

Supplemental Material:

**Helicoidal dynamics of biaxial curved rods in twist-bend nematic
phases unveiled by unsupervised machine learning techniques**

Massimiliano Chiappini, Alessandro Patti, and Marjolein Dijkstra

(Dated: September 8, 2020)

I. THE GENERALISED DYNAMIC MONTE CARLO METHOD FOR HARD CURVED SPHEROCYLINDERS

We consider a single Monte Carlo (MC) step of our generalised Dynamic Monte Carlo (DMC) method for a system of hard curved spherocylinders with diffusion tensor at infinite dilution $\mathbf{D}_0 = \mathbf{Q}\mathbf{D}'_0\mathbf{Q}^T$, where \mathbf{D}'_0 is the diagonalized \mathbf{D}_0 tensor and \mathbf{Q}^T is a change of basis matrix. We randomly select a curved rod and generate a vector $\boldsymbol{\delta}'$ with components δ'_i sampled from a uniform distribution in the interval $[-\Delta x_i, \Delta x_i]$ with $\Delta x_i = \sqrt{2D'_{0,i}\delta t_{MC}}$. We propose a displacement $\Delta\mathbf{x} = \mathbf{Q}\boldsymbol{\delta}'$ of the 6-dimensional particle state \mathbf{x} , comprising both translational and rotational degrees of freedom. If the move $\Delta\mathbf{x} = \mathbf{Q}\boldsymbol{\delta}'$ is accepted, the correlation between translational and rotational displacements is given by the cross-correlation matrix

$$\langle\Delta\mathbf{x}\Delta\mathbf{x}^T\rangle = \langle(\mathbf{Q}\boldsymbol{\delta}')(\mathbf{Q}\boldsymbol{\delta}')^T\rangle = \mathbf{Q}\langle\boldsymbol{\delta}'\boldsymbol{\delta}'^T\rangle\mathbf{Q}^T. \quad (1)$$

Since $\langle\boldsymbol{\delta}'\boldsymbol{\delta}'^T\rangle$ is the self cross-correlation matrix of the vector $\boldsymbol{\delta}'$, and the components of $\boldsymbol{\delta}'$ are stochastically independent, we find

$$\langle\boldsymbol{\delta}'\boldsymbol{\delta}'^T\rangle = \begin{bmatrix} \langle\delta_1'^2\rangle & & 0 \\ & \ddots & \\ 0 & & \langle\delta_6'^2\rangle \end{bmatrix}. \quad (2)$$

In the case the components of $\boldsymbol{\delta}'$ are sampled from a uniform distribution in $[-\Delta x_i, \Delta x_i]$, it is easy to derive that $\langle\delta_i'^2\rangle = \Delta x_i^2/3 = 2D'_{0,i}\delta t_{MC}/3$, and $\langle\boldsymbol{\delta}'\boldsymbol{\delta}'^T\rangle = (2\delta t_{MC}/3)\mathbf{D}'_0$, where $\mathbf{D}'_0 = \mathbf{Q}^T\mathbf{D}_0\mathbf{Q}$. Hence

$$\langle\Delta\mathbf{x}\Delta\mathbf{x}^T\rangle = (2\delta t_{MC}/3)\mathbf{Q}\mathbf{D}'_0\mathbf{Q}^T = (2\delta t_{MC}/3)\mathbf{Q}\mathbf{Q}^T\mathbf{D}_0\mathbf{Q}\mathbf{Q}^T. \quad (3)$$

Since \mathbf{Q} is an orthonormal matrix, $\mathbf{Q}^{-1} = \mathbf{Q}^T$, and the cross-correlation of the particle displacements after one accepted DMC move finally reads

$$\langle\Delta\mathbf{x}\Delta\mathbf{x}^T\rangle = (2\delta t_{MC}/3)\mathbf{D}_0, \quad (4)$$

which corresponds to the cross-correlation of the Brownian dynamics

$$\langle\Delta\mathbf{x}\Delta\mathbf{x}^T\rangle = 2\mathbf{D}_0t \quad (5)$$

for $t = \delta t_{MC}/3$.

