



Combining the Band-Limited Parameterization and Semi-Lagrangian Runge–Kutta Integration for Efficient PDE-Constrained LDDMM

Monica Hernandez¹

Received: 30 March 2020 / Accepted: 6 January 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

Abstract

The family of PDE-constrained Large Deformation Diffeomorphic Metric Mapping (LDDMM) methods is emerging as a particularly interesting approach for physically meaningful diffeomorphic transformations. The original combination of Gauss–Newton–Krylov optimization and Runge–Kutta integration shows excellent numerical accuracy and fast convergence rate. However, its most significant limitation is the huge computational complexity, hindering its extensive use in Computational Anatomy applied studies. This limitation has been treated independently by the problem formulation in the space of band-limited vector fields and semi-Lagrangian integration. The purpose of this work is to combine both in three variants of band-limited PDE-constrained LDDMM for further increasing their computational efficiency. The accuracy of the resulting methods is evaluated extensively. For all the variants, the proposed combined approach shows a significant increment of the computational efficiency. In addition, the variant based on the deformation state equation is positioned consistently as the best performing method across all the evaluation frameworks in terms of accuracy and efficiency.

Keywords Physically meaningful diffeomorphic registration · PDE-constrained LDDMM · Gauss–Newton–Krylov · Optimal control optimization · Band-limited vector fields · Semi-Lagrangian Runge–Kutta integration

Abbreviations

PDE	Partial differential equation
LDDMM	Large deformation diffeomorphic metric mapping
SP	Spatial
BL	Band limited
RK	Runge–Kutta
SL	Semi-Lagrangian
PCG	Preconditioned conjugate gradient
DSC	Dice similarity coefficient
SSD	Sum of squared differences
CPU	Central processing unit
GPU	Graphics processing unit
VRAM	Video random access memory

1 Introduction

Computational Anatomy is a powerful interdisciplinary field for the analysis of anatomical shape variability [19,20]. This discipline is based on Sir D’Arcy Thompson’s original ideas for explaining the similarity of the anatomical shape of homologous species using the transformations existing between the anatomical structures [31]. In Computational Anatomy, shape similarity is measured from the diffeomorphic transformations estimated between the anatomies. These transformations yield a generative model for the analysis of shape variability. Diffeomorphisms are computed from the anatomical images using diffeomorphic registration methods [34].

There exists a vast literature on diffeomorphic registration methods with differences in the transformation characterization, regularizers, image similarity metrics, optimization methods, and additional constraints [29]. Although the differentiability and invertibility of the transformations constitute crucial features for Computational Anatomy applications, the diffeomorphic constraint does not necessarily guarantee that a transformation computed with a given method is physically meaningful for the clinical domain of interest.

This work was partially supported by the National Research Grants PID2019-104358RB-I00 (DL-Aging Project) and TIN2016-80347-R (DIAMOND Project).

✉ Monica Hernandez
mhg@unizar.es

¹ University of Zaragoza, Zaragoza, Spain

PDE-constrained Large Deformation Diffeomorphic Metric Mapping (PDE-LDDMM) has become relevant in the last decade for the computation of transformations under plausible physical models of interest [1,7,10,13,14,18,32,33,36].

Our work focuses on the family of PDE-LDDMM methods pioneered by Hart et al. [7] and leading to the relevant contributions in [9,13,14,17,22]. In this family of methods, the registration problem is approached from an optimal control perspective, where different physical models are imposed directly using the physical PDEs that are attached to the LDDMM variational problem using hard constraints. The numerical optimization is approached using gradient descent [7,13,22] or second-order optimization in the form of inexact reduced Newton–Krylov methods [9,13,14,17]. In particular, the combination of Gauss–Newton–Krylov for optimization, with sophisticated multi-level preconditioners, spectral methods for differentiation, and Runge–Kutta schemes for PDE integration, shows excellent numerical accuracy and an extraordinarily fast convergence rate. However, the most significant limitation of Gauss–Newton–Krylov PDE-LDDMM is the huge computational complexity, which hinders the extensive use in Computational Anatomy applied studies. This computational complexity is due to:

1. The formulation of the problem in the spatial domain.
2. The large time sampling needed for the stability of Runge–Kutta integration.

Both issues have been treated independently in the literature yielding to PDE-LDDMM methods with increased efficiency and an assumable cost in accuracy loss.

1.1 Computational Complexity due to Problem Formulation

The computational complexity due to the formulation of the problem in the spatial domain has been successfully reduced using the band-limited vector field parameterization proposed in [35,36]. LDDMM methods, and in particular PDE-LDDMM, involve the action of low-pass filters in the optimization update equations of the velocities. Therefore, the computation of the high-frequency components of high-resolution velocity fields can be omitted since these computations result equal or nearly equal to zero by the action of the low-pass filters. The band-limited vector field parameterization allows a reduction in the dimensionality of the problem that circumvents the high-frequency computations.

The works in [8,9] formulate three different variants of PDE-LDDMM in the space of band-limited vector fields and perform the computations in the GPU. Some configurations of these variants have been really successful, greatly

outperforming the state-of-the-art methods in terms of computational complexity while keeping a competitive accuracy.

1.2 Computational Complexity due to PDE Integration

Runge–Kutta methods are explicit techniques. Hence, they are only conditionally stable. This means that the time sampling should be selected enough to preserve the Courant–Friedrichs–Lewy (CFL) condition. For PDE-LDDMM, the time sampling values that guarantee stability are usually large. As a result, the time and memory requirements of the problem are considerably increased. In particular, the memory requirements of PDE-LDDMM are increased to limits that hinder the execution on limited memory devices such as the GPU. In addition, one can experience that the time sampling needed for the non-stationary parameterization is much higher than for the stationary parameterization, increasing the complexity of an already not particularly memory efficient configuration. On the other side, when stability is satisfied, the accuracy of PDE-LDDMM is high [8,9,13,14].

Semi-Lagrangian methods are semi-implicit techniques that are unconditionally stable. Therefore, the time sampling can be selected according to accuracy rather than stability considerations. Semi-Lagrangian methods were originally proposed in the 90's in the context of modeling weather predictions [30]. In the context of diffeomorphic registration, the original LDDMM method proposed in [3] already used semi-Lagrangian integration for the solution of the transport equation. The combination of semi-Lagrangian integration with Runge–Kutta has been recently proposed for solving some time-dependent PDEs. Runge–Kutta has shown to increase the accuracy of first-order schemes in semi-Lagrangian integration [6].

The computational complexity in [13,14] due to the use of Runge–Kutta schemes for PDE integration has been successfully reduced using semi-Lagrangian Runge–Kutta integration [15,17] for the stationary parameterization of diffeomorphisms. For PDE-LDDMM, the selected time sampling is usually much smaller than the time sampling typically selected with explicit schemes, yielding to a considerable reduction in the computational complexity of the problem. On the other hand, the expected accuracy of PDE-LDDMM is lower than with explicit schemes.

Beyond the computational complexity improvement through numerical schemes, Mang et al. proposed an efficient implementation of PDE-LDDMM that exploits massive CPU based parallel computing architectures [16]. The source code has been recently released with [17]. A GPU optimized implementation of the method is being proposed in the ArXiv paper [4].

1.3 Our Contribution

The purpose of this work is to further increase the computational efficiency of BL PDE-LDDMM by combining the two independent methodological approaches of circumventing the huge computational complexity of PDE-LDDMM and to extensively analyze the accuracy of the resulting methods. We have implemented the band-limited methods in [8,9] with the semi-Lagrangian Runge–Kutta integration scheme originally proposed in [15] for the stationary and the non-stationary parameterization of diffeomorphisms. The resulting methods have been evaluated in five different datasets following the evaluation frameworks in [9,12,25]. To our knowledge, this is the first time that semi-Lagrangian Runge–Kutta integration is implemented in the space of band-limited vector fields. It is also the first time that semi-Lagrangian Runge–Kutta integration is used in PDE-LDDMM with the non-stationary parameterization. Moreover, our work first provides the position achieved by benchmark PDE-LDDMM methods in the ranking of Klein et al. evaluation. The best performing method of our work coincides with the best performing variant in [9], PDE-LDDMM based on the deformation state equation. The semi-Lagrangian Runge–Kutta scheme proposed in this work has shown to outperform the Runge–Kutta scheme in [9] in terms of computational efficiency and accuracy. Indeed, the best performing PDE-LDDMM variant in this work has recently reached the highest sensitivity (97% vs a baseline of 88%) in the classification of stable versus progressive mild cognitive impaired converters in the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database using convolutional neural networks [23].

1.4 Manuscript Organization

In the following, Sect. 2 reviews the foundations of PDE-LDDMM, with particular emphasis on the band-limited vector field parameterization. Section 3 presents the proposed semi-Lagrangian Runge–Kutta integration method. Next, Sect. 4 details the experimental setup. Section 5 shows the results and Sect. 6 discusses the most important highlights. Finally, Sect. 7 gathers the most remarkable conclusions of our work.

2 PDE-Constrained LDDMM Methods

2.1 Parameterization in the Spatial Domain

Let $\Omega \subseteq \mathbb{R}^d$ be the image domain. Let $\text{Diff}(\Omega)$ be the LDDMM Riemannian manifold of diffeomorphisms and V the tangent space at the identity element. $\text{Diff}(\Omega)$ is a Lie group, and V is the corresponding Lie algebra [3]. The Riemannian metric of $\text{Diff}(\Omega)$ is defined from the scalar product

in V

$$\langle v, w \rangle_V = \langle Lv, w \rangle_{L^2} = \int_{\Omega} \langle Lv(x), w(x) \rangle dx, \tag{1}$$

where L is the invertible self-adjoint differential operator associated with the differential structure of $\text{Diff}(\Omega)$. In traditional LDDMM methods, $L = (Id - \alpha \Delta)^s, \alpha > 0, s \in \mathbb{N}$ [3]. This is the operator used in this work.

Let I_0 and I_1 be the source and the target images. LDDMM is formulated from the minimization of the variational problem

$$E(v) = \frac{1}{2} \int_0^1 \langle Lv_t, v_t \rangle_{L^2} dt + \frac{1}{\sigma^2} \|I_0 \circ (\phi_1^v)^{-1} - I_1\|_{L^2}^2. \tag{2}$$

The LDDMM variational problem [3] is posed in the space of time-varying smooth flows of velocity fields, $v \in L^2([0, 1], V)$. Given the smooth flow $v : [0, 1] \rightarrow V, v_t : \Omega \rightarrow \mathbb{R}^d$, the solution at time $t = 1$ to the evolution equation

$$\partial_t (\phi_t^v)^{-1} = -v_t \circ (\phi_t^v)^{-1} \tag{3}$$

with initial condition $(\phi_0^v)^{-1} = id$ is a diffeomorphism, $(\phi_1^v)^{-1} \in \text{Diff}(\Omega)$. The transformation $(\phi_1^v)^{-1}$, computed from the minimum of $E(v)$, is the diffeomorphism that solves the LDDMM registration problem between I_0 and I_1 . The problem can be straightforwardly restricted to the space of steady flows of velocity fields [11].

LDDMM can be formulated from a dynamical systems point of view. Thus, PDE-LDDMM arises from LDDMM as an optimal control approach to diffeomorphic registration [7, 13]. PDE-LDDMM is formulated from the minimization of the PDE-constrained variational problem

$$E(v) = \frac{1}{2} \int_0^1 \langle Lv_t, v_t \rangle_{L^2} dt + \frac{1}{\sigma^2} \|m(1) - I_1\|_{L^2}^2, \tag{4}$$

subject to the state equation with state variable $m(t)$

$$\partial_t m(t) + \nabla m(t) \cdot v_t = 0 \text{ in } \Omega \times (0, 1], \tag{5}$$

with initial condition $m(0) = I_0$.

Using optimal control terminology, the flow v is the control of the dynamical system. The system dynamics are driven by the state equation which determines the evolution of the state variable m of the dynamical system. The minimization of Eq. 4 aims at finding the optimal control subject to the system dynamics on the initial state I_0 .

The state equation constraint in Eq. 5 can be imposed in two more different manners, yielding three different variants of PDE-LDDMM [9]. Variant I corresponds with the variational formulation presented above and proposed in [7,13].

The second variant (Variant II) is formulated from the minimization of Eq. 4, where

$$m(t) = I_0 \circ \phi(t) \quad (6)$$

and ϕ is computed from the deformation state equation

$$\partial_t \phi(t) + D\phi(t) \cdot v_t = 0 \text{ in } \Omega \times (0, 1], \quad (7)$$

with initial condition $\phi(0) = id$. The third variant (Variant III) is formulated from the minimization of Eq. 4 subject to the deformation state equation, Eq. 7. It should be noticed that $\phi(t)$ was used in Hart et al. [7] for referring to Beg et al. [3] diffeomorphism path $\phi_{t,0}$, which corresponds to $(\phi_t^v)^{-1}$ in Eqs. 2 and 3.

The advantage of the optimal control approach is that the complex dependence between $I_0 \circ (\phi_1^v)^{-1}$ and v is removed from $E(v)$ and translated to the system dynamics. Thus, the original LDDMM variational formulation is transformed into a PDE-constrained formulation. The optimization is approached using the adjoint method, where the computation of the optimality conditions of the system is performed using the method of Lagrange multipliers, yielding a set of state and adjoint differential equations. The solutions to these equations arise in the expressions of the gradient and the Hessian of the augmented energy used for the update of the control variable. Indeed, the optimal control approach allows imposing different system dynamics in the set of state equations providing a straightforward approach to obtain different families of physically meaningful diffeomorphisms [13,14].

The best optimization method from among the algorithms tested for PDE-LDDMM is Gauss–Newton–Krylov [9,13]. The expressions of the gradient and the Hessian vector product are derived from the augmented Lagrangian of the energy functional subject to the state or the deformation state equations, respectively. The expressions of the augmented Lagrangian, the gradient $\nabla_v E_{\text{aug}}(v)$, and the Hessian vector product $H_v E_{\text{aug}}(v)\delta v$ for each variant are found in appendix.

The update equation has the form

$$v^{n+1} = v^n + \epsilon \delta v^n, \quad (8)$$

where ϵ is the update length and δv^n is computed from preconditioned conjugate gradient (PCG) on the system

$$H_v E_{\text{aug}}(v^n)\delta v^n = -\nabla_v E_{\text{aug}}(v^n), \quad (9)$$

with preconditioner L^{-1} .

2.2 Parameterization in the Space of Band-Limited Vector Fields

Let $\tilde{\Omega}$ be the discrete Fourier domain truncated with frequency bounds K_1, \dots, K_d . We denote with \tilde{V} the space of discretized band-limited vector fields on Ω with these frequency bounds. The elements in \tilde{V} are represented in the Fourier domain as $\tilde{v} : \tilde{\Omega} \rightarrow \mathbb{C}^d$, $\tilde{v}(k_1, \dots, k_d)$. The application $\iota : \tilde{V} \rightarrow V$ denotes the natural inclusion mapping of \tilde{V} in V . The application $\pi : V \rightarrow \tilde{V}$ denotes the projection of V onto \tilde{V} [35,36].

The space \tilde{V} of band-limited vector fields has a finite-dimensional Lie algebra structure using the truncated convolution \star in the definition of the Lie bracket [36]. We denote with $\text{Diff}(\tilde{\Omega})$ to the finite-dimensional Riemannian manifold of diffeomorphisms on $\tilde{\Omega}$ with corresponding Lie algebra \tilde{V} . The Riemannian metric in $\text{Diff}(\tilde{\Omega})$ is defined from the scalar product

$$\langle \tilde{v}, \tilde{w} \rangle_{\tilde{V}} = \langle \tilde{L}\tilde{v}, \tilde{w} \rangle_{L^2}, \quad (10)$$

where \tilde{L} is the projection of operator L in the truncated Fourier domain. Similarly, we will denote with $\tilde{\ast}$ to the projection in the truncated Fourier domain of the differential operators \ast involved in the differential equations.

