Molecular Computational Anatomy: Unifying the Molecular to Tissue Continuum Via Measure Representations of the Brain

Michael Miller¹, Daniel Tward², and Alain Trouvé³

¹Department of Biomedical Engineering & Kavli Neuroscience Discovery Institute & Center for Imaging Science, Johns Hopkins University, Baltimore, USA
²Departments of Computational Medicine & Neurology, University of California Los Angeles, Los Angeles, USA
³Centre Giovanni Borelli (UMR 9010), Ecole Normale Supérieure Paris-Saclay, Université Paris-Saclay, Gif-sur-Yvette, FRANCE

*Corresponding author. Email: mim@jhu.edu

April 2, 2022

Abstract

Objective: The objective of this research is to unify the molecular representations of spatial transcriptomics and cellular scale histology with the tissue scales of Computational Anatomy for brain mapping. Impact statement: We present a unified representation theory for brain mapping based on geometric measures of the micro-scale deterministic structure and function with the statistical ensemble of the spatial aggregated tissue scales. Introduction: Mapping across coordinate systems in computational anatomy allows us to understand structural and functional properties of the brain at the millimeter scale. New measurement technologies in digital pathology and spatial transcriptomics allow us to measure the brain molecule by molecule and cell by cell based on protein and transcriptomic identity. We currently have no mathematical representations for integrating consistently the tissue limits with the molecular particle descriptions. The formalism derived here demonstrates the methodology for transitioning consistently from the molecular scale of quantized particles – as first introduced by Dirac as the class of generalized functions – to the continuum and fluid mechanics scales appropriate for tissue. Methods: We introduce two methods based on notions of generalized function geometric measures and statistical mechanics. We use generalized functions expanded to include functional geometric descriptions - electrophysiology, transcriptomic, molecular histology – to represent the molecular biology scale integrated with a Boltzman like procedure to pass from the sparse deterministic particle descriptions to empirical probability laws on the functional state of the tissue. Results: We demonstrate a unified mapping methodology for transferring molecular information in the transcriptome and histological scales to the human atlas scales for understanding Alzheimer’s disease. Conclusions: We demonstrate a unified brain mapping theory for molecular and tissue scales based on geometric measure representations. We call the consistent aggregation of tissue scales from particle and cellular scales molecular Computational Anatomy.

1 Introduction

One of the striking aspects of the study of the brain in modern neurobiology is the fact that the distributions of discrete structures that make up physical tissue, from neural cells to synapses to genes and molecules, exists across nearly ten orders of magnitude in
spatial scale. This paper focusses on the challenge of building multi-scale representations that simultaneously connect the quantum nano-scales of modern molecular biology and digital pathology for characterizing neural circuits architecture in the functioning brain and disease with the classical continuum representations at the anatomical gross and meso scales.

We have been highly motivated by the Cell Census Network project (BICCN [1]) which brings the nano and micron scales of single cell measures of RNA via spatial transcriptomics [2–4] coupled to the tissue scales of mouse atlases. The recent review on bridging scales from cells to physiology [5] motivates the mathematical framework presented herein. The recent emergence of spatial transcriptomics as Nature method of the year highlights the importance and ascendance of such approaches for understanding the dense metric structure of the brain which represent the coarse physiological atlas scales built up from dense imaging measurements at the cellular scales. Specifically, in our own work digital pathology for the study of Alzheimer’s disease called the BIOCARD study [6], we are examining pathological Tau at both the micro histological and macro atlas scales of Tau particle detections, from 10-100 µm [7, 8] and to human magnetic resonance millimeter scales for examining entire circuits in the medial temporal lobe. In the mouse cell counting project we are examining single-cell spatial transcriptomics using modern RNA sequencing in dense tissue at the micron scale and its representations in the Allen atlas coordinates [9].

Most noteworthy for any representation is that at the finest micro scales nothing is smooth; the distributions of cells and molecules are more well described as random quantum counting processes in space [10]. In contrast, information associated to atlasing methods at the gross anatomical tissue and organ scales of Computational Anatomy extend smoothly [11–16]. Cross-sectionally and even cross-species, gross anatomical labelling is largely repeatable, implying information transfers and changes from one coordinate system to another smoothly. This is built into the representation theory of diffeomorphisms and soft matter tissue models for which advection and transport hold [17–23], principles upon which continuum mechanics and its analogues are based. Also of note is the fact that the brain organizes information on geometric objects, submanifolds of the brains such as the foliation of the cortex and associated coordinates of the cortical columns. Our representations must both represent the quantum to ensemble scales as well as encode the organization of the brain.

The focus of this paper is to build a coherent representation theory across scales. For this we view the micron to millimeter scales via the same representation theory called mathematical geometric measures, building the finest micron scales from discrete units termed particle measures which represent molecules, synapses and cells. The measure representation from fine to coarse scale aggregates forming tissue. This measure representation allows us to understand subsets of tissues that contain discretely positioned and placed functional objects at the finest quantized scales and simultaneously pass smoothly with aggregation to the classical continuum scales at which stable functional and anatomical representations exist. Since the study of the function of the brain on its geometric submanifolds -the gyri, sulci, subnuclei and laminae of cortex- are so important, we extend our general framework to exploit varifold measures [24] arising in the modern discipline of geometric measure theory. Geometric measures are a class of generalized functions which have the basic measure property of additivity on disjoint unions of the experimental probe space and encode the complex physiological functions with the geometric properties of the submanifolds to which they are associated. To be able to compare the brains we use diffeomorphisms as the comparator tool, with their action representing 3D varifold action which we formulate as “copy and paste” so that basic particle quantities that are conserved biologically are combined with greater multiplicity and not geometrically distorted as would be the case for measure transport.

The functional features are represented via generalized Dirac delta functions at the finest micro structure scales. The functional feature is abstracted into a function space rich enough to accomodate the molecular machinery as represented by RNA or Tau particles, as well as electrophysiology associated to spiking neurons, or at the tissue scales of medical imaging dense contrasts of magnetic resonance images (MRIs). We pass to the classical function continuum via introduction of a scale-space that extends the descriptions of cortical micro-circuits to the meso and anatomical scales. This passage from the quantized features to the stochastic laws is in fact akin to the Boltzman program transferring the view from the Newtonian particles to the stable distributions describing them. For this we introduce a scale-space of kernel density transformations which allows us to retrieve the empirical averages.
represented by the determinism of the stochastic law consistent with our views of the macro tissue scales.

The representation provides a recipe for scale traversal in terms of a cascade of linear space scaling composed with non-linear functional feature mapping. Following the cascade implies every scale is a measure so that a universal family of measure norms can be introduced which simultaneously measure the disparity between brains in the orbit independent of the probing technology, RNA identities, Tau or amyloid histology, spike trains, or dense MR imagery.

Our brain measure model implies the existence of a sequence. This scale-space of pairs, the measure representation of the brain and the associated probing measurement technologies we call Brainspace. To formulate a consistent measurement and comparison technology on Brainspace we construct a natural metric upon it allowing us to study its geometry and connectedness. The metric between brains is constructed via a Hamiltonian which defines the geodesic connections throughout scale space, providing for the first time a hierarchical representation that unifies micro to millimeter representation in the brain and makes Brainspace into a metric space. Examples of representation and comparison are given for Alzheimer’s histology integrated to magnetic resonance imaging scales, and spatial transcriptomics. We call the consistent formalism presented here for aggregation of tissue scales from particle and cellular scales molecular Computational Anatomy.

2 Results

2.1 Measure Model of Brain Structures

To build a coherent theory we view the micron to anatomical scales via the same representation theory building upon discrete units termed particles or atoms. As they aggregate they form tissues. This is depicted in Figure 1 in which the top left panel shows mouse imaging of CUX1 labelling of the inner layers of mouse cortex (white) and CTP2 imaging of the outer layers (green) at 2.5 micron in plane resolution. Notice the discrete nature of the cells clearly resolved which form the layers of tissue which are the global macro scale features of layer 2,3,4 which stain more prolifically in white and the outer layers 5,6 which stain more prolifically in green.

Our representation exists simultaneously at both the micro and tissue millimeter scales. A key aspect of anatomy is that at a micro or nano scale, information is encoded as a massive collection of pairs \((x_i, f_i)\) where \(x_i \in \mathbb{R}^d\) \((d = 2, 3)\) describes the position of a “particle” and \(f_i\) is a functional state in a given set \(\mathcal{F}\) attached to it. In our applications \(\mathcal{F}\) are proteins representing RNA signatures or Tau tangles, and for single cell Neurophysiology represents the dynamics of neural spiking. At the micro scale basically everything is deterministic, with every particle attached to its own functional state among possible functional state in \(\mathcal{F}\). But zooming out, the tissue level, say mm scale, appears through the statistical distribution of its constituents with two key quantities, the local density of particles \(\rho\) and the conditional probability distribution of the functional features \(\mu_x(df)\) at any location \(x\). At position \(x\), we no longer have a deterministic functional state but a probability distribution \(\mu_x\) on functional states.

The integration of both descriptions into a common mathematical framework can be done quite naturally in the setting of mathematical measures which are mathematical constructs that are able to represent both the discrete and continuous worlds as well as natural level of approximation between both. We associate the elementary ‘Dirac’ \(\delta_x \otimes \delta_{f}\) which applies to infinitesimal volumes in space and function so that \(\delta_x \otimes \delta_{f}(dx, df) = \delta_x(dx)\delta_{f}(df)\) is equal to 1 if \(x \in dx\) and \(f \in df\) and 0, otherwise. Indeed the set \(\mathcal{M}(\mathbb{R}^d \times \mathcal{F})\) of finite positive measures on \(\mathbb{R}^d \times \mathcal{F}\) contains discrete measures written as

\[
\mu = \sum_{i \in I} w_i \delta_{x_i} \otimes \delta_{f_i},
\]

where \(w_i\) is a positive weight that can encode the collection \((x_i, f_i)\) at micro scale.
Figure 1: Top Left: Tissue from a NexCre+/-;Brn2fl/+ adult mouse mimicking a wild-type mouse with CUX1 labelling of layer ii/iii, iv and CTIP2 in layers v, vi in green. Shows sections at 2.5 μm × 50 μm² 6 tile images, 1433 × 1973 pixels; taken from Uli Mueller. Top Right: Showing the abstraction of a coarse-to-fine hierarchy μℓ−1, μℓ, μℓ+1 with fine molecular scales shown at the bottom with colors depicting F function ascending scales. Bottom: Space and function transformation shown as a composition μ T1 → ν T2 → μℓ.