Taking into account the acceptance ratio \mathcal{A} of the DMC moves, crucial to compare dynamics across different conditions and/or systems, and generalising to systems of N_{HCS} identical hard curved spherocylinders, we obtain the time rescale $t = \mathcal{A}\mathcal{C}_{MC}\delta t_{MC}/3$, where \mathcal{C}_{MC} is the number of DMC cycles. The arbitrary Monte Carlo unit of time δt_{MC} can be used to tune the time resolution of the DMC simulations. In this work δt_{MC} is always equal to 0.1τ , where $\tau = \sigma^3\mu/k_B T$, with σ the unit of length, and μ and T the viscosity coefficient and the temperature of the solvent respectively.

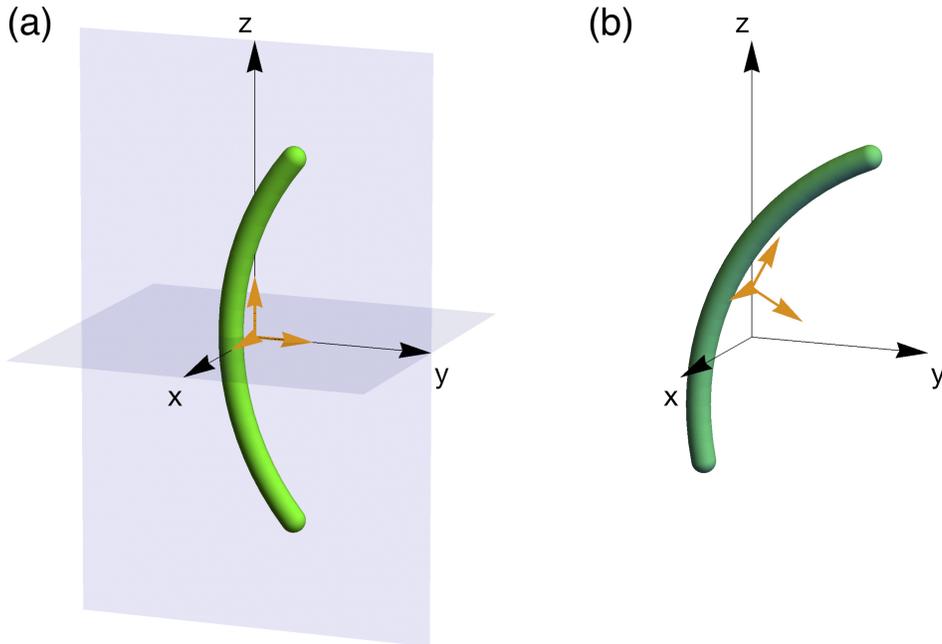


FIG. S1. (a) Sketch of a hard curved spherocylinder of length $L = 16d$ and radius of curvature $R = 11.5d$ with molecular reference frame (in orange) coincident with the laboratory reference frame (in black). The particle body lies in the yz plane of the molecular reference frame, and the xy and yz planes are planes of mirror symmetry. The center-of-mass of the curved rod coincides with the center of the molecular reference frame. (b) Sketch of a hard curved spherocylinder and its molecular reference frame after a displacement via the main mode of diffusion.

II. DIFFUSION TENSOR OF HARD CURVED SPHEROCYLINDERS

We consider a HCS of length $L = 16d$ and radius of curvature $R = 11.5d$, lying in the yz plane of its molecular reference frame as sketched in FIG. S1 (a), with its center-of-mass

in the origin of the molecular reference frame. The diffusion tensor, as calculated via the Hydro++ software package [1], reads

$$\mathbf{D}_0^* = \begin{bmatrix} \frac{\mathbf{D}_0^{tt}}{\tilde{D}^{tt}} & \frac{\mathbf{D}_0^{tr}}{\tilde{D}^{tr}} \\ \frac{\mathbf{D}_0^{rt}}{\tilde{D}^{rt}} & \frac{\mathbf{D}_0^{rr}}{\tilde{D}^{rr}} \end{bmatrix} = \begin{bmatrix} 0.014 & 0 & 0 & 0 & 0 & -0.00019 \\ 0 & 0.014 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.019 & -0.00019 & 0 & 0 \\ 0 & 0 & -0.00019 & 0.00032 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.00031 & 0 \\ -0.00019 & 0 & 0 & 0 & 0 & 0.0017 \end{bmatrix}, \quad (6)$$

where \mathbf{D}_0^{tt} , $\mathbf{D}_0^{tr} = (\mathbf{D}_0^{rt})^T$, and \mathbf{D}_0^{rr} are the translational, roto-translational and rotational parts of the diffusion tensor, respectively, and $\tilde{D}^{tt} = \sigma^2/\tau$, $\tilde{D}^{tr} = \tilde{D}^{rt} = \sigma/\tau$, and $\tilde{D}^{rr} = 1/\tau$. Note that all the terms have been rounded to the second significant decimal and all terms smaller than 10^{-6} have been neglected for simplicity. Note that the diffusion tensor of Eqn. 6 respects the symmetry expected for particles with two mutually orthogonal planes of symmetry (the xy and yz planes) [2].

