Learning from demonstration via two-step diffeomorphic curve matching

Author Name1 ABC@SAMPLE.COM and **Author Name2** XYZ@SAMPLE.COM
Address

Editor: Editor's name

Abstract

We propose a novel approach to learn a nonlinear dynamical systems from demonstrations using a combination of diffeomorphic curve matching and Gaussian mixture models. The learned systems are guaranteed to be globally asymptotically stable, and the proposed method is able to learn and generalise multiple movements from an unlabeled set of demonstrations and provides robustness when presented with contradicting demonstrations.

Keywords: Learning from demonstration, diffeomorphic matching, Lyapunov stability

1. Introduction

The ability to construct goal-driven behaviour from demonstration is the key idea of the learning from demonstratin (LfD) paradigm and allows non-experts users to conveniently *teach* the robot new skills. The LfD approach can be applied to a broad variety of use cases, ranging from high-level skills to low-level motions. Our approach aims at the latter. More precisely we focus on motions converging towards a single target configuration, such as grasping motions, as they are the most common motion primitive.

Learning from demonstration has a rich history in the robotics community, for surveys on the subject we refer the reader to Atkeson and Schaal (1997) or more recent Argall et al. (2009), and is also more and more often used in commercial applications, for instance with the robots Baxter and Sawyer. An interesting way to represent motions is to use dynamical systems (DS) ($\dot{\boldsymbol{x}} = f(\boldsymbol{x}, t)$) or autonomous dynamical systems (aDS) ($\dot{\boldsymbol{x}} = f(\boldsymbol{x})$). Representing motions in this way has the advantage of making the system naturally robust against spatial perturbations and in the case of autonomous, i.e. time invariant, systems the robustness is extended to temporal perturbations.

A possible approach to construct or learn the dynamical system is to use statistical models, such as Gaussian Mixture Models (GMM), and train them with the data given in the form of the demonstrations available. This approach however has the drawback that no guarantees usually can be given regarding the stability of the learned system. The resulting vector field may indeed have divergent zones, limit cycles and spurious attractors. These phenomena especially occur when trying to generalise the movement to regions where no demonstrations are available. A second, less frequently discussed issue associated with statistical models, is what we term drift error. Since the model is trained on the data pairs position and velocity, small errors between the learned and real velocities do not guarantee a small *distance* between the resulting trajectories and the given demonstrations. A small but consistent error can result in large differences due to the integration. To the best of our knowledge no integral constraints, so constraints on the closeness of demonstration trajectories and obtained trajectories, has been derived for statistical models.

To tackle the first problem, different approaches have been developed based on Lyapunov theory that either directly constrain the parameters of the dynamical system like SEDS (Khansari-Zadeh and Billard (2011)) or learn a suitable Lyapunov function and a correction signal like WASQF (Khansari-Zadeh and Billard (2014)). The first approach using diffeomorphic transformations is given by Neumann and Steil (2015). In this work the authors derive the transformation from a compatible Lyapunov function. All these approaches have in common that the expressiveness of the Lyapunov functions have to be limited in order to render the problem computationally tractable or due to inherent constraints of the approach.

This limitation is overcome in Perrin and Schlehuber-Caissier (2016), by directly constructing a diffeomorphic transformation based on the given demonstrations from which a stable non-linear dynamical system reproducing the demonstrations as well as a corresponding Lyapunov function can be deduced. This approach is the most comparable to ours with two main differences: first our matching algorithm seeks to learn the shape of a curve, instead of pure point matching, secondly we couple the diffeomorphic matching with GMMs and statistical considerations in order to be able to handle multiple demonstrations at the same time. Our main contribution is to overcome some of the drawbacks of the method presented in Perrin and Schlehuber-Caissier (2016), which we list in Section 4. A more minor contribution is to attempt to tackle the issue related to the drift error. In our approach, we divide the problem into two parts, and compute what we call *control-space dynamics* and *demonstration-space dynamics*. We partially address the drift error issue in the way we construct the control-space dynamics. To the best of our knowledge, no work directly linked to this construction exists. A newly proposed method called Mixture of Attractors Manschitz et al. (2018) derives drift error-related constraints in convex fashion constructing a time-dependent nonlinear dynamical system, however it is not clear how to integrate such constraints in our case.

