



Published in final edited form as:

Inverse Probl. 2017 March ; 33(3): . doi:10.1088/1361-6420/aa55af.

SHAPE BASED IMAGE RECONSTRUCTION USING LINEARIZED DEFORMATIONS*

OZAN ÖKTEM[†], CHONG CHEN[‡], NEVZAT ONUR DOMANIÇ[§], PRADEEP RAVIKUMAR[§], and CHANDRAJIT BAJAJ[§]

[†]Department of Mathematics, KTH – Royal Institute of Technology, 100 44 Stockholm, Sweden

[‡]Department of Mathematics, KTH – Royal Institute of Technology, 100 44 Stockholm, Sweden and LSEC, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

[§]Department of Computer Science, University of Texas at Austin, Austin, TX 78712, USA

Abstract

We introduce a reconstruction framework that can account for shape related a priori information in ill-posed linear inverse problems in imaging. It is a variational scheme that uses a shape functional defined using deformable templates machinery from shape theory. As proof of concept, we apply the proposed shape based reconstruction to 2D tomography with very sparse measurements, and demonstrate strong empirical results.

Keywords

Tomography; shape analysis; inverse problems; regularization; reconstruction; electron tomography

AMS subject classifications

65R32; 65R30; 94A08; 94A12; 44A12; 47A52; 65F22; 92C55; 54C56; 57N25; 57R27; 37C05

1. Introduction and motivation

It is well known that utilizing a priori information is critical in addressing inverse problems where data is very noisy or the problem is highly ill-posed, with the latter often due to inappropriate sampling of data. Regularization by providing an approximate inverse is far from the best choice for solving such challenging problems. Most current approaches focus on enforcing sparsity and/or regularity within a variational setting, see [33] for a survey.

*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 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 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.

Recognizing and interpreting shapes of structures in images is essential for translating images into knowledge. Furthermore, in many imaging problems the user has a priori information about what shapes to expect, whereas noise and reconstruction artifacts are less likely to fit into the a priori shape information. To illustrate this, consider electron tomography (ET) that uses an electron microscope to image the internal 3D structural arrangement at nano-scale of a specimen. Imaging by ET leads to a severely ill-posed inverse problem with incomplete and highly noisy data [27]. However, scientists often have an idea of the rough overall shape of the structures they seek to image in ET. As an example, consider a biological specimen containing antibodies that we seek to image. Here, one knows beforehand that antibodies are approximately “Y”-shaped. In a similar way, if the specimen contains virions that we seek to image, then often one has a priori shape information about these virions, *e.g.*, they can be “near-spherical” with certain symmetry (like icosahedral) or they may form helical rod-like structures. This can often be seen in data.

To summarize, it makes sense to account for qualitative a priori shape information while performing the reconstruction in imaging. It is clear that enforcing an exact spatial match between sub-structures and a given template is asking for too much since realistic shape information is almost always approximate as in ET. More concretely, consider imaging near-spherical viruses embedded in aqueous buffer by ET. It would be erroneous to perform reconstruction in ET while enforcing that these sub-structures are spheres. Instead, a more natural approach is to perform reconstruction assuming the structures are ‘near-spherical’ in some sense. This highlights the challenge associated with accounting for a priori shape information, namely to properly quantify shape similarity.

2. Survey of the field

The idea of accounting for shape information in reconstruction is gaining increasing interest within the inverse problems community. This is part of an ongoing development in which the reconstruction and feature extraction steps are combined when solving ill-posed inverse problems. Features are in this context specific structures in images relevant for the interpretation. Feature extraction is usually part of image analysis and powerful tools have been developed over the years to interpret images.

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 minimizing 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 [3]. This approach was applied to 2D inversion of the ray transform in [28] 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

generalized and applied to very noisy limited angle and region of interest tomography data in [16]. See also [15, 35] 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 [26, 17, 29] 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 [19]. Here the approximate inverse method is extended to the setting where one recovers the features directly. It assumes 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 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.* [21, 1] and [40, Chapter 3]. There is in fact an axiomatic treatment for constructing invariant image features [18] useful for recognizing objects from different viewpoints and whose numerical values are equal or only moderately affected by basic image transformations. Returning to image reconstruction, [10] shows how such integral invariants can be used in reconstruction. The idea is to encode the a priori 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 proof of existence of minimizers of the corresponding goal functional. The optimization is performed with a gradient descent approach that requires differentiating the goal functional and the actual numerical implementation is based on minimizing a smooth approximation functional, which converges (Γ -convergence) to the original functional.

Another approach for variational image reconstruction that includes a priori shape information is given in [11]. 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 (5). 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 spaces (RKHSs). Furthermore, [2] contains no analysis of existence or uniqueness.

3. Contribution of the paper

Our approach is heavily inspired by the deformable templates framework where shape is modeled as a deformation of a template. Shape based reconstruction can then be seen as an image matching problem in which a shape template in the image domain is matched against a given target representing indirect observations (data) in the imaging problem. There is by now a rich literature for variational methods for image matching, see, *e.g.*, [5, 32, 31, 8]. Some parts of this theory is also used for reconstruction, most notably for spatiotemporal imaging as outlined in subsection 1.6. On the other hand, the usage for shape based reconstruction is yet to be explored more systematically.

This paper initiates such a study and the starting point is linearized deformations, which are among the simplest forms of deformations. Our main mathematical result is a proof of existence of minimizers. As with most other variational schemes for shape based reconstruction, we cannot expect uniqueness in general due to lack of convexity. We also explicitly calculate the shape derivatives of the functionals, both in the continuum setting and when the objective functional is discretized through finite span approximations of linearized deformations contained in a reproducing kernel Hilbert space. The latter can be of interest in actual numerical implementation since minimizers of objective functionals in the finite dimensional setting will in this case also give a minimizer to the objective functionals in the infinite dimensional setting (Remark 4). We also show the impact of using a priori shape information. The experiments strongly point to the improvement of using shape based information in reconstruction. Furthermore, they also show that the a priori shape information doesn't have to be that accurate, which is important for the applications targeted by this paper. None of these claims are however proved formally by mathematical theorems. Finally, our variational framework for shape based reconstruction also extends to other deformation models that allow for large deformations, like the one suggested within the large deformation diffeomorphic metric mapping (LDDMM) framework [36, 6, 24, 13, 40, 37, 25] and variants based on relating the LDDMM to optimal transport for a better handling of grey scale values [20, 30]. These result in smooth deformations that preserve the topology and shape based reconstruction using such, more elaborate, deformation models left for future studies.

We conclude with a short overview of the paper. Section 4 introduces terminology and notation needed for variational reconstruction of inverse problems. Section 5 provides a brief

introduction to shape theory based on the deformable templates framework, and in particular linearized deformations. Section 6 formulates the shape based reconstruction scheme and in section 7 we prove existence. section 8 contains the gradient calculations of the functionals involved in the variational formulation and section 9 is the corresponding gradients expressed when the linearized deformation is contained in a RKHS. section 10 shows some examples of how the shape based reconstruction scheme performs followed by conclusions given in section 11.

4. Inverse problems, ill-posedness and variational regularization

The goal in image reconstruction is to estimate some spatially distributed quantity (*image*) from indirect noisy observations (*measured data*). There are various ways to formalize such a statement and our starting point is the classical (non-statistical) viewpoint on inverse problems.

Image reconstruction as an inverse problem: Recover/reconstruct an *image* $f_{\text{true}} \in \mathcal{X}$ from *data* $g \in \mathcal{H}$ where

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

Here, \mathcal{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, \mathcal{H} (*data space*) is the vector space of all possible data, which for digitized data is a subset of \mathbb{R}^m . Furthermore, $\mathcal{T}: \mathcal{X} \rightarrow \mathcal{H}$ (*forward operator*) models how a given image gives rise to data in absence of noise and measurement errors, and $e \in \mathcal{H}$ (*data noise component*) is a sample of a \mathcal{H} -valued random element E whose probability distribution is (*data noise model*) assumed to be known.

Ill-posedness: A naive approach at recovering 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, modeling errors, *e.t.c.*). This is commonly addressed by relaxing the notion of a solution by considering

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

The mapping $\mathcal{D}: \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}_+$ (*data discrepancy functional*) quantifies the data misfit and a natural candidate is to choose it as a suitable affine transformation of the log likelihood of data. In such case, solving (2) amounts to finding maximum likelihood (ML) solutions.

The above approach 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. These are

the two main issues addressed by means of regularization, *i.e.* to introduce uniqueness and stability by making use of a priori knowledge about f_{true} in the reconstruction.

Variational regularization: This is a class of regularization methods that have gained much attention lately, and especially so for imaging problems [33]. The idea is to add a penalization to (2):

$$\min_{f \in \mathcal{X}} \mu \mathcal{R}(f) + \mathcal{D}(\mathcal{T}(f), g) \quad \text{given } \mu \geq 0. \quad (3)$$

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

5. Shape theory

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 modeling shapes, and the approach we consider is based on *deformable templates*. This approach, which traces back to work pursued in 1917 by D’Arcy Thompson [9], 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 [12] 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 linearized deformations framework.

The starting point is to define a set of *deformable objects* that represent the objects/structures whose shape we seek to analyze. We also define a set of *deformations* that can act on these deformable objects. In our case, a deformation is given as a diffeomorphic perturbation of the identity map (linearized 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 \mathcal{M} of deformable objects are 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 $\mathcal{M} \subset \mathcal{L}^2(\Omega, \mathbb{R})$.

5.2. Set of deformations – Linearized deformations

The set $\mathcal{G}_{\mathcal{V}}$ of deformations will consist of transformations that map the image domain $\Omega \subset \mathbb{R}^n$ into \mathbb{R}^n . These “deform” elements in \mathcal{M} simply by transforming the underlying coordinate grid.

It is natural to be able to reverse and concatenate deformations, which amounts to requiring that $\mathcal{G}_{\mathcal{V}}$ forms a *group* under the group law given by composition of functions and with the identity mapping as the identity element. The group structure also allows us to use the group action, which is a mapping $\mathcal{G}_{\mathcal{V}} \times \mathcal{M} \rightarrow \mathcal{M}$ formally denoted by $(\phi, f) \mapsto \phi.f$, as a canonical way to define how a deformation in $\mathcal{G}_{\mathcal{V}}$ transforms a deformable object in \mathcal{M} . Furthermore, alongside the group structure, we may also require that $\mathcal{G}_{\mathcal{V}}$ is a vector space over \mathbb{R} where the vector space addition and scalar multiplication is given as

$$\begin{aligned} (\phi+\psi)(x) &:= \phi(x) + \psi(x) \\ (\alpha\phi)(x) &:= \alpha\phi(x) \end{aligned} \quad (4)$$

for $\phi, \psi \in \mathcal{G}_{\mathcal{V}}$ and $\alpha \in \mathbb{R}$.

We now consider *linearized deformations* that are given as

$$\phi^{\nu} := \text{Id}_{\Omega} + \nu \quad \text{for some } \nu \in \mathcal{V}. \quad (5)$$

Here, $\mathcal{V} \subset \mathcal{C}_0^1(\Omega, \mathbb{R}^n)$ is a fixed Hilbert space with norm $\|\cdot\|_{\mathcal{V}}$. The vector fields (*deformation vector fields*) in \mathcal{V} are supported in Ω and have a derivative that tends to zero as x tends to infinity.

Remark 1: *The set $\mathcal{C}_0^1(\Omega, \mathbb{R}^n)$ is the Banach space of continuously differentiable vector fields on $\Omega \subset \mathbb{R}^n$ that along with its derivative vanishes on Ω and at infinity. Note also that these vector fields trivially extended to all of \mathbb{R}^n simply by setting them to zero outside Ω .*

Given the Hilbert space $\mathcal{V} \subset \mathcal{C}_0^1(\Omega, \mathbb{R}^n)$, define the set $\mathcal{G}_{\mathcal{V}} \subset \mathcal{C}_0^1(\Omega, \mathbb{R}^n)$ as

$$\mathcal{G}_{\mathcal{V}} := \{ \phi: \Omega \rightarrow \Omega: \phi = \phi^{\nu} \text{ for some } \nu \in \mathcal{V} \text{ and } \phi^{\nu} \text{ is given by (5)} \}. \quad (6)$$

If \mathcal{V} is a vector space closed under composition, then $\mathcal{G}_{\mathcal{V}}$ is also closed under composition. Elements defined as (5) are however not necessarily invertible unless ν is sufficiently small and regular [40, Proposition 8.6]. In conclusion, $\mathcal{G}_{\mathcal{V}}$ forms a group only when the linearized deformations in the vector space \mathcal{V} are small.