Diagonalization of \mathbf{D}_0^* yields the following transformation matrix \mathbf{Q}^* and diagonal diffusion tensor \mathbf{D}'_0^*

$$\mathbf{Q}^* = \begin{bmatrix} 1.0 & 0.015 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.010 & 0.0 & 0.0 \\ 0.0 & 0.0 & -0.010 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \\ -0.015 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}, \quad (7)$$

$$\mathbf{D}'_0^* = \begin{bmatrix} 0.014 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0017 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.019 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.00032 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.014 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.00031 \end{bmatrix}, \quad (8)$$

where again all terms have been rounded to the second significant decimal and all terms smaller than 10^{-6} have been neglected. In the generalised DMC framework, each eigenvalue of \mathbf{D}_0^* corresponds to the diffusion coefficient of a natural coordinate corresponding to a

certain mode of diffusion. Hence, the highest eigenvalue identifies the fastest natural mode of diffusion. In the current case, the highest eigenvalue is $D_{0,c}^* = 0.019$, corresponding to the eigenvector

$$\Psi_c^* = \begin{bmatrix} 0.0 \\ 0.0 \\ 1.0 \\ -0.010 \\ 0.0 \\ 0.0 \end{bmatrix} \quad (9)$$

i.e. to a negative correlation between a displacement along z and a rotation around the x axis which, as sketched in FIG. S1 (b), corresponds to a sliding of the curved spherocylinder following its curved profile.

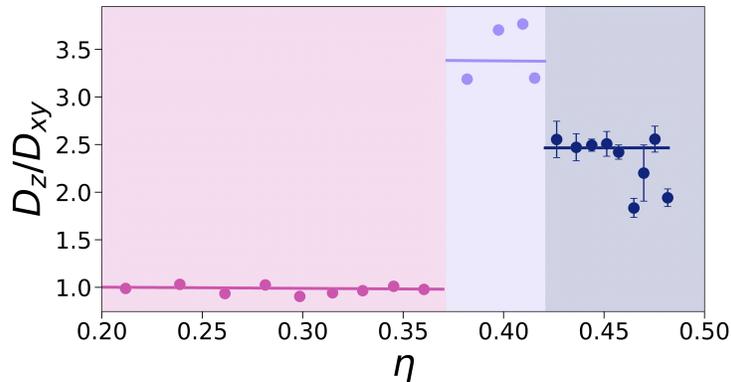


FIG. S2. Anisotropy D_z/D_{xy} of the diffusion HCSs with length $L = 10d$ and curvature $R = 9.5d$ as a function of η . The anisotropy of the dynamics is weakened at the N - N_{TB} transition because of the obliqueness of the nematic director in the N_{TB} phase. The lines are guides to the eye.

III. ANISOTROPY OF THE DIFFUSION

In FIG. S2 we report the anisotropy D_z/D_{xy} of the diffusion longitudinal and perpendicular to the z axis, with $D_{xy} = (D_x + D_y)/2$, as a function of packing fraction η in a system of hard curved spherocylinders with diameter d , length $L = 10d$, and radius of curvature $R = 9.5d$, with a I - N transition at $\eta_{IN} = 0.37 \pm 0.01$ and an N - N_{TB} transition at $\eta_{NN_{TB}} = 0.43 \pm 0.01$.

In the I phase the dynamics is isotropic, *i.e.* $D_x = D_y = D_z$, and $D_z/D_{xy} = 1$. At the I - N transition the system becomes anisotropic as curved rods align along a common nematic director $\hat{\mathbf{n}} \parallel z$. In turn, the single-particle diffusion becomes anisotropic, $D_z/D_{xy} \sim 3.5$ times faster along the nematic director than perpendicular to it. The anisotropy is weaker in the N_{TB} phase, where $D_z/D_{xy} \sim 2.5$, due to the obliqueness of the *local* nematic director with respect to the z axis.