2. Problem statement

We seek to construct an autonomous dynamical system $\boldsymbol{v} = \dot{\boldsymbol{x}} = g(\boldsymbol{x})$ that is guaranteed to be globally asymptotically stable at the origin and which reproduces as well as possible the set of given demonstrations. The demonstrations are given as timed series of positions and velocities $\mathcal{D} = \cup_{i \in [0,N_d]} \mathcal{D}_i;\ \mathcal{D}_i = [\boldsymbol{x}_i, \boldsymbol{v}_i, t_i]$, where i is used to index the demonstration, i.e. a single point-to-point movement, with N_d being the total number of demonstrations. We also wish that the constructed DS generalises the movements into some neighbourhood around the demonstrations. Note that we do not require that the demonstrations show variants of the same movement. The only condition imposed is that all demonstrations terminate at the same point, here without loss of generality taken to be the origin.

3. Definition

In the remainder of this article, several definitions concerning curves are used, which are given here. A curve $\mathcal{C}$ embedded in $\mathbb{R}^d$ is defined as the set of all points generated from its parametric function $c : [0, 1] \mapsto \mathbb{R}^d$, with $c(0)/c(1)$ being the start/end point of the curve. We say that a curve is in its discretized form $\mathcal{C}^{dis}$ when talking about a sorted list of N points that correspond to the evaluation of its parametric form for some $s \in [0, 1]$ resulting in $\mathcal{C}^{dis} = [\boldsymbol{x}_i = c(s_i)]_{i\in[0,N-1]}$ with $0 \leq s_i < s_{i+1} \leq 1$. We say that a curve is naturally parametrized if $\forall s \in [0,1] : s = \int_0^s \left\| \frac{d}{ds}c(s) \right\|_2 ds / \int_0^1 \left\| \frac{d}{ds}c(s) \right\|_2$, so s corresponds to the fraction of the arc length of the curve between the start point and the point corresponding to s. We say that two curves $\mathcal{C}_1$ and $\mathcal{C}_2$ given in their parametric form c_1 and c_2 are *geometrically identical* if there exists a strictly monotonic function $h : [0,1] \mapsto [0,1]$ such that $\forall s \in [0,1] : c_1(s) = c_2(h(s))$. We say that the source and target curves are *compatible* if there exists a (*reasonably regular*[1]) diffeomorphic transformation Φ such that images of the source curves under the transformation approximately match the target curves. We also introduce the weaker notion of *dynamic compatibility*: we say that the source and target curves are dynamically compatible if there exist parametrization functions and a transformation Φ such that the images of the parametrized source curves under Φ approximately match the target curves, see Figure 1. Finally the source and target curves are considered *dynamically incompatible* if no such combination of parametrization functions and a transformation exists (see Figure 1).

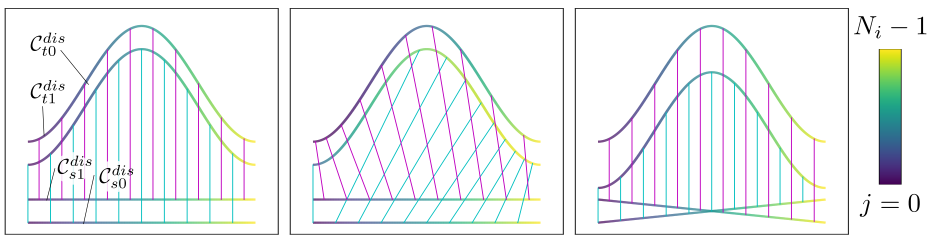

Figure 1: The images show different source ($\mathcal{C}_{s0}^{dis}$, $\mathcal{C}_{s1}^{dis}$) and target ($\mathcal{C}_{t0}^{dis}$, $\mathcal{C}_{t1}^{dis}$) curves in their discretized form. The left image shows a *compatible* configuration, that is there exists a reasonably regular diffeomorphic transformation for which $\Phi(\mathcal{C}_{s0}^{dis}) \approx \mathcal{C}_{t0}^{dis}$ and $\Phi(\mathcal{C}_{s1}^{dis}) \approx \mathcal{C}_{t1}^{dis}$. The image in the middle shows a *dynamically compatible* configuration. That is, there exists no transformation taking the sources to the targets. However if we seek to match the *curves* instead of the points representing the discretized curves, we can find two parametrization functions h_0, h_1 and a transformation Φ for which $\Phi(c_{s0}(h_0(s))) \approx c_{t0}(s)$ and $\Phi(c_{s1}(h_1(s))) \approx c_{t1}(s)$ holds for all $s \in [0,1]$, where c_x is the parametric form of the curve $\mathcal{C}_x$. The image on the right depicts a *dynamically incompatible* configuration. Since the source curves do intersect and the target curves do not intersect, no diffeomorphic transformation can exist between these curves independently of the parametrization. In this case the *best* solution is to find a transformation that minimizes some error criterion.