Remark 2: *If one seeks to ensure that deformations in (5) do not create foldings or holes, then they need to preserve the topology. This naturally leads to the requirement that such*

deformations be diffeomorphisms. In our setting the issue is that deformations in (5) are not necessarily invertible. If they are, then $\mathcal{G}_{\mathcal{V}}$ becomes a group of diffeomorphisms and elements in \mathcal{V} have to be small. To define deformations that are diffeomorphisms without necessarily being small requires considering alternative schemes for generating the deformations. One such approach is provided by LDDMM framework where deformations are generated by an ordinary differential equation with a time dependent velocity field that is contained in \mathcal{V} at each time point. If \mathcal{V} has some basic regularity properties, then $\mathcal{G}_{\mathcal{V}}$ becomes a subgroup to the group of diffeomorphisms, see [40] for the details.

5.3. The deformation operator and shape similarity

Elements in $\mathcal{G}_{\mathcal{V}}$ represent deformations that act on deformable objects, which in our case are grey scale images on Ω represented by square integrable functions $f: \Omega \rightarrow \mathbb{R}$. One way to let a deformation, *i.e.*, element $\phi \in \mathcal{G}_{\mathcal{V}}$, act on a grey scale image f is by a group action, say $\phi.f := f \circ \phi$. Next, each element in $\mathcal{G}_{\mathcal{V}}$ has a corresponding element in \mathcal{V} , *i.e.*, $\phi = \phi^{\nu}$ for some $\nu \in \mathcal{V}$. Hence, elements in \mathcal{V} acts on deformable objects by means of the *deformation operator* $\mathcal{W}_f: \mathcal{V} \rightarrow \mathcal{X}$ given as

$$\mathcal{W}_f(\nu) := \phi^{\nu}.f = f \circ \phi^{\nu} \quad \text{for } \nu \in \mathcal{V} \quad (7)$$

where $\phi^{\nu} \in \mathcal{G}_{\mathcal{V}}$ is given by (5). 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 $\|\nu\|_{\mathcal{V}}^2$. Note that it is zero only if the deformation is the identity mapping (no deformation) and a large value means ν is large, which in turn corresponds to a large deformation.

6. Shape based reconstruction

Consider the inverse problem in (1) where elements in \mathcal{X} represent grey scale images defined over some fixed image domain $\Omega \subset \mathbb{R}^n$, *i.e.*, $\mathcal{M} \subset \mathcal{X} \subset \mathcal{L}^2(\Omega, \mathbb{R})$. Next, assume that the true image f_{true} is approximately a deformation of a known *shape template* $I \in \mathcal{M}$. As we shall see next, the shape functional in subsection 5.3 allows us to use such a priori shape information in the reconstruction of f_{true} in (1).

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

$$f_{\text{true}} = \mathcal{W}_I(\nu) \quad \text{on } \Omega \text{ for some } \nu \in \mathcal{V}. \quad (8)$$

To reconstruct f_{true} in (1), while accounting for a priori shape information given by (8), we formulate the following variational problem:

$$\inf_{\nu \in \mathcal{V}} \left[\lambda \|\nu\|_{\mathcal{V}}^2 + \mu \mathcal{R}(\mathcal{W}_I(\nu)) + \mathcal{D}(\mathcal{T}(\mathcal{W}_I(\nu)), g) \right]. \quad (9)$$

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

7. Mathematical analysis

The goal of a mathematical analysis of a reconstruction scheme is to assess if it is a regularization scheme with key formal properties. The first and most important property is whether the reconstruction scheme always has solutions (existence of solutions), since otherwise the scheme is not suitable for reconstruction. Two following properties are concerned with uniqueness and stability, *i.e.*, whether the scheme renders a unique solution for given data and whether it is stable w.r.t. perturbations in data. Finally, the convergence properties of the scheme are concerned with whether the output from the scheme should converge to a (least-squares/maximum likelihood) solution of (2) when the data error (difference between measured data g and ideal data $\mathcal{T}(f_{\text{true}})$) tends to zero in some norm while reconstruction parameters are chosen accordingly.

We will here only consider the key property of existence, and will follow this analysis by some remarks regarding uniqueness. The starting point is to reformulate the variational problem in (9) as an optimization over \mathcal{V} . Then we prove existence of solutions for this latter equivalent optimization, which is now an optimization over \mathcal{V} . The motivation for this approach is that variational problems over \mathcal{V} are much easier to analyze than those over the group $\mathcal{G}_{\mathcal{V}}$, mainly due to the natural Hilbert space structure of \mathcal{V} . This also has numerical consequences. It is also difficult to perform an optimization over $\mathcal{G}_{\mathcal{V}}$ since it is difficult to parametrize elements in $\mathcal{G}_{\mathcal{V}}$. As we shall see, elements in \mathcal{V} are more easy to parametrize, especially when \mathcal{V} is a RKHS.

7.1. Existence

Here we consider existence of solutions to the reconstruction method given by the variational problem in (9).

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 \mathcal{X} . Then, (9) has a solution in \mathcal{V} .*

Proof: Let $\mathcal{E}: \mathcal{V} \rightarrow \mathbb{R}_+$ denote the goal functional in (9), *i.e.*, for given $\lambda, \mu = 0$ and $I \in \mathcal{X}$, it is defined as

$$\mathcal{E}(\nu) := \lambda \|\nu\|_{\mathcal{V}}^2 + \mu \mathcal{R}(\mathcal{W}_I(\nu)) + \mathcal{D}(\mathcal{T}(\mathcal{W}_I(\nu)), g) \quad \text{with } \nu \in \mathcal{V}. \quad (10)$$

If we can prove that \mathcal{E} is lower semi-continuous on \mathcal{V} , then (9) has solutions.

Let $\{\nu_n\}_n \subset \mathcal{V}$ be a minimizing sequence for (9) so $\|\nu_n\|_{\mathcal{V}}$ are bounded. \mathcal{V} is a Hilbert space, so 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 [40, Theorem A.16]). By [40, Proposition A.15] we then have

$$\|\nu^*\|_{\mathcal{V}} \leq \liminf_{n \rightarrow \infty} \|\nu_n\|_{\mathcal{V}}. \quad (11)$$

Next, by (5) 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

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

Thus, ϕ^{ν} is continuous in the weak topology of \mathcal{V} . Furthermore, for every shape template $I \in \mathcal{M}$, using the result in the proof of [38, Theorem 2.5] gives us \limsup

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

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 \mathcal{X} . Hence,

$$\mathcal{R}(\mathcal{W}_I(\nu^*)) \leq \liminf_{n \rightarrow \infty} \mathcal{R}(\mathcal{W}_I(\nu_n)), \quad (12)$$

$$\mathcal{D}(\mathcal{T}(\mathcal{W}_I(\nu^*)), g) \leq \liminf_{n \rightarrow \infty} \mathcal{D}(\mathcal{T}(\mathcal{W}_I(\nu_n)), g). \quad (13)$$

(11) together with (12)–(13) implies that ν^* is a solution for the variational problem (9) in \mathcal{V} . This concludes the proof.

7.2. Uniqueness

Even though (9) always has a solution (Theorem 1), we do not have uniqueness. The problem is that there may be multiple solutions even when both

$$f \rightarrow \mathcal{R}(f) \quad \text{and} \quad f \rightarrow \mathcal{D}(\mathcal{F}(f), g) \quad (14)$$

are strictly convex. The reason is that the deformation operator $\mathcal{W}_f: \mathcal{V} \rightarrow \mathcal{X}$ in (7) introduces non-convexity. It is furthermore a non-trivial task to work out conditions on \mathcal{V} that would guarantee convexity.

8. Gradient calculations

Performing reconstruction following the scheme outlined in section 6 requires solving the optimization problem in (9). 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 (10). The gradient is calculated w.r.t. the Hilbert structure of \mathcal{V} , which is the natural inner-product space for this minimization. In computing the \mathcal{V} -gradient of the mappings

$$\nu \mapsto \mathcal{R}(\mathcal{W}_I(\nu)) \quad \text{and} \quad \nu \mapsto \mathcal{D}(\mathcal{F}(\mathcal{W}_I(\nu)), g), \quad (15)$$

we assume the mappings in (14) are both Gâteaux differentiable and the template I is differentiable.

8.1. The deformation operator

We here provide an explicit expression for the Gâteaux derivative of the deformation operator defined in (7).

Proposition 2: *Assume $I \in \mathcal{X}$ is differentiable. Then, the deformation operator $\mathcal{W}_f: \mathcal{V} \rightarrow \mathcal{X}$ in (7) is Gâteaux differentiable at $\nu \in \mathcal{V}$ and its Gâteaux derivative is*

$$\partial \mathcal{W}_I(\nu)(\eta)(x) = \langle \nabla I(\phi^\nu(x)), \eta(x) \rangle_{\mathbb{R}^n} \quad \text{for } x \in \Omega \text{ and } \eta \in \mathcal{V}. \quad (16)$$

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

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

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

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

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

$$\partial \mathcal{W}_I(\nu)(\eta)(x) = \left\langle \nabla I(\phi^\nu(x)), \frac{d}{d\varepsilon} (\phi^{\nu + \varepsilon\eta}) \Big|_{\varepsilon=0}(x) \right\rangle_{\mathbb{R}^n} \quad \text{for } x \in \Omega. \quad (17)$$

The second term in the scalar product on the right hand side of (17) is the derivative of a deformation with respect to variations in the associated vector field.

Next, by (5) we have $\phi^{\nu + \varepsilon\eta} = \text{Id}_\Omega + \nu + \varepsilon\eta$, so

$$\frac{d}{d\varepsilon} (\phi^{\nu + \varepsilon\eta}) \Big|_{\varepsilon=0}(x) = \frac{d}{d\varepsilon} (x + \nu(x) + \varepsilon\eta(x)) \Big|_{\varepsilon=0} = \eta(x).$$

Now, (16) follows from inserting the above into (17).

Remark 3: Unless derivatives are to be interpreted in the distribution sense, the template $I \in \mathcal{X}$ has to be differentiable if the expression in (16) is to make sense.

8.2. The $\|\cdot\|_{\mathcal{V}}^2$ -norm

The \mathcal{V} -gradient of $\nu \mapsto \|\nu\|_{\mathcal{V}}^2$ is simply 2ν .

8.3. General matching functionals

Both of the functionals in (15) are of the form $\mathcal{J}_I: \mathcal{V} \rightarrow \mathbb{R}_+$ with

$$\mathcal{J}_I(\nu) := \mathcal{L} \circ \mathcal{W}_I(\nu) \quad \text{for } \nu \in \mathcal{V}, \quad (18)$$

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

We will here assume that $\mathcal{V} \subset \mathcal{C}_0^1(\Omega, \mathbb{R}^n)$ is admissible. In most cases it will be an RKHS with a continuous positive definite reproducing kernel $\mathcal{K}: \Omega \times \Omega \rightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ represented by the continuous positive definite function $K: \Omega \times \Omega \rightarrow \mathbb{R}_+^{n \times n}$, so

$$\langle \eta, \nu \rangle_{\mathcal{L}^2(\Omega, \mathbb{R}^n)} = \langle \nu, \int_{\Omega} \mathcal{K}(\cdot, x)(\eta(x)) \, dx \rangle_{\mathcal{V}}. \quad (19)$$

Theorem 3: Assume $\mathcal{L}: \mathcal{X} \rightarrow \mathbb{R}$ is Gâteaux differentiable on \mathcal{X} and define $\mathcal{J}_I: \mathcal{V} \rightarrow \mathbb{R}$ as in (18) where $I \in \mathcal{L}^2(\Omega, \mathbb{R})$ is differentiable. Then, the Gâteaux derivative of \mathcal{J}_I is given as

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

Furthermore, let $\mathcal{X} \subset \mathcal{L}^2(\Omega, \mathbb{R})$ and \mathcal{V} is a RKHS with a reproducing kernel represented by the symmetric and positive definite function $K: \Omega \times \Omega \rightarrow \mathbb{R}^{n \times n}$. Then

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

Proof: The proof of (20) follows directly from the chain rule and Proposition 2. More precisely, by the chain rule we have

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

Combining this with (16) gives us (20).

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

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

Combining (22) with (23) gives

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

Next, the expression in (16) can be inserted into (24):

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

The last equality makes use of the fact that the inner product in $\mathcal{L}^2(\Omega, \mathbb{R}^n)$ (square integrable \mathbb{R}^n -valued functions) is expressible as the inner product in \mathbb{R}^n followed by the inner product in $\mathcal{L}^2(\Omega, \mathbb{R})$ (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 $\mathcal{L}^2(\Omega, \mathbb{R}^n)$ inner product is well defined. Finally, we use the assumption that \mathcal{V} is a RKHS with reproducing kernel $\mathcal{K}: \Omega \times \Omega \rightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$, so by (19) we have

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

where $\tilde{\nu}: \Omega \rightarrow \mathbb{R}^n$ 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 $K: \Omega \times \Omega \rightarrow \mathbb{R}^{n \times n}_+$ representing the reproducing kernel gives us

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

The last equality is (21), so this concludes the proof of Theorem 3.