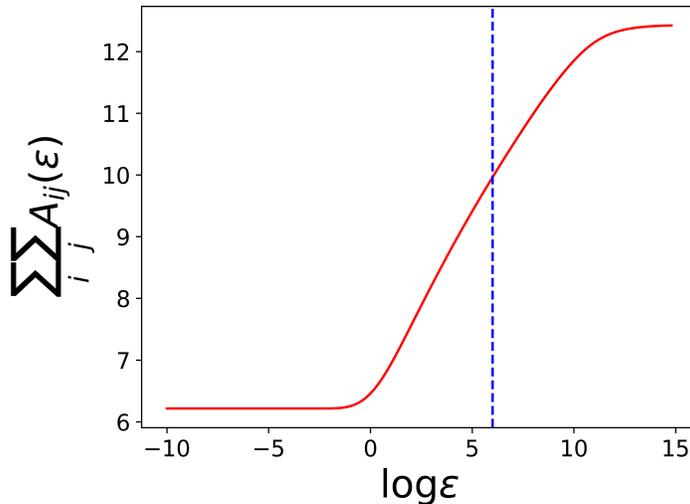


FIG. S3. Example of $\sum_i \sum_j A_{ij}(\epsilon)$ as a function of the Gaussian kernel bandwidth ϵ . The vertical line corresponds to $\epsilon = \exp(6)$, which lies in the linear region of the characteristic sigmoid curve $\sum_i \sum_j A_{ij}(\epsilon)$ for all the analysed trajectories.

IV. DIFFUSION MAPS AS AN UNSUPERVISED DIMENSIONALITY REDUCTION METHOD

The Diffusion Maps (DMs) method is based on the assumption that random walks constructed over a data set have higher probability to diffuse to data points that are "kinetically" close. The data set is then non-linearly reparametrized in a low-dimensional representation that clusters data connected by many kinetic pathways.

To implement DMs, one should first quantify the distance between two data points \mathbf{r}_i and \mathbf{r}_j at time i and j . The choice of this metrics is not trivial, strongly affects the effectiveness of

DMs, and depends dramatically on the system. As our "data set" is a real-space trajectory, a natural choice would be the Euclidean norm $\|\mathbf{r}_i - \mathbf{r}_j\|$. However, configurations close in time are more likely to be kinetically connected and to be part of the same helicoidal path. As the trajectory $\{\mathbf{r}_i\}$ is already ordered in time, we include the temporal information simply by employing the metrics $P_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\| + \sqrt{(i^2 - j^2)}$.

From the matrix \mathbf{P} we build the matrix \mathbf{A} as a convolution of \mathbf{P} with a Gaussian kernel of bandwidth ϵ , *i.e.*

$$A_{ij} = \exp\left(-\frac{P_{ij}^2}{\epsilon}\right). \quad (10)$$

An appropriate choice of ϵ is fundamental to ensure the full connectivity of the random walk that is being built on the data set. To tune the value of ϵ , $\sum_i \sum_j A_{ij}(\epsilon)$ is calculated as a function of ϵ , resulting in a characteristic sigmoid curve in a log-log representation (see FIG. S3). Suitable values of the kernel bandwidth are the ones corresponding to a linear region of this sigmoid [3]. In this work, $\epsilon = \exp(6)$ results to be a suitable value for all the trajectories analysed.

In addition, an estimate of the data set dimensionality, namely the dimensionality of the underlying manifold from which the data is actually sampled, can be inferred as twice the slope of the linear region of the sigmoid curve [3]. In this work, the data set dimensionality fluctuates between 1 and 2 for all the trajectories analysed, confirming that the trajectories explore a manifold of lower dimensionality than the 3D Euclidean space they are embedded in.

The matrix \mathbf{A} is then row-normalized, yielding the transition matrix \mathbf{M} of a Markov process over the data set. The eigenvectors Ψ_i corresponding to the largest eigenvalues of \mathbf{M} provide the leading modes of this Markov process. Neglecting the trivial stationary mode of eigenvalue $\lambda_1 = 1$ and eigenvector Ψ_1 , a reparameterization of the original dataset is obtained as $\mathbf{r}_i \rightarrow \Xi_i = (\Psi_{2,i}, \Psi_{3,i}, \dots)$.

