Template-Based Image Reconstruction Facing Different Topologies

Sebastian Neumayer[1,2] · Antonia Topalovic[3]

Abstract

The reconstruction of images from measured data is an increasing field of research. For highly under-determined problems, template-based image reconstruction provides a way of compensating for the lack of sufficient data. A caveat of this approach is that dealing with different topologies of the template and the target image is challenging. In this paper, we propose a LDDMM-based image-reconstruction model that resolves this issue by adding a source term. On the theoretical side, we show that the model satisfies all criteria for being a well-posed regularization method. For the implementation, we pursue a discretize-then-optimize approach involving the proximal alternating linearized minimization algorithm, which is known to converge under mild assumptions. Our simulations with both artificial and real data confirm the robustness of the method, and its ability to successfully deal with topology changes even if the available amount of data is very limited.

Keyword Indirect image registration · Inverse problems · LDDMM · Metamorphosis · Variational models

1 Introduction

In medical applications such as computed tomography (CT) [31], images are typically observed via indirect and potentially noisy measurements. Especially when the amount of measured data is limited, obtaining meaningful reconstructions is challenging. This is, for instance, the case in limited-angle CT [17, 31], where sparse data is acquired in order to minimize exposure time of organisms to X-radiation. In such settings, it is inevitable to add a priori information about the target into the reconstruction process, e.g., in form of a template image that is somehow close to the expected reconstruction. Template-based methods, outlined in more detail below, encode this closeness assumption directly into the reconstruc-

✉ Sebastian Neumayer
sebastian.neumayer@mathematik.tuchemnitz.de

1 EPFL: Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland

2 Present Address: TU Chemnitz, Chemnitz, Germany

3 Humboldt Universität Berlin, Berlin, Deutschland

tion process. Hence, any reconstruction will strongly depend on the chosen template. If a good template is available, e.g., from an earlier observation, competing methods such as the filtered backprojection [31] or total-variation (TV) regularization [39] are outperformed by large margins [10, Sec. 10]. Clearly, template-based methods can also be applied for other inverse problems such as deblurring or MRI. In the following, we discus *indirect image matching* in more detail. The included examples lead us to the proposed model, which is a simplified extension of the metamorphosis approach proposed in [18].

Indirect image matching

Let $\Omega \subset \mathbb{R}^d$ be a bounded domain. Image matching refers to the task of transforming a template image $T \in L^2(\Omega)$ to match a target image $U \in L^2(\Omega)$ as closely as possible regarding some misfit measure. For indirect image matching, the target U is additionally unknown and only specified to be a solution of the *image-reconstruction* problem

$$K \circ U + n^\delta = g, \tag{1}$$

where $K : L^2(\Omega) \to Y$ is a (not necessarily linear) data acquisition operator (often related to a physical process) that maps into a Banach space Y, $g \in Y$ is the measurement, and the noise n^δ models measurement imperfections. Consequently, one has to jointly solve a motion estimation problem and an image-reconstruction problem. Various indirect image matching models have been proposed in the literature [10, 18, 21, 23, 33, 35]. For all models, the transformation between the template T and the unknown target U can be modeled with a partial differential equation (PDE). Essentially, the template's domain Ω is deformed by a diffeomorphism, usually resulting in transformations that appear natural to humans. This idea originates from the flow of diffeomorphism model [12, 13, 42], in which image pixel intensities are transported along trajectories determined by diffeomorphism paths. For comprehensive overviews, we refer to [27, 49], and for a historic account we refer to [25]. In the following, we discuss two indirect image matching models in more detail.

As the space of diffeomorphisms has no natural vector space structure, the popular class of Large Diffeomorphic Deformation Metric Mapping (LDDMM) image matching models relies on a subset of deformations generated by admissible (smooth) velocity fields $v \in \mathcal{V} := L^2([0, 1], V)$ via the ordinary differential equation (ODE)

$$
\begin{aligned}
\tfrac{\mathrm{d}}{\mathrm{d}t}\varphi(t, x) &= v\big(t, \varphi(t, x)\big) &&\text{for } (t, x) \in [0, 1] \times \Omega, \\
\varphi(0, x) &= x &&\text{for } x \in \Omega.
\end{aligned}
\tag{2}
$$

In the definition of $\mathcal{V}$, V is an admissible vector space that continuously embeds into $C_0^{1,\alpha}(\Omega, \mathbb{R}^d)$, $0 < \alpha \leq 1$, namely the closure of $C_c^\infty(\Omega, \mathbb{R}^d)$ with respect to the Hölder norm $\| \cdot \|_{C^{1,\alpha}}$. Using the diffeomorphism φ, we implicitly define a transformation path $I : [0, 1] \times \Omega \to \mathbb{R}$ starting from the template T via $I(t, \varphi(t, x)) = T(x)$. From an Eulerian perspective, we can link this path I to the velocity v directly [10], namely as the (weak) solution of the transport equation

$$
\begin{aligned}
\tfrac{\partial}{\partial t} I(t, x) + v(t, x)\nabla_x I(t, x) &= 0 &&\text{for} (t, x) \in [0, 1] \times \Omega, \\
I(0, x) &= T(x) &&\text{for} x \in \Omega.
\end{aligned}
\tag{3}
$$

We denote the set of feasible tuples $(I, v) \in L^2([0, 1], L^2(\Omega)) \times \mathcal{V}$ solving this PDE by $\mathcal{A}$. Given the data g and the template T, a reconstruction can now be defined as $R = I(1, \cdot) =$

$T(\varphi^{-1}(1, \cdot))$,[1] where I is a minimizer of the variational problem

$$\min_{(I,v)\in\mathcal{A}} \mathcal{D}\big(K \circ I(1, \cdot), g\big) + \lambda\mathcal{E}(v). \tag{4}$$

Here, the data fidelity term $\mathcal{D}\colon Y \times Y \to \mathbb{R}_{\geq 0}$ quantifies the misfit of $K \circ I(1, \cdot)$ with the measurements g, and the regularizer $\mathcal{E}\colon \mathcal{V} \to \mathbb{R}_{\geq 0}$ enforces the required smoothness of v [10]. Note that (2) can be simplified by using linearized deformations and a Taylor expansion around some initial flow field, leading to optical-flow-based transformation models [1, 6, 22, 35].

However, an image transformation model purely based on diffeomorphisms can yield unsatisfying reconstructions, e.g., if the images have different mass or topological properties [10]. To resolve this issue, Gris et al. [18] proposed to replace the flow of diffeomorphism model (3) underlying (4) with the metamorphosis model [28, 38, 43, 44]. In addition to the transport of pixel intensities, metamorphosis also allows intensity variations along the trajectories based on a source term $\zeta \in L^2([0, 1] \times \Omega)$. Hence, this model can create or remove objects during the transformation process. To put this into formulas, metamorphosis paths I are solutions of the transport equation

$$\begin{aligned}
\tfrac{\partial}{\partial t} I(t, x) + v(t, x)\nabla_x I(t, x) &= \zeta(t, x) &&\text{for}(t, x) \in [0, 1] \times \Omega,\\
I(0, x) &= T(x) &&\text{for}x \in \Omega.
\end{aligned} \tag{5}$$

We denote the set of feasible tuples $(I, v, \zeta) \in L^2([0, 1] \times \Omega) \times \mathcal{V} \times L^2([0, 1] \times \Omega)$ by $\mathcal{B}$. From the Lagrangian perspective, $I(t, x)$ can be equivalently defined using the solution φ of (2)via

$$I\big(t, \varphi(t, x)\big) = T(x) + \int_0^t \zeta(s, \varphi(s, x))\, \mathrm{d}s \qquad \text{for } (t, x) \in [0, 1] \times \Omega. \tag{6}$$

The relation between (5) and (6) is commonly known as the method of characteristics [15]. Given the data g and the template T, a reconstruction can then be defined as $R = I(1, \cdot)$, where I solves

$$\min_{(I,v,\zeta)\in\mathcal{B}} \mathcal{D}\big(K \circ I(1, \cdot), g\big) + \lambda_1\mathcal{E}_1(v) + \lambda_2\mathcal{E}_2(\zeta). \tag{7}$$

Here, the additional regularizer $\mathcal{E}_2\colon L^2([0, 1] \times \Omega) \to \mathbb{R}_{\geq 0}$ enforces the necessary regularity of ζ.

Proposed model If we take a closer look at (7), we notice that the objective takes only $I(1, \cdot)$ into account and is blind to the path I at all other times. Further, it holds that $I(1, \cdot) = T(\varphi^{-1}(1, \cdot))+z$, where $z \in L^2(\Omega)$ is defined via $z(\varphi(1, x)) = \int_0^1 \zeta(s, \varphi(s, x))\, \mathrm{d}s$. Consequently, any reconstruction R consists of a template deformation and some intensity change z that depends on ζ and φ. In (7), we have a complicated regularization of z in terms of both $\mathcal{E}_1$ and $\mathcal{E}_2$. To obtain a simpler reconstruction model, we propose to replace the underlying transformation model (5) in (7) by

$$\begin{aligned}
\tfrac{\mathrm{d}}{\mathrm{d}t}\varphi(t, x) &= v\big(t, \varphi(t, x)\big) &&\text{for } (t, x) \in [0, 1] \times \Omega,\\
\varphi(0, x) &= x &&\text{for } x \in \Omega,\\
I(x) &= T\big(\varphi^{-1}(1, x)\big) + z(x) &&\text{for } x \in \Omega.
\end{aligned} \tag{8}$$

The set of tuples $(I, v, z) \in L^2(\Omega) \times \mathcal{V} \times L^2(\Omega)$ satisfying (8) is denoted by $\mathcal{C}$. Loosely speaking, we first deform the template T based on φ. Afterwards, we modify the pixel

---

[1] Throughout the manuscript, the inverse of φ is always with respect to the spatial coordinate only.

values with the source z. Based on the simplified transformation model (8), we propose the reconstruction model

$$\min_{(R,v,z)\in\mathcal{C}} \mathcal{D}\big(K \circ R, g\big) + \lambda_1\mathcal{E}_1(v) + \lambda_2\mathcal{E}_2(z). \tag{9}$$

Instead of implicitly regularizing z as in (7), we directly regularize it with $\mathcal{E}_2\colon L^2(\Omega) \to \mathbb{R}_{\geq 0}$. To this end, we can rely on any well-established (convex) regularizer for images. The reconstruction R in (9) is composed of two summands, where one depends on v and the other on z. Therefore, designing efficient numerical schemes for (8) appears simpler than for (6), where we have a highly nonlinear dependence of R on ζ. Additionally, the dimensionality of z is lower than that of ζ as we do not have a time dependence. Since each tuple (v, z) gives rise to a unique reconstruction $R_{v,z}$, we can eliminate the set $\mathcal{C}$ from (9) and end up with

$$\min_{(v,z)\in\mathcal{V}\times L^2(\Omega)} J_{\lambda,g}(v, z) := \mathcal{D}\big(K \circ R_{v,z}, g\big) + \lambda_1\mathcal{E}_1(v) + \lambda_2\mathcal{E}_2(z), \tag{10}$$

where the system (8) is implicitly encoded in $R_{v,z}$.

Contributions We study the regularizing properties of the proposed model (10), and develop efficient numerical schemes for solving it. To the best of our knowledge, the only other template-based reconstruction model with a source term is [18], but their approach involves a nonlinear coupling between the deformation and the source part. Further, their regularizer choice is restricted to differentiable ones. The simulations in [18] are based on L^2 regularization for ζ in (7), which in our experiments turned out to be unsuitable for sparse data. Better regularizers seem to be necessary to avoid reconstruction artifacts related to the source term in both (7) and (10). Hence, we propose to use TV regularization instead, which is known to yield good results for CT problems at reasonable computational cost. With this choice, we obtain meaningful reconstructions even for sparse data, which is a setting that [18] does not target. Compared to [23], we achieve reconstructions of similar quality, but with the advantage that topology changes are possible due to z. The detail of reconstructable structures that are not present in the template T depends on the available amount of data g, i.e., we cannot reconstruct detailed structures out of nothing. Hence, the proposed method is most useful if a good template T is available.