As in Boltzmann modelling we describe the features statistically at a fixed spatial scale transferring our attention to their stochastic laws modelled as conditional probabilities in ℳP(F) with integral 1. For this we factor the measures into the marginal density measure on space ρ on Rd with ρ(x) = μ(dx, df), and the field of probability distributions on F conditioned on x. We use the convention dx and df as events in space and function, respectively, implying the classical factorization

\[ \mu(dx, df) = \rho(x) \, \mu_x(df), \quad \text{on } \mathbb{R}^d \times \mathcal{F}, \tag{2a} \]

with field of conditional probabilities:

\[ x \mapsto \mu_x \in \mathcal{M}_P(F). \tag{2b} \]

Dense tissue is modelled as brain measures μ with marginal density ρμ(dx) = ρ(x)dx continuous with Lebesgue measure on Rd:

\[ \mu(dx, df) = \rho(x)dx \, \mu_x(df). \]

A fundamental link between the molecular and continuum tissue can be addressed through the law of large numbers since if \((x_i, f_i)_{i \geq 0}\) is an independent and identically distributed sample drawn from law \(\mu/M\) of \(\mathbb{R}^d \times \mathcal{F}\) where \(M = \int_{\mathbb{R}^d \times \mathcal{F}} \mu(dx, df)\) is the total mass of such \(\mu\), then we have almost surely the weak convergence

\[ \mu_N := \frac{M}{N} \sum_{i=1}^{N} \delta_{x_i} \otimes \delta_{f_i} \rightharpoonup \mu. \]
Passing from the tissue scales to the molecular-cellular scales behooves us to introduce a scale-space so that empirical averages which govern it are repeatable. Figure 1 (right) depicts our multi-scale model of a brain as a sequence of measures:

$$\mu = (\mu^\ell)_{\ell=0, 1, \ldots}.$$  

(3)

Our idealization of Brainspace as a sequence of measures as depicted in Figure 1 descends from the the coarse tissue scale (top) to the finest particle representation (bottom), with color representing function \(f \in F\), and radius space-scale. Throughout the range of scales is denoted shorthand \(\ell \geq 0\) to mean \(0 \leq \ell < \ell_{\text{max}}\).

2.2 Nonlinear Transformation Model for Crossing Scales

The brain being a multi-scale collection of measures requires us to be able to transform from one scale to another. We do this by associating a scale-space to each particle feature by pairing to each measure a kernel function transforming it from a generalized function to a classical function \(\delta_z : h \mapsto h(z)\). The kernels carry resolution scales \(\sigma\) or reciprocally bandwidths, analogous to Planck’s scale.

We introduce the abstract representation of our system as a collection of descriptive elements \(z \in Z\) made from spatial and functional features. We transform our mathematical measure \(\mu(\cdot)\) on \(Z\) generating new measures \(\mu^\ell(\cdot)\) on \(Z'\) by defining correspondences via kernels \(z \mapsto k(z, dz')\), with the kernel acting on the measures transforming as

$$K[\mu](dz') = \int_Z k(z, dz') \mu(dz).$$  

(4)

This implies the particles transform as \(K[\delta_z](dz') = k(z_i, dz')\).

Figure 1 (bottom row) shows the cascade of operations, the first transforming linearly \(\mu = \sum_{i \in I} w_i \delta_{x_i} \otimes \delta_{f_i} \mapsto \nu\) on \(F\) according to \(\nu(\cdot) = \int_{R^d \times F} k_1((x, f), \cdot) \mu(dx, df)\), the second transforming nonlinearly \(\nu \mapsto \mu^\ell\) on \(F^\ell\) at scale \(\ell\):

$$\begin{cases} T_1 : & \mu \mapsto \nu(\cdot) = \sum_{i \in I} w_i k_1((x_i, f_i), \cdot), \\
& \text{with } \nu(dx, df) = \rho'(dx) \nu_x(df); \\
T_2 : & \nu \mapsto \mu^\ell(\cdot) = \int_{R^d} k_2((x, \nu_x), \cdot) \rho'(dx). \end{cases}$$  

(5)

Space resampling is defined by projecting particles to the continuum smoothly and then resampling to the computational lattice. We use a smooth resampling process defined by \(\pi : x \mapsto \pi(x, y)\) the fraction particle \(x\) shares with \(y\), giving resample kernel:

$$k_1((x, f), \cdot) = \int_{R^d} \pi(x, y) \delta_x \otimes \delta_f(\cdot) dy;$$  

(6a)

$$\nu(dx, df) = w(x)dx \nu_x(df) \quad \text{with } \begin{cases} w(x) = \sum_{i \in I} w_i \pi(x, x_i) \\
\nu_x = \sum_{i \in I} w_i / w(x) \pi(x, x_i) \delta_{f_i}. \end{cases}$$  

(6b)

Feature reduction uses singular kernels with projective feature maps from machine learning, \(\alpha \mapsto \phi(\alpha) \in F^\ell\), \(\sum_F \alpha(df) = 1\):

$$k_2((x, \alpha), \cdot) = \delta_x \otimes \delta_{\phi(\alpha)}(\cdot);$$  

(6c)

$$\mu^\ell(\cdot) = \int_{R^d} w(x) k_2((x, \nu_x), \cdot) dx = \int_{R^d} w(x) \delta_x \otimes \mu^\ell_x(\cdot) dx$$  

with \(\mu^\ell_x = \delta_{\phi(\nu_x)}\).  

(6d)
Computational lattices \( Y_j^\ell \subset \mathbb{R}^d \) are used to interpolate \( \mu = \sum_{i \in I} w_i \delta_{x_i} \otimes \delta_{f_i} \) onto scale \( \ell \), with \( \pi(x, Y_j^\ell) \) sharing its probability concentrated to the lattice centers \( y_j \in Y_j^\ell \):

\[
k_1((x, f), \cdot) = \sum_{j \in I} \pi(x, Y_j^\ell) \delta_{y_j} \otimes \delta_{f(\cdot)}
\]

\[
\mu^\ell(\cdot) = \sum_{j \in I} w_j k_2((y_j, \nu_{y_j}), \cdot) = \sum_{j \in I} w_j \delta_{\nu_{y_j}} \otimes \delta_{\phi(\nu_{y_j})}(\cdot)
\]

with

\[
\begin{align*}
    w_j &= \sum_{i \in I} w_i \pi(x_i, Y_j^\ell) \\
    \nu_{y_j} &= \sum_{i \in I} w_i / w_j \pi(x_i, Y_j^\ell) \delta_{f_i}
\end{align*}
\]

\tag{7}

2.3 Dynamical Systems Model via Varifold Action of Multi-scale Diffeomorphisms

We want to measure and cluster brains by building a metric space structure. We do this by following the original program of D’Arcy Thompson building bijective correspondence. In this setting this must be done at every scale with each scale having different numbers of particles and resolutions. We build correspondence between sample brains via dense connections of the discrete particles to the continuum at all scales using the diffeomorphism group and diffeomorphic transport. For this define the group of \( k \)-times continuously differentiable diffeomorphisms \( \varphi \in G_k \) with group operation function composition \( \varphi^a \circ \varphi^b \). For any brain \( \mu = \sum_{i \in I} w_i \delta_{x_i} \otimes \delta_{f_i} \in \mathcal{M} \), the diffeomorphisms act

\[
(\varphi, \mu) \mapsto \varphi \cdot \mu = \sum_{i \in I} w_i |d\varphi(x_i)| \delta_{\varphi(x_i)} \otimes \delta_{f_i} \ : 
\]

\tag{8a}

tissue has continuous density \( \mu = \int_{\mathbb{R}^d} \rho(x) \delta_x \otimes \delta_{f(x)} \ dx \) with action

\[
\varphi \cdot \mu = \int_{\mathbb{R}^d} \rho(x) |d\varphi(x)| \delta_{\varphi(x)} \otimes \delta_{f(x)} \ dx .
\]

\tag{8b}

Space scales are represented as the group product, \( \varphi := (\varphi^\ell)_{\ell \geq 0} \), acting component-wise with action

\[
\varphi \cdot \mu := (\varphi^\ell \cdot \mu^\ell)_{\ell \geq 0} .
\]

\tag{8c}

The \( |d\varphi(x)| \) term in the action we call the “copy and paste” varifold action. It enables the crucial property that when a tissue is extended to a larger area, the total number of its basic constituents increase accordingly with total integral not conserved, in contrast to classic measure or probability transport.

Dynamics occurs by generating the diffeomorphism as flows \( t \rightarrow (\varphi^\ell_t)_{\ell \geq 0} \), with dynamics controlled by vector fields \( t \rightarrow u_t := (u_t^\ell)_{\ell \geq 0} \) via the ordinary differential equation (ODE) at each scale satisfying

\[
\dot{\varphi}^\ell_t = u_t^\ell \circ \varphi^\ell_t .
\]

\tag{9a}

The controls are coupled by successive refinement \( v^\ell, \ell \geq 0 \), with \( u^{-1} = 0 \):

\[
u_t^\ell = u_t^{\ell - 1} + v_t^\ell .
\]

\tag{9b}

To control smoothness of the maps we force the vector fields to be elements of a reproducing kernel Hilbert spaces (RKHS’s) \( V_\ell \),
norms $\| \cdot \|_{V_\ell}$, with multi-scale $u \in V := \prod_{\ell \geq 0} V_\ell$. Each RKHS is taken to have a diagonal kernel with $g_\ell$ the Green’s function,

$$K_\ell(\cdot, \cdot) = g_\ell(\cdot, \cdot) \text{id}_d,$$

(10)

$id_d$ the $d \times d$ identity matrix; see [25] for non-diagonal kernels. Geodesic mapping flows under a control process along paths of minimum energy respecting boundary conditions. Figure 2 shows the multi-scale control hierarchy:

The multi-scale dynamical control are written

$$\varphi_t : u_t \circ \varphi_t := \left( u_\ell^t \circ \varphi_\ell^t \right)_{\ell \geq 0},$$

with observer and dynamics equation:

$$\left\{ \begin{array}{ll}
\varphi_t : \mu, & t \in [0,1] \\
\dot{\varphi}_t = u_t \circ \varphi_t, & \varphi_0 = \text{id} \end{array} \right. \text{ Observer}$$

(11)

Dynamics translates into a navigation in the orbit of brains and provides a metric distance between brains. Paths of minimum energy connecting the identity $\varphi_0 = \text{id}$ to any fixed boundary condition (BC) $\varphi_1$ where $\varphi_2$ is accessible defines the distance extending LDDMM [26] to a hierarchy of diffeomorphisms, and is a geodesic for an associated Riemannian metric [25].

The metric from $\mu_0$ to $\mu_1$ in the orbit accessible from $\mu_0$ via diffeomorphisms is the shortest length geodesic paths with BCs $\varphi_0 \cdot \mu_0 = \mu_0$ and $\varphi_1 \cdot \mu_0 = \mu_1$. This extension to multi-scale LDDMM Eqn. (23) is given in the Methods 4.3 where we discuss the smoothness required for the geodesics to define a metric and specify the optimal control problem in the state Eqn. (25).

