}^2 & \bar{R} \\ \bar{R} & 1 \end{pmatrix}\boldsymbol{\beta} - 2\boldsymbol{\beta}^{\mathrm{T}}\begin{pmatrix} \mathrm{Cov}\,(R, F^{\mathbf{u}}) + \bar{R}\bar{F^{\mathbf{u}}} \\ \bar{F^{\mathbf{u}}} \end{pmatrix} + \mathrm{Var}\,(F^{\mathbf{u}}) + \bar{F^{\mathbf{u}}}^2}{1 + \beta_1^2}. \tag{A.14}$$

The gradient of M_O with respect to $\boldsymbol{\beta}$ is given by:

$$\nabla M_O(\boldsymbol{\beta}) = \frac{1}{1+\beta_1^2}\left[\nabla M_V(\boldsymbol{\beta}) - 2M_O(\boldsymbol{\beta})\begin{pmatrix} \beta_1 \\ 0 \end{pmatrix}\right]. \tag{A.15}$$

Setting $\nabla M_O\left(\hat{\boldsymbol{\beta}}\right) = 0$ for $\hat{\boldsymbol{\beta}}$ yields the system of equations:

$$\nabla M_V(\boldsymbol{\beta}) = 2M_O(\boldsymbol{\beta})\begin{pmatrix} \beta_1 \\ 0 \end{pmatrix}. \tag{A.16}$$

If we substitute the second equation in (A.7) into the second equation in (A.16), we see that setting $\hat{\beta}_2 = \bar{F^{\mathbf{u}}} - \bar{R}\beta_1$ yields $\partial M_V(\boldsymbol{\beta})/\partial\beta_2 = 0$. Substituting $\hat{\beta}_2$ into the first equation in (A.16) for β_2 and simplifying yields the following quadratic equation in β_1:

$$\beta_1^2 \mathrm{Cov}\,(R, F^{\mathbf{u}}) + \beta_1(\mathrm{Var}\,(R) - \mathrm{Var}\,(F^{\mathbf{u}})) - \mathrm{Cov}\,(R, F^{\mathbf{u}}) = 0. \tag{A.17}$$

If we assume that $\mathrm{Cov}\,(R, F^{\mathbf{u}}) \neq 0$, then (A.17) has two distinct real roots:

$$\hat{\beta}_1 = \frac{\mathrm{Var}\,(F^{\mathbf{u}}) - \mathrm{Var}\,(R) \pm \sqrt{(\mathrm{Var}\,(R) - \mathrm{Var}\,(F^{\mathbf{u}}))^2 + 4\mathrm{Cov}\,(R, F^{\mathbf{u}})^2}}{2\mathrm{Cov}\,(R, F^{\mathbf{u}})}. \tag{A.18}$$

Second-order analysis verifies that the only one of these roots is a local minimum:

$$\hat{\beta}_1 = \frac{\mathrm{Var}\,(F^{\mathbf{u}}) - \mathrm{Var}\,(R) + \sqrt{(\mathrm{Var}\,(R) - \mathrm{Var}\,(F^{\mathbf{u}}))^2 + 4\mathrm{Cov}\,(R, F^{\mathbf{u}})^2}}{2\mathrm{Cov}\,(R, F^{\mathbf{u}})}. \tag{A.19}$$

The other root is a local maximum.

Substituting $\hat{\beta}_1$ from (A.19) and $\hat{\beta}_2$ for $\boldsymbol{\beta}$ in (A.14) and simplifying yields:

$$M_O\left(\hat{\boldsymbol{\beta}}\right) = \frac{|\Omega|\left(\mathrm{Var}\left(R\right) + \mathrm{Var}\left(F^{\mathbf{u}}\right)\right)}{2}\left(1 - \sqrt{\frac{\left(\mathrm{Var}\left(R\right) - \mathrm{Var}\left(F^{\mathbf{u}}\right)\right)^2 + 4\mathrm{Cov}\left(R, F^{\mathbf{u}}\right)^2}{\left(\mathrm{Var}\left(R\right) + \mathrm{Var}\left(F^{\mathbf{u}}\right)\right)^2}}\right), \tag{A.20}$$

or

$$\frac{MSODL\left(R, F^{\mathbf{u}}, \hat{\boldsymbol{\beta}}\right)}{\left(\mathrm{Var}\left(R\right) + \mathrm{Var}\left(F^{\mathbf{u}}\right)\right)} = \frac{|\Omega|}{2}\left(1 - \sqrt{SOCC(R, F^{\mathbf{u}})}\right). \tag{A.21}$$

Since Ω remains the same over the course of registration, the negative of the ratio of $MSODL$ to $\mathrm{Var}\left(R\right) + \mathrm{Var}\left(F^{\mathbf{u}}\right)$ is a convex function of $SOCC$. Hence, the proposition holds. □

Appendix B

Gâteaux Derivatives of Dissimilarity Measures

This appendix relates to Section 2.4 of Chapter 2, in which we present Gâteaux derivatives for the dissimilarity measures. It proves the following six propositions:

Proposition 3. *If the dissimilarity measure* $\mathcal{I}(R, F^{\mathbf{u}})$ *can be expressed as a functional of the form:*

$$\mathcal{I}(R, F^{\mathbf{u}}) = \int_{\Omega} \mathcal{M}(R(\mathbf{x}), F^{\mathbf{u}}(\mathbf{x}))\, d\mathbf{x}, \tag{B.1}$$

then the Gâteaux derivative $d\mathcal{I}(R, F^{\mathbf{u}}; \mathbf{w})$ *can be expressed as:*

$$d\mathcal{I}(R, F^{\mathbf{u}}; \mathbf{w}) = \int_{\Omega} \langle \mathcal{P}(\mathbf{x}; R, F^{\mathbf{u}}), \mathbf{w}(\mathbf{x}) \rangle\, d\mathbf{x}. \tag{B.2}$$

Proposition 4. *The Gâteaux derivative of the entropy* $H(F^{\mathbf{u}})$ *with respect to* $\mathbf{w}$ *can be written in terms of* (2.114)*, with* $\mathcal{P}^g_H(\mathbf{x}; F^{\mathbf{u}})$ *approximated by* (2.146)*.*

Proposition 5. *The Gâteaux derivative of the joint entropy* $H(R, F^{\mathbf{u}})$ *with respect to* $\mathbf{w}$ *can be written in terms of* (2.114)*, with* $\mathcal{P}^g_H(\mathbf{x}; R, F^{\mathbf{u}})$ *approximated by* (2.147)*.*

Proposition 6. *The Gâteaux derivative of the cumulative residual entropy* $\varepsilon(F^{\mathbf{u}})$ *with respect to* $\mathbf{w}$ *can be written in terms of* (2.114)*, with* $\mathcal{P}^g_\varepsilon(\mathbf{x}; F^{\mathbf{u}})$ *approximated by* (2.148)*.*

Proposition 7. *The Gâteaux derivative of the joint cumulative residual entropy* $\varepsilon(R, F^{\mathbf{u}})$ *with respect to* $\mathbf{w}$ *can be written in terms of* (2.114)*, with* $\mathcal{P}^g_\varepsilon(\mathbf{x}; R, F^{\mathbf{u}})$ *approximated by* (2.149)*.*

Proposition 8. *The Gâteaux derivative of the joint cumulative residual entropy* $\varepsilon(F^{\mathbf{u}}, R)$ *with respect to* $\mathbf{w}$ *can be written in terms of* (2.114)*, with* $\mathcal{P}^g_\varepsilon(\mathbf{x}; F^{\mathbf{u}}, R)$ *approximated by* (2.150)*.*

Propositions 3–5 are not new results; however, we include their proofs in order to establish a framework with which to prove Propositions 6–8.

Proof of Proposition 3. By the definition of the Gâteaux derivative, we have:

$$\begin{aligned}
d\mathcal{I}(R, F^{\mathbf{u}}; \mathbf{w}) &= \lim_{h\to 0} \frac{1}{h}\left[\int_\Omega \mathcal{M}\Big(R(\mathbf{x}), F^{\mathbf{u}+h\mathbf{w}}(\mathbf{x})\Big)\, d\mathbf{x} - \int_\Omega \mathcal{M}(R(\mathbf{x}), F^{\mathbf{u}}(\mathbf{x}))\, d\mathbf{x}\right] \\
&= \lim_{h\to 0} \frac{1}{h}\int_\Omega \int_0^1 \frac{d}{ds}\mathcal{M}\Big(R(\mathbf{x}), F^{\mathbf{u}+hs\mathbf{w}}(\mathbf{x})\Big)\, ds\, d\mathbf{x} \\
&= \lim_{h\to 0} \frac{1}{h}\int_\Omega \int_0^1 \frac{\partial}{\partial F}\mathcal{M}\Big(R(\mathbf{x}), F^{\mathbf{u}+hs\mathbf{w}}(\mathbf{x})\Big) \cdot \frac{d}{ds}F^{\mathbf{u}+hs\mathbf{w}}(\mathbf{x})\, ds\, d\mathbf{x} \\
&= \lim_{h\to 0} \frac{1}{h}\int_\Omega \int_0^1 \frac{\partial}{\partial F}\mathcal{M}\Big(R(\mathbf{x}), F^{\mathbf{u}+hs\mathbf{w}}(\mathbf{x})\Big) \Big\langle \nabla F^{\mathbf{u}+hs\mathbf{w}}(\mathbf{x}), -h\mathbf{w}\Big\rangle\, ds\, d\mathbf{x} \\
&= -\int_\Omega \int_0^1 \frac{\partial}{\partial F}\mathcal{M}(R(\mathbf{x}), F^{\mathbf{u}}(\mathbf{x}))\, \langle \nabla F^{\mathbf{u}}(\mathbf{x}), \mathbf{w}\rangle\, ds\, d\mathbf{x} \\
&= -\int_\Omega \frac{\partial}{\partial F}\mathcal{M}(R(\mathbf{x}), F^{\mathbf{u}}(\mathbf{x}))\, \langle \nabla F^{\mathbf{u}}(\mathbf{x}), \mathbf{w}\rangle\, d\mathbf{x} \\
&= \int_\Omega \langle \mathcal{P}(\mathbf{x}; R, F^{\mathbf{u}}), \mathbf{w}(\mathbf{x})\rangle\, d\mathbf{x},
\end{aligned}$$

where

$$\mathcal{P}(\mathbf{x}; R, F^{\mathbf{u}}) = -\left[\frac{\partial}{\partial F}\mathcal{M}(R(\mathbf{x}), F^{\mathbf{u}}(\mathbf{x}))\right] \nabla F^{\mathbf{u}}(\mathbf{x}). \tag{B.3}$$

