IOP Publishing Inverse Problems

Inverse Problems **33** (2017) 035004 (33pp) doi:10.1088/1361-6420/aa55af

Shape-based image reconstruction using linearized deformations

Ozan Öktem[1], Chong Chen[2,5], Nevzat Onur Domaniç[3], Pradeep Ravikumar[4] and Chandrajit Bajaj[3]

[1] Department of Mathematics, KTH—Royal Institute of Technology, 100 44 Stockholm, Sweden
[2] State Key Laboratory of Scientific and Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, People's Republic of China
[3] Department of Computer Science, University of Texas at Austin, Austin, TX 78712, United States of America
[4] Machine Learning Department, Carnegie Mellon University, Pittsburgh PA 15213, United States of America

E-mail: ozan@kth.se

Received 23 March 2016, revised 2 December 2016
Accepted for publication 23 December 2016
Published 1 February 2017

Abstract

We introduce a reconstruction framework that can account for shape related prior information in imaging-related inverse problems. It is a variational scheme that uses a shape functional, whose definition is based on deformable template machinery from computational anatomy. We prove existence and, as a proof of concept, we apply the proposed shape-based reconstruction to 2D tomography with very sparse and/or highly noisy measurements.

Keywords: tomography, image reconstruction, indirect image matching, computational anatomy, shape analysis, inverse problems

(Some figures may appear in colour only in the online journal)

1. Introduction

Utilising prior information is critical in addressing inverse problems where data is very noisy or the problem is highly ill-posed, the latter often due to inappropriate sampling of data. Regularisation is needed to solve such challenging problems and many approaches enforce sparsity or regularity within a variational setting [40].

Translating an image into knowledge is, to a large extent, dependent on recognising and interpreting shapes of structures within the image. This paper introduces a variational

[5] Present address: Department of Mathematics, KTH—Royal Institute of Technology, 100 44 Stockholm, Sweden.

 © 2017 IOP Publishing Ltd Printed in the UK

reconstruction scheme that makes use of such prior shape information. The use cases from imaging for such an approach are outlined in section 1.1, but our emphasis is mainly theoretical and aims to show how shape information can be used in image reconstruction. The 2D tomographic examples in section 11 merely serve as a proof of concept, and even though they do not replicate all of the complexity in the use cases, they do serve to illustrate the benefits and drawbacks of the proposed method. A further discussion on this and other matters related to applicability is given in section 12.2.

1.1. Use cases from imaging that serve as motivation

Despite the emphasis on theory, our work is inspired from use cases from imaging that we describe here. As already mentioned, section 12.2 discusses extensions needed for a more realistic usage.

In some imaging problems, the main aim is to recover the shape of the sub-structure, whereas variations within the structures are of less importance. One example is electron tomography (ET) where a transmission electron microscope is used to image the internal 3D structures at the nano-scale of a specimen. The resulting tomographic imaging problem is severely ill-posed with incomplete and highly noisy data [34], but scientists often have an idea of the rough overall shape of the structures they seek to image with ET. A concrete example is given in [12], where ET is used to study the 3D shape variability of gold nano-particles, which is closely related to their unique optical and catalytical properties. Here, there are reasonable shape templates that can be used as prior knowledge in reconstruction. Another use case is in imaging flexible DNA-origami nano-structures in their natural environment by means of ET [10]. These deformable structures are built up by a folded single long uninterrupted strand. The folding is governed by specific shorter 'stapler' strands that connect the longer strands at various known places, and the structures in the specimen can be seen as the deformation of a known template. In these imaging applications, it makes sense to account for qualitative prior shape information during the reconstruction. Enforcing an exact spatial match between a template and the reconstruction is often asking for too much since realistic shape information is almost always approximate, so the natural approach is to perform reconstruction assuming the structures are 'shape wise similar' to a template.

Another use case is image registration where deformations are used to fit a template image against a target image. There is by now a rich literature on variational methods for image registration [8, 14, 38, 39, 53] with applications to computational vision and medical imaging [27]. Such registration is usually done as a post-processing step, and there are cases where a template needs to be registered against a target that is only indirectly observed (*indirect image matching*). An example is in joint reconstruction and registration. Here, a natural approach is to consider an intertwined scheme where the template and its registration are updated iteratively, as discussed in section 12.2.2. A similar intertwined approach can also be used for joint image and motion reconstruction in spatiotemporal imaging, as outlined in section 12.3. In both these cases, the intertwined schemes involve indirect registration.

2. Survey of the field

The idea of accounting for shape information in reconstruction is gaining increasing interest within the inverse problems community. It is part of an ongoing development where reconstruction and feature extraction steps are combined to solve ill-posed inverse problems. In this context, features are specific structures in images relevant for interpretation, and their extraction is usually part of the image analysis.

2.1. Joint segmentation and reconstruction

A step towards combined reconstruction and feature extraction is joint segmentation and reconstruction. One approach is to consider a level-set based approach where the true (unknown) image is assumed to be piecewise constant. In this setting, one may consider joint segmentation and reconstruction by minimising a Mumford–Shah-like functional over the set of admissible contours and, given a fixed contour, over the space of piecewise constant densities which may be discontinuous across the contour within each level set [5]. This approach was applied to the 2D inversion of the ray transform in [36] from complete data. A key step lies in the calculation of the 'shape sensitivity' in order to find a descent direction for the cost functional, which in turn leads to an update formula for the contour in the level-set framework. Motivated in part by the inverse problem in ET, the approach was later generalised and applied to a very noisy limited angle and region of interest tomography data in [22]. See also [21, 44] for further developments of the Mumford–Shah approach involving a variety of inverse problems in imaging. Another line of development is based on incorporating priors about the desired classes of the segmentation through a probabilistic model. Here, a hidden Markov measure field model has been used—see [23, 31, 37] for variants of this approach.

A drawback with these variational formulations is that they are computationally demanding. Furthermore, they also lead to non-convex problems, so there is *always the issue of non-uniqueness*. An entirely different approach for joint segmentation and reconstruction, that in part circumvents these problems, is provided in [25]. Here, the approximate inverse method is extended to the setting where one recovers the features directly. It assumes that the features in question can be extracted by applying an operator, which may be a differential operator for calculating partial derivatives or the Laplacian, or it may be the solution operator for the heat equation for fixed time, or it may be a wavelet transform. The approach is, however, limited to features that can be extracted by applying a linear extraction operator on the image.

In general, the papers cited above demonstrate that joint reconstruction and segmentation typically leads to better results than performing the two steps successively, especially in inverse problems where there are high noise levels and/or the data is incomplete.

2.2. Shape-based reconstruction

The next natural step that follows joint segmentation and reconstruction is to consider the shape. *A priori* shape information often includes geometrical information about the structure in the image, so the reconstruction scheme has to capture such information.

One approach is based on describing such geometric information by means of invariants that carry geometrical information about the structures. Such invariants have been successfully used for shape-based classification, see, e.g. [1, 26] and [50, ch 3]. There is in fact an axiomatic treatment for constructing invariant image features [24] useful for recognising objects from different viewpoints and whose numerical values are equal or only moderately affected by basic image transformations. Returning to image reconstruction, [15] shows how such integral invariants can be used in reconstruction. The idea is to encode the prior shape information in a variational setting by introducing the 2-norm of the difference between the invariant of the structure and the invariant of a prior. The main mathematical result is the proof of the existence of minimisers of the corresponding goal functional. The optimisation is performed with a gradient descent approach that requires the differentiation of the goal functional, and the actual numerical implementation is based on minimising a smooth approximation functional, which converges (Γ-convergence) to the original functional.

Another approach for variational image reconstruction that includes prior shape information is given in [16]. Here, one constructs an energy functional based on local segmentation information obtained by segmentation and comparison with a known spatial model. The approach is demonstrated on tomography data from digital phantoms, simulated data, and experimental ET data of virus complexes.

An approach very similar to the one considered here is given in [2]. There, the authors propose a variational scheme for shape-based reconstruction using a deformable template. The scheme is applied to emission tomography. Once a template is chosen, the activity map (the function describing the isotope intensity that one seeks to recover) is assumed to be a deformation of the template obtained through composition with various planar mappings, together with a magnitude adjustment. Hence, just like we will do, they consider linear deformations of the form (4). On the other hand, in their approach the linear deformations are given by their wavelet coefficients whereas we will consider linear deformations given by vector fields in a reproducing kernel Hilbert space (RKHS). Furthermore, [2] contains no analysis of existence or uniqueness.

3. Overview of paper and specific contributions

The main contribution is the development of a variational framework for *shape-based reconstruction* (section 6), which can be seen as an indirect image matching problem where a template is matched against a target that is known only through indirect observations (data). Our approach is heavily inspired by computational anatomy where shape is modelled as a diffeomorphic deformation of a template. The necessary parts of this theory are mentioned in section 5 along with section 4 that provides terminology and notation related to inverse problem theory.

An important theoretical result is to prove existence, which is done in section 7. As with most other variational schemes for image matching, we cannot expect to have uniqueness due to the lack of convexity (section 8). This theoretical investigation is followed by explicit calculations of the shape derivative for the objective functional, both in the continuum setting (section 9) and when the linearised deformations are discretised through finite span approximations (section 10). The latter is of interest in numerical implementation since finite span approximations are well suited for discretising such infinite dimensional variational problems (remark 4).

Section 11 contains some numerical experiments from tomography that show the impact of using prior shape information. The outcome strongly supports the claim that shape-based information can improve reconstructions, especially in a noisy highly under-sampled setting. Furthermore, the experiments also suggest that the prior shape information does not have to be that accurate, which is important for the use cases mentioned as a motivation in section 1.1. None of these claims have, however, been proved formally by mathematical theorems.

Finally, section 12 contains a discussion on various extensions, e.g. one may consider other deformation models that allow for large deformations, such as those suggested within the large deformation diffeomorphic metric mapping (LDDMM) framework. The deformations can be large while being smooth and topology preserving, but performing shape-based reconstruction with these, more elaborate, deformation models are left for the future.

4. Inverse problems, ill-posedness and variational regularisation

4.1. Image reconstruction

The goal in image reconstruction is to estimate some spatially distributed quantity (*image*) from indirect noisy observations (*measured data*). Stated mathematically, we seek to reconstruct an *image* $f_{\text{true}} \in \mathscr{X}$ from *data* $g \in \mathscr{H}$ where

$$g = \mathcal{T}(f_{\text{true}}) + e. \tag{1}$$

Here, $\mathscr{X}$ (*reconstruction space*) is the vector space of all possible images, so it is a suitable Hilbert space of functions defined on a fixed domain $\Omega \subset \mathbb{R}^n$. Next, $\mathscr{H}$ (*data space*) is the vector space of all possible data, which for digitised data is a subset of $\mathbb{R}^m$. Furthermore, $\mathcal{T}: \mathscr{X} \to \mathscr{H}$ (*forward operator*) models how a given image gives rise to data in the absence of noise and measurement errors, and $e \in \mathscr{H}$ (*data noise component*) is a sample of a $\mathscr{H}$ –valued random element E whose probability distribution is the (*data noise model*) assumed to be known.

4.2. Ill-posedness

A naive approach at reconstructing the true (unknown) image f_{true} is to try to solve the equation $\mathcal{T}(f) = g$. Often there are no solutions to this equation since measured data is not in the range of $\mathcal{T}$ for a variety of reasons (noise, modelling errors, etc). This is commonly addressed by relaxing the notion of a solution by considering

$$\min_{f \in \mathscr{X}} \mathcal{D}\Big(\mathcal{T}(f), g\Big). \tag{2}$$

The mapping $\mathcal{D} : \mathscr{H} \times \mathscr{H} \to \mathbb{R}_+$ (*data discrepancy functional*) quantifies the data misfit and a natural candidate is to choose it as a suitable affine transformation of the negative log likelihood of data. In such a case, solving (2) amounts to finding maximum likelihood solutions. This works well when (2) has a unique solution (*uniqueness*) that depends continuously on data (*stability*). This is, however, not the case when (1) is ill-posed and regularisation needs to be used in order to introduce uniqueness and stability by making use of prior knowledge about f_{true}.

4.3. Variational regularisation

These regularisation methods have gained much attention lately, and especially so in imaging [40]. The idea is to add a penalisation term to the objective functional in (2), resulting in a variational problem of the type

$$\min_{f \in \mathscr{X}} [\mu \mathcal{R}(f) + \mathcal{D}(\mathcal{T}(f), g)] \quad \text{for some given } \mu \geqslant 0. \tag{3}$$

The functional $\mathcal{R} : \mathscr{X} \to \mathbb{R}_+$ introduces stability, and perhaps uniqueness, by encoding prior information about f_{true}. It often encodes some *a priori* known regularity property of f_{true}, e.g. assuming $\mathscr{X} \subset \mathscr{L}^2(\Omega, \mathbb{R})$, and taking the $\mathscr{L}^2$-norm of the gradient magnitude (Dirichlet energy) is known to produce smooth solutions whereas taking the $\mathscr{L}^1$-norm of the gradient magnitude (total variation) yields solutions that preserve edges while smooth variations may be suppressed [13].

5. Shape theory based on deformable templates

Shape theory seeks to develop quantitative tools to study shapes and their variability. A key objective is to define a *shape metric* that quantifies the shape similarity between two objects/structures in the image.

There are a variety of mathematical approaches for modelling shapes, and the approach we consider is based on *deformable templates*. This approach, which can be traced back to work pursued in 1917 by D'Arcy Thompson [45], is based on the idea that shapes can be represented as a deformation of a template. A quantitative analysis of shape variability can then be based on quantifying the cost of the deformation. Grenander laid the foundations of pattern theory [17] that was necessary for a coherent mathematical and statistical theory for shape and shape variability based on the above idea of deformable templates. We here provide a very brief introduction to shape theory with emphasis on the linearised deformations' framework.

The starting point is to define a set of *deformable objects* representing objects/structures whose shape we seek to analyse. There is also an associated set of *deformations* that can act on the deformable objects. In our case, a deformation is given as a diffeomorphic perturbation of the identity map (linearised deformation). Under certain conditions, the set of deformations becomes a group with a metric that can be used, together with the group action, to quantify the shape similarity between two deformable objects.

5.1. Set of deformable objects

We consider grey scale images as deformable objects, so the set $\mathscr{X}$ of deformable objects is a vector space of real-valued functions defined on a fixed underlying *image domain* $\Omega \subset \mathbb{R}^n$ ($n = 2$ for 2D images or $n = 3$ for 3D images), which we assume to be an open and bounded set. We furthermore assume that $\mathscr{X} \subset \mathscr{L}^2(\Omega, \mathbb{R})$.

5.2. Set of deformations—linearised deformations

The set of deformations will consist of transformations that map the image domain $\Omega \subset \mathbb{R}^n$ into Ω, thereby acting on $\mathscr{X}$ by transforming the underlying coordinate grid. It is natural to form new deformations by concatenating existing ones, and the identity mapping is a natural 'zero' deformation. This implies that the set of deformations should form a semigroup under the group law given by the composition of functions with identity mapping as its identity element.

Linearised deformations are transformations defined as

$$\phi^\nu := \mathrm{Id}_\Omega + \nu \quad \text{for some } \nu \in \mathcal{V}. \tag{4}$$

Here, $\mathcal{V} \subset \mathscr{C}^1_0(\Omega, \mathbb{R}^n)$ is a fixed Hilbert space with norm $\|\cdot\|_\mathcal{V}$ and its elements are vector fields (*displacement fields*) supported in Ω. The set $\mathscr{C}^1_0(\Omega, \mathbb{R}^n)$ above is the Banach space of continuously differentiable vector fields with support in Ω (if Ω is unbounded, then the vector field along with its derivative is assumed to decay to zero at infinity).