Many algorithmic approaches for the flow of diffeomorphism model have been proposed during the last years [2, 20, 26, 34, 40, 47, 48]. For linear forward operators K, these approaches can be often extended to the indirect setting (9) without any complicated modifications. As Lagrangian approaches turned out to be very efficient for indirect image matching, we decided to adapt the methods developed in [23, 24], which build upon the FAIR toolbox [30]. Since the problem is non-smooth due to the TV regularization of z, we cannot deploy their proposed Gauss–Newton–Krylov solver, and we use the iPALM algorithm [37] instead. Similarly as in [23], the ODE in (8) is solved with an explicit Runge–Kutta method, which allows one for efficient algorithmic differentiation of $(v, z) \mapsto R_{v,z}$. By construction of $R_{v,z}$, computing its gradient $\nabla R_{v,z}$ has basically the same cost as for the LDDMM-based model [23], which does not involve a source term z. Further, the approach does not require the storage of multiple space-time vector fields or images at intermediate time instances, as it is often the case when directly solving the PDE (5) with Eulerian methods. We want to emphasize that the proposed scheme can be implemented matrix-free, which is crucial when using dense forward operators K such as the Radon transform. Since iPALM has in general worse convergence rates than second-order methods, we combine it with a Gauss–Newton solver for v as post-processing step. This can only improve the objective function values, and in practice we also observed an improved reconstruction quality.

Outline In Sect. 2, the necessary theoretical background for the flow equation (2) and the total variation is provided. For the proposed model (10), existence of a minimizer, stability with respect to the data, and convergence for vanishing noise are established in Sect. 3. In order to approximate solutions of (10) numerically, we follow a *discretize-then-optimize* approach that involves the iPALM algorithm as outlined in Sect. 4. This allows one to easily exchange the regularizer for v and z if desired. Our implementation builds upon the FAIR toolbox [30], which allows for a simple extension to other distances and regularizers that are already implemented as part of the toolbox. Numerical results for the proposed model are provided in Sect. 5. Finally, conclusions are drawn in Sect. 6.

2 Preliminaries

In this section, we briefly review the necessary theoretical background. **Diffeomorphisms** Recall that the deformations φ are induced by (2). For our theoretical investigations, it is useful to consider different initial times $s \in [0, 1]$, i.e., the modified equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi_{s,v}(t, x) = v\big(t, \varphi_{s,v}(t, x)\big) \qquad \text{for } (t, x) \in [0, 1] \times \Omega,$$
$$\varphi_{s,v}(s, x) = x \qquad \qquad \text{for } x \in \Omega. \tag{11}$$

Here, the subscripts express the dependence on the initial time s and the velocity field v. This generalization enables us to move back and forth on trajectories $\varphi_{s,v}(\cdot, x)$ starting from arbitrary time instants s, allowing us to rely on a unified theoretical result. The following theorem is a reformulation of [43, Thms. 1 and 9] and characterizes the solutions of (11).

Theorem 1 *Let $v \in L^2([0, T], \mathcal{V})$, where $\mathcal{V}$ is continuously embedded into $C_0^{1,\alpha}(\Omega, \mathbb{R}^d)$ for some $0 < \alpha \leq 1$. Given $s \in [0, T]$, there exists a unique global solution $\varphi_{s,v} \in C([0, T], C^1(\overline{\Omega}, \mathbb{R}^d))$ of (11). Further, the solution operator $\Phi_s \colon L^2([0, T], \mathcal{V}) \to C([0, T] \times \overline{\Omega}, \mathbb{R}^d)$ assigning a flow $\varphi_{s,v}$ to a velocity field v is continuous with respect to the weak topology in $L^2([0, T], \mathcal{V})$.*

As $\varphi_{t,v}(0, \varphi_{0,v}(t, x)) = x$, we directly get that $\varphi_{0,v}(t, \cdot)$ is a diffeomorphism for every $t \in [0, 1]$. Now, let us have a closer look at the solutions of (8). Since $\varphi_{0,v}^{-1}(1, \cdot) = \varphi_{1,v}(0, \cdot)$, we conclude by Theorem 1 that $v_i \rightharpoonup v$ in $L^2([0, T], \mathcal{V})$ implies $\varphi_{0,v_i}^{-1}(1, \cdot) \to \varphi_{0,v}^{-1}(1, \cdot) \in C(\overline{\Omega}, \mathbb{R}^d)$. Further, [14, Thm. 3.1.10] implies that $\{\varphi_{0,v_i}^{-1}(1, \cdot)\}_{i \in \mathbb{N}}$ is uniformly bounded in $C^{1,\alpha}(\overline{\Omega}, \mathbb{R}^d)$. Hence, [32, Cor. 3] implies that $T \circ \varphi_{0,v_i}^{-1}(1, \cdot) \to T \circ \varphi_{0,v}^{-1}(1, \cdot)$ in $L^2(\Omega)$. If further $z_i \rightharpoonup z$ in $L^2(\Omega)$, this directly implies $R_{v_i,z_i} \rightharpoonup R_{v,z}$ in $L^2(\Omega)$.

Total variation The total variation [39] is a popular regularizer in imaging as it tends to preserve edges and sharp structures, which is in contrast to other techniques such as linear smoothing or Tikhonov regularization [7]. For any $f \in L^1(\Omega)$, the *distributional gradient* $\nabla f \in C_c^1(\Omega, \mathbb{R}^d)^*$ is given by

$$\nabla f(\psi) := -\int_\Omega f \operatorname{div} \psi \, \mathrm{d}x \qquad \forall \psi \in C_c^1(\Omega, \mathbb{R}^d). \tag{12}$$

Based on this, the *total variation* of $f \in L^1(\Omega, \mathbb{R})$ is introduced as

$$\mathrm{TV}(f) := \sup\big\{\nabla f(\psi) : \psi \in C_c^1(\Omega, \mathbb{R}^d), \|\psi\|_\infty \leq 1\big\}, \tag{13}$$

and the *functions of bounded variation* are defined as

$$\mathrm{BV}(\Omega) := \big\{f \in L^1(\Omega) : \mathrm{TV}(f) < \infty\big\}. \tag{14}$$

Equipped with the norm $\|f\|_{BV} := \|f\|_{L^1(\Omega)} + \text{TV}(f)$, $BV(\Omega)$ becomes a Banach space. A function $\mu \colon \mathcal{B}(\Omega) \to \mathbb{R}^d$ on the Borel σ-algebra $\mathcal{B}(\Omega)$ is called a *vector-valued Radon measure* if every coordinate function $\mu_i \colon \mathcal{B}(\Omega) \to \mathbb{R}$ is a Radon measure. We denote by $\mathcal{M}(\Omega, \mathbb{R}^d)$ the *space of vector-valued finite Radon measures*. Due to the Riesz–Markov– Kakutani representation theorem, it holds $C_0(\Omega, \mathbb{R}^d)^* \cong \mathcal{M}(\Omega, \mathbb{R}^d)$, where $C_0(\Omega, \mathbb{R}^d)$ denotes the continuous functions vanishing at infinity. Hence, we can equip $\mathcal{M}(\Omega, \mathbb{R}^d)$ with the associated *weak* convergence*. For $f \in BV(\Omega, \mathbb{R}^d)$ it holds

$$|\nabla f(\psi)| \le \text{TV}(f) \|\psi\|_\infty \quad \forall \psi \in C_c^1(\Omega, \mathbb{R}^d), \tag{15}$$

and since $C_c^1(\Omega, \mathbb{R}^d)$ is dense and continuously embedded in $C_0(\Omega, \mathbb{R}^d)$, the gradient ∇f can be uniquely extended to a continuous linear functional on $C_0(\Omega, \mathbb{R}^d)$ using the Hahn– Banach theorem. Therefore, we can associate a unique measure $Df \in \mathcal{M}(\Omega, \mathbb{R}^d)$ to ∇f such that

$$\nabla f(\psi) = \sum_{i=1}^{d} \int_\Omega \psi_i \, dDf_i \quad \forall \psi \in C_c^1(\Omega, \mathbb{R}^d). \tag{16}$$

Hence, $BV(\Omega)$ consists of those functions $f \in L^1(\Omega)$ having a distributional gradient that is a finite Radon measures. Note that the domain of TV can be extended to $L^p(\Omega)$, $p \in [1, \infty)$, via

$$\text{TV}(f) := \begin{cases} \text{TV}(f) & \text{for } f \in BV(\Omega), \\ +\infty & \text{for } f \in L^p(\Omega) \backslash BV(\Omega). \end{cases} \tag{17}$$

It is well-known that this extension is proper, convex and (weakly) lower semi-continuous (lsc) [7, Lem. 6.105]. Furthermore, we have for $p \le \frac{d}{d-1}$ the continuous embedding $BV(\Omega) \hookrightarrow L^p(\Omega)$ as well as the *Poincaré–Wirtinger inequality* $\|P_0 f\|_p \le \text{TV}(f)$, where $P_0 f := f - |\Omega|^{-1} \int_\Omega f \, dx$. More precisely, P_0 is the projection onto the complement of the subspace Π_0 of the constant functions. The projection onto Π_0 itself is denoted by Q_0. Consequently, $\text{TV}(f)$ is coercive in the sense that $\|P_0 f_n\|_p \to \infty$ implies $\text{TV}(f_n) \to \infty$. This can be used to prove coercivity of functionals involving TV regularization, provided that the remaining terms are coercive with respect to $\|Q_0 f_n\|_p$.

3 Regularizing Properties

Here, we study the regularizing properties of the reconstruction model (10) following the considerations in [23, Sec. 3] and making the necessary modifications due to additional source term z. **General case** Throughout this section, we assume that V fulfills the regularity requirements from Theorem 1, i.e., $V \hookrightarrow C_0^{1,\alpha}(\Omega, \mathbb{R}^d)$ for some $0 < \alpha \le 1$. Regarding the data fidelity term $\mathcal{D}$, the forward operator K, and the regularizers $\mathcal{E}_1$, $\mathcal{E}_2$, we make the following assumptions:

1. The operator $K \colon L^2(\Omega) \to Y$ is continuous and weak-weak-continuous, i.e., $x_n \rightharpoonup x$ in $L^2(\Omega)$ implies $K(x_n) \rightharpoonup K(x)$ in Y.
2. The functional $\mathcal{D}(\cdot, g)$ is weakly lsc for all $g \in Y$ and $D(f, \cdot)$ is continuous for all $f \in Y$.
3. If $\mathcal{D}(f, g) = 0$, then it holds that $f = g$.
4. There exists $C > 0$ such that it holds $|\mathcal{D}(f, h) - \mathcal{D}(f, g)| \le C\mathcal{D}(g, h)$ for all $f, g, h \in Y$.
5. For fixed $g \in Y$, any bounded sequence $\{f_n\}_{n \in \mathbb{N}} \subset L^2(\Omega)$ and any sequence $\{z_n\}_{n \in \mathbb{N}} \subset L^2(\Omega)$ with $\|z_n\|_2 \to \infty$ it holds that $\mathcal{D}(K(f_n + z_n), g) + \mathcal{E}_2(z_n) \to \infty$. Further, for

$\{v_n\}_{n\in\mathbb{N}} \subset \mathcal{V}$ with $\|v_n\|_\mathcal{V} \to \infty$ it holds that $\mathcal{E}_1(v_n) \to \infty$. This can be interpreted as coercivity in v and z.

6. The regularizers $\mathcal{E}_1$ and $\mathcal{E}_2$ are weakly lsc.

Remark 1 These conditions readily imply that if $\{f_n\}_{n\in\mathbb{N}}, \{g_n\}_{n\in\mathbb{N}}$ are sequences in Y with $f_n \rightharpoonup f$ and $g_n \to g$, then $\liminf_{n\to\infty} \mathcal{D}(f_n, g) = \liminf_{n\to\infty} \mathcal{D}(f_n, g_n)$. Further, it is easy to verify that Conditions 1-4 are fulfilled if $\mathcal{D}$ is a metric and K a bounded linear operator.

First, we prove existence of a minimizer for (10).