□

In order to prove Propositions 4–8, we first introduce and prove the following lemmas:

Lemma 3. *The Gâteaux derivative of the probability density function* $p_{F^{\mathbf{u}}}(f)$ *with respect to* $\mathbf{w}$ *is given by:*

$$dp_{F^{\mathbf{u}}}(f; \mathbf{w}) = -\frac{1}{|\Omega|}\int_\Omega \Big\langle \omega'_\sigma(F^{\mathbf{u}}(\mathbf{x}) - f)\, \nabla F^{\mathbf{u}}(\mathbf{x}), \mathbf{w}(\mathbf{x})\Big\rangle\, d\mathbf{x}, \tag{B.4}$$

where $\omega'_\sigma(t)$ *is the derivative of the Parzen window function* $\omega_\sigma(t)$*.*

Lemma 4. *The Gâteaux derivative of the joint probability density function* $p_{R,F^{\mathbf{u}}}(r, f)$ *with respect to* $\mathbf{w}$ *is given by:*

$$dp_{R,F^{\mathbf{u}}}(r, f; \mathbf{w}) = -\frac{1}{|\Omega|}\int_\Omega \Big\langle \omega_\sigma(R(\mathbf{x}) - r)\, \omega'_\sigma(F^{\mathbf{u}}(\mathbf{x}) - f)\, \nabla F^{\mathbf{u}}(\mathbf{x}), \mathbf{w}(\mathbf{x})\Big\rangle\, d\mathbf{x}. \tag{B.5}$$

Proof of Lemma 3. The probability density function of $F^{\mathbf{u}}$ can be written according to (2.83) as:

$$p_{F^{\mathbf{u}}}(f) = \frac{1}{|\Omega|}\int_{\Omega} \omega_\sigma(F^{\mathbf{u}}(\mathbf{x}) - f)\, d\mathbf{x}. \tag{B.6}$$

The Taylor series expansion of $F^{\mathbf{u}+h\mathbf{w}}(\mathbf{x})$ about $F^{\mathbf{u}}(\mathbf{x})$ is given by:

$$F^{\mathbf{u}+h\mathbf{w}}(\mathbf{x}) = F^{\mathbf{u}}(\mathbf{x}) - h\left\langle \nabla F^{\mathbf{u}}(\mathbf{x}), \mathbf{w}(\mathbf{x})\right\rangle + O\big(h^2\big). \tag{B.7}$$

Substituting (B.7) into (B.6) yields:

$$p_{F^{\mathbf{u}+h\mathbf{w}}}(f) = \frac{1}{|\Omega|}\int_{\Omega} \omega_\sigma\Big(F^{\mathbf{u}+h\mathbf{w}}(\mathbf{x}) - f\Big)\, d\mathbf{x}. \tag{B.8}$$

By another Taylor series expansion, we have:

$$p_{F^{\mathbf{u}+h\mathbf{w}}}(f) = \frac{1}{|\Omega|}\int_{\Omega} \Big[\omega_\sigma(F^{\mathbf{u}}(\mathbf{x}) - f) - h\omega_\sigma^{'}(F^{\mathbf{u}}(\mathbf{x}) - f)\left\langle \nabla F^{\mathbf{u}}(\mathbf{x}), \mathbf{w}(\mathbf{x})\right\rangle + O\big(h^2\big)\Big]\, d\mathbf{x}, \tag{B.9}$$

or

$$\frac{p_{F^{\mathbf{u}+h\mathbf{w}}}(f) - p_{F^{\mathbf{u}}}(f)}{h} = -\frac{1}{|\Omega|}\int_{\Omega} \Big\langle \omega_\sigma^{'}(F^{\mathbf{u}}(\mathbf{x}) - f)\, \nabla F^{\mathbf{u}}(\mathbf{x}), \mathbf{w}(\mathbf{x})\Big\rangle\, d\mathbf{x} + O(h). \tag{B.10}$$

Taking the limit of (B.10) as $h \to 0$ yields the desired result. □

Proof of Lemma 4. The joint density function $p_{R,F^{\mathbf{u}}}$ can be written in terms of (2.84) as:

$$p_{R,F^{\mathbf{u}}}(r, f) = \frac{1}{|\Omega|}\int_{\Omega} \omega_\sigma(R(\mathbf{x}) - r)\, \omega_\sigma(F^{\mathbf{u}}(\mathbf{x}) - f)\, d\mathbf{x}. \tag{B.11}$$

Following the proof of Lemma 3 with (B.11) in place of (B.6) yields the desired result. □

Proof of Proposition 4. By linearity and an application of the chain rule, we find:

$$dH(F^{\mathbf{u}}; \mathbf{w}) = -\int_{-\infty}^{\infty} [1 + \log p_{F^{\mathbf{u}}}(f)]\, dp_{F^{\mathbf{u}}}(f; \mathbf{w})\, df. \tag{B.12}$$

Substituting (B.4) into (B.12) and changing the order of integration yields:

$$dH(F^{\mathbf{u}}; \mathbf{w}) = \int_{\Omega} \left\langle \mathcal{P}_H^g(\mathbf{x}; F^{\mathbf{u}}), \mathbf{w}(\mathbf{x})\right\rangle d\mathbf{x}, \tag{B.13}$$

where

$$\mathcal{P}_H^g(\mathbf{x}; F^{\mathbf{u}}) = \frac{1}{|\Omega|}\left(\int_{-\infty}^{\infty} \omega_\sigma^{'}(F^{\mathbf{u}}(\mathbf{x}) - f)\, [1 + \log p_{F^{\mathbf{u}}}(f)]\, df\right) \nabla F^{\mathbf{u}}(\mathbf{x}). \tag{B.14}$$

Integrating by parts allows us to write (B.14) as:

$$\mathcal{P}_H^g(\mathbf{x}; F^{\mathbf{u}}) = \frac{1}{|\Omega|} \left(\int_{-\infty}^{\infty} \omega_\sigma(F^{\mathbf{u}}(\mathbf{x}) - f) \left[\frac{p'_{F^{\mathbf{u}}}(f)}{p_{F^{\mathbf{u}}}(f)} \right] df \right) \nabla F^{\mathbf{u}}(\mathbf{x}) \,. \tag{B.15}$$

Assuming $\sigma \to 0$ so that $\omega_\sigma(t) \to \delta(t)$ allows us to apply the sifting property of the Dirac delta function to yield:

$$\mathcal{P}_H^g(\mathbf{x}; F^{\mathbf{u}}) \approx \frac{1}{|\Omega|} \left(\frac{p'_{F^{\mathbf{u}}}(F^{\mathbf{u}}(\mathbf{x}))}{p_{F^{\mathbf{u}}}(F^{\mathbf{u}}(\mathbf{x}))} \right) \nabla F^{\mathbf{u}}(\mathbf{x}) \,, \tag{B.16}$$

which is the desired result. □

Proof of Proposition 5. By linearity and an application of the chain rule, we find:

$$dH(R, F^{\mathbf{u}}; \mathbf{w}) = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [1 + \log p_{R,F^{\mathbf{u}}}(r, f)] \, dp_{R,F^{\mathbf{u}}}(r, f; \mathbf{w}) \, dr df. \tag{B.17}$$

Substituting (B.5) into (B.17) and changing the order of integration yields:

$$dH(R, F^{\mathbf{u}}; \mathbf{w}) = \int_\Omega \left\langle \mathcal{P}_H^g(\mathbf{x}; R, F^{\mathbf{u}}) \, , \mathbf{w}(\mathbf{x}) \right\rangle d\mathbf{x}, \tag{B.18}$$

where

$$\begin{aligned} &\mathcal{P}_H^g(\mathbf{x}; R, F^{\mathbf{u}}) = \\ &\quad \tfrac{1}{|\Omega|} \left(\textstyle\int_{-\infty}^{\infty} \omega_\sigma(R(\mathbf{x}) - r) \int_{-\infty}^{\infty} \omega'_\sigma(F^{\mathbf{u}}(\mathbf{x}) - f) \, [1 + \log p_{R,F^{\mathbf{u}}}(r, f)] \, df dr \right) \nabla F^{\mathbf{u}}(\mathbf{x}) \,. \end{aligned} \tag{B.19}$$

Assuming $\sigma \to 0$ so that $\omega_\sigma(t) \to \delta(t)$ allows us to apply the sifting property of the Dirac delta function to yield:

$$\mathcal{P}_H^g(\mathbf{x}; R, F^{\mathbf{u}}) \approx \frac{1}{|\Omega|} \left(\int_{-\infty}^{\infty} \omega'_\sigma(F^{\mathbf{u}}(\mathbf{x}) - f) \, [1 + \log p_{R,F^{\mathbf{u}}}(R(\mathbf{x}) \, , f)] \, df \right) \nabla F^{\mathbf{u}}(\mathbf{x}) \,. \tag{B.20}$$

Integrating by parts allows us to write (B.20) as:

$$\mathcal{P}_H^g(\mathbf{x}; R, F^{\mathbf{u}}) \approx \frac{1}{|\Omega|} \left(\int_{-\infty}^{\infty} \omega_\sigma(F^{\mathbf{u}}(\mathbf{x}) - f) \left(\frac{\frac{\partial}{\partial f} p_{R,F^{\mathbf{u}}}(R(\mathbf{x}) \, , f)}{p_{R,F^{\mathbf{u}}}(R(\mathbf{x}) \, , f)} \right) df \right) \nabla F^{\mathbf{u}}(\mathbf{x}) \,. \tag{B.21}$$