Given a Hilbert space $\mathcal{V}$ of vector fields as above, define $\mathscr{G}_\mathcal{V} \subset \mathscr{C}^1_0(\Omega, \mathbb{R}^n)$ by

$$\mathscr{G}_\mathcal{V} := \{\phi : \Omega \to \mathbb{R}^n : \phi = \phi^\nu \text{ for some } \nu \in \mathcal{V} \text{ and } \phi^\nu \text{ is given by (4)}\}. \tag{5}$$

Then, $\mathscr{G}_\mathcal{V}$ is closed under composition whenever $\mathcal{V}$ is closed under composition (remark 1), in which case it forms a semigroup. On the other hand, mappings defined as in (4) are not necessarily invertible unless ν is sufficiently small and regular [50, proposition 8.6]. It is furthermore difficult to mathematically characterise the Hilbert spaces $\mathcal{V}$ that ensure elements in

$\mathscr{G}_{\mathcal{V}}$ are invertible, i.e. $\mathscr{G}_{\mathcal{V}}$ is not a group. Section 12.1 has a brief discussion on how to generate large diffeomorphic deformations. For our purposes, the semigroup structure is sufficient. As an example, the way a deformation in $\mathscr{G}_{\mathcal{V}}$ transforms an element in $\mathscr{X}$ is naturally described by a semigroup action, which is a mapping $\mathscr{G}_{\mathcal{V}} \times \mathscr{X} \to \mathscr{X}$ denoted by $(\phi, f) \mapsto \phi.f$.

Remark 1. One has to be somewhat careful when concluding that $\mathscr{G}_{\mathcal{V}}$ is closed under composition. Vector fields in $\mathcal{V}$ are given here as $\mathbb{R}^n$-valued mappings supported on $\Omega \subset \mathbb{R}^n$ and if Ω is unbounded, then the vector field along with its derivative is assumed to decay to zero at infinity. Such vector fields trivially extend to all of $\mathbb{R}^n$ by setting them to zero outside Ω, so whenever convenient, we may consider them as mappings defined on all of $\mathbb{R}^n$. This implies that $\mathcal{V}$ is closed under composition and any deformation in $\mathscr{G}_{\mathcal{V}}$ extends to all of $\mathbb{R}^n$, so $\mathscr{G}_{\mathcal{V}}$ is closed under composition.

5.3. Reproducing kernel Hilbert spaces

The theory of RKHS deals with function spaces. The theory was initialised in the 1940s for spaces of scalar-valued functions [3]; it was later extended to spaces of functions with values in locally convex topological spaces [43]. It has lately gained significant interest due to applications in machine learning [11, 32, 42].

The starting point is an abstract Hilbert space $\mathcal{V}$ whose elements are functions defined on a fixed domain $\Omega \subset \mathbb{R}^n$ that take values in a real Hilbert space $\mathscr{X}$. Such a space is an RKHS if evaluation functionals $\delta_x^a : \mathcal{V} \to \mathbb{R}$

$$\delta_x^a(\nu) := \big\langle \nu(x), a \big\rangle_{\mathscr{X}} \quad \text{for } \nu \in \mathcal{V}$$

are bounded for every $x \in \Omega$ and $a \in \mathscr{X}$. One way to construct an RKHS is to specify a reproducing kernel function, which is an operator $\mathcal{K} : \Omega \times \Omega \to \mathscr{L}(\mathscr{X}, \mathscr{X})$ such that

(a) $\mathcal{K}(\cdot, x)(a) \in \mathcal{V}$ for all $x \in \Omega$ and $a \in \mathscr{X}$.
(b) The *reproducing property* holds for $\mathcal{K}$, i.e. if $x \in \Omega$ then

$$\big\langle \nu(x), a \big\rangle_{\mathscr{X}} = \big\langle \nu, \mathcal{K}(\cdot, x)(a) \big\rangle_{\mathcal{V}} \quad \text{for any } \nu \in \mathcal{V} \text{ and } a \in \mathscr{X}. \tag{6}$$

$\mathscr{L}(\mathscr{X}, \mathscr{X})$ above is the Banach space of bounded linear operators on $\mathscr{X}$. A central result is that a Hilbert space $\mathcal{V}$ of $\mathscr{X}$-valued functions is an RKHS if and only if it has a continuous reproducing kernel $\mathcal{K} : \Omega \times \Omega \to \mathscr{L}(\mathscr{X}, \mathscr{X})$.

We consider vector fields in $\mathbb{R}^n$, so $\mathscr{X} = \mathbb{R}^n$ and $\mathcal{V}$ contains $\mathbb{R}^n$-valued functions. It is an RKHS if it has a continuous positive definite reproducing kernel $\mathcal{K} : \Omega \times \Omega \to \mathscr{L}(\mathbb{R}^n, \mathbb{R}^n)$, which in turn can be represented by a continuous positive definite function K: $\Omega \times \Omega \to \mathbb{M}_+^{n \times n}$. Here, $\mathbb{M}^{n,m}$ denotes the vector space of all $(n \times m)$ matrices and $\mathbb{M}_+^{n,m}$ denotes those matrices that are positive definite. Furthermore, if $\mathcal{V} \subset \mathscr{L}^2(\Omega, \mathbb{R}^n)$ then one can relate the inner product on $\mathcal{V}$ to the $\mathscr{L}^2$-inner product by a straightforward calculation:

$$\langle \eta, \nu \rangle_{\mathscr{L}^2(\Omega, \mathbb{R}^n)} = \left\langle \nu, \int_\Omega \mathcal{K}(\cdot, x)(\eta(x))\, \mathrm{d}x \right\rangle_{\mathcal{V}}. \tag{7}$$

Note also that $\|\cdot\|_{\mathcal{V}}$ is different from $\|\cdot\|_{\mathscr{L}^2(\Omega, \mathbb{R}^n)}$.

5.4. The deformation operator and shape similarity

Elements in $\mathscr{G}_{\mathcal{V}}$ represent deformations that act on deformable objects, which in our case are grey scale images on Ω represented by square integrable functions $f : \Omega \to \mathbb{R}$. One way to let

a deformation, i.e. element $\phi \in \mathscr{G}_{\mathcal{V}}$, act on a grey scale image f is by a semigroup action, say $\phi.f := f \circ \phi$.

Next, each element in $\mathscr{G}_{\mathcal{V}}$ has a corresponding element in $\mathcal{V}$, i.e. $\phi = \phi^{\nu}$ for some $\nu \in \mathcal{V}$. Hence, elements in $\mathcal{V}$ act on deformable objects by means of the *deformation operator* $\mathcal{W}_f : \mathcal{V} \to \mathscr{X}$ given as

$$\mathcal{W}_f(\nu) := \phi^{\nu}.f = f \circ \phi^{\nu} \quad \text{for } \nu \in \mathcal{V} \tag{8}$$

where $\phi^{\nu} \in \mathscr{G}_{\mathcal{V}}$ is given by (4). The deformation operator models how a grey scale image is deformed by ϕ^{ν}. The shape similarity between f and its deformed version $\mathcal{W}_f(\nu)$ can now be quantified by $\|\nu\|^2_{\mathcal{V}}$. Note that it is zero only if the deformation is the identity mapping (no deformation).

6. Shape-based reconstruction

Consider the inverse problem in (1) where elements in $\mathscr{X}$ represent grey scale images defined over some fixed image domain $\Omega \subset \mathbb{R}^n$, i.e. $\mathscr{X} \subset \mathscr{L}^2(\Omega, \mathbb{R})$. Next, assume that the true image f_{true} is approximately a deformation of a known *shape template* $I \in \mathscr{X}$. As we shall see next, the shape functional in section 5.4 allows us to use such prior shape information in the reconstruction of f_{true} in (1).

Assume now that the true (unknown) image f_{true} can be written as an admissible deformation of the shape template I, i.e.

$$f_{\text{true}} = \mathcal{W}_I(\nu) \quad \text{on } \Omega \text{ for some } \nu \in \mathcal{V}. \tag{9}$$

To reconstruct f_{true} in (1), while accounting for prior shape information given by (9), we formulate the following variational problem:

$$\inf_{\nu \in \mathcal{V}} [\lambda \|\nu\|^2_{\mathcal{V}} + \mu \mathcal{R}(\mathcal{W}_I(\nu)) + \mathcal{D}(\mathcal{T}(\mathcal{W}_I(\nu)), g)]. \tag{10}$$

Here, the mapping $\mathcal{D} : \mathscr{H} \times \mathscr{H} \to \mathbb{R}_+$ is the data discrepancy functional introduced in (2), $\mathcal{R} : \mathscr{X} \to \mathbb{R}_+$ is the *regularity functional* that encodes regularity properties of f_{true} that are known beforehand, and as already mentioned, $\|\nu\|^2_{\mathcal{V}}$ quantifies the shape similarity between the template I and its deformed variant $\mathcal{W}_I(\nu)$. Finally, $\lambda \geqslant 0$ regulates the influence of the prior shape information and $\mu \geqslant 0$ regulates the influence of the *a priori* regularity properties that f_{true} might possess.

7. Existence of solutions

A basic property that a reconstruction scheme must fulfil is that it always has solutions (the existence of solutions), since otherwise it is of limited usefulness for reconstruction. In our case, this translates into proving the existence of solutions for (10).

Theorem 1. *Let $\mathcal{V}$ be a Hilbert space and assume $f \mapsto \mathcal{D}(\mathcal{T}(f), g)$ and $f \mapsto \mathcal{R}(f)$ are both lower semi-continuous on $\mathscr{X}$. Then, (10) has a solution in $\mathcal{V}$.*

Proof. Let $\mathcal{E} : \mathcal{V} \to \mathbb{R}_+$ denote the goal functional in (10), i.e. for given $\lambda, \mu \geqslant 0$ and $I \in \mathscr{X}$, it is defined as

$$\mathcal{E}(\nu) := \lambda \|\nu\|^2_{\mathcal{V}} + \mu \mathcal{R}(\mathcal{W}_I(\nu)) + \mathcal{D}(\mathcal{T}(\mathcal{W}_I(\nu)), g) \quad \text{with } \nu \in \mathcal{V}. \tag{11}$$

If we can prove that $\mathcal{E}$ is lower semi-continuous on $\mathcal{V}$, then (10) has solutions.

Let $\{\nu_n\}_n \subset \mathcal{V}$ be a minimising sequence for (10) so $\|\nu_n\|_{\mathcal{V}}$ are bounded. Since $\mathcal{V}$ is a Hilbert space, it is in particular a complete metric space with respect to the distance function induced by the inner product. Hence, there is a subsequence, still denoted by $\{\nu_n\}_n$, that converges weakly to some $\nu^* \in \mathcal{V}$ (see [50, theorem A.16]). By [50, proposition A.15] we then have

$$\|\nu^*\|_{\mathcal{V}} \leqslant \liminf_{n\to\infty} \|\nu_n\|_{\mathcal{V}}. \tag{12}$$

Next, by (4) we get

$$\|\phi^{\nu_n}(x) - \phi^{\nu^*}(x)\| = \|\nu_n(x) - \nu^*(x)\| \quad \text{for } x \in \Omega\,.$$

Since $\nu \mapsto \nu(x)$ is a continuous linear form on $\mathcal{V}$, we conclude that

$$\limsup_{x\in\Omega} \left\|\phi^{\nu_n}(x) - \phi^{\nu^*}(x)\right\| = 0.$$

Thus, ϕ^{ν} is continuous in the weak topology of $\mathcal{V}$. Furthermore, for every shape template $I \in \mathscr{X}$, using the result in the proof of [48, theorem 2.5] gives us

$$\limsup_{x\in\Omega} \left\|\mathcal{W}_I(\nu_n)(x) - \mathcal{W}_I(\nu^*)(x)\right\| = 0.$$

Therefore, $\mathcal{W}_I$ is also continuous in the weak topology of $\mathcal{V}$. Now, by assumption both $f \mapsto \mathcal{D}(\mathcal{T}(f), g)$ and $f \mapsto \mathcal{R}(f)$ are lower semi-continuous on $\mathscr{X}$. Hence,

$$\mathcal{R}(\mathcal{W}_I(\nu^*)) \leqslant \liminf_{n\to\infty} \mathcal{R}(\mathcal{W}_I(\nu_n)), \tag{13}$$

$$\mathcal{D}(\mathcal{T}(\mathcal{W}_I(\nu^*)), g) \leqslant \liminf_{n\to\infty} \mathcal{D}(\mathcal{T}(\mathcal{W}_I(\nu_n)), g) \tag{14}$$

(12) together with (13)–(14) implies that ν^* is a solution for the variational problem (10) in $\mathcal{V}$. This concludes the proof. □

8. Uniqueness

Another desirable property of a reconstruction scheme is uniqueness, i.e. the scheme renders a unique solution for given data. Unfortunately, there are no guarantees that (10) has a unique solution, even when both

$$f \to \mathcal{R}(f) \quad \text{and} \quad f \to \mathcal{D}(\mathcal{T}(f), g) \tag{15}$$

are strictly convex. This is because the deformation operator $\mathcal{W}_f : \mathcal{V} \to \mathscr{X}$ in (8) is non-convex.

As with all reconstruction methods that involve solving non-convex optimisation problems, there is always the issue of getting stuck in local extrema. Numerical experiments suggest that this risk is higher if the template needs to undergo a significant translation. Hence, the template should be centred over the object before reconstruction, e.g. by ensuring its centre of gravity coincides reasonably well with the centre of gravity of the (unknown) object we seek to recover. Another observation is that templates with the wrong topology tend to result in reconstructions getting stuck in local minima. In test suite 4, the poor results in figures 7(e), (g) and 8(d) are probably due to iterates getting stuck in local minima. In theory, it should be possible to compensate for erroneous template topology by invoking stronger deformations, as is the case with the corners in the template in figure 1(f) that are hardly visible in the reconstructions in figures 1(g) and (h).

One option to address the issue of non-convexity is to further restrict the set $\mathcal{V}$ of velocity fields, but it is highly non-trivial to work out conditions on $\mathcal{V}$ that would guarantee strict convexity, and thereby uniqueness for (10). There are, however, specific parameterisations for generating deformations that can be worth considering—see discussion in section 12.1.

Remark 2. A complete mathematical analysis of a reconstruction scheme would also include an analysis of stability, i.e. whether the scheme is stable w.r.t. perturbations in data. An even more ambitious endeavour is to establish convergence properties, i.e. to what extent the output from the scheme converges to a (least-squares/maximum likelihood) solution of (2) when the data error tends to zero in some norm while reconstruction parameters are chosen accordingly. Such issues are not considered in this paper.

9. Derivative calculations

Performing reconstruction following the scheme outlined in section 6 requires solving the optimisation problem in (10). Most computationally feasible approaches make use of first order derivative information, so one central topic is to compute the gradient of the goal functional given in (11).

The gradient is calculated w.r.t. the Hilbert structure of $\mathcal{V}$, which is the natural inner-product space for this minimisation. In computing the $\mathcal{V}$-gradient of the mappings

$$\nu \mapsto \mathcal{R}(\mathcal{W}_I(\nu)) \quad \text{and} \quad \nu \mapsto \mathcal{D}(\mathcal{T}(\mathcal{W}_I(\nu)), g), \tag{16}$$

we assume the mappings in (15) are both Gâteaux differentiable and I is differentiable.

9.1. The deformation operator

We here provide an explicit expression for the Gâteaux derivative of the deformation operator defined in (8).

Proposition 1. *Assume $I \in \mathscr{X}$ is differentiable. Then, the deformation operator $\mathcal{W}_I : \mathcal{V} \to \mathscr{X}$ in (8) is Gâteaux differentiable at $\nu \in \mathcal{V}$ and its Gâteaux derivative is*

$$\partial\mathcal{W}_I(\nu)(\eta)(x) = \left\langle \nabla I(\phi^\nu(x)), \eta(x) \right\rangle_{\mathbb{R}^n} \quad \text{for } x \in \Omega \text{ and } \eta \in \mathcal{V}. \tag{17}$$

Proof. The Gâteaux derivative is the linear mapping $\partial\mathcal{W}_I(\nu) : \mathcal{V} \to \mathscr{X}$ defined as

$$\partial\mathcal{W}_I(\nu)(\eta) := \frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{W}_I(\nu + \epsilon\eta)\Big|_{\epsilon=0} \quad \text{for } \eta \in \mathcal{V}.$$

Hence, $\partial\mathcal{W}_I(\nu)(\eta) : \Omega \to \mathbb{R}$ where

$$\partial\mathcal{W}_I(\nu)(\eta)(x) := \frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathcal{W}_I(\nu + \epsilon\eta)(x)\Big|_{\epsilon=0} = \frac{\mathrm{d}}{\mathrm{d}\epsilon}(I \circ \phi^{\nu+\epsilon\eta})(x)\Big|_{\epsilon=0} \quad \text{for } x \in \Omega\,.$$

Now, $I \in \mathscr{X} \subset \mathscr{L}^2(\Omega, \mathbb{R})$ is differentiable so the chain rule gives us

$$\partial\mathcal{W}_I(\nu)(\eta)(x) = \left\langle \nabla I(\phi^\nu(x)), \frac{\mathrm{d}}{\mathrm{d}\epsilon}(\phi^{\nu+\epsilon\eta})\Big|_{\epsilon=0}(x) \right\rangle_{\mathbb{R}^n} \quad \text{for } x \in \Omega\,. \tag{18}$$

The second term in the scalar product on the right hand side of (18) is the derivative of a deformation with respect to variations in the associated vector field.

