CSGB CENTRE FOR **STOCHASTIC GEOMETRY** AND ADVANCED **BIOIMAGING**

Henry O. Jacobs and Stefan Sommer

Higher-order Spatial Accuracy in Diffeomorphic Image Registration

No. 02, January 2015

Higher-order Spatial Accuracy in Diffeomorphic Image Registration

Henry O. Jacobs[1] and Stefan Sommer[2]

[1]Department of Mathematics, Imperial College, h.jacobs@imperial.ac.uk
[2]Department of Computer Science, University of Copenhagen, sommer@di.ku.dk

Abstract

We discretize a cost functional for image registration problems by deriving Taylor expansions for the matching term. Minima of the discretized cost functionals can be computed with no spatial discretization error, and the optimal solutions are equivalent to minimal energy curves in the space of k-jets. We show that the solutions convergence to optimal solutions of the original cost functional as the number of particles increases with a convergence rate of $O(h^{d+k})$ where h is a resolution parameter. The effect of this approach over traditional particle methods is illustrated on synthetic examples and real images.

1 Introduction

The goal of image registration is to place differing images of the same object (e.g. MRI scans) into a shared coordinate system so that they may be compared. One common means of doing this is to deform one image until it matches the other. Typical numerical schemes for implementing this task are particle methods, where particles are used as a finite dimensional representation of a diffeomorphism. If the particles are initialized on a regular grid of resolution h, then the solutions can be $O(h^d)$ accurate at best where d is the dimension of the image domain. Improving this order of accuracy is non-trivial because traditional higher-order numerical schemes are designed on fixed meshes (e.g. higher order finite differences).

In this paper we seek to improve this order of accuracy by considering a more sophisticated class of particles. We will find that by equipping the particles with jet-data, one can achieve registrations with higher orders of accuracy. One impact of the use of higher-order particles is that the improved accuracy per particle permits the use of fewer particles for a desired total accuracy.

1.1 Organization of the paper

We will introduce the higher-order accurate image registration framework through the following steps:

1. We will introduce the hierarchy of jet-particles.
2. We will pose an image registration problem as an optimal control problem on an infinite dimensional space.
3. We will pose a sequence of deformed problems which are easier to solve.
4. We will reduce the deformed optimization problems to optimization problems involving computation of finite dimensional ODEs (i.e. an infinite dimensional reduction).
5. We will prove that the sequence of computed solutions to the deformed problems converges to the solution of the original problem at a rate $O(h^{d+k})$, where $k \geq 0$ depends on the order of the jet-particles used.

Finally, we will display the results of numerical experiments comparing the use of 0-th, 1-st, and 2-nd order jet-particles.

2 Previous work

In this section we attempt a sparse and incomplete overview of large deformation diffeomorphic metric mapping (LDDMM) from its origins in the 1990s, to its recent marriage with geometric mechanics (2000s-present).

2.1 Matching with LDDMM

The notion of seeking deformations for the sake of image registration goes back a long way, see [SDP13, You10] and references therein. One of the first attempts was to consider diffeomorphisms of the form $\varphi(x) = x + f(x)$ for some map $f : \mathbb{R}^d \to \mathbb{R}^d$. When f is "small", φ is a diffeomorphism, but this can fail when f is "large" [You10, Chapter 7].

This breakdown for large f is a result of the fact that the space of diffeomorphisms is a nonlinear space. One of the early obstacles in diffeomorphic image registration entailed dealing with this nonlinearity. A key insight in getting a handle on the nonlinearity of the diffeomorphisms was to consider the linear space of vector fields. Given a time-dependent vector field $v(t)$, one can integrate it to obtain a diffeomorphism φ_t, which is called the *flow of* v [CRM96]. This insight was used to obtain diffeomorphisms for imaging applications by posing an optimal control problem on the space of vector-fields, and then integrating the flow of the optimal vector field to obtain a diffeomorphism. The well-posedness of this approach was studied in [Tro95, DGM98], where the cost functional (i.e. the norm) was identified as a fundamental choice in ensuring well-posedness and controlling properties of the resulting diffeomorphisms. A particle method based upon [DGM98] was implemented for the purpose of medical imaging in [JM00]. The completeness of the Euler-Lagrange equations in [DGM98] was studied thoroughly in [TY05], where the image data was allowed to be of a fairly general type (i.e. any entity upon which diffeomorphism act smoothly). The analytic safe-guards provided by [DGM98] and [TY05] where then utilized in [BMTY05], where a number of examples where numerically investigated.

2.2 Connections with geometric mechanics

Very soon after these early investigations, connections with geometric mechanics began to form. The cost functional chosen in [JM00] was the H^1-norm of the vector-fields. Coincedentally, this is the cost functional the of the n-dimensional Camassa-Holm equation (see for example [HM05] and references therein). In 1-dimension, the particle solutions in [JM00] are identical to the peakon solutions discovered in [CH93], and the numerical scheme reduces to that of [HR06]. The convergence of [JM00] was proven using geometric techniques in [CDTM12]. As images appear as advected quantities, the use of momentum maps became a useful conceptual technique for geometers to understand the numerical scheme of [BMTY05]. The identification of numerous mathematical terms in [BMTY05] as momentum maps was performed in [BGBHR11].

2.3 Jet particles

The particle method implemented in [JM00] allowed only for deformations whose Jacobian was an indentity matrix at each of the particle locations. These deformations can be thought of as "local translations" (See figure 1 (a).). Motivated by a desire to create more general deformations [SNDP13] introduced a hierarchy of particles which advect jet-data. We call the particles *jet-particles* in this paper. The first order jet-particles modify the Jacobian matrix at the particle locations and allow for "locally linear" transformations such as local scalings and local rotations (See figures 1(b-e).). Second order jet-particles allow for deformations which are "locally quadratic" (i.e. transformations with nontrivial Hessians. See figures 1(f-h).). The geometric and hierarchal structure of [SNDP13] was investigated in [Jac13] where the Lie groupoid structure of jet-particles was linked to the Lie group structure of the diffeomorphism group, thus making the case for jet-particles as multi-scale representations of diffeomorphisms. Independently, an incompressible version of this idea was invented for the purpose of incompressible fluid modelling in [DJR13]. Solutions to this fluid model were numerically computed in [CHJM14] based upon the regularized Euler fluid equations developed in [MM13b] and expressions for matrix-valued reproducing kernels derived in [MG14]. The final section of [CHJM14] provides formulas which illustrate how jet-particles in the kth level of the hierarchy yield deformations which are approximated by particles in the $(k-1)$th level of the hierarchy. The approximation being accurate to an order $O(h^k)$ where $h > 0$ is some measure of particle spacing. This approximation is more or less equivalent to the approximation of a partial differential operator by a finite difference, and it will serve as one of the main tools used in this paper in producing higher-order accurate numerical schemes.

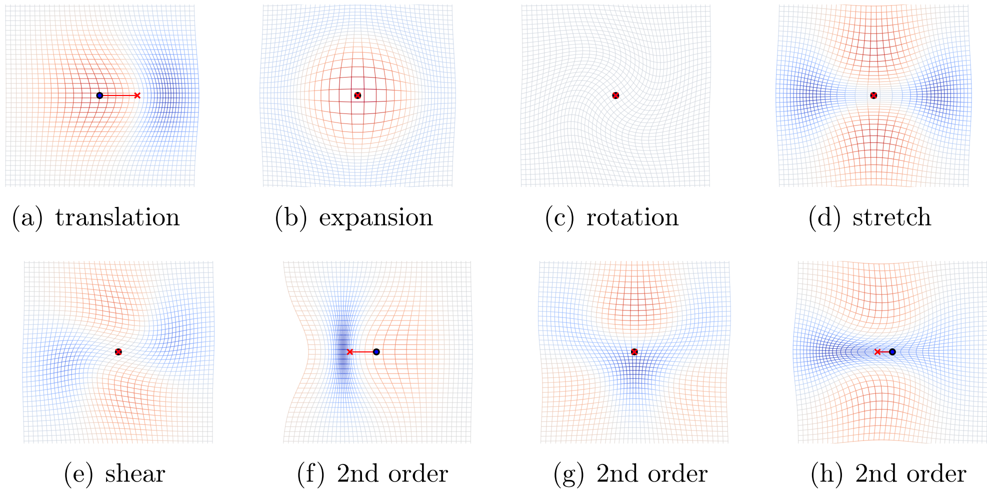


Figure 1: Deformations of initially square grids. (a) 0-th order, (b-e) 1-st order, (f-h) 2-nd order. A single jet-particle is located at the blue dots before moving with the flows to the red crosses. Grids are colored by log-Jacobian determinant.

3 LDDMM

Let M be a manifold and let $V \subset \mathfrak{X}(M)$ be a subspace of the vector-fields on M equipped with an inner-product $\langle \cdot, \cdot \rangle_V : V \times V \to \mathbb{R}$. Let $G_V \subset \mathrm{Diff}(\mathbb{R}^n)$ be the corresponding topological Lie group to which V integrates [You10, Chapter 8]. To do image registration we try to assemble "small" diffeomorphism by minimizing a cost function on the space of curve in G_V. The standard cost function takes a time-dependent diffeomorphism, φ_t, and outputs a real number. Mathematically, the cost function is often take to be a map $E_{G_V} : C^1([0,1] : G_V) \to \mathbb{R}$ given by

$$E_{G_V}[\varphi(\cdot)] := \frac{1}{2} \int_0^1 \ell(v(t))dt + F(\varphi_1),$$

where $v(t) \in V$ is the Eulerian velocity field $v(t,x) = \partial_t \varphi_t(\varphi_t^{-1}(x))$ and ℓ is a "control-cost". Explicitly, $\varphi_t \in G_V$ is obtained from $v(t) \in V$ via the initial value problem

$$\begin{cases} \frac{d}{dt}\varphi_t = v(t) \circ \varphi_t, \\ \varphi_0 = id. \end{cases} \tag{3.1}$$

One then obtains extremizers of E_{G_V} by solving the Euler-Lagrange equations on G_V. However, G_V is a non-commutative group, and can be very difficult to work with. It is typical to express E_{G_V} as a cost function on the vector-space V and incorporate (3.1) as a constraint. This means optimizing a cost function $E : C^1([0,1], V) \to \mathbb{R}$ with respect to constrained variations. Any extremizer, $v(\cdot)$, of E must necessarily satisfy a symmetry reduced form of the Euler-Lagrange equations, known as the Euler-Poincaré equation. In essence, the Euler-Poincaré equations are nothing but the Euler-Lagrange equations pulled to the space V. For a generic ℓ, the Euler-

Poincaré equations take the form

$$\frac{d}{dt}\left(\frac{\delta \ell}{\delta v}\right) + \mathrm{ad}_v^*\left(\frac{\delta \ell}{\delta v}\right) = 0. \tag{3.2}$$

We suggest [MR99] for further information on the Euler-Poincaré equations. Equation (3.2) is an evolution equation, which allows us to search over the space of initial conditions (i.e. V) in place of optimizing over space of curves (i.e. $C^1([0,1];V)$). Explicitly, this is done by considering the map $\mathrm{evol}_{EP} : V \to C^1([0,1];V)$ which sends each $v_0 \in V$ to the curve $v(\cdot) \in C^1([0,1];V)$ obtained by integrating (3.2) with initial condition v_0. We can then pre-compose E with evol_{EP} to produce the function

$$e := E \circ \mathrm{evol}_{EP} : V \to \mathbb{R}.$$

The initial condition $v^* \in V$ minimizes e if and only the solution $v(\cdot) = \mathrm{evol}_{EP}(v^*)$ of (3.2) minimizes E. Generally, solutions to (3.2) are extremizers of E, and one must appeal to higher-order variations in order to obtain sufficient conditions for an extremizer to be a minimizer. However, we will not do pursue these matters in this article.

3.1 Overview of the problem and our solution

Particle methods are typically used to approximate a diffeomorphism in the following way. We usually compute all quantities with respect to an initial condition where all the particles lie on a grid/mesh and prove convergence as the mesh width, h, tends to 0. However, it would be nice to have an order of accuracy as well.

PROBLEM: Can we solve for a minimizer of E with a convergence rate of $O(h^p)$ for some $p \in \mathbb{N}$?

Our strategy for tackling this problem is to approximate E with a sequence of $O(h^p)$-accurate curve energies E_h for which we can compute the minimizers exactly up to time discretization (i.e. the computed solutions have no spatial discretization error). More specifically, the meshsize will determine a continous sequence of subgroups $G_h \subset G$. We will approximate the matching functional F, with a G_h-invariant functional $F_h : G_V \to \mathbb{R}$ such that for a fixed $\varphi \in G_V$

$$F(\varphi) - F_h(\varphi) = O(h^p)$$

for some $p \in \mathbb{N}$. We find the curve energies to be $O(h^p)$ accurate as well, and this accuracy will transfer to the solutions for a sufficiently wide range of scenarios.

4 Reduction theory

In this section we review subgroup reduction of a class of optimization problems using Clebsch variables. In the Hamiltonian context, Clebsch variables are also called symplectic variables, and constitute a Poisson map $\psi : T^*\mathbb{R}^n \to P$. This is useful when $2n < \dim(P)$, since solutions to certain Hamiltonian equations on P can be derived by solving Hamiltonian equations on $T^*\mathbb{R}^n$ first [MW83, Wei83]. The Lagrangian version of this idea was further developed in the context of equations with hydrodynamic background in [HM05]. It is this later perspective which we shall take in this paper, since problem setup is stated in Lagrangian form.

Let G be a group and $G_s \subset G$ be a subgroup. We will denote the homogenous space of right cosets by $Q = G/G_s$, and we will denote the corresponding principal bundle projection by $\pi : G \to Q$. Note that G naturally acts on Q through the formula $g \cdot \pi(\tilde{g}) = \pi(g \cdot \tilde{g})$. Given this action, the corresponding (left) momentum map, $J : T^*Q \to \mathfrak{g}^*$, is defined by the condtio

$$\langle J(q,p), \xi \rangle = \langle p, \xi \cdot q \rangle \qquad \text{for all } \xi \in \mathfrak{g}.$$

Let $L : TG \to \mathbb{R}$ be the Lagrangian and let $F : G \to \mathbb{R}$. We wish to minimize the curve energy or "action"

$$E[g(\cdot)] = \int_0^1 L(g(t), \dot{g}(t)) dt + F(g(1)) \tag{4.1}$$

over the space of curves $g(t) \in G$ on the interval $[0,1]$ with $g(0) = id$. That is to say $E : C^1_{id}([0,1]; G) \to \mathbb{R}$ where $C^1_{id}([0,1]; G)$ denotes the space of C^1 curves in G originating from the identity. Extremization of E means taking a variation in $C^1_{id}([0,1]; G)$, which is a variation of a curve with a fixed end-point at $t = 0$ but *not* at $t = 1$. It is simple to show that any solution must satisfy the boundary value problem

$$\begin{cases} \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{g}}\right) - \frac{\partial L}{\partial g} = 0 \\ g(0) = id, \quad \left.\frac{\partial L}{\partial \dot{g}}\right|_{t=1} + dF(g(1)) = 0. \end{cases} \tag{4.2}$$

If the dimension of G is large, integrating this equation can be troublesome. However, in the presence of a G_s-symmetry a reduction can be applied to reduce the problem to a boundary value problem on Q.