Applying the sifting property a second time yields:

$$\mathcal{P}_H^g(\mathbf{x}; F^{\mathbf{u}}) \approx \frac{1}{|\Omega|} \left(\frac{\frac{\partial}{\partial f} \left[p_{R,F^{\mathbf{u}}}(R(\mathbf{x}) \, , f) \right]_{f = F^{\mathbf{u}}(\mathbf{x})}}{p_{R,F^{\mathbf{u}}}(R(\mathbf{x}) \, , F^{\mathbf{u}}(\mathbf{x}))} \right) \nabla F^{\mathbf{u}}(\mathbf{x}) \,, \tag{B.22}$$

which is the desired result. □

Proof of Proposition 6. By linearity and an application of the chain rule, we find:

$$d\varepsilon(F^{\mathbf{u}};\mathbf{w}) = -\int_{-\infty}^{\infty} \left[1+\log q_{F^{\mathbf{u}}}(f)\right] dq_{F^{\mathbf{u}}}(f;\mathbf{w})\, df, \tag{B.23}$$

where

$$q_{F^{\mathbf{u}}}(f) = P(F^{\mathbf{u}} > f) = \int_{f}^{\infty} p_{F^{\mathbf{u}}}(t)\, dt. \tag{B.24}$$

Taking the Gâteaux derivative of (B.24), substituting in (B.4), and changing the order of integration yields:

$$dq_{F^{\mathbf{u}}}(f;v) = -\frac{1}{|\Omega|}\int_{\Omega}\left\langle\left(\int_{f}^{\infty}\omega_{\sigma}^{'}(F^{\mathbf{u}}(\mathbf{x})-t)\, dt\right)\nabla F^{\mathbf{u}}(\mathbf{x})\,,\mathbf{w}(\mathbf{x})\right\rangle d\mathbf{x}. \tag{B.25}$$

After evaluating the inner integral, (B.25) reduces to

$$dq_{F^{\mathbf{u}}}(f;v) = -\frac{1}{|\Omega|}\int_{\Omega}\left\langle\omega_{\sigma}(F^{\mathbf{u}}(\mathbf{x})-t)\,\nabla F^{\mathbf{u}}(\mathbf{x})\,,\mathbf{w}(\mathbf{x})\right\rangle d\mathbf{x}. \tag{B.26}$$

Substituting (B.26) into (B.23) and changing the order of integration yields:

$$d\varepsilon(F^{\mathbf{u}};v) = \int_{\Omega}\left\langle\mathcal{P}_{\varepsilon}^{g}(\mathbf{x};F^{\mathbf{u}})\,,\mathbf{w}(\mathbf{x})\right\rangle d\mathbf{x}, \tag{B.27}$$

where

$$\mathcal{P}_{\varepsilon}^{g}(\mathbf{x};F^{\mathbf{u}}) = \frac{1}{|\Omega|}\left(\int_{-\infty}^{\infty}\omega_{\sigma}(F^{\mathbf{u}}(\mathbf{x})-f)\left[1+\log q_{F^{\mathbf{u}}}(f)\right]df\right)\nabla F^{\mathbf{u}}(\mathbf{x})\,. \tag{B.28}$$

Assuming $\sigma \to 0$ allows us to apply the sifting property to (B.28) to yield:

$$\mathcal{P}_{\varepsilon}^{g}(\mathbf{x};F^{\mathbf{u}}) \approx \frac{1}{|\Omega|}\left(1+\log q_{F^{\mathbf{u}}}(F^{\mathbf{u}}(\mathbf{x}))\right)\nabla F^{\mathbf{u}}(\mathbf{x})\,, \tag{B.29}$$

which is the desired result. □

Proof of Proposition 7. By linearity and an application of the chain rule, we find:

$$d\varepsilon(R,F^{\mathbf{u}};\mathbf{w}) = -\int_{-\infty}^{\infty} p_R(r)\int_{-\infty}^{\infty}\left[1+\log q_{F^{\mathbf{u}}|R}(f|r)\right] dq_{F^{\mathbf{u}}|R}(f|r;\mathbf{w})\, df dr, \tag{B.30}$$

where

$$q_{F^{\mathbf{u}}|R}(f|r) = P(F^{\mathbf{u}} > f|R=r) = \frac{\int_{f}^{\infty} p_{R,F^{\mathbf{u}}}(r,t)\, dt}{p_R(r)}. \tag{B.31}$$

Taking the Gâteaux derivative of (B.31), substituting in (B.5), and changing the order of integration yields:

$$dq_{F^{\mathbf{u}}|R}(f|r;v) = -\frac{1}{|\Omega|\, p_R(r)} \int_\Omega \left\langle \omega_\sigma(R(\mathbf{x}) - r) \left(\int_f^\infty \omega_\sigma'(F^{\mathbf{u}}(\mathbf{x}) - t)\, dt \right) \nabla F^{\mathbf{u}}(\mathbf{x}) \,, \mathbf{w}(\mathbf{x}) \right\rangle d\mathbf{x}. \tag{B.32}$$

After evaluating the inner integral, (B.32) reduces to

$$dq_{F^{\mathbf{u}}|R}(f|r;v) = -\frac{1}{|\Omega|\, p_R(r)} \int_\Omega \left\langle \omega_\sigma(R(\mathbf{x}) - r)\, \omega_\sigma(F^{\mathbf{u}}(\mathbf{x}) - f)\, \nabla F^{\mathbf{u}}(\mathbf{x}) \,, \mathbf{w}(\mathbf{x}) \right\rangle d\mathbf{x}. \tag{B.33}$$

Substituting (B.33) into (B.30) and changing the order of integration yields:

$$d\varepsilon(R, F^{\mathbf{u}}; v) = \int_\Omega \left\langle \mathcal{P}_\varepsilon^g(\mathbf{x}; R, F^{\mathbf{u}}) \,, \mathbf{w}(\mathbf{x}) \right\rangle d\mathbf{x}, \tag{B.34}$$

where

$$\mathcal{P}_\varepsilon^g(\mathbf{x}; R, F^{\mathbf{u}}) = \\ \frac{1}{|\Omega|} \left(\int_{-\infty}^\infty \omega_\sigma(R(\mathbf{x}) - r) \int_{-\infty}^\infty \omega_\sigma(F^{\mathbf{u}}(\mathbf{x}) - f) \left[1 + \log q_{F^{\mathbf{u}}|R}(f|r) \right] df dr \right) \nabla F^{\mathbf{u}}(\mathbf{x}) \,. \tag{B.35}$$

Assuming $\sigma \to 0$ and applying the sifting property twice to (B.35) yields:

$$\mathcal{P}_\varepsilon^g(\mathbf{x}; R, F^{\mathbf{u}}) \approx \frac{1}{|\Omega|} \left(1 + \log q_{F^{\mathbf{u}}|R}(F^{\mathbf{u}}(\mathbf{x}) \,|R(\mathbf{x})) \right) \nabla F^{\mathbf{u}}(\mathbf{x}) \,, \tag{B.36}$$

which is the desired result. □

Proof of Proposition 8. By linearity and an application of the chain rule, we find:

$$d\varepsilon(F^{\mathbf{u}}, R; \mathbf{w}) = d\varepsilon(F^{\mathbf{u}}; \mathbf{w}) - \int_{-\infty}^\infty \int_{-\infty}^\infty p_{F^{\mathbf{u}}}(f) \left[1 + \log q_{R|F^{\mathbf{u}}}(r|f) \right] dq_{R|F^{\mathbf{u}}}(r|f; \mathbf{w})\, dr df \\ - \int_{-\infty}^\infty \int_{-\infty}^\infty dp_{F^{\mathbf{u}}}(f; v)\, q_{R|F^{\mathbf{u}}}(r|f) \log q_{R|F^{\mathbf{u}}}(r|f)\, dr df \tag{B.37}$$

where

$$q_{R|F^{\mathbf{u}}}(r|f) = P(R > r | F^{\mathbf{u}} = f) = \frac{\int_r^\infty p_{R,F^{\mathbf{u}}}(t, f)\, dt}{p_{F^{\mathbf{u}}}(f)}. \tag{B.38}$$

Taking the Gâteaux derivative of (B.38) yields:

$$dq_{R|F^{\mathbf{u}}}(r|f; v) = \frac{\int_r^\infty dp_{R,F^{\mathbf{u}}}(t, f; \mathbf{w})\, dt}{p_{F^{\mathbf{u}}}(f)} - \frac{q_{R|F^{\mathbf{u}}}(r|f)\, dp_{F^{\mathbf{u}}}(f; v)}{p_{F^{\mathbf{u}}}(f)}. \tag{B.39}$$

Substituting (B.39) into (B.37) and simplifying yields:

$$d\varepsilon(F^{\mathbf{u}}, R; \mathbf{w}) = d\varepsilon(F^{\mathbf{u}}; \mathbf{w}) + \int_{-\infty}^\infty \int_{-\infty}^\infty q_{R|F^{\mathbf{u}}}(r|f)\, dp_{F^{\mathbf{u}}}(f; v)\, dr df \\ - \int_{-\infty}^\infty \int_{-\infty}^\infty \int_r^\infty \left[1 + \log q_{R|F^{\mathbf{u}}}(r|f) \right] dp_{R,F^{\mathbf{u}}}(t, f; v)\, dt dr df \tag{B.40}$$