### 2.4 Geodesic Brain Mapping via the Varifold Measure Norm

The BC matching brains is defined using measure norms with equality meaning brains are equal, with small normed difference meaning brains are similar; for particle brains $\mu = \sum_i w_i \delta_{x_i} \otimes \delta_{f_i}$; for tissue, $\mu = \int_{\mathbb{R}^d} \delta_x \otimes \mu_x \rho(x)dx$. Geodesic mapping controls the flow to minimize energy simultaneously minimizing the norm distance to the target. Every brain has a variable number of particles with no correspondence between particles. Varifold measure norms accomodate these variabilities. The varifold norm is constructed modelling the particle as elements $\delta_x \otimes \delta_f \in W^*$ the dual of an RKHS $W$ with a smooth kernel $K_W$ defining the inner product. We introduce the dual bracket notation for $h \in W, \mu \in W^*$, $(\mu \mid h) := \int_{\mathbb{R}^d \times \mathbb{R}^d} h(x, f) \mu(dx, df)$. Then we have

$$\langle \delta_{x_i} \otimes \delta_{f_i}, \delta_{x_j} \otimes \delta_{f_j} \rangle_{W^*} = \left( \delta_{x_i} \otimes \delta_{f_i} \mid K_W[\delta_{x_j} \otimes \delta_{f_j}] \right) = K_W((x_i, f_i), (x_j, f_j)).$$

(12a)
The norm-square for particle and tissue measures reduces to
\[ \|\mu\|_{W^*}^2 := (\mu \mid K_W[\mu]) = \int_{\mathbb{R}^d \times \mathcal{F}} K_W[\mu](x,f) \mu(dx,df) \]
\[ = \sum_{i,j \in I} w_i w_j K((x_i,f_i),(x_j,f_j)) - \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \rho(x) \rho(y) K((x,f),(y,f')) \mu_x(df) \mu_y(df') dx dy. \]

The hierarchical norms across the scales becomes \[ \|\mu\|_{W^*}^2 := \sum_{\ell \geq 0} \|\mu^\ell\|_{W^*}^2. \]
The optimal control \((u_t)_{0 \leq t \leq 1}\) is square-integrable under the \(V\)-norms for both particles and tissue, satisfying for \(\alpha > 0\):
\[ \min_{(u_t)_{0 \leq t \leq 1} \in L^2([0,1],\mathcal{V})} \frac{1}{2} \sum_{\ell \geq 0} \int_0^1 \|u^\ell_t - u^{\ell-1}_t\|_{V_\ell}^2 dt + \frac{\alpha}{2} \|\varphi^\ell - \mu - \mu_{\text{obs}}\|_{W^*}^2. \]

The control flows the measures \(\mu(q_t) := \varphi_t \cdot \mu\) with state processes \(t \rightarrow q_t\) and endpoint \(U(q_1) = \sum_{t \geq 0} U(q^t_1): \)
\[ q_t := \left\{ \begin{array}{l}
(x_i,t, w_i,t = w_i[d\varphi_t]/(x_i))_{i \in I} \\
(\varphi_t(x), w_t(x) = w_0(x)[d\varphi_t]/(x))_{x \in \mathbb{R}^d}
\end{array} \right., \quad \left(14a\right) \]
\[ U(q^t_1) := \frac{\alpha}{2} \|\mu(q_t) - \mu_{\text{obs}}\|_{W^*}^2 = \frac{\alpha}{2} \|\varphi_t \cdot \mu - \mu_{\text{obs}}\|_{W^*}^2. \]

The endpoint \(q_t \rightarrow U(q_1) = \sum_{t \geq 0} U(q^t_1)\) is modelled as continuously differentiable in the states.

Hamiltonian control of particles reparameterizes (13) in momentum “co-states” \(t \rightarrow p_t = (p^\ell_t, p^w_t)_{\ell \in I}\), with momentum at each scale a function of same scale dynamics (\(\ell\) index suppressed). The optimal control averages Green’s function (10) across scales \(\bar{g}_\ell := \sum_{0 \leq j \leq t} g^j\) with \(\nabla_1 \bar{g}(a,b)\) the gradient on the first variable \(a\); defining \(\varphi_{t,s} = \varphi_s \circ (\varphi_t)^{-1}\) then
\[ u^\ell_t(\cdot) = \sum_{j \geq 0 \in I} \sum_{t \in I} \left( \bar{g}_{\ell,s,j}(x^j_t,\cdot) p^\ell_{i,t} + p^w_{i,t} w^\ell_{i,t} \nabla_1 \bar{g}_{\ell,s,j}(x^j_{i,t},\cdot) \right), \quad \left(15a\right) \]
\[ p^\ell_{i,t} = (d\varphi_{t,s})^T(x_{i,t}) p^\ell_{i,s} + p^w_{i,s} w_{i,s} \int_t^1 (d\varphi_{t,s})^T(x_{i,t}) \nabla (\text{div} u_s)(x_{i,s}) ds \]
\[ p^w_{i,s} = p^w_{i,t} |d\varphi_{t,s}|(x_{i,t}). \quad \left(15b\right) \]

Control of tissue reparameterizes (13) in \(t \rightarrow p_t = (p^\ell_t(x), p^w_t(x))_{x \in \mathbb{R}^d}\); if \((p^\ell_t, p^w_t)\) is absolutely continuous it remains for \(t < 1:\)
\[ u^\ell_t(\cdot) = \sum_{j \geq 0 \in I} \int_{\mathbb{R}^d} \left( \bar{g}_{\ell,s,j}(\varphi_{t}(y),\cdot) p^\ell_{i,t}(y) + p^w_{i,t}(y) w^\ell_{i,t}(y) \nabla_1 \bar{g}_{\ell,s,j}(\varphi_{t}(y),\cdot) \right) dy; \quad \left(16a\right) \]
\[ p^\ell_t = (d\varphi_{t,s})^T \circ \varphi_t p^\ell_{i,s} + p^w_{i,t} w_{i,s} \int_t^1 (d\varphi_{t,s})^T \circ \varphi_t \nabla (\text{div} u_s) \circ \varphi_s ds \]
\[ p^w_t = p^w_{i,t} |d\varphi_{t,s}| \circ \varphi_t. \quad \left(16b\right) \]

The endpoint momentum for particles \((p^\ell_{i,1}, p^w_{i,1})_{\ell \in I}\) and dense tissue \((p^\ell_{i,1}(x), p^w_{i,1}(x))_{x \in \mathbb{R}^d}\) are given by the variation of the varifold.
norm-square endpoint matching condition:

\[ h_{q_1} := \alpha K W [\mu(q_1) - \mu_{\text{obs}}] \text{ on } \mathbb{R}^d \times \mathcal{F} . \]  

\[
\begin{align*}
\rho_{x,1}^w &= -\nabla_{x_i} U (q_1) = -u_{x_i} \nabla_{x_i} h_{q_1} (x_i, f_i) \\
\rho_{w,1}^w &= -\nabla_{w_i} U (q_1) = -h_{q_1} (x_i, f_i) \\
\rho_{x}^w (x) &= -\nabla_{\varphi(x)} U (q_1) = -u_1(x) \rho(x) \int_{\mathcal{F}} \nabla_{\varphi} h_{q_1} (\varphi_1(x), f) \mu_x (df) \\
\rho_{w}^w (x) &= -\nabla_{w(x)} U (q_1) = -\rho(x) \int_{\mathcal{F}} h_{q_1} (\varphi_1(x), f) \mu_x (df)
\end{align*}
\]  

The kernel defined in the Methods 4.2 is as separable Gaussians in space and function. The optimal control at any scale "averages" all the particle/tissue data across scales. The Methods sections 4.4, 4.5 establishes the smoothness for the Hamiltonian equations and the smoothness for the norm gradients. The Methods 4.5 establishes the variation and smoothness for the norm gradients. We emphasize that the varifold action gives the continuum problem unifying with image based LDDMM [26] such as studied by the MRI community. Taking \( I(y) \in \mathbb{R}^+ \) with \( \mu_y = \delta_{I(y)} \), \( \mu^I := \int_{\mathbb{R}^d} \delta_y \otimes \delta_{I(y)} dy \), the action becomes

\[
\varphi_1 \cdot \mu^I = \int_{\mathbb{R}^d} |d\varphi_1(y)| \delta_{I(y)} \otimes \mu_{I(y)} dy = \int_{\mathcal{R}^d} \delta_x \otimes \delta_{I \otimes \varphi_1^{-1}(x)} d\mu .
\]  

2.5 MRI and Digital Pathology for Tau Histology in Alzheimer’s

2.5.1 Bayes Segmentation of MRI

Figure 3 shows the multi-scale data from the clinical BIOCARD study [6] of Alzheimer’s disease within the medial temporal lobe [7, 8, 27]. The top (left) panel shows clinical magnetic resonance imaging (MRI) with the high-field 200 \( \mu m \) MRI scale (right) shown depicting the medial temporal lobe including the collateral sulcus and lateral bank of the entorhinal cortex. Bayes classifiers

Figure 3: Top: Medial lobe at 1 mm and high-field 200 \( \mu m \) MRI; Mai-Paxino atlas section of the MRI with hippocampus and entorhinal cortex; right shows the Bayes compartments. Bottom: Alzheimer 4 \( \mu m \) Tau histology (red) from section depicted via high-field MRI (top row); right shows detected Tau particles in similar section. Box depicts trans-entorhinal region from top row. Right two panels shows mean particle size and standard deviation at \( \mu m \) and tissue scales; deep red color denotes 80 \( \mu m^2 \) Tau area.
for brain parcellation performs feature reduction as a key step for segmentation at tissue scales [28]. Feature reduction maps the distribution on gray levels \( \mathcal{F} = [0, 255] \) to probabilities on \( N \) tissue types, defined by the integration over the decision regions \( \theta_n \subset [0, 255] \):

\[
\phi_n(\mu_x) = \int_{\theta_n} \mu_x(df) = p_n, \quad n = 1, \ldots, N.
\]

Figure 3 (top row, right) depicts a Bayes classifier for gray, white and cerebrospinal fluid compartments generated from the temporal lobe high-field MRI section corresponding to the Mai-Paxinos section (panel 3).

### 2.5.2 Gaussian Scale-Space Resampling of Tau Histology:

For histology at the molecular scales the measure encodes the detected Tau and amyloid particles \( \mu = \sum_i w_i \delta_{x_i} \otimes \delta_{f_i} \) for fine scale particles with function the geometric features \( \mathcal{F} = \mathbb{R}^+ \). Figure 3 (bottom row) shows the detected Tau particles as red dots at 4\( \mu \)m with feature reduction done via moments on tau. We use computational lattices to interpolate between scales reapportioning to the particles with function the geometric features \( F \).

For histology at the molecular scales the measure encodes the detected Tau and amyloid particles \( \mu = \sum_i w_i \delta_{x_i} \otimes \delta_{f_i} \) for fine scale particles with function the geometric features \( \mathcal{F} = \mathbb{R}^+ \). Figure 3 (bottom row) shows the detected Tau particles as red dots at 4\( \mu \)m with feature reduction done via moments on tau. We use computational lattices to interpolate between scales reapportioning to the lattice centers \( (\delta_{y_j}, y_j \in Y_j \subset \mathbb{R}^d)_{j \in I}, \) via Gaussian resampling \( x \mapsto \pi(x, Y_j) \) from 4\( \mu \)m. Feature reduction maps to the first two moments at the tissue scale of mean and variance of particle size \( \mathcal{F}^t \subset \mathbb{R}^2 \):

\[
\pi(x_1, Y_j) = \frac{1}{(2\pi\sigma^2)^{1/2}} \int_{Y_j} e^{-|y-x|^2/2\sigma^2} dy
\]

\[
\mu^t = \sum_{j \in I} w^j \delta_{y_j} \otimes \delta_{\phi(\nu_j)} \text{ with } \nu_j = \frac{1}{w^j} \sum_{i \in I} w_i \pi(x_i, Y_j) \delta_{f_i} ; \phi(\nu_j) = \left( \int_{\mathcal{F}} f \nu_{y_j} (df), \int_{\mathcal{F}} (f^2) \nu_{y_j} (df) \right).
\]

The bottom row of Figure 3 (right two panels) shows the mean and variance of the particle size reconstructed from the 4\( \mu \)m scale. The mm scale depicts the global folding property of the tissue. The color codes the mean tissue Tau area as a function of position at the tissue scales with deep red color denoting 80 \( \mu \)m\(^2\) maximum Tau area for the detected particles.