Throughout this section we will assume that F is G_s invariant. As a result there exists a function $f : Q \to \mathbb{R}$ defined by the condition

$$f(q) = F(g) \qquad \text{for all } q \in Q \text{ , } g \in G \text{ such that } q = \pi(g).$$

More succinctly, $f = F \circ \pi$. We will also assume that $L(g, \dot{g})$ is G-invariant, and comes from a reduced Lagrangian function $\ell : \mathfrak{g} \to \mathbb{R}$. Finally, we will assume that the Legendre transformation, $\frac{\delta \ell}{\delta \xi} : \mathfrak{g} \to \mathfrak{g}^*$, is invertible. The reduced Hamiltonian $h : \mathfrak{g}^* \to \mathbb{R}$ is then given by

$$h(\mu) = \left\langle \mu, \frac{\delta \ell}{\delta \xi}^{-1}(\mu) \right\rangle - \ell\left(\frac{\delta \ell}{\delta \xi}^{-1}(\mu) \right).$$

Theorem 4.1 (see also [BGBHR11])**.** *Let $H = h \circ J : T^*Q \to \mathbb{R}$. If the curve $(q,p)(t) \in T^*Q$ satisfies*

$$\begin{cases} \dot{q} = \frac{\partial H}{\partial p}, \quad \dot{p} = -\frac{\partial H}{\partial q} \\ q(0) = \pi(id), \quad p(1) + df(q(1)) = 0. \end{cases} \tag{4.3}$$

then the curve $g(t)$ obtained by integrating the initial value problem

$$\dot{g}(t) = \xi(t) \cdot g(t), \quad \xi = (\delta\ell/\delta\xi)^{-1}(J(q,p)), \quad g(0) = id$$

satisfies (4.2)*.*

Proof. We can replace E with the (equivalent) curve energy $E_2 : C([0,1]; \mathfrak{g}) \to \mathbb{R}$ given by

$$E_2[\xi] = \int_0^1 \ell(\xi(t))dt + F(g(1)) \tag{4.4}$$

where $g(1) \in G$ is implicitly obtained through the reconstruction equation $\frac{dg}{dt} = \xi \cdot g$ which we view as a constraint. Minimizers of E_2 are related to minimizers of E through the reconstruction equation as well.

We are now going to use the G_s symmetry of (4.4) to reduce the dimensionality of the problem. The G_s invariance of F implies the existence of a function $f : Q \to \mathbb{R}$ such that $F = f \circ \pi$. Therefore, we may equivalently express E_2 as the energy functional

$$E_2[\xi] = \int_0^1 \ell(\xi(t))dt + f(q(1)) \tag{4.5}$$

where $q(1)$ is obtained through the reconstruction equation $\dot{q}(t) = \xi(t) \cdot q(t)$ with the initial condition $q(0) = \pi_s(id)$. Again, the dynamic constraint $\dot{q} = \xi \cdot q$ makes this a constrained optimization problem. We may take the dual of this constrained optimization problem by using Lagrange multipliers to get an equivalent unconstrained optimization problem [BV04]. In our case, the dual problem is that of extremizing the (unconstrained) curve energy $E_3 : C^1([0,1]; \mathfrak{g} \times T^*Q) \to \mathbb{R}$ given by

$$E_3[\xi, q, p] = \int_0^1 \ell(\xi(t)) + \langle p(t), \dot{q}(t) - \xi(t) \cdot q(t) \rangle dt + f(q(1)).$$

Using the definition of J we can re-write this as

$$E_3[\xi, q, p] = \int_0^1 \ell(\xi) + \langle p, \dot{q} \rangle - \langle J(q,p), \xi \rangle dt + f(q(1)).$$

We find that stationarity with respect to arbitrary variations of ξ implies

$$\frac{\delta\ell}{\delta\xi} = J(q(t), p(t)). \tag{4.6}$$

We may view (4.6) as a constraint which defines ξ in terms of the q's and p's. Explicitly, (4.6) tell us

$$\xi = \frac{\delta\ell}{\delta\xi}^{-1}(J(q,p)).$$

We can substitute this into the previous curve energy to eliminate the variable ξ and express E_3 solely in terms if p and q. We thus obtain the curve energy

$$E_4[q,p] = \int_0^1 \ell\left(\frac{\delta\ell}{\delta\xi}^{-1}(J(q,p))\right) + \left\langle p, \dot{q} \rangle - \langle J(q,p), \frac{\delta\ell}{\delta\xi}^{-1}(J(q,p)) \right\rangle dt$$
$$+ f(q(1)).$$

Observing that

$$H(q,p) = h(J(q,p)) = \left\langle J(q,p), \frac{\delta \ell}{\delta \xi}^{-1}(J(q,p)) \right\rangle - \ell\left(\frac{\delta \ell}{\delta \xi}^{-1}(J(q,p)) \right)$$

we can write E_4 as

$$E_4[q,p] = \int_0^1 \langle p, \dot{q} \rangle - H(q,p)dt + f(q(1)).$$

By taking arbitrary variations of q and p we find that extremization of E_4 implies the desired result. □

Theorem 4.1 allows us to minimize curve energies using the following gradient descent algorithm.

Algorithm for general Lie groups

1. Solve for $(q(t), p(t)) \in T^*Q$ in (4.3).
2. Set $\xi(t) = (\delta\ell/\delta\xi)^{-1} \cdot J(q(t), p(t))$
3. Obtain $g(t) \in G$ as a solution to the initial value problem, $\dot{g} = \xi \cdot g$, $g(0) = id$.
4. Evaluate cost function, and backward compute the adjoint equations [Son98] to compute the gradient of the cost function with respect to a new initial condition.
5. If the gradient is below some tolerance, ϵ, then stop. Otherwise use the gradient to create a new initial condition and return to step 1.

The resulting curve $g(t) \in G$ will minimize the original curve energy given in equation (4.1). Moreover, all the minimizers of the original problem are obtained in this way. Again, the advantage of this method is that the bulk of the computation is performed on the lower dimensional space T^*Q rather than TG.

In the next sections we will consider the case where G is a diffeomorphism group, and G_s is a subgroup such that $Q = G/G_s$ is the (finite-dimensional) space of jet-particles.

5 Jets as Homogenous spaces

In order to invoke the findings of the previous section, we must find a way to characterize the space of jet-particles as a homogenous space (i.e. a group modulo a subgroup). This is the content of proposition 5.1, the main result in this section.

Let $\Lambda \subset M$ be a finite set of distinct points in M. If f is any k-differentiable map from a neighborhood of Λ, the k-jet of f is denoted $\mathcal{J}_\Lambda^{(k)}(f)$, and in coordinates is represented by the coefficients of the kth order Taylor expansions of f about each of the points in Λ. We call $\mathcal{J}_\Lambda^{(k)}$ the "k-th order Jet functor about Λ". This is indeed a functor, and can be applied to any k-differentiable map from subsets of M which contain Λ, including real valued functions, diffeomorphisms, and curves supported on Λ [KMS99, Chapter IV].

Let $G = \mathrm{Diff}(M)$ and let $e \in G$ denote the identity transformation on M. We can consider the subgroup

$$G_\Lambda^{(0)} := \{\psi \in G \mid \psi(x) = x \qquad \forall x \in \Lambda\}$$

and the normal subgroups

$$G_\Lambda^{(k)} := \{\psi \in G_\Lambda^{(0)} \mid \mathcal{J}_\Lambda^{(k)}\psi = \mathcal{J}_\Lambda^{(k)}e\}$$

Moreover, the Lie algebra of $G_\Lambda^{(k)}$ is

$$\mathfrak{g}_\Lambda^{(k)} = \{\eta_\Lambda^{(k)} \in \mathfrak{X}(M) \mid \mathcal{J}_\Lambda^{(k)}\eta_\Lambda^{(k)} = \mathcal{J}_\Lambda^{(k)}(0)\}.$$

In other words, $\mathfrak{g}_\Lambda^{(k)}$ is the sub-algebra of $\mathfrak{X}(M)$ consisting of vector fields with vanishing partial derivatives up to order k at the points of Λ.

Proposition 5.1. *The functor, $\mathcal{J}_\Lambda^{(k)}$, is the principal bundle projection from G to $Q^{(k)} = G/G_\Lambda^{(k)}$.*

Proof. This is merely the definition of $\mathcal{J}_\Lambda^{(k)}$, and a more thorough description of this statement can be found in [KMS93]. Nonetheless, we will attempt a skeletal proof here.

If $\varphi_2 = \varphi_1 \circ \psi$ for some $\psi \in G_\Lambda^{(k)}$ then $\mathcal{J}_\Lambda^k(\varphi_2) = \mathcal{J}_\Lambda^k(\varphi_1 \circ \psi)$. However, ψ has absolutely no impact on the kth order Taylor expansion because the Taylor expansion of ψ is trivial to kth order. Thus $\mathcal{J}_\Lambda^{(k)}(\varphi_1) = \mathcal{J}_\Lambda^{(k)}(\varphi_2)$ and so $\mathcal{J}_\Lambda^{(k)}$ is a well defined map on the coset space $Q^{(k)}$. Conversely, for each element $q \in Q^{(k)}$ one can show that the inverse image $(\mathcal{J}_\Lambda^{(k)})^{-1}(q)$ is composed of a single $G_\Lambda^{(k)}$ orbit and no more. $\square$

For example if Λ consists of only two distinct points then

$$Q^{(0)} = \{(y_1, y_2) \in M^2 \mid y_1 \neq y_2\},$$
$$Q^{(1)} = \{(f_1, f_2) \in \mathrm{Fr}(M)^2 \mid \pi_{\mathrm{Fr}}(f_1) \neq \pi_{\mathrm{Fr}}(f_2)\}.$$

where $\pi_{\mathrm{Fr}} : \mathrm{Fr}(M) \to M$ is the frame bundle of M.

Proposition 5.2. *$\mathcal{J}_\Lambda^{(k)}(G_\Lambda^{(0)})$ is a (finite dimensional) Lie group, and the functor $\mathcal{J}_\Lambda^{(k)}$ restricted to $G_\Lambda^{(0)}$ is a group homomorphism. Moreover $\mathcal{J}_\Lambda^{(k)}(G_\Lambda^{(0)})$ is a normal subgroup of $\mathcal{J}_\Lambda^{(k)}(G_\Lambda^{(l)})$ for all $l \in \mathbb{N}$.*

Corollary 5.3. *The space $Q^{(k)}$ is a (finite-dimensional) principal bundle with structure group $\mathcal{J}_\Lambda^{(k)}(G_\Lambda^{(0)})$.*

For $k = 0$ this structure group is trivial. At $k = 1$ this structure group is identifiable with $\mathrm{GL}(d)$ where $d = \dim(M)$.

6 An $O(h^p)$ accurate algorithm

In this section we describe the basic strategy for using jet-particles to get high order accuracy in solutions to LDDMM problems posed on $M = \mathbb{R}^d$. The algorithm uses uses a $O(h^p)$ approximation to the matching term which is $G_\Lambda^{(k)}$ invariant. We then invoke Theorem 4.1 to reduce the problem to a finite dimensional boundary problem on $Q^{(k)}$ which we solve to obtain an approximation of the solution to the original problem.

We will assume that the problem is defined on a reproducing kernel Hilbert space (RKHS), which we denote by $V \subset \mathfrak{X}(\mathbb{R}^n)$ and whose kernel we denote by $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ [You10, Chapter 9]. Moreover, we will assume V satisfies the admissibility condition

$$\|v\|_V \leq \|v\|_{k,\infty} \tag{6.1}$$

for all $v \in V$. We will denote the topological group which integrates V by G_V.

To make precise what we mean by "an $O(h^p)$ approximation" to a matching term we will recall the "big O" notation.

Definition 6.1. *Let $F : G_V \to \mathbb{R}$, and let $F_h : G_V \to \mathbb{R}$ depend on a parameter $h > 0$. We say that F_h is an $O(h^p)$*-approximation *to F if*

$$\lim_{h \to 0} \left(\frac{F(x) - F_h(x)}{h^p} \right) < \infty$$

for all $x \in \mathbb{R}^d$. Moreover, $O(h^p)$ will serve as a place-holder for an arbitrary function within the equivalence class of all functions of h which vanish at a rate of h^p or faster as $h \to 0$. Under this notation F_h is an $O(h^p)$ approximation of F if $F = F_h + O(h^p)$.

To illustrate how we may produce $O(h^p)$-approximations to matching functions we will consider the following example.

Example Let $I_0, I_1 \in C^k(\mathbb{R}^d; [0,1])$ be two greyscale images with compact support. We can consider the matching functional $F : \mathrm{Diff}(\mathbb{R}^d) \to \mathbb{R}$ given by

$$F(\varphi) = \frac{1}{\sigma} \|I_0 - (I_1 \circ \varphi)\|_{L_2}^2 = \frac{1}{\sigma} \int_{\mathbb{R}^d} |I_0(x) - I_1(\varphi(x))|^2 dx.$$

As I_0 and I_1 each have compact support, the integral term can be restricted to a compact domain. We will continue to write our integrals as integrations over $\mathbb{R}^d$, but we will exploit this compactification when we need to.

Consider the regular lattice $\Lambda_h = \mathbb{Z}^d h$ whereopon for sufficiently small $h > 0$ the L_2-integral can be approximated to order $O(h^d)$ with a Riemann sum

$$F_h^{(0)}(\varphi) = \sum_{x \in \Lambda_h} h^d (I_0(x) - I_1(\varphi(x)))^2$$

While the order of the set Λ_h is infinite, the sum over Λ_h used to compute F_h has only finitely many non-zero terms to consider because I_0 and I_1 have compact support. Moreover, F_h is $G_{\Lambda_h}^{(0)}$ invariant because it only depends on $\varphi(x)$ for $x \in \Lambda_h$.

An $O(h^{d+2})$ approximation is given by

$$
\begin{aligned}
F_h^{(2)}(q) = &\sum_{x\in\Lambda_h} h^d(I_0(x) - I_1(\varphi(x)))^2 \\
&+ \sum_\alpha \frac{h^{d+2}}{12}\left[(\partial_\alpha I_0(x) - \partial_\beta I_1(\varphi(x))\partial_\alpha \varphi^\beta(x))^2\right] \\
&\quad + \frac{h^{d+2}}{12}\Big[\big(I_0(x) - I_1(\varphi(x))\big)(\partial_\alpha^2 I_0(x) \\
&\quad - \partial_{\beta\gamma} I_1(\varphi(x))\partial_\alpha\varphi^\beta(x)\partial_\alpha\varphi^\gamma(x) - \partial_\gamma I_1(\varphi(x))\partial_{\alpha\alpha}\varphi^\gamma(x))\Big]
\end{aligned}
$$

and we can observe that $F_h^{(2)}$ is $G_{\Lambda_h}^{(2)}$ invariant because $F^{(2)}(\varphi)$ only depends on the 2nd order Taylor expansion of φ centered at each $x \in \Lambda_h$. $\Diamond$

Given a $G_{\Lambda_h}^{(k)}$-invariant $O(h^p)$-approximation $F_h : \mathrm{Diff}(\mathbb{R}^d) \to \mathbb{R}$ to the matching term F we may consider the alternative curve energy

$$
E_h[\varphi] = \frac{1}{2}\int_0^1 \|v(t)\|_V^2 + F_h(\varphi_1),
$$

where $v(t) \in V$ is the Eulerian velocity field $v(t,x) = \partial_t \varphi_t(\varphi_t^{-1}(x))$. For a fixed curve φ_t, we observe that E_h is an $O(h^p)$-approximation to E. One might surmise that the extremizers of E_h provide good approximations of the extremizers of E. This is important because E_h is $G_{\Lambda_h}^{(k)}$-invariant, and we can invoke Theorem 4.1 to solve for extremizers of E_h, but we can not do this for E. Fortunately, for many choices of F_h the minimizers of E_h will converge to those of E and we even have a convergence rate.

Theorem 6.2. *Let $F : G_V \to \mathbb{R}$ be C^2 with respect to some topology.*[1] *Let $F_h : G_V \to \mathbb{R}$ be C^2 and an $O(h^p)$-approximation for F. Consider the curve energies $E, E_h : C([0,1], V) \to \mathbb{R}$*

$$
\begin{aligned}
E[v(\cdot)] &= \frac{1}{2}\int_0^1 \|v(t)\|_V dt + F(\varphi_1) \\
E_h[v(\cdot)] &= \frac{1}{2}\int_0^1 \|v(t)\|_V dt + F_h(\varphi_1).
\end{aligned}
$$

where $\varphi_1 \in G_V$ is the Lie integration of $v(t)$. Let v^* minimize

$$
e = E \circ \mathrm{evol}_{EP} : V \to \mathbb{R}
$$

If the Hessian, $D^2 e$, is a bounded positive definite operator on V at v^* and the solution of (3.2) *exhibits C^2 dependents upon the initial velocity field, then there exists a minimizer v_h^* of*

$$
e_h = E_h \circ \mathrm{evol}_{EP} : V \to \mathbb{R}
$$

which is an $O(h^p)$-approximation of v^*.