Theorem 2 *(Existence of minimizers) For any $\lambda_1, \lambda_2 > 0$ and $g \in Y$, the problem (10) has a minimizer.*

Proof Let $\{v_n, z_n\}_{n\in\mathbb{N}} \subset \mathcal{V} \times L^2(\Omega)$ be a minimizing sequence for (10). As $\mathcal{E}_1(v_n) \leq J(v_0, z_0)/\lambda_1$ holds for all $n \in \mathbb{N}$, the sequence $\{v_n\}_{n\in\mathbb{N}}$ is bounded in $\mathcal{V}$ by Condition 5. Hence, there exists a subsequence, also denoted with $\{v_n\}_{n\in\mathbb{N}}$, such that $v_n \rightharpoonup v^*$ for some $v^* \in \mathcal{V}$. By Theorem 1 and the discussion thereafter, the sequence $f_n := T \circ \varphi_{0,v_n}^{-1}(1, \cdot)$ converges strongly in $L^2(\Omega)$ to $f := T \circ \varphi_{0,v^*}^{-1}(1, \cdot)$ and is thus bounded. By Condition 5, we conclude that also $\{z_n\}_{n\in\mathbb{N}}$ is bounded. Therefore, we can extract a weakly convergent subsequence, also denoted with $\{z_n\}_{n\in\mathbb{N}}$, with limit $z^* \in L^2(\Omega)$. As K is weak-weak-continuous (Condition 1), we obtain $K(R_{v_n,z_n}) \rightharpoonup K(R_{v^*,z^*})$. Since $\mathcal{D}(\cdot, g), \mathcal{E}_1$ and $\mathcal{E}_2$ are all weakly lsc, we get $J(v^*, z^*) \leq \liminf_{n\to\infty} J(v_n, z_n) = \inf_{v,z} J(v, z)$. Hence, (v^*, z^*) is a minimizer for (10). □

Note that (10) is non-convex. Hence, we cannot expect uniqueness of the minimizer. If K is nonlinear, it appears sensible to require that it is completely continuous, i.e., that it maps weakly convergent sequences to strongly convergent ones. In this case, we can relax the constraints for $\mathcal{D}$ by considering operators that are only lsc. Next, we provide a result regarding the continuous dependence of minimizers for (10) on the data $g \in Y$.

Theorem 3 *(Dependence on the data) Consider a sequence $g_n \to g$ in Y. For each $n \in \mathbb{N}$, let $(v_n, z_n) \in \mathcal{V} \times L^2(\Omega)$ be a minimizer of the functional $J_n := J_{\lambda, g_n}$, where $\lambda = (\lambda_1, \lambda_2)$. Then, there exists a subsequence of $\{v_n, z_n\}_{n\in\mathbb{N}}$ that converges weakly to a minimizer of $J := J_{\lambda, g}$.*

Proof As (v_n, z_n) minimizes J_n, it holds

$$\lambda_1 \mathcal{E}_1(v_n) + \lambda_2 \mathcal{E}_2(z_n) \leq J_n(v_n, z_n) \leq J_n(0, 0)$$
$$= \mathcal{D}\big(K(R_{0,0}), g_n\big) + C \to \mathcal{D}\big(K(R_{0,0}), g\big) + C, \quad (18)$$

where the convergence follows by continuity of $\mathcal{D}$ in its second entry (Condition 1). Similarly as in Theorem 2, we choose a subsequence of $\{v_n\}_{n\in\mathbb{N}}$ with limit v^* such that $f_n := T \circ \varphi_{0,v_n}^{-1}(1, \cdot)$ converges strongly in $L^2(\Omega)$ to $f := T \circ \varphi_{0,v^*}^{-1}(1, \cdot)$, i.e., $\{f_n\}_{n\in\mathbb{N}}$ is bounded. By Condition 4, it holds that

$$\mathcal{D}\big(K(f_n + z_n), g\big) \leq \mathcal{D}\big(K(f_n + z_n), g_n\big) + C\mathcal{D}(g, g_n). \quad (19)$$

Now, Condition 5 implies that also $\{z_n\}_{n\in\mathbb{N}}$ is bounded. Hence, we may extract a weakly convergent subsequence $\{v_n, z_n\}_{n\in\mathbb{N}}$ with limit $(v^*, z^*) \in \mathcal{V} \times L^2(\Omega)$ and $R_{v_n,z_n} \rightharpoonup R_{v,z}$ in $L^2(\Omega)$. By incorporation of (19), we get that it holds

$$\mathcal{D}\big(K(R_{v^*,z^*}), g\big) \leq \liminf_{n\to\infty} \mathcal{D}\big(K(R_{v_n,z_n}), g\big) \leq \liminf_{n\to\infty} \mathcal{D}\big(K(R_{v_n,z_n}), g_n\big). \quad (20)$$

Now, let $(\tilde{v}, \tilde{z}) \in \mathcal{V} \times L^2(\Omega)$ be arbitrary. By utilizing (20) and the weak lower semi-continuity of $\mathcal{E}_1$ and $\mathcal{E}_2$, we have

$$J(v^*, z^*) \le \liminf_{n \to \infty} J_n(v_n, z_n) \le \liminf_{n \to \infty} J_n(\tilde{v}, \tilde{z}). \tag{21}$$

The continuity of $\mathcal{D}$ implies $\mathcal{D}(K(R_{\tilde{v}, \tilde{z}}), g_n) \to \mathcal{D}(K(R_{\tilde{v}, \tilde{z}}), g)$. Thus, it holds that

$$J(v^*, z^*) \le \liminf_{n \to \infty} J_n(\tilde{v}, \tilde{z}) = J(\tilde{v}, \tilde{z}), \tag{22}$$

which concludes the proof. □

We conclude our investigation with a convergence result for vanishing noise, provided that we use an appropriate parameter choice rule $\lambda = \gamma(\delta)$. This enables us to approximate solutions of (1).

Theorem 4 (*Convergence for vanishing noise*) *Let* $T \in L^2(\Omega)$, $g \in Y$ *and assume that there exists* $(\hat{v}, \hat{z}) \in \mathcal{V} \times L^2(\Omega)$ *with* $K(R_{\hat{v}, \hat{z}}) = g$. *Further, assume that* $\gamma_i : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$, $i = 1, 2$, *satisfy* $\gamma_i(\delta) \to 0$, $\delta/\gamma_1(\delta) \to 0$ *for* $\delta \to 0$ *and* $\gamma_2(\delta)/\gamma_1(\delta) \to c > 0$. *Choose a sequence* $\{\delta_n\}_{n \in \mathbb{N}} \subset \mathbb{R}_{>0}$ *that converges to zero, and a sequence* $\{g_n\}_{n \in \mathbb{N}} \subset Y$ *with* $\mathcal{D}(g, g_n) \le \delta_n$. *Furthermore, let* (v_n, z_n) *be a minimizer of* $J_n := J_{\lambda_n, g_n}$ *with* $\lambda_n := (\lambda_{1,n}, \lambda_{2,n}) = (\gamma_1(\delta_n), \gamma_2(\delta_n))$ *for each* $n \in \mathbb{N}$. *Then there exists a subsequence of* $\{v_n, z_n\}_{n \in \mathbb{N}}$ *that converges weakly to a point* (v^*, z^*) *with* $K(R_{v^*, z^*}) = g$.

Proof For every $n \in \mathbb{N}$, it holds that

$$\mathcal{E}_1(v_n) + \frac{\lambda_{2,n}}{\lambda_{1,n}} \mathcal{E}_2(z_n) \le \frac{1}{\lambda_{1,n}} J_n(v_n, z_n) \le \frac{1}{\lambda_{1,n}} J_n(\hat{v}, \hat{z})$$

$$\le \frac{1}{\lambda_{1,n}} \mathcal{D}(g, g_n) + \mathcal{E}_1(\hat{v}) + \frac{\lambda_{2,n}}{\lambda_{1,n}} \mathcal{E}_2(\hat{z})$$

$$\le \frac{\delta_n}{\lambda_{1,n}} + \mathcal{E}_1(\hat{v}) + \frac{\lambda_{2,n}}{\lambda_{1,n}} \mathcal{E}_2(\hat{z}). \tag{23}$$

Hence, $\{v_n\}_{n \in \mathbb{N}}$ and $\{\mathcal{E}_2(z_n)\}_{n \in \mathbb{N}}$ are bounded. Further, $\mathcal{D}(K(R_{v_n, z_n}), g_n)$ is bounded, which as in Theorem 3 implies that $\{z_n\}_{n \in \mathbb{N}}$ is bounded. Therefore, $\{v_n, z_n\}_{n \in \mathbb{N}}$ is bounded and possesses a weakly convergent subsequence (without relabeling) with limit (v^*, z^*), for which the following estimate holds true

$$\mathcal{D}(K(R_{v^*, z^*}), g) \le \liminf_{n \to \infty} \mathcal{D}(K(R_{v_n, z_n}), g)$$

$$\le \liminf_{n \to \infty} \mathcal{D}(K(R_{v_n, z_n}), g_n) \le \liminf_{n \to \infty} J_n(v_n, z_n)$$

$$\le \liminf_{n \to \infty} J_n(\hat{v}, \hat{z}) \le \lim_{n \to \infty} \delta_n + \lambda_{1,n} \mathcal{E}_1(\hat{v}) + \lambda_{2,n} \mathcal{E}_2(\hat{z}) = 0. \tag{24}$$

Finally, by using Condition 3, we deduce $K(R_{v^*, z^*}) = g$. □

Specific Setting In our numerical experiments, we work with 2D images, i.e., $\Omega = (0, 1)^2 \subset \mathbb{R}^2$, and we choose $\mathcal{E}_2(z) := \mathrm{TV}(z)$. Further, the data fidelity term $\mathcal{D}$ is chosen for all examples with synthetic data as $\mathcal{D}(f, g) = \frac{1}{2} \|f - g\|^2$, and the regularizer for v as

$$\mathcal{E}_1(v) := \frac{1}{2} \int_0^T \int_\Omega \|Bv(t, x)\|^2 \, dx \, dt, \tag{25}$$

where B is a differential operator such that Condition (5) is satisfied, i.e., $\mathcal{E}_1$ is coercive in v. Since the Sobolev space $V = H_0^3(\Omega, \mathbb{R}^2)$ can be continuously embedded into $C^{1,0.5}(\Omega, \mathbb{R}^2)$,

we have to choose a matrix B that encodes all third-order derivatives in space. We want to remark that our numerical approach in Sect. 4 works for any regularizers of the form (25), even if it is not coercive.

Now, the specification of problem (10) reads

$$\min_{(v,z)\in\mathcal{V}\times L^2(\Omega)} \frac{1}{2}\|K\circ R_{v,z} - g\|^2 + \frac{\lambda_1}{2}\int_0^T\int_\Omega \|Bv(t,x)\|^2\,\mathrm{d}x\,\mathrm{d}t + \lambda_2\mathrm{TV}(z). \qquad (26)$$

In case that $K\colon L^2(\Omega)\to Y$ is linear, bounded and does not vanish for constant functions, the assumptions underlying our theoretical investigations in the previous paragraph are satisfied:

- As K is linear and bounded, it is also continuous and weak-weak-continuous.
- Due to the norm properties, $\mathcal{D}$ is weakly lsc and continuous in both entries. Further, Condition 3 and 4 also follow from the norm properties.
- We show Condition 5 explicitly. Due to our choice of B, the regularizer $\mathcal{E}_1$ is coercive. Hence, it remains to show the first part of the condition. Let $\{f_n\}_n$ be bounded and let $\|z_n\| \to \infty$. Then, either $\|P_0z_n\| \to \infty$ or $\|Q_0z_n\| \to \infty$. In the first case, we have that $\mathcal{E}_2(z_n) \to \infty$ and the claim follows by positivity of $\mathcal{D}$. For the second one, we get due to linearity of K that

$$K(f_n + z_n) = K(f_n + P_0z_n) + K(Q_0z_n), \qquad (27)$$

  where the first term remains bounded. Since $\|K(Q_0z_n)\| \to \infty$ due to the assumption that K does not vanish for constant functions, we conclude $\|K(f_n + z_n) - g\| \to \infty$ and the claim follows by positivity of $\mathcal{E}_2$.
- The regularizers $\mathcal{E}_1$ and $\mathcal{E}_2$ are chosen such that they are weakly lsc.