### 2.6 Cellular Neurophysiology: Neural Network Temporal Models

Single unit neurophysiology uses temporal models of spiking neurons with a “neural network” \( \mu = \sum_i \delta_{x_i} \otimes \delta_{f_i} \), \( x_i \in \mathbb{R}^d, f_i \in \mathcal{F} \) taking each neuron \( x_i \) modelled as a counting measure in time \( N_i(t), t \geq 0 \) with the spike times the feature \( f_i = (f_k)_{1 \leq k \leq n_i} \):

\[
\mathcal{F} = \{ f = (t_k)_{1 \leq k \leq n_i} : n_f \geq 1, t_1 < t_2 \cdots < t_{n_f} \}.
\]

The Poisson model with intensity \( \lambda(t), t \geq t_0 [10] \) has probabilities \( \text{Pr}(N_i(t) = n) = \frac{(\int_0^t \lambda(s) ds)^n}{n!} e^{-\int_0^t \lambda(s) ds} \).

Post-stimulus time (PST) [29] and interval histograms are used to examine the instantaneous discharge rates and inter-spike interval statistics [30]. The interval histogram abandons the requirement of maintaining the absolute phase of the signal for measuring temporal periodicity and phase locking. Synchrony in the PST is measured using binning \( [b_i, b_{i+1}), i = 1, \cdots, B \) and Fourier transforms, \( j = \sqrt{-1} I:

\[
\phi^{PST}_n(\mu_x) = \sum_i e^{j\omega_n i} \int_{\mathcal{F}} \frac{1}{n_f} \sum_{k=1}^{n_f} 1_{[b_i, b_{i+1})}(t_k) \mu_x(df), \quad n = 0, 1, \ldots.
\]

The \( n = 0 \) frequency computes integrated rate; each phase-locked feature is complex \( \phi_n \in \mathbb{C} \).
2.7 Scale Space Resampling of RNA to Cell and Tissue Scales

Methods in spatial-transcriptomics which have emerged for localizing and identifying cell-types via marker genes and across different cellular resolutions [4, 31–34] presents the opportunity of localizing in spatial coordinates the transcriptionally distinct cell-types. Depicted in Figure 4 are the molecular measurements at the micron scales with MERFISH [35] at three different scales.

The molecular measures represent RNA locations \( p_{\delta x_i} \) with sparse RNA features \( \mu_R = \sum_{i \in I_R} p_{\delta x_i} \). Crossing to cells \( p_{\delta y_j} \) resamples to the cell centers \( p_{\delta y_j} \) partitioning into closest particle subsets defined by the distance \( d(x_i, Y_j) \) of particle \( x_i \) to cell \( Y_j \), accumulating as nonsparse mixtures of RNA within the closest cell. The feature is the conditional probability of the 17 cell-type vector \( F_C \) following [35].

For this example the conditional probabilities on the RNA feature vector was found using principle components followed by Gaussian mixture modelling on \( \nu_{y_j} \) following [35].

Resampling to tissue \( (\delta z_k, z_k) \in Z_k \subset R^2 \) uses Gaussian rescaling. The new feature vector becomes the probability of the cell at any position being one of 10 tissue types \( F_T \subset [0, 1]^{10} \). The probability of tissue type is calculated using 10-means clustering on the cell probabilities. The distance for 10-means clustering is computed using the Fisher-Rao metric [36] between...
the empirical feature laws $\nu_{zk}^T$. The 10-means clustering partitions the cell-type feature space $\cup_{1\leq t \leq 10} F_t = F^C$ into new features:

$$\pi_\sigma(y_j, Z_k) = \frac{1}{2\pi\sigma^2} \int_{Z_k} e^{-|z-y_j|^2/2\sigma^2} dz.$$  \hspace{1cm} (21a)

$$\mu^T = \sum_{k \in I_s} w_k^T \delta_{z_k} \otimes \delta_\phi(\nu_{zk}^T)$$ with

$$\nu_{zk}^T = \frac{1}{w_k} \sum_{j \in I_c} w_j^C \pi_\sigma(y_j, Z_k) \delta f_j^z \hspace{1cm} \text{and} \hspace{1cm} \phi^T(\nu_{z_j}^T) = \left( \Pr\{ T_{z_j} = t \mid \nu_{z_j}^T \} \right)_{t=1,\ldots,10}.$$  \hspace{1cm} (21b)

For this example the conditional probabilities on the tissue-type feature vector are computed as indicator functions under the 10-mean clustering.

Figure 4 (bottom left panels) shows the RNA forming $\mu^R$ depicted as colored markers corresponding to the different gene species (bar scale 1, 10 microns). Panel 2 shows the feature space of 17 cell types making up $\mu^C$ associated to the maximal probability in the PCA projection from a classifier based on the mixtures of RNA at each cell location. Bottom right shows the 10 tissue features of the 10-means procedure. In both scales probabilities are concentrated on single classes.

### 2.8 Geodesic Mapping for Spatial Transcriptomics, Histology and MRI

#### 2.8.1 Navigation Between Sections of Cell Identity in Spatial Transcriptomics.

Figure 5 (top panels 1,2) shows sections (left column) from [4] depicting neuronal cell types via colors including excitatory cells eL2/3 (yellow), eL4 (orange), red eL5 (red), inhibitory cells ST (green), VIP (light blue), each classified via high dimensional gene expression features vectors via spatial transcriptomics.

![Figure 5](image)

Figure 5: Left column: Shows neuronal cell types [4] for two sections: top template, bottom target. Right column: Panels show $\phi_t \cdot \mu$ for $t = 0, 0.5, 1.0$ of neuronal cells mapping cell types (top) and entropy (bottom). The varifold norm $\sigma_\mu^T = \sigma_\phi^T = 10$ pix, 50 pix with the feature a one hot encoding of cell type. The vector fields have RKHS norm induced by the differential operator $L_\ell := ((1 - (\alpha^\ell)^2 \nabla^2)\text{id}_d)^2$, $\alpha^\ell = 5$ pix, 50 pix with full width at half maximum (FWHM) of the Green’s kernel $FWHM = 20$ pix, 145 pix, for $\ell = 1, 2$, respectively.

The measure on cell types $\mu^C = \sum_i w_i \delta_{x_i} \otimes \delta_f^C$, $F^C = \{\text{cell types}\}$ crosses to atlas tissue scales using $\pi_\sigma$ in $\mathbb{R}^2$ of Eqn. (21a)
with feature reduction expectations of moments, \( \mathcal{F}_T = \{ \text{size, variance, entropy} \}:

\[
\mu_T = \sum_{j \in J} w_j \delta y_j \otimes \delta \phi(\nu_{y_j})
\]

with

\[
\phi(\nu_{y_j}) = \left( \sum_{x} f \nu_{y_j}(f), \sum_{x} (f^2 \nu_{y_j}(f), -\sum_{x} \nu_{y_j}(f) \log \nu_{y_j}(f) \right).
\]

The right panels of Figure 5 shows the tissue scale features associated to the cell identity depicted and the entropy. The right panels shows the results of transforming the neuronal cells depicting the cell type (top row) and entropy feature (bottom row). The entropy is a measure of dispersion across the cell identities given by the expectation of the log probability function with zero entropy meaning the space location feature distribution \( \nu \) has all its mass on 1 cell type. Geodesic mapping enforces vector field smoothness via differential operators specifying the norms in the RKHS \( \|v\|^2_{\mathcal{L}_q} := (L_{\ell} v \mid v) \) with \( L_{\ell} := ((1 - (\alpha^\ell)^2 \nabla^2) \rho \downarrow \alpha^\ell)^2 \).

### 2.8.2 Navigation Between Sections of Histology.

Figure 6 (rows 1 and 2) shows navigation between the cortical folds of the 4 \( \mu m \) histology. Shown in panel 1 is a section showing the machine learning detection of the Tau particles. Columns 2, 3, and 4 depict the template, mapped template and target showing the mathematical measure representation of the perirhinal cortex constructed from the positions and sizes at the 4 \( \mu m \) scale (top row) and reconstruction using Gaussian resampling onto the tissue scale (bottom row). The color codes the mean of \( \mu_{x} \) representing Tau area as a function of position at the tissue scales with deep red of 80 \( \mu m^2 \) of Tau area the maximum value. The gradients in tau tangle area between superficial and deep layers is apparent with the deep red as high as 80 \( \mu m^2 \) for the innermost tissue fold. The bottom left panel shows the vector field encoding of the geodesic transformation of the of the perirhinal cortex mapping between the two sections in column 1. The narrowing of the banks of the perirhinal cortex is exhibited at the tissue scale for motions order 1000 \( \mu m \) (brightness on scale bar).

Figure 6 (row 3), shows the collateral sulcus fold at the boundary of the trans-entorhinal cortex region transforming based on the normed distances between sections with deformation motions 1000 \( \mu m \) in size. Shown is the micron scale depicting the transformation of the gyrus with the color representing the entropy of the particle identity distribution.

### 2.8.3 Mapping Digital Pathology from Histology to MRI Scales

All of the examples thus far have created the multi-scale data generated using the resampling kernels from the finest scales. As illustrated in our early figures much of the data is inherently multi-scale, with the measurement technologies generating the coarse scale representations. Shown in Figure 7 is data illustrating our Alzheimer’s study of post mortem MR images that are simultaneously collected with amyloid and Tau pathology sections. MR images have a resolution of approximately 100 \( \mu m \), while pathology images have a resolution of approximately 1 \( \mu m \). For computational purposes the MRI template and target images were downsampled to 759 and 693 particles, respectively with the tau tangles downsampled to 1038 and 1028 particles, respectively. We treated every pixel in the MR image as a coarse scale particle with image intensity as its feature value Eqn. (18), and every detected tau tangle as a fine scale particle with a constant feature value, and performed varifold matching to align to neighboring sections. The endpoint representing the two scales is

\[
U = \frac{1}{2} \sum_{\ell=1,2} \| \mu_{\ell, \text{temp}} - \mu_{\ell, \text{obs}} \|^2_{\mathcal{L}_q}.
\]

For each scale norm we use a varifold kernel given by the products of Gaussian distributions with the varifold measure norm Eqn. (22) at each scale. For the MRI scale, the weights are identically \( w = 1 \) with the function component given by the MRI image value: for the tau particles there is no function component making \( f = g \) with the kernel 1 for all values.