[1] We will assume that there exists some regular Lie group, which is large enough to contain G_V as a set. We will then use the topology of this regular Lie group rather than the topology of G_V induced by V. For example if $M = \mathbb{R}^d$ one may consider the groups defined in [MM13a].

We will employ the following well-known result to approximate vector-fields in V with finite linear combinations of the RKHS kernel K.

Lemma 6.3. *Assume V satisfies the admissibility assumption* (6.1)*. Consider the subspace of vector-fields*

$$V_h^{(k)} = \{v \in \mathfrak{X}(\mathbb{R}^d) \mid v = \sum_{y \in \Lambda_h, |\alpha| \leq k} \alpha_y \partial_\alpha K(x-y)\}.$$

The set $W = \cup_{h>0} V_h^{(0)}$ is dense in V with respect to $\langle \cdot, \cdot \rangle_V$.

Proof. Let $\{h_j > 0\}$ be a sequence such that $\lim_{j\to\infty}(h_j) = 0$. Let $v \in V$ be orthogonal to W. Thus $\langle v, w \rangle_V = 0$ for all $w \in W$. That is to say $v(x) = 0$ for all $x \in \Lambda_{h_j}$ and all $j \in \mathbb{N}$. However, any point $y \in \mathbb{R}^n$ is the limit of a sequence $\{x_j \in \Lambda_{h_j}\}$. Since all members of V are continuous, it must be the case that $v = 0$. □

A direct corollary is that $W^{(k)} = \cup_{h>0} V_h^{(k)}$ is dense in V since $V_h^{(0)} \subset V_h^{(k)}$ for any $k \in \mathbb{N}$. Lemma 6.3 will allow us to approximate our cost functional on V.

Lemma 6.4. *Let h_j be a sequence of positive real numbers which converges to* 0*. Let $v, \delta v \in V$ and $v_j, \delta v_j \in V_{h_j}$ be such that $v_j \to v$ and $\delta v_j \to \delta v$. Then $(\frac{d}{d\epsilon}^{\ell})\big|_{\epsilon=0} e_{h_j}(v_j + \epsilon^\ell \delta v_j) \to (\frac{d}{d\epsilon}^{\ell})\big|_{\epsilon=0} e(v + \epsilon^\ell \delta v)$ and ℓ is only limited by the smoothness of E and E_{h_j}.*

Proof. We will deal first with the case where $\ell = 0$. Thus we seek to prove that $e_{h_j}(v_j)$ converges to $e(v)$. Lemma 6.3 implies that such sequences of v_j's exists and (6.1) implies that v_j converges to v with respect to $\|\cdot\|_{k,\infty}$. Let $\varphi^t_{h_j}$ denote the diffeomorphisms obtained by integrating the solutions of algorithm 2 equations with respect to the initial condition v_{h_j}. We have that $\varphi^t_{h_j} \to \varphi^t$ as $j \to \infty$ from [You10, 11.11] because they are each optimal trajectories (i.e. solutions of the Euler-Lagrange equations). This shows convergence. The differentiability of e and e_{h_j} yeilds ℓ-th order convergence as well by the same argument. □

Corollary 6.5. *If $v^* \in V$ minimizes e, then there exists a sequence of extremizers, $v^*_{h_j} \in V^{(k)}_{h_j}$, of e_{h_j} which converges to v^*.*

Proof of Theorem 6.2. Note that $\Delta e(v) := e(v) - e_h(v) = F(\varphi) - F_h(\varphi)$. We desire to prove that Δe is a C^2 function on $V \subset \mathfrak{X}(M)$. If G is a regular Lie group which contains the same underlying set as G_V then integration of (3.2) is a smooth map [KM96] (see the footnote on page 11). As $\Delta e(v^*) = F(\varphi^*) - F_h(\varphi^*)$, and $F, F_h \in C^2(G)$ we observe that Δe is C^2 at $v^* \in V \subset \mathfrak{X}(M)$. Moreover, we know that $\Delta e(v) = F(\varphi) - F_h(\varphi) = O(h^p)$. We can discard of the "big O" notation and write

$$e(v) - e_h(v) = A(v)h^p + B(v,h)$$

where $A \in C^2(\mathfrak{X}(M))$ is independent of h and $\partial_h^k B = 0$ for $k \leq p$. By Corollary 6.5, there exists a sequence of minimizers of e_h, denoted by v_h, which are parametrized by h and converge to the extremizer v^* of e as $h \to 0$. By the Morse Lemma (suitably generalized to Hilbert Manifolds [Tro83, GM83]) there exists a smooth coordinate chart around v^*, $\Phi : U \to V$, such that $\Phi(v^*) = v^*$ and $\tilde{e}(v^* + w) =$

$\tilde{e}(v^*) + D^2_{v^*}\tilde{e}(w,w)$, where $\tilde{e} := e \circ \Phi$. If v_h^* minimizes e_h then $\tilde{v}_h^* = \Phi(v_h^*)$ minimizes $\tilde{e}_h := e_h \circ \Phi$. If we define $\tilde{w} = \tilde{v}^* - \tilde{v}_h^*$ then we observe

$$
\begin{aligned}
0 = \delta\tilde{e}_h(\tilde{v}_h^*) &= \delta\tilde{e}(\tilde{v}_h^*) + \delta\tilde{A}(\tilde{v}_h^*)h^p + \delta\tilde{B}(\tilde{v},h) \\
&= \delta\tilde{e}(v^*) + \delta^2\tilde{e}(v^*)(\tilde{w},\cdot) + \delta\tilde{A}(\tilde{v}_h^*)h^p + \delta\tilde{B}(\tilde{v},h).
\end{aligned}
$$

Moreover $\delta\tilde{e}(v^*) = 0$ because v^* extremizes $\tilde{e}$. Thus we observe

$$
\delta^2\tilde{e}(v^*)(\tilde{w},\,\cdot\,) = -\delta\tilde{A}(\tilde{v}_h^*)h^p + \delta\tilde{B}(\tilde{v},h)
$$

We can observe that the Hessian $\delta^2\tilde{e}(v^*)$ is related to the Hessian δe via pre-composition by the linear operator $D\Phi(v^*)$, which is a bounded. Thus the Hessian $\delta^2\tilde{e}(v^*)$ is a bounded operator from U into V^*. By assumption, this Hessian is non-degenerate, and thus invertible. Thus we observe $\tilde{w} = -[\delta^2\tilde{e}(v^*)]^{-1} \cdot \left(\delta\tilde{A}(\tilde{v}_h^*)h^p + \delta\tilde{B}(\tilde{v},h)\right)$. In other words, $v^* = \tilde{v}^*h + O(h^p)$. So there exists functions $C(v)$ and $D(v,h)$ such that $v^* = v_h^* + C(v)h^p + D(v,h)$ where $\partial_h^k D = 0$ for $k \le p$. Thus we find

$$
\begin{aligned}
v^* &= \Phi^{-1}(v^*) = \Phi^{-1}(\tilde{v}_h^* + C(v)h^p + D(v,h)) \\
&= \Phi^{-1}(\tilde{v}_h^*) + D\Phi^{-1}(\tilde{v}_h^*) \cdot (C(v)h^p + D(v,h)) + O(h^{2p}) \\
&= v_h^* + O(h^p).
\end{aligned}
$$
□

The assumption that the Hessian of the curve energy be non-degenerate is generally difficult to check in practice. We can still invoke this theorem in specific examples because the minimizer of

$$
E(v) = \frac{1}{2}\int_0^1 \|v(t)\|_V^2 dt \quad , v(t) = \mathrm{evol}_{EP}^t(v)
$$

is $v^* = 0$, and the Hessian is just the twice the identity on V. We can view all relevant examples as perturbations of this curve energy, and use the continuity of the Hessian operator to invoke Theorem 6.2.

Setting $G = G_V$, $Q = Q^{(k)} = G_V/G^{(k)}_{\Lambda_h}$ in algorithm 1, we obtain the special case of algorithm 1 given by

Algorithm 2:

1. Solve for $(q(t), p(t)) \in T^*Q^{(k)}$ in (4.3).
2. Set $u(t) = K * J(q(t), p(t))$.
3. Obtain $\varphi_t \in G_V$ through the reconstruction formula $\dot{\varphi}_t(x) = u(\varphi(x))$ for all $x \in M$.
4. Evaluate cost function, and backward compute the adjoint equations to compute the gradient of the cost function with respect to a new initial condition.
5. If the gradient is below some tolerance, ϵ, then stop. Otherwise use the gradient to create a new initial condition and return to step 1.

Here $J(q,p)$ is a summation of Dirac-delta distributions, and distributional derivates of Dirac-deltas. Thus the convolution $K * J(q,p)$ can be computed in closed form. For detailed equations of motion see the appendix.

The example we will be considering in this paper is where $d = 2, k = 2$. By theorem 6.2 we should be able to approximate minimizers of E with $O(h^4)$ accuracy in the V-norm.

7 Numerical Results

Here we will illustrate the deformations encoded by jet-particles of various orders. We will numerically verify theorem 6.2 by testing the $O(h^{d+k})$ convergence rate of the matching functional approximation for $k = 0, 2$, and we will show that the second order approximation $F_h^{(k)}$, $k = 2$ allows matching of second order image features. We will use simple examples to describe the different capabilities of higher order jet-particles over lower order jet-particles. We do this by illustrating structures that cannot be matched with low numbers of regular 0-th order landmarks, but can still be matched successfully with 1-st and 2-nd order jet-particles. These effects imply more precise matching of small scale features on larger images where more spatial derivatives can be leveraged.

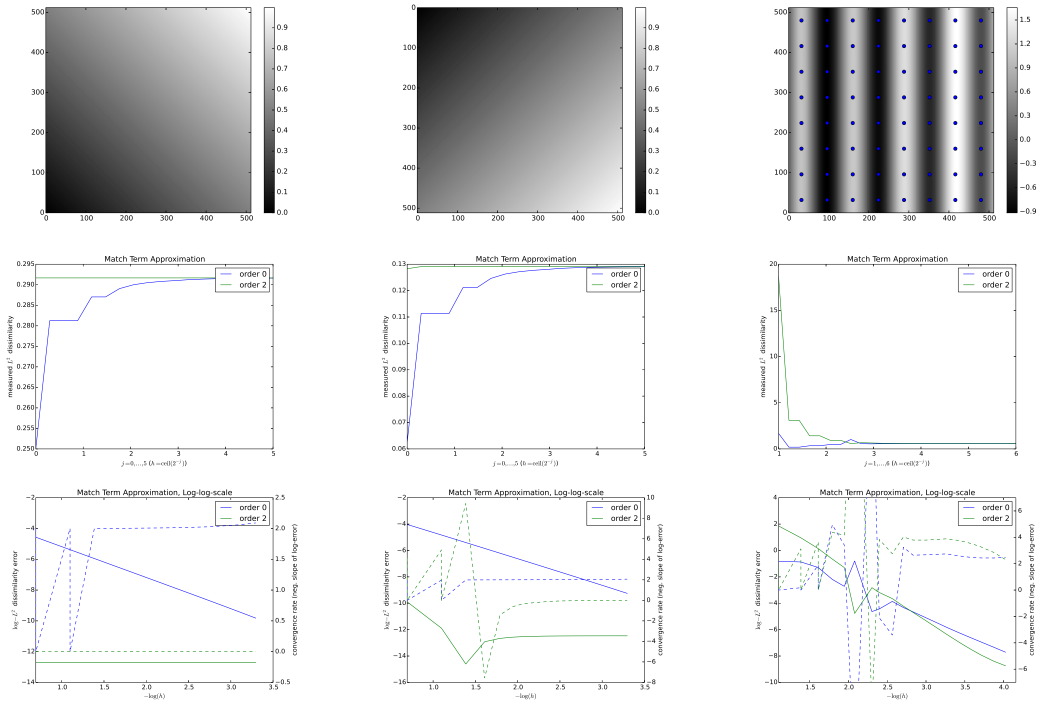
The results are obtained using the *jetflows* code available `http://www.github.com/nefan/jetflows`. The package include scripts for producing the figures displayed in this section. The flow equations are integrated forward and backward using SciPy's odeint solver (`http://scipy.org`) and the optimization is performed with a quasi-Newton BFGS optimizer. The algorithm uses isotropic Gaussian kernels. The images are pre-smoothed with a Gaussian filter, and image derivatives are computed as analytic gradients of a B-spline interpolation of the smoothed images.

7.1 Jet Deformations

Figure 1 shows the deformations encoded by 0-th, 1-st and 2-nd order jet-particles on initially square grids. Note the locally affine deformations arising from the 0-th and 1-st order jet-particles. Up to rotation of the axes, the three 1-st order examples in the figure constitute a basis of the 4 dimensional space of 1-st order jet-particles with fixed lower-order components. Likewise, up to rotation, the three 2-nd order examples constitute a basis for the 6 dimensional space of 2-nd order jet-particles.

7.2 Matching Functional Approximation

We here illustrate and test the convergence rate of the matching functional approximations. In figure 2 (page 15), the approximations $F_h^{(p)}$ are compared for $p = 0, 2$ and varying grid sizes on three synthetic images supported on the unit square. The first two images (a,b) are generated by first and second order polynomials, respectively, while the last image (c) is generated by a trigonometric function and it can therefore only be approximated by a truncated Taylor expansion. The second order approximation $F_h^{(2)}$ models F locally with a second order polynomial and it is thus expected that the error should vanish on the images (a,b). As the mesh width h



(a) $f(x, y) = x + y$ (b) $f(x, y) = (x + y)^2$ (c) $f(x, y) = \sin(6\pi x) + x^2$

Figure 2: Convergence of matching functional $F_h^{(k)}$, $k = 0, 2$. Top row: (a) linear, (b) quadratic, and (c) non-polynomial images. Lower rows, horz. axis: decreasing h (increasing nr. of sample points); vert. axis: $F_h^{(k)}$ (solid, left axis) and convergence rate (dashed, right axis). With linear and quadratic images, the error is vanishing with $k = 2$ and using only one sample point. Average convergence rates, $k = 0$: quadratic; $k = 2$: quartic as expected. (c, top row) sample points for $k = 3$ (2^3 sample points per axis, $h = 2^{-3}$).

decreases, we expect to observe $O(h^2)$ convergence rate for the zeroth order approximation $F_h^{(0)}$ on all three images. Likewise, we expect a convergence rate of $O(h^4)$ for $F_h^{(2)}$ on image (c).

In accordance with these expectations, we see the vanishing error for $F_h^{(2)}$ on (a,b) and decreasing error on (c) (lower row, solid green lines). The non-monotonic convergence seen on (c) is a result of the polynomial approximation being integrated over a compact domain. The zeroth order approximation $F_h^{(0)}$ likewise decreases with h^2 convergence rate (lower row, dashed blue lines). The convergence rate of $F_h^{(2)}$ on image (c) stabilizes at approximately h^4 until it decreases due to numerical errors introduced when the error approaches the machine precision.

