DATA CLUSTERING BASED ON LANGEVIN ANNEALING WITH A SELF-CONSISTENT POTENTIAL

By

KYLE LAFATA (*Department of Radiation Oncology, Physics Division, Department of Physics, and Medical Physics Graduate Program, Duke University, Durham, North Carolina 27705*),

ZHENNAN ZHOU (*Beijing International Center for Mathematical Research, Peking University, Beijing, People's Republic of China 100871*),

JIAN-GUO LIU (*Departments of Mathematics and Physics, Duke University, Durham, North Carolina 27705*),

AND

FANG-FANG YIN (*Department of Radiation Oncology, Physics Division and Medical Physics Graduate Program, Duke University, Durham, North Carolina 27705*)

Abstract. This paper introduces a novel data clustering algorithm based on Langevin dynamics, where the associated potential is constructed directly from the data. To introduce a self-consistent potential, we adopt the potential model from the established Quantum Clustering method. The first step is to use a radial basis function to construct a density distribution from the data. A potential function is then constructed such that this density distribution is the ground state solution to the time-independent Schrödinger equation. The second step is to use this potential function with the Langevin dynamics at subcritical temperature to avoid ergodicity. The Langevin equations take a classical Gibbs distribution as the invariant measure, where the peaks of the distribution coincide with minima of the potential surface. The time dynamics of individual data points

lead to different metastable states, which are interpreted as cluster centers. Clustering is therefore achieved when subsets of the data aggregate—as a result of the Langevin dynamics for a moderate period of time—in the neighborhood of a particular potential minimum. While the data points are pushed towards potential minima by the potential gradient, Brownian motion allows them to effectively tunnel through local potential barriers and escape saddle points into locations of the potential surface otherwise forbidden. The algorithm's feasibility is first established based on several illustrating examples and theoretical analyses, followed by a stricter evaluation using a standard benchmark dataset.

1. Introduction. Searching for meaningful structure within hyper-dimensional datasets is fundamental to modern data science. Cluster analysis, i.e., the grouping of similar data objects together based on their intrinsic properties, is a common approach to understanding otherwise nontrivial data. Although data clustering is a hallmark of many fields (e.g., machine learning, data mining, pattern recognition, etc. [3, 4, 12, 20, 21, 28, 33]), it is, in general, an ill-defined practice [10, 32] that may benefit from physics intuition. In this paper, we propose a novel approach to data clustering based on Langevin annealing. Data points are modeled as particles satisfying the second order Langevin equations with a self-consistent potential, where the Langevin dynamics are used to stochastically propagate them on a potential surface. The time dynamics of individual data points lead to different metastable states, which are interpreted as cluster centers. Clustering is achieved when subsets of the data—as a result of the Langevin dynamics—aggregate in the neighborhood of a particular potential minimum. To demonstrate the novelty of such a technique, we provide the following in this paper: (1) a complete mathematical derivation of the proposed technique, and (2) a numerical validation of the proposed technique, using a common benchmark dataset.

Briefly, the motivation behind our approach to clustering is as follows. A given dataset is sampled from a canonical ensemble, whose distribution takes the form of the classical Gibbs measure,

$$\rho_A(\mathbf{q}) = Z^{-1} e^{-\frac{V(\mathbf{q})}{k_B T}}, \tag{1.1}$$

where k_B is the Boltzmann constant, T is the absolute temperature, Z is the appropriate partition function, $\mathbf{q} \in \mathbb{R}^k$ is a position vector in k dimensions,[1] and $V(\mathbf{q})$ is a potential function that reflects the density of the data distribution. The Gibbs distribution, ρ_A, also serves as the $\mathbf{q}$ marginal distribution of the invariant measure of the following second order Langevin equation,

$$m\frac{d^2\mathbf{q}}{dt^2} = -\nabla V(\mathbf{q}) - \gamma\frac{d\mathbf{q}}{dt} + \Gamma\,\eta(t), \tag{1.2}$$