8.3.1. \mathcal{L}^2 -data discrepancy—Here we consider the special case when $\mathcal{H} = \mathcal{L}^2(Y, \mathbb{R})$ with Y denoting some manifold that introduces coordinates for the measured data. If the data noise component $e \in \mathcal{H}$ is a sample of a \mathcal{H} -valued Gaussian random element with mean zero and known co-variance operator, then a natural data discrepancy functional $\mathcal{D}: \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}$ in (9) is then given by

$$\mathcal{D}(u) := \|u - g\|_{\mathcal{L}^2(Y, \mathbb{R})}^2 = \int_Y (u(y) - g(y))^2 dy \quad \text{for } u, g \in \mathcal{H}. \quad (25)$$

Define $\mathcal{L}(f) := \mathcal{D}(\mathcal{T}(f))$ and consider \mathcal{J}_I in (18), *i.e.*,

$$\mathcal{J}_I(\nu) := \mathcal{L} \circ \mathcal{W}_I(\nu) = \|\mathcal{T}(I \circ \phi^\nu) - g\|_{\mathcal{L}^2(Y, \mathbb{R})}^2. \quad (26)$$

We now seek the Gâteaux derivative of \mathcal{J}_I in (26) assuming $\mathcal{X} \subset \mathcal{L}^2(\Omega, \mathbb{R})$ is a Hilbert space with the usual \mathcal{L}^2 -structure and the forward operator $\mathcal{T}: \mathcal{X} \rightarrow \mathcal{H}$ in (9) is Gâteaux differentiable. First, the Gâteaux derivative of \mathcal{L} and its corresponding gradient are given as

$$\begin{aligned}\partial\mathcal{L}(f)(h) &= 2\langle\partial\mathcal{T}(f)^*(\mathcal{T}(f)-g), h\rangle_{\mathcal{X}} \quad \text{for } f, h \in \mathcal{X}. \\ \nabla\mathcal{L}(f) &= 2\partial\mathcal{T}(f)^*(\mathcal{T}(f)-g)\end{aligned}\quad (27)$$

Here, “*” denotes the Hilbert space adjoint and \mathcal{T} is the Gâteaux derivative of \mathcal{F} . When \mathcal{T} is linear, then $\mathcal{T}(f) = \mathcal{T}$, so (27) simplifies to

$$\begin{aligned}\partial\mathcal{L}(f)(h) &= 2\langle\mathcal{T}^*(\mathcal{T}(f)-g), h\rangle_{\mathcal{X}} \\ \nabla\mathcal{L}(f) &= 2\mathcal{T}^*(\mathcal{T}(f)-g).\end{aligned}\quad (28)$$

Corollary 4: *Let the assumptions in Theorem 3 hold. Furthermore, assume that $\mathcal{J}_I: \mathcal{V} \rightarrow \mathbb{R}$ is given as in (26) with $\mathcal{T}: \mathcal{X} \rightarrow \mathcal{H}$ Gâteaux differentiable. Then*

$$\partial\mathcal{J}_I(\nu)(\eta) = 2\langle\partial\mathcal{T}(\mathcal{W}_I(\nu))^*(\mathcal{T}(\mathcal{W}_I(\nu)-g)(\cdot)\nabla I(\phi^\nu(\cdot)), \eta(\cdot))\rangle_{\mathcal{L}^2(\Omega, \mathbb{R}^n)} \quad (29)$$

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

$$\nabla\mathcal{J}_I(\nu)(x) = 2\int_{\Omega}\partial\mathcal{T}(\mathcal{W}_I(\nu))^*(\mathcal{T}(\mathcal{W}_I(\nu)-g)(y)K(x, y) \cdot \nabla I(\phi^\nu(y)) \, dy. \quad (30)$$

Proof: The proof of (29) follows directly from inserting the expression for \mathcal{L} in (27) into (20). Similarly, (30) follows directly from inserting the expression for $\nabla\mathcal{L}$ in (27) into (21).

8.4. Gradient of goal functional

It is common to make use of a gradient method in the actual minimization of the goal functional in (10). If we assume \mathcal{V} is a RKHS, then a general expression for its gradient $\nabla\mathcal{E}: \mathcal{V} \rightarrow \mathcal{V}$ can be obtained by combining the results in subsection 8.2 and (21). A common special case is however when $\mu = 0$ and \mathcal{D} is given as in (25). Then, (30) gives an explicit expression for the gradient at $x \in \Omega$ as

$$\nabla\mathcal{E}(\nu)(x) = 2\nu(x) + 2\int_{\Omega}\partial\mathcal{T}(\mathcal{W}_I(\nu))^*(\mathcal{T}(\mathcal{W}_I(\nu)-g)(y)K(x, y) \cdot \nabla I(\phi^\nu(y)) \, dy. \quad (31)$$

9. Finite dimensional setting

\mathcal{V} is in general an infinite dimensional Hilbert space of vector fields, so the variational problem in (9) is an optimization over an infinite dimensional Hilbert space. In a computational setting we can only minimize functionals over finite dimensional vector

spaces. One approach is to formulate the optimization scheme for finding a local minima in the infinite dimensional setting. The explicit expressions for the gradient of the goal functional (section 8) can be used for this. One can then discretize this optimization 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 is such natural finite dimensional formulations when \mathcal{V} is a RKHS.

9.1. 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 $K:\Omega \times \Omega \rightarrow \mathbb{R}^{n \times n}_+$ 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

$$\Sigma := \{x_1, \dots, x_N\} \subset \Omega \subset \mathbb{R}^n.$$

Given such a set, define the corresponding finite dimensional vector space as

$$\mathcal{V}_\Sigma := \left\{ \nu \in \mathcal{V} : \nu(x) = \sum_{j=1}^N K(x, x_j) \cdot \alpha_j \text{ for some } \alpha_j \in \mathbb{R}^n \right\}. \quad (32)$$

Elements in \mathcal{V}_Σ are called finite span approximations of elements in \mathcal{V} .

Then $\mathcal{V}_\Sigma \subset \mathcal{V}$ is a finite dimensional RKHS with an inner product given by

$$\begin{aligned} \langle \nu, \eta \rangle_{\mathcal{V}_\Sigma} &:= \left\langle \sum_{j=1}^N K(\cdot, x_j) \cdot \alpha_j, \sum_{k=1}^N K(\cdot, x_k) \cdot \beta_k \right\rangle_{\mathcal{V}} \\ &= \sum_{j,k=1}^N \langle K(\cdot, x_j) \cdot \alpha_j, K(\cdot, x_k) \cdot \beta_k \rangle_{\mathcal{V}} = \sum_{j,k=1}^N \alpha_j^\top \cdot K(x_j, x_k) \cdot \beta_k \end{aligned} \quad (33)$$

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

$$\nu = \sum_{j=1}^N K(\cdot, x_j) \cdot \alpha_j \quad \text{and} \quad \eta = \sum_{j=1}^N K(\cdot, x_j) \cdot \beta_j. \quad (34)$$

Remark 4: *Finite span approximations of the type in (32) are especially suitable for problems involving optimization over \mathcal{V} . This is because minimizers of objective functionals*

that include typical loss-functions are expressible as finite span approximations. More precisely, representer theorems (like [40, Theorem 9.7], see also [34, 41]) show that

$$\nu^* := \arg \min_{\nu \in \mathcal{V}} \left[\lambda \|\nu\|_{\mathcal{V}}^2 + \sum_{j=1}^N L(\nu(x_j)) \right] \Rightarrow \nu^* \in \mathcal{V}_{\Sigma}. \quad (35)$$

Here, $L: \mathbb{R}^n \rightarrow \mathbb{R}$ is a given “loss” function, $\|\nu\|_{\mathcal{V}}^2 := \langle \nu, \nu \rangle_{\mathcal{V}}$ is the RKHS norm, and $x_j \in \Sigma$ is a fixed set of N control points.

Before proceeding, we introduce some convenient vector/matrix notation. For each $\nu \in \mathcal{V}_{\Sigma}$ there are unique $\alpha_1, \dots, \alpha_N \in \mathbb{R}^n$ such that

$$\nu(x) = \sum_{j=1}^N K(x, x_j) \cdot \alpha_j \quad \text{for } x \in \Omega. \quad (36)$$

Hence, each $\nu \in \mathcal{V}_{\Sigma}$ corresponds uniquely to an Nn vector given by

$$\alpha_{\nu} := \text{vec} \left[\begin{pmatrix} \alpha_1 & \cdots & \alpha_N \end{pmatrix} \right] \quad (37)$$

where $\alpha_1, \dots, \alpha_N \in \mathbb{R}^n$ fulfill (36). Finally, define also \mathbf{K} as the $(Nn \times Nn)$ kernel Gram matrix given as

$$\mathbf{K} = [K(x_j, x_k)]_{j,k=1,\dots,N}.$$

The above notion allows us to express (33) as

$$\langle \nu, \eta \rangle_{\mathcal{V}_{\Sigma}} = \alpha_{\nu}^{\top} \cdot \mathbf{K} \cdot \beta_{\eta}.$$

9.2. The variational problem in the finite dimensional setting

We now consider the problem of minimizing the functional $\mathcal{E} : \mathcal{V} \rightarrow \mathbb{R}$ in (10). If data $g \in \mathcal{H}$ is digitized, then this functional is similar to the “loss functional” in (35) (at least when $\mu = 0$) and a minimizer of (9) should be contained in $\mathcal{V}_{\Sigma} \subset \mathcal{V}$. Hence, for finite data we can replace the (infinite dimensional) minimization in (9) with the following finite dimensional variational problem:

$$\arg \min_{\nu \in \mathcal{V}_\Sigma} \mathcal{E}(\nu). \quad (38)$$

Let us now consider (38) 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 (34) holds. Hence, the values of ν at the control points in $\Sigma \subset \Omega$ is expressible through the following matrix equality:

$$\left(\nu(x_1) \quad \dots \quad \nu(x_N) \right)^\top = \mathbf{K} \cdot \boldsymbol{\alpha}_\nu.$$

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

$$\|\nu\|_{\mathcal{V}}^2 = \|\nu\|_{\mathcal{V}_\Sigma}^2 = \left\langle \sum_{j=1}^N \mathbf{K}(\cdot, x_j) \cdot \alpha_j, \sum_{j=1}^N \mathbf{K}(\cdot, x_j) \cdot \alpha_j \right\rangle_{\mathcal{V}_\Sigma} = \boldsymbol{\alpha}_\nu^\top \cdot \mathbf{K} \cdot \boldsymbol{\alpha}_\nu. \quad (39)$$

This motivates introducing the function

$$\mathbf{S}: \mathbb{R}^{Nn} \rightarrow \mathbb{R}_+ \quad \text{given as} \quad \mathbf{S}(\boldsymbol{\alpha}) := \boldsymbol{\alpha}^\top \cdot \mathbf{K} \cdot \boldsymbol{\alpha} \quad \text{for } \boldsymbol{\alpha} \in \mathbb{R}^{Nn}. \quad (40)$$

Then, \mathbf{S} above corresponds to the $\nu \mapsto \|\nu\|_{\mathcal{V}}^2$ term in (10) as shown in (39). Next, define $\mathbf{W}_I: \mathbb{R}^{Nn} \rightarrow \mathcal{X}$ as

$$\mathbf{W}_I(\boldsymbol{\alpha}) := I \left(\cdot + \sum_{j=1}^N \mathbf{K}(\cdot, x_j) \cdot \alpha_j \right) \quad \text{for } \boldsymbol{\alpha} = \text{vec}[(\alpha_1, \dots, \alpha_N)] \in \mathbb{R}^{Nn}. \quad (41)$$

Then, \mathbf{W}_I corresponds to the deformation operator \mathcal{W}_I in (7), i.e., $\mathbf{W}_I(\boldsymbol{\alpha}_\nu) = \mathcal{W}_I(\nu)$ whenever $\boldsymbol{\alpha}_\nu \in \mathbb{R}^{Nn}$ and $\nu \in \mathcal{V}_\Sigma$ are related by (37). Finally, we define the functions $\mathbf{R}, \mathbf{J}_I: \mathbb{R}^{Nn} \rightarrow \mathbb{R}_+$ as

$$\begin{aligned} \mathbf{R}(\boldsymbol{\alpha}) &:= \mathcal{R}(\mathbf{W}_I(\boldsymbol{\alpha})) \\ \mathbf{J}_I(\boldsymbol{\alpha}) &:= \mathcal{D}(\mathcal{T}(\mathbf{W}_I(\boldsymbol{\alpha})), g) \end{aligned} \quad \text{for } \boldsymbol{\alpha} \in \mathbb{R}^{Nn}. \quad (42)$$