The top two rows of Figure 7 shows the imaging data for both sections. The bottom row shows the transformed template image at the fine scale. The high resolution mapping carries the kernels across all the scales as indicated by the geodesic equation (15a). Notice the global motions of the high resolution of the fine particles.
Figure 6: Rows 1,2: Second section 4µm histology (similar to Figure 3) with box depicting perirhinal cortex. Columns 2, 3 and 4 show template, mapped template and target at the molecular (top) and tissue scales (bottom) showing the first moments of Tau size on the perirhinal cortex; saturated red color denotes 80 µm² Tau area. Mapped template shows narrowing 1000 µm of perirhinal sulcus. Second row left panel depicts vector field encoding of the geodesic mapping with associated scale bar. Row 3: Geodesic navigation \( \varphi_t \cdot \mu \) of collateral sulcus (Figure 3, box) showing 1000 µm widening of folds for molecular and tissue scales depicting the mean transforming. The vector field mappings have RKHS norm induced by the differential operator with \( L_\ell := ((1 - (\alpha^2)^2 \nabla^2) \text{id})^2 \), with \( \alpha \) with \( FWHM = 580\mu m, 3300\mu m \), for \( \ell = 1, 2 \) respectively. The varifold norm has \( \sigma^\ell_z = 800 \mu m, 160 \mu m \) for the mappings.

3 Discussion

Computational anatomy was originally formulated as a mathematical orbit model for representing medical images at the tissue scales. The model generalizes linear algebra to the group action on images by the diffeomorphism group, a non-linear algebra, but one that inherits a metric structure from the group of diffeomorphisms. The formulation relies on principles of continuity of medical images as classical functions, generalizing optical flow and advection of material to diffeomorphic flow of material, the material represented by the contrast seen in the medical imaging modality such as fiber orientation for diffusion tensor imaging, and or bold contrast for gray matter content. Unifying this representation to images built at the particle and molecular biological scale has required us to move away from classical functions, to the more modern 20th century theory of non-classical generalized functions. Mathematical measures are the proper representation as they generally reflect the property that probes from molecular biology associated to disjoints sets are additive, the basic starting point of measure theory. Changing the model from a focus on groups acting on functions to groups acting on measures allows for a unified representation that has both a metric structure at the finest scales, as well as a unification with the tissue imaging scales.

The brain measure formulation, carries with it implicitly the notion of scale-space, i.e. the existence of a sequence of pairs across scales, the measure representation of the brain and the associated scale-space reproducing kernel Hilbert space of functions which correspond to the probing measurement technologies. As such part of the prescription of the theory is a method for crossing...
scales and carrying information from one scale to the other. Important to this approach is that at every scale we generate a new measure, therefore the recipe of introducing “measure norms” built from RKHS’s for measuring brain disparity is universal across the hierarchy allowing us to work simultaneously with common data structures and a common formalism. Interestingly, the measure norms do not require identical particle numbers across brains in brain space at the molecular scales.

The key modelling element of brain function is that the conditional feature probability is manipulated from the quantized features to the stochastic laws. These are the analogues of the Boltzman distributions generalized to the complex feature spaces representing function. As they correspond to arbitrary feature spaces not necessarily Newtonian particles, we represent them simply as empirical distributions on the feature space, with the empirical measure constructed from the collapse of the fine scale to the resampled coarse scale. To model rescaling through scale-space explicitly, the two kernel transformation are used allowing us to retrieve the empirical averages represented by the determinism of the stochastic law consistent with our views of the macro tissue scales. This solves the dilemma that for the quantized atomic and micro scales cell occurrence will never repeat, i.e. there is zero probability of finding a particular cell at a particular location, and conditioned on finding it once it will never be found again in the exact same location in another preparation. The properties that are stable are the probability laws with associated statistics that may transfer across organisms and species.

Importantly, our introduction of the $|d\varphi(x)|$ term in the action enables the crucial property that when a tissue is extended to
a larger area, the total number of its basic constituents should increase accordingly and not be conserved. This is not traditional measure transport which is mass preserving which is not a desirable feature for biological samples. Rather we have defined a new action on measures that is reminiscent of the action on $d$-dimensional varifolds [37, 38]. We call this property “copy and paste”, the notion being that the brain is built on basic structuring elements that are conserved.

We believe that many different diffeomorphism methods can be used at multi-scale. The proper coupling of the vector fields would have to be derived to determined how the different scales mix as we have done for the multi-scale LDDMM formulation here based on the total kinetic energy. Also, successive refinement for the small deformation setting has been introduced in many areas associated to multigrid and basis expansions. The notion of building multi-scale representation in the large deformation LDDMM setting was originally explored F. Riesser et al. [39] in which the kernels are represented as a sum of kernels and Sommer et al. [40] in which the kernel is represented as vector bundles. In their multi-scale setting there is a post-optimization decomposition in which the contribution of the velocity field into its different components can then each be integrated. In that multi-scale setting the basic Euler-Lagrange equation termed EPDIFF remains that of LDDMM [41]. In the setting proposed here we separate the scales before optimization via the hierarchy of layered diffeomorphisms and use a multi-scale representation of the brain hierarchy itself which is directly associated to the diffeomorphism at that scale. This gives the fundamental setting of the product group of diffeomorphisms with the Euler-Lagrange equation corresponding to the sequence of layered diffeomorphisms for multi-scale LDDMM [25].

In terms of the efficiency of the multi-scale representation, Figure 6 shows clearly the power of the multi-scale diffeomorphism transferring micro and macro scale properties of the brain as well as the power of crossing of scales transferring information consistently from particles to the continuum. What is striking in Figure 6 is the deformations of the particles at the micron scale results in consistent motions of the cortical surface as a smooth global manifold. Also, the functional feature being transferred consistently from the particles via $T_1, T_2$ of Eqns. (5) shows the clear pattern of the functional Tau-particle size being layered a property that is hardly noticed at the molecular scale but is clearly associated to the coarse scale.

The aggregation across scales from particle to tissue scales on lattices provides the essential link to inference on graphs. It is natural for these aggregated features with associated conditional probability laws to become the nodes in Markov random field modelling for spatial inference; see examples in spatial transcriptomics and tissue segmentation [42]. Building neighborhood relations as conditional probabilities between lattice sites from which global probabilities laws are constructed with the Hammersley-Clifford theorem links us to Grenander’s metric pattern theory formalisms with the atoms and conditional laws $\langle x_i, \mu_{x_i} \rangle_{i \in \mathcal{I}}$ at any scale playing the roles of the generators.

4 Materials and Methods

4.1 Experimental and Technical Design

The objective of this research is to unify the molecular representations of spatial transcriptomics and cellular scale histology with the tissue scales of Computational Anatomy for brain mapping. To accomplish this we designed a mathematical framework for representing data at multiple scales using generalized functions, and mapping data using geodesic flows of multiple diffeomorphisms. We illustrate the method using several examples from human MRI and digital pathology, as well as mouse spatial transcriptomics.

4.2 Gaussian Kernel Varifold Norm

Our varifold norm construction models the measures as elements of a Hilbert space $W^*$ which is dual to an RKHS $W$ with a kernel $K_W$. Using the dual bracket notation for $h \in W, \mu \in W^*$, $(\mu \mid h) := \int_{\mathbb{R}^d \times \mathcal{F}} \mu(dx, df)h(x, f)$, the norm becomes
the integration against the kernel Eqn. (12b) written as \( \| \mu \|_{W^*}^2 := (\mu \mid K_W[\mu]) \); the multi-scale norm is given by \( \| \mu \|_{W^*}^2 := \sum_{\ell \geq 0} \| \mu^{(\ell)} \|_{W^*}^2 \).

To ensure the brain measures are elements of \( W^* \) dual to the RKHS \( W \), the kernel \( K_W \) is chosen to densely and continuously embed in bounded continuous functions \( C_b(\mathbb{R}^d \times \mathcal{F}, \mathbb{R}) \) so that the signed measure spaces \( M_+(\mathbb{R}^d \times \mathcal{F}) \) of brains are continuously embedded in the dual spaces \( C_b(\mathbb{R}^d \times \mathcal{F}, \mathbb{R})^* \). An example is the Gaussian kernel (22) which satisfies this condition, the kernel taken as separable Gaussians with \( | \cdot | \) Euclidean distance:

\[
k_W((x, f), (y, g)) := \exp \left( -\frac{1}{2\sigma_x^2} |x - y|^2 \right) \exp \left( -\frac{1}{2\sigma_f^2} |f - g|^2 \right).
\]  

(22)

For measures \( \mu^a = \sum_{i \in I} u_i^a \delta_{x_i^a} \otimes \delta_{f_i^a} \), \( \mu^b = \sum_{j \in J} w_j^b \delta_{x_j^b} \otimes \delta_{f_j^b} \) gives

\[
\| \mu^a - \mu^b \|_{W^*}^2 = \sum_{i, j \in I, J} u_i^a w_j^b \exp \left( -\frac{1}{2\sigma_x^2} |x_i^a - x_j^b|^2 \right) \exp \left( -\frac{1}{2\sigma_f^2} |f_i^a - f_j^b|^2 \right) - 2 \sum_{i \in I, j \in J} u_i^a w_j^b \exp \left( -\frac{1}{2\sigma_x^2} |x_i^a - x_j^b|^2 \right) \exp \left( -\frac{1}{2\sigma_f^2} |f_i^a - f_j^b|^2 \right) + \sum_{j, j' \in J} w_j^b w_j'^b \exp \left( -\frac{1}{2\sigma_x^2} |x_j^b - x_{j'}^b|^2 \right) \exp \left( -\frac{1}{2\sigma_f^2} |f_j^b - f_{j'}^b|^2 \right).
\]

For data with position information but no features (tau tangle locations), each \( f_i, f_j \) is constant with exponential terms all 1.

### 4.3 The Riemannian Distance Metric on the Multi-Scale Group

The diffeomorphism group acts on the hierarchy \( \varphi \cdot \mu \) component-wise Eqn. (8c) with the multi-scale group the product

\[
G_k = G_k \times \cdots \times G_k,
\]

\( \ell_{\text{max}} \) times

with elements \( \varphi \in G_k \) satisfying the law of composition component-wise \( \varphi \circ \varphi' = (\varphi^k \circ \varphi'^k) \). The group \( G_k \) supporting \( k \)-derivatives of the diffeomorphisms builds from \( \mathcal{C}_0^k(\mathbb{R}^d, \mathbb{R}^d) \) a space of \( k \)-times continuously differentiable vector fields vanishing at infinity and its partial derivatives of order \( p \leq k \) intersecting with diffeomorphisms with 1-derivative:

\[
G_k = (\text{id} + \mathcal{C}_0^k(\mathbb{R}^d, \mathbb{R}^d)) \cap \text{Diff}^1(\mathbb{R}^d, \mathbb{R}^d).
\]

Dynamics occurs via group action generated as a dynamical system in which the multi-scale control \( t \mapsto u_t := (u_t^\ell)_{\ell \geq 0} \) flows the hierarchy \( t \mapsto \varphi_t \) satisfying \( \dot{\varphi}_t = u_t \odot \varphi_t \) of (9a). The control is in the product \( V = \prod_{\ell \geq 0} V_{\ell} \), each space an RKHS with norm-square \( \| \cdot \|_V^2 = \sum_{\ell \geq 0} \| \cdot \|_{V_{\ell}}^2 \) selected to control the smoothness of the vector fields. The hierarchy of spaces are organized as a sequence of continuous embeddings, \( V_0 \hookrightarrow \cdots \hookrightarrow V_{\ell_{\text{max}}} \), where \( V_{\ell_{\text{max}}} \) is an additional layer containing the others with \( V_{\ell_{\text{max}}} = \mathcal{C}_0^m(\mathbb{R}^d, \mathbb{R}^d) \) defined as a space of \( m \)-times continuously differentiable vector fields vanishing at infinity as well all its partial derivatives of order \( p \leq m \).