---

[1]The number of dimensions in the problem is inherently defined based on the number of data attributes within a particular dataset.

where m is the particle's mass, γ is a damping coefficient, $\eta(t)$ is a k-dimensional noise term, and $\Gamma = \sqrt{2\gamma k_B T}$ is an amplitude scaling coefficient [7]. According to the Fluctuation-Dissipation Theorem [11, 24, 37], the damping term will *dissipate* kinetic energy as it evolves in time, which is balanced with corresponding *fluctuations* in the form of Brownian motion. Langevin dynamics are equivalently formulated as the following stochastic differential equations of motion:

$$d\mathbf{q} = \mathbf{p}\,dt \qquad (1.3)$$

and

$$d\mathbf{p} = -\nabla V(\mathbf{q})\,dt - \gamma\mathbf{p}\,dt + \sqrt{2\gamma k_B T}\,dB, \qquad (1.4)$$

where the noise term is specified as k-dimensional Brownian motion, $B(t)$, $\mathbf{p} \in \mathbb{R}^k$ is a time-dependent momentum vector, and we let $m = 1$ for simplicity. Formally speaking, while the forces from the potential surface push data points towards potential minima, corresponding Brownian fluctuations allow them to jump small potential barriers and escape saddle points into locations of the potential surface otherwise forbidden. The key observation is the following. Although each trajectory is ergodic with respect to the classical Gibbs distribution regardless of where it is initiated, when the potential barrier between two portions of the configuration space is big, it is highly likely that a trajectory initiated at one portion of the space will stay within the same portion for a moderate period of time. Therefore, we are motivated to propose the following clustering mechanism. When the potential function $V(\mathbf{q})$ has several distinct minima, if we propagate the trajectories starting from the data points of a given data set, with a high probability, those trajectories will aggregate into several groups within a short time. Further, if the profile of the potential function intrinsically indicates the density distribution of the data points, the data points that end up in the neighborhood of the same minima of $V(\mathbf{q})$ after $O(1)$ time can be considered to be in the same cluster.

Clearly, the mechanics summarized via equations (1.1)–(1.4) require a potential function, V. While V can take on any well-defined functional form, we opt in this paper to use the Schrödinger equation to model a self-consistent potential based on the method developed by Horn and Gottlieb [19]. As such, Euclidean data are mapped into a Hilbert space of square intergrable functions [17] by associating with a given data set a corresponding wave function, $\psi(\mathbf{x})$. One can then interpret $\psi(\mathbf{x})$ as the ground-state solution associated with the quantum mechanical Hamiltonian operator,

$$\mathcal{H}\psi(\mathbf{x}) = -\frac{\varepsilon^2}{2}\nabla^2\psi(\mathbf{x}) + V(\mathbf{x})\psi(\mathbf{x}), \qquad (1.5)$$

and inversely solve for the potential function, $V(\mathbf{x})$, whose minima correspond to cluster centers. Here, ε is a rescaled, dimensionless analog of the reduced Planck constant, $\hbar \to \varepsilon$, such that $\varepsilon \ll 1$ defines the semi-classical regime: a transition regime between quantum mechanics and classical mechanics [15, 29]. Minima of $V(\mathbf{x})$ intuitively correspond to amplified regions of high probability density within the original data set, leading to its high-resolution functional representation [19].

The *dynamic* nature of the proposed methodology is consistent with recent research interests that study clustering as a dynamic problem [5, 25, 31, 41]. In general, dynamic frameworks provide the unique capability to investigate *how* clusters are formed, and may potentially provide important meta-stable structural information about a particular dataset. In particular, Dynamic Quantum Clustering (DQC) extends quantum mechanics intuition to a dynamic framework [41]. By coupling data points with Gaussian wave-packets in a Hilbert space of wave functions, DQC interprets the data as solutions to the time-dependent Schrödinger equation,

$$i\varepsilon\frac{\partial}{\partial t}\left|\Psi\right\rangle = \mathcal{H}\left|\Psi\right\rangle, \tag{1.6}$$

such that clustering is achieved by temporally propagating the system based on $\mathcal{H}$, and tracking the wave-packets on their oscillatory trajectory about potential minima [43].