As to be expected, \mathbf{R} corresponds to \mathcal{R} in (15) and \mathbf{J}_I corresponds to \mathcal{J}_I in (18). Hence, solving (38) is equivalent to solving

$$\min_{\alpha \in \mathbb{R}^{Nn}} [\lambda S(\alpha) + \mu R(\alpha) + J_I(\alpha)] \quad \text{for fixed } \lambda > 0. \quad (43)$$

If α^* denotes a solution to (43), then we can construct $\nu^* \in \mathcal{V}_\Sigma$ from α^* by using the relations in (34) and (37):

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

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

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

Remark 5—*The reconstruction f^* depends on a variety of choices. First are the functionals $\mathcal{D}: \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}_+$ and $\mathcal{R}: \mathcal{X} \rightarrow \mathbb{R}_+$. How to choose these, and the (regularization) parameter $\mu > 0$, is a common topic in variational regularization. More precisely, the choice of \mathcal{D} is dictated by the data noise model and choice of $\mu > 0$ is governed by the “magnitude of the noise” in data. As an example, (25) is a natural choice for \mathcal{D} for additive Gaussian noise and if there is a reasonably accurate estimate of the noise level, then one may use the Morozov discrepancy principle to select μ . Furthermore, the choice of \mathcal{R} depends on what a priori regularity properties that one seeks to enforce on the reconstruction. The part in the scheme suggested here that differs from traditional variational regularization is the parts that controls the a priori shape information, which are the choice of template $I \in \mathcal{X}$, kernel $K: \Omega \times \Omega \rightarrow \mathbb{R}_+^{n \times n}$, the control points $\Sigma \subset \Omega$, and the (shape) regularization parameter $\lambda > 0$.*

9.3. Solving the finite dimensional minimization problem

We solve (43) using a gradient descent method, e.g., gradient descent. Such methods involves computing the gradients of

$$\alpha \mapsto S(\alpha), \quad \alpha \mapsto R(\alpha) \quad \text{and} \quad \alpha \mapsto J_I(\alpha).$$

Computing gradients of the above functions amounts to calculating the partial derivative of them with respect to every element $\alpha_{j,k}$ ($j = 1, \dots, N$ and $k = 1, \dots, n$) in the Nn vector α .

We will here present an example of such calculations in a special case. In doing so, we omit the regularization functional $\mathcal{R}: \mathcal{X} \rightarrow \mathbb{R}_+$, i.e., we omit $\alpha \mapsto R(\alpha)$. Furthermore, we focus on the special case when $\mathcal{H} = \mathbb{R}^m$, i.e., there are m data points and $\mathcal{T}: \mathcal{X} \rightarrow \mathbb{R}^m$. The data discrepancy $\mathcal{D}: \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}_+$ is the corresponding finite dimensional version of (25):

$$\mathcal{D}(u) := \|u - g\|_2^2 = \sum_{\iota=1}^m (u_\iota - g_\iota)^2 \quad \text{for } u, g \in \mathcal{H}. \quad (44)$$

9.3.1. The gradient of $\alpha \mapsto \mathbf{S}(\alpha)$ —We know from (39) that

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

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

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

9.3.2. The gradient of $\alpha \mapsto \mathbf{J}_I(\alpha)$ —When $\mathcal{D}: \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}_+$ is given by (44), then $\mathbf{J}_I: \mathbb{R}^{Nn} \rightarrow \mathbb{R}_+$ defined in (42) becomes

$$\mathbf{J}_I(\alpha) := \mathcal{D}(\mathcal{T}(W_I(\alpha)), g) = \|\mathcal{T}(W_I(\alpha)) - g\|_2^2 = \sum_{\iota=1}^m (\mathbf{T}(\alpha)_\iota - g_\iota)^2 \quad (45)$$

where $\mathbf{T}: \mathbb{R}^{Nn} \rightarrow \mathbb{R}^m$ is defined as

$$\mathbf{T}(\alpha) := \mathcal{T}(W_I(\alpha)) = \mathcal{T} \left(I \left(\cdot + \sum_{j=1}^N \mathbf{K}(\cdot, x_j) \cdot \alpha_j \right) \right) \quad \text{for } \alpha \in \mathbb{R}^{Nn}. \quad (46)$$

Then

$$\frac{\partial \mathbf{J}_I}{\partial \alpha_{j,k}}(\alpha) = \frac{\partial}{\partial \alpha_{j,k}} \left[\sum_{\iota=1}^m (\mathbf{T}(\alpha)_\iota - g_\iota)^2 \right] = 2 \sum_{\iota=1}^m \left[(\mathbf{T}(\alpha)_\iota - g_\iota) \frac{\partial}{\partial \alpha_{j,k}} \mathbf{T}(\alpha)_\iota \right]$$

for $j = 1, \dots, N$ and $k = 1, \dots, n$. A more explicit expression for the above partial derivatives requires one to calculate

$$\frac{\partial}{\partial \alpha_{j,k}} \mathbf{T}(\alpha)_\iota := \frac{\partial}{\partial \alpha_{j,k}} \left[\mathcal{T} \left(I \left(\cdot + \sum_{l=1}^N \mathbf{K}(\cdot, x_l) \cdot \alpha_l \right) \right)_\iota \right]$$

for $j = 1, \dots, N$, $k = 1, \dots, n$, and $\iota = 1, \dots, m$. The above can be calculated through the chain rule and the resulting expression will involve derivatives of the forward model \mathcal{F} and the template I . Below, we do this calculation for linear forward operators.

Let $\mathcal{T}_\iota: \mathcal{X} \rightarrow \mathbb{R}$ denote the ι :th coordinate of the forward operator $\mathcal{F}: \mathcal{X} \rightarrow \mathbb{R}^m$, which we now assume is linear. Then

$$\mathbf{T}(\boldsymbol{\alpha})_\iota = \mathcal{T}_\iota \left(I \left(\cdot + \sum_{l=1}^N \mathbf{K}(\cdot, x_l) \cdot \alpha_l \right) \right) \quad \text{for } \boldsymbol{\alpha} \in \mathbb{R}^{Nn} \text{ and } \iota = 1, \dots, m.$$

For linear \mathcal{T}_ι we get

$$\begin{aligned} \frac{\partial}{\partial \alpha_{j,k}} \mathbf{T}(\boldsymbol{\alpha})_\iota &= \frac{\partial}{\partial \alpha_{j,k}} \left[\mathcal{T}_\iota \left(I \left(\cdot + \sum_{l=1}^N \mathbf{K}(\cdot, x_l) \cdot \alpha_l \right) \right) \right] \\ &= \mathcal{T}_\iota \left(\frac{\partial}{\partial \alpha_{j,k}} \left[I \left(\cdot + \sum_{l=1}^N \mathbf{K}(\cdot, x_l) \cdot \alpha_l \right) \right] \right) \\ &= \mathcal{T}_\iota \left(\nabla I \left(\cdot + \sum_{l=1}^N \mathbf{K}(\cdot, x_l) \cdot \alpha_l \right)^\top \cdot (\mathbf{K}(\cdot, x_j))_k \right) \\ &= \mathcal{T}_\iota \left((\mathbf{K}(\cdot, x_j))_k^\top \cdot \nabla I \left(\cdot + \sum_{l=1}^N \mathbf{K}(\cdot, x_l) \cdot \alpha_l \right) \right). \end{aligned}$$

Note that $(\mathbf{K}(\cdot, x_j))_k$ is the k :th column of the matrix $\mathbf{K}(\cdot, x_j)$. Hence,

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

where

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

10. Results for 2D tomography

10.1. The inverse problem

In the continuum setting, image reconstruction in 2D tomography is the inverse problem of reconstructing $f_{\text{true}} \in \mathcal{X}$ from (1) where data $\mathbf{g} = \mathcal{F}(f)$ and \mathcal{F} is the 2D ray/Radon transform:

$$\mathcal{F}(f)(\ell) := \int_\ell f(x) \, d\mu_\ell(x) \quad \text{for } f \in \mathcal{X} \text{ and } \ell \subset \mathbb{R}^2 \text{ is a line.}$$

Here, $d\mu_\ell$ is the 1-dimensional Lebesgue measure on the line ℓ

\mathcal{X} is some suitable Hilbert space of real valued functions defined on $\Omega \subset \mathbb{R}^2$, for our purposes it is enough to require that $\mathcal{X} \subset \mathcal{L}_c^2(\Omega, \mathbb{R})$. Moreover, in the continuum setting, tomographic data is a real-valued function $g: Y \rightarrow \mathbb{R}$ defined on some sub-manifold Y of lines in \mathbb{R}^2 that correspond to the tomographic data acquisition geometry.

It is possible to introduce explicit coordinates on the manifold Y . Any line ℓ in \mathbb{R}^2 can be described by a direction $\omega \in \mathcal{S}^1$ and a point $x \in \omega^\perp$ that it passes through, *i.e.*, $\ell: t \mapsto t\omega + x$. Since $\omega = (\sin\theta, \cos\theta)$ for some $0 \leq \theta < \pi$ and $x = p(-\cos\theta, \sin\theta)$ for some $p \in \mathbb{R}$, we can describe any line ℓ in \mathbb{R}^2 by a pair $(\theta, p) \in [0, \pi] \times \mathbb{R}$ so that

$$\ell: t \mapsto t(\sin\theta, \cos\theta) + p(-\cos\theta, \sin\theta).$$

Now, for parallel beam geometry data acquisition (*i.e.*, all lines are parallel for a given angle θ), $Y = \Gamma \times \mathbb{R}$ for some subset $\Gamma \subset [0, \pi]$ and

$$\mathcal{T}(f)(\theta, p) := \int_{-\infty}^{\infty} f(t(\sin\theta, \cos\theta) + p(-\cos\theta, \sin\theta)) dt$$

for $(\theta, p) \in Y$. Furthermore, we also have that

$$\partial \mathcal{T}(u)^*(g)(x) = \int_{\Gamma} g(\theta, x \cdot (-\cos\theta, \sin\theta)) d\theta \quad \text{for } x \in \Omega.$$

Remark 6: For digitized data, Y is a finite set of pairs (θ, p) where

$$\theta \in \{\theta_1, \dots, \theta_k\} \subset [0, \pi] \quad \text{and} \quad p \in \{p_1, \dots, p_l\} \subset \mathbb{R}.$$

Hence, in this setting we would have k directions and l measurements for each direction, *i.e.*, data is a vector $g \in \mathbb{R}^m$ where $m = kl$ is the total number of data points.

10.2. The reconstruction scheme

We will estimate $f_{\text{true}} \in \mathcal{X}$ by shape based reconstruction, *i.e.*, by minimizing $\mathcal{E}: \mathcal{V} \rightarrow \mathbb{R}$ in (10) with $\mu = 0$ and \mathcal{D} as in (25), *i.e.*,

$$\min_{v \in \mathcal{V}} \left[\lambda \|v\|_{\mathcal{V}}^2 + \|\mathcal{T}(I \circ \phi^\nu) - g\|_{\mathcal{L}^2(Y, \mathbb{R})}^2 \right]. \quad (47)$$

In the above, the shape template $I \in \mathcal{X}$ and regularization parameter $\lambda > 0$ is fixed.

Next, the space of vector fields \mathcal{V} is chosen as the RKHS with a reproducing kernel represented by the following symmetric and positive definite Gaussian function $K: \Omega \times \Omega \rightarrow \mathbb{R}_+^{2 \times 2}$ that, for fixed $\sigma > 0$, is given as

$$K(x, y) := \exp\left(-\frac{1}{2\sigma^2} \|x-y\|_2\right) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{for } x, y \in \mathbb{R}^2. \quad (48)$$

Considering finite span approximations of the type in (32) allows us to recast the minimization in (47) as the finite dimensional variational problem (43) with $\mu = 0$ and $J_I: \mathbb{R}^{2N} \rightarrow \mathbb{R}_+$ as in (45), i.e.,

$$\min_{\alpha \in \mathbb{R}^{2N}} \left[\lambda(\alpha^\top \cdot \mathbf{K} \cdot \alpha) + \sum_{i=1}^l \sum_{j=1}^k \left(T(\alpha)_{i,j} - g_{i,j} \right)^2 \right] \quad (49)$$

where $T: \mathbb{R}^{2N} \rightarrow \mathbb{R}^m$ with $m = lk$ is defined as

$$T(\alpha)_{i,j} = \int_{-\infty}^{\infty} I[t(\sin\theta_i, \cos\theta_i) + p_j(-\cos\theta_i, \sin\theta_i) + \sum_{\tau=1}^N K(t(\sin\theta_i, \cos\theta_i) + p_j(-\cos\theta_i, \sin\theta_i), x_\tau) \cdot \alpha_\tau] dt$$

for $i = 1, \dots, k$ and $j = 1, \dots, l$.

