

TOPOLOGICAL AND DYNAMICAL METHODS FOR DATA ANALYSIS

KELLY SPENDLOVE

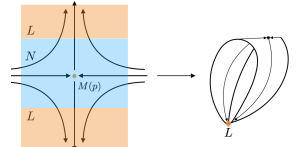
Summary *My interests lie in the development of theories, algorithms and software for the analysis of nonlinear data and systems. My recent work is in Conley index theory, where I developed both a categorical and computational formulation of the connection matrix theory; these are the first steps toward building a constructive, computational homological theory of dynamical systems. My current focus is the development of a computational sheaf theory for describing both multiscale dynamics and topology over different parameters and datasets.*

THE MULTI-SCALE DYNAMICS AND TOPOLOGY OF DATA

High-throughput technologies and high-performance computing infrastructures are providing unprecedented amounts of experimental and simulated data. As science moves toward a data-driven era, with data being harnessed to inform experiment and advance theory, it is increasingly important to recognize that such data are often generated by a nonlinear system. This must be reconciled with 20th-century dynamical systems theory, which established that nonlinear systems can exhibit intricate behavior at all scales with respect to both system variables and parameters. Experimental data are often noisy; simulated data may arise from models where parameters or nonlinearities are not known precisely. This suggests that any analysis designed to synthesize theory, experimentation, computation and data must be based on robust, multi-scale and multi-parameter features. That is, one is interested in the multi-scale topology of the data, as well as a robust description of the behavior of the unknown nonlinear system; in both the *topology* and *dynamics* of the data.

Topology and dynamics are intimately linked: topology constrains or forces the existence of certain dynamics. Loosely put, a dynamical system engenders topological data: both local (e.g. fixed points) and global (e.g. attractors); these data have algebraic invariants (e.g. homology) and a relationship between local and global is codified in the algebra. Morse theory is an influential instantiation of this idea [1], giving a link between the dynamics of a gradient system on a manifold and its topology, often expressed as a homology theory. The local data (nondegenerate fixed points) in the gradient system $\dot{x}(t) = -\nabla f(x(t))$ of generic map $f: M \rightarrow \mathbb{R}$ are graded by their Morse index and assemble into a chain complex (C_\bullet, ∂) . The boundary operator is determined by the connecting orbits and the Morse homology $H_\bullet(C_\bullet, \partial)$ is isomorphic to the singular homology $H_\bullet(M)$. Conley theory is a topological generalization of Morse theory to continuous semiflows and self-maps on a compact metric space X . The global dynamics are organized via a *Morse decomposition*: a finite collection of mutually disjoint, compact, isolated invariant sets $M(p) \subset X$, called *Morse sets* and indexed by a partial order (P, \leq) , such that if $x \in X \setminus \bigcup_{p \in P} M(p)$, then in forward time x limits to a Morse set $M(p)$, in backward time an orbit through x limits to a Morse set $M(q)$, and $p < q$. A Morse set $M(p)$ is isolated with pair $L \subseteq N$ and quantified via $CH_\bullet(p) := H_\bullet(N, L)$, a coarse description of the (local) unstable dynamics. This homology is called the *Conley index* as it is independent of index pair (N, L) , hence an invariant of the Morse set. A crucial property is *continuation*: the index is invariant under perturbation [4].

My recent work [15–17] concerns computing the connection matrix, a Conley-theoretic generalization of the Morse boundary operator [9, 25]. The connection matrix provides a description of the global dynamics over a set of parameters in the terms of homological algebra and order theory. My current focus twofold: (i) the development of a computational sheaf theory for relating connection matrices, (ii) the advancement of the connection matrix as a tool for data analysis.