Remark 2 Although this is not covered theoretically, we use a normalized-cross-correlation-based distance $\mathcal{D}_{\mathrm{NCC}}\colon Y\backslash\{0\} \times Y\backslash\{0\} \to [0,1]$ given by

$$\mathcal{D}_{\mathrm{NCC}}(f,g) = 1 - \frac{\langle f,g\rangle^2}{\|f\|_Y^2\|g\|_Y^2} \qquad (28)$$

for our numerical experiments with real data as proposed in [23]. This modification is necessary as the gray value scale between template and target is often different for real data. It holds that $\mathcal{D}_{\mathrm{NCC}}(f,g) = 0$ only implies $f = cg$ with $c\in\mathbb{R}$, i.e., Condition 3 is violated. Further, we do not have the required coercivity with respect to z (Condition 5). Hence, this setup is not covered by our theory. Nevertheless, our numerical experiments indicate that $\mathcal{D}_{\mathrm{NCC}}$ combined with $\mathcal{E}_1 = \mathrm{TV}$ leads to good reconstructions.

If we choose $\mathcal{E}_2 = \|\cdot\|_{L^2(\Omega)}^2$ instead, we can derive the same theoretical results as for $\mathcal{D}(f,g) = \|f - g\|^2$. In this case, Condition 4 and 5 are obsolete, as they are only required to infer boundedness of $z\in L^2(\Omega)$ from the boundedness of the objective, which now follows directly. The remaining conditions are met by $\mathcal{D}_{\mathrm{NCC}}$, except for Condition 3. Hence, the convergence in Theorem 4 holds only up to a scalar.

4 Numerical Approach

In this section, we present a numerical scheme for solving (26). Our approach is based on the Lagrangian method developed in [23, 24] as well as the iPALM algorithm [4, 37]. The actual implementation relies upon the FAIR toolbox [29].

So far, we have not specified the differential operator B in (26). As our approach is guided by the modular framework of [23, 24], we could use *curvature regularization*, defined by $B = \nabla_x$, or *diffusion regularization*, where $B = \Delta_x$. These choices correspond to the H^1 and H^2 semi-norm, respectively. However, they do not satisfy the coercivity requirement, i.e., Condition 5. Instead, we use the H^3 semi-norm, which has been proposed in [23]. Although our theoretical investigations for (26) in Sect. 3 only hold for the 2D case, we provide the algorithm in general form, and also deploy it for 3D images later. In the following, we briefly sketch the components of our approach.

4.1 Discretization

We pursue a *discretize-then-optimize* strategy. By partitioning every coordinate in m blocks of length $h_X = 1/m$, the domain $(0, 1)^d$ is split into m^d equally sized cubes. Then, the template $T \in L^2(\Omega)$ and the source $z \in L^2(\Omega)$ are both sampled at the cell-centered nodes $\mathbf{x_c} \in \mathbb{R}^{m^d}$, resulting in discrete versions $\mathbf{T(x_c)}$ and $\mathbf{z(x_c)} \in \mathbb{R}^{m^d}$, respectively. Their values are interpolated by cubic B-splines if off grid values are required. Further, the time domain $[0, 1]$ is uniformly partitioned into m_t units of length $h_t = 1/m_t$. Then, the velocity $v: [0, 1] \times \Omega \to \mathbb{R}^d$ is sampled over cell-centered locations in space and at the nodes in time, resulting in a discrete velocity vector $\mathbf{v} \in \mathbb{R}^N$ with $N = d \cdot (m_t + 1) \cdot m^d$.

Lagrangian solver for $\mathbf{R_{v,z}}$ In order to compute the solution map $(\mathbf{v, z}) \mapsto \mathbf{R_{v,z}}$, we need to solve the flow equation (11). Here, every function $\varphi_{s,v}(\cdot, x_0): [0, 1] \to \Omega$ can be interpreted as a trajectory of some particle with position x_0 at initial time s. We compute $\mathbf{R_{v,z}}$ as follows:

1. *Computing the characteristics*: For numerically solving (11), we employ a fourth order Runge–Kutta scheme (RK4). Since we require $\varphi_{1,\mathbf{v}}(0, \cdot)$, we solve (11) backwards in time with N_t equidistant steps of size $\Delta t = -\frac{1}{N_t}$ and initial condition $\varphi_{1,\mathbf{v}}(1, \mathbf{x_c}) = \mathbf{x_c}$. To simplify the notation, the remaining discussion is instead based on the explicit Euler scheme

$$\varphi_{1,\mathbf{v}}(t_{k+1}, \mathbf{x_c}) = \varphi_{1,\mathbf{v}}(t_k, \mathbf{x_c}) + \Delta t \, I\big(\mathbf{v}, t_k, \varphi_{1,\mathbf{v}}(t_k, \mathbf{x_c})\big), \tag{29}$$

where $k = 0, \ldots, N_t - 1$ and $t_k = 1 - k\Delta t$. Here, I interpolates the velocity at time t_k and transformed positions $\varphi_{1,\mathbf{v}}(t_k, \mathbf{x_c})$. This is necessary as the points $\varphi_{1,\mathbf{v}}(t_k, \mathbf{x_c})$ are in general not on the grid. Note that the time discretization parameters N_t and m_t differ in general: The first determines the accuracy of the ODE solver, whereas the second is related to the discretization.

2. *Deforming the template T*: Based on the output $\varphi_{1,\mathbf{v}}(0, \mathbf{x_c})$ of the RK4-scheme, we can evaluate the template $\mathbf{T}$ at the deformed grid using interpolation.

3. *Computing $\mathbf{R_{v,z}}$*: Finally, we add the source term z and the deformed template T, i.e., $\mathbf{R_{v,z}(x_c)} = \mathbf{T} \circ \varphi_{1,\mathbf{v}}(0, \mathbf{x_c}) + \mathbf{z(x_c)}$.

Note that actually all steps of this procedure are independent of the forward operator K, the data fidelity term $\mathcal{D}$ and the chosen regularizers $\mathcal{E}_1, \mathcal{E}_2$. Further, the gradient $\nabla_{\mathbf{v}} \mathbf{R_{v,z}}$ can be explicitly computed within the Runge–Kutta scheme, which is important for computational efficiency.

Forward operator Let us denote by $\mathbf{K}: \mathbb{R}^{m^d} \to \mathbb{R}^M$, $M \in \mathbb{N}$, a finite-dimensional, Fréchet differentiable approximation of the operator $K: L^2(\Omega, \mathbb{R}) \to Y$. With the application to CT in mind, we discuss a discretization of the d-dimensional Radon transform. More

generally, if a discretization $\mathbf{K}$ of some operator K is given, we can simply insert it into the model (26).

For given $\theta \in \mathbb{S}^{n-1}$ and $s \in \mathbb{R}$, the Radon transform of $f : \mathbb{R}^d \to \mathbb{R}$ is defined pointwise by

$$Rf(\theta, s) = \int_{\theta^\perp} f(s\theta + y) \, \mathrm{d}y, \tag{30}$$

where $\theta^\perp$ is the orthogonal complement of span$\{\theta\}$ [31, Chap. 2]. The Radon transform is linear and thus also Fréchet differentiable. For $f \in L^2(\Omega)$, the value $R(f) \in Y$ is a function that maps from the cylinder $\mathbb{S}^{d-1} \times \mathbb{R}$ to $\mathbb{R}$. We discretize this cylinder as follows: Take $p \in \mathbb{N}$ directions in $\mathbb{S}^{d-1}$. For simplicity, we say that we take one measurement in each direction. Furthermore, we take the interval $(0, 1)$ instead of $\mathbb{R}$, and split it into $q \in \mathbb{N}$ equally sized cells of length $1/q$, as we have also done with Ω. Depending on the dimension d and the diameter of Ω, the intervals length requires adjustment. Then, the data is sampled at cell-centered points $\mathbf{y_c}$ for each angle, resulting in vectors $\mathbf{g_i}(\mathbf{y_c}) \in \mathbb{R}^q$, $i = 1, \ldots, p$, and the entire data vector is represented as $\mathbf{g} \in \mathbb{R}^M$ with $M = p \cdot q$. A discrete Radon transform is implemented for both CPUs and GPUs as part of the ASTRA toolbox [36, 45, 46].

Data fidelity term and regularizers We discretize the data fidelity term $\mathcal{D}(x, y) = \frac{1}{2}\|x - y\|_2^2$ using the midpoint-rule for numerical integration, which results in

$$\mathcal{D}_{SSD}(\mathbf{x}, \mathbf{y}) = \frac{1}{2} h_Y (\mathbf{x} - \mathbf{y})^T (\mathbf{x} - \mathbf{y}) \tag{31}$$

with $h_Y = 1/q$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^M$, see also [29, Chap. 6.2]. As we consider the Radon transform for few directions $\theta \in \mathbb{S}^{n-1}$ only, we disregard the necessary modifications related to the integration over the unit sphere. Similarly, the discretization of the regularizer $\mathcal{E}_1$ is given by

$$\mathcal{E}_1(\mathbf{v}) = \frac{1}{2} h_t h_X^d \mathbf{v}^T \mathbf{B}^T \mathbf{B} \mathbf{v}, \tag{32}$$

where $\mathbf{B}$ is a finite-difference counterpart of the chosen differential operator B. To mitigate boundary effects caused by the discretization of B, we pad the spatial domain and impose zero Neumann boundary conditions. For the TV regularizer, the norm is again discretized based on the midpoint-rule and the gradient is replaced by a finite difference counterpart, resulting in

$$\mathrm{TV}(\mathbf{z}) = h_X^d \sum_{i=1}^{m^d} \|(\nabla_{h_X}(\mathbf{z}))_i\|. \tag{33}$$

In our experiments, we employ backward differences and the boundary is extended by 0. To this end, we denote by $z \in \bigotimes_{i=1}^d \mathbb{R}^m$ the tensor representation of $\mathbf{z} \in \mathbb{R}^{m^d}$. For any $i = 1, \ldots, m^d$, $\mathbf{z}_i$ corresponds to $z_{(i)} = z_{i_1,\ldots,i_d}$ with $i_k = 1, \ldots m$. Then, it holds that $(\nabla_{h_X}(\mathbf{z}))_i = (\partial_k z_{(i)})_{k=1}^d$ with

$$(\partial_k z_{(i)}) = \begin{cases} \frac{1}{h_x}\left(z_{i_1,\ldots,i_k+1,\ldots,i_d} - z_{(i)}\right) & \text{if } i_k < m, \\ 0 & \text{else.} \end{cases} \tag{34}$$

4.2 iPALM

Putting all parts from Sect. 4.1 together, we obtain the discrete problem

$$\min_{(\mathbf{v},\mathbf{z})\in\mathbb{R}^N\times\mathbb{R}^{m^d}} \mathcal{D}_{SSD}\big(\mathbf{K}(\mathbf{R}_{\mathbf{v},\mathbf{z}}),\mathbf{g}\big) + \frac{\lambda_1}{2}h_t h_X^d \mathbf{v}^T \mathbf{B}^T \mathbf{B}\mathbf{v} + \lambda_2 h_X^d \sum_{i=1}^{d\cdot m^d} \|(\nabla_{h_X}(\mathbf{z}))_i\|. \quad (35)$$

In the following, we deploy the *inertial proximal alternating linearized minimization* (iPALM) algorithm [4, 37] for solving (35). This scheme can solve generic problems of the form

$$\min_{x,y\in E_1\times E_2} \Psi(x,y) := G_1(x) + G_2(y) + H(x,y), \quad (36)$$

where E_1, E_2 are Euclidean spaces, $H \in C^1(E_1 \times E_2)$ and $G_i \in \Gamma_0(E_i)$, $i = 1, 2$, namely the proper, convex, and lsc functions on E_i. The corresponding iterations are given by

$$
\begin{aligned}
\bar{x}_k &= x_k + \alpha_k(x_k - x_{k-1}) \\
x_{k+1} &= \text{prox}_{\sigma_k^1 G_1}\left(\bar{x}_k - \frac{1}{\sigma_k^1}\nabla_x H(\bar{x}_k, y_k)\right) \\
\bar{y}_k &= y_k + \alpha_k(y_k - y_{k-1}) \\
y_{k+1} &= \text{prox}_{\sigma_k^2 G_2}\left(\bar{y}_k - \frac{1}{\sigma_k^2}\nabla_y H(x_{k+1}, \bar{y}_k)\right),
\end{aligned} \quad (37)
$$

where $\text{prox}_{\tau f}(x) = \arg\min_y \frac{1}{2}\|x - y\|_2^2 + \tau f(y)$ is the proximal mapping of f. The step-sizes $\sigma_k^1, \sigma_k^2 > 0$ are chosen according to the respective partial Lipschitz-constants of ∇H, and $\alpha_k > 0$ is an inertia parameter, which helps to escape from local minima and boosts the convergence speed. In [37], the convergence of the iterations (37) is proven under the assumption that the objective has the *Kurdyka-łojasiewicz (KL) property*. Among others, this property holds for *semi algebraic* functions, which include real polynomials, the $\|\cdot\|_p$-norm with rational, non-negative p, indicator-functions of semi-algebraic sets, as well as functions of the form $x \mapsto \sup\{g(x,y): y \in S\}$, where S is semi-algebraic and g is a semi-algebraic function, see [4]. Furthermore, compositions of semi-algebraic functions are semi-algebraic.