Next, by (4) we have $\phi^{\nu+\epsilon\eta} = \mathrm{Id}_\Omega + \nu + \epsilon\eta$, so

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}(\phi^{\nu+\epsilon\eta})\bigg|_{\epsilon=0}(x) = \frac{\mathrm{d}}{\mathrm{d}\epsilon}(x+\nu(x)+\epsilon\eta(x))\bigg|_{\epsilon=0} = \eta(x).$$

Now, (17) follows from inserting the above into (18). □

Remark 3. Unless derivatives are to be interpreted in the distribution sense, the template $I \in \mathscr{X}$ in (17) has to be differentiable.

9.2. General matching functionals

Both of the functionals in (16) are of the form $\mathcal{J}_I : \mathcal{V} \to \mathbb{R}_+$ with

$$\mathcal{J}_I(\nu) := \mathcal{L} \circ \mathcal{W}_I(\nu) \quad \text{for } \nu \in \mathcal{V}, \tag{19}$$

where $\mathcal{L} : \mathscr{X} \to \mathbb{R}$ is a fixed functional on $\mathscr{X}$ that is sufficiently regular, e.g. we require it to be Gâteaux differentiable.

Theorem 2. *Assume $\mathcal{L} : \mathscr{X} \to \mathbb{R}$ is Gâteaux differentiable on $\mathscr{X}$ and define $\mathcal{J}_I : \mathcal{V} \to \mathbb{R}$ as in (19) where $I \in \mathscr{L}^2(\Omega, \mathbb{R})$ is differentiable. Then, the Gâtueaux derivative of $\mathcal{J}_I$ is given as*

$$\partial\mathcal{J}_I(\nu)(\eta) = \partial\mathcal{L}(\mathcal{W}_I(\nu))(\langle \nabla I(\phi^\nu(x)), \eta(x)\rangle_{\mathbb{R}^n}) \quad \text{for } \nu, \eta \in \mathcal{V}. \tag{20}$$

Furthermore, let $\mathscr{X} \subset \mathscr{L}^2(\Omega, \mathbb{R})$ and $\mathcal{V}$ is an RKHS with a reproducing kernel represented by the symmetric and positive definite function $\mathsf{K}\colon \Omega \times \Omega \to \mathbb{M}_+^{n\times n}$. Then

$$\nabla\mathcal{J}_I(\nu)(x) = \int_\Omega \nabla\mathcal{L}(\mathcal{W}_I(\nu))(y)\mathsf{K}(x,y)\cdot\nabla I(\phi^\nu(y))\,\mathrm{d}y \quad \text{for } x \in \Omega\,. \tag{21}$$

Proof. The proof of (20) follows directly from the chain rule and proposition 1. More precisely, by the chain rule we have

$$\partial\mathcal{J}_I(\nu)(\eta) = \partial\mathcal{L}(\mathcal{W}_I(\nu))(\partial\mathcal{W}_I(\nu)(\eta)) \quad \text{for } \nu, \eta \in \mathcal{V}. \tag{22}$$

Combining this with (17) gives us (20).

To prove (21) we need to use the assumption $\mathscr{X} \subset \mathscr{L}^2(\Omega, \mathbb{R})$. Then, the Riesz representation theorem allows us to define the gradient of $\mathcal{L}$ as the mapping $\nabla\mathcal{L} : \mathscr{X} \to \mathscr{X}$ given implicitly by the relation

$$\partial\mathcal{L}(f)(h) = \langle \nabla\mathcal{L}(f), h\rangle_{\mathscr{L}^2(\Omega,\mathbb{R})} \quad \text{for } h \in \mathscr{X}. \tag{23}$$

Combining (22) with (23) gives

$$\partial\mathcal{J}_I(\nu)(\eta) = \langle \nabla\mathcal{L}(\mathcal{W}_I(\nu)), \partial\mathcal{W}_I(\nu)(\eta)\rangle_{\mathscr{L}^2(\Omega,\mathbb{R})} \quad \text{for } \nu, \eta \in \mathcal{V}. \tag{24}$$

Next, the expression in (17) can be inserted into (24):

$$\partial \mathcal{J}_I(\nu)(\eta) = \Big\langle \nabla \mathcal{L}(\mathcal{W}_I(\nu))(\,\cdot\,), \Big\langle \nabla I(\phi^\nu(\,\cdot\,)), \eta(\,\cdot\,) \Big\rangle_{\mathbb{R}^n} \Big\rangle_{\mathscr{L}^2(\Omega,\mathbb{R})}$$
$$= \Big\langle \nabla \mathcal{L}(\mathcal{W}_I(\nu))(\,\cdot\,), \Big\langle \nabla I(\phi^\nu(\,\cdot\,)), \eta(\,\cdot\,) \Big\rangle_{\mathbb{R}^n} \Big\rangle_{\mathscr{L}^2(\Omega,\mathbb{R})}$$
$$= \Big\langle \eta(\,\cdot\,), \nabla \mathcal{L}(\mathcal{W}_I(\nu))(\,\cdot\,) \nabla I(\phi^\nu(\,\cdot\,)) \Big\rangle_{\mathscr{L}^2(\Omega,\mathbb{R}^n)}.$$

The last equality makes use of the fact that the inner product in $\mathscr{L}^2(\Omega, \mathbb{R}^n)$ (square integrable $\mathbb{R}^n$-valued functions) is expressible as the inner product in $\mathbb{R}^n$ followed by the inner product in $\mathscr{L}^2(\Omega, \mathbb{R})$ (square integrable real-valued functions). Note also that

$$x \mapsto \nabla \mathcal{L}(\mathcal{W}_I(\nu))(x) \nabla I(\phi^\nu(x)) \quad \text{and} \quad x \mapsto \eta(x)$$

are both mappings from Ω into $\mathbb{R}^n$, so the $\mathscr{L}^2(\Omega, \mathbb{R}^n)$ inner product is well defined. Finally, we use the assumption that $\mathcal{V}$ is an RKHS with reproducing kernel $\mathcal{K} : \Omega \times \Omega \to \mathscr{L}(\mathbb{R}^n, \mathbb{R}^n)$, so by (7) we have

$$\partial \mathcal{J}_I(\nu)(\eta) = \left\langle \eta(\,\cdot\,), \int_\Omega \mathcal{K}(\,\cdot\,, x)(\tilde{\nu}(x))\, \mathrm{d}x \right\rangle_{\mathcal{V}}$$

where $\tilde{\nu} : \Omega \to \mathbb{R}^n$ is defined as

$$\tilde{\nu}(x) := \nabla \mathcal{L}(\mathcal{W}_I(\nu))(x) \nabla I(\phi^\nu(x)) \quad \text{for } x \in \Omega\,.$$

We can now read off the expression for the $\mathcal{V}$-gradient of $\mathcal{J}_I$ and inserting the matrix valued function $\mathsf{K}\colon \Omega \times \Omega \to \mathbb{M}^{n\times n}_+$ representing the reproducing kernel gives us

$$\nabla \mathcal{J}_I(\nu)(x) = \int_\Omega \mathcal{K}(x, y)(\tilde{\nu}(y))\, \mathrm{d}y = \int_\Omega \nabla \mathcal{L}(\mathcal{W}_I(\nu))(y) \mathsf{K}(x, y) \cdot \nabla I(\phi^\nu(y))\, \mathrm{d}y.$$

The last equality is (21), so this concludes the proof of theorem 2. □

In tomography, data is typically represented by $\mathscr{H} = \mathscr{L}^2(Y, \mathbb{R})$ where Y is a manifold of lines given by the data acquisition geometry. If measured data has additive Gaussian noise, then a natural matching functional is

$$\mathcal{J}_I(\nu) := \Big\| \mathcal{T}(I \circ \phi^\nu) - g \Big\|^2_{\mathscr{L}^2(Y,\mathbb{R})}. \tag{25}$$

Corollary 1. *Let the assumptions in theorem 2 hold. Furthermore, assume that $\mathcal{J}_I : \mathcal{V} \to \mathbb{R}$ is given as in (25) with $\mathcal{T}\colon \mathscr{X} \to \mathscr{H}$ Gâteaux differentiable. Then*

$$\partial \mathcal{J}_I(\nu)(\eta) = 2 \Big\langle \partial \mathcal{T}(\mathcal{W}_I(\nu))^* \mathcal{T}(\mathcal{W}_I(\nu) - g)(\,\cdot\,) \nabla I(\phi^\nu(\,\cdot\,)), \eta(\,\cdot\,) \Big\rangle_{\mathscr{L}^2(\Omega,\mathbb{R}^n)} \tag{26}$$

for $\nu, \eta \in \mathcal{V}$. Furthermore, if $\mathscr{X} \subset \mathscr{L}^2(\Omega, \mathbb{R})$ and $\mathcal{V}$ is an RKHS with a reproducing kernel represented by the symmetric and positive definite function $\mathsf{K}\colon \Omega \times \Omega \to \mathbb{M}^{n\times n}_+$, then for $x \in \Omega$ we have

$$\nabla \mathcal{J}_I(\nu)(x) = 2 \int_\Omega \partial \mathcal{T}(\mathcal{W}_I(\nu))^* \mathcal{T}(\mathcal{W}_I(\nu) - g)(y) \mathsf{K}(x, y) \cdot \nabla I(\phi^\nu(y))\, \mathrm{d}y. \tag{27}$$

9.3. Gradient of goal functional

Many optimisation schemes rely on computing the gradient of (11). If we assume $\mathcal{V}$ is an RKHS, then a general expression for its gradient $\nabla\mathcal{E}: \mathcal{V} \to \mathcal{V}$ can be obtained by combining (21) with the $\mathcal{V}$-gradient of $\nu \mapsto \|\nu\|_{\mathcal{V}}^2$, which is simply 2ν.

Corollary 1 can be used to get an explicit expression for the gradient of the objective functional in the case when the matching functional is given by (25) and $\mu = 0$:

$$\nabla\mathcal{E}(\nu)(x) = 2\nu(x) + 2\int_{\Omega} \partial\mathcal{T}(\mathcal{W}_I(\nu))^*\mathcal{T}(\mathcal{W}_I(\nu) - g)(y)\mathsf{K}(x,y)\cdot\nabla I(\phi^{\nu}(y))\,\mathrm{d}y. \tag{28}$$

10. Finite dimensional setting

$\mathcal{V}$ is in general an infinite dimensional Hilbert space of vector fields, so the variational problem in (10) is an optimisation over an infinite dimensional Hilbert space. In a computational setting we can only minimise functionals over finite dimensional vector spaces. One approach is to formulate the optimisation scheme for finding a local minima in the infinite dimensional setting. The explicit expressions for the gradient of the goal functional (section 9) can be used for this. One can then discretise this optimisation scheme. Another approach is to look for 'natural' finite dimensional formulations. In particular, elements in $\mathcal{V}$ can only be evaluated at a finite number of points, so we need to consider suitable finite dimensional sub-spaces of $\mathcal{V}$. It now turns out that there are such natural finite dimensional formulations when $\mathcal{V}$ is an RKHS.

10.1.Finite span approximations of vector fields

If $\mathcal{V}$ is an RKHS, then there is a natural way to construct admissible finite dimensional sub-spaces of vector fields by considering *finite span approximations*. Let K: $\Omega \times \Omega \to \mathbb{M}_+^{n\times n}$ denote the symmetric and positive definite function that represents the reproducing kernel of $\mathcal{V}$. Next, define the *(geometric) control points* as the fixed set of points $\Sigma := \{x_1, \ldots, x_N\} \subset \Omega \subset \mathbb{R}^n$. Given such a set, define the corresponding finite dimensional vector space as

$$\mathcal{V}_\Sigma := \left\{ \nu \in \mathcal{V} : \nu(x) = \sum_{j=1}^{N} \mathsf{K}(x, x_j)\cdot\alpha_j \text{ for some } \alpha_j \in \mathbb{R}^n \right\}. \tag{29}$$

Elements in $\mathcal{V}_\Sigma$ are called finite span approximations of elements in $\mathcal{V}$ and $\mathcal{V}_\Sigma \subset \mathcal{V}$ is a finite dimensional RKHS with an inner product given by

$$\langle \nu, \eta \rangle_{\mathcal{V}_\Sigma} = \sum_{j,k=1}^{N} \alpha_j^{\top}\cdot\mathsf{K}(x_j, x_k)\cdot\beta_k \tag{30}$$

for $\nu, \eta \in \mathcal{V}_\Sigma$ given as

$$\nu = \sum_{j=1}^{N} \mathsf{K}(\cdot, x_j)\cdot\alpha_j \quad \text{and} \quad \eta = \sum_{j=1}^{N} \mathsf{K}(\cdot, x_j)\cdot\beta_j. \tag{31}$$

Remark 4. Finite span approximations of the type in (29) are especially suitable for problems involving optimisation over $\mathcal{V}$. More precisely, representer theorems (like [50, theorem 9.7], see also [41, 52]) show that

$$\nu^* := \arg\min_{\nu\in\mathcal{V}}\left[\lambda\|\nu\|_{\mathcal{V}}^2 + \sum_{j=1}^{N} L\Big(\nu(x_j)\Big)\right] \quad \Longrightarrow \quad \nu^* \in \mathcal{V}_\Sigma \ \text{ with } \ \Sigma := \{x_j\}_{j=1}^N. \tag{32}$$

Here, $L : \mathbb{R}^n \to \mathbb{R}$ is a given function (usually called loss-function) and $\|\nu\|_{\mathcal{V}}^2 := \big\langle \nu, \nu \big\rangle_{\mathcal{V}}$ is the RKHS norm.

10.1.1.Notation. Before proceeding, we introduce some convenient vector/matrix notations. The 'vec'-operator $\mathrm{vec} : \mathbb{M}^{n\times m} \to \mathbb{R}^{nm}$ stacks all the columns of a matrix, i.e.

$$\mathrm{vec}[\mathbf{A}] := \begin{pmatrix} a_{1,1} & \dots & a_{n,1} & \dots & a_{1,m} & \dots & a_{n,m} \end{pmatrix}^\top$$

whenever $\mathbf{A}$ is an $(n \times m)$ matrix

$$\mathbf{A} := \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,m} \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,m} \end{pmatrix}.$$

Next, for each $\nu \in \mathcal{V}_\Sigma$ there are unique $\alpha_1, \dots, \alpha_N \in \mathbb{R}^n$ such that

$$\nu(x) = \sum_{j=1}^{N} \mathsf{K}(x, x_j) \cdot \alpha_j \quad \text{for } x \in \Omega\,. \tag{33}$$

Hence, each $\nu \in \mathcal{V}_\Sigma$ corresponds to a unique Nn vector that we denote by $\boldsymbol{\alpha}_\nu$, i.e.

$$\boldsymbol{\alpha}_\nu := \mathrm{vec}\left[\begin{pmatrix} \alpha_1 & \cdots & \alpha_N \end{pmatrix}\right] \quad \text{where } \alpha_1, \dots, \alpha_N \in \mathbb{R}^n \text{fulfil (33).} \tag{34}$$

Finally, introducing the $(Nn \times Nn)$ kernel Gram matrix $\mathbf{K} := [\mathsf{K}(x_j, x_k)]_{j,k=1,\dots,N}$ allows us to express (30) as $\langle \nu, \eta \rangle_{\mathcal{V}_\Sigma} = \boldsymbol{\alpha}_\nu^\top \cdot \mathbf{K} \cdot \boldsymbol{\beta}_\eta$.