4. Diffeomorphic matching

The method proposed in Perrin and Schlehuber-Caissier (2016) can be summed up into the following steps: first, a representative demonstration is generated for each distinct movement (label) by averaging over all given demonstrations with the same label. In the second step a simple globally stable autonomous dynamical system $\dot{\boldsymbol{y}} = f(\boldsymbol{y})$ (defining what we call the control-space dynamics) is defined to generate converging trajectories, called $\mathcal{T}$. In the last step a diffeomorphic transformation Φ is constructed such that the distance, defined as a suitable cost function Cost, between the image of the control space trajectories $\mathcal{T}$ under the diffeomorphism Φ and the representative demonstrations is minimized. Note that this approach corresponds to diffeomorphic point matching. Finally the demonstration-space

1. We keep this notion of reasonable regularity vague: it roughly corresponds to diffeomorphisms that produce only limited deformations.

dynamics are given as $\dot{\boldsymbol{x}} = g(\boldsymbol{x}) = J_\Phi(\Phi^{-1}(\boldsymbol{x})).f(\Phi^{-1}(\boldsymbol{x}))$, with $J_\Phi(\boldsymbol{y})$ being the Jacobian matrix $J_\Phi(\boldsymbol{y}) = \frac{\partial \Phi}{\partial \boldsymbol{y}}(\boldsymbol{y})$ and Φ^{-1} denoting the inverse transformation of Φ. This approach yields good results for the publicly available LASA-Handwriting dataset (Khansari-Zadeh and Billard (2011)), suffers however from a couple of drawbacks which we will detail here and address throughout the rest of the paper.

(1) Each demonstration has to be labelled according to the movement shown
(2) The algorithm cannot handle well multiple demonstrations for the same movement
(3) The expressiveness is reduced if multiple movements are given
(4) The dynamics found cannot be improved or modified by providing new demonstrations
(5) Overly deformed state-spaces
    (5.1) The a priori fixed control-space dynamics necessitates a deformation of the state-space even in regions naturally converging
    (5.2) Constructing the diffeomorphism as a long composition of local translations

In this work we therefore propose a new approach based on diffeomorphic curve matching and statistical learning. The main ideas behind this approach are to separate the learning of the norm (magnitude) of the velocity from the direction of velocity and to search for a diffeomorphic transformation that seeks to match curves regarding their geometry. We propose to learn the magnitude using classic statical methods whereas the direction is *learned* by the transformation in regions where the movement does not naturally converge with respect to the l_2-norm[2]. Finally we seek to improve the structure of the transformation by grouping several locally weighted translations into one diffeomorphic (sub-)transformation in order to increase the regularity if the transformation.
Separating the norm of the velocity from the direction directly allows us to cope with dynamically compatible demonstrations, partially resolving drawback (2). A diffeomorphic transformation matching the geometry in such cases can be found using Algorithm 2 introduced in the next section. As we use statistical models to learn the magnitude of the velocity, contradicting values do not pose a problem. Moreover our approach provides higher robustness to dynamical incompatibility and directly takes advantage of multiple demonstrations showing the same movement to achieve better generalisation.
In order to structure the rest of the paper, we first present how we compute the demonstration-space dynamics in Algorithm 1 and then detail the different parts and show how to generate them from the demonstrations later on.

Algorithm 1: demonstration-space dynamics

Input: demonstration-space point $\boldsymbol{x}$
Output: demonstration-space velocity $\dot{\boldsymbol{x}} = g(\boldsymbol{x})$
$\boldsymbol{y} \leftarrow \Phi^{-1}(\boldsymbol{x})$ Compute the preimage of $\boldsymbol{x}$ unde Φ
$\dot{\boldsymbol{y}} \leftarrow f(\boldsymbol{y})$ Evaluate the constrol-space dynamics
$\dot{\boldsymbol{y}} \leftarrow \textsc{EnsureConv}(\boldsymbol{y},\dot{\boldsymbol{y}})$ Ensure convergence
$\dot{\boldsymbol{x}} \leftarrow J_\Phi(\boldsymbol{y}).\dot{\boldsymbol{y}}$ Transform into demonstration space
$\dot{\boldsymbol{x}} \leftarrow \dot{\boldsymbol{x}}\frac{h(\boldsymbol{x})}{\|\dot{\boldsymbol{x}}\|_2}$ Scale the direction to obtain a velocity
return $\dot{\boldsymbol{x}}$

with $h(\boldsymbol{x})$ being the magnitude model $h : \mathbb{R}^d \mapsto [0 < \epsilon, \inf[$ whose parameters are directly learned in the demonstration-space. In this paper we use the maximum a posteriori (MAP) of a GMM trained on the concatenation of position and norm of velocity of all measurement