Theorem 5 *([37, Thm. 4.1]) Let E_1, E_2 be Euclidean spaces, $H \in C^1(E_1 \times E_2)$, and $G_i \in \Gamma_0(E_i)$, $i = 1, 2$, such that Ψ in (36) has the KL property. Further, let all functions have finite infima. Assume that ∇H is locally Lipschitz continuous and that both $x_i \mapsto \nabla_{x_i} H(x_1, x_2)$ are globally Lipschitz, where the constants $L_1(x_2)$, $L_2(x_1)$ possibly depend on the fixed variable and are bounded on compact sets. Finally, assume*

$$\sigma_k^1 = \frac{1 + 2\alpha_k}{2(1 - \alpha_k)}L_1(y_k) \quad and \quad \sigma_k^2 = \frac{1 + 2\alpha_k}{2(1 - \alpha_k)}L_2(x_{k+1}) \quad (38)$$

and $\alpha_k < 0.5$ for every $k \in \mathbb{N}$. If the sequence generated by (37) is bounded, then it converges to a critical point.

Remark 3 Although this is not supported by Theorem 5, it turned out that choosing $\alpha_k = (k-1)/(k-2)$, $\sigma_k^1 = L_1(y_k)$ and $\sigma_k^2 = L_2(x_{k+1})$ works very well in practice [37]. Unfortunately, the Lipschitz-moduli $L(x_k)$, $L(y_k)$ are often unknown or difficult to compute. Instead, a backtracking scheme can be used to ensure this condition, see [37] for details.

Next, we want to apply iPALM to problem (35). Both regularizers and the data fidelity term have the KL property since they are computed via composition of an affine function and a permissible norm function. Further, the operation $(\mathbf{v}, \mathbf{z}) \mapsto \mathbf{R}_{\mathbf{v},\mathbf{z}}$ is polynomial in $(\mathbf{v}, \mathbf{z})$ for every component as the RK4 scheme and linear interpolators provide a composition of polynomial functions. Hence, the objective in (35) has the KL property. We propose to use the splitting

$$H: \mathbb{R}^N \times \mathbb{R}^{m^d} \to \mathbb{R}, \quad (\mathbf{v}, \mathbf{z}) \mapsto \mathcal{D}_{SSD}\big(\mathbf{K}(\mathbf{R}_{\mathbf{v},\mathbf{z}}), \mathbf{g}\big), \tag{39}$$

$$G_1: \mathbb{R}^N \to \mathbb{R}, \qquad \mathbf{v} \mapsto \frac{\lambda_1}{2} h_t h_X^d \mathbf{v}^T \mathbf{B}^T \mathbf{B} \mathbf{v}, \tag{40}$$

$$G_2: \mathbb{R}^{m^d} \to \mathbb{R}, \qquad \mathbf{z} \mapsto \lambda_2 h_X^d \sum_i \|(\nabla_{h_X}(\mathbf{z}))_i\|. \tag{41}$$

The partial gradients of H are given by

$$\nabla_{\mathbf{v}} H(\mathbf{v}, \mathbf{z}) = h_Y \big(\mathbf{K}(\mathbf{R}_{\mathbf{v},\mathbf{z}}) - \mathbf{g}\big)^{\mathrm{T}} \partial\mathbf{K}(\mathbf{R}_{\mathbf{v},\mathbf{z}}) \frac{\partial}{\partial\mathbf{v}}\mathbf{R}_{\mathbf{v},\mathbf{z}}, \tag{42}$$

$$\nabla_{\mathbf{z}} H(\mathbf{v}, \mathbf{z}) = h_Y \big(\mathbf{K}(\mathbf{R}_{\mathbf{v},\mathbf{z}}) - \mathbf{g}\big)^{\mathrm{T}} \partial\mathbf{K}(\mathbf{R}_{\mathbf{v},\mathbf{z}}), \tag{43}$$

where $\partial\mathbf{K}$ refers to the Fréchet-derivative of the operator $\mathbf{K}$. If $\mathbf{K}$ is linear, its derivative coincides with the operator itself. The derivative $\frac{\partial}{\partial\mathbf{v}}\mathbf{R}_{\mathbf{v},\mathbf{z}}$ of the solution map is given by

$$\frac{\partial}{\partial\mathbf{v}}\mathbf{R}_{\mathbf{v},\mathbf{z}} = \nabla_x \mathbf{T}\big(\varphi_{1,\mathbf{v}}(0, \mathbf{x_c})\big) \frac{\partial}{\partial\mathbf{v}}\varphi_{1,\mathbf{v}}(0, \mathbf{x_c}). \tag{44}$$

Here, $\nabla_x \mathbf{T}$ is the gradient of the interpolated template T. The derivative $\frac{\partial}{\partial\mathbf{v}}\varphi_{1,\mathbf{v}}(0, \mathbf{x_c})$ can be computed recursively within the ODE solver for (11). Exemplary, we obtain for the explicit Euler scheme (29) that

$$\frac{\partial}{\partial\mathbf{v}}\varphi_{1,\mathbf{v}}(t_{k+1}, \mathbf{x_c}) = \frac{\partial}{\partial\mathbf{v}}\varphi_{1,\mathbf{v}}(t_k, \mathbf{x_c}) + \Delta t \frac{\partial}{\partial\mathbf{v}} I\big(\mathbf{v}, t_k, \varphi_{1,\mathbf{v}}(t_k, \mathbf{x_c})\big)$$
$$+ \Delta t \frac{\partial}{\partial\varphi} I\big(\mathbf{v}, t_k, \varphi_{1,\mathbf{v}}(t_k, \mathbf{x_c})\big) \frac{\partial}{\partial\mathbf{v}}\varphi_{1,\mathbf{v}}(t_k, \mathbf{x_c}) \tag{45}$$

for all $k = 0, \ldots, N_t - 1$, see also [30].

Further, we require the proximal mappings of G_1 and G_2. For G_1, it holds that

$$\mathrm{prox}_{\sigma_1 G_1}(\mathbf{v}) = \arg\min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2\sigma_1}\|\mathbf{x} - \mathbf{v}\|^2 + \frac{1}{2}\gamma h_t h_X^d \mathbf{x}^T \mathbf{B}^T \mathbf{B} \mathbf{x}, \tag{46}$$

where the minimum is determined by the linear system of equations $(\mathrm{Id} + \sigma_1 \gamma h_t h_X^d \mathbf{B}^T \mathbf{B})\mathbf{x} = \mathbf{v}$. This system is sparse and efficiently solvable with a preconditioned conjugate gradient method. Regarding G_2, we need to compute $\mathrm{prox}_{\sigma_2 \mathrm{TV}}$ with the discrete TV (33). As outlined in [7], this can be done using the primal dual hybrid gradient method [9].

The step-sizes in the iPALM scheme (37) are chosen as the partial Lipschitz constants of $\nabla_{\mathbf{v}} H(\mathbf{v}, \mathbf{z})$ and $\nabla_{\mathbf{z}} H(\mathbf{v}, \mathbf{z})$, respectively. For a linear $\mathbf{K}$, it holds $\nabla_{\mathbf{z}} H(\mathbf{v}, \mathbf{z}) = h_Y \mathbf{K}^T \mathbf{K}(\mathbf{f}(\mathbf{R}_{\mathbf{v},\mathbf{z}}) - \mathbf{g})$. Hence, $\mathbf{z} \to \nabla_{\mathbf{z}} H(\mathbf{v}, \mathbf{z})$ is Lipschitz with $L_2(\mathbf{v}) = h_Y \|\mathbf{K}^T \mathbf{K}\|$ for every $\mathbf{v}$. Unfortunately, an upper bound for L_1 cannot be derived explicitly. Hence, we have to rely on backtracking instead.

Remark 4 (Normalized-cross-correlation-based distance) For $\mathcal{D}_{\mathrm{NCC}}$, we use the discretization

$$\mathcal{D}_{\mathrm{NCC}}(\mathbf{x}, \mathbf{y}) = 1 - \frac{(\mathbf{x}^\top \mathbf{y})^2}{\|\mathbf{x}\|^2 \|\mathbf{y}\|^2}, \tag{47}$$

for which the derivative is given by

$$\frac{\partial}{\partial \mathbf{x}} \mathcal{D}_{\text{NCC}}(\mathbf{x}, \mathbf{y}) = -2\frac{(\mathbf{x}^\top \mathbf{y})\mathbf{y}}{\|\mathbf{x}\|^2 \|\mathbf{y}\|^2} + 2\frac{(\mathbf{x}^\top \mathbf{y})^2 \mathbf{x}}{\|\mathbf{x}\|^4 \|\mathbf{y}\|^2}. \tag{48}$$

Incorporating (48), we can perform the according gradient steps in the iPALM scheme (37). Then, also a line search for estimating the Lipschitz constant with respect to z is needed. As discussed in Remark 2, $\mathcal{D}_{\text{NCC}}$ should be paired with an L^2-regularizer for the source variable to get theoretical guarantees. Then, we obtain $\mathcal{E}_2(z) = \|z\|^2_{L^2(\Omega)}$, which we discretize similarly as $\mathcal{D}_{\text{SSD}}$, namely as

$$G_2 \colon \mathbb{R}^{m^d} \to \mathbb{R}, \quad \mathbf{z} \to \frac{1}{2} h_X \mathbf{z}^T \mathbf{z}. \tag{49}$$

This results in the proximal operator

$$\text{prox}_{\sigma G_2}(\mathbf{z}) = \arg\min_{\mathbf{x} \in \mathbb{R}^{m^d}} \frac{1}{2\sigma}\|\mathbf{x} - \mathbf{z}\|^2 + \frac{1}{2} h_X \mathbf{x}^T \mathbf{x} = \frac{1}{1 + \sigma/h_X}\mathbf{z}. \tag{50}$$

Multi-level approach and post-processing Due to the non-convexity of (35), we have to cope with local minima. To avoid being trapped in one, we follow a multi-level strategy [30], which also helps to reduce the computational cost. Its different levels refer to different resolutions of the template and the target. We apply iPALM at each level, starting with the coarsest resolution. Each computed minimizer is bilinearly interpolated to the next finer scale to serve as initialization.

This approach requires multi-level versions of the operator **K** and a method for downsampling the measurements g. If these are not available, iPALM can still be performed with only one scale. For the 2D Radon transform, multi-level versions of the operator **K** can be obtained with any backend that takes the discretization of the measurement geometry as an input. More precisely, assume that the number of grid cells used to discretize $\Omega \subset \mathbb{R}^2$ at the finest level is $m = 2^l$, $l \in \mathbb{N}$. In our experiments, we set the number of cells for discretizing the measurement domain $(0, 1)$ at level $k \leq l$ to $q^{(k)} = 1.5 \cdot 2^k$ and the length of each cell to $h_Y^{(k)} = 1/q^{(k)}$. Then, a multilevel representation of each measurement $\mathbf{g}_i$, $i \leq p$, at cell-centered grid points $\mathbf{y}_j = (j - 1/2)h_Y^{(k-1)}$ is given by

$$\mathbf{g}_i^{(k-1)}(\mathbf{y}_j) = \left(\mathbf{g}_i^{(k)}(\mathbf{y}_j) + \mathbf{g}_i^{(k)}\big(\mathbf{y}_j + h_Y^{(k)}\big)\right)/4, \tag{51}$$

where the denominator arises from averaging over two neighboring grid points and dividing the edge length of the image domain Ω in each coordinate direction in half.