10.2.Finite dimensional formulation

Consider the problem of minimising the functional $\mathcal{E} : \mathcal{V} \to \mathbb{R}$ in (11). If data $g \in \mathscr{H}$ is digitised, then this functional is similar to the 'loss functional' in (32) (at least when $\mu = 0$) and a minimiser of (10) should be contained in $\mathcal{V}_\Sigma \subset \mathcal{V}$. Hence, for finite data we can replace the (infinite dimensional) minimisation in (10) with the following finite dimensional variational problem:

$$\arg\min_{\nu\in\mathcal{V}_\Sigma} \mathcal{E}(\nu). \tag{35}$$

Let us now consider (35) more closely. As already stated, to each $\nu \in \mathcal{V}_\Sigma$ there are $\alpha_1, \dots, \alpha_N \in \mathbb{R}^n$ such that (31) holds. Hence, the values of ν at the control points in $\Sigma \subset \Omega$ are expressible through the following matrix equality:

$$\begin{pmatrix} \nu(x_1) & \dots & \nu(x_N) \end{pmatrix}^\top = \mathbf{K} \cdot \boldsymbol{\alpha}_\nu.$$

In particular, the RKHS norm of a displacement field $\nu \in \mathcal{V}_\Sigma$ can be written as

$$\|\nu\|_{\mathcal{V}}^2 = \|\nu\|_{\mathcal{V}_\Sigma}^2 = \left\langle \sum_{j=1}^N \mathsf{K}(\cdot, x_j)\cdot\alpha_j, \sum_{j=1}^N \mathsf{K}(\cdot, x_j)\cdot\alpha_j \right\rangle_{\mathcal{V}_\Sigma} = \boldsymbol{\alpha}_\nu^\top \cdot \mathbf{K}\cdot\boldsymbol{\alpha}_\nu. \tag{36}$$

This motivates the introduction of the function

$$\mathsf{S}\colon \mathbb{R}^{Nn} \to \mathbb{R}_+ \quad \text{given as} \quad \mathsf{S}(\boldsymbol{\alpha}) := \boldsymbol{\alpha}^\top \cdot \mathbf{K}\cdot\boldsymbol{\alpha} \quad \text{for } \boldsymbol{\alpha}\in\mathbb{R}^{Nn}. \tag{37}$$

Then, S above corresponds to the $\nu \mapsto \|\nu\|_{\mathcal{V}}^2$ term in (11) as shown in (36). Next, define $\mathsf{W}_I : \mathbb{R}^{Nn} \to \mathscr{X}$ as

$$\mathsf{W}_I(\boldsymbol{\alpha}) := I\left(\cdot + \sum_{j=1}^N \mathsf{K}(\cdot, x_j)\cdot\alpha_j\right) \quad \text{for } \boldsymbol{\alpha} = \mathrm{vec}[(\alpha_1, \ldots, \alpha_N)] \in \mathbb{R}^{Nn}. \tag{38}$$

Then, W_I corresponds to the deformation operator $\mathcal{W}_I$ in (8), i.e. $\mathsf{W}_I(\boldsymbol{\alpha}_\nu) = \mathcal{W}_I(\nu)$ whenever $\boldsymbol{\alpha}_\nu \in \mathbb{R}^{Nn}$ and $\nu \in \mathcal{V}_\Sigma$ are related by (34). Finally, we define the functions $\mathsf{R}, \mathsf{J}_I : \mathbb{R}^{Nn} \to \mathbb{R}_+$ as

$$\begin{aligned} \mathsf{R}(\boldsymbol{\alpha}) &:= \mathcal{R}(\mathsf{W}_I(\boldsymbol{\alpha})) \\ \mathsf{J}_I(\boldsymbol{\alpha}) &:= \mathcal{D}(\mathcal{T}(\mathsf{W}_I(\boldsymbol{\alpha})), g) \end{aligned} \qquad \text{for } \boldsymbol{\alpha}\in\mathbb{R}^{Nn}. \tag{39}$$

As was to be expected, R corresponds to $\mathcal{R}$ in (16) and J_I corresponds to $\mathcal{J}_I$ in (19). Hence, solving (35) is equivalent to solving

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^{Nn}} [\lambda \mathsf{S}(\boldsymbol{\alpha}) + \mu \mathsf{R}(\boldsymbol{\alpha}) + \mathsf{J}_I(\boldsymbol{\alpha})] \quad \text{for fixed } \lambda > 0. \tag{40}$$

If $\boldsymbol{\alpha}^*$ denotes a solution to (40), then we can construct $\nu^* \in \mathcal{V}_\Sigma$ from $\boldsymbol{\alpha}^*$ by using the relations in (31) and (34):

$$\nu^* := \sum_{j=1}^N \mathsf{K}(\cdot, x_j)\cdot\alpha_j^*$$

in which the n-vector $\alpha_j^* \in \mathbb{R}^n$ contains elements j to $j+n-1$ of $\boldsymbol{\alpha}^*$. The corresponding reconstruction f^* that approximates $f_{\text{true}} \in \mathscr{X}$ in (1) is then given by

$$f^*(x) := \mathsf{W}_I(\nu^*) = I(x + \nu^*(x)) \quad \text{for all } x \in \Omega\,.$$

Computing the reconstruction f^* is associated with a variety of choices. First is the choice of the functionals $\mathcal{D} : \mathscr{H}\times\mathscr{H} \to \mathbb{R}_+$ and $\mathcal{R} : \mathscr{X} \to \mathbb{R}_+$. Next is the choice of the (regularisation) parameter $\mu \geqslant 0$. The choice of $\mathcal{D}$ is ideally dictated by the statistical model for the noise in data, and the choice of $\mu \geqslant 0$ is governed by the 'magnitude' of this noise. As an example, the 2-norm is a natural choice for $\mathcal{D}$ for additive Gaussian noise and one may use the Morozov discrepancy principle to select μ if there are reasonably accurate estimates of the noise level in data. The choice of $\mathcal{R}$ depends on the *a priori* regularity properties that one seeks to enforce during reconstruction. These choices are identical to those that need to be made in traditional variational regularisation. Our scheme does, however, differ in that there is also a component that controls the *a priori* shape information, which is given by a template $I \in \mathscr{X}$, kernel $\mathsf{K}\colon \Omega\times\Omega \to \mathbb{M}_+^{n\times n}$, the control points $\Sigma \subset \Omega$, and the (shape) regularisation parameter $\lambda \geqslant 0$.

10.3. Solving the finite dimensional minimisation problem

We solve (40) using a gradient descent method, e.g. gradient descent. Such methods involve computing the gradients of $\boldsymbol{\alpha} \mapsto \mathsf{S}(\boldsymbol{\alpha})$, $\boldsymbol{\alpha} \mapsto \mathsf{R}(\boldsymbol{\alpha})$, and $\boldsymbol{\alpha} \mapsto \mathsf{J}_I(\boldsymbol{\alpha})$. Computing the gradients

of these functions amounts to calculating their partial derivative with respect to every element $\alpha_{j,k}$ ($j = 1, \ldots, N$ and $k = 1, \ldots, n$) in the Nn vector $\boldsymbol{\alpha}$.

We will here present an example of such calculations in a special case, and for simplicity, we omit the regularisation functional $\mathcal{R} : \mathscr{X} \to \mathbb{R}_+$, i.e. we omit $\boldsymbol{\alpha} \mapsto \mathsf{R}(\boldsymbol{\alpha})$. Furthermore, we focus on the special case when $\mathscr{H} = \mathbb{R}^m$, i.e. there are m data points and $\mathcal{T}: \mathscr{X} \to \mathbb{R}^m$. The data discrepancy $\mathcal{D} : \mathscr{H} \times \mathscr{H} \to \mathbb{R}_+$ is the corresponding finite dimensional version of the squared 2-norm:

$$\mathcal{D}(u, g) := \left\| u - g \right\|_2^2 = \sum_{\iota=1}^{m} (u_\iota - g_\iota)^2 \quad \text{for } u, g \in \mathscr{H}. \tag{41}$$

10.3.1.The gradient of $\boldsymbol{\alpha} \mapsto \mathsf{S}(\boldsymbol{\alpha})$. We know from (36) that

$$\mathsf{S}(\boldsymbol{\alpha}) := \boldsymbol{\alpha}^\top \cdot \mathbf{K} \cdot \boldsymbol{\alpha} = \|\nu\|_{\mathcal{V}}^2 \quad \text{whenever } \boldsymbol{\alpha} = \boldsymbol{\alpha}_\nu.$$

Hence, for $j = 1, \ldots, N$ and $k = 1, \ldots, n$ we get

$$\frac{\partial \mathsf{S}}{\partial \alpha_{j,k}}(\boldsymbol{\alpha}) = \frac{\partial}{\partial \alpha_{j,k}} [\boldsymbol{\alpha}^\top \cdot \mathbf{K} \cdot \boldsymbol{\alpha}] \quad \Longrightarrow \quad \nabla \mathsf{S}(\boldsymbol{\alpha}) = 2\mathbf{K} \cdot \boldsymbol{\alpha} \qquad \text{for } \boldsymbol{\alpha} \in \mathbb{R}^{Nn}.$$

10.3.2.The gradient of $\boldsymbol{\alpha} \mapsto \mathsf{J}_I(\boldsymbol{\alpha})$. When $\mathcal{D} : \mathscr{H} \times \mathscr{H} \to \mathbb{R}_+$ is given by (41), then $\mathsf{J}_I : \mathbb{R}^{Nn} \to \mathbb{R}_+$ defined in (39) becomes

$$\mathsf{J}_I(\boldsymbol{\alpha}) := \mathcal{D}(\mathcal{T}(\mathsf{W}_I(\boldsymbol{\alpha})), g) = \|\mathcal{T}(\mathsf{W}_I(\boldsymbol{\alpha})) - g\|_2^2 = \sum_{\iota=1}^{m} (\mathsf{T}(\boldsymbol{\alpha})_\iota - g_\iota)^2 \tag{42}$$

where $\mathsf{T}: \mathbb{R}^{Nn} \to \mathbb{R}^m$ is defined as

$$\mathsf{T}(\boldsymbol{\alpha}) := \mathcal{T}(\mathsf{W}_I(\boldsymbol{\alpha})) = \mathcal{T}\left(I\left(\cdot + \sum_{j=1}^{N} \mathsf{K}(\cdot, x_j) \cdot \alpha_j \right) \right) \quad \text{for } \boldsymbol{\alpha} \in \mathbb{R}^{Nn}. \tag{43}$$

If $\mathcal{T}: \mathscr{X} \to \mathbb{R}^m$ is linear, then a straightforward computation gives

$$\nabla \mathsf{J}_I(\boldsymbol{\alpha}) = 2 \sum_{\iota=1}^{m} [(\mathsf{T}(\boldsymbol{\alpha})_\iota - g_\iota) \nabla \mathsf{T}(\boldsymbol{\alpha})_\iota]$$

where $\mathcal{T}_\iota : \mathscr{X} \to \mathbb{R}$ is the the ι:th coordinate of the forward operator $\mathcal{T}$, so

$$\begin{aligned} \nabla \mathsf{T}(\boldsymbol{\alpha})_\iota &= \nabla \left[\mathcal{T}_\iota \left(I\left(\cdot + \sum_{l=1}^{N} \mathsf{K}(\cdot, x_l) \cdot \alpha_l \right) \right) \right] \\ &= \mathcal{T}_\iota \left(\begin{pmatrix} \mathsf{K}(\cdot, x_1) & \ldots & \mathsf{K}(\cdot, x_N) \end{pmatrix}^\top \cdot \nabla I\left(\cdot + \sum_{l=1}^{N} \mathsf{K}(\cdot, x_l) \cdot \alpha_l \right) \right). \end{aligned}$$

11. Results for 2D tomography

To illustrate the impact of shaped-based reconstruction, we consider 2D parallel beam tomography with very sparse and highly noisy data. All the data is simulated and the phantoms are highly simplistic, but we believe it serves the purpose of illustrating the impact of using prior

shape information. An implementation applicable to the use cases mentioned in section 1.1 would need to address several other issues, some of which are discussed in section 12.2.

11.1. The inverse problem

The goal is to reconstruct a real-valued function in $\mathscr{X} = \mathscr{L}^2_c(\Omega, \mathbb{R})$ on a domain $\Omega \subset \mathbb{R}^2$ from its transmission projection images. Hence, data in (1) is represented by a real-valued function in $\mathscr{H} = \mathscr{L}^2(Y, \mathbb{R})$ where Y denotes a 2D manifold of parallel lines in $\mathbb{R}^2$ and the forward operator $\mathcal{T}$ is the 2D ray/Radon transform—see [33] for details.

11.2. The reconstruction scheme

The reconstruction is based on solving (10) where the shape template I is a given 2D image and the data discrepancy $\mathcal{D}$ is the the squared 2-norm in $\mathscr{H}$. We do not include any additional regularisation, so $\mu = 0$ and (10) reduces to

$$\min_{\nu \in \mathcal{V}} \left[\lambda \|\nu\|^2_{\mathcal{V}} + \left\| \mathcal{T}\left(I \circ \phi^{\nu} \right) - g \right\|^2_{\mathscr{L}^2(Y,\mathbb{R})} \right]. \tag{44}$$

In the above, the shape template $I \in \mathscr{X}$ and regularisation parameter $\lambda > 0$ are fixed.

The space $\mathcal{V}$ of vector fields is chosen as an RKHS where its reproducing kernel is given by a symmetric and positive definite Gaussian function $\mathsf{K}\colon \Omega \times \Omega \to \mathbb{M}^{2\times 2}_+$ that, for fixed $\sigma > 0$, is defined as

$$\mathsf{K}(x, y) := \exp\left(-\frac{1}{2\sigma^2} \|x - y\|_2 \right) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{for } x, y \in \mathbb{R}^2. \tag{45}$$

11.3. Finite dimensional formulation

To replace the infinite dimensional minimisation in (44) with a finite dimensional counterpart, we consider finite span approximations of the type in (29)

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^{2N}} \left[\lambda \left(\boldsymbol{\alpha}^{\top} \cdot \mathbf{K} \cdot \boldsymbol{\alpha} \right) + \sum_{i=1}^{l} \sum_{j=1}^{k} \left(\mathsf{T}(\boldsymbol{\alpha})_{i,j} - g_{i,j} \right)^2 \right]. \tag{46}$$

In the above, $\mathsf{T}\colon \mathbb{R}^{2N} \to \mathbb{R}^m$ is the fully discretised 2D Radon transform so m denotes the total number of lines. In parallel beam geometry, $m = lk$ where k is the number of directions and l is the number of lines per direction.

The variational problem (46) is solved using a gradient descent scheme where the gradient of the objective is computed from the expressions in section 10.3. The kernel is given in (45) with $\sigma^2 = 1$ for all cases. Other reconstruction parameters, like the value of λ and the number of iterations, vary.

11.4. The data acquisition and reconstruction protocols

Data is generated by evaluating the 2D parallel beam ray transform from a number of directions surrounding a digital phantom and then adding additive Gaussian noise. The noise level is quantified in terms of the signal-to-noise ratio (SNR) expressed using the logarithmic decibel (dB) scale:

$$\mathrm{SNR}(g) = 10\log_{10}\left[\frac{\sum_{j=1}^{m}|g_j^{\mathrm{ideal}} - \mu_{\mathrm{ideal}}|^2}{\sum_{j=1}^{m}|e_j - \mu_{\mathrm{noise}}|^2}\right]. \tag{47}$$

Here, $g = g^{\mathrm{ideal}} + e$ in (1) where $g^{\mathrm{ideal}} \in \mathbb{R}^m$ is the noise-free component of data g and e is the noise component. Furthermore, μ_{ideal} is the arithmetic average of g^{ideal} and μ_{noise} is the arithmetic average of e. Since we work with simulated data, there is access to the noise-free component of data.