The relative distances between the centroids of different states changes with time according to Ehrenfest's Theorem, thus representing a measure of similarity between data points. Wave-packets which oscillate around the same minima of $V(\mathbf{x})$ are considered to be in the same cluster. In general, deriving weights from distance metrics such as this has been a hallmark of data clustering algorithms. As the convergence of many points to a common location is a clear and intuitive indication of clustered data, many approaches advocate constructing similarity functions. This is a way to characterize the distance between data points and is an essential aspect of DQC, as well as methods of spectral clustering [6, 31, 40], diffuse interface models [2, 13], and graph-clustering based on diffusion geometry [8, 25, 31, 38].

Evolving the quantum system according to equation (1.6) leads to closed-system physics driven by a completely conserved Hamiltonian, such that clustering is only a transient behavior. In our methodology, while we adopt the method used in [19] to generate $V(\mathbf{x})$ explicitly from the data, we propagate the trajectories based on Langevin equations. Compared to [41], this leads to a fundamentally different dynamic scheme influenced by the interaction between data points and a temperature-characterized environment. In fact, we derive (in Section 2 of this paper) a critical temperature, T_0, in the regime that equation (1.1) is asymptotically matched with the quantum mechanics formalism for probability density,

$$\rho_B(\mathbf{x}) = \left|\psi(\mathbf{x})\right|^2, \tag{1.7}$$

such that

$$Z^{-1}e^{-\frac{V(\mathbf{x})}{k_B T_0}} \approx \left|\psi(\mathbf{x})\right|^2. \tag{1.8}$$

In practice, it is possible that at such a critical temperature, clustering of data is unlikely due to relatively large stochastic fluctuations. However, at subcritical temperatures (that is, when $T < T_0$ is sufficiently small), the time-evolution is more strongly

influenced by dissipative and systematic forces.[2] Accordingly, the position of each data point tends to first move *towards*, and subsequently move locally *around*, nearby potential minima. As will be demonstrated in this paper, this thermal annealing process is the primary phenomenon which results in decoding meaningful structure about a given dataset.

Data clustering is therefore a direct consequence of Langevin dynamics at subcritical temperatures for a moderate period of time (in contrast to ergodic sampling, which is a long-time asymptotic behavior). In fact, in the regime that $T \ll T_0$, momentum decays to a small magnitude due to nominal stochastic influence (i.e., minimal Brownian motion where $\Gamma \ll 1$), resulting in a very high likelihood of data-point localization. Appropriately adjusting the subcritical temperature in the thermostat provides a means to tailor the clustering phenomenon to that of a particular data set. This subtle yet important distinction offers a couple of key advantages. First, the Langevin dynamics framework is fundamentally different from time-evolution based on Ehrenfest's Theorem. The latter results in oscillatory trajectories about potential minima, which can often lead to practically difficult implementation. Subcritical Langevin dynamics, however, lead to damped trajectories towards potential minima, resulting in the intuitive aggregation of data points. Second, Brownian motion allows data points to thermally jump local potential barriers and escape saddle points into locations of the potential surface otherwise forbidden. Such a tunneling phenomenon makes it possible for nearly degenerate local minima of $V(\mathbf{x})$ to merge into a single cluster. Datasets can therefore be explored with a high resolution kernel, while still maintaining a reasonably narrow impulse response.

Given the proposed dynamic framework, there is a natural transition away from cluster analysis based solely on geometric intuition. By interpreting the potential landscape from a kinematic perspective—rather than solely as a geometric entity—it can be easily generalized to higher-dimensional space [23]. Under such a generalization, the landscape of $V(\mathbf{x})$ is characterized by the net systematic force felt by each data point at a given location contributing to the overall Langevin dynamics. In particular, there are three important scenarios on the surface of $V(\mathbf{x})$ where such an interpretation may be used to evaluate the dynamic response of a test particle. These scenarios, which are detailed in Section 2 of this paper, include (1) evaluation of a test particle at a saddle point on the surface of $V(\mathbf{x})$; (2) evaluation of a test particle far from all stationary points of $V(\mathbf{x})$; and (3) evaluation of a test particle approaching minima of $V(\mathbf{x})$.