2. Due to the lack of a better terms, we use *dynamics* for the magnitude of the velocity and *geometry* for the direction.

points in the demonstration set using the greedy search strategy presented by Verbeek et al. (2003). The control-space dynamic $f(\boldsymbol{y})$ is constructed based on some statistical model trained on a converging version of the demonstrations. The procedure EnsureConv takes the current control-space position $\boldsymbol{y}$ and the desired direction $\dot{\boldsymbol{y}}$ and ensures the global asymptotic stability by enforcing ${}^{\boldsymbol{y}^\intercal \cdot \dot{\boldsymbol{y}}}/_{\|\boldsymbol{y}\|_2 \cdot \|\dot{\boldsymbol{y}}\|_2} \leq \beta$ where $\beta < 0$ corresponds to a minimal convergence angle. Then we use the Jacobian of the diffeomorphic transformation to transpose the control-space direction into the demonstration-space direction. Finally the direction is transformed into a real velocity by normalizing it, then scaling it using the magnitude function $h(.)$.

4.1. Curve matching algorithm

In this section we present a new type of diffeomorphic transformation and the curve matching algorithm used in this work.

Locally weighted diffeomorphic multitranslations

To improve the structure of the diffeomorphic transformation and obtain a more compact representation, the concept of locally weighted multitranslations is introduced. This allows us to group several locally weighted translations into one diffeomorphic sub-transformation of the form

$$\Psi(\boldsymbol{x}) = \boldsymbol{x} + \sum_i k_{p_i,b_i,\boldsymbol{c}_i}(\boldsymbol{x}).\boldsymbol{u}_i \tag{1}$$

where $k_{p_i,b_i,\boldsymbol{c}_i}$ denotes the weight function for which we use piecewise polynomial radial basis functions $\mathbb{R}^d \mapsto \mathbb{R}_0^+$ centred at $\boldsymbol{c}_i$ and base b_i. For these weight functions sufficient conditions ensuring that Ψ is indeed diffeomorphic can be derived similar to those presented in Perrin and Schlehuber-Caissier (2016). We refer the reader to the appendix for a detailed discussion. To keep the notation more compact, we write Ψ_θ when we talk about a locally weighted multitranslation that is defined by the vector of parameters θ, which holds all necessary information.
The inverse for this type of transformation can not be given in closed form, but efficiently computed using Newton's method.

Heuristic algorithm

In Perrin and Schlehuber-Caissier (2016) the heuristic to successively construct the transformation Φ composed locally weighted translations that, at each iteration, seek to minimize $\textsc{Cost}(\Psi_i(\text{source}_i), \text{target})$, with source_i denoting the current image of the source points after $i-1$ transformations $\text{source}_i = \Psi_{i-1} \circ \cdots \circ \Psi_{i-1}(\text{source})$. This heuristic also fixed the center and translation of the transformation Ψ_i using the largest current error between the current source and the target. This approach can not handle well dynamically compatible configurations and fails when confronted with dynamical incompatibilities (see drawback (2)). To resolve these problems we propose Algorithm 2 which performs approximative heuristic curve matching instead of point matching.

The procedure Naturalize takes curves in their discretized form $\mathcal{C}^{dis}$ (a list of points) and returns the discretized curve in its natural parametrisation with the same number of evaluation points. The function ComputeCenterAndDir uses the current source curves and errors to deduce a center and translation that is likely to significantly reduce the error