The band-limited PDE-constrained variational problem is given by the minimization of

$$E(\tilde{v}) = \frac{1}{2} \int_0^1 \langle \tilde{L}\tilde{v}_t, \tilde{v}_t \rangle_{L^2} dt + \frac{1}{\sigma^2} \|m(1) - I_1\|_{L^2}^2. \quad (11)$$

The band-limited version of Variant I is formulated from the minimization of Eq. 11 subject to

$$\partial_t m(t) + \nabla m(t) \cdot \iota(\tilde{v}_t) = 0 \text{ in } \Omega \times (0, 1], \quad (12)$$

with initial condition $m(0) = I_0$. For Variant II, the diffeomorphism is computed from $\phi(t) = id - \iota(\tilde{u})(t)$ where $\tilde{u}(t)$ is computed from the deformation state equation formulated in displacement field form

$$\partial_t \tilde{u}(t) + \tilde{D}\tilde{u}(t)\tilde{\ast}\tilde{v}_t = \tilde{v}(t) \text{ in } \tilde{\Omega} \times (0, 1]. \quad (13)$$

Variant III is formulated analogously to the spatial case from the minimization of Eq. 11 subject to the deformation state equation, Eq. 13 [8,9].

The optimization is approached using Gauss–Newton–Krylov methods in \tilde{V} with preconditioner \tilde{L}^{-1} . The update equation has the form

$$\tilde{v}^{n+1} = \tilde{v}^n + \epsilon \delta \tilde{v}^n, \quad (14)$$

where $\delta\tilde{v}^n$ is computed from

$$(H_{\tilde{v}}\widetilde{E_{\text{aug}}}(\tilde{v}^n))\delta\tilde{v}^n = -(\nabla_{\tilde{v}}\widetilde{E_{\text{aug}}}(\tilde{v}^n)). \tag{15}$$

In the next section, we provide the expressions of the gradient and the Hessian for each variant.

2.3 BL PDE-Constrained LDDMM Equations

2.3.1 Original BL PDE-Constrained LDDMM (Variant I)

Originally proposed BL PDE-LDDMM uses the state equation in the augmented Lagrangian for the derivation of the state and adjoint equations and their incremental counterparts [8]

$$E_{\text{aug}}(\tilde{v}) = E(\tilde{v}) + \int_0^1 \langle \lambda(t), \partial_t m(t) + Dm(t) \cdot \iota(\tilde{v}_t) \rangle_{L^2} dt + \langle \eta, m(0) - I_0 \rangle_{L^2}, \tag{16}$$

where λ and η are the Lagrangian multipliers associated with the state equation (Eq. 12) and its initial condition.

The gradient and the Gauss–Newton approximation of the Hessian vector product are computed from the first- and second-order optimality conditions, derived from vanishing the formal computations of δE_{aug} and $\delta^2 E_{\text{aug}}$

$$(\nabla_{\tilde{v}}\widetilde{E_{\text{aug}}}(\tilde{v}))_t = \tilde{L}\tilde{v}_t + \tilde{\lambda}(t)\star\tilde{\nabla}\tilde{m}(t) \tag{17}$$

$$(H_{\tilde{v}}\widetilde{E_{\text{aug}}}(\tilde{v}))_t \delta\tilde{v}(t) = \tilde{L}\delta\tilde{v}(t) + \delta\tilde{\lambda}(t)\star\tilde{\nabla}\tilde{m}(t), \tag{18}$$

where the projected state variable \tilde{m} and the projected adjoint variable $\tilde{\lambda}$ are computed from $\pi(m)$ and $\pi(\lambda)$, and m and λ are computed from

$$\partial_t m(t) + \nabla m(t) \cdot \iota(\tilde{v}_t) = 0 \tag{19}$$

$$-\partial_t \lambda(t) - \nabla \cdot (\lambda(t) \cdot \iota(\tilde{v}_t)) = 0. \tag{20}$$

The incremental counterparts $\delta\tilde{m}$ and $\delta\tilde{\lambda}$ are the solutions of

$$\partial_t \delta\tilde{m}(t) + \tilde{\nabla}\delta\tilde{m}(t)\star\tilde{v}_t + \tilde{\nabla}\tilde{m}(t)\star\delta\tilde{v}(t) = 0 \tag{21}$$

$$-\partial_t \delta\tilde{\lambda}(t) - \tilde{\nabla} \cdot (\delta\tilde{\lambda}(t)\star\tilde{v}_t) = 0 \tag{22}$$

in the BL domain. The initial conditions are, respectively, $m(0) = I_0$, $\lambda(1) = -\frac{2}{\sigma^2}(m(1) - I_1)$, $\delta\tilde{m}(0) = 0$, $\delta\tilde{\lambda}(1) = -\frac{2}{\sigma^2}\delta\tilde{m}(1)$. Algorithm 1 shows the pseudocode for Variant I.

2.3.2 BL PDE-Constrained LDDMM Based on the State Equation (Variant II)

This method departs from the original BL PDE-LDDMM by using $m(t) = I_0 \circ \phi(t)$ and $\lambda(t) = J(t)\lambda(1) \circ \psi(t)$,

where ϕ is the direct map, ψ is the inverse map, and J is the Jacobian determinant of ψ [7,9]. The transformations ϕ and ψ and the scalar field J are computed from the inclusion of the truncated displacement fields ($\tilde{u}(t)$ and $\tilde{\tau}(t)$) and the corresponding Jacobian

$$\phi(t) = id - \iota(\tilde{u}(t)), \tag{23}$$

$$\psi(t) = id - \iota(\tilde{\tau}(t)), \tag{24}$$

$$J(t) = 1 - \iota(\tilde{U}(t)), \tag{25}$$

where

$$\partial_t \tilde{u}(t) + \tilde{D}\tilde{u}(t)\star\tilde{v}_t = \tilde{v}_t \tag{26}$$

$$-\partial_t \tilde{\tau}(t) - \tilde{D}\tilde{\tau}(t)\star\tilde{v}_t = -\tilde{\tau}_t \tag{27}$$

$$-\partial_t \tilde{U}(t) - \tilde{v}_t\star\tilde{\nabla}\tilde{U}(t) = -\tilde{\nabla}\cdot\tilde{v} + \tilde{U}(t)\star\tilde{\nabla}\cdot\tilde{v}(t) \tag{28}$$

with initial conditions $\tilde{u}(0) = 0$, $\tilde{\tau}(1) = 0$, and $\tilde{U}(1) = 0$.

The incremental state and adjoint variables are computed from the differential of the expressions given for $m(t)$ and $\lambda(t)$

$$\delta\tilde{m}(t) = \nabla I_0 \circ \phi(t) \cdot \iota(\delta\tilde{u}(t)) \tag{29}$$

$$\delta\tilde{\lambda}(t) = J(t)\delta\tilde{\lambda}(1) \circ \psi(t) \cdot \iota(\delta\tilde{\tau}(t)) \tag{30}$$

where the precedence of ∇ , δ , and \circ operators reads $\nabla I_0 \circ \phi(t) = (\nabla I_0) \circ \phi(t)$ and $\delta\tilde{\lambda}(1) \circ \psi(t) = (\delta\tilde{\lambda}(1)) \circ \psi(t)$. The incremental expressions $\delta\tilde{u}$ and $\delta\tilde{\tau}$ are computed from the differentiation of Eqs. 26 and 27, yielding

$$\partial_t \delta\tilde{u}(t) + \tilde{D}\delta\tilde{u}(t)\star\tilde{v}_t + \tilde{D}\tilde{u}(t)\star\delta\tilde{v}_t = \delta\tilde{v}_t \tag{31}$$

$$-\partial_t \delta\tilde{\tau}(t) - \tilde{D}\delta\tilde{\tau}(t)\star\tilde{v}_t - \tilde{D}\tilde{\tau}(t)\star\delta\tilde{v}_t = -\delta\tilde{\tau}_t \tag{32}$$

with initial conditions $\delta\tilde{u}(0) = 0$ and $\delta\tilde{\tau}(1) = 0$. Algorithm 1 also shows the pseudocode for Variant II, which shares the steps with Variant I.

2.3.3 BL PDE-Constrained LDDMM Based on the Deformation State Equation (Variant III)

The third method is formulated from the minimization of Eq. 11 subject to the truncated displacement state equation (Eq. 13) [9]

$$\partial_t \tilde{u}(t) + \tilde{D}\tilde{u}(t)\star\tilde{v}_t = \tilde{v}_t. \tag{33}$$

In this variant, the augmented Lagrangian is given by

$$E_{\text{aug}}(\tilde{v}) = E(\tilde{v}) + \int_0^1 \langle \tilde{\rho}(t), \partial_t \tilde{u}(t) + \tilde{D}\tilde{u}(t)\star\tilde{v}_t - \tilde{v}_t \rangle_{L^2} dt + \langle \tilde{\mu}, \tilde{u}(0) \rangle_{L^2}, \tag{34}$$

Algorithm 1. Band-limited PDE-LDDMM. Variants I and II.

Data: $I_0, I_1, \tilde{v}_t^0 \in \tilde{V}, \tilde{L}, \sigma, \epsilon, PCG_its$.
Results: $\tilde{v}_t \in \tilde{V}, m(1)$ solutions of Equation 11 subject to 12.
for $n \leftarrow 0$ **to** convergence **do**
 1) Compute $\tilde{m}^{n+1}(t)$ forward in time from Equation 19 and $m(0) = I_0$ or $m(t) = I_0 \circ \phi(t)$.
 2) Compute $\lambda^{n+1}(t)$ backward in time from Equation 20 or $\lambda(t) = J(t)\lambda(1) \circ \psi(t)$, where $\lambda(1) = -\frac{2}{\sigma^2}(m(1) - I_1)$.
 3) Compute $(\nabla_{\tilde{v}} \widetilde{E_{aug}}(\tilde{v}^{n+1}))$ from Equation 17.
 4) Compute $\delta\tilde{v}^{n+1}$ from Equation 15.
for $k \leftarrow 0$ **to** PCG convergence or PCG_its **do**
 4.1) Compute $\delta\tilde{m}^{k+1}(t)$ from Equation 21 or Equation 29.
 $\delta\tilde{m}(1)$ is needed in the computation of $\delta\tilde{\lambda}(1)$ used as initial condition in step 4.2.
 4.2) Compute $\delta\tilde{\lambda}^{k+1}(t)$ from Equation 22 or Equation 30.
 4.3) Compute $(H_{\tilde{v}} \widetilde{E_{aug}}(\tilde{v}^n))\delta\tilde{v}^k$ from Equation 18.
 4.4) Compute $\delta\tilde{v}^{k+1}$ from PCG at iteration k .
 4.5) Check PCG convergence.
end
 5) Compute \tilde{v}^{n+1} from Equation 14.
 6) Check algorithm convergence.
end

Algorithm 2. Band-limited PDE-LDDMM. Variant III.

Data: $I_0, I_1, \tilde{v}_t^0 \in \tilde{V}, \tilde{L}, \sigma, \epsilon, PCG_its$.
Results: $\tilde{v}_t \in \tilde{V}, m(1)$ solutions of Equation 11 subject to 13.
for $n \leftarrow 0$ **to** convergence **do**
 1) Compute $\tilde{u}^{n+1}(t)$ forward in time from Equation 37 and $\tilde{u}(0) = 0$.
 2) Compute $\tilde{\rho}^{n+1}(t)$ backward in time from Equation 38, where $\tilde{\rho}(1) = \pi(-\frac{2}{\sigma^2}(m(1) - I_1))\nabla m(1)$.
 3) Compute $(\nabla_{\tilde{v}} \widetilde{E_{aug}}(\tilde{v}^{n+1}))$ from Equation 35.
 4) Compute $\delta\tilde{v}^{n+1}$ from Equation 15.
for $k \leftarrow 0$ **to** PCG convergence or PCG_its **do**
 4.1) Compute $\delta\tilde{u}^{k+1}(t)$ from Equation 31.
 4.2) Compute $\delta\tilde{\rho}^{k+1}(t)$ from Equation 40.
 4.3) Compute $(H_{\tilde{v}} \widetilde{E_{aug}}(\tilde{v}^n))\delta\tilde{v}^k$ from Equation 36.
 4.4) Compute $\delta\tilde{v}^{k+1}$ from PCG at iteration k .
 4.5) Check PCG convergence.
end
 5) Compute \tilde{v}^{n+1} from Equation 14.
 6) Check algorithm convergence.
end

where $\tilde{\rho}$ and $\tilde{\mu}$ are the Lagrangian multipliers associated with the BL deformation state equation (Eq. 33) and its initial condition.

The gradient and the Hessian vector product, computed from the first- and second-order optimality conditions, are given by the equations

$$(\nabla_{\tilde{v}} \widetilde{E_{aug}}(\tilde{v}))_t = \tilde{L}\tilde{v}_t + \tilde{\rho}(t) - \tilde{D}\tilde{u}(t)\star\tilde{\rho}(t) \tag{35}$$

$$(H_{\tilde{v}} \widetilde{E_{aug}}(\tilde{v}))_t \delta\tilde{v}(t) = \tilde{L}\delta\tilde{v}(t) + \delta\tilde{\rho}(t) - \tilde{D}\delta\tilde{u}(t)\star\tilde{\rho}(t) \tag{36}$$

where the displacement state variable \tilde{u} , the adjoint variable $\tilde{\rho}$, and their incremental counterparts $\delta\tilde{u}$ and $\delta\tilde{\rho}$ are computed

from

$$\partial_t \tilde{u}(t) + \tilde{D}\tilde{u}(t)\star\tilde{v}_t = \tilde{v}_t \tag{37}$$

$$-\partial_t \tilde{\rho}(t) - \tilde{\nabla} \cdot (\tilde{\rho}(t)\star\tilde{v}_t) = 0 \tag{38}$$

$$\partial_t \delta\tilde{u}(t) + \tilde{D}\delta\tilde{u}(t)\star\tilde{v}_t + \tilde{D}\tilde{u}(t)\star\delta\tilde{v}(t) = \delta\tilde{v}(t) \tag{39}$$

$$-\partial_t \delta\tilde{\rho}(t) - \tilde{\nabla} \cdot (\delta\tilde{\rho}(t)\star\tilde{v}_t) = 0 \tag{40}$$

with initial conditions $\tilde{u}(0) = 0, \tilde{\rho}(1) = \pi(-\frac{2}{\sigma^2}(m(1) - I_1))\nabla m(1), \delta\tilde{u}(0) = 0$, and $\delta\tilde{\rho}(1) = \pi(-\frac{2}{\sigma^2}\delta m(1))\nabla m(1)$. Algorithm 2 shows the pseudocode for Variant III.

3 Semi-Lagrangian Runge–Kutta Integration

3.1 Semi-Lagrangian Integration in a Spatial Domain

Semi-Lagrangian (SL) integration methods [30] allow solving transport equations of the general form

$$D_t u = f(u, v), \tag{41}$$

where $u : \Omega^d \times [0, 1] \rightarrow \mathbb{R}$ is a scalar or a vector function varying in time, and

$$D_t u = \partial_t u + Du \cdot v.$$

SL methods combine the most interesting properties of Eulerian and Lagrangian schemes. On the one hand, SL methods involve following the characteristic lines of the differential equation, similarly to Lagrangian approaches. On the other hand, the equation is solved on the regular grid, similarly to Eulerian approaches. As a result, SL methods are unconditionally stable as Lagrangian schemes. This means that the time sampling can be selected according to accuracy considerations rather than stability considerations. SL methods allow selecting a number of time steps usually much smaller than Eulerian methods yielding a sensible reduction in the computational complexity.