10.3. Results

This section illustrates some results of using shaped based reconstruction in 2D parallel beam tomography with very sparse data. Even though these tests are not exhaustive enough to constitute a formal validation study, they do give an indication of the impact of using a priori shape information.

10.3.1. Phantoms and data acquisition protocol—All phantoms reside in a fixed image domain given as $\Omega = [-2.5, 2.5] \times [-2.5, 2.5]$ that is digitized into 101×101 pixels. Phantoms are either step functions or smoothed versions of step functions. The smoothing is done by applying a Gaussian kernel function. Hence, smoothed phantoms may present some grey-scale variations also in regions away from the edges, but in general the phantoms are rather homogenous and do not present any texture information.

Data is generated by evaluating the 2D parallel beam ray transform from equally distributed directions surrounding the aforementioned phantoms and then adding noise. All tests except for test suite 4 use three directions: 0° , 45° and 90° . In test suite 4 there are four directions. Each direction samples 151 parallel lines uniformly distributed and noise is additive Gaussian with varying noise levels. The noise level in data is quantified in terms of the signal-to-noise ratio (SNR) expressed using the logarithmic decibel (dB) scale:

$$\text{SNR}(g) = 10 \log_{10} \left[\frac{\sum_{j=1}^m |g_j^{\text{ideal}} - \mu_{\text{ideal}}|^2}{\sum_{j=1}^m |e_j - \mu_{\text{noise}}|^2} \right]. \quad (50)$$

Here, $g = g^{\text{ideal}} + e$ in (1) where $g^{\text{ideal}} \in \mathbb{R}^m$ 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 . Note here that we assume one has access to the noise-free component of data.

10.3.2. Reconstruction protocol—Shape based reconstructions are obtained by solving (49) using a gradient descent method. Here, $N = 2601$ and kernel is as in (48) with $\sigma^2 = 1$ for all cases. Other reconstruction parameters, like value of λ and number of iterations, vary. Test suites 1 and 3 involve comparisons against filtered back projection (FBP), classical Tikhonov, and TV reconstruction.

- FBP reconstruction is obtained by applying the `iRadon` command to data. This is the implementation of FBP in the MATLAB Image Processing Tool-box [22, pp. 9–29–9–34] and we have used linear interpolation for the back projection and the Ram-Lak filter (full frequency range).
- Classical Tikhonov reconstruction is obtained by solving (3) with \mathcal{R} as the Dirichlet's energy:

$$\mathcal{R}(f) := \int_{\Omega} |\nabla f(x)|^2 dx.$$

Gradient descent is used to solve the resulting quadratic optimization problem. Reconstruction parameters, like value of μ and number of iterations, vary.

- TV reconstruction is obtained by solving (3) with \mathcal{R} as the total variation functional:

$$\mathcal{R}(f) := \int_{\Omega} |\nabla f(x)| dx.$$

The resulting non-smooth optimization problems is solved using a linearized split Bregman scheme [7]. Reconstruction parameters, like value of μ and number of iterations, vary.

10.3.3. Test suites—The test suites seek to illustrate different properties of the shape based reconstruction scheme. One is to investigate how the smoothness of the shape template influences the reconstruction. Another is to consider the sensitivity of the reconstruction scheme against the choice of regularization parameter λ . We also have tests for assessing the robustness against noise in data. Finally, we consider the impact of using a shape template with an erroneous topology.

Test suite 1: Smoothness of template: Here we seek to investigate how smoothness of the shape template influences the final reconstruction, so the test suite involves two phantoms and two shape templates with different regularity properties. One phantom is a disc with a sharp edge (Figure 1b) so it has a distinct singularity. The other phantom is a smoothed version of the first one (Figure 2b), *i.e.*, it lacks singularities. Likewise, the two shape templates have different regularity properties. One template is a disc with a sharp edge centered at the origin with diameter equal to 1/4 of domain size (Figure 1f or Figure 2f). The other is a small square with a sharp edge and corners centered at the origin and rotated 45° (Figure 1g or Figure 2g). Data generated from these two phantoms has moderate noise level (SNR 13.7dB), see Figure 1a and Figure 2a.

The results are summarized in Figure 1 and Figure 2. We clearly see that shape based reconstructions inherit edge smoothness from the template. Likewise, corners in the template also leave traces in the reconstruction. This is to be expected since shape based reconstructions are given as a smooth deformation of a shape template, so the smoothness properties of the template is transferred to the reconstruction.

Test suite 2: Sensitivity w.r.t. regularization parameter: The parameter λ in (49) regulates the influence of the shape information, so in some sense it acts as a regularization and its choice should be dictated by the noise characteristics in data (noise level, statistical properties of noise, *e.t.c.*). A natural question is to empirically investigate the sensitivity of the reconstruction w.r.t. the choice of λ for data with fixed noise level. Hence, we consider a fixed phantom (Figure 3b) that is a smoothed concatenation of an ellipse and a rectangle and associated data (Figure 3a) is moderately noisy (SNR 13.7dB).

The results are summarized in Figure 3 where reconstructions use the same shape template, a disc with a smoothed edge centered at the origin with diameter equal to 1/4 of domain size (Figure 3c). The shape based reconstructions for varying values of λ are shown in Figures 3d to 3j. It is clear that once λ is small enough to allow for data to have an impact, it does not seem to influence the outcome that much.

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 -1.8dB) to low (SNR 25.35dB). All four data sets are generated using the same phantom (see, *e.g.*, Figure 4b), which is a smoothed concatenation of an ellipse and a rectangle. The data sets are shown in Figure 4a, Figure 5a, Figure 6a, and Figure 7a.

Shape based reconstructions are all based on the same shape template, a disc with smooth edge centered at the origin with diameter equal to 1/4 of domain size (see, *e.g.*, Figure 4g). The results are summarized in Figures 4 to 7 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.

Test suite 4: The topology of the template: Here we seek to investigate the influence of topology of 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, one a disc with sharp edge centered at the origin with diameter equal to $1/3$ of domain size (Figure 8b). The other is a U-shaped object (Figure 9b). Both are topologically equivalent, but the latter phantom has a more complex shape.

Next, there are two shape templates with different topologies. One is a disc with sharp edge centered at the origin with diameter equal to $1/4$ of domain size (see, *e.g.*, Figure 8e), so it has genus 0. The other an annulus centered at the origin (see, *e.g.*, Figure 8f), so it has genus 1.

The results, summarized in Figure 8 and Figure 9, show that the topology of the template is transferred to the reconstruction. Again, this is to be expected since shape based reconstructions are given as a smooth deformation of a shape template, so the topology of the template is transferred to the reconstruction.

10.3.4. Conclusions—The tests clearly show that shape information is powerful in suppressing the streak artifacts that typically appears in reconstructions obtained from sparsely, or unevenly, sampled data. Furthermore, shape based reconstructions outperform other methods when the goal is to reconstruct “step-function like” objects. Test suites 1 and 4 show that apart from the topology, a priori shape information doesn’t have to be that accurate. This is a very important feature since a priori shape information is expected to be approximate at best. Furthermore, test suite 2 shows that reconstructions are not that sensitive to precise choice of λ . Finally, test suite 3 shows that shape based reconstructions are remarkably robust against noise in data. Taken together, these empirical observations clearly indicate that shape based reconstruction as suggested in (9), with its finite dimensional variant in (49), is an interesting approach for image reconstruction problems where texture is not of prime interest.

The tests in section 10 do however not constitute a proper validation and comparison study, which would include several components that are currently missing. Examples of such components are:

- Evaluation protocol: Define quantitative figures of merits for assessing reconstruction quality.
- Reconstruction protocol: Define schemes for selecting the “optimal” reconstruction parameters (like regularization parameter, choice of kernel function, number of iterates, *e.t.c.*) in reconstruction methods given different noise levels and sampling schemes (typically by training against a training set). The idea is to ensure a fair comparison between methods.
- Sensitivity: Asses sensitivity of reconstructions against variations in the reconstruction parameters and variations in the a priori assumptions (shape template).

Shape based reconstruction, as defined in (9), is applicable to a rather limited class of image reconstruction problems, see discussion in subsection 11.1. Hence, a proper validation as indicated above is probably not worth the effort. Time is better spent on addressing this shortcoming and thereby widen the applicability of the approach, at least so that it can be

applied to some biomedical imaging problems of interest, like the one in ET. One can then conduct a proper validation and comparison study.

11. Discussion and remarks

11.1. Applicability

The reconstruction is a smooth deformation of a shape template in the scheme that is based on solving (9). Hence, only the texture present in the template will enter into the reconstruction, *i.e.*, texture (intensity variations) not present in the template will not be accounted for during reconstruction. The suggested approach is therefore only feasible for image reconstruction problems where intensity variations are not important.

11.2. Optimization

As with all reconstruction methods that involve solving non-convex optimization problems, there is always the issue of getting stuck in local extrema. Empirical experience suggests that this risk is higher if the shape reconstruction includes a significant translation of the template. Hence, it is better to first centre the template, *i.e.*, ensure its centre of gravity coincides reasonably well with the centre of gravity of the (unknown) image we seek to recover.

The shape based reconstructions in subsection 10.3 were obtained using a gradient descent method with a fixed step-length for solving (49). Convergence is slow and depends heavily on the fixed step size. More sophisticated approaches with better convergence properties are called for if one is to apply shape based reconstruction on larger problems. The expressions in section 8 for the gradient of a general regularization and discrepancy functionals in (9) can be used to derive alternative discretization schemes.

11.3. Regularity properties

We have seen that the regularity of I is preserved by the deformation operator (*i.e.*, essentially any regularity preserved when composing with ν) will show up in the reconstruction. This can be used to derive reconstructions that are suited for specific tasks, *e.g.*, one may consider using a sharp template if the reconstructed image is to be automatically segmented. In some cases knowledge about f_{true} may give an indication regarding the edge-smoothness of the shape template. As an example, in ET of a cryo-fixed 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, see [39, 27] for further details.

11.4. Extensions to handle intensity values

The influence of the shape template also poses a drawback regarding the handling of intensity values (texture). Assume f_{true} has texture but the shape template lacks texture, *i.e.*,

think of it as a set-function (with possibly smoothed edge). The reconstruction will lack texture as well since the only a priori information we make use of is shape related. Choosing an appropriate $\mathcal{R}: \mathcal{X} \rightarrow \mathbb{R}_+$ and setting $\mu > 0$ in (9) may have an impact since this makes it possible to account for intensity related regularity properties of f_{true} .

One approach to address this is to avoid enforcing the a priori assumption in (8) when defining the iterates that yield the reconstruction, *e.g.*, by updating the template as well. The latter amounts to replacing (9) with

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

An alternative, more feasible, formulation is to decouple the updating of the template from the updating of the deformation by considering the following intertwined recursive scheme:

$$\begin{cases} \nu_n & := \arg \inf_{\nu \in \mathcal{V}} \left[\lambda \|\nu\|_{\mathcal{V}}^2 + \mathcal{D}(\mathcal{T}(\mathcal{W}_{I_{n-1}}(\nu)), g) \right] \\ I_n & := \arg \inf_{I \in \mathcal{X}} \left[\mu \mathcal{R}(I) + \mathcal{D}(\mathcal{T}(\mathcal{W}_I(\nu_n)), g) \right] \end{cases} \quad (51)$$

Another approach would be to mimic the way metamorphosis extends the LDDMM framework to the case when the diffeomorphic deformations act simultaneously on both the shape and image intensity, see [40, Chapter 13] for further details on metamorphosis.

11.5. Deformation models

We consider linearized deformations, so each element in $\mathcal{G}_{\mathcal{V}}$ is generated as in (5) by a smooth vector field in \mathcal{V} . A drawback with this approach is that it becomes difficult to mathematically characterize those Hilbert spaces \mathcal{V} that ensure elements in $\mathcal{G}_{\mathcal{V}}$ are invertible. A natural approach that addresses this is to generate deformations using the LDDMM approach (see Remark 2) instead of (5). Then, elements in $\mathcal{G}_{\mathcal{V}}$ are diffeomorphisms if the Hilbert space \mathcal{V} fulfills a rather weak condition. Next, all of the analysis done here, including results in section 7, generalize to the LDDMM setting. The expressions for the gradients in section 8 do however become significantly more involved. Furthermore, the software implementation for shape based reconstruction used in section 10 is based on deformations generated as in (5). For these reasons, the theoretical and computational work using the LDDMM framework is part of a forthcoming paper. Finally, there are other deformation models that one may consider, *e.g.*, see [23, 4] for an overview of various groups of diffeomorphisms that may act as deformation models.