The hierarchy is connected via successive refinements \( u^\ell = u^{\ell-1} + v^\ell, u^0 = v^0 \) expressed via the continuous linear operator
A : V → V with v = Au. The control process (u_{t})_{0 \leq t \leq 1} \in L^2([0, 1], V) has finite square-integral with total energy

$$E_A((u_{t})_{0 \leq t \leq 1}) = \frac{1}{2} \int_{0}^{1} \|Au_{t}\|_{V}^{2} dt .$$

Optimal curves which minimize the integrated energy \(E_A\) between any two fixed boundary conditions (BC) \(\varphi_0 = \text{Id}\) and \(\varphi_1\) which is accessible with a path of finite energy extends the LDDMM setting [26] to a hierarchy of diffeomorphisms and describes a geodesic for an associated Riemannian metric and multi-scale LDDMM [25] on \(G_k:\)

$$d_{G_k}(\text{Id}, \varphi_1)^2 := \min_{\varphi_t \in \mathcal{L}(\{0, 1\}, V); \varphi_0 = \text{Id}, \varphi_1} \int_{0}^{1} \|Au_{t}\|_{V}^{2} dt . \quad (23)$$

Existence of solutions for minimizers over \(u\) of (23) when \(d_{G_k}(\text{Id}, \varphi_1)^2\) is finite can be established when \(m \geq k \geq 1\).

### 4.4 Geodesic Multi-Scale LDDMM via Hamiltonian Control

The shape of Brainspace is given by its geodesics. We use Hamiltonian control to generate the geodesics.

The Hamiltonian method reduces the parameterization of the vector field to the dynamics of the particles that encode the flow of states (14a). We write the dynamics explicitly as a linear function of the control, which defines the flow of the measures:

$$\dot{q}_{t} = \xi_{q_{t}}(u_{t}) := (\xi_{q_{t}}^{l}(u_{t}^{l}))_{l \geq 0}, \text{ with } \dot{q}_{t} = \left\{ \begin{array}{ll}
(u_{t}(x_{i,t}), w_{i,t,}, \text{div}u_{t}(x_{i,t}))_{i \in I} & \\
(u_{t}(\varphi_{t}(x)), w_{t}(x)\text{div}u_{t}(\varphi_{t}(x)))_{x \in \mathbb{R}^d}
\end{array} \right.;$$

$$\mu(q_{l}) := (\mu(q_{l}^{l}))_{l \geq 0}, \text{ with } \mu(q_{l}) := \left\{ \begin{array}{ll}
\sum_{i \in I} w_{i,t,} \delta_{x_{i,t}} \otimes \delta_{f_{i}} & \\
\int_{\mathbb{R}^d} \rho(x) w_{t}(x) \delta_{\varphi_{t}(x)} \otimes \delta_{f(x)} dx
\end{array} \right. . \quad (24a)$$

The control problem satisfying (13) reparameterized in the states becomes, for \(\alpha > 0\):

$$\min_{(u_{t})_{0 \leq t \leq 1} \in L^2([0, 1], V); \varphi_0 = \text{Id}, \varphi_1} \frac{1}{2} \int_{0}^{1} \|Au_{t}\|_{V}^{2} dt + \frac{\alpha}{2} \|\mu(q_{l}) - \mu_{obs}\|_{W^*_\alpha}^{2} . \quad (25)$$

Hamiltonian control for particles and tissues introduces the co-states \(t \rightarrow p_{t}\) via the Hamiltonian

$$\mathcal{H}(q, p, u) := (p | \xi_{q}(u)) - \frac{1}{2} \|Au\|_{V}^{2} , \quad (26a)$$

with \((p | \xi_{q}(u)) = \begin{cases} \sum_{l \geq 0} \left( \sum_{i \in I} \langle p^{x_{i,t}}(x_{i,t}), u_{t}(x_{i,t}) \rangle_{\mathbb{R}^d} + p^{w_{i,t}} w_{i,t} \text{div}u_{t}(x_{i,t}) \right) & \\
\sum_{l \geq 0} \int_{\mathbb{R}^d} \langle (p^{\varphi_{t}(x), u}(x_{i,t})(\varphi_{t}(x)), u_{t}(x_{i,t})(\varphi_{t}(x)) \rangle_{\mathbb{R}^d} + p^{w_{i,t}}(x) w_{t}(x) \text{div}u_{t}(\varphi_{t}(x)) \rangle_{\mathbb{R}^d} \right) dx . \end{cases} \quad (26b)$$

Under the assumption \(V_{t_{\text{max}}} \hookrightarrow \mathcal{C}_{0}^{k+2}([\mathbb{R}^d, \mathbb{R}^d])\) the Pontryagin maximum [22] with \(k \geq 1\) gives the optimal control for every \(\ell:\)

$$\begin{align*}
\dot{p}^{l} &= -\partial_{q^{l}} \mathcal{H}(p, q, u) \\
\dot{q}^{l} &= -\partial_{u^{l}} \mathcal{H}(p, q, u) \\
\dot{p}^{l} &= -\partial_{u^{l}} \mathcal{H}(p, q, u) .
\end{align*} \quad (27)$$
Statement 1.

Geodesics of Particles. Assume that \( V_{t_{\text{max}}} = \mathcal{C}_0^m(\mathbb{R}^d, \mathbb{R}^d) \) with \( m \geq k + 2 \), \( k \geq 1 \). If \((u_t)_{0 \leq t \leq 1}\) is a solution of the optimal control problem (25) then there exists time-dependent co-state \((t \mapsto p_t) = (p^l_{i,t})_{i \in I, t \leq t_{\text{max}}}\) for all \( t \) satisfies
\[
\begin{aligned}
\dot{q}_i^l(t) &= (u^l(x^l_{i,t}), w^l_{i,t}) \text{div } u^l_i(x^l_{i,t})) \\
\dot{p}^x_{i,t}^l &= -(du^l)^T(x^l_{i,t}) p^x_{i,t}^l - p^w_{i,t}^l w^l_{i,t} \nabla (\text{div } u^l_i(x^l_{i,t})) \\
\dot{p}^w_{i,t}^l &= -p^w_{i,t}^l \text{div } u^l_i(x^l_{i,t})
\end{aligned}
\tag{28a}
\]

The optimal controls for particles satisfying \( \partial_{u^l} \mathcal{H}(p, q, u) = 0 \) for any \( t \geq 0 \) and \( v = Au \) \((v_{t_{\text{max}}} = 0)\) is given by (15a).

Geodesics of Tissue. Assume that \( V_{t_{\text{max}}} = \mathcal{C}_0^m(\mathbb{R}^d, \mathbb{R}^d) \) with \( m \geq k + 2 \), \( k \geq 1 \). If \((u_t)_{0 \leq t \leq 1}\) is a solution of the optimal control problem (25) then there exists time-dependent co-state \((t \mapsto p_t) = (p^l_{i}(x))_{x \in \mathbb{R}^d, t \leq t_{\text{max}}}\) for all \( t \) satisfies
\[
\begin{aligned}
\dot{q}^l_i(x) &= (u^l_i(\varphi^l_i(x)), w^l_i(x) \text{div } u^l_i(\varphi^l_i(x))) \\
\dot{p}^x_{i,t}^l &= -(du^l)^T(\varphi^l_i(x)) p^x_{i,t}^l - p^w_{i,t}^l w^l_{i,t} \nabla (\text{div } u^l_i(\varphi^l_i(x))) \\
\dot{p}^w_{i,t} &= -p^w_{i,t} \text{div } u^l_i(\varphi^l_i(x))
\end{aligned}
\tag{28b}
\]

The optimal controls for tissues satisfying \( \partial_{u^l} \mathcal{H}(p, q, u) = 0 \) for any \( t \geq 0 \) and \( v = Au \) \((v_{t_{\text{max}}} = 0)\) is given by (16a).

See Appendix (A) for proof of differential equations.

Statement 2 (Integral equations for Hamiltonian-Momentum). Assuming \( q \rightarrow U(q) := \frac{a}{2} \| \mu(q_1) - \mu_{\text{obs}} \|^2_{W^*} \) is \( \mathcal{C}^1 \) in \( q \), the geodesic co-state flowing from \( t = 0 \) satisfies
\[
\begin{aligned}
\dot{p}^x_{i,t} &= (d\varphi^l_0)^T(x^l_{i,t}) p_{i,t,0}^x - p^w_{i,t,0} w^l_{i,t} \int_0^t (d\varphi^l_{s,t})^T(x^l_{i,t}) \nabla (\text{div } u^l_s(x^l_{i,s}))ds \\
\dot{p}^w_{i,t} &= p^w_{i,t} (d\varphi^l_0)(x^l_{i,t}) \\
\dot{p}^x_{i,t} &= (d\varphi^l_1)^T(x^l_{i,t}) p_{i,t,1}^x + p^w_{i,t,1} w^l_{i,t} \int_0^1 (d\varphi^l_{s,t})^T(x^l_{i,t}) \nabla (\text{div } u^l_s(x^l_{i,s}))ds \\
\dot{p}^w_{i,t} &= p^w_{i,t} (d\varphi^l_1)(x^l_{i,t})
\end{aligned}
\tag{29}
\]

As proven in Appendix (B) the integral Eqns. (29) corresponding to (15b), (16b) satisfy (28a), (28b), respectively.

4.5 Gradients of the Endpoint Varifold Matching Norm

This gradients of the matching endpoints require us to compute the variation \( \frac{d}{dz} U(\varphi_1(z), w_1(z))|_{z=0} \) requiring \( U \) as a function of \( \varphi \) is \( \mathcal{C}^1 \) for \( \varphi \in G_k \subset \text{Diff}_k^1(\mathbb{R}^d, \mathbb{R}^d) \); this requires \( k \geq 1 \).

The gradients of (17b) are efficiently rewritten using the state \( q_t = (x_{i,t}, w_{i,t})_{i \in I} \) to define the norm-square in terms of \( h_q \) continuously differentiable in \( x \) and bounded \( \mathcal{C}^{1,0}_0 \) giving the smooth endpoint term determining the smooth gradients:
\[
h_q(y, z) := \alpha K_{W}[\mu(q_1) - \mu_{\text{obs}}](y, z), (y, z) \in \mathbb{R}^d \times \mathcal{F}.
\tag{30}
\]
We take the variation \( \frac{d}{dz} U(q(z)) \) varying each term \( q_1(\varepsilon) = (x_{i,1} + \varepsilon \psi_i^T, w_{i,1} + \varepsilon \psi_i^w) \) with dependence on \( \ell \)-scale implied.

\[
\frac{d}{dz} \mu(q_1(\varepsilon)) - \mu_{obs} \left|_{\ell = 0} \right. = \frac{d}{dz} \mu(q_1(\varepsilon)) - \mu_{obs} \left|_{\ell = 0} \right. = \frac{d}{dz} (\mu(q_1(\varepsilon)) - \mu_{obs}) \left|_{\ell = 0} \right. = \sum_{i \in I} \frac{d}{dz} (\langle w_{i,1} + \varepsilon \psi_i^w \rangle h_{q_1}(x_{i,1} + \varepsilon \psi_i^T, f_i) \rangle_{\ell = 0}.
\]

The gradients (17c) for the tissue continuum has \( \mu = \int_{\mathbb{R}^d} \rho(x) \delta_x \otimes \mu_x dx \) and \( \mu(q_1) = \int_{\mathbb{R}^d} w_1(x) \rho(x) \delta_{\varphi_1(x)} \otimes \mu_x dx \) with \( q_1 := (\varphi_1, w_1 = w_0 |_{d_x}) \). The average of \( h_{q_1} \) over the feature space determines the boundary term variation.