SL schemes involve two steps. First, the departure points are computed solving the characteristic equation

$$D_t X(t) = v(X(t), t), \tag{42}$$

with initial condition $X(0) = x$. The direction of the time integration can be forward or backward, depending on the direction of the time integration of the transport equation. From several methods proposed in the literature for solving the characteristic equation, we use the approach given by Mang et al. in [15]

$$X_* = x - \delta t \cdot v \tag{43}$$

$$v_* = v \circ X_* \tag{44}$$

$$X_* = x - 0.5 \delta t \cdot (v_* + v). \tag{45}$$

Second, the transport equation (Eq. 41) is solved in the Eulerian grid

$$D_t u(X(t), t) = f(u(X(t), t), v(X(t), t)) \tag{46}$$

along the characteristic line X . The use of Runge–Kutta (RK) integration has been recently proposed in this step, yielding a higher-order accurate SL-RK method [6]. The velocity field needs to be estimated at points that do not belong to the Eulerian grid. Therefore, an interpolator is needed. Cubic interpolation is the method of choice for SL schemes [24].

Algorithm 3. Semi-Lagrangian Runge-Kutta integration of BL PDE-LDDMM equations.

Data: \tilde{u}_0 or $\tilde{u}_1, \tilde{v}, X$

Results: \tilde{u}_t solution of Equation 50.

- 1) Compute $\tilde{f}(\tilde{u}, \tilde{v})$.
- 2) Project \tilde{f} on $\Omega, f = \pi(\tilde{f})$.
- 3) Compute $f \circ X$ using cubic interpolation.
- 4) Include $f \circ X$ in $\tilde{\Omega}$, providing $\tilde{f}(\tilde{X})$.
- 5) Use \tilde{f} as the right hand side of the Runge-Kutta solver.

end

3.2 Semi-Lagrangian Integration in a Band-Limited Domain

In $\tilde{\Omega}$, the transport equations are of the general form

$$D_t \tilde{u} = \tilde{f}(\tilde{u}, \tilde{v}), \tag{47}$$

where

$$D_t \tilde{u} = \partial_t \tilde{u} + \tilde{D} \tilde{u} \star \tilde{v}. \tag{48}$$

The characteristics are computed from

$$D_t X(t) = \iota(\tilde{v})(X(t), t), \tag{49}$$

and the transport equation is solved from

$$D_t \tilde{u}(\tilde{X}(t), t) = \tilde{f}(\tilde{u}(\tilde{X}(t), t), \tilde{v}(\tilde{X}(t), t)). \tag{50}$$

3.3 Semi-Lagrangian Runge–Kutta Integration in PDE-LDDMM

In this work, SL-RK integration has been implemented in Ω and $\tilde{\Omega}$ for the spatial and band-limited versions of the three

PDE-LDDMM variants. To be able to apply SL integration, the differential equations need to be written in the shape of Eqs. 46 or 50, respectively. We focus on the derivation for the BL domain $\tilde{\Omega}$. The derivation for the spatial domain can be performed analogously and it is provided in appendix.

The state equations, the deformation state equations, and their incremental counterparts (Eqs. 19, 26, 21, 31) are already in the shape of Eq. 50 by just moving to the right-hand side of the equation a remaining term. For the adjoint and the incremental adjoint equations (Eqs. 20, 38, 22, 40), we use the identity

$$\tilde{\nabla} \cdot (\tilde{u} \star \tilde{v}) = \tilde{u} \tilde{\nabla} \cdot \tilde{v} + \tilde{v} \star \tilde{\nabla} \tilde{u} \text{ in } \tilde{\Omega} \tag{51}$$

and move the divergence term to the right-hand side of the transformed equation. Table 1 gathers the expressions of the resulting differential equations, needed for the implementation of BL PDE-LDDMM methods in SL form. For SL-RK, the right-hand side expressions can be directly plugged into an RK differential solver. Algorithm 3 shows the pseudocode for SL-RK integration.

4 Experimental Setup

In this work, we evaluate the performance of SL-RK integration in all the variants of PDE-LDDMM (see Table 2). The evaluation has been performed consistently with our previous work [8,9], in order to show the improvement in the proposed integration method over RK integration. In addition, we have performed an extensive evaluation of the most memory efficient stationary methods in the frameworks of Klein et al. [12] and Rohlfing et al. [25] in order to establish the position achieved by PDE-LDDMM methods in these evaluation rankings. Finally, we show some complementary experiments justifying the selection of SL-RK as integration scheme for PDE-LDDMM.

4.1 Datasets

We have used five different databases in our evaluation:

NIREP16 contains 16 skull-stripped brain images with the segmentation of 32 gray matter structures. The dimension of the images is $256 \times 300 \times 256$ with a voxel size of $0.7 \times 0.7 \times 0.7$ mm. The acquisition and post-processing details can be found at the web page (<http://www.nirep.org>). The most remarkable features of this dataset are the excellent image quality and the ventricle sizes that are usually small. The geometry of the segmentations provides a specially challenging framework for deformable registration evaluation.

LPBA40 contains 40 skull-stripped brain images without the cerebellum and the brain stem. LPBA40 is provided with the segmentation of 50 gray matter structures together

Table 1 Original PDEs involved in BL PDE-LDDMM and corresponding PDEs written in SL form

Equation in original form	Equation in SL form
$\partial_t m(t) + \nabla m(t) \cdot \iota(\tilde{v}_t) = 0$ (Eq. 19)	$D_t m(t) = 0$
$-\partial_t \lambda(t) - \nabla \cdot (\lambda(t) \cdot \iota(\tilde{v}_t)) = 0$ (Eq. 20)	$-D_t \lambda(t) = \lambda(t) \nabla \cdot \iota(\tilde{v}_t)$
$\partial_t \tilde{u}(t) + \tilde{D} \tilde{u}(t) \star \tilde{v}_t = \tilde{v}_t$ (Eq. 26)	$\tilde{D}_t \tilde{u}(t) = \tilde{v}_t$
$-\partial_t \tilde{\rho}(t) - \tilde{\nabla} \cdot (\tilde{\rho}(t) \star \tilde{v}_t) = 0$ (Eq. 38)	$-\tilde{D}_t \tilde{\rho}(t) = \tilde{\rho}(t) \star \tilde{\nabla} \cdot \tilde{v}_t$
$\partial_t \delta \tilde{m}(t) + \tilde{\nabla} \delta \tilde{m}(t) \star \tilde{v}_t + \tilde{\nabla} \tilde{m}(t) \star \delta \tilde{v}(t) = 0$ (Eq. 21)	$\tilde{D}_t \delta \tilde{m}(t) = -\tilde{\nabla} \tilde{m}(t) \star \delta \tilde{v}_t$
$-\partial_t \delta \tilde{\lambda}(t) - \tilde{\nabla} \cdot (\delta \tilde{\lambda}(t) \star \tilde{v}_t) = 0$ (Eq. 22)	$-D_t \delta \tilde{\lambda}(t) = \tilde{\lambda}(t) \star \tilde{\nabla} \cdot \tilde{v}_t$
$\partial_t \delta \tilde{u}(t) + \tilde{D} \delta \tilde{u}(t) \star \tilde{v}_t + \tilde{D} \tilde{u}(t) \star \delta \tilde{v}(t) = \delta \tilde{v}(t)$ (Eq. 39)	$\tilde{D}_t \delta \tilde{u}(t) = \delta \tilde{v}_t - \tilde{D} \tilde{u}(t) \star \delta \tilde{v}(t)$
$-\partial_t \delta \tilde{\rho}(t) - \tilde{\nabla} \cdot (\delta \tilde{\rho}(t) \star \tilde{v}_t) = 0$ (Eq. 40)	$-\tilde{D}_t \delta \tilde{\rho}(t) = \delta \tilde{\rho}(t) \star \tilde{\nabla} \cdot \tilde{v}_t$

with the caudate, putamen, and hippocampus. LPBA40 protocols can be found at: <http://www.loni.ucla.edu/Protocols/LPBA40>.

The image quality in LPBA40 is, overall, acceptable. The variability of the ventricle sizes is high.

IBSR18 contains 18 brain images with the segmentation of 96 cerebral structures. The masks for skull-stripping are available with the dataset. In addition, the release **IBSR_V2.0** skull-stripped NIFTI [25] contains 18 skull-stripped brain images with the segmentation of 62 cerebral structures. This dataset provides the segmentation of brain structures of interest for the evaluation of image registration methods. The image quality is low. For example, most of the images show motion artifacts. The variability of the ventricle sizes is high.

CUMC12 contains 12 full brain images with the segmentation of 130 cerebral structures. The masks for skull-stripping are available with the dataset. Overall, the image quality is acceptable, although some of the images are noisy. The contrast of the images is low. The variability of the ventricle sizes is high.

MGH10 contains 10 full brain images with the segmentation of 106 cerebral structures. The masks for skull-stripping are available with the dataset. Overall, the image quality is acceptable, although some of the images are noisy. The contrast of the images is low. Ventricle sizes are usually all big.

4.2 Image Registration Pipeline

The evaluation consistent with our previous work was performed in a subsampled NIREP16 database. The registrations were carried out from the first subject to every other subject in the database, yielding to a total of 15 registrations per method. The subsampled NIREP16 database was obtained from the resampling of the images into volumes of size $180 \times 210 \times 180$ with a voxel size of $1.0 \times 1.0 \times 1.0$ mm after the alignment to a common coordinate system using affine transformations. The images were scaled between 0 and 1. The affine alignment and subsampling were performed using the Insight Toolkit (ITK). The PDE-constrained registration methods were executed directly on this dataset. For bench-

marking, we run single- and multi-resolution versions of the SyN version of ANTS diffeomorphic registration [2] with L^2 image similarity (ANTS-SSD).

The evaluation in the framework of Klein et al. was performed in NIREP16, LPBA40, IBSR18, CUMC12, and MGH10 databases. The IBSR18, CUMC12, and MGH10 images normalized with respect to the MNI152 space were used as input data. The registrations were carried out from every subject to every other subject in each database yielding to a total of 2328 registrations per method. The evaluation in the framework of Rohlfing et al. was performed in IBSR18 database, with a total of 306 registrations per method. The NIREP16, LPBA40, IBSR18, CUMC12, and MGH10 images were preprocessed similarly to [12]. In the first place, N4 bias field correction and histogram matching were applied to all the images. To perform these preprocessing steps we used the algorithms available in ITK. The images were scaled between 0 and 1. Next, we performed an affine registration between all the image pairs. Instead of using the affine registered images as input of our non-rigid registration methods, we used the affine transformation as input, and it was included in the parameterization of the diffeomorphic transformations.

Subsampled NIREP16 experiments were run on a cluster equipped with one NVidia Titan RTX with 24 GBS of video memory and an Intel Core i7 with 64 GBS of DDR3 RAM. NIREP16, LPBA40, IBSR18, CUMC12, and MGH10 experiments were run on a cluster equipped with four NVidia GeForce GTX 1080 ti with 11 GBS of video memory and an Intel Core i7 with 64 GBS of DDR3 RAM. The codes were developed in the GPU with MATLAB 2017a and Cuda 8.0. Since MATLAB lacks a 3D GPU cubic interpolator, we implemented in a Cuda MEX file the GPU cubic interpolator with prefiltering proposed in [26].

4.3 Parameter Configuration

Regularization parameters were selected from a search of the optimal parameters in the registration experiments performed in our previous work [9]. We selected the parameters $\sigma^2 =$

Table 2 List of the PDE-LDDMM methods compared in this work

Method	Abbreviation	Variant	Integration	Publication
PDE-constrained LDDMM	PDE-LDDMM RK	I	RK	Mang et al. [13]
PDE-constrained LDDMM based on the state equation	PDE-LDDMM st. RK	II	RK	Hernandez [9]
PDE-constrained LDDMM based on the deformation state equation	PDE-LDDMM def. RK	III	RK	Hernandez [9]
PDE-constrained LDDMM	PDE-LDDMM SL	I	SL-RK	Mang et al. [15]
PDE-constrained LDDMM based on the state equation	PDE-LDDMM st. SL	II	SL-RK	This work
PDE-constrained LDDMM based on the deformation state equation	PDE-LDDMM def. SL	III	SL-RK	This work
band-limited PDE-constrained LDDMM	BL PDE-LDDMM RK	I	RK	Hernandez [8]
band-limited PDE-constrained LDDMM based on the state equation	BL PDE-LDDMM st. RK	II	RK	Hernandez [9]
band-limited PDE-constrained LDDMM based on the deformation state equation	BL PDE-LDDMM def. RK	III	RK	Hernandez [9]
band-limited PDE-constrained LDDMM	BL PDE-LDDMM SL	I	SL-RK	This work
band-limited PDE-constrained LDDMM based on the state equation	BL PDE-LDDMM st. SL	II	SL-RK	This work
band-limited PDE-constrained LDDMM based on the deformation state equation	BL PDE-LDDMM def. SL	III	SL-RK	This work

The table gathers the name of the method given in Sect. 2, the abbreviations used in this experimental section, the variant identifier, the integration scheme, and the publication where the method first appeared