Algorithm 2: Approximative curve matching

Input: List of source $\left[\mathcal{C}_{si}^{dis}\right]_i$ and target $\left[\mathcal{C}_{ti}^{dis}\right]_i$ curves in their discretized form, max. nbr of translations per multitranslation N_t, max. nbr of multitransformations N_m
Output: Diffeomorphic transformation Φ
$\Phi \leftarrow Id$ Initialize transformation to identity
$\forall i : \mathcal{C}_{ti}^{dis} \leftarrow \textsc{Naturalize}(\mathcal{C}_{ti}^{dis})$ Naturalize the curve representation
for $k \leftarrow 0$ **to** $N_m - 1$ **by** 1 **do**
  $\forall i : \mathcal{C}_{ci}^{dis} \leftarrow \textsc{Naturalize}(\Phi(\mathcal{C}_{si}^{dis}))$ Compute current natural image of the sources
  $\theta_k = [\,]$ Initialize multitranslation to identity$\leftrightarrow$ empty parameter vector
  for $j \leftarrow 0$ **to** $N_t - 1$ **by** 1 **do**
    $\forall i : \mathcal{C}_{csi}^{dis} \leftarrow \textsc{Naturalize}(\Psi_{\theta_k}(\mathcal{C}_{ci}^{dis}))$ Update current source curve
    $\forall i : e_i \leftarrow \mathcal{C}_{ti}^{dis} - \mathcal{C}_{csi}^{dis}$ Compute the current distance
    $\boldsymbol{c}_j, \boldsymbol{u}_j \leftarrow \textsc{ComputeCenterAndDir}(\left[\mathcal{C}_{ci}^{dis}\right]_i, [e_i]_i)$
    $\alpha_j^*, b_j^* \leftarrow \underset{\alpha_j, b_j}{\operatorname{argmin}} \textsc{Cost}\left([e_i]_i - k_{p_j, b_j, \boldsymbol{c}_j}(\left[\mathcal{C}_{ci}^{dis}\right]_i).(\alpha_j \boldsymbol{u}_j)\right)$
    $\theta_k \leftarrow \textsc{EnsureDiffeo}(\textsc{Join}(\theta_k, \boldsymbol{c}_j, \alpha_j^* \boldsymbol{u}_j, b_j^*))$ Add the new parameters
  end
  $\Phi \leftarrow \Psi_{\theta_k} \circ \cdots \circ \Psi_{\theta_0}$
  if *converged*$\left(\Phi, \left[\mathcal{C}_{sk}^{dis}\right]_k, \left[\mathcal{C}_{tk}^{dis}\right]_k\right)$ **then**
    return Φ
  end
end
return Φ

when used to construct a locally weighted translation. In this work we concatenate the current source curves and errors in order to cluster them using k-means algorithm[3] with an *a priori* fixed number of clusters. The means of the resulting clusters can then be interpreted as a concatenation of centers and translations likely to improve the matching. Finally the function selects the *best* center and translation pair using a heuristic and returns it. To find a good locally weighted translation given the suggested center and translation, we perform a grid search over the scaling factor α_j and for each value of α_j in the grid we perform a minimization to find the best base size b_j. The function Cost evaluates the quality of the matching. Here we use the mean squared euclidean norm of the error, regularized by a small term giving preference to larger bases. We have a preference for large bases as this usually indicates a higher regularity of the transformation.
Since the choice of the center and translation direction of each locally weighted translation is based on statistical considerations, the robustness of the algorithm when faced with dynamical compatibilities is improved, while the repeated naturalization of the source curves automatically reparametrizes the underlying function of each curve dealing with dynamic compatibilities.
Please note that we are therefore no longer in a real point matching setting but seek to match the geometry of the curves. There exist other methods to perform curve matching, notably the approach to construct a flow that minimizes a predefined measure (see for in-

3. For this work we used the implementation provided by Pedregosa et al. (2011)

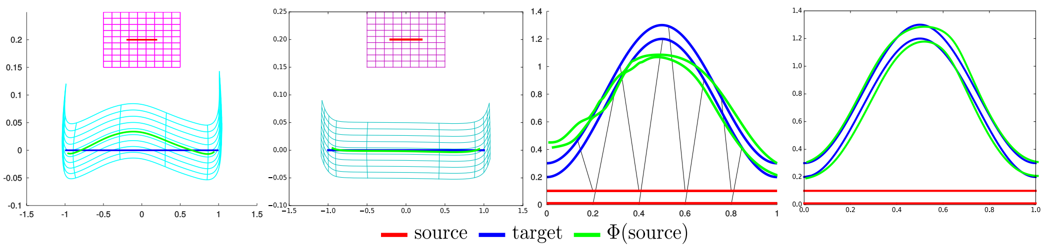

Figure 2: These images compare the resulting transformations of our approach (2nd and 4th image) and the approach proposed in Perrin and Schlehuber-Caissier (2016) (1st and 3rd image) in different settings. The results are obtained for the same number of local translations. The first two images depict a feasible setting. Our approach achieves better overall matching of the curves while maintaining a higher regularity of the transformation. The second two images depict a setting with dynamic compatibility comparable to the one in Figure 1, which can be handled by the proposed approach.