With \( q_1(\varepsilon) = (\varphi_1(\varepsilon), w_1(\varepsilon)) \) take the variation \( \varphi_1(\varepsilon) = \varphi_1 + \varepsilon \psi^T, w_1(\varepsilon) = w_1 + \varepsilon \psi^w \):

\[
\frac{d}{dz} (\mu(q_1(\varepsilon)) - \mu_{obs}) \left|_{\ell = 0} \right. = \frac{d}{dz} (\mu(q_1(\varepsilon)) - \mu_{obs}) \left|_{\ell = 0} \right. = \sum_{i \in I} \frac{d}{dz} (\langle w_{i,1} \nabla_x h_{q_1}(x_{i,1}, f_i) \rangle_{\ell = 0} + h_{q_1}(x_{i,1}, f_i) \psi_i^w) = 0.
\]

**Acknowledgements**

**Author Contributions**

MM and AT designed the theoretical framework and DT built the computational implementation. All authors contributed to writing and editing the manuscript.

**Funding**

This work was supported by the National Institutes of Health (NIH) (www.nih.gov) grants R01EB020062 (MM), R01NS102670 (MM), U19AG033655 (MM), and R01MH105660 (MM), the National Science Foundation (NSF) (www.nsf.gov) 16-569 NeurOx contract 1707298 (MM), the Computational Anatomy Science Gateway (DT and MM) as part of the Extreme Science and Engineering Discovery Environment (XSEDE Towns et al., 2014), which is supported by the NSF grant ACI1548562, Johns Hopkins University Alzheimer’s Disease Research Center with NIH grant P50AG05146, the Dana Foundation’s (www.dana.org) clinical neuroscience research program, and the Kavli Neuroscience Discovery Institute (kavlijhu.org) supported by the Kavli Foundation (www.kavlifoundation.org) (DT, MM, and JT).
Conflict of Interest

MM owns a founder share of Anatomy Works with the arrangement being managed by Johns Hopkins University in accordance with its conflict of interest policies. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability

The major contribution of this work is a mathematical and computational framework for modeling hierarchical neuroimaging data. Specific datasets shown in examples are for illustrative purposes, but can be made available upon request.

A Hamiltonian Control Statement

We take for our RKHS $V_\ell$, $\ell \geq 0$ with the operators $L_\ell : V_\ell \to V_\ell^*$ defining the isometries such that the inner products $\langle \cdot, \cdot \rangle_{V_\ell}$ satisfy for any $u^\ell, v^\ell \in V_\ell$:

$$\langle u^\ell, v^\ell \rangle_{V_\ell} = (L_\ell u^\ell | v^\ell).$$

**Appendix Statement 1.** We assume that $V_{\ell_{\text{max}}} = \mathcal{C}_0^m(\mathbb{R}^d, \mathbb{R}^d)$ with $m \geq k + 2$ and $k \geq 1$.

If $u_\ell$ is a solution of the optimal control problem (25) then for particles there exists time-dependent co-state $(t \mapsto p_t = (p_t^i)_{i \in I, t < \ell_{\text{max}}})$ such that for all $\ell$ satisfies

$$\begin{align*}
\hat{q}_t^\ell &= (u_t^\ell(x_{i,t}), w_t^\ell \div u_t^\ell(x_{i,t})) \\
\hat{p}_t^\ell &= -(du_t^\ell)^T(x_{i,t}) p_t^\ell + w_t^\ell u_t^\ell \nabla (\div u_t^\ell)(x_{i,t}) \\
\hat{p}_t^w &= -p_t^\ell \div u_t^\ell(x_{i,t})
\end{align*}$$

the optimal control for particles satisfies for every scale $\partial u_\ell \mathcal{H}(p, q, u) = 0$ with $v = Au (v^{\ell_{\text{max}}} = 0)$:

$$\begin{align*}
u_t^\ell(\cdot) &= u_t^\ell(\cdot) - u_t^{\ell-1}(\cdot) = \sum_{j \geq \ell} \sum_{i \in I_j} (g_t(x_{j,t}, \cdot)p_t^{j,i} + p_t^{j,i} w_t^{\ell,j} \nabla_1 g_t(x_{j,t}, \cdot)) \\
u_t^\ell(\cdot) &= \sum_{j \geq 0} \sum_{i \in I_j} (\tilde{g}_t(x_{j,t}, \cdot)p_t^{j,i} + p_t^{j,i} w_t^{\ell,j} \nabla_1 \tilde{g}_t(x_{j,t}, \cdot))
\end{align*}$$

with $\tilde{g}_t := \sum_{j=0}^{\ell} g^j$ and $\nabla_1 g(a, b)$ denotes the gradient of $g_\ell$ with respect to the first variable $a$, and with shorthand notation $\geq \ell$ to mean upper bounded by $\ell_{\text{max}}$.

**Appendix Statement 2.** We assume that $V_{\ell_{\text{max}}} = \mathcal{C}_0^m(\mathbb{R}^d, \mathbb{R}^d)$ with $m \geq k + 2$ and $k \geq 1$. If $u_\ell$ is a solution of the optimal control problem (25) for tissue there exists time-dependent co-state $(t \mapsto p_t = (p_t^i(x))_{x \in \mathbb{R}^d, t < \ell_{\text{max}}})$ such that for all $\ell$ satisfies

$$\begin{align*}
\hat{q}_t^\ell(x) &= (u_t^\ell(\varphi_t^\ell(x)), w_t^\ell(x) \div u_t^\ell(\varphi_t^\ell(x))) \\
\hat{p}_t^\ell(x) &= -(du_t^\ell)^T(\varphi_t^\ell(x)) p_t^\ell(x) - p_t^\ell w_t^\ell(x) \nabla (\div u_t^\ell)(\varphi_t^\ell(x)) \\
\hat{p}_t^w &= -p_t^\ell \div u_t^\ell(\varphi_t^\ell(x))
\end{align*}$$
the optimal control for tissue satisfies for every scale $\partial_u H(p, q, u) = 0$ with $v = Au$ ($v^{\ell_{\text{max}} = 0}$):

$$
\begin{align*}
\begin{cases}
\psi^{l'}_t(\cdot) = u^{l'}_t(\cdot) - u^{l'}_{t-\varepsilon}(\cdot) = \sum_{j\geq 0} \int_{\mathbb{R}^d} \left( g_{l',j} (\varphi^{l'}_t(y), \cdot) p^{l',j}_i(y) w^l_j(y) \nabla_x g_{l',j} (\varphi^{l'}_t(y), \cdot) \right) dy \\
u^{l'}_t(\cdot) = \sum_{j \geq 0} \int_{\mathbb{R}^d} \left( g_{l',j} (\varphi^{l'}_t(y), \cdot) p^{l',j}_i(y) w^l_j(y) \nabla_x g_{l',j} (\varphi^{l'}_t(y), \cdot) \right) dy .
\end{cases}
\end{align*}
$$

(32b)

**Proof.** Under the assumption $V_{\ell_{\text{max}}} \rightarrow \mathcal{C}^{\psi^2}(\mathbb{R}^d, \mathbb{R}^d)$ then we have $(u, q) \rightarrow \xi_q(u)$ is $\psi^2$ and standard results of optimal control theory applying the Pontryagin maximum principle [22] gives

$$
\begin{align*}
\begin{cases}
\dot{q} = \partial_q H \\
\dot{p} = -\partial_p H \\
\partial_u H = 0
\end{cases}
\end{align*}
$$

Showing Statement 1 the state and Hamiltonian momentum variations are at the same scale for all $\ell \geq 0$, therefore we omit superscripts $\ell$. For particles the variations at scale $\ell$, $\partial_{l'} H$ and $\partial_u H$, varies $p^{l'}_i(\varepsilon) = p^{l'}_i + \varepsilon \psi^{l'}_i$, $p^{w}_{l'}(\varepsilon) = p^{w}_{l'} + \varepsilon \psi^{w}_{l'}$ and $x_i(\varepsilon) = x_i + \varepsilon \psi^{x}_{l'}$, $w_i(\varepsilon) = w_i + \varepsilon \psi^{w}_{l'}$, respectively, giving the state velocities and momentum differential equations (31a):

$$
\begin{align*}
\frac{d}{d\varepsilon} \left( \langle p \mid \xi_q(u) \rangle \bigg|_{\varepsilon = 0} \right) = \frac{d}{d\varepsilon} \sum_{i \in I} \left( \langle p^{l'}_i + \varepsilon \psi^{l'}_i, u(x_i) \rangle_{\mathbb{R}^d} + \langle p^{w}_{l'} + \varepsilon \psi^{w}_{l'}, w_i \rangle_{\mathbb{R}^d} \right) |_{\varepsilon = 0} \\
= \sum_{i \in I} \left( \langle p^{l'}_i, u(x_i) \rangle_{\mathbb{R}^d} + \psi^{l'}_i \right) \text{div} \left( u(x_i) \right) \\
\frac{d}{d\varepsilon} \left( \langle p \mid \xi_q(u) \rangle \bigg|_{\varepsilon = 0} \right) = \frac{d}{d\varepsilon} \sum_{i \in I} \left( \langle p^{l'}_i, u(x_i(\varepsilon)) \rangle_{\mathbb{R}^d} + \langle p^{w}_{l'}, w_i(\varepsilon) \rangle_{\mathbb{R}^d} \right) |_{\varepsilon = 0} \\
= \sum_{i \in I} \left( \langle p^{l'}_i, du(x_i) \psi^{x}_{l'} \rangle_{\mathbb{R}^d} + \langle p^{w}_{l'}, w_i(\varepsilon) \rangle_{\mathbb{R}^d} \right) \left( \nabla \text{div} \left( u(x_i) \right), \psi^{x}_{l'} \right)_{\mathbb{R}^d} + \langle p^{w}_{l'}, \psi^{w}_{l'} \rangle_{\mathbb{R}^d} .
\end{align*}
$$

(33a)

(33b)

Calculating $\partial_{l'} H = 0$, define $u'^{(\varepsilon)}(\varepsilon) = u^l + \varepsilon \psi^{x}_{l'}$ for $\psi^{x}_{l'} \in V_{l'}$, implying $v^{l+1}(\varepsilon) = v^{l+1} - \varepsilon \psi^{x}_{l'}$, $v^{l'}(\varepsilon) = v^{l'} + \varepsilon \psi^{w}_{l'}$:

$$
\begin{align*}
\frac{d}{d\varepsilon} \left( \langle p^{l'} \mid \xi_{q^{l'}}(u'^{(\varepsilon)}(\varepsilon)) \rangle \bigg|_{\varepsilon = 0} \right) - \frac{1}{2} \sum_{j \geq \ell} \left( \langle L_j v^{l'}(\varepsilon) \mid v^{l'}(\varepsilon) \rangle \right) |_{\varepsilon = 0} \\
= \sum_{i \in I} \left( \langle p^{x^{l'},l'}_i, u^{x^{l'},\varepsilon}(x_i^{l'}) \rangle_{\mathbb{R}^d} + p^{w}_{l'} w_i^{l'} \text{div} u'^{l'}(\varepsilon, x_i^{l'}) \right) - \left( L_{\ell} v^{l'} - L_{\ell+1} v^{l+1} \mid \psi^{x}_{l'} \right) = 0 .
\end{align*}
$$