7.3 Matching Simple Structures

With the following set of examples, we wish to illustrate the effects of including second order information in the matching term approximation. We visualize this using simple test images. In all examples, we will employ the approximations $F_h^{(k)}$ for $k = 0, 2$. In addition, we will match using only zeroth and first order information with a matching term that results from dropping the second order terms from $F_h^{(2)}$. While this approximation does not arise naturally from a Taylor expansion of F, it

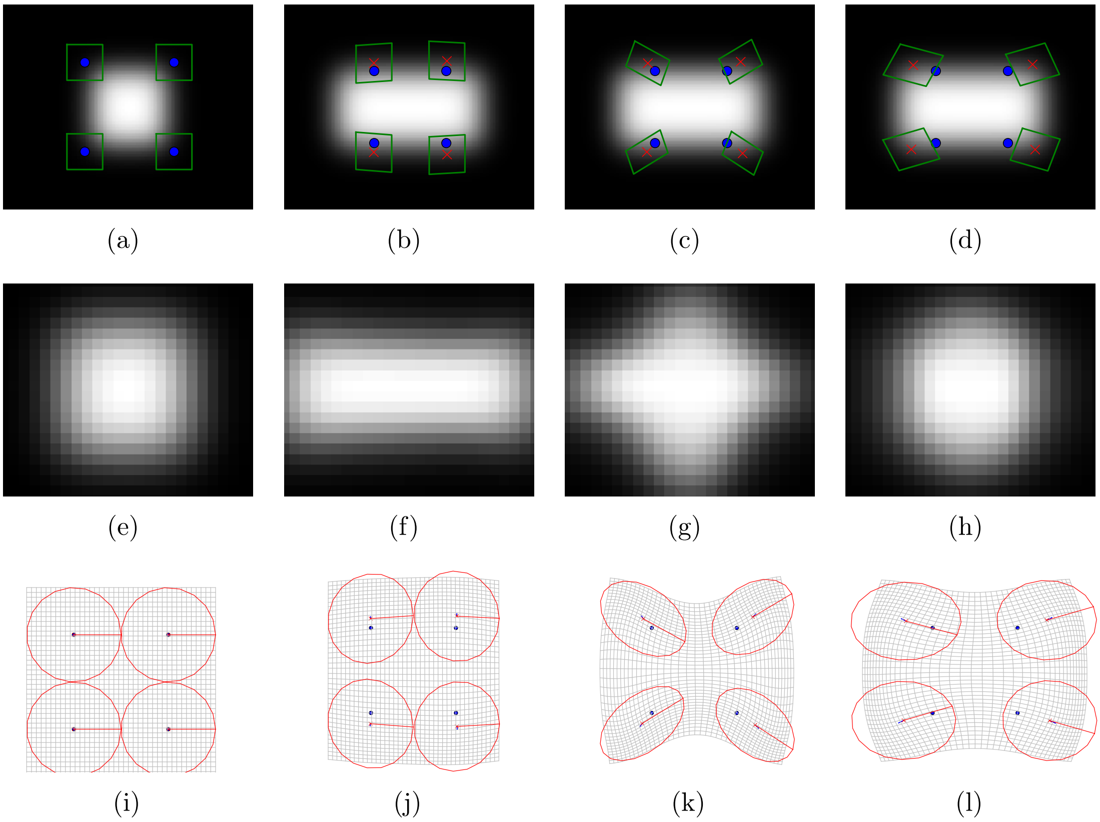


Figure 3: Matching moving images (b–d) to fixed image (a) using four jet-particles (blue points). Enlarged fixed image and moving images after warping (e–h). Corresponding deformations of an initially square grid (i–l). (b/f/j) Order 0; (c/g/k) order 1; (d/h/l) order 2. Red crosses mark location of jet-particles in moving images after matching, green boxes deformed by the warp Jacobian at the particle positions. Moving images at the red crosses should match fixed image at blue dots; second row images should match the fixed image (a/e).

allows visualization of the differences between including first and second order image information in the match.

In figure 3, a bar (moving image) is matched to a square (fixed image). The figure shows how four jet-particles move from their positions on a grid in the fixed image (a) to positions in the moving image that contain features matching the fixed image up to the order of the approximation. For zeroth order (b), only pointwise intensity is matched and the jet-particles move vertically (red crosses) resulting in only a slight deformation. With first order matching (c), the jet-particles locally rotate the domain (warp Jacobian matrices shown with green boxes) to account for the image gradient at the corners of the square. This produces a diamond-like shape. With second order (d), the corners are matched and the jet-particles move towards the corners of the moving image bar. The middle row shows the warped moving images enlarged. The second order match (h) is close to the fixed image (a) while both first and zeroth order fail to produce satisfying matches.

Figure 4 shows the result of matching images differing by an affine transformation with either one 1-st order jet or multiple 0-th order jet-particles. While three 0-th

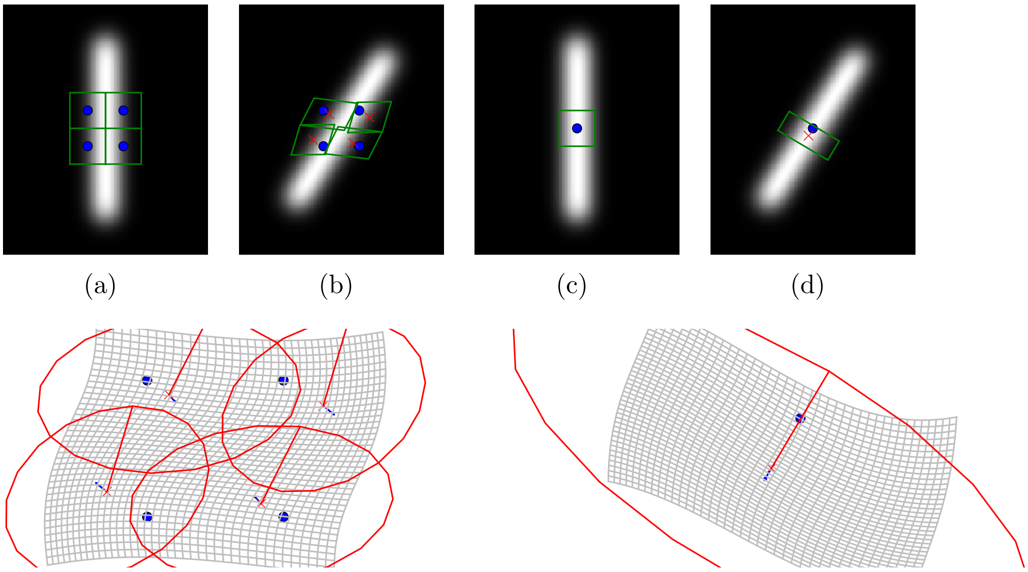


Figure 4: First order (linear/affine) deformations of an image can be matched with multiple 0-th order jet-particles (a,b) or one 1-st order jet-particle (c,d). A rotated bar (b/d) is matched to a bar (a/c). The warps that transform the moving images (b/d) to the fixed images (a/c) are applied to initially square grids in (e/f). Red circles are deformed with warp derivative at the particle positions.

order jet-particles can approximate a first order deformation in 2D, four particles are used to produce a symmetric picture. The warp Jacobians deform the initially square green boxes displayed at the jet positions. The resulting warps in both cases approximate an affine transformation.

With translation only, including second order information in the match does not change the result as illustrated in figure 5 where the match is performed on an image and a translated version of the image.

7.4 Real image data

We illustrate the effect of the increased order on real images by matching two mid-sagittal slices of 3D MRI from the MGH10 dataset[2]. In figure 6, red boxes mark the ventricle area of the brain on which the matching is performed. We perform the match with 9 jet-particles (3 per axis), 16 jet-particles (4 per axis) and 64 jet-particles (8 per axis) and $k = 0, 2$. With 9 2-nd order jet-particles (e), the moving image (d) approaches the fixed (b). A visually good match is obtained with 16 or more jet-particles. 9 and 16 0-th order jet-particles are not sufficient to correctly encode the expansion of the ventricle. With 64 0-th order jet-particles, the transformed image is close the results of the second order matches.

[2]http://www.mindboggle.info/papers/evaluation_NeuroImage2009/data/MGH10.php

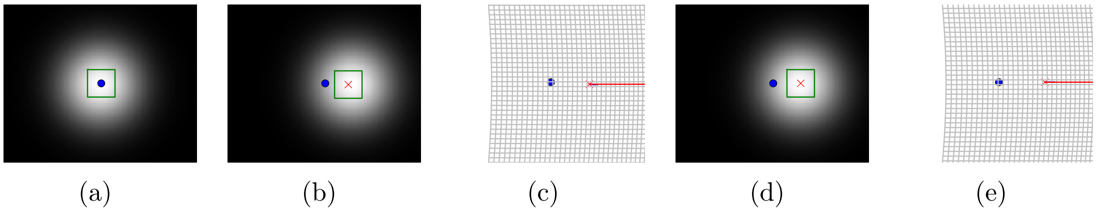


Figure 5: Without higher order features, 2nd order jet-particles do not change the match: A blob (a) is translated and matched in moving images (b,d) with red crosses marking positions of jet-particles after match. Grids (c,e) illustrate the deformations that are equivalent for 0th order (b,c) and 2nd order (d,e).

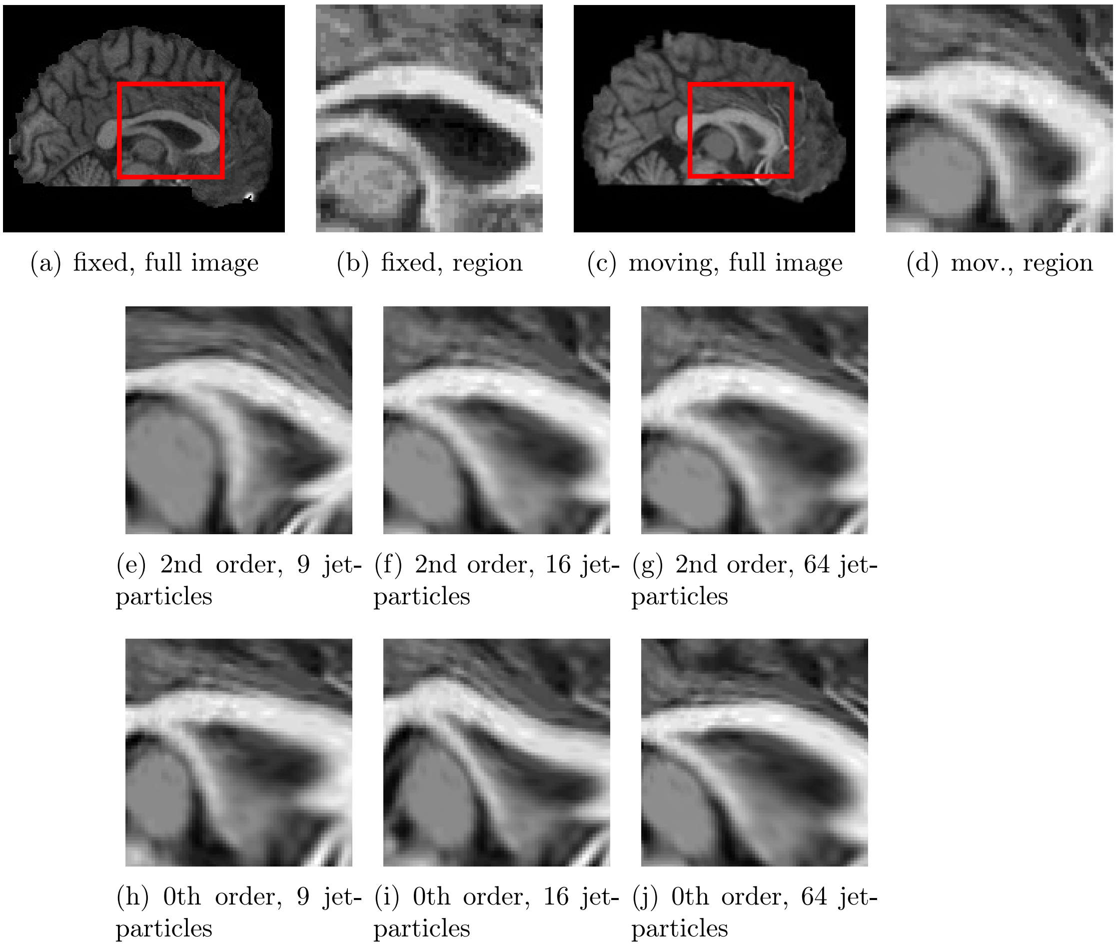


Figure 6: 2D registration of MRI slices, (a-b) fixed image, (c-d) moving image, red boxes: regions to be matched. Lower rows: matching results using 2nd order jet-particles (e-g), 0th order jet-particles (h-j). Images in lower rows should be close to (b). With 9 2nd order jet-particles (3 per axis), the moving image approaches the fixed. The match is visually good with 16 jet-particles (4 per axis). The ventricle region can equivalently be inflated with 64 0th order jet-particles.

8 Conclusion and Future Work

A priori, the LDDMM framework of image registration poses an optimization problem on the space of Diffeomorphism. Here, we introduced a family of discretized cost functions on a finite dimensional phase space that can be minimized numerically. The solutions of the discretized problem can be related to solutions of the full infinite-dimensional problem with $O(h^{d+k})$ accuracy, where h is a grid spacing and k is the order of approximation.

We provided numerical examples of deformations parametrized by 0-th, 1-st, and 2-nd order jet-particles, and we show examples of the higher order convergence of the similarity measure. The higher-order similarity measure allows matching of higher order features, and we use this fact to register various shapes and images with low numbers of jet-particles.

A C^k image requires much less information than a of C^0 images. Heuristically, the impact of this for computation is that we may use different techniques to approximate and advect smooth images with a sparse set of parameters. The higher-order accuracy schemes here constitutes a particular example of using reduction by symmetry to remove redundant information, and specialize advection to the data at hand. In this case, we reduce the dimensionality from infinite to finite for a given discretization, and we specialize the discretization to C^2 images.

While the applicability of this specialization is limited to images of sufficient regularity, the bigger point of this article is the notion of tailoring discretizations to data. This approach is applicable for reducing the dimensionality of data beyond images. For example, accurate discretizations of curves with tangents, surfaces with tangent planes, and higher-order tensors can be derived with corresponding reduction in dimensionality. The present framework thus points to a general approach for higher-order accurate discretizations of general classes of matching problems. Future work will constitute testing these areas of wider applicability.

Acknowledgments

Henry O. Jacobs is supported by the European Research Council Advanced Grant 267382 FCCA. Stefan Sommer is supported by the Danish Council for Independent Research with the project “Image Based Quantification of Anatomical Change”. The research was in addition supported by the Centre for Stochastic Geometry and Advanced Bioimaging, funded by a grant from the Villum Foundation.