Additionally, after applying iPALM, we deploy the second-order inexact Gauss–Newton method from [23, 24] for refining the velocity **v**. This method utilizes the same discretization and Lagrangian solver that is used for iPALM. Due to the linearity of **K** and the structure of $\mathcal{D}_{\text{SSD}}$, we can indeed fix **z** and consider the data $\tilde{\mathbf{g}} = \mathbf{g} - \mathbf{K}(\mathbf{z})$ within the corresponding LDDMM Gauss–Newton scheme for **v**. We observed that this can compensate the slower convergence rate of iPALM. Finally, note the described modifications can also be applied within the setting of Remark 4.

Remark 5 There is a vast literature on alternating minimization and forward-backward schemes with variable metrics [3, 5, 11, 16, 41], which attempt to improve the convergence speed. For the iPALM scheme (37), our simulations indicate that an adaption of the metric is most promising for **v**. To pursue this idea further, we choose a different splitting of (35) and add G_1 to H instead. Hence, $G_1 = 0$ and $\text{prox}_{G_1} = \text{Id}$ is independent of the

(a) Template image **T**.   (b) Target image **U**.   (c) Noisy data **g**.

(d) L^2-TV model (52).   (e) Our method with E_2 = TV.   (f) Our method with $E_2 = \|\cdot\|^2$.
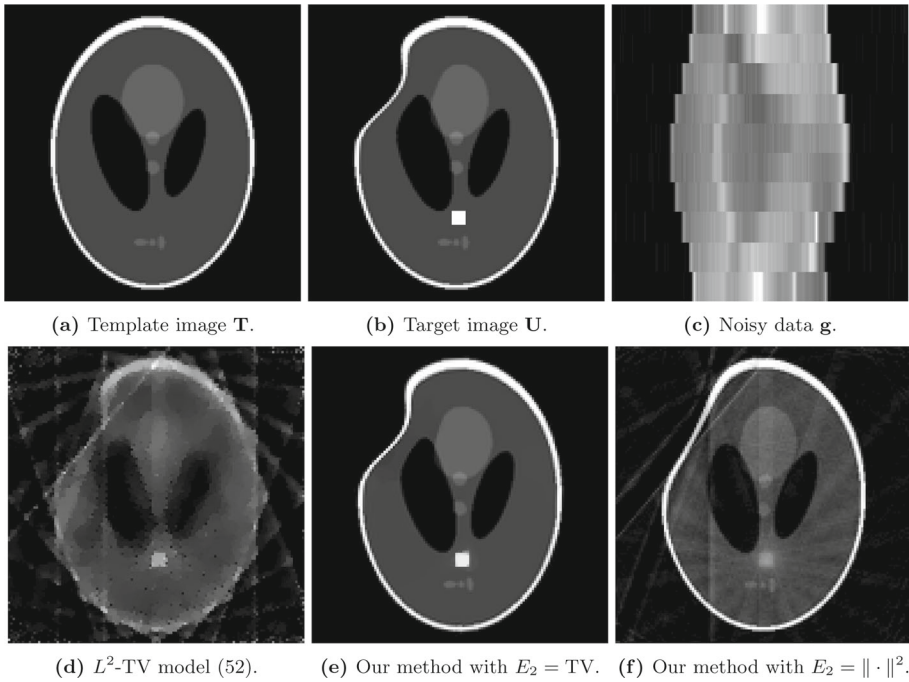
Fig. 1 Reconstruction for a cartoon like image with 10 measurement angles using our model (26) and the L^2-TV model (52) as a baseline comparison. The reconstruction **R** obtained with (26) can be decomposed into a deformation and a source part

chosen metric for the coordinate **v**. There are different strategies for constructing metrics such that convergence guarantees can be obtained, e.g., minimize-maximize strategies [11] or sparse approximations of the Hessian [3]. For a small benchmark, we deployed the same metric for **v** as proposed in [23], and solved (35) with a variable metric version of PALM, i.e., without the inertia steps. The sufficient decrease of the objective with respect to **v** is ensured with the same Armijo line search as in [23]. Experimentally, this has led to similar results as our post-processing scheme. A more thorough comparison of the approaches could be an interesting direction of future research.

5 Numerical Examples

Here, we present numerical results for our model (26), discretized and solved numerically as described in Sect. 4. The code for all experiments is available on Github.[2] Since we are mainly interested in a proof-of-concept, we only investigate CT and leave other inverse problems for future work. Our examples mostly rely on synthetic data generated from target images **U** of size 128×128 with range [0, 1]. A corresponding sinogram **g** is obtained by applying the Radon transform with ten equally distributed angles in [0, 180] to **U**. Then, we add 5% Gaussian noise to get **g**. Further, we also include one example with real data from a CT scanner. For all examples, we use a time-dependent velocity field **v** with a single

---

[2] https://github.com/anttop/FAIR.m.

time point, and 5 steps in the Runge–Kutta method that solves the associated Eq. (11). The multi-level procedure starts with resolution 32×32 at the coarsest scale. Throughout this section, all parameters are optimized via grid search.

For the first experiment, the template $\mathbf{T}$ is chosen as the Shepp-Logan phantom and the target $\mathbf{U}$ is a diffeomorphically deformed version of $\mathbf{T}$ with an additional small white square, see Fig. 1. As TV regularization favors constant areas, this image pair is well-suited for our model (26), and an almost perfect reconstruction is expected. Due to the additional structure, a good reconstruction with purely diffeomorphic approaches such as [10, 23] is impossible. The regularization parameter λ_1 splits into $\lambda_1 = [\lambda_1^1, \lambda_1^2, \lambda_1^3]$ for the spatial, temporal and $L^2(\Omega, \mathbb{R})$ regularization of $\mathbf{v}$, respectively. We have chosen $\lambda_1 = [0.001, 0.001, 10^{-6}]$ and $\lambda_2 = 0.1$. Recall that the reconstruction $\mathbf{R}$ can be decomposed into a deformation part $\mathbf{T}(\varphi^{-1}(1, \mathbf{x_c}))$ and a source part $\mathbf{z}$. We observe that the main structure is reconstructed as deformation of the template $\mathbf{T}$, whereas $\mathbf{z}$ reconstructs the additional square, see Fig. 1. This is indeed the expected behavior for properly chosen parameters. As comparison, we included a reconstruction with the standard L^2-TV model [39], i.e., the solution of

$$\min_{\mathbf{R}} \|\mathbf{K}(\mathbf{R}) - \mathbf{g}\|_2^2 + \lambda \, \mathrm{TV}(\mathbf{R}), \tag{52}$$

where we have chosen $\lambda = 0.1$ based on a grid search. For (52), we observe the typical reconstruction artifacts related to the Radon transform, i.e., rays crossing the reconstruction. If we increase λ to avoid these artifacts, the reconstruction looses details. Last, we also included a reconstruction with $E_2 = \| \cdot \|^2$ and $\lambda_2 = 0.0001$ instead of $E_2 = \mathrm{TV}$. This regularization has been investigated for the metamorphosis model (7) in [18]. As expected, we get similar artifacts as in (52) with small λ.

In our second experiment, we deal with a pair of images that contain finer details. More precisely, we have chosen an artificial brain image [19] as template $\mathbf{T}$, which is diffeomorphically deformed into the target $\mathbf{U}$ [23]. Further, we added a structure in $\mathbf{U}$ to get a non-diffeomorphic setting. We also varied the smoothness of this structure, which leads to two sub-experiments. In the first one, we added a circle with constant intensity and in the second one a 2D Gaussian. The obtained results for (26) with $\lambda_1 = [0.001, 0.1, 10^{-6}]$ and $\lambda_2 = 0.2$ are depicted in Fig. 2. Note that we increased the contrast for the error maps in order to improve the visibility. Our method is able to reconstruct all the major structures in $\mathbf{U}$. Most of the errors occur at the boundaries of the structures, which is partially due to the employed interpolation. Similarly as reported in [23], our approach struggles with the swirl in the middle of $\mathbf{U}$. This is most likely due to the large, almost non-diffeomorphic deformation and the limited amount of data. If we compare the $\mathbf{z}$ for the two sub-experiments, we observe that the reconstruction of the Gaussian is worse, which is not surprising as TV-regularization favors piece-wise constant images. In absence of topological changes, [23] demonstrated that a LDDMM-based approach without the source term can yield satisfactory results. A reconstruction with our method for $\mathbf{U}$ without the additional structure is provided in Fig. 3. For $\lambda_1 = [0.001, 0.001, 10^{-5}]$ and $\lambda_2 = 1$, the reconstruction consists only of a deformation part. However, if λ_2 is chosen too small, artifacts related to the source z appear.

Next, we comment on the robustness with respect to parameter changes. As the comparative examples in Fig. 3 and 7 show, the choice of parameters significantly impacts the quality of the reconstruction. For the images from Fig. 2, the influence of the parameter choice in terms of SSD error and SSIM value is provided in Table 1. Some corresponding reconstructions are given in Fig. 4. Changing each parameter by an order of magnitude leads to clearly visible changes in the reconstruction, see Fig. 4. In particular, too large regularization parameters can repress the effect of their respective component. In contrast, changes on a
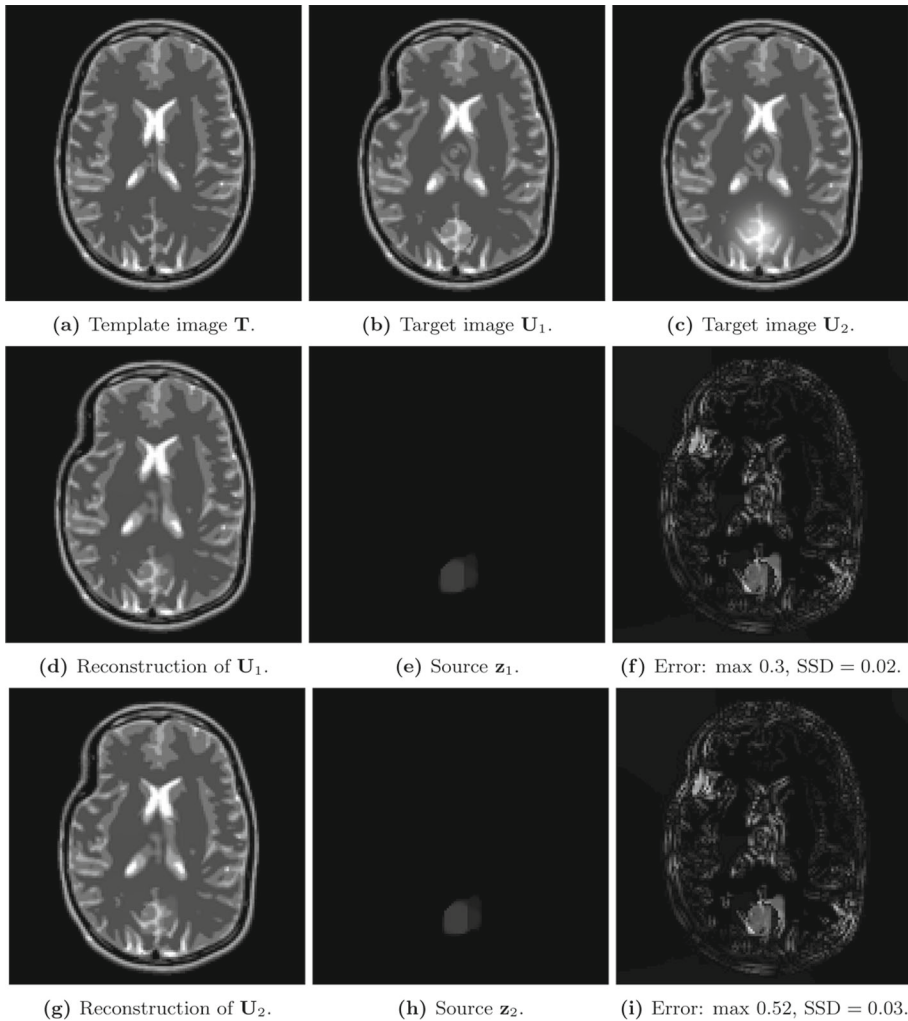
(a) Template image $\mathbf{T}$.   **(b)** Target image $\mathbf{U}_1$.   **(c)** Target image $\mathbf{U}_2$.

(d) Reconstruction of $\mathbf{U}_1$.   **(e)** Source $\mathbf{z}_1$.   **(f)** Error: max 0.3, SSD = 0.02.

(g) Reconstruction of $\mathbf{U}_2$.   **(h)** Source $\mathbf{z}_2$.   **(i)** Error: max 0.52, SSD = 0.03.