Substituting in (B.5) and (B.4), and changing the order of integration yields:

$$d\varepsilon(F^{\mathbf{u}}, R; \mathbf{w}) = d\varepsilon(F^{\mathbf{u}}; \mathbf{w}) + \frac{1}{|\Omega|} \int_{\Omega} \langle \mathcal{K}_{\varepsilon}^{g}(\mathbf{x}; F^{\mathbf{u}}, R)\, \nabla F^{\mathbf{u}}(\mathbf{x})\,, \mathbf{w}(\mathbf{x}) \rangle\, d\mathbf{x}, \tag{B.41}$$

where

$$\begin{aligned} &\mathcal{K}_{\varepsilon}^{g}(\mathbf{x}; F^{\mathbf{u}}, R) = \\ &\quad \textstyle\int_{-\infty}^{\infty} \omega_{\sigma}^{'}(F^{\mathbf{u}}(\mathbf{x})) \int_{-\infty}^{\infty} \left(\left[1 + \log q_{R|F^{\mathbf{u}}}(r|f) \right] \left(\int_{r}^{\infty} \omega_{\sigma}(R(\mathbf{x}) - t)\, dt \right) - q_{R|F^{\mathbf{u}}}(r|f) \right) dr df \end{aligned} \tag{B.42}$$

Noting that $\int_{r}^{\infty} \omega_{\sigma}(R(\mathbf{x}) - t)\, dt \approx \mathbf{1}_{r \leq R(\mathbf{x})}$ if $\sigma \to 0$, we see that (B.42) simplifies to:

$$\begin{aligned} \mathcal{K}_{\varepsilon}^{g}(\mathbf{x}; F^{\mathbf{u}}, R) \quad \approx \quad & \int_{-\infty}^{\infty} \omega_{\sigma}^{'}(F^{\mathbf{u}}(\mathbf{x})) \int_{-\infty}^{R(\mathbf{x})} \left(1 - q_{R|F^{\mathbf{u}}}(r|f) + \log q_{R|F^{\mathbf{u}}}(r|f) \right) dr df \\ & - \int_{-\infty}^{\infty} \omega_{\sigma}^{'}(F^{\mathbf{u}}(\mathbf{x})) \int_{R(\mathbf{x})}^{\infty} q_{R|F^{\mathbf{u}}}(r|f)\, dr df \end{aligned} \tag{B.43}$$

Noting that $1 - q_{R|F^{\mathbf{u}}}(r|f) = P(R < r | F^{\mathbf{u}} = f)$, integrating by parts, and applying the sifting property yields the desired result. □

Appendix C

Gâteaux Derivatives Under Changing Overlap

This appendix relates to Section 3.1 of Chapter 3, in which we present Gâteaux derivatives for the dissimilarity measures after accounting for the changing overlap region. It proves the following two propositions (both new results):

Proposition 9. *The Gâteaux derivative of* $\mathrm{C}(\mathbf{u}) := |\Omega|$ *can be approximated by:*

$$d\mathrm{C}(\mathbf{u};\mathbf{w}) \approx -\int_{\mathbb{R}^n} \langle \mathrm{H}_\epsilon(\Theta_R(\mathbf{x})) \cdot \delta_\epsilon(\Theta_{F^\mathbf{u}}(\mathbf{x})) \cdot \nabla\Theta_{F^\mathbf{u}}(\mathbf{x}) \,, \mathbf{w}(\mathbf{x})\rangle \, d\mathbf{x}. \tag{C.1}$$

Proposition 10. *The Gâteaux derivative of* $\mathcal{I}^g_{\Psi,r}(R, F^\mathbf{u}) := \mathrm{M}_{\Psi,r}(R, F^\mathbf{u}) / \mathrm{C}(\mathbf{u})$*, where* $\mathrm{M}_{\Psi,r}(R, F^\mathbf{u})$ *is given by (3.9) and* $\mathrm{C}(\mathbf{u})$ *is given by (3.1) can be approximated by:*

$$d\mathcal{I}^g_{\Psi,r}(R, F^\mathbf{u};\mathbf{w}) \approx \int_{\mathbb{R}^n} \langle \mathcal{P}_{\Psi,r}(\mathbf{x}; R, F^\mathbf{u}) \,, \mathbf{w}(\mathbf{x})\rangle \, d\mathbf{x}, \tag{C.2}$$

where

$$\begin{aligned}\mathcal{P}_{\Psi,r}(\mathbf{x}; R, F^\mathbf{u}) = &-\mathrm{H}_\epsilon(\Theta_R(\mathbf{x})) \cdot \delta_\epsilon(\Theta_{F^\mathbf{u}}(\mathbf{x})) \cdot \left(\frac{\Psi(r(\mathbf{x}; R, F^\mathbf{u})) - \mathcal{I}^g_{\Psi,r}(R, F^\mathbf{u})}{\mathrm{C}(\mathbf{u})} \right) \cdot \nabla\Theta_{F^\mathbf{u}}(\mathbf{x}) \\ &- \mathrm{H}_\epsilon(\Theta_R(\mathbf{x})) \cdot \mathrm{H}_\epsilon(\Theta_{F^\mathbf{u}}(\mathbf{x})) \cdot \left(\frac{\frac{\partial}{\partial F}\Psi(r(\mathbf{x}; R, F^\mathbf{u}))}{\mathrm{C}(\mathbf{u})} \right) \cdot \nabla F^\mathbf{u}(\mathbf{x}) \,. \end{aligned} \tag{C.3}$$

Proof of Proposition 9. The region content $\mathrm{C}(\mathbf{u})$ can be approximated by:

$$\mathrm{C}(\mathbf{u}) \approx \int_{\mathbb{R}^n} \mathrm{H}_\epsilon(\Theta_R(\mathbf{x})) \cdot \mathrm{H}_\epsilon(\Theta_{F^\mathbf{u}}(\mathbf{x})) \, d\mathbf{x}, \tag{C.4}$$

where H_ϵ is the regularized Heaviside function (3.4). By the definition of the Gâteaux derivative, we then have:

$$d\mathrm{C}(\mathbf{u};\mathbf{w}) \approx \lim_{h\to 0} \frac{1}{h} \int_{\mathbb{R}^n} \mathrm{H}_\epsilon(\Theta_R(\mathbf{x})) \cdot [\mathrm{H}_\epsilon(\Theta_{F^\mathbf{u}+h\mathbf{w}}(\mathbf{x})) - \mathrm{H}_\epsilon(\Theta_{F^\mathbf{u}}(\mathbf{x}))] \, d\mathbf{x}. \tag{C.5}$$

By a Taylor series approximation, we see that:

$$\Theta_{F^{\mathbf{u}+h\mathbf{w}}}(\mathbf{x}) \approx \Theta_{F^{\mathbf{u}}}(\mathbf{x}) - h\left\langle \nabla\Theta_{F^{\mathbf{u}}}(\mathbf{x}), \mathbf{w}(\mathbf{x})\right\rangle + O\big(h^2\big), \tag{C.6}$$

and therefore:

$$\mathrm{H}_\epsilon(\Theta_{F^{\mathbf{u}+h\mathbf{w}}}(\mathbf{x})) \approx \mathrm{H}_\epsilon\big(\Theta_{F^{\mathbf{u}}}(\mathbf{x}) - h\left\langle \nabla\Theta_{F^{\mathbf{u}}}(\mathbf{x}), \mathbf{w}(\mathbf{x})\right\rangle + O\big(h^2\big)\big). \tag{C.7}$$

Another Taylor series approximation yields:

$$\begin{aligned}\mathrm{H}_\epsilon(\Theta_{F^{\mathbf{u}+h\mathbf{w}}}(\mathbf{x})) &\approx \mathrm{H}_\epsilon(\Theta_{F^{\mathbf{u}}}(\mathbf{x})) - h\left[\frac{\partial}{\partial\Theta_{F^{\mathbf{u}}}}\mathrm{H}_\epsilon(\Theta_{F^{\mathbf{u}}}(\mathbf{x}))\right]\left\langle \nabla\Theta_{F^{\mathbf{u}}}(\mathbf{x}), \mathbf{w}(\mathbf{x})\right\rangle + O\big(h^2\big) &\text{(C.8)}\\ &= \mathrm{H}_\epsilon(\Theta_{F^{\mathbf{u}}}(\mathbf{x})) - h\delta_\epsilon(\Theta_{F^{\mathbf{u}}}(\mathbf{x}))\left\langle \nabla\Theta_{F^{\mathbf{u}}}(\mathbf{x}), \mathbf{w}(\mathbf{x})\right\rangle + O\big(h^2\big), &\text{(C.9)}\end{aligned}$$

where $\delta_\epsilon(\Theta)$ is given by (3.7). Hence, (C.5) can be approximated by:

$$d\mathrm{C}(\mathbf{u};\mathbf{w}) \approx \lim_{h\to 0}\left[\int_{\mathbb{R}^n} \mathrm{H}_\epsilon(\Theta_R(\mathbf{x}))\cdot\left[-\delta_\epsilon(\Theta_{F^{\mathbf{u}}}(\mathbf{x}))\left\langle \nabla\Theta_{F^{\mathbf{u}}}(\mathbf{x}), \mathbf{w}(\mathbf{x})\right\rangle\right] d\mathbf{x} + O(h)\right]. \tag{C.10}$$

Taking the limit and rearranging yields the desired result. □

Proof of Proposition 10. By the quotient rule, we see that the Gâteaux derivative of $\mathcal{I}^g_{\Psi,r}$ can be written in terms of the Gâteaux derivatives of $\mathrm{M}_{\Psi,r}$ and C in the following way:

$$\begin{aligned}d\mathcal{I}^g_{\Psi,r}(R, F^{\mathbf{u}};\mathbf{w}) &= \frac{d\mathrm{M}_{\Psi,r}(R, F^{\mathbf{u}};\mathbf{w})}{\mathrm{C}(\mathbf{u})} - \frac{\mathrm{M}_{\Psi,r}(R, F^{\mathbf{u}})\, d\mathrm{C}(\mathbf{u};\mathbf{w})}{\mathrm{C}(\mathbf{u})^2}\\ &= \frac{d\mathrm{M}_{\Psi,r}(R, F^{\mathbf{u}};\mathbf{w}) - \mathcal{I}^g_{\Psi,r}(R, F^{\mathbf{u}})\, d\mathrm{C}(\mathbf{u};\mathbf{w})}{\mathrm{C}(\mathbf{u})}.\end{aligned} \tag{C.11}$$