Shape-based reconstructions are obtained by solving (46) using a gradient descent method. Unless otherwise stated, the RKHS kernel is the one in (45) with $\sigma^2 = 1$. Other reconstruction parameters, like the value of λ and the number of iterations, vary. Some test suites involve comparisons against other reconstruction methods, like filtered back projection (FBP), classical Tikhonov, and TV reconstruction. These are computed using routines from the Operator Discretization Library, which is available from http://github.com/odlgroup/odl

11.5. The tests

The tests seek to illustrate some basic properties of shape-based reconstruction and as such, they do not constitute a proper validation/comparison study. Despite their simplicity, these tests do provide quite interesting information that is sufficient as a proof-of-concept in a study whose main purpose is to develop a theoretical framework. More extensive tests are called for when the proposed approach for shape-based reconstruction is applied to specific imaging problems, such as in some of the use cases mentioned in section 1.1.

Test suite 1 investigates how the smoothness of the shape template influences the reconstruction. Test suite 2 considers the sensitivity of the reconstruction scheme against the choice of the regularisation parameter, whereas test suite 3 assesses the robustness against noise in data. Test suite 4 considers the impact of using an erroneous shape template. Finally, test suite 5 investigates performance in a multi-object setting.

11.5.1. Test suite 1: regularity of template. The aim is to investigate how regularity properties of the shape template influence the final reconstruction, so the test suite involves two phantoms and two shape templates with different regularity properties. The phantom in figure 1(a) has distinct (sharp) edges, the other in figure 1(b) lacks singularities. Moderately noisy (SNR 13.7 dB) data is generated from these two phantoms (not shown) and reconstructions are based on two shape templates, one in figure 1(c) with sharp edges, the other in figure 1(f) that also has corners.

The results are summarised in figure 1 and the shape-based reconstructions are nearly indistinguishable. The edge regularity is directly reflected in the reconstruction, whereas traces from corners in the template are harder to identify. This is to be expected since reconstructions are obtained by smoothly deforming the shape template, so template smoothness properties (like edges and/or corners) get transferred to the reconstruction.

It is worth commenting on how such a preservation of regularity can be utilised to obtain reconstructions that are suited to specific imaging tasks. An obvious example is when the reconstructed image is to be automatically segmented. In such a case, one may consider using a sharp template. Another more subtle example when one has *a priori* knowledge about the edge characteristics comes from ET. When imaging a cryo-fixated aqueous specimen, we expect to have isolated molecular assemblies embedded in ice. Here the shape template typically corresponds to the 3D electrostatic map of a molecular assembly with a shape similar to the structures being studied. The smoothness of this 3D electrostatic map is given by how the

electrostatic potential behaves at the interface between the molecular assembly and the aqueous buffer. The blurriness of the edge is thus a consequence of the decay of the electrostatic potential as one moves from the molecule into the buffer. This can be computed in advance, as shown in [34, 49], and it provides an idea of how blurry the edges are in the true unknown structure f_{true}.

11.5.2. Test suite 2: sensitivity w.r.t. regularisation parameter. The parameter λ in (46) is the regularisation parameter that dictates the influence of the shape information. Its choice should be dictated by the noise characteristics in data (noise level, statistical properties of noise, etc) and here we empirically investigate the sensitivity of the reconstruction w.r.t. the choice of λ at a fixed noise level.

Figure 2(b) shows the phantom and figure 2(a) is the moderately noisy (SNR 13.7dB) data. Shape reconstructions use the template in figure 2(c) and the result for varying values of λ are shown in figures 2(d)–(j). It is clear from this test that reconstructions in figures 2(d) and (e) are clearly over-regularised, i.e. the shape template has too much influence on the outcome. Once λ is small enough to allow data to have an impact, it does not seem to influence the outcome that much. In fact, setting $\lambda = 0$ gives essentially the same result, which would imply that the whole problem does not need to be regularised! This is of course not the case. There is another regularisation parameter, the RKHS kernel parameter σ in (45). When λ is too small, what remains is the effect of the choice σ.

11.5.3. Test suite 3: influence of noise. The tests here focus on robustness against noise in data, so there are four different data sets with noise levels ranging from very high (SNR $-$ 1.8 dB) to low (SNR 25.35 dB). All four data sets, shown in figures 3(a), 4(a), 5(a) and 6(a), are generated using the same phantom in figure 3(b).

All shape-based reconstructions use the shape template in figure 3(g) and we see a rather remarkable robustness against noise. In fact, only shape-based reconstruction provides a result that bears any similarity to the original phantom.

11.5.4. Test suite 4: the topology of the template. Here we seek to investigate the influence of the topology of the template. We have already seen in test suite 1 that smoothness properties of the template are essentially transferred to the reconstruction. The test suite involves two phantoms with different topologies (genus 0 and 1), shown in figures 7(a) and (b). Likewise, there are two shape templates, shown in figures 7(c) and (f) with genus 0 and 1. A similar test is conducted in figure 8 with a somewhat more complex phantom.

The results show that the topology of the template is transferred to the reconstruction. Again, this is to be expected since reconstructions are obtained by smoothly deforming the shape template, so the topology of the template is transferred to the reconstruction.

11.5.5. Test suite 5: multi-object reconstruction. The previous tests always used a phantom consisting of a single object. Here we consider the well-known Shepp–Logan phantom shown in figure 9(b), which consists of ten ellipsoids with grey-values in [0,1]. Data shown in figure 9(a) is parallel beam tomographic data along five directions that are not uniformly distributed over $[0°, 180°]$, so besides being under-sampled, data is also limited angle. The noise level is SNR 4.34 dB.

The shape-based reconstruction makes use of two templates shown in figures 9(c) and (d). Both have the same topology as the phantom, which is a prerequisite for our approach, as shown in test suite 4. The results, presented in figure 9, show that shape-based reconstruction performs fairly well.

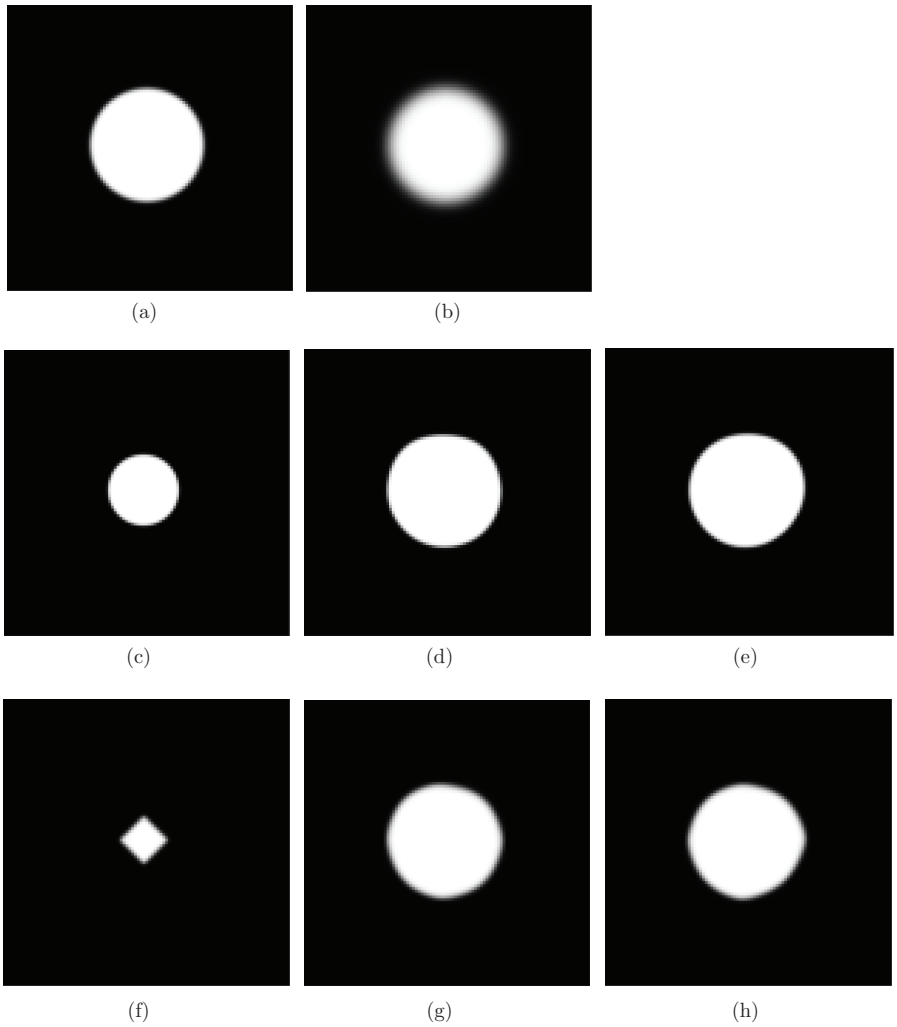

Figure 1. Regularity of template (test suite 1). Shape-based reconstructions of two 101×101 pixel phantoms from moderately noisy (SNR 13.8 dB) parallel beam 2D tomographic data. Data has three directions $0°, 45°$ and $90°$, each with 151 data points. Each phantom is reconstructed using two templates with slightly varying regularity. It is clear that regularity properties of the template have a strong influence on the reconstruction. (a) Phantom with sharp edges. (b) Phantom (smooth). (c) Template with sharp edges but no corners. (d) Reconstruction of phantom in (a) using the template in (c) (400 iterations, $\lambda = 10^{-5}$). (e) Reconstruction of phantom in (b) using the template in (c) (400 iterations, $\lambda = 10^{-5}$). (f) Template with sharp edges and corners. (g) Reconstruction of phantom in (a) using the template in (f) (1000 iterations, $\lambda = 10^{-3}$). (h) Reconstruction of phantom in (a) using the template in (f) (1000 iterations, $\lambda = 10^{-3}$).

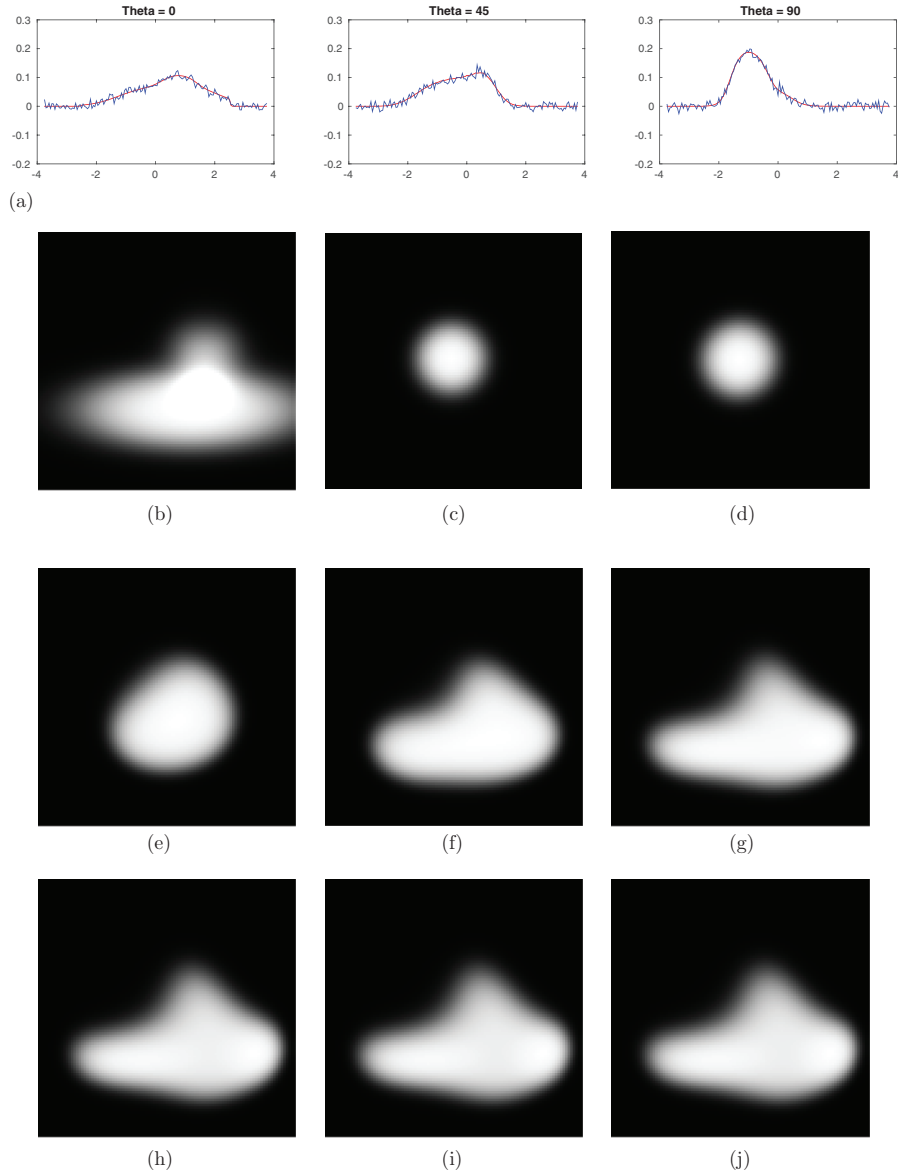

Figure 2. Sensitivity w.r.t. regularisation parameter (test suite 2). Figures (d)–(j) show shape-based reconstructions obtained from data in (a) after 400 iterations using a shape template in (c) and different values for λ. (a) Data: Parallel beam 2D tomographic data with moderate noise level (SNR 13.7 dB) generated from the phantom in (b) using three directions $0°, 45°$ and $90°$ with 151 data points for each direction. The jagged blue curve is the data used as the input for reconstruction, the red curve is the corresponding noise-free data. (b) The phantom (101×101 pixel). (c) Shape template. (d) $\lambda = 1.0$. (e) $\lambda = 10^{-1}$. (f) $\lambda = 10^{-2}$. (g) $\lambda = 10^{-3}$. (h) $\lambda = 10^{-4}$. (i) $\lambda = 10^{-5}$. (j) $\lambda = 10^{-7}$.

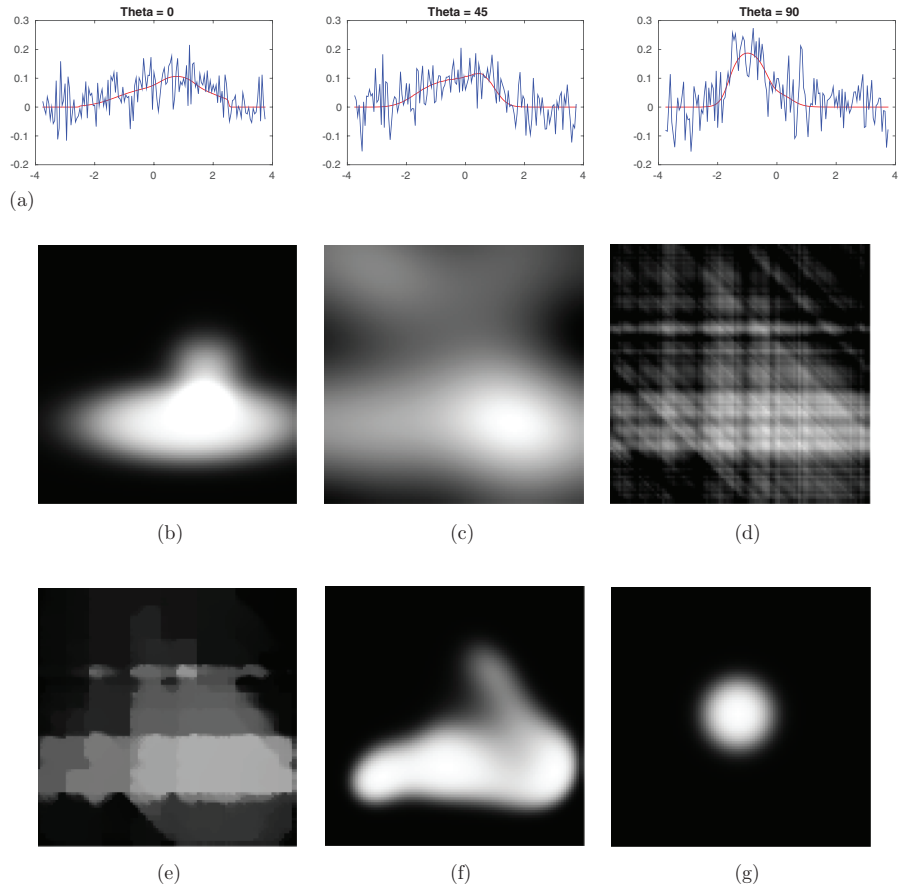

Figure 3. *Robustness against noise (test suite 3)—very high noise.* Data, shown in (a), has very high noise (SNR -1.8dB). The phantom is shown in (b) and reconstructions, obtained using different methods, are shown in (c)–(f). Only shape-based reconstruction provides a result (f) that bears any similarities to the original phantom (b). (a) Data: parallel beam 2D tomographic data with a very high noise level (SNR -1.8 dB) generated from the phantom in (b) using three directions $0°, 45°$ and $90°$ with 151 data points for each direction. The jagged blue curve is the data used as the input for reconstructions shown below, the red curve is the corresponding noise-free data. (b) The phantom, a 101×101 pixel image. (c) FBP reconstruction (frequency scaling $= 0.04$). (d) Classical Tikhonov reconstruction (4000 iterations, $\mu = 240$). (e) TV reconstruction (4000 iterations, $\mu = 60$). (f) Shape-based reconstruction using the template in (g) (400 iterations, $\lambda = 10^{-5}$). (g) The shape template.