Table 3 Subsampled NIREP16

Method ingredients	Variant	n_t	E_{total}	E_{reg}	E_{img}	MSE_{rel}	$\ g\ _{\infty, \text{rel}}$
St. RK	I	50	$(2.43 \pm 0.19)e^{-3}$	$(1.48 \pm 0.18)e^{-3}$	$(1.69 \pm 0.20)e^{-3}$	18.29 ± 2.83	0.07 ± 0.05
St. RK	II	30	$(2.48 \pm 0.36)e^{-3}$	$(1.51 \pm 0.20)e^{-3}$	$(1.72 \pm 0.37)e^{-3}$	18.42 ± 2.71	0.18 ± 0.19
St. RK	III	30	$(2.36 \pm 0.19)e^{-3}$	$(1.59 \pm 0.11)e^{-3}$	$(1.56 \pm 0.17)e^{-3}$	16.83 ± 1.53	0.16 ± 0.08
St. SL	I	5	$(2.66 \pm 0.22)e^{-3}$	$(1.69 \pm 0.11)e^{-3}$	$(1.82 \pm 0.20)e^{-3}$	19.55 ± 1.76	0.14 ± 0.13
St. SL	II	5	$(2.57 \pm 0.22)e^{-3}$	$(1.73 \pm 0.20)e^{-3}$	$(1.70 \pm 0.16)e^{-3}$	18.34 ± 1.68	0.23 ± 0.21
St. SL	III	5	$(2.46 \pm 0.18)e^{-3}$	$(1.74 \pm 0.17)e^{-3}$	$(1.59 \pm 0.15)e^{-3}$	17.10 ± 1.50	0.12 ± 0.05
NSt. RK	I	50	$(2.62 \pm 0.29)e^{-3}$	$(1.04 \pm 0.52)e^{-3}$	$(2.10 \pm 0.47)e^{-3}$	22.68 ± 5.49	0.26 ± 0.12
NSt. RK	II	25*	$(3.22 \pm 0.30)e^{-3}$	$(0.89 \pm 0.36)e^{-3}$	$(2.78 \pm 0.26)e^{-3}$	29.92 ± 2.41	<i>0.32 ± 0.10</i>
NSt. RK	III	30	$(2.18 \pm 0.19)e^{-3}$	$(1.36 \pm 0.15)e^{-3}$	$(1.50 \pm 0.21)e^{-3}$	16.10 ± 1.70	0.20 ± 0.08
NSt. SL	I	5	$(3.14 \pm 0.49)e^{-3}$	$(2.11 \pm 0.54)e^{-3}$	$(2.09 \pm 0.64)e^{-3}$	22.44 ± 6.26	<i>0.51 ± 0.18</i>
NSt. SL	II	5	$(2.99 \pm 0.24)e^{-3}$	$(1.91 \pm 0.37)e^{-3}$	$(2.03 \pm 0.31)e^{-3}$	21.82 ± 2.51	<i>0.49 ± 0.19</i>
NSt. SL	III	5	$(2.47 \pm 0.18)e^{-3}$	$(1.81 \pm 0.25)e^{-3}$	$(1.57 \pm 0.20)e^{-3}$	16.86 ± 1.91	0.13 ± 0.04
BL 40x St. RK	I	25	$(2.09 \pm 0.28)e^{-3}$	$(0.27 \pm 0.04)e^{-3}$	$(1.95 \pm 0.29)e^{-3}$	20.99 ± 2.59	0.04 ± 0.03
BL 32x St. RK	II	25	$(1.85 \pm 0.15)e^{-3}$	$(0.26 \pm 0.03)e^{-3}$	$(1.72 \pm 0.14)e^{-3}$	18.53 ± 1.71	0.02 ± 0.01
BL 32x St. RK	III	25	$(1.75 \pm 0.15)e^{-3}$	$(0.29 \pm 0.03)e^{-3}$	$(1.61 \pm 0.14)e^{-3}$	17.32 ± 1.68	0.03 ± 0.01
BL 40x St. SL	I	5	$(2.02 \pm 0.15)e^{-3}$	$(0.25 \pm 0.03)e^{-3}$	$(1.89 \pm 0.14)e^{-3}$	20.41 ± 1.89	0.02 ± 0.00
BL 32x St. SL	II	5	$(1.96 \pm 0.14)e^{-3}$	$(0.22 \pm 0.02)e^{-3}$	$(1.84 \pm 0.13)e^{-3}$	19.89 ± 1.76	0.01 ± 0.00
BL 32x St. SL	III	5	$(1.79 \pm 0.15)e^{-3}$	$(0.29 \pm 0.02)e^{-3}$	$(1.65 \pm 0.15)e^{-3}$	17.77 ± 1.66	0.04 ± 0.01
BL 40x NSt. RK	I	25	$(2.79 \pm 0.39)e^{-3}$	$(0.12 \pm 0.05)e^{-3}$	$(2.73 \pm 0.40)e^{-3}$	29.30 ± 3.50	0.21 ± 0.08
BL 32x NSt. RK	II	25	$(2.89 \pm 0.18)e^{-3}$	$(0.32 \pm 0.18)e^{-3}$	$(2.73 \pm 0.19)e^{-3}$	29.21 ± 3.96	0.24 ± 0.11
BL 32x NSt. RK	III	25	$(1.57 \pm 0.15)e^{-3}$	$(0.23 \pm 0.03)e^{-3}$	$(1.46 \pm 0.14)e^{-3}$	15.68 ± 1.52	0.04 ± 0.01
BL 40x NSt. SL	I	5	$(2.03 \pm 0.20)e^{-3}$	$(0.50 \pm 0.08)e^{-3}$	$(1.78 \pm 0.20)e^{-3}$	19.14 ± 2.06	0.11 ± 0.05
BL 32x NSt. SL	II	5	$(2.10 \pm 0.16)e^{-3}$	$(0.39 \pm 0.04)e^{-3}$	$(1.90 \pm 0.16)e^{-3}$	20.49 ± 1.75	0.17 ± 0.15
BL 32x NSt. SL	III	5	$(1.70 \pm 0.15)e^{-3}$	$(0.29 \pm 0.02)e^{-3}$	$(1.56 \pm 0.14)e^{-3}$	16.80 ± 1.57	0.07 ± 0.03

Convergence results. Mean and standard deviation of the total energy E_{total} , regularization E_{reg} , image similarity energy E_{img} , relative image similarity error expressed in % (MSE_{rel}), and the relative gradient magnitude ($\|g\|_{\infty, \text{rel}}$). The method ingredients are the diffeomorphism parameterization in the spatial or band-limited domain and the integration scheme. St. stands for the stationary and NSt. for the non-stationary parameterization of diffeomorphisms. The best performing E_{total} , E_{img} , and MSE_{rel} values in each group are highlighted with boldface. The bad performing values are set in italics. The $\|g\|_{\infty, \text{rel}}$ values indicating convergence stagnation are also highlighted in italics. (*) The number of time steps n_t selected for NSt. PDE-LDDMM, st. eq. with RK integration was due to a memory load greater than the GPU maximum capacity (24 GBS)

1.0, $\alpha = 0.0025$, and $s = 2$ and a unit-domain discretization of the image domain Ω [3].

The optimization was run a maximum of 10 iterations with the stopping conditions used in [13]. The maximum number of PCG iterations was selected equal to 5. These parameters were selected as optimal in our previous work since the methods achieved state of the art accuracy at a reasonable amount of time [8].

The experiments were performed with band sizes of 32×32 for BL PDE-LDDMM based on the state and on the deformation state equations (Variant II and III), and band sizes of $40 \times 40 \times 40$ for original BL PDE-LDDMM (Variant I). This selection was found as optimal for each method in our previous work [8,9].

For SL-RK integration, the number of time steps n_t was selected equal to 5 for all the methods. For RK integration, n_t was selected equal to 25 for the BL PDE-LDDMM based on

the state and on the deformation state equations, and 50 for the spatial methods due to stability issues. In the evaluation with LPBA40, IBSR18, CUMC12, and MGH10 datasets, $n_t = 25$ showed stability issues in a considerable number of experiments and it was raised to 50.

ANTS-SSD was run with the following parameters for the single-resolution experiments

```
$synconvergence=" [50, 1e-6, 10] ",
$synshrinkfactors=" 1 ",
and $synsmoothingsigmas=" 3vox ".
```

For the multi-resolution experiments the parameters were set to

```
$synconvergence=" [50x50x50, 1e-6, 10] ",
$synshrinkfactors=" 4x2x1 ",
and $synsmoothingsigmas=" 3x2x1vox ".
```

Table 4 Subsampled NIREP16

Method ingredients	Variant	$\min(J(\phi_1^v)^{-1})$	$\max(J(\phi_1^v)^{-1})$
St. RK	I	0.16 ± 0.05	3.70 ± 0.51
St. RK	II	0.07 ± 0.06	2.95 ± 0.41
St. RK	III	0.12 ± 0.05	4.12 ± 1.41
St. SL	I	0.03 ± 0.04	3.25 ± 0.29
St. SL	II	0.03 ± 0.04	3.46 ± 0.69
St. SL	III	0.14 ± 0.05	5.02 ± 1.01
NSt. RK	I	0.16 ± 0.05	3.38 ± 0.90
NSt. RK	II	0.11 ± 0.07	2.93 ± 0.78
NSt. RK	III	0.09 ± 0.03	4.55 ± 0.69
NSt. SL	I	0.02 ± 0.02	3.96 ± 0.65
NSt. SL	II	0.01 ± 0.02	4.20 ± 0.94
NSt. SL	III	0.09 ± 0.02	7.35 ± 2.63
BL 40x St. RK	I	0.21 ± 0.04	3.32 ± 0.41
BL 32x St. RK	II	0.11 ± 0.07	3.10 ± 0.35
BL 32x St. RK	III	0.13 ± 0.05	3.76 ± 0.49
BL 40x St. SL	I	0.32 ± 0.02	9.65 ± 4.77
BL 32x St. SL	II	0.29 ± 0.03	7.94 ± 2.81
BL 32x St. SL	III	0.21 ± 0.03	8.23 ± 3.24
BL 40x NSt. RK	I	0.24 ± 0.05	2.25 ± 0.51
BL 32x NSt. RK	II	0.15 ± 0.06	2.71 ± 0.48
BL 32x NSt. RK	III	0.09 ± 0.04	4.77 ± 0.78
BL 40x NSt. SL	I	0.01 ± 0.07	5.25 ± 0.80
BL 32x NSt. SL	II	0.07 ± 0.03	4.51 ± 0.61
BL 32x NSt. SL	III	0.09 ± 0.03	7.37 ± 1.43

Mean and standard deviation of the Jacobian determinant extrema associated with the transformation $(\phi_1^v)^{-1}$. The method ingredients are the diffeomorphism parameterization in the spatial or band-limited domain and the integration scheme. St. stands for the stationary and NSt. for the non-stationary parameterization of diffeomorphisms

The selection of the number of iterations was in agreement with the number of outer × inner iterations used in Gauss–Newton–Krylov optimization.

5 Results

5.1 Subsampled NIREP16 Evaluation Results

5.1.1 Convergence Analysis

Table 3 shows, averaged by the number of experiments, the mean and standard deviation of the total, regularization, and image similarity energies after registration (E_{total} , E_{reg} , and E_{img}), the relative image similarity error,

$$MSE_{rel} = \frac{\|m(1) - I_1\|_{L^2}^2}{\|I_0 - I_1\|_{L^2}^2},$$

and the relative gradient magnitude,

$$\|g\|_{\infty,rel} = \frac{\|\nabla_v E(v^n)\|_{\infty}}{\|\nabla_v E(v^0)\|_{\infty}},$$

obtained with PDE-LDDMM in the subsampled NIREP16 dataset. In addition, Table 4 shows the mean and the standard deviation of the extrema of the Jacobian determinant.

Overall, the worst-performing methods show high values for the relative gradient in Table 3, which indicate the stagnation of the convergence. For most of the BL methods, the relative gradient was reduced to average values ranging from 0.02 to 0.04, which means that the optimization was stopped in acceptable energy values. All the Jacobians remained above zero.

Next, we group the analysis of Table 3 by integration scheme, variant, image domain, and diffeomorphism parameterization:

- RK versus SL-RK. In absolute terms, the E_{total} values obtained with SL-RK methods tend to be greater than those achieved by RK integration. Both E_{reg} and E_{img} values contribute to the greater E_{total} values for the SL-RK methods. However, in relative terms, the MSE_{rel} values achieved by SL-RK methods at convergence are close or even improve RK methods. It drives our attention to the bad performance of Variant II for the non-stationary parameterization and RK integration which is indeed improved by SL-RK integration.
- Variants. The best performing variant, with the best E_{total} , E_{img} , and MSE_{rel} values, is Variant III. This result is persistent for different integration schemes, the spatial or BL parameterization, and the diffeomorphic parameterization.
- SP versus BL. Due to the high-frequency suppression property of the BL parameterization, the E_{reg} values are all smaller for the BL methods. The MSE_{rel} values obtained with the spatial methods are slightly degraded by the BL methods as expected. The degradation is only shown in some cases.
- St. versus NSt. The E_{reg} values for the stationary parameterization are greater than for the non-stationary parameterization. On the one hand, the stationary parameterization yields one-parameter subgroups that are not geodesics due to the non-bi-invariance of the metric. On the other hand, the minimizing E_{reg} property of geodesics is expected for the solutions with the non-stationary parameterization. Therefore, the obtained E_{reg} results are consistent with these two facts. The non-stationary methods do not outperform the stationary methods in a consistent manner. The (out)performance depends on the variant.

5.1.2 Evaluation

The evaluation is based on the accuracy of the registration results for template-based segmentation. The Dice Similarity Coefficient (DSC) is selected as the evaluation metric. Given two segmentations S and T , the DSC is defined as

$$\text{DSC}(S, T) = \frac{2\text{Vol}(S \cap T)}{\text{Vol}(S) + \text{Vol}(T)}. \quad (52)$$

This metric provides the value of 1 if S and T exactly overlap and gradually decreases toward 0 depending on the overlap of the two volumes.

Figure 1 shows, in the shape of box and whisker plots, the statistical distribution of the DSC values obtained after the registration across the 32 segmented structures. For the single-resolution experiments, the performance of the benchmark method ANTS-SSD was under 50%. For the multi-resolution experiments, the average DSC value achieved by ANTS-SSD equals to 55.59%. We have selected this value as a baseline of good registration accuracy for methods with L^2 -based image similarity.

A great number of PDE-LDDMM methods showed similar values or even improved ANTS-SSD performance. The best performing variant was our PDE-LDDMM based on the deformation state equation, Variant III (boxes in pink tones). This variant showed similar results for RK and SL-RK integration regardless the image domain and diffeomorphism parameterization.

For the variant associated with the PDE-constrained benchmark methods [13,15], Variant I (boxes in blue tones), RK integration slightly outperformed SL-RK integration for the stationary parameterization. For the non-stationary parameterization the median DSC value for RK integration was under the value for SL-RK integration. Similarly, RK slightly outperformed SL-RK integration for Variants I and II of stationary BL PDE-LDDMM. On the contrary, SL-RK integration greatly outperformed RK integration for the non-stationary parameterization.

Variant II (boxes in green tones) performed similarly to benchmark Variant I for the stationary parameterization for the same integration scheme. However, it is remarkable the low performance achieved by this variant for the non-stationary parameterization and RK integration in both image domains.

Table 5 shows the results of the analysis of variance (ANOVA) for the effects of variant (I, II, III), integration scheme (RK, SL-RK), domain (SP, BL), and parameterization (St, NSt) selection on the distribution of the DSC values obtained in the subsampled NIREP16 evaluation experiments. The tests showed statistical significance in all the considered factors except for the domain factor. This means that the accuracy of the methods using the spatial domain is

statistically indistinguishable from the accuracy of the corresponding methods using the band-limited domain. From the analysis for each separated variant of the effects of integration scheme, domain, and parameterization, the tests showed statistical significance for Variants I and II. For the best performing variant (Variant III), no factor showed any statistical significance.

Figure 2 shows the p values of pairwise right-tailed Wilcoxon rank-sum tests for the assessment of the statistical significance of the difference of medians for the distribution of the DSC values obtained in the registration experiments. The alternative hypothesis is that the median of the first distribution is higher than the median of the second one. For increasing the interpretability of the tests, we have grouped the comparisons into the spatial methods, the band-limited methods, and Variant III methods. The figure shows statistical significance for the better performance of Variant III methods over ANTS-SSD and the combinations of Variants I and II underperforming ANTS-SSD. For Variant III, the differences in the distribution of the DSC for the different combinations of integration and parameterization are not statistically significant.

5.1.3 Computational Complexity

The analysis of the memory complexity performed in [9] for RK integration reported an $O(TN)$ for Variant I, and an $O(TN^d)$ for Variants II and III, where T represents the time sampling selected for PDE integration, N is the size of the discretized domain, and d is the dimensionality of the image domain (3D). The memory complexity analysis holds for SL-RK integration. Therefore, it is still expected a reduction in the VRAM usage since the n_t for SL-RK is considerably smaller than the n_t for RK.

The time complexity for RK integration is $O(TN \log N)$ for Variant I, and $O(TN^d \log N)$ for Variants II and III. The time complexity analysis also holds for SL-RK integration. Despite the extra computations of the departure points, the cubic interpolation of the right-hand-side values of the PDEs, and the extra projection-inclusion from \tilde{V} to V needed in the BL version of the variants, it is expected a reduction in the computation time for SL-RK with respect to RK integration due to the dependence of the complexity with T .