First, we approximate $\mathrm{M}_{\Psi,r}$ using the regularized Heaviside function; i.e.,

$$\mathrm{M}_{\Psi,r}(R, F^{\mathbf{u}}) \approx \int_{\mathbb{R}^n} \mathrm{H}_\epsilon(\Theta_R(\mathbf{x}))\cdot\mathrm{H}_\epsilon(\Theta_{F^{\mathbf{u}}}(\mathbf{x}))\cdot\Psi(r(\mathbf{x}; R, F^{\mathbf{u}}))\, d\mathbf{x} \tag{C.12}$$

Applying the definition of the Gâteaux derivative to (C.12) yields:

$$d\mathrm{M}_{\Psi,r}(R, F^{\mathbf{u}};\mathbf{w}) \approx \lim_{h\to 0}\frac{1}{h}\left[\begin{array}{c}\int_{\mathbb{R}^n} \mathrm{H}_\epsilon(\Theta_R(\mathbf{x}))\cdot\mathrm{H}_\epsilon(\Theta_{F^{\mathbf{u}+h\mathbf{w}}}(\mathbf{x}))\cdot\Psi\big(r\big(\mathbf{x}; R, F^{\mathbf{u}+h\mathbf{w}}\big)\big)\, d\mathbf{x}\\ -\int_{\mathbb{R}^n} \mathrm{H}_\epsilon(\Theta_R(\mathbf{x}))\cdot\mathrm{H}_\epsilon(\Theta_{F^{\mathbf{u}}}(\mathbf{x}))\cdot\Psi(r(\mathbf{x}; R, F^{\mathbf{u}}))\, d\mathbf{x}\end{array}\right]. \tag{C.13}$$

Using an derivation similar to that in the proof of Proposition 3 in Appendix B, we find that (C.13) can be expressed as:

$$d\mathrm{M}_{\Psi,r}(R, F^{\mathbf{u}};\mathbf{w}) \approx \int_{\mathbb{R}^n}\left\langle \mathcal{P}_{\mathrm{M}}(\mathbf{x}; R, F^{\mathbf{u}}), \mathbf{w}(\mathbf{x})\right\rangle d\mathbf{x}, \tag{C.14}$$

where

$$\begin{aligned}\mathcal{P}_{\mathrm{M}}(\mathbf{x}; R, F^{\mathbf{u}}) = &- \mathrm{H}_{\epsilon}(\Theta_R(\mathbf{x})) \cdot \delta_{\epsilon}(\Theta_{F^{\mathbf{u}}}(\mathbf{x})) \cdot \Psi(r(\mathbf{x}; R, F^{\mathbf{u}})) \cdot \nabla\Theta_{F^{\mathbf{u}}}(\mathbf{x}) \\ &- \mathrm{H}_{\epsilon}(\Theta_R(\mathbf{x})) \cdot \mathrm{H}_{\epsilon}(\Theta_{F^{\mathbf{u}}}(\mathbf{x})) \cdot \left[\frac{\partial}{\partial F}\Psi(r(\mathbf{x}; R, F^{\mathbf{u}}))\right] \cdot \nabla F^{\mathbf{u}}(\mathbf{x})\,.\end{aligned} \tag{C.15}$$

Substituting (C.1) and (C.14)–(C.15) into (C.11) and simplifying yields the desired result.

□

Appendix D

Gâteaux Derivatives of Regularizers

This appendix relates to Section 4.4 of Chapter 4, in which we present Gâteaux derivatives of the various regularizers. It proves the following propositions:

Proposition 11. *The Gâteaux derivative of the homogeneous second-order elastic regularizer has partial differential operator given by:*

$$\mathcal{A}(\mathbf{u}) = \mu\Delta^2\mathbf{u} + (\lambda + 2\mu)\Delta\nabla\mathrm{div}\,\mathbf{u} \tag{D.1}$$

with boundary conditions:

$$\mathcal{B}_1[\mathbf{u}] = \lambda(\mathrm{div}\,\mathbf{u})\vec{\mathbf{n}} + \mu\langle\nabla\mathbf{u}, \vec{\mathbf{n}}\rangle + 2\mu\langle(\nabla\mathbf{u})^{\mathrm{T}}, \vec{\mathbf{n}}\rangle \tag{D.2}$$

$$\mathcal{B}_2[\mathbf{u}] = \lambda(\Delta\mathrm{div}\,\mathbf{u})\vec{\mathbf{n}} + \mu\langle\Delta\nabla\mathbf{u}, \vec{\mathbf{n}}\rangle + 2\mu\langle\Delta\,(\nabla\mathbf{u})^{\mathrm{T}}, \vec{\mathbf{n}}\rangle \tag{D.3}$$

Proposition 12. *The Gâteaux derivative of the image-based isotropic version of the diffusion regularizer has partial differential operator given by:*

$$\mathcal{A}(\mathbf{u}) = -\mathrm{div}\,(\beta(\mathbf{x})\,\nabla\mathbf{u}) \tag{D.4}$$

with boundary conditions:

$$\mathcal{B}[\mathbf{u}] = \langle\beta(\mathbf{x})\,\nabla\mathbf{u}, \vec{\mathbf{n}}\rangle \tag{D.5}$$

Proof of Proposition 11. First, we expand and simplify the second-order elastic regularizer

in the following manner:

$$\begin{aligned}\mathcal{R}(\mathbf{u}) &= \frac{1}{2}\int_\Omega \left(\frac{\mu}{3}\sum_{j,k,l=1}^{n}\left(\partial^2_{x_jx_k}u_l+\partial^2_{x_jx_l}u_k+\partial^2_{x_kx_l}u_j\right)^2+\lambda\left\|\nabla\mathrm{div}\,\mathbf{u}\right\|^2\right)d\mathbf{x}\\ &= \frac{1}{2}\int_\Omega \mu\sum_{j,k,l=1}^{n}\left(\left(\partial^2_{x_jx_k}u_l\right)^2+2\left(\partial^2_{x_jx_k}u_l\right)\left(\partial^2_{x_jx_l}u_k\right)\right)d\mathbf{x}\\ &\quad+\frac{1}{2}\int_\Omega \lambda\left\|\nabla\mathrm{div}\,\mathbf{u}\right\|^2 d\mathbf{x}. \end{aligned} \tag{D.6}$$

Applying the definition of the Gâteaux derivative and taking the limit yields:

$$\begin{aligned}d\mathcal{R}(\mathbf{u};\mathbf{w}) &= \int_\Omega \mu\sum_{j,k,l=1}^{n}\left(\left(\partial^2_{x_jx_k}u_l\right)\left(\partial^2_{x_jx_k}w_l\right)+2\left(\partial^2_{x_jx_k}u_l\right)\left(\partial^2_{x_jx_l}w_k\right)\right)d\mathbf{x}\\ &\quad+\int_\Omega \lambda\left\langle\nabla\mathrm{div}\,\mathbf{u},\nabla\mathrm{div}\,\mathbf{w}\right\rangle d\mathbf{x}. \end{aligned} \tag{D.7}$$

Integrating by parts yields:

$$\begin{aligned}d\mathcal{R}(\mathbf{u};\mathbf{w}) &= \int_{\partial\Omega} \mu\sum_{j,k,l=1}^{n}\left(\left(\partial^2_{x_jx_k}u_l\right)(\partial_{x_k}w_l)\,\vec{n}_j+2\left(\partial^2_{x_jx_k}u_l\right)(\partial_{x_l}w_k)\,\vec{n}_j\right)d\sigma\\ &\quad+\int_{\partial\Omega}\lambda\left(\mathrm{div}\,\mathbf{w}\right)\left\langle\nabla\mathrm{div}\,\mathbf{u},\vec{\mathbf{n}}\right\rangle d\sigma\\ &\quad-\int_\Omega \mu\sum_{j,k,l=1}^{n}\left(\left(\partial^3_{x_jx_jx_k}u_l\right)(\partial_{x_k}w_l)+2\left(\partial^3_{x_jx_jx_k}u_l\right)(\partial_{x_l}w_k)\right)d\mathbf{x}\\ &\quad-\int_\Omega\lambda\left(\mathrm{div}\,\mathbf{w}\right)\left(\Delta\mathrm{div}\,\mathbf{w}\right)d\mathbf{x}. \end{aligned} \tag{D.8}$$

Integrating a second time by parts yields:

$$\begin{aligned}d\mathcal{R}(\mathbf{u};\mathbf{w}) &= \int_{\partial\Omega} \mu\sum_{j,k,l=1}^{n}\left(\left(\partial^2_{x_jx_k}u_l\right)(\partial_{x_k}w_l)\,\vec{n}_j+2\left(\partial^2_{x_jx_k}u_l\right)(\partial_{x_l}w_k)\,\vec{n}_j\right)d\sigma\\ &\quad+\int_{\partial\Omega}\lambda\left\langle\nabla\mathrm{div}\,\mathbf{u},\left(\mathrm{div}\,\mathbf{w}\right)\vec{\mathbf{n}}\right\rangle d\sigma\\ &\quad-\int_{\partial\Omega} \mu\sum_{j,k,l=1}^{n}\left(\left(\partial^3_{x_jx_jx_k}u_l\right)w_l\vec{n}_k+2\left(\partial^3_{x_jx_jx_k}u_l\right)w_k\vec{n}_l\right)d\sigma\\ &\quad-\int_{\partial\Omega}\lambda\left\langle\left(\Delta\mathrm{div}\,\mathbf{u}\right)\vec{\mathbf{n}},\mathbf{w}\right\rangle d\sigma\\ &\quad+\int_\Omega \mu\sum_{j,k,l=1}^{n}\left(\left(\partial^4_{x_jx_jx_kx_k}u_l\right)w_l+2\left(\partial^3_{x_jx_jx_kx_l}u_l\right)w_k\right)d\mathbf{x}\\ &\quad+\int_\Omega\lambda\left\langle\Delta\nabla\mathrm{div}\,\mathbf{u},\mathbf{w}\right\rangle d\mathbf{x}. \end{aligned} \tag{D.9}$$