12. Discussion

The first part of the discussion considers alternative deformations that one may consider for shape-based reconstruction. Some of these could be of interest in addressing issues related to non-uniqueness (section 8). The second part in section 12.2 deals with extensions motivated by the use cases mentioned in section 1.1. Finally, in section 12.3, we consider using shape-based reconstruction as a component in spatiotemporal imaging which was also mentioned as an important application area in section 1.1

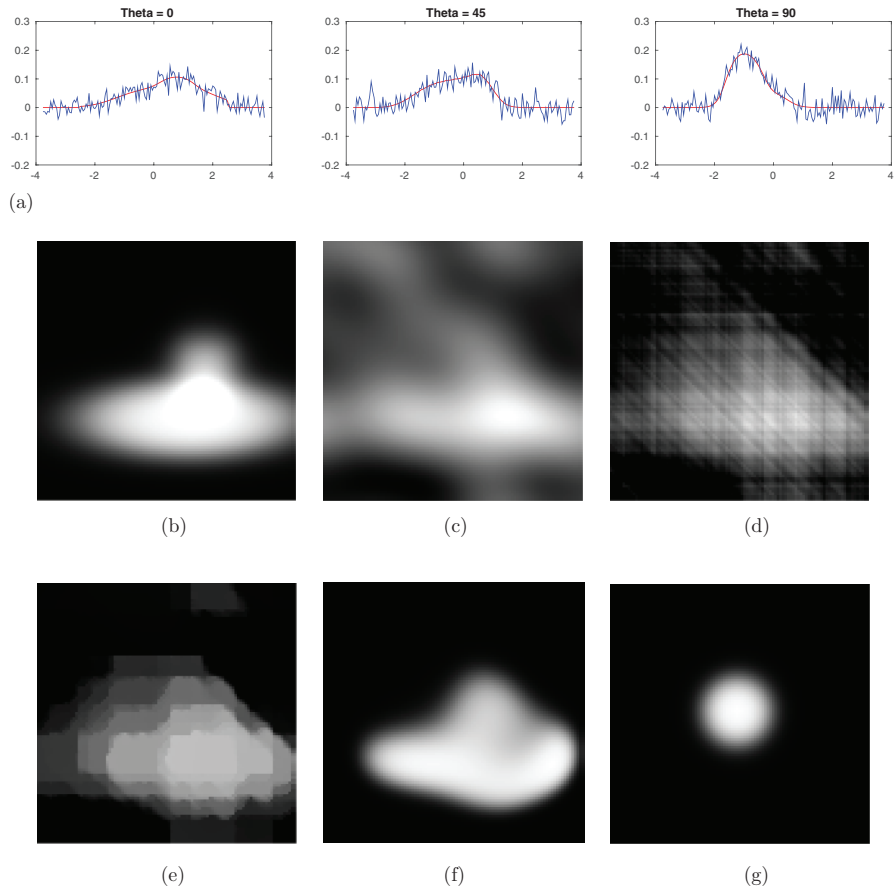

Figure 4. *Robustness against noise (test suite 3)—high noise.* Data, shown in (a), has high noise (SNR 5.39 dB). The phantom is shown in (b) and reconstructions, obtained using different methods, are shown in (c)–(f). Only shape-based reconstruction provides a result (f) that bears any similarities to the original phantom (b). (a) Data: parallel beam 2D tomographic data with high noise level (SNR 5.39 dB) generated from the phantom in (b) using three directions $0°, 45°$ and $90°$ with 151 data points for each direction. The jagged blue curve is the data used as the input for reconstructions shown below, the red curve is the corresponding noise-free data. (b) The phantom, a 101×101 pixel image. (c) FBP reconstruction (frequency scaling $= 0.06$). (d) Classical Tikhonov reconstruction (4000 iterations, $\mu = 150$). (e) TV reconstruction (4000 iterations, $\mu = 30$). (f) Shape-based reconstruction using the template in (g) (400 iterations, $\lambda = 10^{-5}$). (e)The shape template.

12.1. Alternative deformation models

As mentioned in section 5.2, linearised deformations are not necessarily invertible so $\mathscr{G}_{\mathcal{V}}$ is only a semigroup, not a group. There can be cases when the invertibility of deformations is important, such as when one seeks diffeomorphic deformations. This requires the consideration of alternative schemes for generating deformation.

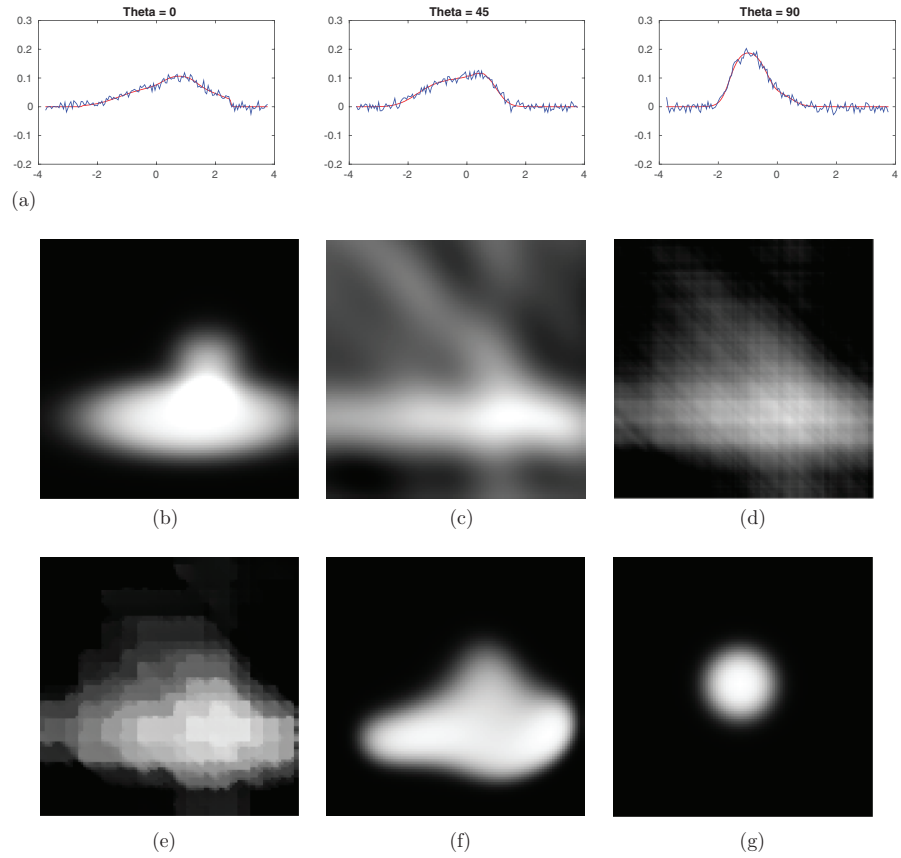

Figure 5. *Robustness against noise (test suite 3)—moderate noise.* Data, shown in (a), has a moderate noise level (SNR 13.49dB). The phantom is shown in (b) and reconstructions, obtained using different methods, are shown in (c)–(f). Only shape-based reconstruction provides a result (f) that bears any similarities to the original phantom (b). (a) Data: parallel beam 2D tomographic data with a moderate noise level (SNR 13.49 dB) generated from the phantom in (b) using three directions $0°, 45°$ and $90°$ with 151 data points for each direction. The jagged blue curve is the data used as the input for reconstructions shown below, the red curve is the corresponding noise-free data. (b) The phantom, a 101×101 pixel image. (c) FBP reconstruction (frequency scaling $= 0.08$). (d) Classical Tikhonov reconstruction (4000 iterations, $\mu = 100$). (e) TV reconstruction (4000 iterations, $\mu = 1$). (f) Shape-based reconstruction using the template in (g) (400 iterations, $\lambda = 10^{-5}$). (g) The shape template.

One such alternative is provided by the LDDMM framework [9, 19, 29, 30, 46, 47, 50] where deformations are generated by a flow equation, which is given as an ordinary differential equation with a velocity field contained in $\mathcal{V}$ at each time point. Hence, linearised deformations can be seen as a single time step within the LDDMM framework. The velocity fields need to be $\mathscr{C}^1$ in order to give $\mathscr{G}_{\mathcal{V}}$ a topological group structure. With further regularity ($\mathcal{V}$ is admissible), $\mathscr{G}_{\mathcal{V}}$ becomes a subgroup of the group of diffeomorphisms [50]. See also [7, 28] for an overview of various other groups of diffeomorphisms whose elements may serve as deformation models. Next, all of the analysis done here, including results in section 7,

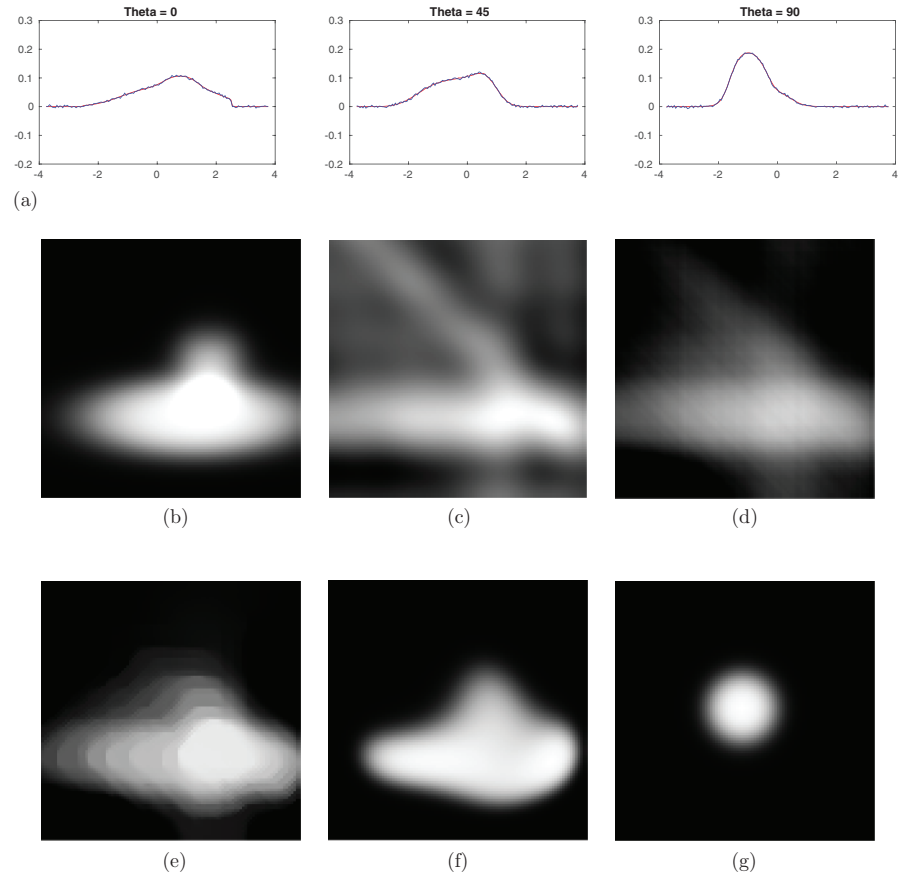

Figure 6. *Robustness against noise (test suite 3)—low noise.* Data, shown in (a), has low noise (SNR 25.35 dB). The phantom is shown in (b) and reconstructions, obtained using different methods, are shown in (c)–(f). Only shape-based reconstruction provides a result (f) that bears any similarities to the original phantom (b). (a) Data: parallel beam 2D tomographic data with a low noise level (SNR 25.35 dB) generated from the phantom in (b) using three directions $0°, 45°$ and $90°$ with 151 data points for each direction. The jagged blue curve is the data used as the input for reconstructions shown below, the red curve is the corresponding noise-free data. (b) The phantom, a 101×101 pixel image. (c) FBP reconstruction (frequency scaling $= 0.1$). (d) Classical Tikhonov reconstruction (4000 iterations, $\mu = 80$). (e) TV reconstruction (4000 iterations, $\mu = 0.1$). (f) Shape-based reconstruction using the template in (g) (400 iterations, $\lambda = 10^{-5}$). (g) The shape template.

generalise to the LDDMM setting. The expressions for the gradients in section 9 become significantly more involved and implementation is more complex. For these reasons, the theoretical and computational work using the LDDMM framework is part of a forthcoming paper.