A Equations of motion

The equations of motion are expressible as Hamiltonian equations with respect to a non-canonical Poisson bracket. If we denote $q^{(0)}$ simply by q and $p^{(0)}$ simply by p

then the Hamiltonian is

$$\begin{aligned}H(q,p,\mu^{(1)},\mu^{(2)}) = \tfrac{1}{2}p_{i\alpha}p_{j\beta}K^{\alpha\beta}(q_i-q_j) - p_{i\alpha}[\mu_j^{(1)}]_\beta{}^\gamma\partial_\gamma K^{\alpha\beta}(q_i-q_j)\\
+ p_{i\alpha}[\mu_j^{(2)}]_\beta{}^{\gamma\delta}\partial_{\delta\gamma}K^{\alpha\beta}(q_i-q_j) - \tfrac{1}{2}[\mu_i^{(1)}]_\alpha{}^\delta[\mu_j^{(1)}]_\beta{}^\gamma\partial_{\delta\gamma}K^{\alpha\beta}(q_i-q_j)\\
+ [\mu_i^{(1)}]_\alpha{}^\epsilon[\mu_j^{(2)}]_\beta{}^{\gamma\delta}\partial_{\epsilon\gamma\delta}K^{\alpha\beta}(q_i-q_j)\\
+ \tfrac{1}{2}[\mu_i^{(2)}]_\alpha{}^{\epsilon\phi}[\mu_j^{(2)}]_\beta{}^{\gamma\delta}\partial_{\gamma\delta\epsilon\phi}K^{\alpha\beta}(q_i-q_j)\end{aligned}$$

Where $K^{\alpha\beta}(x) = \delta^{\alpha\beta}e^{-\|x\|^2/2\sigma^2}$. Hamiton's equations are then given in short by

$$\dot{q} = \frac{\partial H}{\partial p} \tag{A.1}$$

$$\dot{p} = -\frac{\partial H}{\partial q} \tag{A.2}$$

$$\xi = \frac{\partial H}{\partial \mu} \tag{A.3}$$

$$\dot{\mu} = -\operatorname{ad}^*_\xi(\mu). \tag{A.4}$$

More explicitly, equation (A.1) is given by

$$\dot{q}_i^\alpha = p_{j\beta}K^{\alpha\beta}(q_i-q_j) - [\mu_j^{(1)}]_\beta{}^\gamma\partial_\gamma K^{\alpha\beta}(q_i-q_j) + [\mu_j^{(2)}]_\beta{}^{\gamma\delta}\partial_{\gamma\delta}K^{\alpha\beta}(q_i-q_j)$$

equation (A.2) is given by the sum

$$\dot{p}_{i\alpha} = T^{00}_{i\alpha} + T^{01}_{i\alpha} + T^{02}_{i\alpha} + T^{12}_{i\alpha} + T^{11}_{i\alpha} + T^{22}_{i\alpha}$$

Where we define the six terms in this sum as

$$\begin{aligned}T^{00}_{i\alpha} =& -p_{i\gamma}p_{j\beta}\partial_\alpha K^{\gamma\beta}(q_i-q_j),\\
T^{01}_{i\alpha} =& (p_{i\delta}[\mu_j^{(1)}]_\beta{}^\gamma - p_{j\delta}[\mu_i^{(1)}]_\beta{}^\gamma)\partial_{\gamma\alpha}K^{\delta\beta}(q_i-q_j),\\
T^{02}_{i\alpha} =& -(p_{i\epsilon}[\mu_j^{(2)}]_\beta{}^{\gamma\delta} + p_{j\epsilon}[\mu_i^{(2)}]_\beta{}^{\gamma\delta})\partial_{\gamma\delta\alpha}K^{\epsilon\beta}(q_i-q_j),\\
T^{12}_{i\alpha} =& -([\mu_i^{(1)}]_\phi{}^\epsilon[\mu_j^{(2)}]_\beta{}^{\gamma\delta} - [\mu_j^{(1)}]_\phi{}^\epsilon[\mu_i^{(2)}]_\beta{}^{\gamma\delta})\partial_{\epsilon\gamma\delta\alpha}K^{\phi\beta}(q_i-q_j),\\
T^{11}_{i\alpha} =& [\mu_i^{(1)}]_\epsilon{}^\delta[\mu_j^{(1)}]_\beta{}^\gamma\partial_{\delta\gamma\alpha}K^{\epsilon\beta}(q_i-q_j),\\
T^{22}_{i\alpha} =& -[\mu_i^{(2)}]_\zeta{}^{\epsilon\phi}[\mu_j^{(2)}]_\beta{}^{\gamma\delta}\partial_{\epsilon\delta\gamma\phi\alpha}K^{\zeta\beta}(q_i-q_j).\end{aligned}$$

Next, we calculate the quantities $\xi^{(i)} = \partial H/\partial\mu^{(i)}$ for $i = 1, 2$ of equation (A.3) to be

$$\begin{aligned}[][\xi_i^{(1)}]^\alpha{}_\beta &= p_{j,\gamma}\partial_\beta K^{\alpha\gamma}(q_i-q_j) - [\mu_j^{(1)}]_\delta{}^\gamma\partial_{\beta\gamma}K^{\alpha\delta}(q_i-q_j) + [\mu_j^{(2)}]_\epsilon{}^{\gamma\delta}\partial_{\beta\gamma\delta}K^{\alpha\epsilon}(q_i-q_j)\\
[\xi_i^{(2)}]^\alpha{}_{\beta\gamma} &= p_{j\delta}\partial_{\beta\gamma}K^{\alpha\delta}(q_i-q_j) - [\mu_j^{(1)}]_\delta{}^\epsilon\partial_{\epsilon\beta\gamma}K^{\alpha\delta}(q_i-q_j) + [\mu_j^{(2)}]_\epsilon{}^{\phi\delta}\partial_{\beta\gamma\phi\delta}K^{\alpha\epsilon}(q_i-q_j)\end{aligned}$$

which allows us to compute $\dot{\mu}^{(i)}$ in equation (A.4) as

$$\begin{aligned}[][\dot{\mu}_i^{(1)}]_\alpha{}^\beta =& [\mu_i^{(1)}]_\alpha{}^\gamma[\xi_i^{(1)}]^\beta{}_\gamma - [\mu_i^{(1)}]_\gamma{}^\beta[\xi_i^{(1)}]^\gamma{}_\alpha\\
&+ [\mu_i^{(2)}]_\alpha{}^{\delta\gamma}[\xi_i^{(2)}]^\beta{}_{\delta\gamma} - [\mu_i^{(2)}]_\delta{}^{\beta\gamma}[\xi_i^{(2)}]^\delta{}_{\alpha\gamma} - [\mu_i^{(2)}]_\delta{}^{\gamma\beta}[\xi_i^{(2)}]^\delta{}_{\gamma\alpha}\\
[\dot{\mu}_i{}^{(2)}]_\alpha{}^{\beta\gamma} =& [\mu_i^{(2)}]_\alpha{}^{\delta\gamma}[\xi_i^{(1)}]^\beta{}_\delta + [\mu_i^{(2)}]_\alpha{}^{\beta\delta}[\xi_i^{(1)}]^\gamma{}_\delta - [\mu_i^{(2)}]_\delta{}^{\beta\gamma}[\xi_i^{(1)}]^\delta{}_\alpha\end{aligned}$$

A.1 Computing $\dot{q}$ as a function of ξ

The action of ξ on q is given by $\xi \cdot q$. We set $\dot{q} = \xi \cdot q$. We've already calculated $\dot{q}^{(0)}$. We need only calculate $\dot{q}^{(1)}$ and $\dot{q}^{(2)}$. Componentwise we calculate these to be

$$\begin{aligned}
[\dot{q}^{(1)}]^{\alpha}{}_{\beta} &= [\xi^{(1)}]^{\alpha}{}_{\gamma}[q^{(1)}]^{\gamma}{}_{\beta} \\
[\dot{q}^{(2)}]^{\alpha}{}_{\beta\gamma} &= [\xi^{(2)}]^{\alpha}{}_{\delta\epsilon} \cdot [q^{(1)}]^{\delta}{}_{\beta} \cdot [q^{(1)}]^{\epsilon}{}_{\gamma} + [\xi^{(1)}]^{\alpha}{}_{\delta} \cdot [q^{(2)}]^{\delta}{}_{\beta\gamma}
\end{aligned}$$

B First variation equations

The first variation equations are equivalent to applying the tangent functor to our evolutions. We find the velocites:

$$\begin{aligned}
\frac{d}{dt}\delta q_i^{\alpha} &= \delta p_{j\beta} K^{\alpha\beta}(q_i - q_j) + p_{j\beta}(\delta q_i^{\gamma} - \delta q_j^{\gamma})\partial_{\gamma} K^{\alpha\beta}(q_i - q_j) \\
&\quad - [\delta\mu_j^{(1)}]_{\beta}{}^{\gamma}\partial_{\gamma} K^{\alpha\beta}(q_i - q_j) - [\mu_j^{(1)}]_{\beta}{}^{\gamma}\partial_{\gamma\delta} K^{\alpha\beta}(q_i - q_j)(\delta q_i^{\delta} - \delta q_j^{\delta}) \\
&\quad + [\delta\mu_j^{(2)}]_{\beta}{}^{\gamma\delta}\partial_{\gamma\delta} K^{\alpha\beta}(q_i - q_j) + [\mu_j^{(2)}]_{\beta}{}^{\gamma\delta}\partial_{\gamma\delta\epsilon} K^{\alpha\beta}(q_i - q_j)(\delta q_i^{\epsilon} - \delta q_j^{\epsilon}) \\
[\delta\xi_i^{(1)}]^{\alpha}{}_{\beta} &= \delta p_{j\gamma}\partial_{\beta} K^{\gamma\alpha}(q_i - q_j) + p_{j\gamma}(\delta q_i^{\delta} - \delta q_j^{\delta})\partial_{\delta\beta} K^{\alpha\gamma}(q_i - q_j) \\
&\quad - [\delta\mu_j^{(1)}]_{\delta}{}^{\gamma}\partial_{\beta\gamma} K^{\alpha\delta}(q_i - q_j) - [\mu_j^{(1)}]_{\delta}{}^{\gamma}(\delta q_i^{\epsilon} - \delta q_j^{\epsilon})\partial_{\beta\gamma\epsilon} K^{\alpha\delta}(q_i - q_j) \\
&\quad + [\delta\mu_j^{(2)}]_{\phi}{}^{\gamma\delta}\partial_{\beta\gamma\delta} K^{\alpha\phi}(q_i - q_j) + [\mu_j^{(2)}]_{\phi}{}^{\gamma\delta}(\delta q_i^{\epsilon} - \delta q_j^{\epsilon})\partial_{\beta\gamma\delta\epsilon} K^{\alpha\phi}(q_i - q_j) \\
[\delta\xi_i^{(2)}]^{\alpha}{}_{\beta\gamma} &= \delta p_{j\delta}\partial_{\beta\gamma} K^{\alpha\delta}(ij) + p_{j\epsilon}(\delta q_i^{\delta} - \delta q_j^{\delta})\partial_{\beta\gamma\delta} K^{\alpha\epsilon}(ij) \\
&\quad - [\delta\mu_j^{(1)}]_{\epsilon}{}^{\delta}\partial_{\beta\gamma\delta} K^{\alpha\epsilon}(ij) - [\mu_j^{(1)}]_{\phi}{}^{\delta}(\delta q_i^{\epsilon} - \delta q_j^{\epsilon})\partial_{\beta\gamma\delta\epsilon} K^{\alpha\phi}(ij) \\
&\quad + [\delta\mu_j^{(2)}]_{\phi}{}^{\delta\epsilon}\partial_{\beta\gamma\delta\epsilon} K^{\alpha\phi}(ij) + [\mu_j^{(2)}]_{\lambda}{}^{\delta\epsilon}(\delta q_i^{\phi} - \delta q_j^{\phi})\partial_{\beta\gamma\delta\epsilon\phi} K^{\alpha\lambda}(ij)
\end{aligned}$$

and the momenta:

$$\frac{d}{dt}\delta p_{i\alpha} = \delta T_{i\alpha}^{00} + \delta T_{i\alpha}^{01} + \delta T_{i\alpha}^{02} + \delta T_{i\alpha}^{12} + \delta T_{i\alpha}^{11} + \delta T_{i\alpha}^{22}.$$

The first-variation equation for $\mu^{(1)}$ is

$$\begin{aligned}
\frac{d}{dt}[\delta\mu^{(1)}]_{\alpha}{}^{\beta} &= [\delta\mu^{(1)}]_{\alpha}{}^{\gamma}[\xi^{(1)}]^{\beta}{}_{\gamma} + [\mu^{(1)}]_{\alpha}{}^{\gamma}[\delta\xi^{(1)}]^{\beta}{}_{\gamma} \\
&\quad - [\delta\mu^{(1)}]_{\gamma}{}^{\beta}[\xi^{(1)}]^{\gamma}{}_{\alpha} - [\mu^{(1)}]_{\gamma}{}^{\beta}[\delta\xi^{(1)}]^{\gamma}{}_{\alpha} \\
&\quad + [\delta\mu^{(2)}]_{\alpha}{}^{\delta\gamma}[\xi^{(2)}]^{\beta}{}_{\delta\gamma} + [\mu^{(2)}]_{\alpha}{}^{\delta\gamma}[\delta\xi^{(2)}]^{\beta}{}_{\delta\gamma} \\
&\quad - [\delta\mu^{(2)}]_{\delta}{}^{\beta\gamma}[\xi^{(2)}]^{\delta}{}_{\alpha\gamma} - [\mu^{(2)}]_{\delta}{}^{\beta\gamma}[\delta\xi^{(2)}]^{\delta}{}_{\alpha\gamma} \\
&\quad - [\delta\mu^{(2)}]_{\delta}{}^{\gamma\beta}[\xi^{(2)}]^{\delta}{}_{\gamma\alpha} - [\mu^{(2)}]_{\delta}{}^{\gamma\beta}[\delta\xi^{(2)}]^{\delta}{}_{\gamma\alpha}
\end{aligned}$$

and finally

$$\begin{aligned}
\frac{d}{dt}[\delta\mu^{(2)}]_{\alpha}{}^{\beta\gamma} &= [\delta\mu^{(2)}]_{\alpha}{}^{\delta\gamma}[\xi^{(1)}]^{\beta}{}_{\delta} + [\mu^{(2)}]_{\alpha}{}^{\delta\gamma}[\delta\xi^{(1)}]^{\beta}{}_{\delta} \\
&\quad + [\delta\mu^{(2)}]_{\alpha}{}^{\beta\delta}[\xi^{(1)}]^{\gamma}{}_{\delta} + [\mu^{(2)}]_{\alpha}{}^{\beta\delta}[\delta\xi^{(1)}]^{\gamma}{}_{\delta} \\
&\quad - [\delta\mu^{(2)}]_{\delta}{}^{\beta\gamma}[\xi^{(1)}]^{\delta}{}_{\alpha} - [\mu^{(2)}]_{\delta}{}^{\beta\gamma}[\delta\xi^{(1)}]^{\delta}{}_{\alpha}
\end{aligned}$$