11.6. 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 \mathcal{X} are functions $f: [t_0, t_1] \times \Omega \rightarrow \mathbb{R}$ where $x \mapsto f(t, x) \in \mathcal{L}^2([t_0, t_1], \mathbb{R})$ and $t \mapsto f(t, x) \in \mathcal{X}_0$, where \mathcal{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 a image point $x \in \Omega$. Now, it often it makes sense to explicitly

separate the spatial and temporal components of f . Such a separation can be achieved by introducing a time evolution operator

$$\mathcal{W}: \mathcal{V} \times \mathcal{X}_0 \rightarrow \mathcal{X}. \quad (52)$$

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

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

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 $I^* \in \mathcal{X}_0$ and the true evolution parameter $\nu^* \in \mathcal{V}$ from data $g(t, \cdot) \in \mathcal{H}$ where

$$g(t, \cdot) = \mathcal{T}(\mathcal{W}(\nu^*, I^*)(t, \cdot)) + e(t, \cdot) \quad \text{for } t \in [t_0, t_1]. \quad (53)$$

Note that \mathcal{X}_0 is the reconstruction space for the spatial component of the spatiotemporal signals, \mathcal{H} is the data space, which is common for all data across time, $\mathcal{T}: \mathcal{X}_0 \rightarrow \mathcal{H}$ is the Fréchet differentiable (spatial) forward operator, and \mathcal{W} in (52) is the time evolution operator modeling the evolution, governed by evolution parameter in \mathcal{V} , across time. Finally, $e(t, \cdot) \in \mathcal{H}$ are samples of independent (as t varies) \mathcal{H} -valued random process $\{E_t\}_t$.

Now, if \mathcal{V} is a normed space, then one scheme for solving the spatiotemporal inverse problem in (53) is to consider

$$\inf_{\nu \in \mathcal{V}, I \in \mathcal{X}_0} \left[\lambda \|\nu\|_{\mathcal{V}}^2 + \int_{t_0}^{t_1} \{ \mu \mathcal{R}(\mathcal{W}(\nu, I)(t, \cdot)) + \mathcal{D}(\mathcal{T}(\mathcal{W}(\nu, I)(t, \cdot)), g(t, \cdot)) \} dt \right] \quad (54)$$

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

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

Solving (54) is quite challenging. Similar to (51), it may be simpler to consider the following intertwined recursive scheme where the evolution model is updated separately from the template:

$$\begin{cases} \nu^n & := \arg \inf_{\nu \in \mathcal{V}} \left[\lambda \|\nu\|_{\mathcal{V}}^2 + \int_{t_0}^{t_1} \mathcal{D}(\mathcal{T}(\mathcal{W}(\nu, I^{n-1})(t, \cdot)), g(t, \cdot)) dt \right] \\ I^n & := \arg \inf_{I \in \mathcal{X}_0} \left[\int_{t_0}^{t_1} \{ \mu \mathcal{R}(\mathcal{W}(\nu^n, I)(t, \cdot)) + \mathcal{D}(\mathcal{T}(\mathcal{W}(\nu^n, I)(t, \cdot)), g(t, \cdot)) \} dt \right] \end{cases} \quad (55)$$

The above is computationally more feasible, but it is unclear how these intertwined iterates relate to a solution of (54). Furthermore, it is highly non-trivial to understand whether (54), or (55), constitutes a regularization of the inverse problem in (53). As shown in [14], one may use the LDDMM framework (subsection 11.5) to define evolution operators with diffeomorphisms parametrized by admissible Hilbert spaces of vector fields $\mathcal{V} \subset \mathcal{C}_0^1(\Omega, \mathbb{R}^n)$.

Finally, the notion of “time” above does not have to correspond to physical time. It could be a mere parametrization 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 $I^* \in \mathcal{X}_0$. In such 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.

References

1. Amanatiadis A, Kaburlasos VG, Gasteratos A, Papadakis SE. Evaluation of shape descriptors for shape-based image retrieval. *IET Image Processing*. 2011; 5:493–499.
2. Amit Y, Manbeck KM. Deformable template models for emission tomography. *IEEE Transactions on Medical Imaging*. 1993; 12:260–268. [PubMed: 18218413]
3. Bar, L., Chan, TF., Chung, G., Jung, M., Vese, LA., Kiryati, N., Sochen, N. Mumford and Shah model and its applications to image segmentation and image restoration. In: Scherzer, O., editor. *Handbook of Mathematical Methods in Imaging*. 2. Springer-Verlag; 2015. p. 1539-1597.
4. Bauer M, Bruveris M, Michor PW. Overview of the geometries of shape spaces and diffeomorphism groups. *Journal of Mathematical Imaging and Vision*. 2014; 50:60–97.
5. Becker, F., Petra, S., Schnörr, C. Optical flow. In: Scherzer, O., editor. *Handbook of Mathematical Methods in Imaging*. 2. Springer-Verlag; 2015. p. 1945-2004.
6. Beg FM, Miller MI, Trouné A, Younes L. Computing large deformation metric mappings via geodesic flow of diffeomorphisms. *International Journal of Computer Vision*. 2005; 61:139–157.
7. Chen C, Xu G. A new linearized split Bregman iterative algorithm for image reconstruction in sparse-view X-ray computed tomography. *Computers and Mathematics with Applications*. 2016 In press.
8. Cremers, D. Image segmentation with shape priors: Explicit versus implicit representations. In: Scherzer, O., editor. *Handbook of Mathematical Methods in Imaging*. 2. Springer-Verlag; 2015. p. 1909-1944.
9. D’Arcy, T. *On Growth and Form*. Cambridge University Press; New York: 1945.
10. Fidler T, Grasmair M, Scherzer O. Shape reconstruction with a priori knowledge based on integral invariants. *SIAM Journal of Imaging Sciences*. 2012; 5:726–745.
11. Gopinath A, Xu G, Ress D, Öktem O, Subramaniam S, Bajaj C. Shape-based regularization of electron tomographic reconstruction. *IEEE Transactions on Medical Imaging*. 2012; 31:2241–2252. [PubMed: 22922711]

12. Grenander, U. General pattern theory: A Mathematical Study of Regular Structures. Oxford Mathematical Monographs, Clarendon Press; Oxford: 1994.
13. Grenander, U., Miller, M. Pattern Theory. From Representation to Inference. Oxford University Press; 2007.
14. Hinkle J, Szegedi M, Wang B, Salter B, Joshi S. 4D CT image reconstruction with diffeomorphic motion model. *Medical Image Analysis*. 2012; 16:1307–1316. [PubMed: 22766457]
15. Hohm K, Storath M, Weinmann A. An algorithmic framework for Mumford-Shah regularization of inverse problems in imaging. *Inverse Problems*. 2015; 31:115011 (30pp).
16. Klann E. A Mumford–Shah-like method for limited data tomography with an application to electron tomography. *SIAM Journal of Imaging Sciences*. 2011; 4:1029–1048.
17. Layer T, Blaickner M, Knaüsl B, Georg D, Neuwirth J, Baum RP, Schuchardt C, Wiessalla S, Matz G. PET image segmentation using a Gaussian mixture model and Markov random fields. *EJNMMI Physics*. 2015; 2
18. Lindeberg T. Generalized axiomatic scale-space theory. *Advances in Imaging and Electron Physics*. 2013; 178:1–96.
19. Louis AK. Feature reconstruction in inverse problems. *Inverse Problems*. 2011; 27:065010 (21pp).
20. Maas J, Rumpf M, Schönlieb C, Simon S. A generalized model for optimal transport of images including dissipation and density modulation. Tech Report. 2015 arXiv:1504.01988v1 [math.NA], ArXiv e-prints.
21. Manay, S., Cremers, D., Hong, B-W., Yezzi, A., Soatto, S. Integral invariants and shape matching. In: Krim, H., Yezzi, A., editors. *Statistics and Analysis of Shapes*. Springer-Verlag; 2006. p. 137-166. Modeling and Simulation in Science, Engineering and Technology
22. The MathWorks. *Image Processing Toolbox User’s Guide*. 2015. Revised for Version 9.3 (Release 2015b)
23. Michor PW, Mumford D. A zoo of diffeomorphism groups on \mathbb{R}^n . *Annals of Global Analysis and Geometry*. 2013; 44:529–540.
24. Miller MI, Trouné A, Younes L. Geodesic shooting for computational anatomy. *Journal of Mathematical Imaging and Vision*. 2006; 24:209–228. [PubMed: 20613972]
25. Miller MI, Trouné A, Younes L. Hamiltonian systems and optimal control in computational anatomy: 100 years since D’arcy Thompson. *Annual Review of Biomedical Engineering*. 2015; 17:447–509.
26. Mohammad-Djafari A. Gauss-Markov-Potts priors for images in computer tomography resulting to joint optimal reconstruction and segmentation. *International Journal of Tomography and Statistics*. 2009; 11:76–92.
27. Öktem, O. Mathematics of electron tomography. In: Scherzer, O., editor. *Handbook of Mathematical Methods in Imaging*. 2. Springer–Verlag; New York: 2015. p. 937-1031.II of Springer Reference
28. Ramlau R, Ring W. A Mumford–Shah level-set approach for the inversion and segmentation of X-ray tomography data. *Journal of Computational Physics*. 2007; 221:539–557.
29. Romanova M, Dahla AB, Donga Y, Hansena PC. Simultaneous tomographic reconstruction and segmentation with class priors. *Inverse Problems in Science and Engineering*. 2015
30. Rottman, C., Bauer, M., Modin, K., Joshi, S. Weighted diffeomorphic density matching with applications to thoracic image registration [arxiv,]. *Proceedings of the 5th MICCAI Workshop on Mathematical Foundations of Computational Anatomy (MFCA)*; Munich, Germany. October 9, 2015; 2015.
31. Rumpf, M., Wirth, B. Variational methods in shape analysis. In: Scherzer, O., editor. *Handbook of Mathematical Methods in Imaging*. 2. Springer-Verlag; 2015. p. 1819-1858.
32. Ruthotto, L., Modersitzki, J. Non-linear image registration. In: Scherzer, O., editor. *Handbook of Mathematical Methods in Imaging*. 2. Springer-Verlag; 2015. p. 2005-2051.
33. Scherzer, O., Grasmair, M., Grossauer, H., Haltmeier, M., Lenzen, F. *Variational Methods in Imaging*, vol. 167 of Applied Mathematical Sciences. Springer-Verlag; New York: 2009.
34. Schölkopf, B., Herbrich, R., Smola, AJ. A generalized representer theorem. In: Helmbold, D., Williamson, B., editors. *Computational Learning Theory Volume; Proceedings of the 14th Annual*

- Conference on Computational Learning Theory, COLT 2001 and 5th European Conference on Computational Learning Theory, EuroCOLT 2001; Amsterdam, The Netherlands. July 16–19, 2001; Berlin: Springer-Verlag; 2001. p. 416-426.2111 of Lecture Notes in Computer Science
35. Storath M, Weinmann A, Jürgen Friel J, Unser M. Joint image reconstruction and segmentation using the Potts model. *Inverse Problems*. 2015; 31:025003 (29pp).
 36. Trouvé A. Diffeomorphism groups and pattern matching in image analysis. *International Journal of Computer Vision*. 1998; 28:213–221.
 37. Trouvé, A., Younes, L. Shape spaces. In: Scherzer, O., editor. *Handbook of Mathematical Methods in Imaging*. 2. Springer-Verlag; 2015. p. 1759-1817.
 38. Vialard, F-X. Computational anatomy from a geometric point of view. Presentation at the Geometry, Mechanics and Control, 8th International Young Researchers workshop 2013; Barcelona. December 2013;
 39. Vulovi M, Ravelli RBG, van Vliet LJ, Koster AJ, Lazić I, Lücken U, Rullgård H, Öktem O, Rieger B. Image formation modeling in cryo electron microscopy. *Journal of Structural Biology*. 2013; 183:19–32. [PubMed: 23711417]
 40. Younes, L. Shapes and Diffeomorphisms, vol. 171 of Applied Mathematical Sciences. Springer-Verlag; 2010.
 41. Yu, Y-L., Cheng, H., Schuurmans, D., Szepesvári, C. Characterizing the representer theorem. *Journal of Machine Learning Research*; Special issue for the Proceedings of the 30th International Conference on Machine Learning; Atlanta, Georgia, USA. 2013; 2013.