Evaluating the summations enables us to write (D.9) as:

$$\begin{aligned} d\mathcal{R}(\mathbf{u};\mathbf{w}) &= \int_{\partial\Omega} \mu \left\langle \nabla \mathrm{div}\, \mathbf{u}, \langle \nabla \mathbf{w}, \vec{\mathbf{n}} \rangle \right\rangle + 2\mu \left\langle \nabla \mathrm{div}\, \mathbf{u}, \langle (\nabla \mathbf{w})^{\mathrm{T}}, \vec{\mathbf{n}} \rangle \right\rangle d\sigma \\ &+ \int_{\partial\Omega} \lambda \left\langle \nabla \mathrm{div}\, \mathbf{u}, (\mathrm{div}\, \mathbf{w})\, \vec{\mathbf{n}} \right\rangle d\sigma \\ &- \int_{\partial\Omega} \mu \left\langle \langle \Delta \nabla \mathbf{u}, \vec{\mathbf{n}} \rangle, \mathbf{w} \right\rangle + 2\mu \left\langle \langle \Delta (\nabla \mathbf{u})^{\mathrm{T}}, \vec{\mathbf{n}} \rangle, \mathbf{w} \right\rangle d\sigma \\ &- \int_{\partial\Omega} \lambda \left\langle (\Delta \mathrm{div}\, \mathbf{u})\, \vec{\mathbf{n}}, \mathbf{w} \right\rangle d\sigma \\ &+ \int_{\Omega} \mu \left(\left\langle \Delta^2 \mathbf{u}, \mathbf{w} \right\rangle + 2 \left\langle \Delta \nabla \mathrm{div}\, \mathbf{u}, \mathbf{w} \right\rangle \right) d\mathbf{x} \\ &+ \int_{\Omega} \lambda \left\langle \Delta \nabla \mathrm{div}\, \mathbf{u}, \mathbf{w} \right\rangle d\mathbf{x}. \end{aligned} \tag{D.10}$$

But (D.10) is simply:

$$d\mathcal{R}(\mathbf{u};\mathbf{w}) = \int_{\partial\Omega} \left\langle \nabla \mathrm{div}\, \mathbf{u}, \mathcal{B}_1[\mathbf{w}] \right\rangle d\sigma - \int_{\partial\Omega} \left\langle \mathcal{B}_2[\mathbf{u}], \mathbf{w} \right\rangle d\sigma + \int_{\Omega} \left\langle \mathcal{A}(\mathbf{u}), \mathbf{w} \right\rangle d\mathbf{x}, \tag{D.11}$$

with $\mathcal{A}$, $\mathcal{B}_1$ and $\mathcal{B}_2$ given by (D.1)–(D.3). □

Proof of Proposition 12. Applying the definition of the Gâteaux derivative to (4.14) and taking the limit yields:

$$d\mathcal{R}(\mathbf{u};\mathbf{w}) = \int_{\Omega} \beta(\mathbf{x}) \left(\sum_{j=1}^{n} \langle \nabla u_j, \nabla w_j \rangle \right) d\mathbf{x}. \tag{D.12}$$

Integrating by parts yields:

$$d\mathcal{R}(\mathbf{u};\mathbf{w}) = \int_{\partial\Omega} \left(\sum_{j=1}^{n} \langle \beta(\mathbf{x})\, \nabla u_j, \vec{\mathbf{n}} \rangle\, w_j \right) d\sigma - \int_{\Omega} \left(\sum_{j=1}^{n} \left\langle \mathrm{div}\, (\beta(\mathbf{x})\, \nabla u_j), \mathbf{w} \right\rangle \right) d\mathbf{x}. \tag{D.13}$$

Evaluating the summations allows us to write (D.13) as:

$$d\mathcal{R}(\mathbf{u};\mathbf{w}) = \int_{\partial\Omega} \left\langle \langle \beta(\mathbf{x})\, \nabla \mathbf{u}, \vec{\mathbf{n}} \rangle, \mathbf{w} \right\rangle d\sigma - \int_{\Omega} \left\langle \mathrm{div}\, (\beta(\mathbf{x})\, \nabla \mathbf{u}), \mathbf{w} \right\rangle d\mathbf{x}. \tag{D.14}$$

But (D.14) is simply

$$d\mathcal{R}(\mathbf{u};\mathbf{w}) = \int_{\partial\Omega} \left\langle \mathcal{B}[\mathbf{u}], \mathbf{w} \right\rangle d\sigma + \int_{\Omega} \left\langle \mathcal{A}(\mathbf{u}), \mathbf{w} \right\rangle d\mathbf{x}, \tag{D.15}$$

with $\mathcal{A}$ and $\mathcal{B}$ given by (D.4)–(D.5). □

Appendix E

Fourier Methods for Elastic and Second-order Elastic Registration

This appendix relates to Section 5.4 of Chapter 5, in which we present linear operators to transform the operators arising from the elastic and second-order elastic regularizers. It proves the following four propositions (all new results):

Proposition 13. *If $\mathcal{L}_1$ is given by (5.51), and if $\mathcal{A}_{elastic}$ and $\mathcal{A}_{curvature}$ are defined as in Table 4.1, then $\mathcal{L}_1 \circ \mathcal{A}_{elastic}$ is a scaled version of $\mathcal{A}_{curvature}$.*

Proposition 14. *If $\mathcal{L}_2$ is given by (5.55), and if $\mathcal{A}_{elastic}$ is defined as in Table 4.1, then the following statement holds:*

$$\mathcal{L}_2 \circ [\mathbf{I} + \tau \mathcal{A}_{elastic}] = \mathbf{I} - \tau (\lambda + 3\mu) \Delta + \tau^2 \mu (\lambda + 2\mu) \Delta^2. \tag{E.1}$$

Proposition 15. *If $\mathcal{L}_3$ is given by (5.58), and if $\mathcal{A}_{elastic\text{-}2}$ is defined as in Table 4.1, then $\mathcal{L}_3 \circ \mathcal{A}_{elastic\text{-}2}$ is a scaled version of $(-\Delta)^3$.*

Proposition 16. *If $\mathcal{L}_4$ is given by (5.62), and if $\mathcal{A}_{elastic\text{-}2}$ is defined as in Table 4.1, then the following statement holds:*

$$\mathcal{L}_4 \circ [\mathbf{I} + \tau \mathcal{A}_{elastic\text{-}2}] = \mathbf{I} + \tau (\lambda + 4\mu) \Delta^2 + \tau^2 \mu (\lambda + 3\mu) \Delta^4. \tag{E.2}$$

Proof of Proposition 13.

$$\begin{aligned}
\mathcal{L}_1 \circ \mathcal{A}_{elastic} &= \left[-\left(\lambda+2\mu\right)\Delta\mathbf{I} + \left(\lambda+\mu\right)\nabla\mathrm{div}\right] \circ \left[-\mu\Delta\mathbf{I} - \left(\lambda+\mu\right)\nabla\mathrm{div}\right] \\
&= \mu\left(\lambda+2\mu\right)\Delta^2\mathbf{I} + \left(\lambda+\mu\right)\left(\lambda+2\mu\right)\Delta\nabla\mathrm{div} \\
&\quad - \mu\left(\lambda+\mu\right)\nabla\mathrm{div}\,\Delta - \left(\lambda+\mu\right)^2\nabla\mathrm{div}\,\nabla\mathrm{div} \\
&= \mu\left(\lambda+2\mu\right)\Delta^2\mathbf{I} + \left[\left(\lambda+\mu\right)\left(\lambda+2\mu\right) - \mu\left(\lambda+\mu\right) - \left(\lambda+\mu\right)^2\right]\Delta\nabla\mathrm{div} \\
&= \mu\left(\lambda+2\mu\right)\Delta^2\mathbf{I} + 0\cdot\Delta\nabla\mathrm{div} \\
&= \mu\left(\lambda+2\mu\right)\mathcal{A}_{curvature}
\end{aligned}$$

□

Proof of Proposition 14.

$$\begin{aligned}
\mathcal{L}_2 \circ \left[\mathbf{I} + \tau\mathcal{A}_{elastic}\right] &= \left[\mathbf{I} + \tau\mathcal{L}_1\right] \circ \left[\mathbf{I} + \tau\mathcal{A}_{elastic}\right] \\
&= \mathbf{I} + \tau\left(\mathcal{L}_1 + \mathcal{A}_{elastic}\right) + \tau^2\left(\mathcal{L}_1 \circ \mathcal{A}_{elastic}\right) \\
&= \mathbf{I} - \tau\left(\lambda+3\mu\right)\Delta + \tau^2\mu\left(\lambda+2\mu\right)\Delta^2
\end{aligned}$$

□

Proof of Proposition 15.

$$\begin{aligned}
\mathcal{L}_3 \circ \mathcal{A}_{elastic\text{-}2} &= \left[-\left(\lambda+3\mu\right)\Delta\mathbf{I} + \left(\lambda+2\mu\right)\nabla\mathrm{div}\right] \circ \left[\mu\Delta^2\mathbf{I} + \left(\lambda+2\mu\right)\Delta\nabla\mathrm{div}\right] \\
&= -\mu\left(\lambda+3\mu\right)\Delta^3\mathbf{I} - \left(\lambda+2\mu\right)\left(\lambda+3\mu\right)\Delta^2\nabla\mathrm{div} \\
&\quad + \mu\left(\lambda+2\mu\right)\nabla\mathrm{div}\,\Delta^2 + \left(\lambda+2\mu\right)^2\nabla\mathrm{div}\,\Delta\nabla\mathrm{div} \\
&= \mu\left(\lambda+3\mu\right)\left(-\Delta\right)^3\mathbf{I} - \left[\left(\lambda+2\mu\right)\left(\lambda+3\mu\right) - \mu\left(\lambda+2\mu\right) - \left(\lambda+2\mu\right)^2\right]\Delta^2\nabla\mathrm{div} \\
&= \mu\left(\lambda+3\mu\right)\left(-\Delta\right)^3\mathbf{I} - 0\cdot\Delta^2\nabla\mathrm{div} \\
&= \mu\left(\lambda+3\mu\right)\left(-\Delta\right)^3\mathbf{I}
\end{aligned}$$

□

Proof of Proposition 16.

$$\begin{aligned}\mathcal{L}_4 \circ [\mathbf{I} + \tau \mathcal{A}_{\textit{elastic-2}}] &= [\mathbf{I} - \tau \Delta \circ \mathcal{L}_3] \circ [\mathbf{I} + \tau \mathcal{A}_{\textit{elastic-2}}] \\ &= \mathbf{I} + \tau \left(\mathcal{A}_{\textit{elastic-2}} - \Delta \circ \mathcal{L}_3 \right) - \tau^2 \left(\Delta \circ \mathcal{L}_3 \circ \mathcal{A}_{\textit{elastic-2}} \right) \\ &= \mathbf{I} + \tau \left(\lambda + 4\mu \right) \Delta^2 + \tau^2 \mu \left(\lambda + 3\mu \right) \Delta^4\end{aligned}$$

□

Bibliography

[1] L. Alvarez, J. Esclarín, M. Lefébure, and J. Sánchez. A PDE model for computing the optical flow. In *Proceedings XVI Congreso de Ecuaciones Diferenciales y Aplicaciones*, pages 1349–1356, September 1999.