Fig. 2 Artifical brain image with data from 10 equally distributed angles in [0, 180]. The first target contains an additional circle, which is well-suited for TV regularization of z. The second one contains an additional 2D Gaussian, which is more challenging for this setting

smaller scale lead to robust reconstruction results. For the other examples from this section, we observed a similar behavior, although the precise scale depends on the underlying set of images.

For our third experiment, see Fig. 5, we have chosen two X-ray images that are not diffeomorphic to each other and contain some noise structures. Finding the correct deformation for this pair is challenging as the deformation is relatively large and irregular. In this experiment, we compare the third-order regularizer ($\lambda_1 = [0.5, 0.1, 10^{-6}]$, $\lambda_2 = 1$) with the curvature one from [24] ($\lambda_1 = [6, 6]$, $\lambda_2 = 1$), which imposes less regularity on $\mathbf{v}$. For both choices, we observe that the model struggles to bend the hand to correctly match the target. This is most noticeable at the right corner of the palm and at some of the fingers. For the third-order regularizer, we also notice that the fingers are not spread sufficiently from each other. The additional or disappearing structures are very fine, such as the noise beside the hands or the
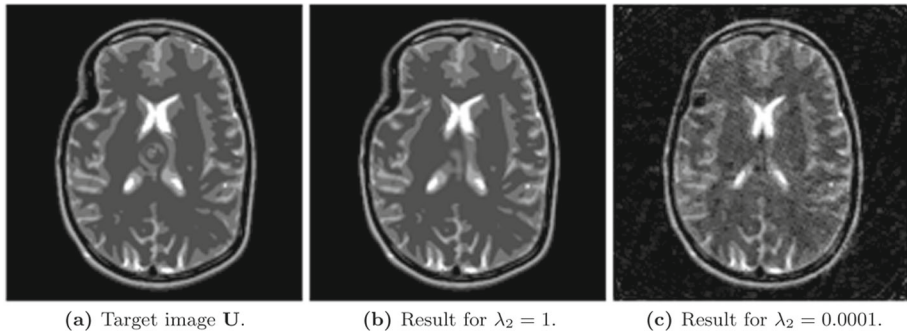
(a) Target image **U**.   (b) Result for $\lambda_2 = 1$.   (c) Result for $\lambda_2 = 0.0001$.

Fig. 3 Diffeomorphic counterpart to Fig. 2, where **U** does not contain an additional structure

(a) $\lambda_1 = 10, \lambda_2 = 2$ (b) $\lambda_1 = 0.01, \lambda_2 = 0.001$ (c) $\lambda_1 = 0.1, \lambda_2 = 0.05$ (d) $\lambda_1 = 0.1, \lambda_2 = 0.5$
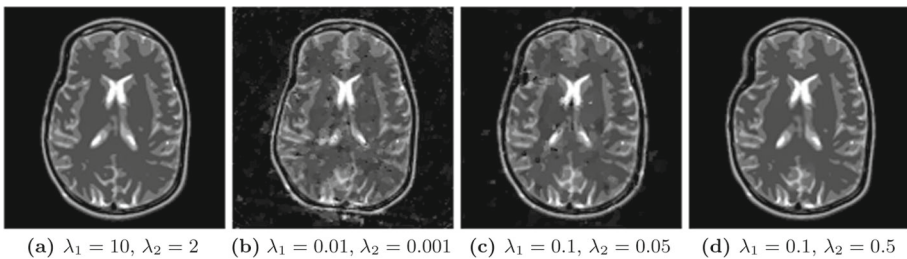
Fig. 4 Reconstructions of an artificial brain image with various parameters. Drastic changes lead to artifacts (4a–b), whereas smaller ones lead to similar reconstructions (4c–d) as in Fig. 5d

slight change in intensity on the edges of the bones. Hence, these are not well-suited for TV regularization and not reconstructed by our method. However, even if we would use a different E_2, it is questionable if the amount of data suffices to reconstruct these structures. More precisely, reconstructing fine details without sufficient data or an appropriate prior is in general impossible. Nevertheless, this experiment shows that our model (26) allows to align image pairs with large deformations between them, even under the presence of noise.

In our last synthetic experiment, we demonstrate that our method is in principal also applicable to 3D CT reconstruction problems, see Fig. 6. To this end, we use an image pair from [30] and the same 10 simulated 2D measurements as in [23], which correspond to an rotation around the third coordinate axis with angles equally distributed in [0, 180]. Due to the problem size, we use curvature regularization on **v** with $\lambda_1 = [0.07, 0.07]$ for the spatial and temporal components, respectively, and for the source **z** we use the regularization parameter $\lambda_2 = 0.01$. Overall, we obtain a satisfying reconstruction with a SSIM of 0.9060, which is significantly better than the SSIM of 0.8807 obtained by [23].

In our final experiment, we tackle real CT data from a lotus root cross-section [8]. Since the recorded data is dense, the underlying target **U** can be reconstructed via filtered back-projection. Retroactively, we deformed the computed target **U** and removed a hole in the lotus root to obtain a template **T** with a different topology, see Fig. 7. To get a sparse setting, we subsampled **g** with 12 uniformly distributed angles in [0, 180]. As we are already dealing with real data, no additional noise is added. Unfortunately, the intensity range of the given sinogram does not match the range of our Radon transform operator **K**. Therefore, $\mathcal{D}_{SSD}$ is no sensible choice for comparing the real data with the simulated data **K**(**R**). Instead, we use $\mathcal{D}_{NCC}$, which is invariant to the scaling of the images' intensity, see Remark 2. Although our theoretical results in Sect. 3 do not apply in this setting, we obtain satisfying numerical

Table 1 SSD-error and SSIM for various parameters with the images from Fig. 2. For greater clarity, the deformation regularization parameter is chosen as $\lambda = \lambda_1[0.001, 0.1, 10^{-6}]$

$\lambda_1\lambda_2$	0.02	0.2	2	20
0.01	SSD: 0.0073 SSIM: 0.28	SSD: 0.0066 SSIM: 0.30	SSD: 0.0061 SSIM: 0.31	SSD:0.0054 SSIM: 0.34
0.1	SSD: 0.0019 SSIM: 0.44	SSD: 0.0007 SSIM: 0.54	SSD: 0.0011 SSIM: 0.51	SSD: 0.0013 SSIM: 0.50
1	SSD: 0.0027 SSIM: 0.42	SSD: 0.0008 SSIM: 0.54	SSD: 0.0012 SSIM: 0.51	SSD: 0.0012 SSIM: 0.51
10	SSD: 0.0042 SSIM: 0.37	SSD: 0.0017 SSIM: 0.49	SSD: 0.0016 SSIM: 0.51	SSD: 0.0020 SSIM: 0.48
100	SSD: 0.0057 SSIM: 0.32	SSD: 0.0028 SSIM: 0.44	SSD: 0.0033 SSIM: 0.45	SSD: 0.0033 SSIM: 0.45

(a) Template image **T**.   (b) Target image **U**.   (c) Reconstruction (third order).

(d) Reconstruction (curvature).   (e) Error (third order):   (f) Error (curvature):
                                        max 0.54, SSD = 0.05.   max 0.44, SSD = 0.05.
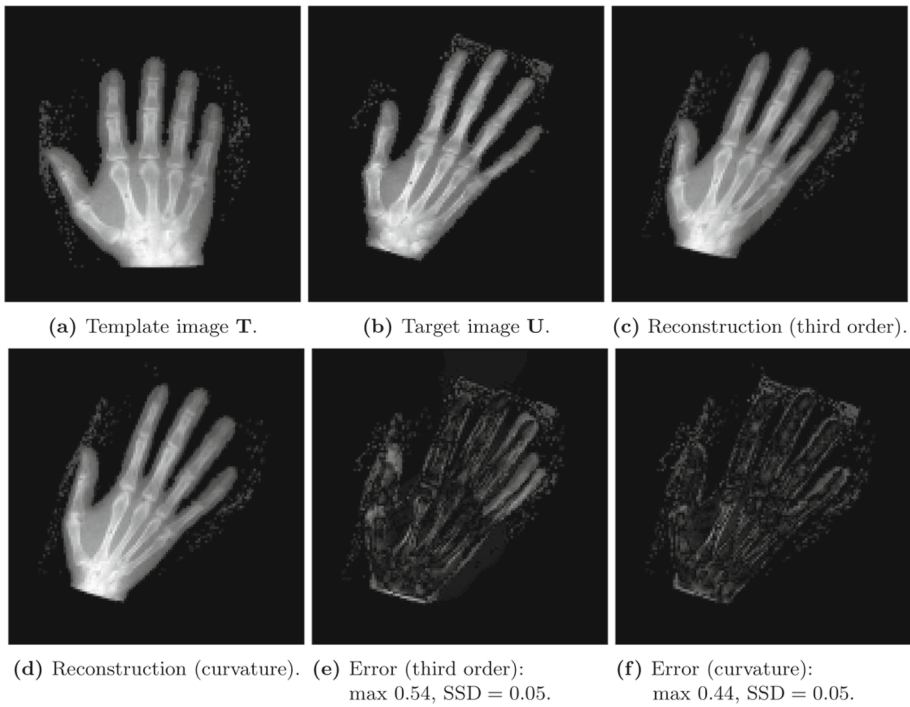
Fig. 5 Reconstruction of a human hand [30] with measurements for 10 angles in [0, 180]. The underlying deformation is relatively large and the images are non-diffeomorphic. Especially the small noise structures outside of the bone areas are hard to reconstruct

(a) Template image **T**.   (b) Target **U**.   (c) Reconstruction of **U**.
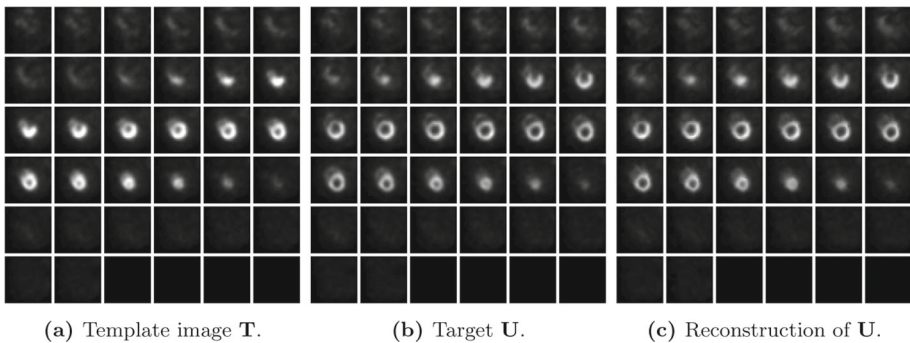
Fig. 6 Reconstruction of a 3D volume using only ten measurement directions. Each image depicts a slice of the volume along the third coordinate axis