CATEGORICAL CONNECTION MATRIX THEORY

The connection matrix theory was first introduced by R. Franzosa in [9]. The *connection matrix* is a boundary operator defined on Conley indices associated to a Morse decomposition, i.e. a map

$$(1) \quad \Delta: \bigoplus_{p \in P} CH_{\bullet}(p) \rightarrow \bigoplus_{p \in P} CH_{\bullet}(p).$$

Its basic utility is to identify and give lower bounds on the structure of the connecting orbits. Ultimately, the connection matrix completes Conley theory to a homology theory for dynamical systems. In contrast with Morse theory, the connection matrix may not be unique [9, 24] and the non-uniqueness often reflects bifurcations or changes in the connecting orbits.

Part of my work was to adapt Conley theory to computation by introducing the proper data structures and categories. To sketch an instance of this, let P be a finite poset. Data often arise as a cell complex \mathcal{X} (e.g. CW, simplicial) graded by P , i.e. there is an order-preserving map $\nu: (\mathcal{X}, \leq) \rightarrow (P, \leq)$ where (\mathcal{X}, \leq) is the face poset. This is a combinatorial analogue of the Morse decomposition.¹ A P -graded complex $(C_{\bullet}(P), \partial)$ is a chain complex (C_{\bullet}, ∂) admitting a decomposition into subspaces $C_{\bullet} = \bigoplus_{p \in P} C(p)$ with boundary map determined by its components $\partial_{pq}: C(q) \rightarrow C(p)$ subject to the condition that if $\partial_{pq} \neq 0$ then $p \leq q$. P -graded chain complexes arise as algebra associated to P -graded cell complexes. A morphism $\phi: C_{\bullet}(P) \rightarrow B_{\bullet}(P)$ of graded complexes is a chain map $\phi: C_{\bullet} \rightarrow B_{\bullet}$ that is P -graded, i.e. if $\phi_{qp} \neq 0$ then $p \leq q$. The category of P -graded chain complexes over a field k is $\mathbf{GCh}(P, k)$. The graded homotopy category $\mathbf{GK}(P, k)$ is obtained by localizing about P -graded chain equivalences. The goal of connection matrix theory is to replace a P -graded complex with a simple representative of its isomorphism class in $\mathbf{GK}(P, k)$. The simple representatives are *cyclic-graded* complexes: graded complexes such that $\partial_{pp} = 0$ for each $p \in P$. If $(M_{\bullet}(P), \partial)$ is cyclic-graded then ∂ is a P -graded boundary map, in the form of (1):

$$\partial: \bigoplus_{p \in P} H_{\bullet}(C(p), \partial_{pp}) \rightarrow \bigoplus_{p \in P} H_{\bullet}(C(p), \partial_{pp}).$$

This leads to our new perspective in [15]: a *Conley complex* $(M_{\bullet}(P), \partial^M)$ for $C_{\bullet}(P)$ is a graded-cyclic complex isomorphic to $C_{\bullet}(P)$ in $\mathbf{GK}(P)$; the boundary map ∂^M is a *connection matrix* for $C_{\bullet}(P)$. The new perspective is encapsulated by the following theorem of [15]:

Theorem A. *The inclusion functor $\mathfrak{I}: \mathbf{GK}_0(P, k) \rightarrow \mathbf{GK}(P, k)$ is a categorical equivalence, and there is a inverse functor (a Conley functor) taking each graded complex to a Conley complex,*

$$\mathfrak{C}: \mathbf{GK}(P, k) \rightarrow \mathbf{GK}_0(P, k)$$

This gives a functorial foundation for the connection matrix theory; results in the next section show that the functor \mathfrak{C} may be computed using algebraic-discrete Morse theory.