V. CLUSTERING OF PARTICLE TRAJECTORIES TO IDENTIFY DISTINCT HELICOIDAL PATHS

Using diffusion maps, a 3-dimensional particle trajectory consisting of data points $\mathbf{r}_i = (x_i, y_i, z_i)$ with $i = 1, \dots, M$ is mapped onto a 2-dimensional trajectory $\Xi_i = (\Psi_{2,i}, \Psi_{3,i})$. We apply the Hierarchical Agglomerative Clustering (HAC) method [4] on the unfolded 2D

trajectory. HAC is a bottom-up clustering approach: all data are initially assigned to a different cluster, and the clusters are then gradually merged. The merging of clusters is stopped when a pre-determined number of clusters N_c is reached. To decide which clusters have to be merged at each iteration, a measure of the similarity of the clusters, *i.e.* a so-called *linkage criterion*, is required. To this end, we first define the distance $d(\Xi_i, \Xi_j)$ between two data points Ξ_i and Ξ_j . As a metrics, we simply use the Euclidean distance in the 2-dimensional space

$$d(\Xi_i, \Xi_j) = \sqrt{(\Psi_{2,i} - \Psi_{2,j})^2 + (\Psi_{3,i} - \Psi_{3,j})^2}. \quad (11)$$

We then determine the dissimilarity $\mathcal{D}(A, B)$ between two clusters A and B

$$\mathcal{D}(A, B) = \frac{1}{|A||B|} \sum_{i \in A} \sum_{j \in B} d(\Xi_i, \Xi_j), \quad (12)$$

where $|A|$ and $|B|$ are the cardinalities of clusters A and B , respectively. At each iteration, the two least dissimilar clusters, *i.e.* with the smallest distance $\mathcal{D}(A, B)$, are merged.

The optimal choice of N_c is the minimal number of clusters that yields the best fit of the helicoidal trajectories, *i.e.* such that using $N_c + 1$ clusters does not entail a significant improvement of the quality of the fit, quantified as the R^2 of the fit. In order to systematically determine the optimal N_c according to this criterion, we employ the L -method by Salvador and Chan [5] to locate an elbow in R^2 as a function of N_c . We perform the clustering with various number of clusters $\{N_c\} = \{1, 2, \dots, 10\}$, and fit the resulting clusters as discussed in the main paper. We measure for each N_c the total R^2 on the fits. Given $R^2(N_c)$, we find the division point e of $\{N_c\}$ in left and right subsets $\{N_c^l\} = \{1, \dots, e - 1\}$ and $\{N_c^r\} = \{e, \dots, 10\}$ which yields the best linear fits of $R^2(N_c^l)$ and $R^2(N_c^r)$, thereby locating the elbow of $R^2(N_c)$ as shown in FIG. S01. We then use $N_c = e + 2$ as the number of clusters.

VI. DERIVATION OF THE PSEUDOHELICES PITCH AND RADIUS

In the main paper, we define pseudohelices as manifolds $\mathbf{p}(z)$ to which the nematic director field $\hat{\mathbf{n}}(z) = n_x(z)\hat{\mathbf{e}}_x + n_y(z)\hat{\mathbf{e}}_y + n_z(z)\hat{\mathbf{e}}_z$ of the N_{TB} phase is tangent at any z , namely $\mathbf{p}'(z) = \nu\hat{\mathbf{n}}(z)$, with ν a constant norm $\|\mathbf{p}'(z)\| = \nu$. We can thus obtain the