results using $\lambda_1 = [0.01, 0.01, 10^{-6}]$ and $\lambda_2 = 0.001$, see Fig. 7. More precisely, our method manages to find the main deformation and the additional hole. In this real data setting, we also investigate how a reconstruction with $E_2 = \| \cdot \|^2$ instead of $E_2 = $ TV performs. Here, our theoretical results from Sect. 3 hold again. To this end, we included two reconstructions corresponding to $\lambda_2 = 0.1$ and $\lambda_2 = 0.01$, respectively. In the first reconstruction, we have almost no artifcats, but can also only guess the new hole. For the second one, the hole can be clearly seen, but the artifcats are much stronger. Either way, the results are inferior to those generated with $E_2 = $ TV.
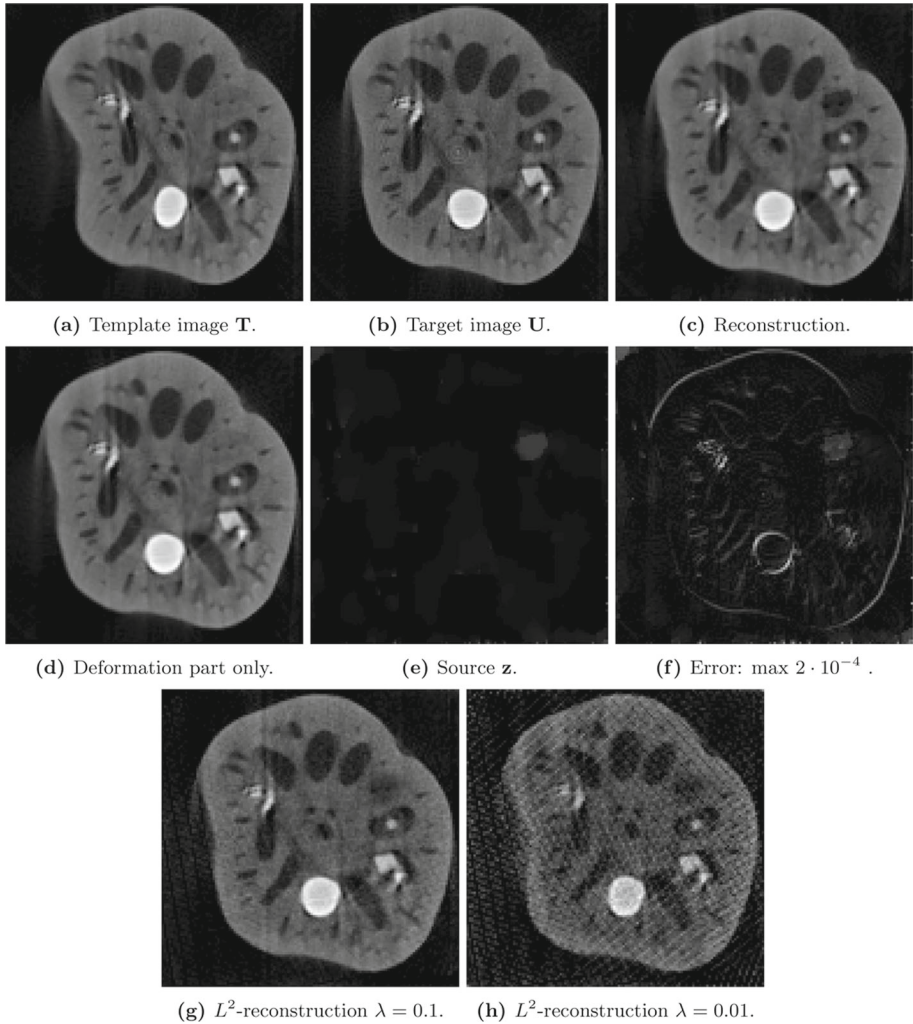
(a) Template image **T**.   (b) Target image **U**.   (c) Reconstruction.

(d) Deformation part only.   (e) Source **z**.   (f) Error: max $2 \cdot 10^{-4}$.

(g) L^2-reconstruction $\lambda = 0.1$.   (h) L^2-reconstruction $\lambda = 0.01$.

Fig. 7 Reconstructions for data obtained by a CT scanner, namely 12 measurements with angles equally distributed in [0, 180]

6 Conclusions

In this paper, we extended the reconstruction model from [23] to also cope with topology changes. On the theoretical side, we were able to carry over all of the previous results. Compared to [18], we utilize a simplified metamorphosis approach, which allows one to use non-convex regularizers at lower computational cost. The chosen TV regularization enabled us to obtain satisfying reconstructions even for very limited data without suffering from the typical artifacts. So far, our experiments are a proof-of-concept. In the future, we also want to work with larger real data. To this end, it could be necessary to use more sophisticated (maybe even problem-tailored) regularization methods for **z** or different data terms $\mathcal{D}$, which can be incorporated into our model without much effort. Again, we stress that the method can be easily extended to higher dimensions and other forward operators **K**. Even without

the scope of real data, this seems to be a natural direction of future research as our method is designed in a modular way.

Data Availability Click here to download links to supporting data https://drive.google.com/file/d/1hwoHSDCu TsVOd-0vG5bEGmRwev1TdatV/view

References

1. Andreev, R., Scherzer, O., Zulehner, W.: Simultaneous optical flow and source estimation: space–time discretization and preconditioning. Appl. Numer. Math. **96**, 72–81 (2015)
2. Ashburner, J.: A fast diffeomorphic image registration algorithm. Neuroimage **38**(1), 95–113 (2007)
3. Becker, S., Fadili, J., Ochs, P.: On quasi-Newton forward-backward splitting: proximal calculus and convergence. SIAM J. Optim. **29**(4), 2445–2481 (2019)
4. Bolte, J., Sabach, S., Teboulle, M.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. Math. Program. **146**(1–2), 459–494 (2014)
5. Bonettini, S., Loris, I., Porta, F., Prato, M., Rebegoldi, S.: On the convergence of a linesearch based proximal-gradient method for nonconvex optimization. Inv. Prob. **33**(5), 055005 (2017)
6. Borzì, A., Ito, K., Kunisch, K.: Optimal control formulation for determining optical flow. SIAM J. Sci. Comput. **24**(3), 818–847 (2003)
7. Bredies, K., Lorenz, D.: Mathematical Image Processing. Applied and Numerical Harmonic Analysis. Birkhäuser, Cham (2018)
8. Bubba, T.A., Hauptmann, A., Huotari, S.J., Rimpeläinen, S. Siltanen. Tomographic X-ray data of a lotus root filled with attenuating objects. arXiv:1609.07299, (2016)
9. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imag. Vis. **40**(1), 120–145 (2011)
10. Chen, C., Öktem, O.: Indirect image registration with large diffeomorphic deformations. SIAM J. Imag. Sci. **11**(1), 575–617 (2018)
11. Chouzenoux, E., Pesquet, J.-C., Repetti, A.: A block coordinate variable metric forward-backward algorithm. J. Global Optim. **66**(3), 457–485 (2016)
12. Christensen, G.E., Rabbitt, R.D., Miller, M.I.: Deformable templates using large deformation kinematics. IEEE Trans. Image Process. **5**(10), 1435–1447 (1996)
13. Dupuis, P., Grenander, U., Miller, M.I.: Variational problems on flows of diffeomorphisms for image matching. Q. Appl. Math. **56**(3), 587–600 (1998)
14. Effland, A.: Discrete Riemannian calculus and a posteriori error control on shape spaces. PhD thesis, University of Bonn, 2017
15. Evans, L.: Partial differential equations. Graduate Studies in Mathematics, vol. 19. Amer. Math. Soc, Providence (1998)
16. Frankel, P., Garrigos, G., Peypouquet, J.: Splitting methods with variable metric for Kurdyka-łojasiewicz functions and general convergence rates. J. Optim. Theory Appl. **165**(3), 874–900 (2015)
17. Frikel, J.: Sparse regularization in limited angle tomography. Appl. Comput. Harmon. Anal. **34**(1), 117–141 (2013)
18. Gris, B., Chen, C., Öktem, O.: Image reconstruction through metamorphosis. Inverse Prob. **36**(2), 025001 (2020)

19. Guerquin-Kern, M., Lejeune, L., Pruessmann, K.P., Unser, M.: Realistic analytical phantoms for parallel magnetic resonance imaging. IEEE Trans. Med. Imag. **31**(3), 626–636 (2012)
20. Hernandez, M.: Primal-dual convex optimization in large deformation diffeomorphic metric mapping: LDDMM meets robust regularizers. Phys. Med. Biol. **62**(23), 9067–9098 (2017)
21. Hinkle, J., Szegedi, M., Wang, B., Salter, B., Joshi, S.: 4D CT image reconstruction with diffeomorphic motion model. Med. Image Anal. **16**(6), 1307–1316 (2012)
22. Horn, B.K., Schunck, B.G.: Determining optical flow. Artif. Intell. **17**(1–3), 185–203 (1981)
23. Lang, L., Neumayer, S., Öktem, O., Schönlieb, C.-B.: Template-based image reconstruction from sparse tomographic data. Appl. Math. Optim. **82**, 1081–1109 (2020)
24. Mang, A., Ruthotto, L.: A Lagrangian Gauss-Newton-Krylov solver for mass- and intensity-preserving diffeomorphic image registration. SIAM J. Sci. Comput. **39**(5), 860–885 (2017)
25. Miller, M.I., Trouvé, A., Younes, L.: On the metrics and Euler-Lagrange equations of computational anatomy. Annu. Rev. Biomed. Eng. **4**(1), 375–405 (2002)
26. Miller, M.I., Trouvé, A., Younes, L.: Geodesic shooting for computational anatomy. J. Math. Imaging Vis. **24**(2), 209–228 (2006)
27. Miller, M.I., Trouvé, A., Younes, L.: Hamiltonian systems and optimal control in computational anatomy: 100 years since D'Arcy Thompson. Annu. Rev. Biomed. Eng. **17**, 447–509 (2015)
28. Miller, M.I., Younes, L.: Group actions, homeomorphisms, and matching: a general framework. Int. J. Comput. Vis. **41**(1–2), 61–84 (2001)
29. Modersitzki, J.: Numerical Methods for Image Registration. Oxford University Press, Oxford (2004)
30. Modersitzki, J.: FAIR: Flexible Algorithms for Image Registration. SIAM, Philadelphia (2009)
31. Natterer, F.: The Mathematics of Computerized Tomography. Classics in Applied Mathematics, vol. 32. SIAM, Philadelphia (2001)
32. Neumayer, S., Persch, J., Steidl, G.: Morphing of manifold-valued images inspired by discrete geodesics in image spaces. SIAM J. Imag. Sci. **11**(3), 1898–1930 (2018)
33. Neumayer, S., Persch, J., Steidl, G.: Regularization of inverse problems via time discrete geodesics in image spaces. Inverse Prob. **35**(5), 055005 (2019)
34. M. Niethammer, G. L. Hart, D. F. Pace, P. M. Vespa, A. Irimia, J. D. Van Horn, and S. R. Aylward. Geometric metamorphosis. In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011, pages 639–646, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg
35. Öktem, O., Chen, C., Domaniç, N.O., Ravikumar, P., Bajaj, C.: Shape-based image reconstruction using linearized deformations. Inverse Prob. **33**(3), 035004 (2017)
36. Palenstijn, W., Batenburg, K., Sijbers, J.: Performance improvements for iterative electron tomography reconstruction using graphics processing units (GPUs). J. Struct. Biol. **176**(2), 250–253 (2011)
37. Pock, T., Sabach, S.: Inertial proximal alternating linearized minimization (iPALM) for nonconvex and nonsmooth problems. SIAM J. Imag. Sci. **9**(4), 1756–1787 (2016)
38. Richardson, C.L., Younes, L.: Metamorphosis of images in reproducing kernel Hilbert spaces. Adv. Comput. Math. **42**(3), 573–603 (2016)
39. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**(1–4), 259–268 (1992)
40. N. Singh, J. Hinkle, S. Joshi, and P. T. Fletcher. A vector momenta formulation of diffeomorphisms for improved geodesic regression and atlas construction. In Proc. of the 10th International Symposium on Biomedical Imaging, pages 127–142. IEEE, New York, 2013
41. Stella, L., Themelis, A., Patrinos, P.: Forward-backward quasi-Newton methods for nonsmooth optimization problems. Comput. Optim. Appl. **67**(3), 443–487 (2017)
42. A. Trouvé. An infinite dimensional group approach for physics based models in pattern recognition. Technical report, 1995
43. Trouvé, A., Younes, L.: Local geometry of deformable templates. SIAM J. Math. Anal. **37**(2), 17–59 (2005)
44. Trouvé, A., Younes, L.: Metamorphoses through Lie group action. Found. Comput. Math. **5**(2), 173–198 (2005)
45. van Aarle, W., Palenstijn, W.J., Cant, J., Janssens, E., Bleichrodt, F., Dabravolski, A., De Beenhouwer, J., Batenburg, K.J., Sijbers, J.: Fast and flexible X-ray tomography using the ASTRA toolbox. Opt. Express **24**(22), 25129–25147 (2016)
46. van Aarle, W., Palenstijn, W.J., De Beenhouwer, J., Altantzis, T., Bals, S., Batenburg, K.J., Sijbers, J.: The ASTRA toolbox: A platform for advanced algorithm development in electron tomography. Ultramicroscopy **157**, 35–47 (2015)
47. Vialard, F.-X., Risser, L., Rueckert, D., Cotter, C.J.: Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation. Int. J. Comput. Vis. **97**(2), 229–241 (2011)

48. Yang, X., Kwitt, R., Styner, M., Niethammer, M.: Quicksilver: Fast predictive image registration-a deep learning approach. Neuroimage **158**, 378–396 (2017)
49. Younes, L.: Shapes and Diffeomorphisms. Springer, Berlin (2010)