stance Dupuis et al. (1998), Glaunès et al. (2008)) expressing the closeness or resemblance of two curves. We could use such measures as COST, avoiding the repeated naturalization of the curves, however these measures are either prohibitively expensive to compute when dealing with many points, susceptible to local minima or also take into account the *dynamics* of the curve.
The advantages of this new approach, which results in faster convergence (in the sense of the number of locally weighted translations used) and higher regularity of the transformation, are showcased in Figure 2. However these advantages come at the price of a longer computation time (10 to 40 times longer) to construct Φ, which however is not a real issue as we can still learn the transformation in less than one minute. The significant operations are the ones to be performed repeatedly in the control loop, the evaluation of Φ and Φ^{-1}, for which computation time is comparable or even reduced.

5. Learning nonlinear dynamical systems with reduced space distortion

To ensure global asymptotic convergence we impose that the movement has to converge with respect to the l_2-norm along the trajectory in the control-space, which is equivalent to imposing the Lyapunov function candidate $V(\boldsymbol{y}) = \boldsymbol{y}^\intercal.\boldsymbol{y}$ as Lyapunov function for the DS. Note that one could also seek to impose other Lyapunov function candidates. However the squared euclidean distance is a *natural* choice in absence of a generic method to construct suitable Lyapunov functions. In contrast to the approach taken in Perrin and Schlehuber-Caissier (2016) where the control-space dynamics are (almost) entirely defined *a priori*, we seek to construct it from the given demonstrations, which resolves the drawbacks (1) and (3) and helps to tackle the drawbacks (2) and (5.1). This yields the advantage that we do not have to deform the state-space in regions where the demonstrations already naturally converge and the necessary overall deformation is reduced. We achieve this by first creating a converging curve $\mathcal{C}_{fs}^{dis}$ for every demonstration by, loosely speaking, replacing non-converging sections of the curve by a converging arc as shown in fig 3. We then construct a globally stable vector field based on some statistical model, as for instance SEDS (Khansari-Zadeh

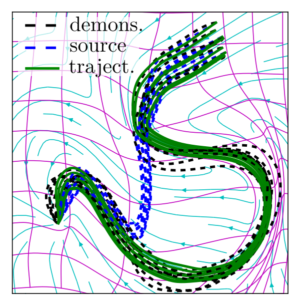

Figure 3: Depiction of the demonstration-space trajectory and dynamics.

and Billard (2011)), by training it on the data pairs position and (tangential) direction of the curves. The so obtained control-space dynamics $\dot{\boldsymbol{y}} = f(\boldsymbol{y})$ faithfully reproduces the geometry of the demonstrations. As SEDS suffers from what we termed earlier as drift error, we use a different approach that constructs a weighted sum of dynamics that converge locally towards the demonstrations, see appendix for details. In the second step we use the forward trajectories of the start point of each demonstration under this vector or better direction field as source curves. As in this approach only the direction is of interest, we can demand $\forall \boldsymbol{y} : \|f(\boldsymbol{y})\|_2 = 1$. Please note that one could also directly use the feasible curves as source curves, however this leads to inferior results.

Results on the LASA-Dataset

To showcase the resulting dynamics obtained with our approach we present the results for a selection of movements that do not naturally converge with respect to the l_2-norm from the publicly available LASA-Dataset. Moreover, we present results for combined demonstrations in the right column of Figure 4. The composed demonstration in the bottom ($Worm$-dataset with y-axis scaled by 3 and hee) further restricts the choice of the control space dynamics for the approach proposed in Perrin and Schlehuber-Caissier (2016), whereas the demonstration in the top row (composed of two $SharpC$-trajectories, one scaled by a factor of two) cannot be learned.