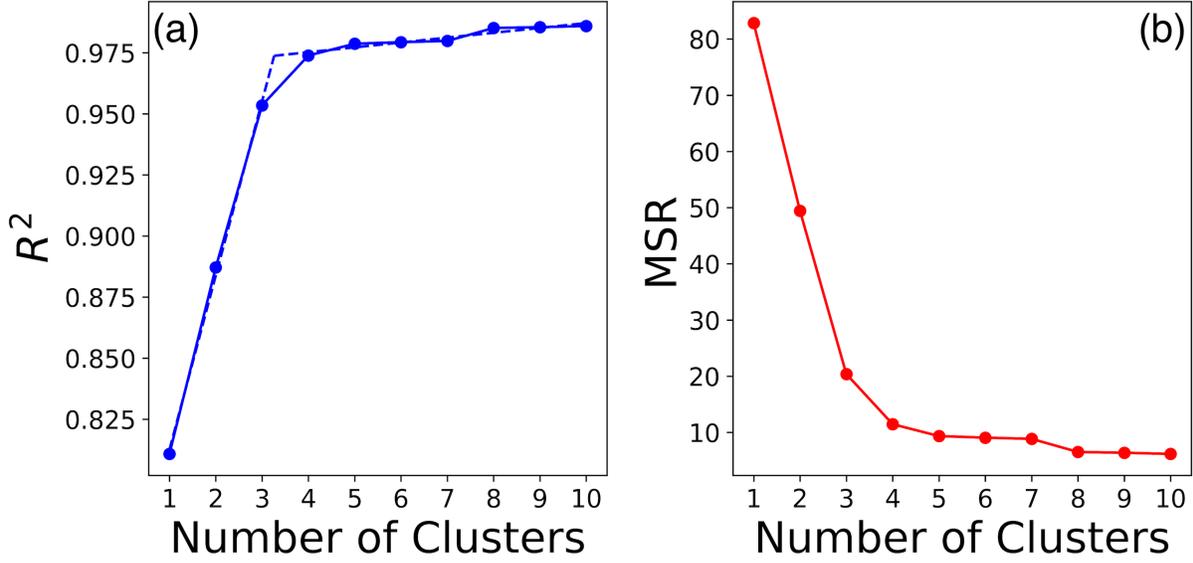


FIG. S4. Examples of (a) R^2 and (b) Mean Squared Residuals (MSR) on the fits of the distinct helicoidal paths of particle trajectories of HCSs as a function of the number of clusters N_c . The L -method by Salvator and Chan [5] is applied to locate the elbow e of the function $R^2(N_c)$ by determining the partition of $R^2(N_c)$ which yields the best pair of linear fits, as shown in (a).

equation $\mathbf{p}(z)$ of a pseudohelix by integration of the nematic director field $\hat{\mathbf{n}}(z)$

$$\mathbf{p}(z) = \begin{cases} p_x(z) = x_0 + \nu \int_{x_0}^z dz' n_x(z) \\ p_y(z) = y_0 + \nu \int_{y_0}^z dz' n_y(z) , \\ p_z(z) = z_0 + \nu \int_{z_0}^z dz' n_z(z) \end{cases} , \quad (13)$$

which, for an N_{TB} phase with pitch p and conical angle θ_0 reads

$$\mathbf{p}(z) = \begin{cases} p_x(z) = x_0 + \nu \int_{x_0}^z dz' \sin \theta_0 \cos (q(z - z_0)) \\ p_y(z) = y_0 + \nu \int_{y_0}^z dz' \sin \theta_0 \sin (q(z - z_0)) . \\ p_z(z) = z_0 + \nu \int_{z_0}^z dz' \cos \theta_0 \end{cases} . \quad (14)$$

The integral of Eq. 14 yields an infinite number of identical manifolds with different starting points $\mathbf{r}_0 = (x_0, y_0, z_0)$. If we consider, for example, the manifold passing through the origin,

i.e. we set $\mathbf{r}_0 = (0, 0, 0)$, we obtain

$$\mathbf{p}(z) = \begin{cases} p_x(z) = \frac{\nu}{q} \sin \theta_0 \sin(qz) \\ p_y(z) = -\frac{\nu}{q} \sin \theta_0 \cos(qz) . \\ p_z(z) = \nu \cos \theta_0 z \end{cases} \quad (15)$$

But $p_z(z)$ must be equal to z , yielding $\nu = 1/\cos\theta_0$ and

$$\mathbf{p}(z) = \begin{cases} p_x(z) = \frac{1}{q} \tan \theta_0 \sin(qz) \\ p_y(z) = -\frac{1}{q} \tan \theta_0 \cos(qz) , \\ p_z(z) = z \end{cases} \quad (16)$$

i.e. the equation of a helix of pitch $\mathcal{P} = p$ and radius $\mathcal{R} = |p| \tan \theta_0 / 2\pi$.

- [1] J. G. de la Torre, G. del Rio Echenique, and A. Ortega, *The Journal of Physical Chemistry B* **111**, 955 (2007).
- [2] H. Brenner, *Journal of Colloid and Interface Science* **23**, 407 (1967).
- [3] R. R. Coifman, Y. Shkolnisky, F. J. Sigworth, and A. Singer, *IEEE Transactions on Image Processing* **17**, 1891 (2008).
- [4] L. Rokach and O. Maimon, "Clustering methods," in *Data Mining and Knowledge Discovery Handbook*, edited by O. Maimon and L. Rokach (Springer US, Boston, MA, 2005) pp. 321–352.
- [5] S. Salvador and P. Chan, in *16th IEEE International Conference on Tools with Artificial Intelligence* (2004) pp. 576–584.