ALGORITHMS FOR CONNECTION MATRICES

Connection matrix theory is concerned with replacement with simple representatives. A *reduction* is a general method for finding simpler representatives up to chain equivalence [7, 26]. In [15] we introduced a *graded reduction*, a diagram of P -graded complexes and P -graded maps

$$\begin{array}{c} \gamma \\ \curvearrowright \\ (C_{\bullet}(P), \partial) \xrightleftharpoons[\phi]{\psi} (M_{\bullet}(P), \partial^M) \end{array}$$

¹An \mathbb{R} -graded cell complex is the data structure for persistent homology. If $P \subset \mathbb{R}^n$ and $O(P)$ is the lattice of downsets of P , the collection $\{\nu^{-1}(a)\}_{a \in O(P)}$ is a *one-critical multi-filtration* in multiparameter persistence.

with ϕ, ψ chain maps and γ a degree +1 linear map satisfying the identities:

$$\psi\phi = \text{id}_M \quad \phi\psi = \text{id}_C - (\gamma\partial + \partial\gamma) \quad \gamma^2 = \gamma\phi = \psi\gamma = 0$$

The identities show that ϕ is a monomorphism; M_\bullet is the *reduced complex* as it often has much smaller cardinality than C_\bullet . The replacement of C_\bullet with M_\bullet is simplification without loss of homology. Discrete Morse theory [8] is a wildly popular technique in applied topology for simplifying complexes [6, 14, 22]. The main tool is a *discrete vector field*: a pairing \mathcal{V} which partitions the cells of a complex \mathcal{X} into pairs $V_\alpha = (x_\alpha \leq y_\alpha)$ where x_α is a codimension-1 face of y_α . Cells not paired by \mathcal{V} are *critical*. A *discrete flowline* is a sequence (V_i) of distinct pairs which may be written as $x_1 \leq y_1 \geq x_2 \leq y_2 \geq \dots \geq x_N \leq y_N$. As in the smooth case, critical cells and their relationship via discrete flowlines determine a chain complex $(M_\bullet(\mathcal{X}, \mathcal{V}), \partial)$ that is chain equivalent to $C_\bullet(\mathcal{X})$. A vector field is *graded with respect to ν* : $\mathcal{X} \rightarrow \mathbf{P}$ if it satisfies the (additional) property: if x and y are paired, then $\nu(x) = \nu(y)$. There are efficient algorithms for producing (graded) vector fields, e.g. [14, 22]. In [15] we show that a graded vector field \mathcal{V} induces a graded reduction, with the Morse complex $M_\bullet(\mathcal{V})$ assuming the role of the reduced complex. This gives an algorithmic foundation for the categorical theory, obtained by iterating discrete Morse theory, as in the next result:

Theorem B. *Let \mathcal{X} be cell complex with grading $\nu: \mathcal{X} \rightarrow \mathbf{P}$. There exists a sequence $\{\mathcal{V}_i\}$ of graded vector fields and an associated tower of \mathbf{P} -graded reductions*

$$\begin{array}{ccccccc} \begin{array}{c} \curvearrowright \gamma_0 \\ C(\mathcal{X})(\mathbf{P}) \end{array} & \begin{array}{c} \xrightarrow{\psi_0} \\ \xleftarrow{\phi_0} \end{array} & \begin{array}{c} \curvearrowright \gamma_1 \\ M(\mathcal{V}_0)(\mathbf{P}) \end{array} & \begin{array}{c} \xrightarrow{\psi_1} \\ \xleftarrow{\phi_1} \end{array} & \dots & \begin{array}{c} \xrightarrow{\psi_{n-1}} \\ \xleftarrow{\phi_{n-1}} \end{array} & \begin{array}{c} \curvearrowright \gamma_n \\ M(\mathcal{V}_{n-1})(\mathbf{P}) \end{array} & \begin{array}{c} \xrightarrow{\psi_n} \\ \xleftarrow{\phi_n} \end{array} & M(\mathcal{V}_n)(\mathbf{P}) \end{array}$$

such that $M(\mathcal{V}_n)(\mathbf{P})$ is a Conley complex.