Appendix A. Basic notation

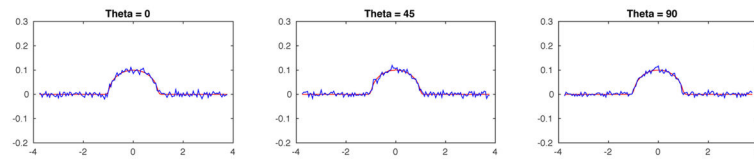
\mathbb{R}^n denotes the n -dimensional Euclidian space and $\mathbb{M}^{n,m}$ is the vector space of all $(n \times m)$ matrices. $\mathbb{M}_+^{n,m}$ will then refer to positive definite $(n \times m)$ matrices. Next, if \mathcal{X} and \mathcal{Y} are two vector spaces, then $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ is the vector space of linear mappings defined on \mathcal{X} with values in \mathcal{Y} .

We also make use of the *vec*-operator that transforms a matrix into a (column) vector by stacking all the columns of the matrix one underneath the other. Hence, $\text{vec} : \mathbb{M}^{n \times m} \rightarrow \mathbb{R}^{nm}$ is defined as

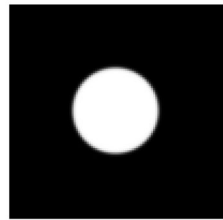
$$\text{vec}[\mathbf{A}] := (a_{1,1} \quad \dots \quad a_{n,1} \quad \dots \quad a_{1,m} \quad \dots \quad a_{n,m})^\top$$

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

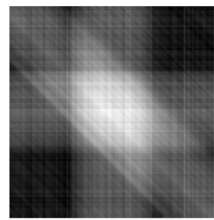
$$\mathbf{A} := \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,m} \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,m} \end{pmatrix}.$$



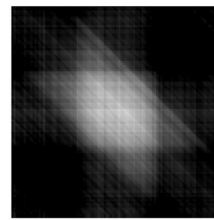
(a) Data: Parallel beam 2D tomographic data with moderate noise level (SNR 13.8 dB) generated from the phantom in (b) using three directions 0° , 45° and 90° . There are 151 data points for each of the three directions, so $m = kl = 3 \times 151$. Jagged blue curve is the data used as input for reconstructions shown below, red curve is the corresponding noise-free data.



(b) The phantom, a 101×101 pixel image with sharp edge (but no corners).



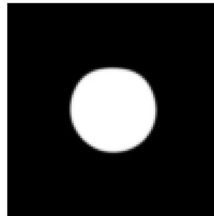
(c) FBP reconstruction.



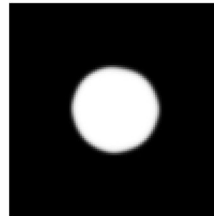
(d) Classical Tikhonov reconstruction (4000 iterations, $\mu = 10$).



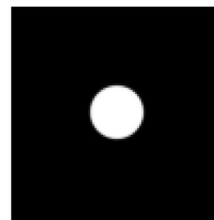
(e) TV reconstruction (4000 iterations, $\mu = 8$).



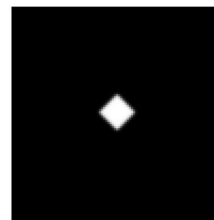
(f) Shape based reconstruction after 400 iterations using template in (h) and $\lambda = 10^{-5}$.



(g) Shape based reconstruction after 1000 iterations using template in (i) and $\lambda = 10^{-3}$.

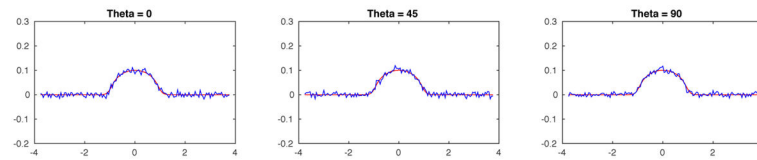


(h) Shape template, a disc with a sharp edge (no corners).

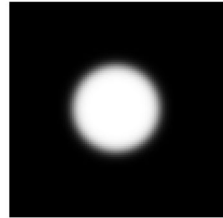


(i) Shape template, a rotated square with a sharp edge and corners.

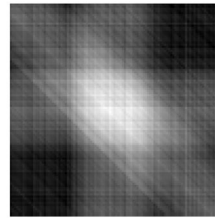
Fig. 1. Test suite 1 – Smoothness of template. The phantom (b) is a disc with a sharp edge centered at the origin (diameter = $1/3$ of domain). Reconstructions, obtained using different methods, are shown in (c)–(g). Shape based reconstructions clearly



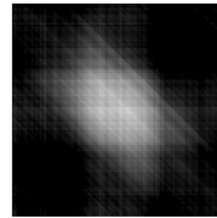
(a) Data: Parallel beam 2D tomographic data with moderate noise level (SNR 13.7 dB) generated from the phantom in (b) using three directions 0° , 45° and 90° . There are 151 data points for each of the three directions, so $m = kl = 3 \times 151$. Jagged blue curve is the data used as input for reconstructions shown below, red curve is the corresponding noise-free data.



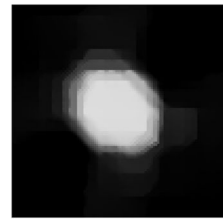
(b) The phantom, a 101×101 pixel image (smoothed version of Figure 1b).



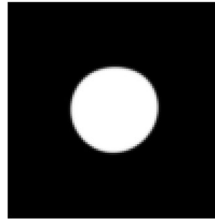
(c) FBP reconstruction.



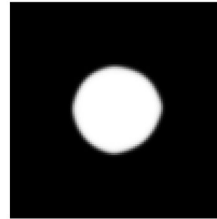
(d) Classical Tikhonov reconstruction (4000 iterations, $\mu = 12$).



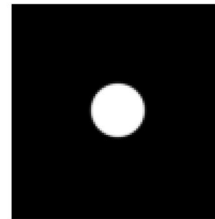
(e) TV reconstruction (4000 iterations, $\mu = 10$).



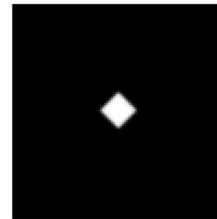
(f) Shape based reconstruction using template in (h) (400 iterations, $\lambda = 10^{-5}$).



(g) Shape based reconstruction using template in (i) (4000 iterations, $\lambda = 10^{-3}$).

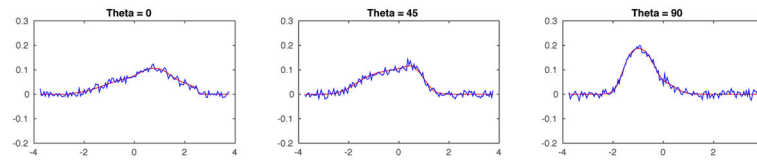


(h) Shape template with sharp edge but no corners.

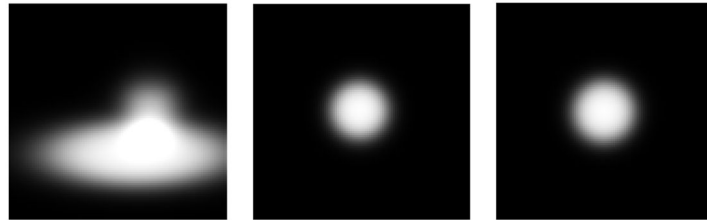


(i) Shape template with sharp edge and corners.

Fig. 2. Test suite 1 – Smoothness of template. The phantom, shown in (b), has smoothed edge and reconstructions, obtained using different methods, are shown in (c)–(g). Despite a phantom with smoothed edge, the shape based reconstructions have a sharp edge since the templates have those properties. To some extent the same also applies to corners.



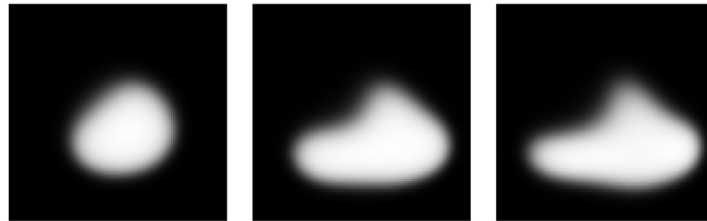
(a) Data: Parallel beam 2D tomographic data with moderate noise level (SNR 13.7dB) generated from the phantom in (b) using three directions 0° , 45° and 90° . There are 151 data points for each of the three directions, so $m = kl = 3 \times 151$. Jagged blue curve is the data used as input for reconstructions shown below, red curve is the corresponding noise-free data.



(b) The phantom, a 101×101 pixel image.

(c) Shape template.

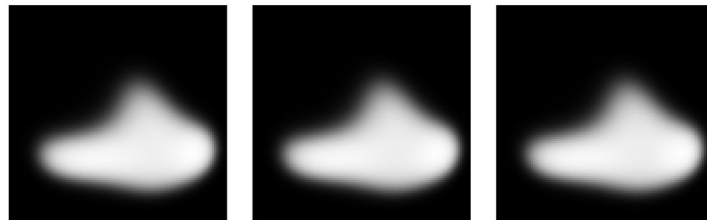
(d) Shape based reconstruction, $\lambda = 1.0$.



(e) Shape based reconstruction, $\lambda = 10^{-1}$.

(f) Shape based reconstruction, $\lambda = 10^{-2}$.

(g) Shape based reconstruction, $\lambda = 10^{-3}$.



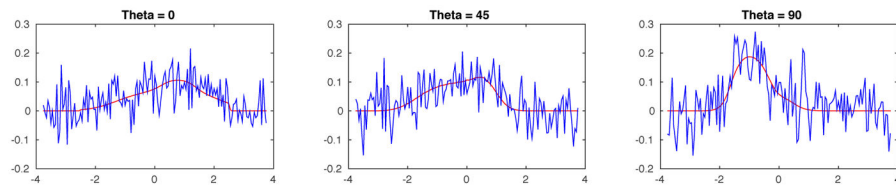
(h) Shape based reconstruction, $\lambda = 10^{-4}$.

(i) Shape based reconstruction, $\lambda = 10^{-5}$.

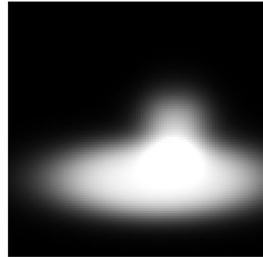
(j) Shape based reconstruction, $\lambda = 10^{-7}$.

Fig. 3.

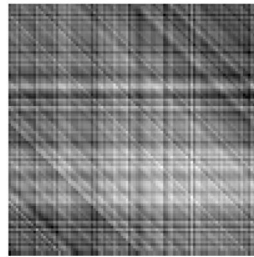
Test suite 2 – Sensitivity w.r.t. regularization parameter. Figures (d)–(j) show shape based reconstructions obtained from data in (a) after 400 iterations using shape template in (c) and different values for λ . The phantom is shown in (b). Reconstructions in (d) and (e) are clearly over-regularized, *i.e.*, shape template has too much influence on the outcome. Once λ is small enough to allow for data to have an impact, it does not seem to influence the outcome that much.



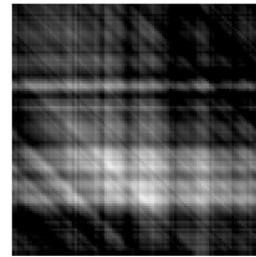
(a) Data: Parallel beam 2D tomographic data with very high noise level (SNR -1.8dB) generated from the phantom in (b) using three directions 0° , 45° and 90° . There are 151 data points for each of the three directions, so $m = kl = 3 \times 151$. Jagged blue curve is the data used as input for reconstructions shown below, red curve is the corresponding noise-free data.



(b) The phantom, a 101×101 pixel image.



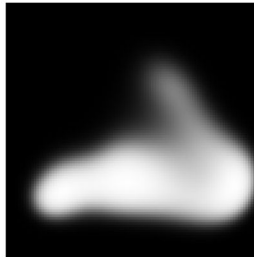
(c) FBP reconstruction.



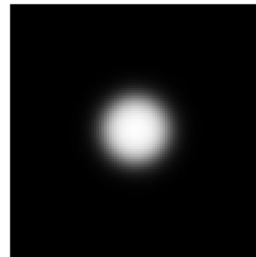
(d) Classical Tikhonov reconstruction (4000 iterations, $\mu = 220$).



(e) TV reconstruction (4000 iterations, $\mu = 50$).