Table 6 shows the VRAM peak memory reached through the computations, and the average and standard deviation of the total computation time in the subsampled NIREP16 experiments. For the spatial methods, SL-RK integration achieved a substantial time and memory reduction over RK integration, as expected. The time and memory reduction achieved by SL-RK over RK integration was also considerable for the BL parameterized methods. For the stationary parameterization, the BL parameterization further decreased the complexity of spatial SL-RK integration methods, as

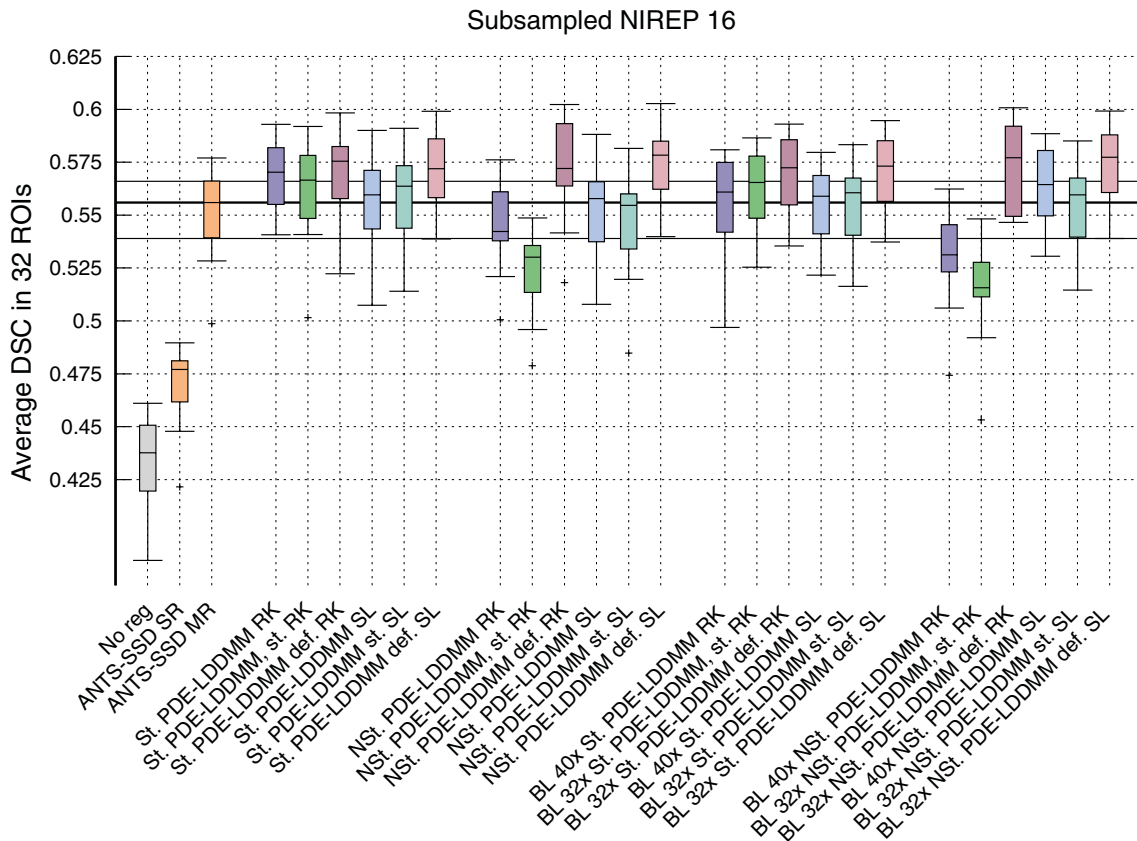


Fig. 1 Subsampled NIREP16. Volume overlap obtained by the registration methods measured in terms of the DSC between the warped and the corresponding manual target segmentations. Box and whisker plots show the distribution of the DSC values averaged over the 32 NIREP manual segmentations. The whiskers indicate the minimum and maximum

of the DSC values. The horizontal lines in the plot indicate the first, second, and third quartiles of multi-resolution ANTs-SSD. Variant I corresponds with boxes in blue tones, Variant II in green tones, and Variant III in pink tones

Table 5 Subsampled NIREP16

Variant	Integration	Domain	Param.
All variants			
0.0001	0.0093	0.2495	0.0004
Separated variants			
I	0.0712	0.3339	0.0041
II	0.0016	0.8407	0.0000
III	0.3235	0.3908	0.5469

Results of ANOVA tests for the effects of variant, integration, domain, and parameterization selection. Up table shows the results of the four factors. The down table shows the results of the effects of integration, domain, and parameterization for each separated variant. Statistical significance is highlighted in boldface

expected. However, SL-RK integration did not reduced the VRAM memory usage for the non-stationary parameterization. The total computation time was effectively reduced.

5.1.4 Qualitative Assessment

For a qualitative assessment of the proposed registration methods, we show the registration results obtained by Mang et al. benchmark methods [13,18] (Variant I), and PDE-LDDMM based on the deformation state equation (Variant III) in a selected experiment representative of a difficult deformable registration problem. For the non-stationary parameterization the images for a qualitative assessment were similar to the stationary parameterization. Figure 3 shows the warped images, the difference between the warped and the target images after registration, and the velocity field. All the methods provide visually acceptable results.

5.2 NIREP16, LPBA40, IBSR18, CUMC12, and MGH10 Evaluation Results

From the evaluation measurements used in Klein et al. framework, we focus on the accuracy of the registration results for template-based segmentation. Since [12], this has been

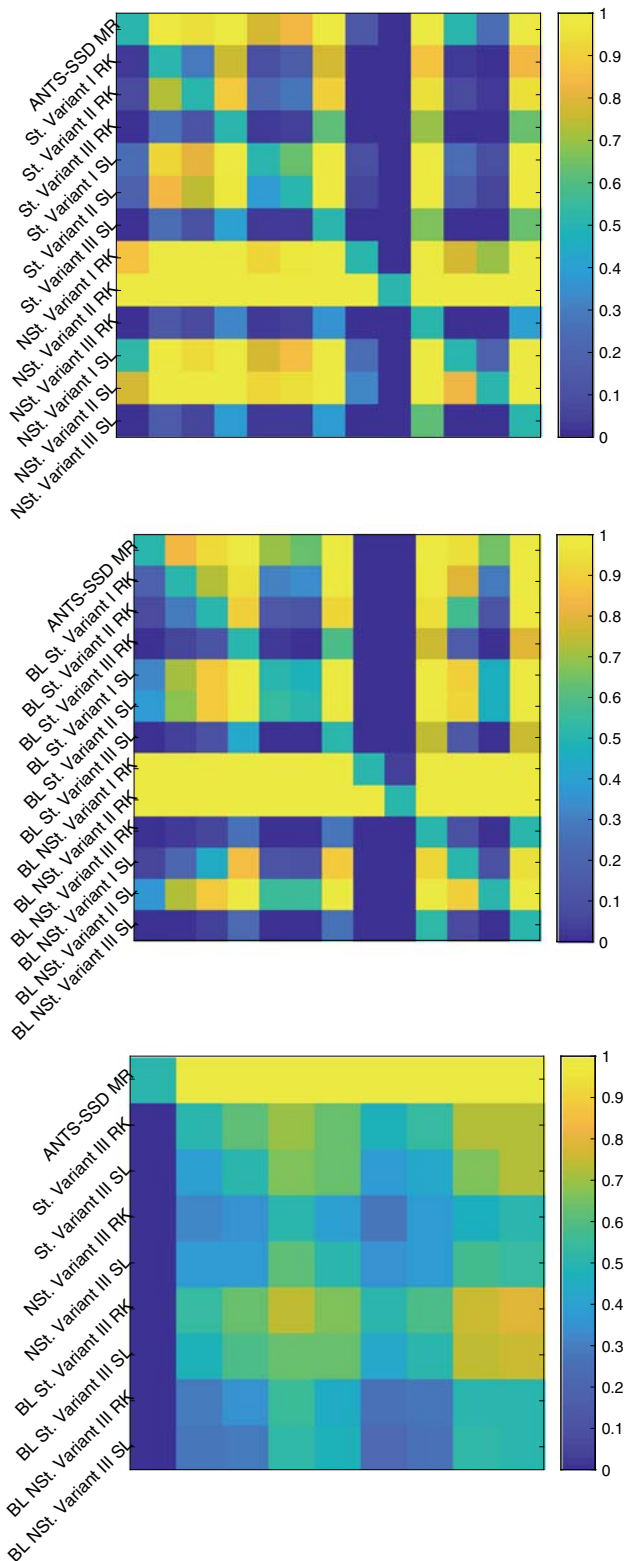


Fig. 2 Subsampled NIREP16. Results of the pairwise right-tailed Wilcoxon rank-sum tests. Up figure, methods with the spatial parameterization. Center figure, methods with the band-limited parameterization. Down figure, Variant III methods (Color figure online)

adopted as a widely extended criterion for non-rigid registration evaluation. From the metrics proposed in [12], we select the Dice Similarity Coefficient (DSC) as evaluation metric. Figures 4, 5 and 6 show the statistical distribution of the DSC values obtained after the registration across the manually segmented structures for the five databases. For the NIREP16 dataset, we show the results obtained with ANTS-SSD. For the remaining databases, we include the results reported in [12] for affine registration (FLIRT), and three diffeomorphic registration methods: Diffeomorphic Demons, SyN, and Dartel.

The results from NIREP16 show how PDE-LDDMM based on the deformation state equation (Variant III) outperformed the other variants of PDE-LDDMM. The distribution of the method parameterized in the BL domain resulted almost identical to the distribution of the method parameterized in the spatial domain. As happened with subsampled NIREP16 evaluation, ANTS-SSD was among the worst performing methods. The band-limited versions of Variant I and II with RK integration exceeded the maximum VRAM capacity of our GPUs for this dataset.

The results obtained from LPBA40, IBSR18, CUMC12, and MGH10 show that, from the PDE-LDDMM methods, the best performing method was BL PDE-LDDMM based on the deformation state equation (Variant III) and SL-RK integration. The performance of the method with RK integration was slightly lower. The performance of the spatial versions of Variant I and II and their band-limited versions was significantly lower in IBSR18, CUMC12, and MGH10 databases. For these methods, RK integration performed slightly better than SL-RK integration. These results are consistent with NIREP16 evaluation results.

In IBSR18, CUMC12, and MGH10 databases, our PDE-LDDMM methods were not able to reach SyN or Dartel performance. This is probably because the image similarity metrics used in these methods (Cross-Correlation and multinomial model, respectively) favor the accuracy in template-based segmentation. In contrast, PDE-LDDMM uses SSD, which is known to restrict the performance in template-based segmentation. However, in LPBA40 databases, our best performing PDE-LDDMM methods overpass Dartel and almost reached SyN performance, while showing a significantly reduced number of outliers.

Our methods significantly outperformed FLIRT and Diffeomorphic Demons, where the third quartile in the distribution of our best performing method was close to the median of Demons for the four databases. It should be noticed that Diffeomorphic Demons also uses SSD as image similarity metric.

Table 6 Computational complexity, GPU peak memory usage and mean and standard deviation of the total computation time. Experiments run in an NVidia Titan RTX with 24 GBS of video memory

Method ingredients	Variant	VRAM (MBS)	Total time (s)
St. RK	I	10,311	2281.40 ± 378.56
St. RK	II	17,229	877.85 ± 176.03
St. RK	III	15,321	1298.50 ± 151.19
St. SL	I	3901	118.29 ± 4.77
St. SL	II	6267	139.91 ± 22.06
St. SL	III	5935	221.01 ± 2.40
NSt. RK	I	14,913	3443.17 ± 932.45
NSt. RK	II	20,635	816.79 ± 480.61
NSt. RK	III	19,065	2222.04 ± 369.96
NSt. SL	I	11,155	191.96 ± 45.53
NSt. SL	II	12,429	214.09 ± 48.54
NSt. SL	III	12,065	355.47 ± 4.38
BL 40x St. RK	I	5709	312.45 ± 5.22
BL 32x St. RK	II	4743	315.00 ± 2.73
BL 32x St. RK	III	1819	377.56 ± 4.26
BL 40x St. SL	I	2685	68.66 ± 1.32
BL 32x St. SL	II	2877	80.76 ± 0.35
BL 32x St. SL	III	2365	116.99 ± 1.04
BL 40x NSt. RK	I	6657	825.15 ± 19.93
BL 32x NSt. RK	II	4789	535.52 ± 279.65
BL 32x NSt. RK	III	1863	905.48 ± 25.13
BL 40x NSt. SL	I	6131	158.79 ± 15.29
BL 32x NSt. SL	II	6171	159.49 ± 20.59
BL 32x NSt. SL	III	5577	249.60 ± 10.85

The most efficient methods in each group are highlighted in bold

5.3 IBSR18 V2.0 Evaluation Results

Figure 7 shows the statistical distribution of the DSC values obtained by our proposed registration methods in the regions of interest of Rohlfing et al. evaluation framework [25]. Consistently with the rest of the evaluation results, the best performing method was BL PDE-LDDMM based on the deformation state equation (Variant III), which significantly outperformed the others in the great majority of regions.

5.4 Some Insights of Runge–Kutta and Semi-Lagrangian Integration

Finally, we show some interesting experiments justifying the selection of SL-RK as integration scheme for PDE-LDDMM beyond the evaluation based on the accuracy for template-based segmentation shown in this experimental section. The

experiments have been performed with the spatial version of Variant III.

5.4.1 Euler Versus RK Versus SL-RK Integration

Euler integration is one of the simplest ODE integration methods. Compared with more sophisticated RK schemes, it is noticeably less accurate. The local truncation error of Euler method is h^2 , while for RK method, it is h^5 . This does not seem to be a problem for LDDMM methods based on gradient descent. For example, Euler integration has been extensively used even in geodesic methods based on the solution of the EPDiff equation [36]. However, PDE-LDDMM with Gauss–Newton–Krylov optimization shows convergence problems when combined with Euler method. Table 7 compares the registration results obtained with Euler, RK, and SL-RK integration in a selected subsampled NIREP16 experiment. With Euler integration and $n_t = 10$, the optimization gets stag-