where the δT's are given by

$$
\begin{aligned}
\delta T^{00}_{i\alpha} =& -\delta p_{i\gamma} p_{j\beta} \partial_\alpha K^{\gamma\beta}(q_i - q_j) - p_{i\gamma} \delta p_{j\beta} \partial_\alpha K^{\gamma\beta}(q_i - q_j) \\
& - p_{i\gamma} p_{j\beta} \partial_{\alpha\delta} K^{\gamma\beta}(q_i - q_j)(\delta q_i^\delta - \delta q_j^\delta) \\
\delta T^{01}_{i\alpha} =& -\delta p_{j\delta} [\mu_i^{(1)}]_\beta{}^\gamma \partial_{\gamma\alpha} K^{\delta\beta}(ij) - p_{j\delta} [\delta\mu_i^{(1)}]_\beta{}^\gamma \partial_{\gamma\alpha} K^{\delta\beta}(ij) \\
& - p_{j\delta} [\mu_i^{(1)}]_\beta{}^\gamma (\delta q_i^\epsilon - \delta q_j^\epsilon) \partial_{\epsilon\gamma\alpha} K^{\delta\beta}(ij) \\
& + \delta p_{i\delta} [\mu_j^{(1)}]_\beta{}^\gamma \partial_{\gamma\alpha} K^{\delta\beta}(ij) + p_{i\delta} [\delta\mu_j^{(1)}]_\beta{}^\gamma \partial_{\gamma\alpha} K^{\delta\beta}(ij) \\
& + p_{i\delta} [\mu_j^{(1)}]_\beta{}^\gamma (\delta q_i^\epsilon - \delta q_j^\epsilon) \partial_{\epsilon\gamma\alpha} K^{\delta\beta}(ij) \\
\delta T^{02}_{i\alpha} =& -\delta p_{i\epsilon} [\mu_j^{(2)}]_\beta{}^{\gamma\delta} \partial_{\gamma\delta\alpha} K^{\epsilon\beta}(ij) - p_{i\epsilon} [\delta\mu_j^{(2)}]_\beta{}^{\gamma\delta} \partial_{\gamma\delta\alpha} K^{\epsilon\beta}(ij) \\
& - p_{i\epsilon} [\mu_j^{(2)}]_\beta{}^{\gamma\delta} (\delta q_i^\phi - \delta q_j^\phi) \partial_{\gamma\delta\phi\alpha} K^{\epsilon\beta}(ij) \\
& - \delta p_{j\epsilon} [\mu_i^{(2)}]_\beta{}^{\gamma\delta} \partial_{\gamma\delta\alpha} K^{\epsilon\beta}(ij) - p_{j\epsilon} [\delta\mu_i^{(2)}]_\beta{}^{\gamma\delta} \partial_{\gamma\delta\alpha} K^{\epsilon\beta}(ij) \\
& - p_{j\epsilon} [\mu_i^{(2)}]_\beta{}^{\gamma\delta} (\delta q_i^\phi - \delta q_j^\phi) \partial_{\gamma\delta\phi\alpha} K^{\epsilon\beta}(ij) \\
\delta T^{12}_{i\alpha} =& -[\delta\mu_i^{(1)}]_\phi{}^\epsilon [\mu_j^{(2)}]_\beta{}^{\gamma\delta} \partial_{\epsilon\gamma\delta\alpha} K^{\phi\beta}(ij) \\
& - [\mu_i^{(1)}]_\phi{}^\epsilon [\delta\mu_j^{(2)}]_\beta{}^{\gamma\delta} \partial_{\epsilon\gamma\delta\alpha} K^{\phi\beta}(ij) \\
& - [\mu_i^{(1)}]_\phi{}^\epsilon [\mu_j^{(2)}]_\beta{}^{\gamma\delta} (\delta q_i^\zeta - \delta q_j^\zeta) \partial_{\zeta\epsilon\gamma\delta\alpha} K^{\phi\beta}(ij) \\
& + [\delta\mu_j^{(1)}]_\phi{}^\epsilon [\mu_i^{(2)}]_\beta{}^{\gamma\delta} \partial_{\epsilon\gamma\delta\alpha} K^{\phi\beta}(ij) \\
& + [\mu_j^{(1)}]_\phi{}^\epsilon [\delta\mu_i^{(2)}]_\beta{}^{\gamma\delta} \partial_{\epsilon\gamma\delta\alpha} K^{\phi\beta}(ij) \\
& + [\mu_j^{(1)}]_\phi{}^\epsilon [\mu_i^{(2)}]_\beta{}^{\gamma\delta} (\delta q_i^\zeta - \delta q_j^\zeta) \partial_{\zeta\epsilon\gamma\delta\alpha} K^{\phi\beta}(ij) \\
\delta T^{11}_{i\alpha} =& ([\delta\mu_i^{(1)}]_\epsilon{}^\delta [\mu_j^{(1)}]_\beta{}^\gamma + [\mu_i^{(1)}]_\epsilon{}^\delta [\delta\mu_j^{(1)}]_\beta{}^\gamma) \partial_{\delta\gamma\alpha} K^{\epsilon\beta}(ij) \\
& + [\mu_i^{(1)}]_\epsilon{}^\delta [\mu_j^{(1)}]_\beta{}^\gamma (\delta q_i^\phi - \delta q_j^\phi) \partial_{\phi\delta\gamma\alpha} K^{\epsilon\beta}(ij) \\
\delta T^{22}_{i\alpha} =& -([\delta\mu_i^{(2)}]_\zeta{}^{\epsilon\phi} [\mu_j^{(2)}]_\beta{}^{\gamma\delta} + [\mu_i^{(2)}]_\zeta{}^{\epsilon\phi} [\delta\mu_j^{(2)}]_\beta{}^{\gamma\delta}) \partial_{\epsilon\delta\gamma\phi\alpha} K^{\zeta\beta}(ij) \\
& - [\mu_i^{(2)}]_\zeta{}^{\epsilon\phi} [\mu_j^{(2)}]_\beta{}^{\gamma\delta} (\delta q_i^\lambda - \delta q_j^\lambda) \partial_{\lambda\epsilon\delta\gamma\phi\alpha} K^{\zeta\beta}(ij)
\end{aligned}
$$

Finally, we compute the variation equations for $\delta q^{(1)}$ and $\delta q^{(2)}$ to be

$$
\begin{aligned}
\delta[\dot{q}_i^{(1)}]^\alpha{}_\beta =& [\delta\xi_i^{(1)}]^\alpha{}_\gamma [q_i^{(1)}]^\gamma{}_\beta + [\xi_i^{(1)}]^\alpha{}_\gamma [\delta q_i^{(1)}]^\gamma{}_\beta \\
\delta[\dot{q}^{(2)}]^\alpha{}_{\beta\gamma} =& [\delta\xi^{(2)}]^\alpha{}_{\delta\epsilon} \cdot [q^{(1)}]^\delta{}_\beta \cdot [q^{(1)}]^\epsilon{}_\gamma + [\xi^{(2)}]^\alpha{}_{\delta\epsilon} \cdot [\delta q^{(1)}]^\delta{}_\beta \cdot [q^{(1)}]^\epsilon{}_\gamma \\
& + [\xi^{(2)}]^\alpha{}_{\delta\epsilon} \cdot [q^{(1)}]^\delta{}_\beta \cdot [\delta q^{(1)}]^\epsilon{}_\gamma + [\delta\xi^{(1)}]^\alpha{}_\delta \cdot [q^{(2)}]^\delta{}_{\beta\gamma} + [\xi^{(1)}]^\alpha{}_\delta \cdot [\delta q^{(2)}]^\delta{}_{\beta\gamma}
\end{aligned}
$$

C Computation of the adjoint equations

Given any ODE on M given by $\dot{x} = f(x)$ we may consider the equations of motion for variations, given by $\frac{d}{dt}\delta x = T_x f \cdot \delta x$. In particular, $T_x f$ is a linear operator over the point x which has a dual operator. The adjoint equations are and ODE on T^*M given by

$$\frac{d\lambda}{dt} = -T_x^* f \cdot \lambda.$$

This is useful for us in the following way. Given an integral curve, $x(t)$, and a variation in the initial condition, δx_0, we see that the quantity $\langle \lambda(t), \delta x(t) \rangle$ is constant when $\delta x(t)$ satisfies the first variation equation with initial condition δx_0 and $\lambda(t)$ satisfies the adjoint equation. In our case we are able to compute the gradient of the energy with respect to varying an initial condition in this way. More explicitly, we should be able to express $T_x f$ as a matrix $M(x)^B{}_A$ so that the first variation equations are

$$\frac{d}{dt}\delta x^A = M(x)^A{}_B \delta x^B$$

and the adjoint equations can be written as

$$\dot{\lambda}_A = -\lambda_B M(x)^B{}_A$$

where λ_A is the covector associated to the A-th coordinate and $M(x)^B{}_A$ is the coefficient for δx^A in the equation for $\frac{d}{dt}\delta^B$. More specifically, the elements of M^B_A is the partial derivative of $\delta \dot{B}$ with respect to δA. So we compute all these (36) quantities below.

$$\begin{aligned}\frac{\partial[\delta \dot{q}_i^{(0)}]^\alpha}{\partial[\delta q_j^{(0)}]^\beta} = {} & \left(p_{k\gamma}\partial_\beta K^{\alpha\gamma}(jk) - [\mu_k^{(1)}]_\delta{}^\gamma \partial_{\gamma\beta} K^{\alpha\delta}(jk) + [\mu_k^{(2)}]_\epsilon{}^{\gamma\delta}\partial_{\gamma\delta\beta}K^{\alpha\epsilon}(jk)\right)\delta_i^j \\ & - p_{j\gamma}\partial_\beta K^{\alpha\gamma}(ij) + [\mu_j^{(1)}]_\delta{}^\gamma \partial_{\gamma\beta}K^{\alpha\delta}(ij) - [\mu_j^{(2)}]_\epsilon{}^{\gamma\delta}\partial_{\gamma\delta\beta}K^{\alpha\epsilon}(ij),\end{aligned}$$

$$\frac{\partial[\delta \dot{q}_i^{(0)}]^\alpha}{\partial[\delta q_j^{(1)}]^\beta{}_\gamma} = 0, \quad \frac{\partial[\delta \dot{q}_i^{(0)}]^\alpha}{\partial[\delta q^{(2)}]^\beta{}_{\gamma\delta}} = 0,$$

$$\frac{\partial[\delta \dot{q}_i^{(0)}]^\alpha]}{\partial[\delta p_j^{(0)}]_\beta} = K^{\alpha\beta}(ij), \quad \frac{\partial[\delta \dot{q}_i^{(0)}]^\alpha}{\partial[\delta \mu_j^{(1)}]_\beta{}^\gamma} = -\partial_\gamma K^{\alpha\beta}(ij), \quad \frac{\partial[\delta \dot{q}_i^{(0)}]^\alpha}{\partial[\delta \mu_j^{(2)}]_\beta{}^{\gamma\delta}} = \partial_{\gamma\delta}K^{\alpha\beta}(ij),$$

$$\frac{\partial[\delta \dot{q}_i^{(1)}]^\alpha{}_\beta}{\partial[\delta q_j^{(0)}]^\gamma} = \frac{\partial[\delta \xi_i^{(1)}]^\alpha{}_\delta}{\partial[\delta q_j^{(0)}]^\gamma}[q_i^{(1)}]^\delta{}_\beta \quad, \frac{\partial[\delta \dot{q}_i^{(1)}]^\alpha{}_\beta}{\partial[\delta q_j^{(1)}]^\gamma{}_\delta} = [\xi_i^{(1)}]^\alpha{}_\gamma \delta_\beta^\delta \delta_i^j, \quad , \frac{\partial[\delta \dot{q}_i^{(1)}]^\alpha{}_\beta}{\partial[\delta q_j^{(2)}]^\gamma{}_{\delta\epsilon}} = 0,$$

$$\frac{\partial[\delta \dot{q}_i^{(1)}]^\alpha{}_\beta}{\partial[\delta p_j^{(0)}]_\gamma} = \frac{\partial[\delta \xi_i^{(1)}]^\alpha{}_\delta}{\partial[\delta p_j^{(0)}]_\gamma}[q_i^{(1)}]^\delta{}_\beta \quad, \frac{\partial[\delta \dot{q}_i^{(1)}]^\alpha{}_\beta}{\partial[\delta \mu_j^{(1)}]_\gamma{}^\delta} = \frac{\partial[\delta \xi_i^{(1)}]^\alpha{}_\epsilon}{\partial[\delta \mu_j^{(1)}]_\gamma{}^\delta}[q_i^{(1)}]^\epsilon{}_\beta,$$

$$\frac{\partial[\delta \dot{q}_i^{(1)}]^\alpha{}_\beta}{\partial[\delta \mu_j^{(2)}]_\gamma{}^{\delta\epsilon}} = \frac{\partial[\delta \xi_i^{(1)}]^\alpha{}_\phi}{\partial[\delta \mu_j^{(2)}]_\gamma{}^{\delta\epsilon}}[q_i^{(1)}]^\phi{}_\beta,$$

$$\begin{aligned}\frac{\partial[\delta \dot{q}_i^{(2)}]^\alpha{}_{\beta\gamma}}{\partial[\delta q_j^{(0)}]^\delta} &= \frac{\partial[\delta \xi_i^{(2)}]^\alpha{}_{\phi\epsilon}}{\partial[\delta q_j^{(0)}]^\delta}[q_i^{(1)}]^\phi{}_\beta [q_i^{(1)}]^\epsilon{}_\gamma + \frac{\partial[\delta \xi_i^{(1)}]^\alpha{}_\epsilon}{\partial[\delta q_j^{(0)}]^\delta}[q_i^{(2)}]^\epsilon{}_{\beta\gamma}, \\ \frac{\partial[\delta \dot{q}_i^{(2)}]^\alpha{}_{\beta\gamma}}{\partial[\delta q_j^{(1)}]^\delta{}_\epsilon} &= ([\xi_i^{(2)}]^\alpha{}_{\delta\phi}[q_i^{(1)}]^\phi{}_\gamma \delta_\beta^\epsilon + [\xi_i^{(2)}]^\alpha{}_{\phi\delta}[q_i^{(1)}]^\phi{}_\beta \delta_\gamma^\epsilon)\delta_i^j, \\ \frac{\partial[\delta \dot{q}_i^{(2)}]^\alpha{}_{\beta\gamma}}{\partial[\delta q_j^{(2)}]^\delta{}_{\epsilon\phi}} &= [\xi_i^{(1)}]^\alpha{}_\delta \delta_\beta^\epsilon \delta_\gamma^\phi \delta_i^j,\end{aligned}$$

$$\frac{\partial[\delta\dot{q}_i^{(2)}]^\alpha{}_{\beta\gamma}}{\partial[\delta p_j^{(0)}]_\delta} = \frac{\partial[\delta\xi_i^{(2)}]^\alpha{}_{\phi\epsilon}}{\partial[\delta p_j^{(1)}]_\delta}[q_i^{(1)}]^\phi{}_\beta[q_i^{(1)}]^\epsilon{}_\gamma + \frac{\partial[\delta\xi_i^{(1)}]^\alpha{}_\epsilon}{\partial[\delta p_j^{(0)}]_\delta}[q_i^{(2)}]^\epsilon{}_{\beta\gamma},$$
$$\frac{\partial[\delta\dot{q}_i^{(2)}]^\alpha{}_{\beta\gamma}}{\partial[\delta\mu_j^{(1)}]_\delta{}^\epsilon} = \frac{\partial[\delta\xi_i^{(2)}]^\alpha{}_{\phi\zeta}}{\partial[\delta\mu_j^{(1)}]_\delta{}^\epsilon}[q_i^{(1)}]^\phi{}_\beta[q_i^{(1)}]^\zeta{}_\gamma + \frac{\partial[\delta\xi_i^{(1)}]^\alpha{}_\phi}{\partial[\delta\mu_j^{(1)}]_\delta{}^\epsilon}[q_i^{(2)}]^\phi{}_{\beta\gamma},$$
$$\frac{\partial[\delta\dot{q}_i^{(2)}]^\alpha{}_{\beta\gamma}}{\partial[\delta\mu_j^{(2)}]_\delta{}^{\epsilon\phi}} = \frac{\partial[\delta\xi_i^{(2)}]^\alpha{}_{\zeta\lambda}}{\partial[\delta\mu_j^{(2)}]_\delta{}^{\epsilon\phi}}[q_i^{(1)}]^\zeta{}_\beta[q_i^{(1)}]^\lambda{}_\gamma + \frac{\partial[\delta\xi_i^{(1)}]^\alpha{}_\zeta}{\partial[\delta\mu_j^{(2)}]_\delta{}^{\epsilon\phi}}[q_i^{(2)}]^\zeta{}_{\beta\gamma},$$

$$\frac{\partial[\delta\dot{p}_i^{(0)}]_\alpha}{\partial[\delta q_j^{(0)}]^\beta} = \frac{\partial[\delta T_i^{00}]_\alpha}{\partial[\delta q_j^{(0)}]^\beta} + \frac{\partial[\delta T_i^{01}]_\alpha}{\partial[\delta q_j^{(0)}]^\beta} + \frac{\partial[\delta T_i^{11}]_\alpha}{\partial[\delta q_j^{(0)}]^\beta} + \frac{\partial[\delta T_i^{12}]_\alpha}{\partial[\delta q_j^{(0)}]^\beta} + \frac{\partial[\delta T_i^{02}]_\alpha}{\partial[\delta q_j^{(0)}]^\beta} + \frac{\partial[\delta T_i^{22}]_\alpha}{\partial[\delta q_j^{(0)}]^\beta},$$
$$\frac{\partial[\delta\dot{p}_i^{(0)}]_\alpha}{\partial[\delta q_j^{(1)}]^\beta{}_\gamma} = 0, \quad \frac{\partial[\delta\dot{p}_i^{(0)}]_\alpha}{\partial[\delta q_j^{(2)}]^\beta{}_{\gamma\delta}} = 0,$$