An issue associated with the LDDMM framework is its non-parametric nature, i.e. it is difficult to decompose the velocity field generating the diffeomorphism into interpretable components. Furthermore, it is not possible to impose a certain *a priori* knowledge regarding which type of deformations to exclude. In many imaging applications, there is prior information

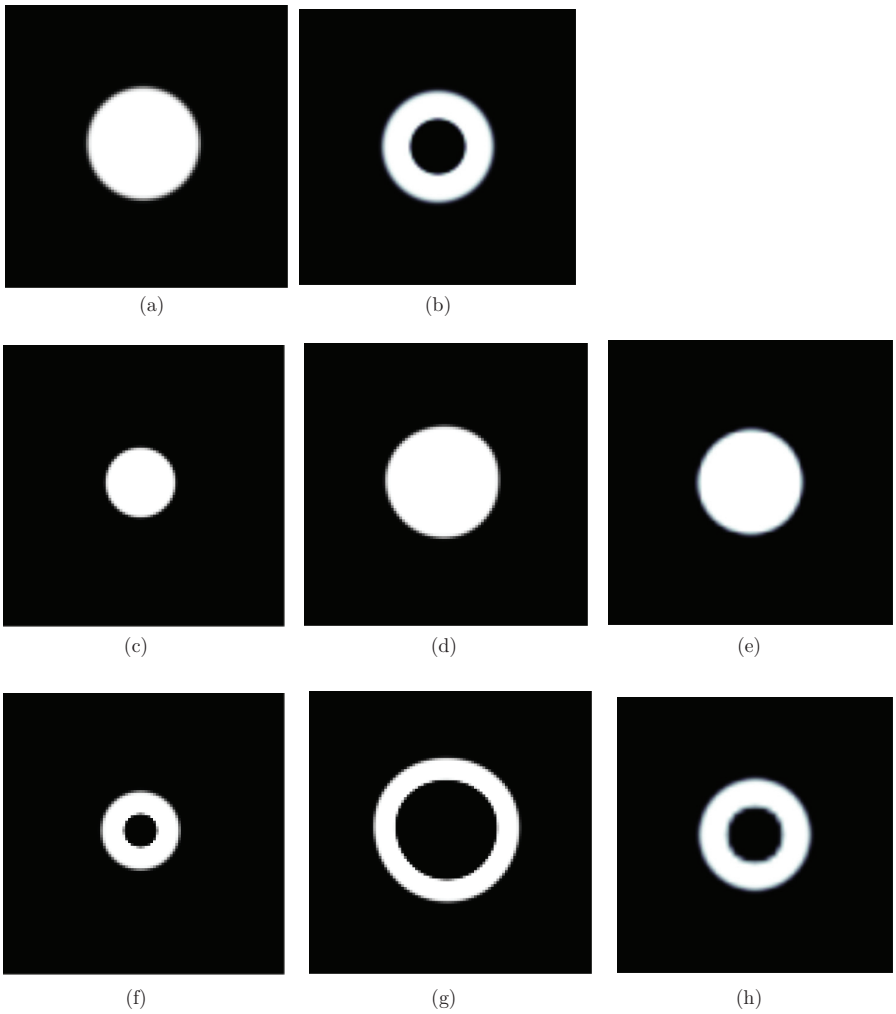

Figure 7. *Topology of phantom (test suite 4).* Shape-based reconstructions of two 101×101 pixel phantoms from moderately noisy (SNR 15.22 dB) parallel beam 2D tomographic data. The data has four directions $0°, 45°, 90°$ and $135°$ with 151 data points for each direction. Each phantom is reconstructed using two templates with varying topology. It is clear that the reconstruction inherits the same topology as the template. (a) Phantom with genus 0. (b) Phantom with genus 1. (c) Template with genus 0. (d) Reconstruction of phantom in (a) using the template in (c) (400 iterations, $\lambda = 10^{-5}$). (e) Reconstruction of phantom in (b) using the template in (c) (400 iterations, $\lambda = 10^{-5}$). (f) Template with genus 1. (g) Reconstruction of phantom in (a) using the template in (f) (600 iterations, $\lambda = 10^{-5}$). (h) Reconstruction of phantom in (b) using the template in (f) (1000 iterations, $\lambda = 10^{-7}$).

regarding which deformations are unfeasible. For this reason, a number of alternative frameworks have been developed where deformations are structured beforehand through parametrisation. One such is the poly-affine framework [4], another is the growth by random iterated diffeomorphisms model [18, 35], and yet another is the Diffeons framework [51]. It may also

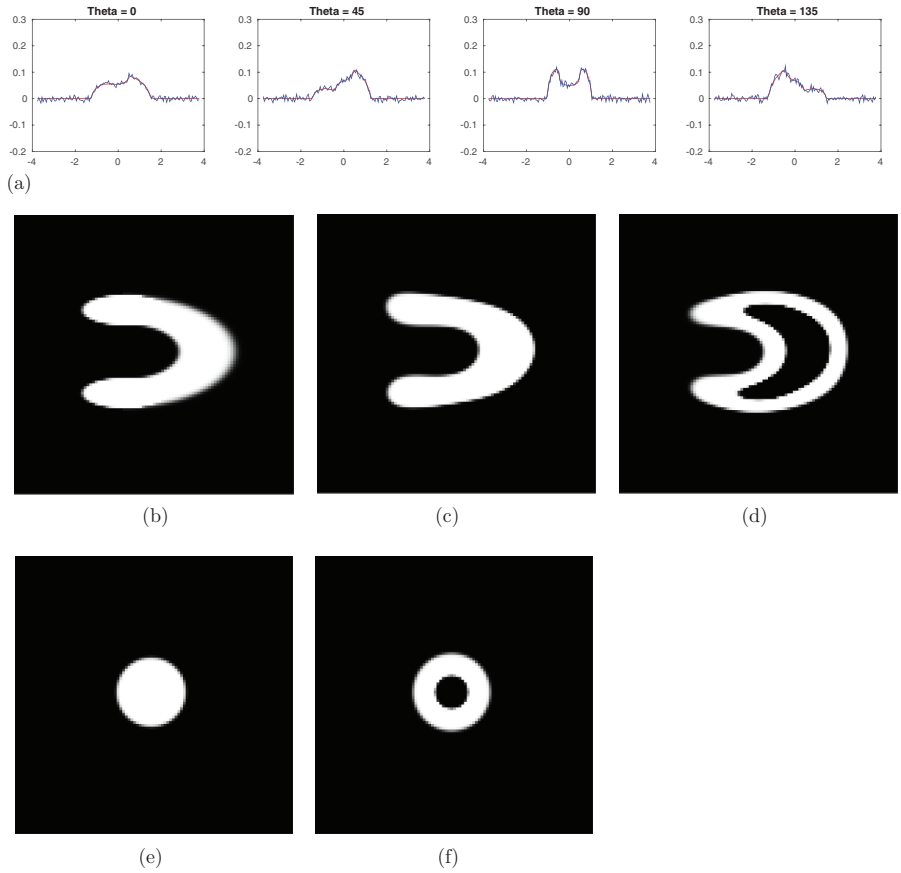

Figure 8. *Test suite 4—topology of phantom.* The phantom, shown in (b), is a 101×101 pixel image that is a 'U'-shaped object with genus 0. The parallel beam data is shown in (a), shape-based reconstructions are shown in (c) and (d) obtained using a template with genus 0 and 1, respectively. It is clear that the reconstruction inherits the topology of the template. (a) Data: parallel beam 2D tomographic data with a moderate noise level (SNR 12.95 dB) generated from the phantom in (b) using four directions $0°, 45°$, $90°$ and $135°$ with 151 data points for each direction. The jagged blue curve is the data used as the input for reconstruction, the red curve is the corresponding noise-free data. (b) A phantom with genus 0. (c) Shape-based reconstruction using the template in (e) (600 iterations, $\lambda = 10^{-5}$). (d) Shape-based reconstruction using the template in (f) (600 iterations, $\lambda = 10^{-5}$). (e) Template with genus 0. (f) Template with genus 1.

be of interest to consider such parametrised frameworks for indirect image matching (shape-based reconstruction).

12.2. Extensions motivated by use cases

The current approach for shape-based reconstruction limits its usage to imaging problems where the goal is to recover the shape of a *single* object given a template with the correct topology. Intensity variations are of little or no importance. As the tests in section 11 clearly show, under these circumstances, shape-based reconstruction performs fairly well, especially

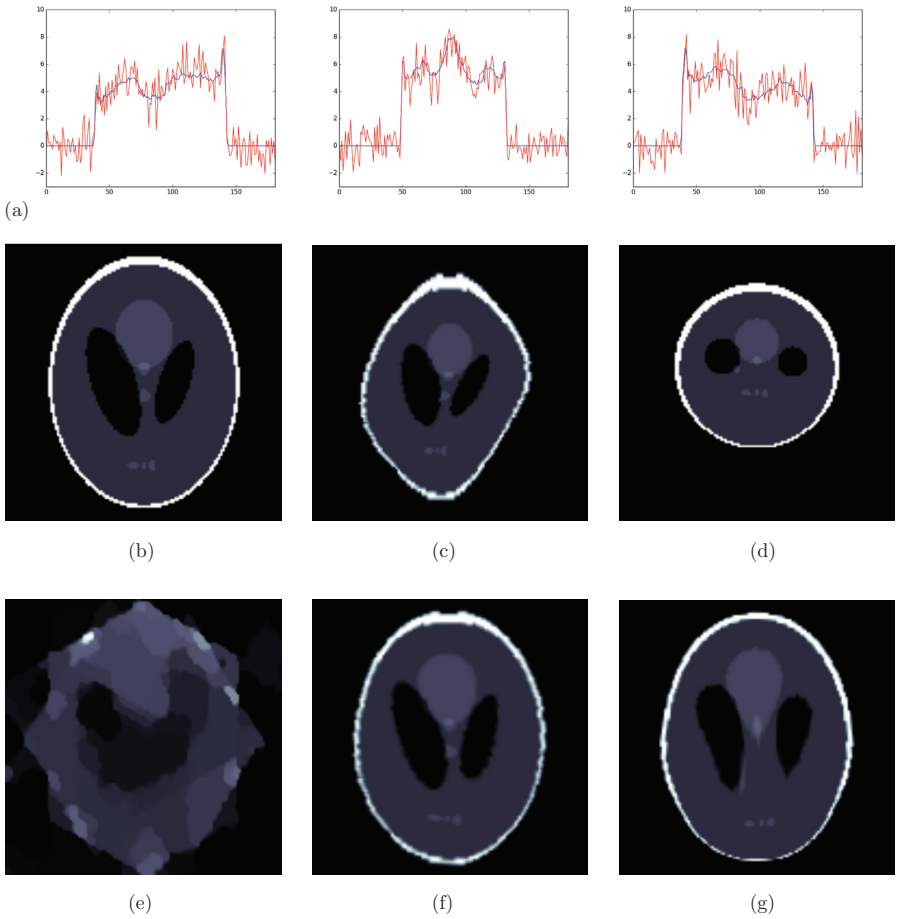

Figure 9. *Multi-object reconstruction (test suite 5).* The phantom shown in (b) is the well-known Shepp–Logan phantom. Both shape-based reconstructions are obtained from 2000 iterations using kernel parameter $\sigma = 4.0$ and $\lambda = 10^{-7}$. They are shown in (f) and (g), respectively. Note that both templates have a topology that matches the phantom, which is important, as shown in test suite 4 (figures 7 and 8). As a comparison, we show the corresponding TV reconstruction in (e). (a) Data: parallel beam 2D tomographic data with a high noise level (SNR 4.34 dB) generated from the phantom in (b) using five directions $30°, 60°, 90°, 120°$ and $150°$ with 182 data points for each direction. The images above show data for the directions $30°, 90°$ and $150°$, respectively. The jagged blue curve is the data used as the input for reconstruction, the red curve is the corresponding noise-free data. (b) The 128×128 pixel Shepp–Logan phantom. (c) Template 1. (d) Template 2. (e) TV-based reconstruction (1000 iterations, $\mu = 0.6$). (f) Shape-based reconstruction using the template in (c). (g) Shape-based reconstruction using the template in (d).

for problems where data is very sparsely, or unevenly, sampled and/or highly noisy. Such imaging problems can also be addressed using discrete tomography techniques, but an advantage of using shape-based reconstruction is that there is no need to *a priori* set the number of grey-scale levels, which is usually the case for discrete tomography [6].

The above is, however, a rather limited category of inverse problems, and the use cases mentioned in section 1.1 do not fall under this category. Hence, there is a clear need to extend the current approach for shape-based reconstruction in order to broaden its applicability. One extension aims to handle *multiple* isolated objects embedded in some background medium (section 12.2.1), another is to allow for changes in intensity (section 12.2.2).

12.2.1. Multiple objects. Many of the use cases mentioned in section 1.1, especially those related to ET, deal with imaging multiple isolated objects, each with its own shape, that are embedded in a fairly uniform background medium. Applying the current implementation for shape-based reconstruction, as in test suite 5, requires finding a single deformation that jointly fits a single known template to all these objects. This is clearly unfeasible unless there is a template with the correct topology, i.e. one containing the correct number of objects, each with their correct topology. There are very few use cases where these conditions are fulfilled, so clearly one needs to extend the current approach for shape-based reconstruction.

One approach to handle multiple objects is to apply shape deformations locally. More precisely, consider (1) and assume a user can *a priori* single out sub-domains $\Omega_1, \ldots, \Omega_N \subset \Omega$ that contain a single isolated object and assign templates $I_1, \ldots, I_N \in \mathscr{X}$ to each of these. Hence, the *a priori* information in (9) is now replaced by

$$f_{\text{true}}\Big|_{\Omega_j} = \mathcal{W}_{I_j}(\nu) + f_0 \quad \text{for some } \nu \in \mathcal{V} \text{ supported in } \Omega_j\,.$$

In the above, $f_0 \in \mathscr{X}$ is the background medium and $\mathcal{W}_{I_j}$ is given by the semigroup action in (8). Many of the ET related use cases in section 1.1 allow for such *a priori* information to be obtained along with an estimate of f_0, typically from an initial reconstruction. The shape-based reconstruction scheme extending (10) is then

$$\inf_{\nu_1,\ldots,\nu_N \in \mathcal{V}} \left[\lambda \sum_{j=1}^{N} \|\nu_j\|_{\mathcal{V}}^2 + \mu \mathcal{R}\left(\mathcal{W}_{\text{tot}}(\nu_1, \ldots, \nu_N) + f_0 \right) + \mathcal{D}\left(\mathcal{W}_{\text{tot}}(\nu_1, \ldots, \nu_N) + f_0, g \right) \right]. \tag{48}$$

Note that data $g \in \mathscr{H}$ in (1) has to be matched against the entire multi-object structure. Furthermore, $\mathcal{W}_{\text{tot}} : \mathcal{V} \times \ldots \times \mathcal{V} \to \mathscr{X}$ is the joint deformation given by sub-domains $\Omega_1, \ldots, \Omega_N \subset \Omega$ and corresponding templates $I_1, \ldots, I_N \in \mathscr{X}$ through

$$\mathcal{W}_{\text{tot}}(\nu_1, \ldots, \nu_N) := \sum_{i=1}^{N} \mathcal{W}_{I_j}(\nu_j).$$

The existence of solutions for (48) follows from the same arguments used to prove theorem 1, and issues related to non-uniqueness mentioned in section 8 also naturally persist for this case. The implementation is, however, much more difficult. A specific challenge is to numerically handle boundary information to ensure $\mathcal{W}_{I_j}(\nu_j)$ is supported in Ω_j. This is probably the main challenge in multiple object shape-based reconstruction.

12.2.2. Modifying intensity values. The deformation operator in (8) that governs how the template is modified is given by a semigroup action that does not change any intensity values, it merely moves intensity by a vector field. Hence, a shape-based reconstruction will have the same texture as its template, regardless of the data. This is not feasible for many imaging problems, like most arising in medical imaging. There are three natural ways to address this shortcoming. One may include the template as an unknown, or use a semigroup action that modifies intensities, or let the semigroup $\mathscr{S}_{\mathcal{V}}$ act on both the coordinate and the intensity.

Including the template as an unknown leads to the following extension of (10):

$$\inf_{\nu\in\mathcal{V},I\in\mathscr{X}}[\lambda\|\nu\|_{\mathcal{V}}^2+\mu\mathcal{R}(I)+\mathcal{D}(\mathcal{T}(\mathcal{W}_I(\nu)),g)].$$

An alternative, more feasible, formulation is to decouple the template and shape variables by considering the following intertwined recursive scheme:

$$\begin{cases}\nu_n := \arg\inf_{\nu\in\mathcal{V}}[\lambda\|\nu\|_{\mathcal{V}}^2+\mathcal{D}(\mathcal{T}(\mathcal{W}_{I_{n-1}}(\nu)),g)]\\ I_n := \arg\inf_{I\in\mathscr{X}}[\mu\mathcal{R}(I)+\mathcal{D}(\mathcal{T}(\mathcal{W}_I(\nu_n)),g)].\end{cases}\tag{49}$$

In the above, the step that updates the template is a 'usual' reconstruction. The difficult step is to one that updates the deformation, which in turn is precisely the shape-based reconstruction problem solved in this paper.

Regarding alternate semigroup actions, one natural alternative is to consider a mass-preserving action. This implies that the deformation operator in (8) is given by

$$\mathcal{W}_f(\nu)(x) := \left|\mathrm{Jac}(\phi^\nu)(x)\right|(f\circ\phi^\nu)(x)\quad\text{for }\nu\in\mathcal{V}\text{ and }x\in\Omega\,.\tag{50}$$

All of the analysis done here, including results in sections 7 and 9, can also be carried out for this new group action.

Finally, one may also mimic the way metamorphosis extends LDDMM by allowing deformations $\phi^\nu\in\mathscr{G}_{\mathcal{V}}$ to act simultaneously on both the shape and image intensity—see [50, Chapter 13] for further details on metamorphosis.