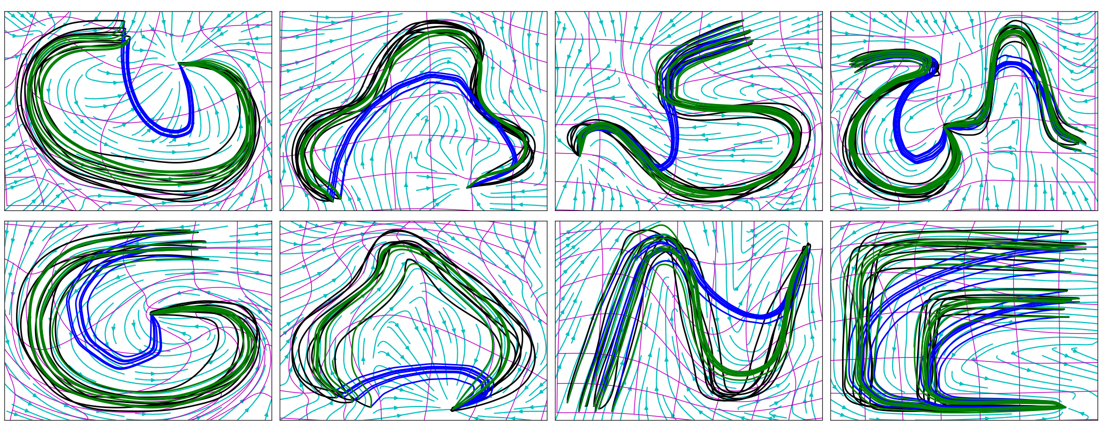

Figure 4: The obtained vector fields and trajectories for the datasets $Sharpc$, $Leaf_1$, $Leaf_2$, $DoubleBendedLine$, $NShape$, $Snake$ as well as for the composed demonstrations in the right column are shown. The cyan coloured grid corresponds to the image of a regular grid under the transformation Φ.

6. Conclusion

In this paper we presented a new approach to learn globally asymptotically nonlinear dynamics from presentation by evolving the idea of using diffeomorphic transformations developed in Perrin and Schlehuber-Caissier (2016) and coupling it with statistical considerations and machine learning. This approach is able to learn movements from as few as one demonstration while being able to deduce better generalization when presented with more demonstrations. The compact structure of the proposed transformation as composition of multitranslations moreover paves the way to resolve the problem of integrating new information, as in the use case of lifelong or active learning Maeda et al. (2017), using known techniques from deep-learning like back propagation, which is part of our ongoing work.

References

Brenna D Argall, Sonia Chernova, Manuela Veloso, and Brett Browning. A survey of robot learning from demonstration. *Robotics and autonomous systems*, 57(5):469–483, 2009.

Christopher G Atkeson and Stefan Schaal. Robot learning from demonstration. In *ICML*, volume 97, pages 12–20, 1997.

Paul Dupuis, Ulf Grenander, and Michael I Miller. Variational problems on flows of diffeomorphisms for image matching. *Quarterly of applied mathematics*, pages 587–600, 1998.

Joan Glaunès, Anqi Qiu, Michael I Miller, and Laurent Younes. Large deformation diffeomorphic metric curve mapping. *International journal of computer vision*, 80(3):317, 2008.

Seyed Mohammad Khansari-Zadeh and Aude Billard. Learning stable nonlinear dynamical systems with gaussian mixture models. *IEEE Transactions on Robotics*, 27(5):943–957, 2011.

Seyed Mohammad Khansari-Zadeh and Aude Billard. Learning control lyapunov function to ensure stability of dynamical system-based robot reaching motions. *Robotics and Autonomous Systems*, 62(6):752–765, 2014.

Guilherme Maeda, Marco Ewerton, Takayuki Osa, Baptiste Busch, and Jan Peters. Active incremental learning of robot movement primitives. In *Conference on Robot Learning (CORL)*, 2017.

Simon Manschitz, Michael Gienger, Jens Kober, and Jan Peters. Mixture of attractors: A novel movement primitive representation for learning motor skills from demonstrations. *IEEE Robotics and Automation Letters*, 2018.

Klaus Neumann and Jochen J Steil. Learning robot motions with stable dynamical systems under diffeomorphic transformations. *Robotics and Autonomous Systems*, 70:1–15, 2015.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python . *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Nicolas Perrin and Philipp Schlehuber-Caissier. Fast diffeomorphic matching to learn globally asymptotically stable nonlinear dynamical systems. *Systems & Control Letters*, 96: 51–59, 2016.

Jakob J Verbeek, Nikos Vlassis, and Ben Kröse. Efficient greedy learning of gaussian mixture models. *Neural computation*, 15(2):469–485, 2003.