$$\frac{\partial[\delta\dot{p}_i^{(0)}]_\alpha}{\partial[\delta p_j^{(0)}]_\beta} = \frac{\partial[\delta T_i^{00}]_\alpha}{\partial[\delta p_j^{(0)}]_\beta} + \frac{\partial[\delta T_i^{01}]_\alpha}{\partial[\delta p_j^{(0)}]_\beta} + \frac{\partial[\delta T_i^{02}]_\alpha}{\partial[\delta p_j^{(0)}]_\beta},$$
$$\frac{\partial[\delta\dot{p}_i^{(0)}]_\alpha}{\partial[\delta\mu_j^{(1)}]_\beta{}^\gamma} = \frac{\partial[\delta T_i^{01}]_\alpha}{\partial[\delta\mu_j^{(1)}]_\beta{}^\gamma} + \frac{\partial[\delta T_i^{11}]_\alpha}{\partial[\delta\mu_j^{(1)}]_\beta{}^\gamma} + \frac{\partial[\delta T_i^{12}]_\alpha}{\partial[\delta\mu_j^{(1)}]_\beta{}^\gamma},$$
$$\frac{\partial[\delta\dot{p}_i^{(0)}]_\alpha}{\partial[\delta\mu_j^{(2)}]_\beta{}^{\gamma\delta}} = \frac{\partial[\delta T_i^{02}]_\alpha}{\partial[\delta\mu_j^{(2)}]_\beta{}^{\gamma\delta}} + \frac{\partial[\delta T_i^{12}]_\alpha}{\partial[\delta\mu_j^{(2)}]_\beta{}^{\gamma\delta}} + \frac{\partial[\delta T_i^{22}]_\alpha}{\partial[\delta\mu_j^{(2)}]_\beta{}^{\gamma\delta}},$$

$$\frac{\partial[\delta\dot{\mu}_i^{(1)}]_\alpha{}^\beta}{\partial[\delta q_j^{(0)}]^\gamma} = [\mu_i^{(1)}]_\alpha{}^\delta\frac{\partial[\delta\xi_i^{(1)}]^\beta{}_\delta}{\partial[\delta q_j^{(0)}]^\gamma} - [\mu_i^{(1)}]_\delta{}^\beta\frac{\partial[\delta\xi_i^{(1)}]^\delta{}_\alpha}{\partial[\delta q_j^{(0)}]^\gamma}$$
$$+ [\mu_i^{(2)}]_\alpha{}^{\delta\epsilon}\frac{\partial[\delta\xi_i^{(2)}]^\beta{}_{\delta\epsilon}}{\partial[\delta q_j^{(0)}]^\gamma} - [\mu_i^{(2)}]_\delta{}^{\beta\epsilon}\frac{\partial[\delta\xi_i^{(2)}]^\delta{}_{\alpha\epsilon}}{\partial[\delta q_j^{(0)}]^\gamma} - [\mu_i^{(2)}]_\delta{}^{\epsilon\beta}\frac{\partial[\delta\xi_i^{(2)}]^\delta{}_{\epsilon\alpha}}{\partial[\delta q_j^{(0)}]^\gamma},$$
$$\frac{\partial[\delta\dot{\mu}_i^{(1)}]_\alpha{}^\beta}{\partial[\delta q_j^{(1)}]^\gamma} = 0, \quad \frac{\partial[\delta\dot{\mu}_i^{(1)}]_\alpha{}^\beta}{\partial[\delta q_j^{(2)}]^\gamma} = 0,$$

$$\frac{\partial[\delta\dot{\mu}_i^{(1)}]_\alpha{}^\beta}{\partial[\delta p_j^{(0)}]_\gamma} = [\mu_i^{(1)}]_\alpha{}^\delta \frac{\partial[\delta\xi_i^{(1)}]^\beta{}_\delta}{\partial[\delta p_j^{(0)}]_\gamma} - [\mu_i^{(1)}]_\delta{}^\beta \frac{\partial[\delta\xi_i^{(1)}]^\delta{}_\alpha}{\partial[\delta p_j^{(0)}]_\gamma}$$

$$+ [\mu_i^{(2)}]_\alpha{}^{\delta\epsilon} \frac{\partial[\delta\xi_i^{(2)}]^\beta{}_{\delta\epsilon}}{\partial[\delta p_j^{(0)}]_\gamma} - [\mu_i^{(2)}]_\delta{}^{\beta\epsilon} \frac{\partial[\delta\xi_i^{(2)}]^\delta{}_{\alpha\epsilon}}{\partial[\delta p_j^{(0)}]_\gamma} - [\mu_i^{(2)}]_\delta{}^{\epsilon\beta} \frac{\partial[\delta\xi_i^{(2)}]^\delta{}_{\epsilon\alpha}}{\partial[\delta p_j^{(0)}]_\gamma},$$

$$\frac{\partial[\delta\dot{\mu}_i^{(1)}]_\alpha{}^\beta}{\partial[\delta\mu_j^{(1)}]_\gamma{}^\delta} = \delta_i^j\delta_\alpha^\gamma[\xi_i^{(1)}]^\beta{}_\delta + [\mu_i^{(1)}]_\alpha{}^\epsilon \frac{\partial[\delta\xi_i^{(1)}]^\beta{}_\epsilon}{\partial[\delta\mu_j^{(1)}]_\gamma{}^\delta} - \delta_i^j\delta_\delta^\beta[\xi_i^{(1)}]^\gamma{}_\alpha - [\mu_i^{(1)}]_\epsilon{}^\beta \frac{\partial[\delta\xi_i^{(1)}]^\epsilon{}_\alpha}{\partial[\delta\mu_j^{(1)}]_\gamma{}^\delta}$$

$$+ [\mu_i^{(2)}]_\alpha{}^{\epsilon\phi} \frac{\partial[\delta\xi_i^{(2)}]^\beta{}_{\epsilon\phi}}{\partial[\delta\mu_j^{(1)}]_\gamma{}^\delta} - [\mu_i^{(2)}]_\phi{}^{\beta\epsilon} \frac{\partial[\delta\xi_i^{(2)}]^\phi{}_{\alpha\epsilon}}{\partial[\delta\mu_j^{(1)}]_\gamma{}^\delta} - [\mu_i^{(2)}]_\phi{}^{\epsilon\beta} \frac{\partial[\delta\xi_i^{(2)}]^\phi{}_{\epsilon\alpha}}{\partial[\delta\mu_j^{(1)}]_\gamma{}^\delta},$$

$$\frac{\partial[\delta\dot{\mu}_i^{(1)}]_\alpha{}^\beta}{\partial[\delta\mu_j^{(2)}]_\gamma{}^{\delta\epsilon}} = [\mu_i^{(1)}]_\alpha{}^\phi \frac{\partial[\delta\xi_i^{(1)}]^\beta{}_\phi}{\partial[\delta\mu_j^{(2)}]_\gamma{}^{\delta\epsilon}} - [\mu_i^{(1)}]_\phi{}^\beta \frac{\partial[\delta\xi_i^{(1)}]^\phi{}_\alpha}{\partial[\delta\mu_j^{(2)}]_\gamma{}^{\delta\epsilon}} + \delta_i^j\delta_\alpha^\gamma[\xi_i^{(2)}]^\beta{}_{\delta\epsilon},$$

$$+ [\mu_i^{(2)}]_\alpha{}^{\phi\lambda} \frac{\partial[\delta\xi_i^{(2)}]^\beta{}_{\phi\lambda}}{\partial[\delta\mu_j^{(2)}]_\gamma{}^{\delta\epsilon}} - \delta_i^j\delta_\delta^\beta[\xi_i^{(2)}]^\gamma{}_{\alpha\epsilon}$$

$$- [\mu_i^{(2)}]_\phi{}^{\beta\lambda} \frac{\partial[\delta\xi_i^{(2)}]^\phi{}_{\alpha\lambda}}{\partial[\delta\mu_j^{(2)}]_\gamma{}^{\delta\epsilon}} - \delta_i^j\delta_\epsilon^\beta[\xi_i^{(2)}]^\gamma{}_{\delta\lambda} - [\mu_i^{(2)}]_\phi{}^{\lambda\beta} \frac{\partial[\delta\xi_i^{(2)}]^\phi{}_{\lambda\alpha}}{\partial[\delta\mu_j^{(2)}]_\gamma{}^{\delta\epsilon}},$$

$$\frac{\partial[\delta\dot{\mu}_i^{(2)}]_\alpha{}^{\beta\gamma}}{\partial[\delta q_j^{(0)}]^\delta} = [\mu_i^{(2)}]_\alpha{}^{\epsilon\gamma} \frac{\partial[\delta\xi_i^{(1)}]^\beta{}_\epsilon}{\partial[\delta q_j^{(0)}]^\delta} + [\mu_i^{(2)}]_\alpha{}^{\beta\epsilon} \frac{\partial[\delta\xi_i^{(1)}]^\gamma{}_\epsilon}{\partial[\delta q_j^{(0)}]^\delta} - [\mu_i^{(2)}]_\epsilon{}^{\beta\gamma} \frac{\partial[\delta\xi_i^{(1)}]^\epsilon{}_\alpha}{\partial[\delta q_j^{(0)}]^\delta}$$

$$\frac{\partial[\delta\dot{\mu}^{(2)}]_\alpha{}^{\beta\gamma}}{\partial[\delta q^{(1)}]^\delta{}_\epsilon} = 0, \quad \frac{\partial[\delta\dot{\mu}^{(2)}]_\alpha{}^{\beta\gamma}}{\partial[\delta q^{(2)}]^\delta{}_{\epsilon\phi}} = 0,$$

$$\frac{\partial[\delta\dot{\mu}_i^{(2)}]_\alpha{}^{\beta\gamma}}{\partial[\delta p_j^{(0)}]_\delta} = [\mu_i^{(2)}]_\alpha{}^{\epsilon\gamma} \frac{\partial[\delta\xi_i^{(1)}]^\beta{}_\epsilon}{\partial[\delta p_j^{(0)}]_\delta} + [\mu_i^{(2)}]_\alpha{}^{\beta\epsilon} \frac{\partial[\delta\xi_i^{(1)}]^\gamma{}_\epsilon}{\partial[\delta p_j^{(0)}]_\delta} - [\mu_i^{(2)}]_\epsilon{}^{\beta\gamma} \frac{\partial[\delta\xi_i^{(1)}]^\epsilon{}_\alpha}{\partial[\delta p_j^{(0)}]_\delta}$$

$$\frac{\partial[\delta\dot{\mu}_i^{(2)}]_\alpha{}^{\beta\gamma}}{\partial[\delta\mu_j^{(1)}]_\delta{}^\epsilon} = [\mu_i^{(2)}]_\alpha{}^{\phi\gamma} \frac{\partial[\delta\xi_i^{(1)}]^\beta{}_\phi}{\partial[\delta\mu_j^{(1)}]_\delta{}^\epsilon} + [\mu_i^{(2)}]_\alpha{}^{\beta\phi} \frac{\partial[\delta\xi_i^{(1)}]^\gamma{}_\phi}{\partial[\delta\mu_j^{(1)}]_\delta{}^\epsilon} - [\mu_i^{(2)}]_\phi{}^{\beta\gamma} \frac{\partial[\delta\xi_i^{(1)}]^\phi{}_\alpha}{\partial[\delta\mu_j^{(1)}]_\delta{}^\epsilon}$$

$$\frac{\partial[\delta\dot{\mu}_i^{(2)}]_\alpha{}^{\beta\gamma}}{\partial[\delta\mu_j^{(2)}]_\delta{}^{\epsilon\phi}} = \delta_i^j\delta_\phi^\gamma\delta_\alpha^\delta[\xi_i^{(1)}]^\beta{}_\epsilon + [\mu_i^{(2)}]_\alpha{}^{\lambda\gamma} \frac{\partial[\delta\xi_i^{(1)}]^\beta{}_\lambda}{\partial[\delta\mu_j^{(2)}]_\delta{}^{\epsilon\phi}} + \delta_i^j\delta_\alpha^\delta\delta_\epsilon^\beta[\xi_i^{(1)}]^\gamma{}_\phi$$

$$+ [\mu_i^{(2)}]_\alpha{}^{\beta\lambda} \frac{\partial[\delta\xi_i^{(1)}]^\gamma{}_\lambda}{\partial[\delta\mu_j^{(2)}]_\delta{}^{\epsilon\phi}} - \delta_i^j\delta_\epsilon^\beta\delta_\phi^\gamma[\xi_i^{(1)}]^\delta{}_\alpha - [\mu_i^{(2)}]_\lambda{}^{\beta\gamma} \frac{\partial[\delta\xi_i^{(1)}]^\lambda{}_\alpha}{\partial[\delta\mu_j^{(2)}]_\delta{}^{\epsilon\phi}}$$