12.3. Connection to spatiotemporal imaging

In spatiotemporal imaging, the image we seek to reconstruct will have a temporal and a spatial component, i.e. elements in the reconstruction space $\mathscr{X}$ are functions $f:[t_0,t_1]\times\Omega\to\mathbb{R}$ where $x\mapsto f(t,x)\in\mathscr{L}^2([t_0,t_1],\mathbb{R})$ and $t\mapsto f(t,x)\in\mathscr{X}_0$, where $\mathscr{X}_0$ is some suitable Hilbert space of functions defined on Ω. Hence, $f(t,\cdot)$ is the image at time $t\in[t_0,t_1]$ and $f(\cdot,x)$ is the time evolution of an image point $x\in\Omega$. Now, it often makes sense to explicitly separate the spatial and temporal components of f. Such a separation can be achieved by introducing a time evolution operator

$$\mathcal{W}:\mathcal{V}\times\mathscr{X}_0\to\mathscr{X}\,.\tag{51}$$

Here, $\mathcal{V}$ is a fixed parameter set for the time evolution and $\mathscr{X}_0$ is the reconstruction space for the spatial component. Spatiotemporal signals are now assumed to be of the form

$$f(t,x)=\mathcal{W}(\nu,I)(t,x)\quad\text{for }x\in\Omega\text{ and }t\in[t_0,t_1]\quad\text{for given }\nu\in\mathcal{V}\text{ and }I\in\mathscr{X}_0.$$

Here, I is the template, which is the time independent spatial component of the spatiotemporal signal. The evolution parameter $\nu\in\mathcal{V}$ governs the time evolution of the template I. In this setting, the *spatiotemporal inverse problem* is to estimate *both* the true template $I^*\in\mathscr{X}_0$ and the true evolution parameter $\nu^*\in\mathcal{V}$ from data $g(t,\cdot)\in\mathscr{H}$ where

$$g(t,\cdot)=\mathcal{T}(\mathcal{W}(\nu^*,I^*)(t,\cdot))+e(t,\cdot)\quad\text{for }t\in[t_0,t_1].\tag{52}$$

Note that $\mathscr{X}_0$ is the reconstruction space for the spatial component of the spatiotemporal signals, $\mathscr{H}$ is the data space, which is common for all data across time, $\mathcal{T}:\mathscr{X}_0\to\mathscr{H}$ is the Fréchet differentiable (spatial) forward operator, and $\mathcal{W}$ in (51) is the time evolution operator

modelling the evolution, governed by the evolution parameter in $\mathcal{V}$, across time. Finally, $e(t,\cdot) \in \mathscr{H}$ are samples of independent (as t varies) $\mathscr{H}$–valued random process $\{\mathsf{E}_t\}_t$.

Now, if $\mathcal{V}$ is a normed space, then one scheme for solving the spatiotemporal inverse problem in (52) is to consider

$$\inf_{\nu\in\mathcal{V},I\in\mathscr{I}_0}\left[\lambda\|\nu\|_{\mathcal{V}}^2+\int_{t_0}^{t_1}\left\{\mu\mathcal{R}\Big(\mathcal{W}(\nu,I)(t,\,\cdot\,)\Big)+\mathcal{D}\Big(\mathcal{T}\Big(\mathcal{W}(\nu,I)(t,\,\cdot\,)\Big),g(t,\,\cdot\,)\Big)\right\}\mathrm{d}t\right] \tag{53}$$

for fixed $\lambda,\mu\geqslant 0$ and given operators $\mathcal{R}:\mathscr{I}_0\to\mathbb{R}_+$ and $\mathcal{D}:\mathscr{H}\times\mathscr{H}\to\mathbb{R}_+$. The rationale for the above scheme is based on the following assumptions:

(i) Regularity properties of image intensities can be encoded by the functional $\mathcal{R}$. These are assumed to be the same for all evolved templates $\mathcal{W}(\nu,I)(t,\,\cdot\,)\in\mathscr{I}_0$ with $t\in[t_0,t_1]$.
(ii) The stochastic process $\{\mathsf{E}_t\}_t$ modelling noise in data has elements that are independent and equally distributed so we may use the same data discrepancy functional $\mathcal{D}:\mathscr{H}\times\mathscr{H}\to\mathbb{R}_+$.

Solving (53) is quite challenging. Similar to (49), it may be simpler to consider the following intertwined recursive scheme where the evolution model is updated separately from the template:

$$\begin{cases}\nu^n:=\arg\inf_{\nu\in\mathcal{V}}\left[\lambda\|\nu\|_{\mathcal{V}}^2+\int_{t_0}^{t_1}\mathcal{D}(\mathcal{T}(\mathcal{W}(\nu,I^{n-1})(t,\,\cdot\,)),g(t,\,\cdot\,))\,\mathrm{d}t\right]\\ I^n:=\arg\inf_{I\in\mathscr{I}_0}\left[\int_{t_0}^{t_1}\{\mu\mathcal{R}(\mathcal{W}(\nu^n,I)(t,\,\cdot\,))+\mathcal{D}(\mathcal{T}(\mathcal{W}(\nu^n,I)(t,\,\cdot\,)),g(t,\,\cdot\,))\}\,\mathrm{d}t\right].\end{cases} \tag{54}$$

The above is computationally more feasible, but it is unclear how these intertwined iterates relate to a solution of (53). Furthermore, it is highly non-trivial to understand whether (53), or (54), constitutes a regularisation of the inverse problem in (52). As shown in [20], one may use the LDDMM framework (section 12.1) to define evolution operators with diffeomorphisms parametrised by admissible Hilbert spaces of vector fields $\mathcal{V}\subset\mathscr{C}_0^1(\Omega,\mathbb{R}^n)$.

Finally, the notion of 'time' above does not have to correspond to physical time. It could be a mere parametrisation of the evolution. The shape-based reconstruction we considered in (3) is now a special case with only one dataset at say $t_1=1$ (stationary data) and known template $I^*\in\mathscr{I}_0$. In such a case, the time evolution operator $\mathcal{W}$ models the (shape) deformation of the template and the goal is to estimate the evolution parameter $\nu\in\mathcal{V}$ from indirect noisy data.

Acknowledgments

The work by Ozan Öktem and Chong Chen has been supported by the Swedish Foundation for Strategic Research grant AM13-0049. Chen was also supported in part by the National Natural Science Foundation of China (NSFC) for youth under the grant 11301520 and Öktem was supported in part by the J. Tinsley Oden Faculty Fellowship from the Institute for Computational Engineering and Sciences (ICES) at The University of Texas at Austin. The research of Chandrajit Bajaj was supported in part by NIH grant R01-GM117594-0 and Sandia subcontract SNL-1439100. Finally, Pradeep Ravikumar acknowledges the support of NIH via R01 GM117594-01 and NSF via IIS-1149803.

References

[1] Amanatiadis A, Kaburlasos V G, Gasteratos A and Papadakis S E 2011 Evaluation of shape descriptors for shape-based image retrieval *IET Image Process.* **5** 493–9

[2] Amit Y and Manbeck K M 1993 Deformable template models for emission tomography *IEEE Trans. Med. Imaging* **12** 260–8

[3] Aronszajn N 1950 Theory of reproducing kernels *Trans. Am. Math. Soc.* **68** 337–404

[4] Arsigny V, Commowick O, Ayache N and Pennec X 2009 A fast and log-euclidean polyaffine framework for locally linear registration *J. Math. Imaging Vis.* **33** 222–38

[5] Bar L, Chan T F, Chung G, Jung M, Vese L A, Kiryati N and Sochen N 2015 Mumford and Shah model and its applications to image segmentation and image restoration *Handbook of Mathematical Methods in Imaging* 2nd edn, ed O Scherzer (Berlin: Springer) pp 1539–97

[6] Batenburg K J and Sijbers J 2011 DART: a practical reconstruction algorithm for discrete tomography *IEEE Trans. Image Process.* **20** 2542–53

[7] Bauer M, Bruveris M and Michor P W 2014 Overview of the geometries of shape spaces and diffeomorphism groups *J. Math. Imaging Vis.* **50** 60–97

[8] Becker F, Petra S and Schnörr C 2015 Optical flow *Handbook of Mathematical Methods in Imaging* 2nd edn, ed O Scherzer (Berlin: Springer) pp 1945–2004

[9] Beg F M, Miller M I, Trouvé A and Younes L 2005 Computing large deformation metric mappings via geodesic flow of diffeomorphisms *Int. J. Comput. Vis.* **61** 139–57

[10] Benson E, Mohammed A, Gardell J, Masich S, Czeizler E, Orponen P and Högberg B 2015 DNA rendering of polyhedral meshes at the nanoscale *Nature* **523** 441–4

[11] Berlinet A and Thomas-Agnan C 2004 *Reproducing Kernel Hilbert Spaces in Probability and Statistics* (Berlin: Springer) (doi: 10.1007/978-1-4419-9096-9)

[12] Bladt E, Pelt D M, Bals S and Batenburg K J 2015 Electron tomography based on highly limited data using a neural network reconstruction technique *Ultramicroscopy* **158** 81–8

[13] Burger M and Osher S 2013 A guide to the TV zoo *Level Set and PDE Based Reconstruction Methods in Imaging* (*Lecture Notes in Mathematics* vol 2090) ed M Burger and S Osher (Berlin: Springer) pp 1–70

[14] Cremers D 2015 Image segmentation with shape priors: explicit versus implicit representations *Handbook of Mathematical Methods in Imaging* 2nd edn, ed O Scherzer (Berlin: Springer) pp 1909–44

[15] Fidler T, Grasmair M and Scherzer O 2012 Shape reconstruction with *a priori* knowledge based on integral invariants *SIAM J. Imaging Sci.* **5** 726–45

[16] Gopinath A, Xu G, Ress D, Öktem O, Subramaniam S and Bajaj C 2012 Shape-based regularization of electron tomographic reconstruction *IEEE Trans. Med. Imaging* **31** 2241–52

[17] Grenander U 1994 *General Pattern Theory: a Mathematical Study of Regular Structures* (*Oxford Mathematical Monographs*) (Oxford: Clarendon)

[18] Grenander U 2007 On the mathematics of growth *Q. Appl. Math.* **65** 205–57

[19] Grenander U and Miller M 2007 *Pattern Theory. From Representation to Inference* (Oxford: Oxford University Press)

[20] Hinkle J, Szegedi M, Wang B, Salter B and Joshi S 2012 4D CT image reconstruction with diffeomorphic motion model *Med. Image Anal.* **16** 1307–16

[21] Hohm K, Storath M and Weinmann A 2015 An algorithmic framework for Mumford-Shah regularization of inverse problems in imaging *Inverse Problems* **31** 115011

[22] Klann E 2011 A Mumford–Shah-like method for limited data tomography with an application to electron tomography *SIAM J. Imaging Sci.* **4** 1029–48

[23] Layer T, Blaickner M, Knäusl B, Georg D, Neuwirth J, Baum R P, Schuchardt C, Wiessalla S and Matz G 2015 PET image segmentation using a Gaussian mixture model and Markov random fields *EJNMMI Phys.* **2** 1

[24] Lindeberg T 2013 Generalized axiomatic scale-space theory *Adv. Imaging Electron Phys.* **178** 1–96

[25] Louis A K 2011 Feature reconstruction in inverse problems *Inverse Problems* **27** 065010

[26] Manay S, Cremers D, Hong B-W, Yezzi A and Soatto S 2006 Integral invariants and shape matching *Statistics and Analysis of Shapes* (*Modeling and Simulation in Science, Engineering and Technology*) ed H Krim and A Yezzi (Berlin: Springer) pp 137–66

[27] Markelj P, Tomaževič D, Likar B and Pernuš F 2012 A review of 3D/2D registration methods for image-guided interventions *Med. Image Anal.* **16** 642–61

[28] Michor P W and Mumford D 2013 A zoo of diffeomorphism groups on R^n *Ann. Glob. Anal. Geom.* **44** 529–40
[29] Miller M I, Trouvé A and Younes L 2006 Geodesic shooting for computational anatomy *J. Math. Imaging Vis.* **24** 209–28
[30] Miller M I, Trouvé A and Younes L 2015 Hamiltonian systems and optimal control in computational anatomy: 100 years since D'arcy Thompson *Ann. Rev. Biomed. Eng.* **17** 447–509
[31] Mohammad-Djafari A 2009 Gauss–Markov–Potts priors for images in computer tomography resulting to joint optimal reconstruction and segmentation *Int. J. Tomogr. Stat.* **11** 76–92
[32] Murphy K P 2012 *Machine Learning: a Probabilistic Perspective* (Cambridge, MA: MIT Press)
[33] Natterer F and Wübbeling F 2001 *Mathematical Methods in Image Reconstruction* (*Mathematical Modeling and Computation*) (Philadelphia: SIAM)
[34] Öktem O 2015 Mathematics of electron tomography *Handbook of Mathematical Methods in Imaging* (*Springer Reference* vol II) 2nd edn, ed O Scherzer (New York: Springer) pp 937–1031
[35] Portman N and Vrscay E R 2011 Existence and uniqueness of solutions to the grid macroscopic growth equation *Appl. Math. Comput.* **217** 8318–27
[36] Ramlau R and Ring W 2007 A Mumford–Shah level-set approach for the inversion and segmentation of x-ray tomography data *J. Comput. Phys.* **221** 539–57
[37] Romanova M, Dahla A B, Donga Y and Hansena P C 2016 Simultaneous tomographic reconstruction and segmentation with class priors *Inverse Problems Sci. Eng.* **24** 1432–53
[38] Rumpf M and Wirth B 2015 Variational methods in shape analysis *Handbook of Mathematical Methods in Imaging* 2nd edn, ed O Scherzer (Berlin: Springer) pp 1819–58
[39] Ruthotto L and Modersitzki J 2015 Non-linear image registration *Handbook of Mathematical Methods in Imaging* 2nd edn, , ed O Scherzer (Berlin: Springer) pp 2005–51
[40] Scherzer O, Grasmair M, Grossauer H, Haltmeier M and Lenzen F 2009 *Variational Methods in Imaging* (*Applied Mathematical Sciences* vol 167) (New York: Springer)
[41] Schölkopf B, Herbrich R and Smola A J 2001 A generalized representer theorem *Computational Learning Theory Volume. Proc. of the 14th Annual Conf. on Computational Learning Theory, COLT 2001 and 5th European Conf. on Computational Learning Theory, EuroCOLT (Amsterdam, The Netherlands, 16–19 July 2001) (Lecture Notes in Computer Science* vol 2111) ed D Helmbold and B Williamson (Berlin: Springer) pp 416–26
[42] Scholkopf B and Smola A J 2001 *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* (Cambridge, MA: MIT)
[43] Schwartz L 1964 Sous-espaces Hilbertiens d'espaces vectoriels topologiques et noyaux associés (noyaux reproduisants) *J. Anal. Math.* **13** 115–256
[44] Storath M, Weinmann A, Jürgen Frikel J and Unser M 2015 Joint image reconstruction and segmentation using the Potts model *Inverse Problems* **31** 025003
[45] Thompson D W 1917 *On Growth and Form* (Cambridge: Cambridge University Press)
[46] Trouvé A 1998 Diffeomorphism groups and pattern matching in image analysis *Int. J. Comput. Vis.* **28** 213–21
[47] Trouvé A and Younes L 2015 Shape spaces *Handbook of Mathematical Methods in Imaging* 2nd edn, ed O Scherzer (Berlin: Springer) pp 1759–817
[48] Vialard F-X Computational anatomy from a geometric point of view *Presentation at the Geometry, Mechanics and Control, 8th Int. Young Researchers workshop (Barcelona, December 2013)*
[49] Vulović M, Ravelli R B G, van Vliet L J, Koster A J, Lazic I, Lücken U, Rullgård H, Öktem O and Rieger B 2013 Image formation modeling in cryo electron microscopy *J. Struct. Biol.* **183** 19–32
[50] Younes L 2010 *Shapes and Diffeomorphisms* (*Applied Mathematical Sciences* vol 171) (Berlin: Springer)
[51] Younes L 2012 Constrained diffeomorphic shape evolution *Found. Comput. Math.* **12** 295–325
[52] Yu Y-L, Cheng H, Schuurmans D and Szepesvàri C 2013 Characterizing the representer theorem *J. Mach. Learn. Res.* **28** 570–8 (*Special issue for the Proc. of the 30th Int. Conf. on Machine Learning (Atlanta, GA, USA)*
[53] Zitová B and Flusser J 2003 Image registration methods: a survey *Image Vis. Comput.* **21** 977–1000