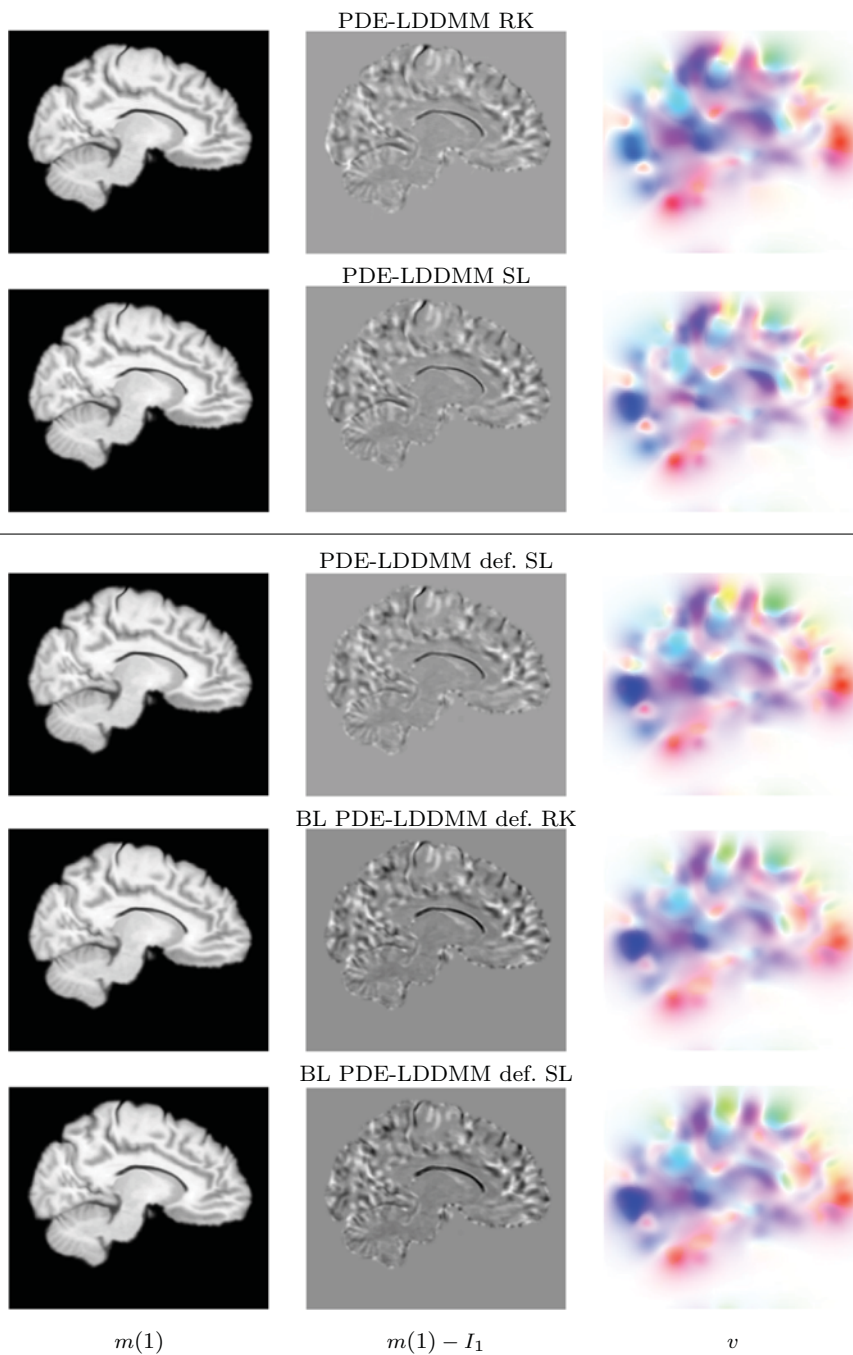
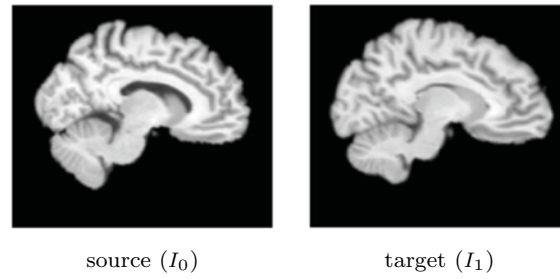
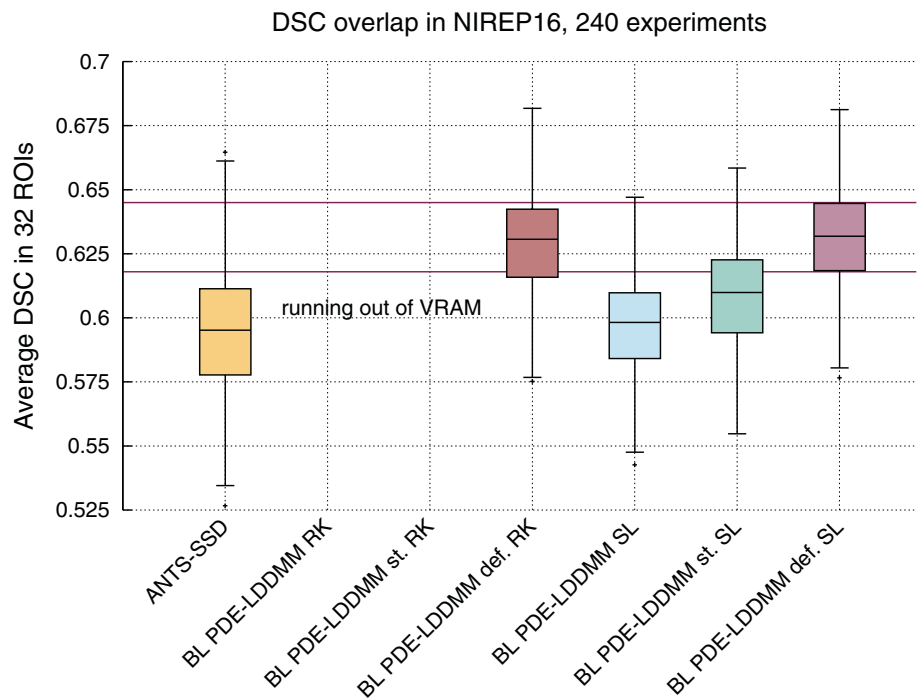


Fig. 3 Subsampled NIREP16. Sagittal view of the warped sources, the intensity differences, and the velocity field after registration for Mang et al. benchmark methods (Variant I) and PDE-LDDMM based on the

deformation state equation (Variant III). Experiments with the stationary parameterization (Color figure online)

Fig. 4 NIREP16. Distribution of the DSC values averaged over the 32 NIREP manual segmentations in the 240 experiments. The whiskers indicate the minimum and maximum of the DSC values. The methods running out of 11 GBS VRAM are indicated in the plots. The horizontal red lines indicate the first and third quartiles of BL PDE-LDDMM based on the deformation state equation



nated in the initial iterations. For $n_t = 25$, $n_t = 30$, and even $n_t = 50$, PCG detects a definite negative Hessian. With RK integration and $n_t = 10$, PCG detects a definite negative Hessian. With RK integration and $n_t = 25$ and $n_t = 30$, the method shows an appropriate convergence behavior. With SL-RK integration and $n_t = 5$, the method shows an appropriate convergence behavior. The best performing method is SL-RK.

5.4.2 Small into Big Parallelepiped Experiment

Figure 8 shows the results of a simulated experiment consisting in the registration of a small into a big parallelepiped noisy image. This example is provided as test images with Mermaid software (<http://mermaid.readthedocs.io>).

The MSE_{rel} reached by RK integration after the registration was equal to 18.37% while, for SL-RK, it was 2.77%. Both integration schemes do not seem to have problems with noise. In the figure, it can be appreciated that RK integration shows problems to adjust the diffeomorphic warp in the corners of the structure while SL-RK integration is much more accurate in these difficult locations. Therefore, in this particular experiment, SL-RK integration overpasses RK in accuracy.

5.4.3 Stability Beyond $t = 1$

In LDDMM and PDE-LDDMM literature, the time domain is typically selected to be $[0, 1]$. This domain is discretized in a number of time samplings enough to achieve the stability in

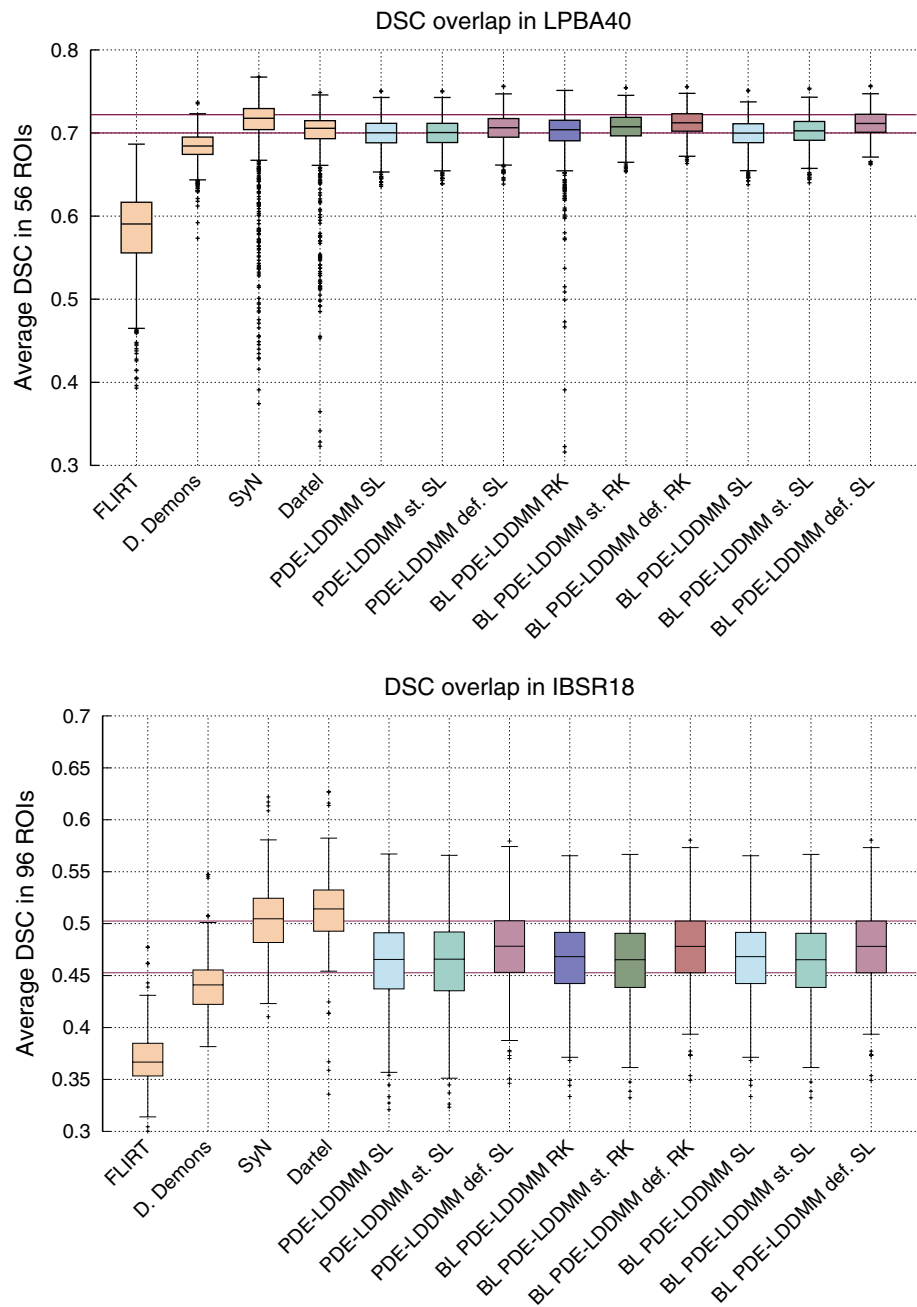
the numerical solvers involved in the computations. However, there are applications where it may be of interest to extend the time domain beyond $t = 1$. For example, time extrapolation may be interesting to predict the anatomical evolution of subjects across time beyond the time limits of temporal deformation models [27]. Integrating beyond $t = 1$ will eventually lead to instabilities of the advected magnitudes. For example, in the deformation state equation, it would mean reaching non-diffeomorphic solutions. This experiment is intended to show the feasibility of using RK and SL-RK integration for time extrapolation.

Figures 9 and 10 show the results of composing the source image with the transformations resulting from the integration of the deformation state equation beyond $t = 1$. For RK integration, we observe that for small extensions of the temporal domain such as $t = 1.25$, the equation has developed instabilities leading to unacceptable results at $t = 1.5$. On the contrary, SL-RK is able to integrate beyond $t = 1$, reaching $t = 3$ with recognizable warped images. This may be due to the unconditionally stable property of SL schemes. Therefore, SL-RK may be appropriate for extrapolation in temporal deformation models.

6 Discussion

The increase in the computational efficiency achieved by the combined over the split BL and SL-RK approaches was significant in terms of computation time and memory. The reduction in the memory requirements allowed us to perform

Fig. 5 LPBA40 and IBSR18. Distribution of the DSC values averaged over the manual segmentations in the registration experiments. The whiskers indicate the minimum and maximum of the DSC values. The horizontal red lines indicate the first and third quartiles of BL PDE-LDDMM based on the deformation state equation (Color figure online)



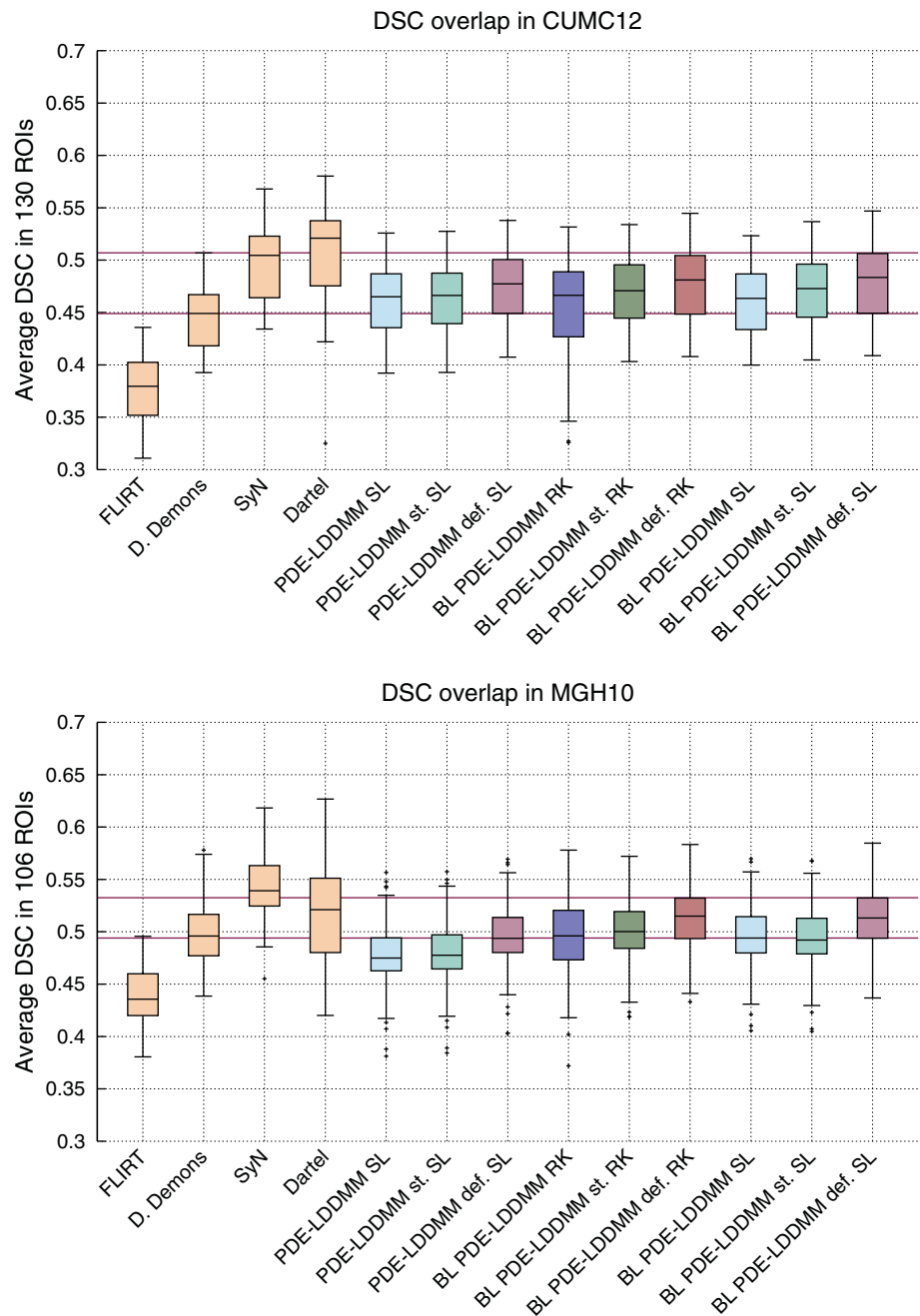
the evaluation of the SL-RK PDE-LDDMM methods extensively, even in the highest-resolution level of NIREP16.

In all the evaluation frameworks, BL PDE-LDDMM based on the deformation state equation with SL-RK integration resulted our best performing method. This method achieved an identical DSC distribution compared with RK integration in the NIREP16 database. The method greatly outperformed ANTS-SSD in this database. In LPBA40, IBSR18, CUMC12, and MGH10 databases, the method outperformed Diffeomorphic Demons. In addition, the evaluation results in the regions of interest of Rohlfing et al. corroborated its excellent performance.

For Mang et al. benchmark PDE-LDDMM methods [13, 15], our evaluation results reported a significant loss of accuracy between RK and SL-RK integration. Interestingly, this loss of accuracy was not observed for our best performing method.