(34)

After summation of (34) for $j \geq \ell$, we get for any $\psi^{x}_{l'} \in V_{l'}$ that

$$
\langle L_{\ell} v^{l'} \mid \psi^{x}_{l'} \rangle = \sum_{j \geq \ell} \sum_{i \in I} \left( \langle p^{x^{l'},j}_i, \psi^{x}(x_i^{l'}) \rangle_{\mathbb{R}^d} + p^{w}_{l'} w_i^{l'} \text{div} \psi^{x}(x_i^{l'}) \right) \\
$$

(35)

Now, for any $x, \alpha \in \mathbb{R}^d$, consider $\alpha \delta_x \in V^*_{l'}$ such that $\langle \alpha \delta_x \mid v^{l'} \rangle = \langle v^{l'}(x), \alpha \rangle_{\mathbb{R}^d}$ for any $v^{l'} \in V_{l'}$. The reproducing property
on $V_\ell$ gives $\langle K_\ell, \alpha \delta_x, K_\ell \beta \delta_y \rangle_{V_\ell} = g_\ell(x, y) \langle \alpha, \beta \rangle_{R^d}$. We get from (35) for $\psi_w = K_\ell \alpha \delta_x = g_\ell (\cdot, x) \alpha$ that

$$
\langle v^\ell(x), \alpha \rangle_{R^d} = (L_\ell v^\ell \mid K_\ell, \alpha \delta_x) = \sum_{j \geq \ell \in J_i} \sum_{i \in I_j} \left( g_{\ell}(x_i^j, x) \langle p_{\ell}^{j,j}, \alpha \rangle_{R^d} + p_{\ell}^{w,j} w_j^i \langle \nabla_g(x_i^j, x), \alpha \rangle_{R^d} \right)
$$

so that we get the first equality above of (31b) for $v^\ell$ for any $\ell \geq 0$.

Now since $u^\ell = \sum_{j \leq \ell} v^j$ we deduce the equality for $u^\ell$, $\ell \geq 0$ given in (31b).

Showing Statement 2 for tissue, the variation of the momentum has co-states and flows at the same scale for any $\ell \geq 0$, therefore we omit superscripts $\ell$. The variations at scale $\ell$ is $\partial_\ell \mathcal{H}$ and $\partial_{\ell \ell} \mathcal{H}$ varies $p^\ell (\varepsilon) = p^\ell + \varepsilon \psi p^\ell$, $p^w (\varepsilon) = p^w + \varepsilon \psi p^w$ and $\varphi (\varepsilon) = \varphi + \varepsilon \psi \varphi$, $w (\varepsilon) = w + \varepsilon \psi w$, respectively, giving the state velocities and momentum differential equations (32a):

\[
\begin{align*}
\frac{d}{d\varepsilon} & \left( p(\varepsilon) \mid \xi(\varepsilon) (u) \right)_{|\varepsilon=0} = \frac{d}{d\varepsilon} \left( \int \left( \langle p^\ell + \varepsilon \psi p^\ell, u \circ \varphi \rangle_{R^d} + \langle p^w + \varepsilon \psi p^w \rangle w \operatorname{div} u \circ \varphi \rangle dx \right) \bigg|_{\varepsilon=0} \\
& = \int_{\mathbb{R}^d} \langle \psi p^\ell, u \circ \varphi \rangle_{R^d} + \psi p^w w \operatorname{div} u \circ \varphi dx . \tag{36a}
\end{align*}
\]

\[
\begin{align*}
\frac{d}{d\varepsilon} & \left( p \mid \xi(\varepsilon) (u) \right)_{|\varepsilon=0} = \frac{d}{d\varepsilon} \left( \int \left( \langle p^\ell, u \circ \varphi \rangle_{R^d} + p^w w (\varepsilon) \operatorname{div} u \circ \varphi (\varepsilon) \rangle dx \right) \bigg|_{\varepsilon=0} \\
& = \int_{\mathbb{R}^d} \langle \psi p^\ell, du \circ \varphi \rangle_{R^d} + \psi p^w \langle \nabla \operatorname{div} u \circ \varphi, \psi \rangle_{R^d} + p^w \operatorname{div} u \circ \varphi \psi \rangle dx . \tag{36b}
\end{align*}
\]

Calculating $\partial_{u^\ell} \mathcal{H} = 0$, define $u^\varepsilon (\varepsilon) = u^\ell + \varepsilon \psi u^\ell$ for $\psi u^\ell \in V_\ell$, implying $v^{\ell + 1} (\varepsilon) = v^{\ell + 1} - \varepsilon \psi u^\ell$, $v^\ell (\varepsilon) = v^\ell + \varepsilon \psi u^\ell$:

\[
\begin{align*}
\frac{d}{d\varepsilon} & \left( \left( p^\ell \mid \xi(\varepsilon) (u^\ell (\varepsilon)) \right) - \frac{1}{2} \sum_{j=\ell}^{\ell+1} \left( L_j v^j (\varepsilon) \mid v^j (\varepsilon) \right) \right)_{|\varepsilon=0} \\
& = \frac{d}{d\varepsilon} \left( \int (\langle p^\ell, u^\ell \circ \varphi \rangle_{R^d} + p^w \langle \nabla \operatorname{div} u \circ \varphi, \psi \rangle_{R^d} + p^w \operatorname{div} \psi \varphi \rangle dx - \frac{1}{2} \sum_{j=\ell}^{\ell+1} \left( L_j v^j (\varepsilon) \mid v^j (\varepsilon) \right) \right)_{|\varepsilon=0} \\
& = \int_{\mathbb{R}^d} (\langle p^\ell, \psi u^\ell \circ \varphi \rangle_{R^d} + p^w \langle \nabla \operatorname{div} \psi u^\ell \circ \varphi, \psi \rangle_{R^d} + p^w \operatorname{div} \psi u^\ell \circ \varphi \rangle dx - (L_{\ell+1} v^{\ell + 1} \mid \psi u^\ell) = 0 .
\end{align*}
\]

### B Hamiltonian co-state momentum integrable dynamics

We omit the superscripts $\ell$ below since co-states and states and flows are at the same scale.

**Appendix Statement 3.** Assume $q \to U(q) := \frac{1}{2} \| \mu (q_t) - \mu_{\text{obs}} \|_{L^2_\mu}$ is $\mathcal{C}^1$ in $q$, then the co-state integral equations (15b) flowing from $t = 1$ solves the Hamiltonian differential equations $(\dot{p}^\ell, \dot{p}^w)$ of (31a) at every scale $\ell \geq 0$; the integral equations flowing from $t = 0$ satisfy:

\[
\begin{align*}
p^\ell_{1,t} &= (d \varphi_{1,0})^T (x_{i,1}) p^\ell_{0,0} - p^\ell_{1,0} w_{i,0} \int_0^t (d \varphi_{1,s})^T (x_{i,t}) \nabla (\operatorname{div} u_{s}) (x_{i,s}) ds \\
p^w_{1,t} &= p^w_{1,0} (d \varphi_{1,0}) (x_{i,t}) .
\end{align*}
\]

**Proof.** First take $p^w_{1,t} = p^w_{1,0} | d \varphi_{1,0} | (x_{i,t})$ and show it satisfies (31a). For $w_{i,1} = w_i | d \varphi_{1}| (x_i), x_{i,t} = \varphi_t (x_i)$ then

\[
\frac{d}{dt} w_{i,t} p^w_{1,t} = \frac{d}{dt} w_{i,0} | d \varphi_{1}| (x_i) p^w_{1,0} | d \varphi_{1}^{-1}| (x_{i,t}) = 0
\]
which implies \( p_{i,t}^w = -p_{i,t}^w w_{i,t} \)\( \frac{w_{i,t}}{w_{i,t}} = -p_{i,t}^w \text{div} u_t(x_{t,i}). \)

The \( p_{i,t}^\tau \) satisfies (31a): rewrite the integral solution using \( \varphi_{t,s} := \varphi_t \circ \varphi_t^{-1} \) and the identities:

\[
p_{i,t}^\tau = (d\varphi_t^{-1})^T \varphi_t(x_i)p_{i,t}^\tau 0 - p_{i,t}^w w_{i,t} \int_0^t (d\varphi_t^{-1})^T \varphi_t(x_i)d\varphi_t^T(x_i)\nabla(\text{div } u_s)(x_{i,s})dx .
\]

Differentiating requires the identity \( \frac{d}{dt}(d\varphi_t^{-1})^T \circ \varphi_t = -(du_t)^T (d\varphi_t^{-1})^T \circ \varphi_t \) (see below):

\[
p_{i,t}^\tau = -(du_t)^T (x_{i,t}) \left( (d\varphi_t^{-1})^T (x_{i,t})p_{i,t}^\tau 0 - p_{i,t}^w w_{i,t} \int_0^t (d\varphi_t^{-1})^T (x_{i,t})d\varphi_t^T(x_i)\nabla(\text{div } u_t)(x_{i,t}) \right) \\
- p_{i,t}^w w_{i,t} \int_0^t (d\varphi_t^{-1})^T (x_{i,t})d\varphi_t^T(x_i)\nabla(\text{div } u_t)(x_{i,t}) \\
- -(du_t)^T (x_{i,t})p_{i,t}^\tau - p_{i,t}^w w_{i,t} \nabla(\text{div } u_t)(x_{i,t}),
\]

where the last equality uses (40), \( w_{i,t}^w = w_{i,t} p_{i,t}^w \). The remaining identity follows \( \frac{d}{dt}(d\varphi_t^{-1}) = 0 \):

\[
\frac{d}{dt}(d\varphi_t^{-1}) = -(du_t)^T (d\varphi_t^{-1})(du_t) \\
= -(du_t)^T (du_t) \circ \varphi_t = (d\varphi_t^{-1}) \circ \varphi_t du_t \circ \varphi_t .
\]

The optimal control (15a) is written in the endpoint \((p_{i,t}^\tau, p_{i,t}^w)\) with the boundary conditions (17b). Using the form of \( \frac{d}{dt}(d\varphi_t^{-1})\circ \varphi_t \) gives \( p_{i,t}^\tau \) in the boundary \( t = 1 \) satisfies the differential equation. For \( p_{i,t}^w \) then constancy \( p_{i,t}^w w_{i,t} = p_{i,t}^w w_{i,t} \) with \( |d\varphi_{1,1}^w(x_{i,t})| = |d\varphi_1^w(x_1)|d\varphi_{1,1}^w - 1|)(x_{i,t}) \) gives the endpoint:

\[
p_{i,t}^w = \frac{p_{i,t}^w w_{i,1}}{w_{i,t}} = \frac{p_{i,t}^w |d\varphi_1^w(x_1)|}{|d\varphi_1^w(x_1)|} = p_{i,t}^w |d\varphi_{1,1}^w(x_{i,t})|.
\]

\[\square\]

References