[2] L. Amodei and M. N. Benbourhim. A vector spline approximation. *Journal of Approximation Theory*, **67**:51–79, 1991.

[3] M. Arigovindan. *Variational Reconstruction of Scalar and Vector Images from Non-Uniform Samples*. PhD thesis, École Polytechnique Fédérale de Lausanne, 2005.

[4] M. Arigovindan, M. Sühling, C. Jansen, P. Hunziker, and M. Unser. Full motion and flow field recover from echo Doppler data. *IEEE Transactions on Medical Imaging*, **26**(1):31–45, January 2007.

[5] K. V. Arya, P. Gupta, P. K. Kalra, and P. Mitra. Image registration using robust M-estimators. *Pattern Recognition Letters*, **28**:1957–1968, 2007.

[6] U. M. Ascher. *Numerical Methods for Evolutionary Differential Equations*. Society for Industrial and Applied Mathematics, 2008.

[7] J. Ashburner and K. J. Friston. Nonlinear spatial normalization using basis functions. *Human Brain Mapping*, **7**:254–266, 1999.

[8] J. Astola and I. Virtanen. A measure of overall statistical dependence based on the entropy concept. In *Proceedings of the University of Vaasa*, number 91, 1983.

[9] M. J. Atallah. Faster image template matching in the sum of the absolute value of differences measure. *IEEE Transactions on Image Processing*, **10**(4):659–663, April 2001.

[10] R. Bajcsy and S. Kovacic. Multiresolution elastic matching. *Computer Vision, Graphics, and Image Processing*, **46**:1–21, 1989.

[11] R. A. Banvard. The Visible Human Project® image data set from inception to completion and beyond. In *Proceedings CODATA 2002: Frontiers of Scientific and Technical Data, Track I-D-2: Medical and Health Data*, October 2002.

[12] R. Bhagalia, J. A. Fessler, and B. Kim. Accelerated nonrigid intensity-based image registration using importance sampling. *IEEE Transactions on Medical Imaging*, **28**(8):1208–1216, August 2009.

[13] M. Bro-Nielsen. *Medical Image Registration and Surgery Simulation.* PhD thesis, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, 1996. IMM-PHD-1996-25.

[14] C. Broit. *Optimal Registration of Deformed Images.* PhD thesis, University of Pennsylvania, 1981.

[15] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *Proceedings of the Eighth European Conference on Computer Vision*, **3024** of *Lecture Notes in Computer Science*, pages 25–36. Springer-Verlag, 2004.

[16] A. Bruhn, J. Weickert, T. Kohlberger, and C. Schnörr. A multigrid platform for real-time motion computation with discontinuity-preserving variational methods. *International Journal of Computer Vision*, **70**(3):257–277, 2006.

[17] A. Bruhn, J. Weickert, and C. Schnörr. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, **61**(3):211–231, 2005.

[18] P. Cachier and N. Ayache. Isotropic energies, filters and splines for vector field regularization. *Journal of Mathematical Imaging and Vision*, **20**(3):251–265, 2004.

[19] P. Cachier and X. Pennec. 3d non-rigid registration by gradient descent on a gaussian weighted similarity measure using convolutions. In *Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, pages 182–189, 2000.

[20] N. D. Cahill, J. A. Noble, and D. J. Hawkes. Examining numerical solutions to the discrete Navier-Lamé equations for nonrigid registration. In *Proceedings of the IEEE Western New York Image Processing Workshop*, September 2006.

[21] N. D. Cahill, J. A. Noble, and D. J. Hawkes. Fourier methods for nonparametric image registration. In *Proceedings of the CVPR Workshop on Image Registration and Fusion*, June 2007.

[22] N. D. Cahill, J. A. Noble, and D. J. Hawkes. A Demons algorithm for image registration with locally adaptive regularization. In *Proceedings of the Twelfth International Conference on Medical Image Computing and Computer-Assisted Intervention*, September 2009.

[23] M. F. Calitz and H. Rüther. Least absolute deviation (LAD) image matching. *ISPRS Journal of Photogrammetry and Remote Sensing*, **51**(5):223–229, 1996.

[24] T. F. Chan, G. H. Golub, and P. Mulet. A nonlinear primal–dual method for total variation–based image restoration. In *Proceedings of the 12th International Conference on Analysis and Optimization of Systems Images, Wavelets and PDEs*, **219** of *Lecture Notes in Control and Information Sciences*, pages 241–252. Springer-Verlag, 1996.

[25] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, **10**(2):266–277, February 2001.

[26] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Two determinstic half-quadratic regularization algorithms for computed imaging. In *Proc. 1994 IEEE International Conference on Image Processing*, **2**, pages 168–172, 1994.

[27] J.-H. Chen, C.-S. Chen, and Y.-S. Chen. Fast algorithm for robust template matching with M-estimators. *IEEE Transactions on Signal Processing*, **51**(1):230–243, January 2003.

[28] G. E. Christensen. *Deformable Shape Models for Anatomy*. PhD thesis, Washington University, St. Louis, Missouri, March 1994.

[29] G. E. Christensen and H. J. Johnson. Consistent image registration. *IEEE Transactions on Medical Imaging*, **20**(7):568–582, July 2001.

[30] G. E. Christensen, R. D. Rabbitt, and M. I. Miller. 3D brain mapping using a deformable neuroanatomy. *Physics in Medicine and Biology*, **39**(3):609–618, 1994.

[31] G. E. Christensen, R. D. Rabbitt, and M. I. Miller. Deformable templates using large deformation kinematics. *IEEE Transactions on Image Processing*, **5**(10):1435–1447, October 1996.

[32] A. C. S. Chung, R. Gan, and W. M. Wells III. Robust multi-modal image registration based on prior joint intensity distributions and minimization of Kullback-Leibler distance. *HKUST CSE Technical Report, HKUST-CS07-01*, 2007.

[33] A. Collignon. *Multi-modality Medical Image Registration by Maximization of Mutual Information*. PhD thesis, Catholic University of Leuven, Leuven, Belgium, 1998.

[34] A. Collignon, D. Vandermeulen, P. Suetens, and G. Marchal. 3D multi-modality medical image registration using feature space clustering. In *Proceedings of the First International Conference on Computer Vision, Virtual Reality and Robotics*

in Medicine, **905** of *Lecture Notes in Computer Science*, pages 193–204. Springer Berlin, 1995.

[35] W. R. Crum, C. Tanner, and D. J. Hawkes. Anisotropic multi-scale fluid registration: Evaluation in magnetic resonance breast imaging. *Physics in Medicine and Biology*, **50**(21):5153–5174, November 2005.

[36] C. Davatzikos, J. L. Prince, and R. N. Bryan. Image registration based on boundary mapping. *IEEE Transactions on Medical Imaging*, **15**(1):112–115, February 1996.

[37] N. Drissi, T. Chonavel, and J. M. Boucher. Generalized cumulative residual entropy for distributions with unrestricted supports. *Research Letters in Signal Processing*, **2008**(790607), 2008.

[38] M. Droske and M. Rumpf. A variational approach to non-rigid morphological image registration. *SIAM Journal on Applied Mathematics*, **64**(2):668–687, 2004.

[39] B. Fischer and J. Modersitzki. Fast inversion of matrices arising in image processing. *Numerical Algorithms*, **22**(1):1–11, October 1999.

[40] B. Fischer and J. Modersitzki. Curvature based image registration. *Journal of Mathematical Imaging and Vision*, **18**(1):81–85, January 2003.

[41] P. A. Freeborough and N. C. Fox. Modeling brain deformations in Alzheimer disease by fluid registration of serial 3D MR images. *Journal of Computer Assisted Tomography*, **22**(5):838–843, 1998.

[42] C. Frohn-Schauf, S. Henn, L. Hömke, and K. Witsch. Total variation based image registration. In *Proceedings of the International Conference on PDE-Based Image Processing and Related Inverse Problems*, Mathematics and Visualization, pages 343–361. Springer, 2005.

[43] I. M. Gelfand and S. V. Fomin. *Calculus of Variations*. Courier Dover Publications, 2000.

[44] S. Geman and D. E. McClure. Statistical methods for tomographic image reconstruction. In *Proceedings of the 46th Session of the International Statistical Institute, Bulletin of the ISI*, **52**, 1987.

[45] M. S. Gockenbach. *Partial Differential Equations*. SIAM, 2002.

[46] U. Grenander and M. I. Miller. Computational anatomy: An emerging discipline. *Quarterly of Applied Mathematics*, **LVI**:617–694, 1998.

[47] E. Haber and J. Modersitzki. Intensity gradient based registration and fusion of multi-modal images. *Methods of Information in Medicine*, **46**(3):292–299, 2007.

[48] J. V. Hajnal, N. Saeed, A. Oatridge, E. J. Williams, I. R. Young, and G. M. Bydder. Detection of subtle brain changes using subvoxel registration and subtraction of serial MR images. *Journal of Computer Assisted Tomography*, **19**:677–691, 1995.