In IBSR18, CUMC12, and MGH10 databases, our PDE-LDDMM methods were not able to reach SyN or Dartel performance. This is because SSD image similarity metric restricts the performance of the methods in template-based segmentation. This problem will be tackled in future work by formulating the PDE-constrained problem with other image similarity metrics such as Normalized Cross-Correlation,

Fig. 6 CUMC12 and MGH10. Distribution of the DSC values averaged over the manual segmentations in the registration experiments. The whiskers indicate the minimum and maximum of the DSC values. The horizontal red lines indicate the first and third quartiles of BL PDE-LDDMM based on the deformation state equation (Color figure online)

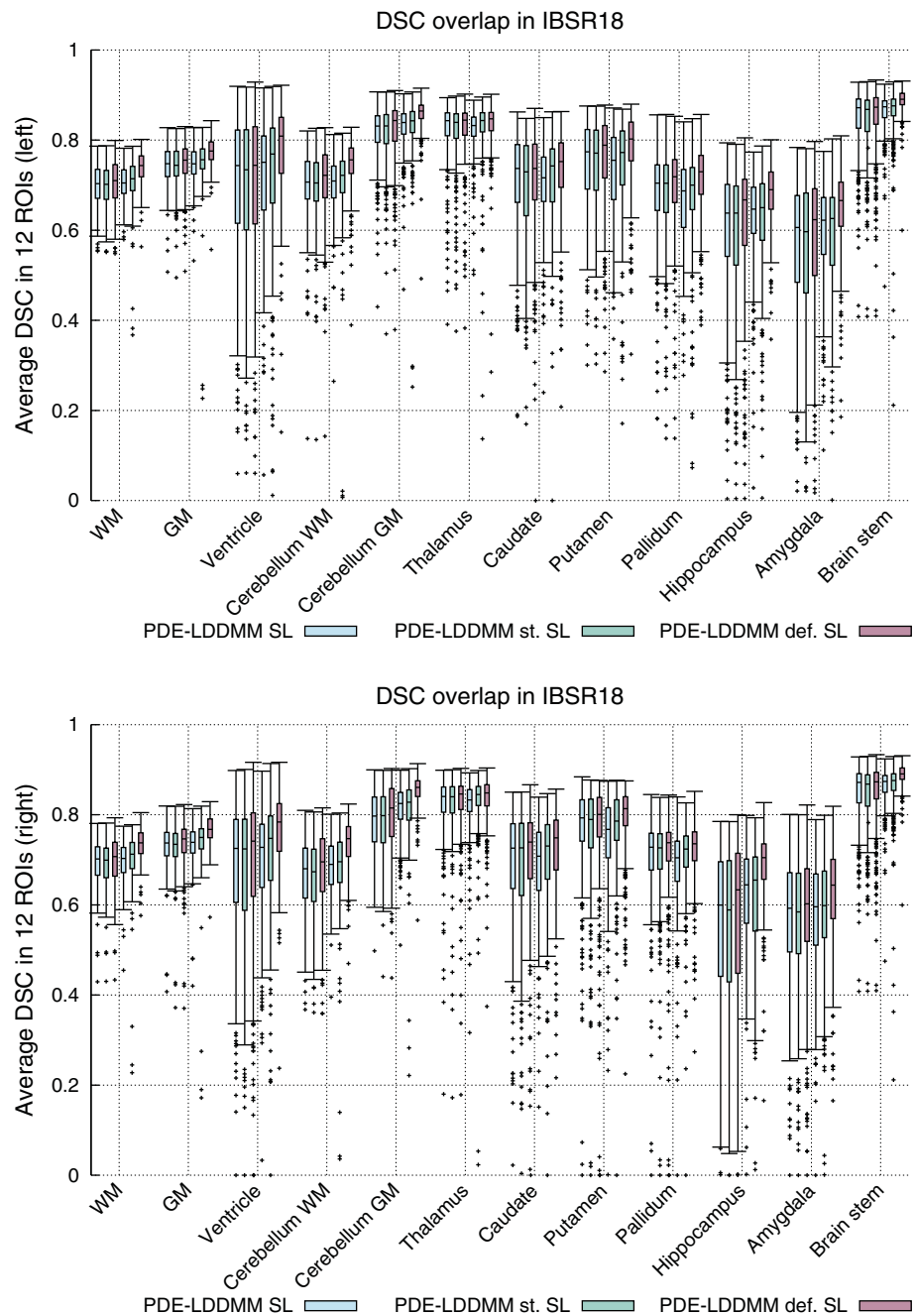


local Normalized Cross-Correlation, or Mutual Information [21]. We expect that this change in the formulation of the problem will increase the performance results of PDE-LDDMM. In addition, it will allow us to apply these methods to other clinical applications involving multi-modal registration.

Simultaneously to the development of this work, Mang et al. released Claire software [17]. The software is intended to exploit massive CPU based parallel computing architectures to accelerate the computation time of PDE-LDDMM. The codes implement original PDE-LDDMM (Variant I)

with a variational extension to nearly incompressible fluids and include H^1 and H^2 regularization terms. The software is restricted to the stationary parameterization of diffeomorphisms. The PDE integration scheme is SL-RK. The software includes a sophisticated multi-level preconditioner that shows to improve the convergence of PCG with respect to the original proposal in [13]. The massive computation allows increasing the number of inner and outer iterations and use the norm of the gradient as stopping condition for achieving an extraordinary accuracy at convergence in a simulated experiment.

Fig. 7 IBSR18 V2.0. Volume overlap obtained by the proposed registration methods. The box and whisker plots show the distribution of the DSC values averaged over the manual segmentations for each region. The whiskers indicate the minimum and maximum of the DSC values. For each group of plots, the first three correspond with the spatial domain parameterization and the three last with the band-limited domain parameterization. WM and GM stand for white matter and grey matter, respectively



In contrast with Claire, our BL methods are intended to run completely in the VRAM of commodity GPUs ($< 4\text{GBS}$). We have limited the variational formulation to the one proposed in [13], although it is straightforwardly extendible to the nearly incompressible fluid problem. We have limited our study to the traditional LDDMM regularizer. Our software works for the stationary and the non-stationary parameterization of diffeomorphisms. We have limited the preconditioner to the one proposed in [13] since we are interested on the comparison of the three different variational variants. We used the stopping conditions suggested in [21] and used for

PDE-LDDMM in [8–10,13,18]. The variety of methods, the extensive evaluation conducted in this work, and our modest hardware capacity hindered us the use of the inner and outer iteration values needed for achieving the stopping conditions based on the norm of the gradient suggested in [17]. In fact, we observed in a selected NIREP experiment that increasing the number of inner iterations in PCG resulted into a faster initial convergence that finally stagnated in greater MSE_{rel} values and lower DSC scores than our considered stopping conditions. This stagnation was also reported in [17] for the simulated experiment. Instead, our selected inner and outer

Table 7 Euler versus RK versus SL-RK results

Euler				
n_t	10	25	30	50
it_{stop}	2	3	3	3
MSE_{rel}	30.04	27.76	22.54	22.84
DSC	55.02	55.44	56.09	56.90
RK				
n_t	10	25	30	
it_{stop}	2	6	10	
MSE_{rel}	30.36	17.50	15.44	
DSC	54.98	58.92	59.83	
SL-RK				
n_t				5
it_{stop}				10
MSE_{rel}				15.91
DSC				59.91

Registration results in a selected subsampled NIREP16 experiment. The table shows the number of time steps n_t , the iteration where the method stopped, and MSE_{rel} and DSC values for assessing the accuracy of the solutions

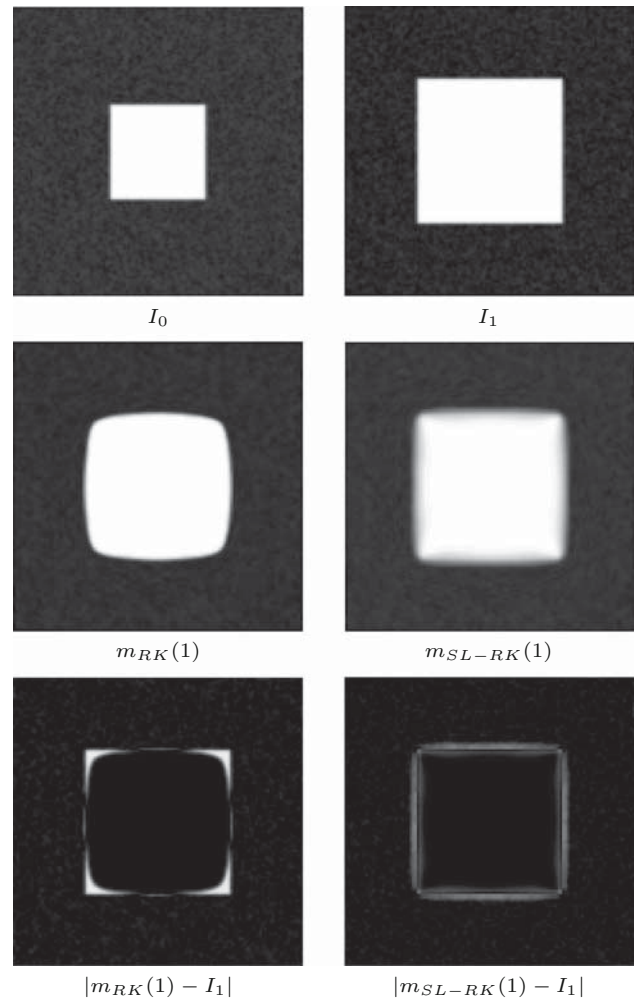
values consumed a reasonable amount of time while obtaining state of the art results for the evaluation metrics.

Comparing Claire and our results, we believe that it would be of interest to implement our best performing variant as a part of Claire's software. In the other direction, it would be very interesting to adopt the multi-level preconditioners in our methods.

With the arise of FlowNet architecture for learning the optical flow in [5], there has been an explosion of methods for non-rigid registration based on deep-learning. These data-driven approaches learn how to build a generative model of deformations from the source and target images. Mermaid (mermaid.readthedocs.io) provides a library of methods where data-driven solutions are based on optimal-transport loss functions, highly related to Variants I and II of PDE-LDDMM formulation. We believe that Mermaid methods may benefit from the parameterization of the problem in the space of band-limited vector fields and the semi-Lagrangian Runge-Kutta schemes proposed in this work. In the other direction, PDE-LDDMM may also benefit from the ingredients of the loss functions defined within these data-driven approaches [28].

7 Conclusions

In this work, we have proposed to combine the two different methodological approaches used to circumventing the huge computational complexity of Gauss-Newton-Krylov PDE-

**Fig. 8** Small into big parallelepiped registration results

LDDMM. In particular, we have included semi-Lagrangian Runge-Kutta integration [15] in the variants of band-limited PDE-LDDMM proposed in [8,9] for further increasing the computational efficiency of these methods. The resulting methods have been extensively evaluated in five different datasets following three different evaluation frameworks. To our knowledge, this is the first time that SL-RK integration is implemented in the framework of PDE-LDDMM for the non-stationary parameterization and in the space of band-limited vector fields. Moreover, our work first provides the position achieved by PDE-LDDMM methods in the ranking of Klein et al. evaluation.

This study positions the formulation of BL PDE-LDDMM based on the deformation state equation and SL-RK integration as the best performing among all PDE-LDDMM methods in terms of accuracy and efficiency. The proposed method has reached the highest sensitivity in the classification of stable versus progressive mild cognitive impaired converters in the Alzheimer's Disease Neuroimaging Initia-

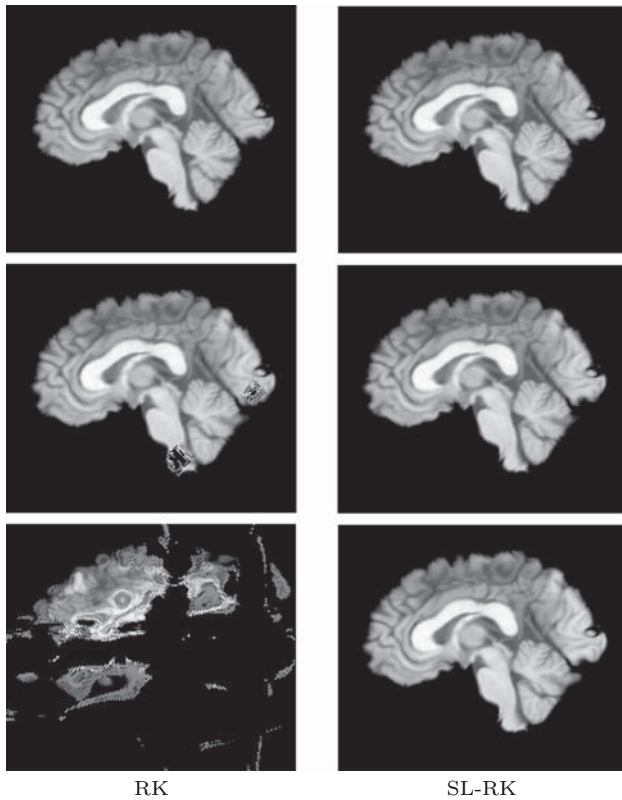


Fig. 9 Integration scheme stability beyond $t = 1$ for RK and SL-RK. Warped images at time samples $t = 1$ (first row), $t = 1.25$ (second row), and $t = 1.5$ (third row)

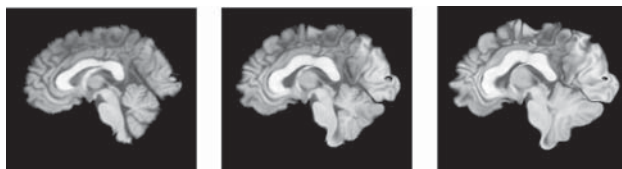


Fig. 10 Integration scheme stability beyond $t = 1$ for SL-RK. Warped images at time samples $t = 1$, $t = 2$, and $t = 3$

tive (ADNI) database using convolutional neural networks. This result has been recently published in [23].

In future work, we will extend this formulation to other relevant physically meaningful LDDMM approaches such as the nearly incompressible method in [14], and the geodesic shooting approach in [10]. We will explore the advantages of using the multi-level preconditioner in [17]. We will adapt our methods for the use of alternative image similarity metrics that usually outperform SSD in registration evaluation rankings. We will try to bridge the gap between constrained variational approaches and data-driven solutions based on optimal-transport loss functions. Finally, we will work in the understanding of which of the features of PDE-LDDMM allow the exceptional classification rates related with Alzheimer’s disease conversion shown in [23].

Acknowledgements The author would like to acknowledge the anonymous reviewers for their valuable revision of the manuscript. The author would like to give special thanks to Wen Mei Hwu from the University of Illinois for interesting ideas in the GPU implementation of the methods, and Nacho Navarro and Rosa Badia from the Barcelona Supercomputing Center (BSC) for their help. This work was partially supported by the National Research Grant TIN2016-80347-R (DIAMOND Project), PID2019-104358RB-I00 (DL-Aging Project), and Government of Aragon Group Reference T64_20R (COS2MOS research group). In addition, this work was supported by NVIDIA through the Polytechnical University of Catalonia/Barcelona Supercomputing Center (UPC/BSC) GPU Center of Excellence.

Appendix

Appendix gathers the expressions of the gradient and the Hessian for the PDE-LDDMM variants defined in the spatial domain and the method for SL-RK integration.

A.1 Original PDE-Constrained LDDMM (Variant I)

Let $E(v)$ be the PDE-constrained variational problem given in Eq. 4. Let us define the Lagrange multipliers $\lambda : \Omega \times [0, 1] \rightarrow \mathbb{R}$ and $\eta : \Omega \rightarrow \mathbb{R}$ associated with the state equation (Eq. 5) and its initial condition. The augmented Lagrangian corresponds with the expression

$$E_{\text{aug}}(v) = E(v) + \int_0^1 \langle \lambda(t), \partial_t m(t) + Dm(t) \cdot v_t \rangle_{L^2} dt + \langle \eta, m(0) - I_0 \rangle_{L^2}. \quad (53)$$

The first- and second-order optimality conditions are derived from the formal computations of

$$\delta E_{\text{aug}}(v, m, \lambda, \eta; dv, dm, d\lambda, d\eta) \quad (54)$$

and

$$\delta^2 E_{\text{aug}}(v, m, \lambda, \eta; dv, dm, d\lambda, d\eta). \quad (55)$$

The details of the formal derivations can be found in [9].