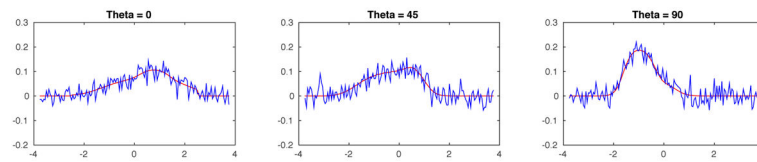
(f) Shape based reconstruction using template in (g) (400 iterations, $\lambda = 10^{-5}$).



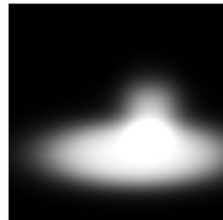
(g) The shape template.

Fig. 4.

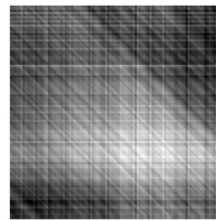
Test suite 3 – Very high noise. Data, shown in (a), has very high noise (SNR – 1.8dB). The phantom is shown in (b) and reconstructions, obtained using different methods, are shown in (c)–(f). Only shape based reconstruction provides a result (f) that bears any similarities to the original phantom (b).



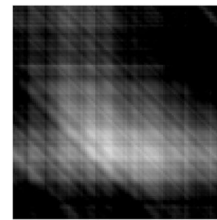
(a) Data: Parallel beam 2D tomographic data with high noise level (SNR 5.39dB) generated from the phantom in (b) using three directions 0° , 45° and 90° . There are 151 data points for each of the three directions, so $m = kl = 3 \times 151$. Jagged blue curve is the data used as input for reconstructions shown below, red curve is the corresponding noise-free data.



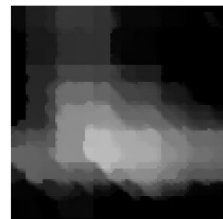
(b) The phantom, a 101×101 pixel image.



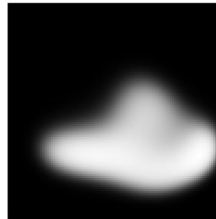
(c) FBP reconstruction.



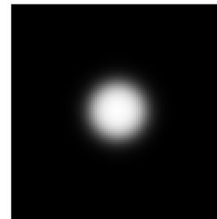
(d) Classical Tikhonov reconstruction (4000 iterations, $\mu = 150$).



(e) TV reconstruction (4000 iterations, $\mu = 25$).



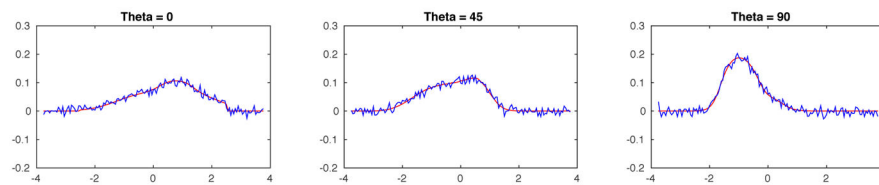
(f) Shape based reconstruction using template in (g) (400 iterations, $\lambda = 10^{-5}$).



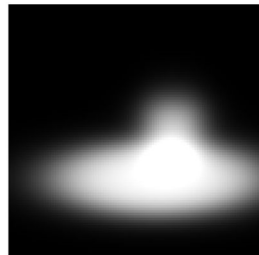
(g) The shape template.

Fig. 5.

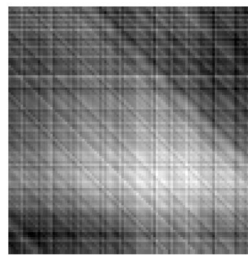
Test suite 3 – High noise. Data, shown in (a), has very high noise (SNR 5.39dB). The phantom is shown in (b) and reconstructions, obtained using different methods, are shown in (c)–(f). Only shape based reconstruction provides a result (f) that bears any similarities to the original phantom (b).



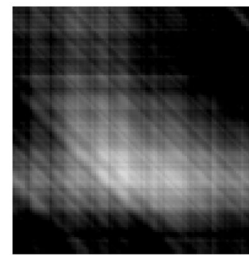
(a) Data: Parallel beam 2D tomographic data with moderate noise level (SNR 13.49dB) generated from the phantom in (b) using three directions 0° , 45° and 90° . There are 151 data points for each of the three directions, so $m = kl = 3 \times 151$. Jagged blue curve is the data used as input for reconstructions shown below, red curve is the corresponding noise-free data.



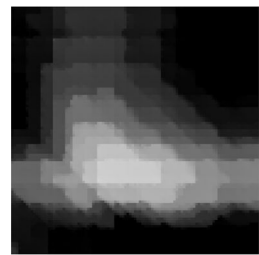
(b) The phantom, a 101×101 pixel image.



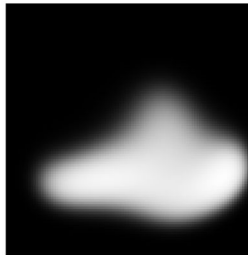
(c) FBP reconstruction.



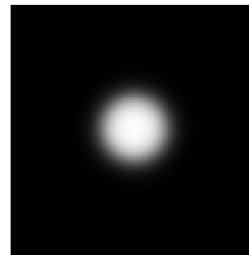
(d) Classical Tikhonov reconstruction (4000 iterations, $\mu = 120$).



(e) TV reconstruction (4000 iterations, $\mu = 1$).



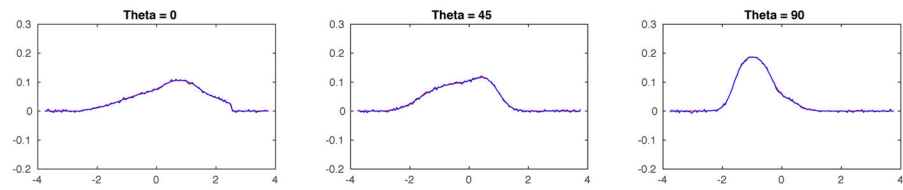
(f) Shape based reconstruction using template in (g) (400 iterations, $\lambda = 10^{-5}$).



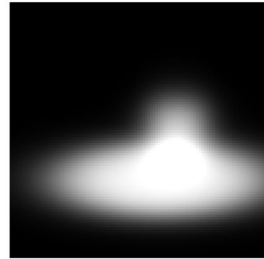
(g) The shape template.

Fig. 6.

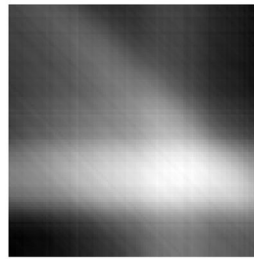
Test suite 3 – Moderate noise. Data, shown in (a), has very high noise (SNR 13.49dB). The phantom is shown in (b) and reconstructions, obtained using different methods, are shown in (c)–(f). Only shape based reconstruction provides a result (f) that bears any similarities to the original phantom (b).



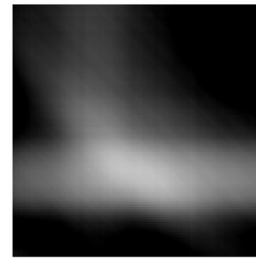
(a) Data: Parallel beam 2D tomographic data with low noise level (SNR 25.35dB) generated from the phantom in (b) using three directions 0° , 45° and 90° . There are 151 data points for each of the three directions, so $m = kl = 3 \times 151$. Jagged blue curve is the data used as input for reconstructions shown below, red curve is the corresponding noise-free data.



(b) The phantom, a 101×101 pixel image.



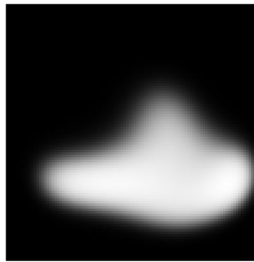
(c) FBP reconstruction.



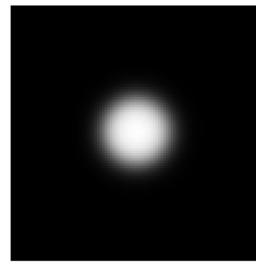
(d) Classical Tikhonov reconstruction (4000 iterations, $\mu = 100$).



(e) TV reconstruction (4000 iterations, $\mu = 0.05$).

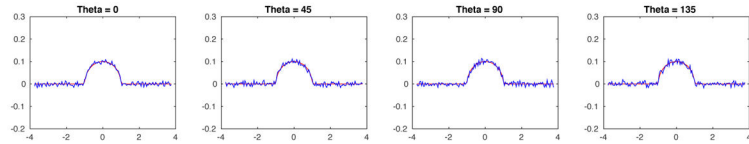


(f) Shape based reconstruction using template in (g) (400 iterations, $\lambda = 10^{-5}$).

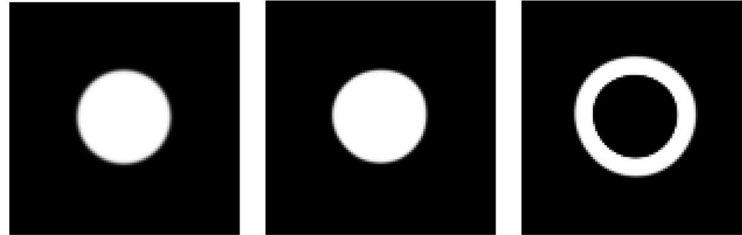


(g) The shape template.

Fig. 7. Test suite 3 – Low noise. Data, shown in (a), has very high noise (SNR 25.35dB). 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).



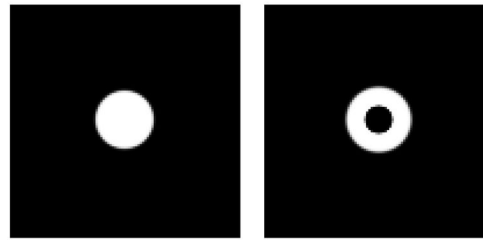
(a) Data: Parallel beam 2D tomographic data with moderate noise level (SNR 14.16dB) generated from the phantom in (b) using four directions 0° , 45° , 90° and 135° . Blue jagged line is the data used as input for reconstruction, red curve is the corresponding noise-free data. There are 151 values for each direction, so $m = kl$ with $k = 4$ and $l = 151$.



(b) A phantom with genus 0.

(c) Shape based reconstruction using template in (e) (400 iterations, $\lambda = 10^{-5}$).

(d) Shape based reconstruction using template in (f) (400 iterations, $\lambda = 10^{-5}$).

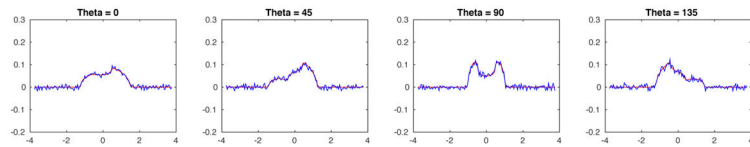


(e) A shape template with genus 0.

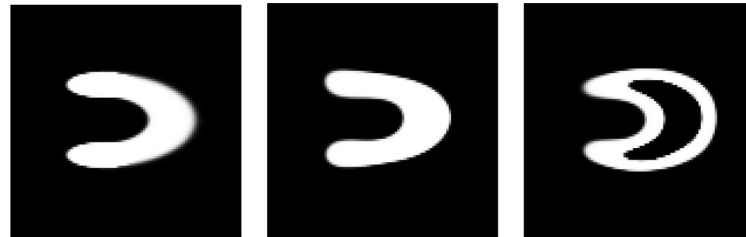
(f) A shape template with genus 1.

Fig. 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) and (c) and (d) show shape based reconstructions obtained using a template with genus 0 and 1, respectively. It is clear that the reconstruction inherits the topology of the template.



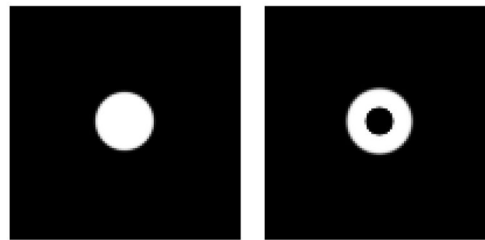
(a) Data: Parallel beam 2D tomographic data with moderate noise level (SNR 12.95dB) generated from the phantom in (b) using four directions 0° , 45° , 90° and 135° . Blue jagged line is the data used as input for reconstruction, red curve is the corresponding noise-free data. There are 151 values for each direction, so $m = kl$ with $k = 4$ and $l = 151$.



(b) A phantom with genus 0.

(c) Shape based reconstruction using template in (e) (400 iterations, $\lambda = 10^{-5}$).

(d) Shape based reconstruction using template in (f) (400 iterations, $\lambda = 10^{-5}$).



(e) A shape template with genus 0.

(f) A shape template with genus 1.

Fig. 9.

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) and (c) and (d) show shape based reconstructions obtained using a template with genus 0 and 1, respectively. It is clear that the reconstruction inherits the topology of the template.