COMPUTATIONAL CONNECTION MATRIX THEORY

With S. Harker, K. Mischaikow and R. Vandervorst, I am developing a *computational connection matrix theory* within the context of the computational Conley theory [5, 18–21]. The ultimate goal is to promote the computational Conley theory into a computational homological theory of dynamics. One selects a cell complex \mathcal{X} (e.g. cubical, simplicial) on a compact global attractor $X \subseteq \mathbb{R}^n$ of a flow φ . The (topological) closure of the top-cells \mathcal{X}_n of \mathcal{X} are the atoms of finite subalgebra \mathbf{X} of the Boolean algebra of regular closed sets on X . Together $(\mathcal{X}, \mathbf{X})$ reflect a choice of scale, often determined a constraint (e.g. confidence in the model). The continuous dynamics of φ are approximated combinatorially by a directed graph \mathcal{F} on \mathcal{X}_n . The quotient of \mathcal{F} by reachability is the poset $\text{SC}(\mathcal{F})$ of strongly connected components; the quotient map $\nu_0: \mathcal{X}_n \rightarrow \text{SC}(\mathcal{F})$ takes a top-cell to its associated strongly connected component. A $\text{SC}(\mathcal{F})$ -graded complex is produced by extending the inclusion $\mathcal{X}_n \hookrightarrow \mathcal{X}$ to a map ν so that the following diagram commutes:

$$(2) \quad \begin{array}{ccc} \mathcal{X}_n & \xrightarrow{\quad} & \mathcal{X} \\ \nu_0 \downarrow & \swarrow \nu & \\ \text{SC}(\mathcal{F}) & & \end{array}$$

In [17] we develop the concept of a *transversality model* for a flow φ : a choice of $(\mathcal{X}, \mathbf{X}, \mathcal{F})$ such that if edge $\xi \rightarrow \xi' \notin \mathcal{F}$ then φ is transverse along the intersection $\xi' \cap \xi$ in the direction ξ' to ξ .

Theorem C. *If $(\mathcal{X}, \mathbf{X}, \mathcal{F})$ is a transversality model for φ then there is an extension ν in (2) so*

- (1) *the collection $\mathbf{A} = \{\nu^{-1}(a)\}_{a \in \mathbf{O}(\mathbf{P})}$, where $\mathbf{O}(\mathbf{P})$ is the lattice of downsets of \mathbf{P} , is a lattice of attracting blocks for φ ,*
- (2) *a connection matrix obtained from $C(\mathcal{X})(\mathbf{P})$ via Theorem B is a connected matrix for \mathbf{A} .*

To my knowledge, the experiments of [17] are among of the highest dimensional computations performed in the applied topology community. The current goal is to examine a 4-fold cover of a pseudo-Anosov braid, a 12-D cubical complex with 1 trillion cells, to help give lower bounds on the entropy. Such massive computations are enabled by an implicit scheme for graded discrete Morse theory on cubical complexes that I developed in collaboration with S. Harker [16]. The long term goal for the braids project twofold: (i) resolve the conjectures of [30]; (ii) develop a computational version of braid Floer homology theory [2], which would prove exceptionally useful as Floer homology is notoriously difficult to calculate.

ONGOING WORK

Connection Matrices for Data Analysis Persistent homology is the most popular tool in topological data analysis. In [15] we show that persistent homology can be computed from the connection matrix, which is nicely summarized via the following theorem from [28].

Theorem D. *Given a P-graded reduction*

$$\begin{array}{c} \gamma \\ \curvearrowright \\ (C_\bullet(P), \partial) \end{array} \begin{array}{c} \xrightarrow{\psi} \\ \xleftarrow{\phi} \end{array} (M_\bullet(P), \partial^M),$$

- (1) *the graded complex splits as $C_\bullet(P) = M_\bullet(P) \oplus \ker \psi$,*
- (2) *if $M_\bullet(P)$ is a Conley complex then $\ker \psi$ contains all pairs of zero persistence.*

As a corollary, a Conley complex is the smallest complex for recovering persistent homology. The paradigm for using the Conley complex in topological data analysis would be to enrich the persistence diagram by instead focusing on the Conley complex as the fundamental tool for analysis. Not only does the Conley complex contain the persistent homology, but generators of both persistence pairs and Conley indices can be lifted to the data through the chain equivalences.