Since δE_{aug} needs to vanish for any dv , dm , and $d\eta$, we get the necessary first-order optimality conditions for Variant I. In particular, the expression of the gradient is given by

$$(\nabla_v E_{\text{aug}}(v))_t = Lv_t + \lambda(t) \cdot \nabla m(t), \quad (56)$$

where m and λ are computed from the state and the adjoint equations

$$\partial_t m(t) + \nabla m(t) \cdot v_t = 0 \quad (57)$$

$$-\partial_t \lambda(t) - \nabla \cdot (\lambda(t) \cdot v_t) = 0 \quad (58)$$

with their corresponding initial conditions $m(0) = I_0$ and $\lambda(1) = -\frac{2}{\sigma^2}(m(1) - I_1)$.

The necessary second-order optimality conditions are obtained vanishing $\delta^2 E_{\text{aug}}$ for any dv, dm , and $d\eta$. Thus, the Gauss–Newton approximation of the Hessian vector product is given by

$$(H_v E_{\text{aug}}(v))_t \delta v(t) = L \delta v_t + \delta \lambda(t) \cdot \nabla m(t), \tag{59}$$

where $\delta \lambda$ is computed from the incremental adjoint equation

$$-\partial_t \delta \lambda(t) - \nabla \cdot (\delta \lambda(t) \cdot v_t) = 0 \tag{60}$$

with initial condition $\delta \lambda(1) = -\frac{2}{\sigma^2} \delta m(1)$, where $\delta m(1)$ is computed from the incremental state equation

$$\partial_t \delta m(t) + \nabla \delta m(t) \cdot v_t + \nabla m(t) \cdot \delta v(t) = 0 \tag{61}$$

with initial condition $\delta m(0) = 0$.

A.2 PDE-Constrained LDDMM Based on the State Equation (Variant II)

Variant II consists in replacing the computation of the state and the adjoint variables, m and λ , from the solution of the state and adjoint PDEs to the identities $m(t) = I_0 \circ \phi(t)$ and $\lambda(t) = J(t)\lambda(1) \circ \psi(t)$, where $\phi(t)$ is the direct map, $\psi(t)$ is the inverse map, and J is the Jacobian determinant of ψ . As a result, Variants I and II are two theoretically but not numerically equivalent formulations of the original PDE-LDDMM problem.

For Variant II, the derivation of the gradient and the Hessian vector product proceeds as for Variant I. However, the computation of the state and adjoint variables is performed using their identities, transferring PDE resolution to the deformation state equation for ϕ and ψ

$$\partial_t \phi(t) + D\phi(t) \cdot v_t = 0 \tag{62}$$

$$-\partial_t \psi(t) - D\psi(t) \cdot v_t = 0 \tag{63}$$

with initial condition $\phi(0) = id$ and $\psi(1) = id$, and the Jacobian equation for J

$$-\partial_t J(t) - v_t \cdot \nabla J(t) = -J(t) \nabla \cdot v_t \tag{64}$$

with initial condition $J(1) = 1$. The incremental state and adjoint variables are computed from the incremental expression of the identities

$$\delta m(t) = \nabla I_0 \circ \phi(t) \cdot \delta \phi(t) \tag{65}$$

$$\delta \lambda(t) = J(t) \nabla \lambda(1) \circ \psi(t) \cdot \delta \psi(t), \tag{66}$$

and, again, the PDE resolution is transferred to the incremental deformation state equations for $\delta \phi$ and $\delta \psi$

$$\partial_t \delta \phi(t) + D\delta \phi(t) \cdot v_t + D\phi(t) \cdot \delta v(t) = 0 \tag{67}$$

$$-\partial_t \delta \psi(t) - D\delta \psi(t) \cdot v_t - D\psi(t) \cdot \delta v(t) = 0. \tag{68}$$

A.3 PDE-Constrained LDDMM Based on the Deformation State Equation (Variant III)

For Variant III, the Lagrange multipliers are $\rho : \Omega \times [0, 1] \rightarrow \mathbb{R}^d$, associated with the deformation state equation (Eq. 7), and $\mu : \Omega \rightarrow \mathbb{R}^d$, associated with its initial condition. The augmented Lagrangian corresponds with

$$E_{\text{aug}}(v) = E(v) + \int_0^1 \langle \rho(t), \partial_t \phi(t) + D\phi(t) \cdot v_t \rangle_{L^2} dt + \langle \mu, \phi(0) - id \rangle_{L^2}. \tag{69}$$

The first- and second-order optimality conditions are derived from the formal computations of

$$\delta E_{\text{aug}}(v, \phi, \rho, \mu; dv, d\phi, d\rho, d\mu) \tag{70}$$

and

$$\delta^2 E_{\text{aug}}(v, \phi, \rho, \mu; dv, d\phi, d\rho, d\mu). \tag{71}$$

The necessary first- and second-order optimality conditions are obtained from the need to vanish δE_{aug} and $\delta^2 E_{\text{aug}}$ for any $dv, d\phi, d\rho$, and $d\mu$, yielding

$$(\nabla_v E_{\text{aug}}(v))_t = Lv_t + D\phi(t) \cdot \rho(t)$$

$$(H_v E_{\text{aug}}(v))_t \delta v(t) = L \delta v_t + D\phi(t) \cdot \delta \rho(t). \tag{72}$$

where ϕ is computed from the deformation state equation, ρ from the deformation adjoint equation, and $\delta \rho$ from the incremental deformation adjoint equation

$$\partial_t \phi(t) + D\phi(t) \cdot v_t = 0 \tag{73}$$

$$-\partial_t \rho(t) - \nabla \cdot (\rho(t) \cdot v_t) = 0 \tag{74}$$

$$-\partial_t \delta \rho(t) - \nabla \cdot (\delta \rho(t) \cdot v_t) = 0 \tag{75}$$

with initial conditions $\phi(0) = id$, $\rho(1) = \lambda(1) \nabla m(1)$, $\delta \rho(1) = \delta \lambda(1) \nabla m(1)$. It should be noticed that the divergence operator acting on tensors operates row-wise.

A.4 Semi-Lagrangian Runge–Kutta Integration

As we mentioned in Sect. 3, to be able to apply SL integration, the differential equations for different spatial variants need to be written in the shape of Eq. 46. The state equations, the deformation state equations, and their incremental

Table 8 Original PDEs involved in PDE-LDDMM and corresponding PDEs written in SL form

Equation in original form	Equation in SL form
$\partial_t m(t) + \nabla m(t) \cdot v_t = 0$ (Eq. 57)	$D_t m(t) = 0$
$-\partial_t \lambda(t) - \nabla \cdot (\lambda(t) \cdot v_t) = 0$ (Eq. 58)	$-D_t \lambda(t) = \lambda(t) \nabla \cdot v_t$
$\partial_t \phi(t) + D\phi(t) \cdot v_t = 0$ (Eq. 62)	$D_t \phi(t) = 0$
$-\partial_t \rho(t) - \nabla \cdot (\rho(t) \cdot v_t) = 0$ (Eq. 74)	$-D_t \rho(t) = \rho(t) \nabla \cdot v_t$
$\partial_t \delta m(t) + \nabla \delta m(t) \cdot v_t + \nabla m(t) \cdot \delta v(t) = 0$ (Eq. 61)	$D_t \delta m(t) = -\nabla m(t) \cdot \delta v_t$
$-\partial_t \delta \lambda(t) - \nabla \cdot (\delta \lambda(t) \cdot v_t) = 0$ (Eq. 60)	$-D_t \delta \lambda(t) = \lambda(t) \cdot \nabla \cdot v_t$
$\partial_t \delta \phi(t) + D\delta \phi(t) \cdot v_t + D\phi(t) \cdot \delta v(t) = 0$ (Eq. 67)	$D_t \delta \phi(t) = D\phi(t) \cdot \delta v(t)$
$-\partial_t \delta \rho(t) - \nabla \cdot (\delta \rho(t) \cdot v_t) = 0$ (Eq. 75)	$-D_t \delta \rho(t) = \delta \rho(t) \cdot \nabla \cdot v_t$

counterparts (Eqs. 57, 62, 61, 67) are already in the shape of Eq. 46 by just moving to the right-hand side of the equation a remaining term. For the adjoint and the incremental adjoint equations (Eqs. 58, 74, 60, 75), we use the identity

$$\nabla \cdot (u \cdot v) = u \nabla \cdot v + v \nabla u \quad (76)$$

and move the divergence term to the right-hand side the transformed equation. Table 8 gathers the expressions of the resulting differential equations, needed for the implementation of PDE-LDDMM methods in SL form. For SL-RK, the right-hand side expressions can be directly plugged into an RK differential solver. Algorithm A1 shows the pseudocode for SL-RK integration.

Algorithm A1. Semi-Lagrangian Runge-Kutta integration of PDE-LDDMM equations.

Data: u_0 or u_1 , v , X

Results: u_t solution of Equation 46.

- 1) Compute $f(u, v)$.
- 2) Compute $f \circ X$ using cubic interpolation.
- 3) Use f as the right hand side of the Runge-Kutta solver.

end

References

1. Ashburner, J., Friston, K.J.: Diffeomorphic registration using geodesic shooting and Gauss–Newton optimisation. *Neuroimage* **55**(3), 954–967 (2011)
2. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* **12**, 26–41 (2008)
3. Beg, M.F., Miller, M.I., Trounev, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. J. Comput. Vis.* **61**(2), 139–157 (2005)
4. Brunn, M., Himthani, N., Biros, G., Mehl, M.: Fast gpu 3d diffeomorphic image registration. *ArXiv* (2020)
5. Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., v.d. Smagt, P., Cremers, D., Brox, T.: FlowNet: learning optical flow with convolutional networks (2015)
6. Guo, D.X.: A Semi-Lagrangian Runge–Kutta method for time-dependent partial differential equations. *J. Appl. Anal. Comput.* **3**(3), 251–263 (2013)
7. Hart, G.L., Zach, C., Niethammer, M.: An optimal control approach for deformable registration. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR’09)* (2009)
8. Hernandez, M.: Band-limited stokes large deformation diffeomorphic metric mapping. *IEEE J. Biomed. Health Inform.* **23**(1), 362–373 (2019)
9. Hernandez, M.: A comparative study of different variants of Newton–Krylov PDE-constrained Stokes-LDDMM parameterized in the space of band-limited vector fields. *SIAM J. Imaging Sci.* **12**, 1038–1070 (2019)
10. Hernandez, M.: PDE-constrained LDDMM via geodesic shooting and inexact Gauss–Newton–Krylov optimization using the incremental adjoint Jacobi equations. *Phys. Med. Biol.* **64**(2), 025002 (2019)
11. Hernandez, M., Bossa, M.N., Olmos, S.: Registration of anatomical images using paths of diffeomorphisms parameterized with stationary vector field flows. *Int. J. Comput. Vis.* **85**(3), 291–306 (2009)
12. Klein, A., et al.: Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* **46**(3), 786–802 (2009)
13. Mang, A., Biros, G.: An inexact Newton–Krylov algorithm for constrained diffeomorphic image registration. *SIAM J. Imaging Sci.* **8**(2), 1030–1069 (2015)
14. Mang, A., Biros, G.: Constrained H1 regularization schemes for diffeomorphic image registration. *SIAM J. Imaging Sci.* **9**(3), 1154–1194 (2016)
15. Mang, A., Biros, G.: A semi-Lagrangian two-level preconditioned Newton–Krylov solver for constrained diffeomorphic image registration. *SIAM J. Sci. Comput.* **39**(6), B1064–B1101 (2017)
16. Mang, A., Gholami, A., Biros, G.: Distributed-memory large-deformation diffeomorphic 3D image registration. In: *Proceedings of ACM/IEEE Super Computing conference (SC16)* (2016)
17. Mang, A., Gholami, A., Davatzikos, C., Biros, G.: Claire: a distributed-memory solver for constrained large deformation diffeomorphic image registration. *SIAM J. Sci. Comput.* **41**(5), C548–C584 (2019)
18. Mang, A., Ruthotto, L.: A Lagrangian Gauss–Newton–Krylov solver for mass- and intensity-preserving diffeomorphic image registration. *SIAM J. Sci. Comput.* **39**(5), B860–B885 (2017)
19. Miller, M.I.: Computational anatomy: shape, growth, and atrophy comparison via diffeomorphisms. *Neuroimage* **23**, 19–33 (2004)

20. Miller, M.I., Qiu, A.: The emerging discipline of computational functional anatomy. *Neuroimage* **45**(1), 16–39 (2009)
21. Modersitzki, J.: *FAIR: Flexible Algorithms for Image Registration*. SIAM, Philadelphia (2009)
22. Polzin, T., Niethammer, M., Heinrich, M.P., Handels, H., Modersitzki, J.: Memory efficient LDDMM for lung CT. In: Proc. of the 19th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'16), Lecture Notes in Computer Science, pp. 28–36 (2014)
23. Ramon-Julvez, U., Hernandez, M., Mayordomo, E., ADNI: Analysis of the influence of diffeomorphic normalization in the prediction of stable vs progressive MCI conversion with convolutional neural networks. In: Proceedings of the 17th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'20) (2020)
24. Riishojgaard, L.P., Cohn, S.E., Li, Y., Menard, R.: The use of spline interpolation in semi-Lagrangian transport models. *Mon. Weather Rev.* **126**(7), 2008–2016 (1998)
25. Rohlfing, T.: Image similarity and tissue overlaps as surrogates for image registration accuracy: widely used but unreliable. *IEEE Trans. Med. Imaging* **31**(2), 153–163 (2012)
26. Ruijters, D., Thevenaz, P.: GPU prefilter for accurate cubic B-spline interpolation. *Comput. J.* **55**(1), 15–20 (2012)
27. Schiratti, J.B., Allasonniere, S., Colliot, O., Durrleman, S.: Learning spatiotemporal trajectories from manifold-valued longitudinal data. *Adv. Neural Inf. Process. Syst.* **28**, 2404–2412 (2015)
28. Shen, Z., Vialard, F.X., Niethammer, M.: Region-specific diffeomorphic metric mapping. In: Advances in Neural Information Processing Systems (NIPS 2019) (2019)
29. Sotiras, A., Davatzikos, C., Paragios, N.: Deformable medical image registration: a survey. *IEEE Trans. Med. Imaging* **32**(7), 1153–1190 (2013)
30. Staniforth, A., Cote, J.: Semi-Lagrangian integration schemes for atmospheric models—a review. *Mon. Weather Rev.* **119**, 2206–2223 (1991)
31. Thompson, D.W.: *On Growth and Form*. Cambridge University Press, Cambridge (1917)
32. 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)
33. Younes, L.: Jacobi fields in groups of diffeomorphisms and applications. *Q. Appl. Math.* **65**, 113–134 (2007)
34. Younes, L.: *Shapes and Diffeomorphisms*. Springer, Berlin (2010)
35. Zhang, M., Fletcher, P.T.: Finite-dimensional Lie algebras for fast diffeomorphic image registration. In: Proceedings of International Conference on Information Processing and Medical Imaging (IPMI'15), Lecture Notes in Computer Science (2015)
36. Zhang, M., Fletcher, T.: Fast diffeomorphic image registration via Fourier-Approximated Lie algebras. *Int. J. Comput. Vis.* **127**, 61–73 (2018)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



generative diseases.

Monica Hernandez received the B.Sc. and M.Sc. degree in mathematics from the University of Zaragoza in 2000. In 2008, she received the Ph.D. degree in computer science in the University of Zaragoza. Since 2011, she is associate professor in the Department of Informatics and Systems Engineering. Her current research interests include diffeomorphic registration, Computational Anatomy applications, and deep and interpretable machine learning toward the early diagnosis of neurode-