$$
\begin{aligned}
\frac{\partial[\delta T_i^{00}]_\alpha}{\partial[\delta q_j^{(0)}]^\beta} &= -p_{j\gamma}p_{k\delta}\partial_{\alpha\beta}K^{\gamma\delta}(jk)\delta_j^i + p_{i\gamma}p_{j\delta}\partial_{\alpha\beta}K^{\gamma\delta}(ij), \\
\frac{\partial[\delta T_i^{01}]_\alpha}{\partial[\delta q_j^{(0)}]^\beta} &= \delta_i^j(p_{j\delta}[\mu_k^{(1)}]_\epsilon{}^\gamma - p_{k\delta}[\mu_j^{(1)}]_\epsilon{}^\gamma)\partial_{\alpha\beta\gamma}K^{\delta\epsilon}(jk) \\
&\quad - (p_{i\delta}[\mu_j^{(1)}]_\epsilon{}^\gamma - p_{j\delta}[\mu_i^{(1)}]_\epsilon{}^\gamma)\partial_{\beta\gamma\alpha}K^{\delta\epsilon}(ij) \\
\frac{\partial[\delta T_i^{02}]_\alpha}{\partial[\delta q_j^{(0)}]^\beta} &=(p_{i\epsilon}[\mu_j^{(2)}]_\phi{}^{\gamma\delta} + p_{j\epsilon}[\mu_i^{(2)}]_\phi{}^{\gamma\delta})\partial_{\gamma\delta\beta\alpha}K^{\epsilon\phi}(ij) \\
&\quad - \delta_i^j(p_{j\epsilon}[\mu_k^{(2)}]_\phi{}^{\gamma\delta} + p_{k\epsilon}[\mu_j^{(2)}]_\phi{}^{\gamma\delta})\partial_{\gamma\delta\beta\alpha}K^{\epsilon\phi}(jk) \\
\frac{\partial[\delta T_i^{12}]_\alpha}{\partial[\delta q_j^{(0)}]^\beta} &= \delta_j^i([\mu_k^{(1)}]_\phi{}^\epsilon[\mu_j^{(2)}]_\lambda{}^{\gamma\delta} - [\mu_j^{(1)}]_\phi{}^\epsilon[\mu_k^{(2)}]_\lambda{}^{\gamma\delta})\partial_{\beta\epsilon\gamma\delta\alpha}K^{\phi\lambda}(jk) \\
&\quad - ([\mu_j^{(1)}]_\phi{}^\epsilon[\mu_i^{(2)}]_\lambda{}^{\gamma\delta} - [\mu_i^{(1)}]_\phi{}^\epsilon[\mu_j^{(2)}]_\lambda{}^{\gamma\delta})\partial_{\beta\epsilon\gamma\delta\alpha}K^{\phi\lambda}(ij) \\
\frac{\partial[\delta T_i^{11}]_\alpha}{\partial[\delta q_j^{(0)}]^\beta} &= \delta_i^j[\mu_j^{(1)}]_\epsilon{}^\delta[\mu_k^{(1)}]_\phi{}^\gamma\partial_{\beta\gamma\delta\alpha}K^{\epsilon\phi}(jk) - [\mu_i^{(1)}]_\epsilon{}^\delta[\mu_j^{(1)}]_\phi{}^\gamma\partial_{\beta\delta\gamma\alpha}K^{\epsilon\phi}(ij) \\
\frac{\partial[\delta T_i^{22}]_\alpha}{\partial[\delta q_j^{(0)}]^\beta} &= -\delta_i^j[\mu_j^{(2)}]_\lambda{}^{\epsilon\phi}[\mu_k^{(2)}]_\zeta{}^{\gamma\delta}\partial_{\beta\epsilon\delta\gamma\phi\alpha}K^{\lambda\zeta}(jk) + [\mu_i^{(2)}]_\lambda{}^{\epsilon\phi}[\mu_j^{(2)}]_\zeta{}^{\gamma\delta}\partial_{\beta\epsilon\delta\gamma\phi\alpha}K^{\lambda\zeta}(ij)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial\delta T_{i\alpha}^{00}}{\partial[\delta p_j^{(0)}]_\beta} &= -\delta_i^j p_{k\gamma}\partial_\alpha K^{\beta\gamma}(jk) - p_{i\gamma}\partial_\alpha K^{\gamma\beta}(ij) \\
\frac{\partial\delta T_{i\alpha}^{01}}{\partial[\delta p_j^{(0)}]_\beta} &= -[\mu_i^{(1)}]_\delta{}^\gamma\partial_{\gamma\alpha}K^{\beta\delta}(ij) + \delta_i^j[\mu_k^{(1)}]_\delta{}^\gamma\partial_{\gamma\alpha}K^{\beta\delta}(jk) \\
\frac{\partial\delta T_{i\alpha}^{02}}{\partial[\delta p_j^{(0)}]_\beta} &= -\delta_i^j[\mu_k^{(2)}]_\epsilon{}^{\gamma\delta}\partial_{\gamma\delta\alpha}K^{\beta\epsilon}(jk) - [\mu_i^{(2)}]_\epsilon{}^{\gamma\delta}\partial_{\gamma\delta\alpha}K^{\beta\epsilon}(ij)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial[\delta T_i^{01}]_\alpha}{\partial[\delta\mu_j^{(1)}]_\beta{}^\gamma} &= -\delta_i^j p_{k\delta}\partial_{\gamma\alpha}K^{\delta\beta}(jk) + p_{i\delta}\partial_{\gamma\alpha}K^{\delta\beta}(ij) \\
\frac{\partial[\delta T_i^{11}]_\alpha}{\partial[\delta\mu_j^{(1)}]_\beta{}^\gamma} &= (\delta_i^j[\mu_k^{(1)}]_\epsilon{}^\delta\partial_{\gamma\delta\alpha}K^{\beta\epsilon}(jk) + [\mu_i^{(1)}]_\epsilon{}^\delta\partial_{\delta\gamma\alpha}K^{\epsilon\beta}(ij) \\
\frac{\partial[\delta T_i^{12}]_\alpha}{\partial[\delta\mu_j^{(1)}]_\beta{}^\gamma} &= -\delta_i^j[\mu_k^{(2)}]_\phi{}^{\epsilon\delta}\partial_{\gamma\epsilon\delta\alpha}K^{\beta\phi}(jk) + [\mu_i^{(2)}]_\phi{}^{\epsilon\delta}\partial_{\gamma\epsilon\delta\alpha}K^{\beta\phi}(ij)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial[\delta T_i^{02}]_\alpha}{\partial[\delta\mu_j^{(2)}]_\beta{}^{\gamma\delta}} &= -(\delta_i^j[\mu_k^{(2)}]_\epsilon{}^{\gamma\delta}\partial_{\gamma\delta\alpha}K^{\beta\epsilon}(jk) + [\mu_i^{(2)}]_\epsilon{}^{\gamma\delta}\partial_{\gamma\delta\alpha}K^{\beta\epsilon}(ij)) \\
\frac{\partial[\delta T_i^{12}]_\alpha}{\partial[\delta\mu_j^{(2)}]_\beta{}^{\gamma\delta}} &= \delta_i^j[\mu_k^{(1)}]_\phi{}^\epsilon\partial_{\epsilon\gamma\delta\alpha}K^{\phi\beta}(jk) - [\mu_i^{(1)}]_\phi{}^\epsilon\partial_{\epsilon\gamma\delta\alpha}K^{\phi\beta}(ij) \\
\frac{\partial[\delta T_i^{22}]_\alpha}{\partial[\delta\mu_j^{(2)}]_\beta{}^{\gamma\delta}} &= -\delta_i^j[\mu_k^{(2)}]_\zeta{}^{\epsilon\phi}\partial_{\gamma\phi\epsilon\delta\alpha}K^{\beta\zeta}(jk) - [\mu_i^{(2)}]_\zeta{}^{\epsilon\phi}\partial_{\epsilon\delta\gamma\phi\alpha}K^{\zeta\beta}(ij)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial[\delta\xi_i^{(1)}]^\alpha{}_\beta}{\partial[\delta q_j^{(0)}]^\gamma} = &\Big(p_{k\delta}\partial_{\beta\gamma}K^{\alpha\delta}(jk) - [\mu_k^{(1)}]_\epsilon{}^\delta\partial_{\beta\gamma\delta}K^{\alpha\epsilon}(jk) + [\mu_k^{(2)}]_\phi{}^{\epsilon\delta}\partial_{\beta\gamma\delta\epsilon}K^{\alpha\phi}(jk)\Big)\,\delta_j^i \\
&- p_{j\delta}\partial_{\beta\gamma}K^{\alpha\delta}(ij) + [\mu_j^{(1)}]_\epsilon{}^\delta\partial_{\beta\gamma\delta}K^{\alpha\epsilon}(ij) - [\mu_j^{(2)}]_\phi{}^{\epsilon\delta}\partial_{\beta\gamma\delta\epsilon}K^{\alpha\phi}(ij) \\
&\frac{\partial[\delta\xi_i^{(1)}]^\alpha{}_\beta}{\partial[\delta q_j^{(1)}]^\gamma{}_\delta} = 0, \quad \frac{\partial[\delta\xi_i^{(1)}]^\alpha{}_\beta}{\partial[\delta q_j^{(2)}]^\gamma{}_{\delta\epsilon}} = 0, \quad \frac{\partial[\delta\xi_i^{(1)}]^\alpha{}_\beta}{\partial[\delta p_j^{(0)}]_\gamma} = \partial_\beta K^{\gamma\alpha}(ij) \\
&\frac{\partial[\delta\xi_i^{(1)}]^\alpha{}_\beta}{\partial[\delta\mu_j^{(1)}]_\gamma{}^\delta} = -\partial_{\delta\beta}K^{\gamma\alpha}(ij) \quad , \quad \frac{\partial[\delta\xi_i^{(1)}]^\alpha{}_\beta}{\partial[\delta\mu_j^{(2)}]_\gamma{}^{\delta\epsilon}} = \partial_{\epsilon\delta\beta}K^{\gamma\alpha}(ij)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial[\xi_i^{(2)}]^\alpha{}_{\beta\gamma}}{\partial[q_j^{(0)}]^\delta} = &\Big[[p_k^{(0)}]_\epsilon\partial_{\beta\gamma\delta}K^{\alpha\epsilon}(jk) - [\mu_k^{(1)}]_\phi{}^\epsilon\partial_{\beta\gamma\delta\epsilon}K^{\alpha\phi}(jk) + [\mu_k^{(2)}]_\lambda{}^{\epsilon\phi}\partial_{\beta\gamma\delta\epsilon\phi}K^{\alpha\lambda}(jk)\Big]\,\delta_j^i \\
&- [p_j^{(0)}]_\epsilon\partial_{\beta\gamma\delta}K^{\alpha\epsilon}(ij) + [\mu_j^{(1)}]_\phi{}^\epsilon\partial_{\beta\gamma\delta\epsilon}K^{\alpha\phi}(ij) - [\mu_j^{(2)}]_\lambda{}^{\epsilon\phi}\partial_{\beta\gamma\delta\epsilon\phi}K^{\alpha\lambda}(ij), \\
&\frac{\partial[\xi_i^{(2)}]^\alpha{}_{\beta\gamma}}{\partial[q_j^{(1)}]^\delta{}_\epsilon} = 0, \quad \frac{\partial[\xi_i^{(2)}]^\alpha{}_{\beta\gamma}}{\partial[q_j^{(2)}]^\delta{}_{\epsilon\phi}} = 0, \quad \frac{\partial[\xi_i^{(2)}]^\alpha{}_{\beta\gamma}}{\partial[p_j^{(0)}]_\delta} = \partial_{\beta\gamma}K^{\alpha\delta}(ij) \\
&\frac{\partial[\xi_i^{(2)}]^\alpha{}_{\beta\gamma}}{\partial[\mu_j^{(1)}]_\delta{}^\epsilon} = -\partial_{\beta\gamma\epsilon}K^{\alpha\delta}(ij), \quad \frac{\partial[\xi_i^{(2)}]^\alpha{}_{\beta\gamma}}{\partial[\mu_j^{(2)}]_\delta{}^{\epsilon\phi}} = \partial_{\beta\gamma\epsilon\phi}K^{\alpha\delta}(ij)
\end{aligned}
$$

The adjoint equation are then given by

$$
\begin{aligned}
\frac{d}{dt}[\lambda_{q_i^{(0)}}]_\alpha = &- [\lambda_{q_j^{(0)}}]_\beta\frac{\partial[\delta\dot{q}_j^{(0)}]^\beta}{\partial[\delta q_i^{(0)}]^\alpha} - [\lambda_{q_j^{(1)}}]_\beta{}^\gamma\frac{\partial[\delta\dot{q}_j^{(1)}]^\beta{}_\gamma}{\partial[\delta q_i^{(0)}]^\alpha} - [\lambda_{q_j^{(2)}}]_\beta{}^{\gamma\delta}\frac{\partial[\delta\dot{q}_j^{(2)}]^\beta{}_{\gamma\delta}}{\partial[\delta q_i^{(0)}]^\alpha} \\
&- [\lambda_{p_j^{(0)}}]^\beta\frac{\partial[\delta\dot{p}_j^{(0)}]_\beta}{\partial[\delta q_i^{(0)}]^\alpha} - [\lambda_{\mu_j^{(1)}}]^\beta{}_\gamma\frac{\partial[\delta\dot{\mu}_j^{(1)}]_\beta{}^\gamma}{\partial[\delta q_i^{(0)}]^\alpha} - [\lambda_{\mu_j^{(2)}}]^\beta{}_{\gamma\delta}\frac{\partial[\delta\dot{\mu}_j^{(2)}]_\beta{}^{\gamma\delta}}{\partial[\delta q_i^{(0)}]^\alpha}
\end{aligned}
$$

References

[BGBHR11] M. Bruveris, F. Gay-Balmaz, D. D. Holm, and T. S. Ratiu, *The momentum map representation of images.*, Journal of Nonlinear Science **21** (2011), no. 1, 115–150.

[BMTY05] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes, *Computing large deformation metric mappings via geodesic flows of diffeomorphisms*, International journal of computer vision **61** (2005), no. 2, 139–157.

[BV04] Stephen Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge University Press, Cambridge, 2004. MR 2061575 (2005d:90002)

[CDTM12] Alina Chertock, Philip Du Toit, and Jerrold Eldon Marsden, *Integration of the EPDiff equation by particle methods*, ESAIM Math. Model. Numer. Anal. **46** (2012), no. 3, 515–534. MR 2877363

[CH93] Roberto Camassa and Darryl D. Holm, *An integrable shallow water equation with peaked solitons*, Phys. Rev. Lett. **71** (1993), no. 11, 1661–1664. MR 1234453 (94f:35121)

[CHJM14] C J Cotter, D D Holm, H O Jacobs, and D M Meier, *A jetlet hierarchy for ideal fluid dynamics*, Journal of Physics A: Mathematical and Theoretical **47** (2014), no. 35, 352001.

[CRM96] GE Christensen, RD Rabbitt, and MI Miller, *Deformable templates using large deformation kinematics*, Image Processing, IEEE Transactions on **5** (1996), no. 10.

[DGM98] Paul Dupuis, Ulf Grenander, and Michael I Miller, *Variational problems on flows of diffeomorphisms for image matching*, Q. Appl. Math. **LVI** (1998), no. 3, 587–600.

[DJR13] M Desbrun, H O Jacobs, and T S Ratiu, *On the coupling between an ideal fluid and immersed particles*, Physica D: Nonlinear Phenomena **265** (2013), no. 0, 40 – 56.

[GM83] Martin Golubitsky and Jerrold Marsden, *The morse lemma in infinite dimensions via singularity theory*, SIAM journal on mathematical analysis **14** (1983), no. 6, 1037–1044.

[HM05] D D Holm and J E Marsden, *Momentum maps and measure-valued solutions (peakons, filaments, and sheets) for the EPDiff equation*, The breadth of symplectic and Poisson geometry, Progr. Math., vol. 232, Birkhäuser Boston, Boston, MA, 2005, pp. 203–235. MR 2103008 (2005g:37144)

[HR06] Helge Holden and Xavier Raynaud, *A convergent numerical scheme for the Camassa-Holm equation based on multipeakons*, Discrete Contin. Dyn. Syst. **14** (2006), no. 3, 505–523. MR 2171724 (2006d:35226)

[Jac13] H O Jacobs, *Symmetries of higher-order momentum distributions in the LDDMM formalism*, Mathematical foundations of computational anatomy, Sep 2013, a MICCAI workshop.

[JM00] Sarang C Joshi and Michael I Miller, *Landmark matching via large deformation diffeomorphisms*, IEEE Transactions on Image Processing **9** (2000), no. 8, 1357–1370.

[KM96] Andreas Kriegl and Peter W Michor, *Regular infinite dimensional lie groups*, J. LIE THEORY **7** (1996), http://www.emis.de/j.

[KMS93] Ivan Kolář, Peter W. Michor, and Jan Slovák, *Natural operations in differential geometry*, Springer-Verlag, Berlin, 1993. MR 1202431 (94a:58004)

[KMS99] I. Kolář, P. W. Michor, and J. Slovák, *Natural operations in differential geometry*, Springer Verlag, 1999.

[MG14] M. Micheli and J. A. Glaunès, *Matrix-valued kernels for shape deformation analysis*, Geometry, Imaging, and Computing **1** (2014), no. 1, 57–39.

[MM13a] Peter W. Michor and David Mumford, *A zoo of diffeomorphism groups on* $\mathbb{R}^n$, Ann. Global Anal. Geom. **44** (2013), no. 4, 529–540. MR 3132089

[MM13b] D Mumford and P W Michor, *On Euler's equation and 'EPDiff'*, Journal of Geometric Mechanics **5** (2013), no. 3, 319–344, arXiv:1209.6576 [math.AP].

[MR99] J. E. Marsden and T. S. Ratiu, *Introduction to mechanics and symmetry*, 2nd ed., Texts in Applied Mathematics, vol. 17, Springer Verlag, 1999.

[MW83] J E Marsden and A Weinstein, *Coadjoint orbits, vortices, and clebsch variables for incompressible fluids*, Physica D: Nonlinear Phenomena **7** (1983), no. 1–3, 305 – 323.

[SDP13] A. Sotiras, C. Davatzikos, and N. Paragios, *Deformable medical image registration: A survey*, IEEE Transactions on Medical Imaging **32** (2013), no. 7, 1153–1190.

[SNDP13] S. Sommer, M. Nielsen, S. Darkner, and X. Pennec, *Higher-order momentum distributions and locally affine lddmm registration*, SIAM Journal on Imaging Sciences **6** (2013), no. 1, 341–367.

[Son98] Eduardo D. Sontag, *Mathematical control theory*, second ed., Texts in Applied Mathematics, vol. 6, Springer-Verlag, New York, 1998, Deterministic finite-dimensional systems. MR 1640001 (99k:93001)

[Tro83] A. J. Tromba, *A sufficient condition for a critical point of a functional to be a minimum and its application to Plateau's problem*, Math. Ann. **263** (1983), no. 3, 303–312. MR 704296 (84i:58033)

[Tro95] Alain TrouvâĹŽÂľ, *An infinite dimensional group approach for physics based models in patterns recognition*, 1995.

[TY05] A. Trouvé and L. Younes, *Local geometry of deformable templates*, SIAM J. Math. Anal. **37** (2005), no. 1, 17–59.

[Wei83] A Weinstein, *The local structure of Poisson manifolds*, J. Differential Geom. **18** (1983), no. 3, 523–557. MR 723816 (86i:58059)

[You10] Laurent Younes, *Shapes and diffeomorphisms*, vol. 171, Springer, 2010.