Computational Transition Matrix Theory Often data or dynamics come parameterized by a base space. In such cases, it is crucial to characterize how the global structure and dynamics of the data vary with respect to parameters. My aim is to develop a computational sheaf theory based on Conley index theory as a tool for the robust analysis of global dynamics over parameters and scales. Classically, *transition matrices* relate connection matrices at different parameters. There have been different definitions of transition matrix (e.g. [10, 23]), all recently unified [11]. Transition matrices are typically used to track changes to orbit structure and are classically defined as to require a Morse decomposition to continue. My preliminary work suggests that they can be formulated more abstractly and analyzed using the conventional tools of homological algebra, viz. mapping cones and mapping cylinders, generalized to the graded category. Transition matrices will be the restriction maps of the Conley sheaf. The first step to a sheaf theory is developing the computational transition matrix theory: placing transition matrices within the framework of reductions, homotopy categories and algebraic-discrete Morse theory and using the categorical language to address non-uniqueness.

Application: Sheaf-Theoretic Systems Biology Conley-theoretic methods enable a database approach to dynamics, cataloging behavior over parameters (e.g. [5]). In [5] the input is a switching system model for a gene regulatory network. Parameter space is decomposed into semi-algebraic sets and encoded as an undirected graph PG: each set is a vertex with an edge between vertices of adjacent sets. Associated with each vertex is a directed graph representation of dynamics \mathcal{F} which is valid over the set of parameters. The graph $\text{SC}(\mathcal{F})$ gives a coarse description of dynamics, allowing for querying spaces of networks and parameters for desired behavior, e.g. bistability.

Incorporating computational connection matrix theory into the database will enrich the poset $\text{SC}(\mathcal{F})$ with the algebra of a $\text{SC}(\mathcal{F})$ -graded Conley complex. Regarding PG as a 1-D cell complex, Conley complexes can be assigned to the vertices and edges. Considering PG as a category whose objects are the cells and morphisms are inclusions, these data assemble into a cellular sheaf [6], i.e. a functor $\text{PG} \rightarrow \mathbf{GK}_0$. Additional higher-order connectivity promotes PG to a regular cell complex, endowing the sheaf with more compatibility data. Interrogation of the sheaf gives information on local and global critical transitions, the existence of multistability, and potential monodromies. For instance, hysteresis is a phenomena dependent upon the behavior of global dynamics over paths in parameter space, which is a sheaf-theoretic query about the existence of certain sections.