[49] G. Hermosillo, C. Chefd'Hotel, and O. Faugeras. Variational methods for multimodal image matching. *International Journal of Computer Vision*, **50**(3):329–343, December 2002.

[50] G. Hermosillo Valadez. *Variational Methods for Multimodal Image Matching*. PhD thesis, Université de Nice - Sophia Antipolis, 2002.

[51] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, **17**:185–203, 1981.

[52] P. J. Huber. *Robust Statistics*. Wiley Interscience, February 1981.

[53] S. Kabus. *Multiple-Material Variational Image Registration*. PhD thesis, der Universität zu Lübeck, October 2006.

[54] S. Kabus, A. Franz, and B. Fischer. Variational image registration with local properties. In J. P. W. Pluim, B. Likar, and F. A. Gerritsen, editors, *Proc. Third International Workshop on Biomedical Image Registration*, **4057** of *Lecture Notes in Computer Science*, pages 92–100, July 2006.

[55] H. Kestelman. Riemann-Stieltjes integration. In *Modern Theories of Integration*, chapter 11, pages 247–269. Dover, 2nd edition, 1960.

[56] J. Kim and J. A. Fessler. Intensity-based image registration using robust correlation coefficients. *IEEE Transactions on Medical Imaging*, **23**(11):1430–1444, November 2004.

[57] D. Lee, M. Hofmann, F. Steinke, Y. Altun, N. D. Cahill, and B. Schölkopf. Learning similarity measure for multi-modal 3D image registration. In *Proc. Computer Vision and Pattern Recognition*, June 2009.

[58] L. Lemieux, R. Jagoe, D. R. Fish, N. D. Kitchen, and D. G. T. Thomas. A patient-to-computed-tomography image registration method based on digitally reconstructed radiographs. *Medical Physics*, **21**(11):1749–1760, November 1994.

[59] H. Lester and S. R. Arridge. A survey of hierarchical non-linear medical image registration. *Pattern Recognition*, **32**:129–149, January 1999.

[60] M. Leventon and W. Grimson. Multi-modal volume registration using joint intensity distribution. In *Proceedings of the First International Conference on Medical Image Computing and Computer-Assisted Intervention*, **1496** of *Lecture Notes in Computer Science*, pages 1057–1066. Springer-Verlag, 1998.

[61] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, **16**(2):187–198, April 1997.

[62] C. R. Maurer, Jr., R. Qi, and V. Raghavan. A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **25**(2):265–270, February 2003.

[63] J. Modersitzki. *Numerical Methods for Image Registration.* Oxford University Press, 2004.

[64] H.-H. Nagel. Constraints for the estimation of displacement vector fields from image sequences. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, **2**, pages 945–951, August 1983.

[65] H.-H. Nagel. On the estimation of optical flow: Relations between different approaches and some new results. *Artificial Intelligence*, **33**:299–324, 1987.

[66] T. Netsch, P. Rosch, A. van Muiswinkel, and J. Weese. Towards real-time multimodality 3d medical image registration. In *Proceedings of the 8th International Conference on Computer Vision*. IEEE Computer Society Press, 2001.

[67] E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, **33**(3):1065–1076, September 1962.

[68] X. Pennec, P. Cachier, and N. Ayache. Understanding the "Demon's algorithm": 3D non-rigid registration by gradient descent. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, **1679** of *Lecture Notes in Computer Science*, pages 597–606. Springer-Verlag, 1999.

[69] T. Pock, M. Urschler, C. Zach, R. Beichel, and H. Bischof. A duality based algorithm for TV-l^1-optical-flow image registration. In *Proceedings of the Tenth International Conference on Medical Image Computing and Computer-Assisted Intervention*, **4792** of *Lecture Notes in Computer Science*, pages 511–518. Springer-Verlag, 2007.

[70] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, 2007.

[71] M. Rao, Y. Chen, B. C. Vemuri, and F. Wang. Cumulative residual entropy: A new measure of information. *IEEE Transactions on Information Theory*, **50**(6):1220–1228, June 2004.

[72] F. J. P. Richard, P. R. Bakić, and A. D. A. Maidment. Non-rigid registration of mammograms obtained with variable breast compression: A phantom study. In *Proceedings of the Workshop on Biomedical Image Registration*, **2717** of *Lecture Notes in Computer Science*, pages 281–290. Springer-Verlag, 2003.

[73] R. D. Richtmyer and K. W. Morton. *Difference Methods for Initial-Value Problems*. John Wiley & Sons, Inc., 2nd edition, 1967.

[74] A. Roche, G. Malandain, X. Pennec, and N. Ayache. The correlation ratio as a new similarity measure for multimodal image registration. In *Proceedings of the First International Conference on Medical Image Computing and Computer-Assisted Intervention*, **1496** of *Lecture Notes in Computer Science*, pages 1115–1124. Springer-Verlag, 1998.

[75] A. Roche, G. Malandain, X. Pennec, and N. Ayache. Multimodal image registration by maximization of the correlation ratio. *INRIA Research Report*, **3378**, August 1998.

[76] A. Roche, X. Pennec, M. Rudolph, D. P. Auer, G. Malandain, S. Ourselin, L. M. Auer, and N. Ayache. Generalized correlation ratio for rigid registration of 3D ultrasound with MR images. In *Proceedings of the Third International Conference on Medical Image Computing and Computer-Assisted Intervention*, **1935** of *Lecture Notes in Computer Science*, pages 567–577. Springer-Verlag, 2000.

[77] M. Sabuncu and P. Ramadge. Gradient based nonuniform subsampling for information-theoretic alignment methods. In *Proc. Int. Conf. IEEE Engineering in Medicine and Biology Society*, **1**, pages 1683–1686, 2004.

[78] M. R. Sabuncu and P. Ramadge. Using spanning graphs for efficient image registration. *IEEE Transactions on Image Processing*, **17**(5):788–797, May 2008.

[79] M. N. Safran, M. Freiman, M. Werman, and L. Joskowicz. Curvelet-based sampling for accurate and efficient multi-modal image registration. In J. P. W. Pluim

and B. M. Dawant, editors, *Proc. SPIE Medical Imaging 2009: Image Processing*, **7259**, February 2009.

[80] C. Schnörr. Unique reconstruction of piecewise-smooth images by minimizing strictly convex nonquadratic functionals. *Journal of Mathematical Imaging and Vision*, **4**:189–198, 1994.

[81] C. Ó. S. Sorzano, P. Thévenaz, and M. Unser. Elastic registration of biological images using vector-spline regularization. *IEEE Transactions on Biomedical Engineering*, **52**(4):652–663, April 2005.

[82] R. Stefanescu, X. Pennec, and N. Ayache. Grid powered nonlinear image registration with locally adaptive regularization. *Medical Image Analysis*, **8**(3):325–342, September 2004.

[83] C. Studholme, D. L. G. Hill, and D. J. Hawkes. Multiresolution voxel similarity measures for MR-PET registration. In Y. Bizais, editor, *Proceedings of the International Conference on Information Processing in Medical Imaging*, pages 287–298. Kluwer, June 1995.

[84] C. Studholme, D. L. G. Hill, and D. J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, **32**:71–86, 1999.

[85] H. Sundar, D. Shen, G. Biros, C. Xu, and C. Davatzikos. Robust computation of mutual information using spatially adaptive meshes. In *Proceedings of the Tenth International Conference on Medical Image Computing and Computer-Assisted Intervention*, **4791** of *Lecture Notes in Computer Science*, pages 950–958. Springer-Verlag, 2007.

[86] J.-P. Thirion. Fast non-rigid matching of 3D medical images. Technical Report 2547, Institut National de Recherche en Informatique et en Automatique, Sophia-Anitpolis, May 1995.

[87] J.-P. Thirion. Image matching as a diffusion process: an analogy with Maxwell's demons. *Medical Image Analysis*, **2**(3):243–260, 1998.

[88] U. Trottenberg, A. Schuller, and C. Oosterlee. *Multigrid*. Academic Press, 2000.

[89] J. W. Tukey. *Exploratory Data Analysis*. Addison-Wesley, 1977.

[90] S. Uras, F. Girosi, A. Verri, and V. Torre. A computational approach to motion perception. *Biological Cybernetics*, **60**(2):79–87, December 1988.

[91] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *Neuroimage*, **45**(1), March 2009.

[92] P. Viola and W. M. Wells, III. Alignment by maximization of mutual information. In *Proceedings of the Fifth International Conference on Computer Vision*, pages 16–23, June 1995.

[93] F. Wang and B. C. Vemuri. Non-rigid multi-modal image registration using cross-cumulative residual entropy. *International Journal of Computer Vision*, **74**(2):201–205, August 2007.

[94] F. Wang, B. C. Vemuri, M. Rao, and Y. Chen. A new & robust information theoretic measure and its application to image alignment. In *Proceedings of the International Conference of Information Processing in Medical Imaging (IPMI 2003)*, pages 388–400, July 2003.

[95] J. Weickert, A. Bruhn, T. Brox, and N. Papenberg. A survey on variational optic flow methods for small displacements. In *Mathematical Models for Registration and Applications to Medical Imaging*, pages 103–136. Springer Berlin Heidelberg, 2006.

[96] J. Weickert and C. Schnörr. A theoretical framework for convex regularizers in PDE-based computation of image motion. *International Journal of Computer Vision*, **45**(3):245–264, December 2001.

[97] J. West *et al.* Comparison and evaluation of retrospective intermodality brain image registration techniques. *Journal of Computer Assisted Tomography*, **21**(4):554–566, Jul–Aug 1997.

[98] A. Wong and W. Bishop. Efficient and robust approach to simultaneous non-rigid image registration and gamma estimation in the frequency domain. In *Proc. International Conference on Image Processing, Computer Vision, and Pattern Recognition*, July 2008.

[99] R. P. Woods, J. C. Mazziotta, and S. R. Cherry. MRI-PET registration with automated algorithm. *Journal of Computer Assisted Tomography*, **17**(4):536–546, Jul–Aug 1993.