REFERENCES CITED

- [1] Bott, R., 1988. Morse theory indomitable. *Publications Mathematiques de l'IHES*, 68, pp.99-114.
- [2] van den Berg, J.B., Ghrist, R., Vandervorst, R.C. and Wojcik, W., 2015. Braid Floer homology. *Journal of Differential Equations*, 259(5), pp.1663-1721.
- [3] van den Berg, J.B., Vandervorst, R.C. and Wjcik, W., 2007. Chaos in orientation reversing twist maps of the plane. *Topology and its Applications*, 154(13), pp.2580-2606.
- [4] C. Conley. *Isolated invariant sets and the Morse index*, American Mathematical Society, 1978.
- [5] Cummins, B., Gedeon, T., Harker, S., Mischaikow, K. and Mok, K., 2016. Combinatorial Representation of Parameter Space for Switching Systems. *SIAM Journal on Applied Dynamical Systems*, 15(4), pp.2176-2212.
- [6] Curry, J., Ghrist, R. and Nanda, V., 2016. Discrete Morse Theory for Computing Cellular Sheaf Cohomology. *Foundations of Computational Mathematics*, 16(4):pp.875-897.
- [7] Eilenberg, S. and Lane, S.M., 1953. On the groups $H(\pi, n)$, I. *Annals of Mathematics*, pp.55-106.
- [8] Forman, R., 1998. Morse theory for cell complexes. *Advances in mathematics*, 134(1), pp.90-145.
- [9] R. Franzosa. 1989. The connection matrix theory for Morse decompositions, *Transactions of the AMS*, 311(2), pp.561 – 592.
- [10] Franzosa, R. and Mischaikow, K., 1998. Algebraic transition matrices in the Conley index theory. *Transactions of the American Mathematical Society*, 350(3), pp.889-912.
- [11] Franzosa, R. and Vieira, E., 2017. Transition matrix theory. *Transactions of the American Mathematical Society*, 369(11), pp.7737-7764.
- [12] Ghrist, R., van den Berg, J.B. and Vandervorst, R.C.A.M., 2003. Morse theory on spaces of braids and Lagrangian dynamics. *Inventiones mathematicae*, 152(2), pp.369-432.
- [13] Ghrist, R. and Vandervorst, R.C.A.M., 2009. Scalar parabolic PDEs and braids. *Transactions of the American Mathematical Society*, 361(5), pp.2755-2788.
- [14] Harker, S., Mischaikow, K., Mrozek, M. and Nanda, V., 2014. Discrete Morse theoretic algorithms for computing homology of complexes and maps. *Foundations of Computational Mathematics*, 14(1), pp.151-184.
- [15] Harker, S., Mischaikow, K. and Spendlove, K. 2018. A Computational Framework for the Connection Matrix Theory. Preprint. <https://arxiv.org/abs/1810.04552>.
- [16] Harker, S., Mischaikow, K. and Spendlove, K. 2018. Discrete Morse Theory for Computation of Connection Matrices. Preprint.
- [17] Harker, S., Mischaikow, K., Spendlove, K. and van der Vorst R. 2018. Computational Connection Matrix Theory with Applications to Transversality Models. Preprint.
- [18] Kalies, W.D., Mischaikow, K. and Vandervorst, R.C., 2005. An algorithmic approach to chain recurrence. *Foundations of Computational Mathematics*, 5(4), pp.409-449.
- [19] Kalies, W.D., Mischaikow, K. and Vandervorst R.C., 2014. Lattice structures for attractors I. *Journal of Computational Dynamics* 1(2), pp. 307-338.
- [20] Kalies, W.D., Mischaikow, K. and Vandervorst, R.C., 2016. Lattice structures for attractors II. *Foundations of Computational Mathematics*, 16(5), pp. 1151-1191.
- [21] Kalies, W.D., Kasti, D. and Vandervorst, R., 2018. An Algorithmic Approach to Lattices and Order in Dynamics. *SIAM Journal on Applied Dynamical Systems*, 17(2), pp.1617-1649.
- [22] Mischaikow, K. and Nanda, V., 2013. Morse theory for filtrations and efficient computation of persistent homology. *Discrete & Computational Geometry*, 50(2), pp.330-353.
- [23] Reineck, J.F., 1988. Connecting orbits in one-parameter families of flows. *Ergodic theory and dynamical systems*, 8(8), pp.359-374.
- [24] Reineck, J.F., 1990. The connection matrix in Morse-Smale flows. *Transactions of the American Mathematical Society*, 322(2), pp.523-545.
- [25] Robbin, J.W. and Salamon, D.A., 1992. Lyapunov maps, simplicial complexes and the Stone functor. *Ergodic Theory and Dynamical Systems*, 12(1), pp.153-183.
- [26] Rubio, J. and Sergeraert, F., 2002. Constructive algebraic topology. *Bulletin des Sciences Mathematiques*, 126(5), pp.389-412.
- [27] Spendlove, K., 2018. On the Homotopy Category for Connection Matrix Theory. *In preparation*.
- [28] Spendlove, K., 2018. A Homotopy-Theoretic Framework for (Lattice)-Filtered Chain Complexes. *In preparation*.
- [29] Spendlove, K., 2018. <https://github.com/kelspendlove/pyCHomP>
- [30] Spendlove, K. and R. Vandervorst 2018. High-Dimensional Conley Theory on Braids. *In preparation*.