The remainder of this paper proceeds as follows. In Section 2, we detail the theoretical framework of our proposed clustering algorithm, define the spatial and thermal conditions for its effective implementation, and derive a critical temperature to match the quantum and classical perspectives of a given dataset. In Section 3, we briefly demonstrate the machinery used to numerically integrate the Langevin dynamics that drives the clustering process. Numerical results are presented in Section 4. Using a common benchmark dataset, several computational experiments are designed to evaluate the methodology and characterize its key parameters. Concluding remarks and future directions are provided in Section 5.

---

[2]Except for particles near stationary points of the potential surface.

2. Data clustering with Langevin annealing. In this section, we aim to establish a general mathematical framework for data clustering based on Langevin dynamics. The Schrödinger equation is first used to model a self-consistent potential function that reflects the probability density of a given dataset. Langevin dynamics are then formulated for such a potential, where clustering is expected to be a relatively low-temperature phenomenon. Finally, an analytical relationship is derived between the scale of the potential function and the temperature in the Langevin thermostat. The authors stress that, while quantum mechanics intuition is borrowed to establish self-consistency, the novelty of this paper is in the Langevin annealing. That is, the approach can be applied to any well-defined potential function.

2.1. *Constructing a self-consistent potential.* Given a dataset, $\{\mathbf{x}_i\}_{i=1}^N \subset \mathbb{R}^k$, each $\mathbf{x}_i$ vector is first mapped from its original k-dimensional Euclidean space into a Hilbert space of square integrable functions,

$$\psi_i(\mathbf{x}) = C e^{-\frac{|\mathbf{x}-\mathbf{x}_i|^2}{2\sigma^2}} = C e^{-\frac{|\mathbf{x}-\mathbf{x}_i|^2}{2\varepsilon}}, \tag{2.1}$$

where σ is the resolution of the chosen radial basis function kernel whose variance, $\varepsilon = \sigma^2$, is interpreted as the semi-classical parameter (a rescaled analog of the reduced Planck constant), and C is the appropriate normalization constant. Similar to [19], a wavefunction is then constructed as the superposition of these N wave-packets,

$$\psi_0(\mathbf{x}) = C_N \sum_{i=1}^N e^{-\frac{|\mathbf{x}-\mathbf{x}_i|^2}{2\varepsilon}}, \tag{2.2}$$

where C_N is a normalization constant ensuring that

$$\langle \psi_0 | \psi_0 \rangle = \|\psi_0\|_{L^2}^2 = 1. \tag{2.3}$$

To measure the initial overlap observed amongst the set of wave-packets at a given $\sqrt{\varepsilon}$-resolution, we introduce a set of weight functions,

$$w_i(\mathbf{x}) = \frac{e^{-\frac{|\mathbf{x}-\mathbf{x}_i|^2}{2\varepsilon}}}{\sum_{j=1}^N e^{-\frac{|\mathbf{x}-\mathbf{x}_j|^2}{2\varepsilon}}}, \tag{2.4}$$

such that

$$\sum_{i=1}^N w_i(\mathbf{x}) = 1. \tag{2.5}$$

Next, by requiring $\psi_0(\mathbf{x})$ to be the ground state solution to the semi-classical Hamiltonian [19],

$$\mathcal{H}\psi_0(\mathbf{x}) \equiv \left[-\frac{\varepsilon^2}{2}\nabla^2 + V(\mathbf{x}) \right]\psi_0(\mathbf{x}) = E_0\psi_0(\mathbf{x}), \tag{2.6}$$

the following potential function can be inversely calculated up to a constant:

$$V(\mathbf{x}) = \frac{\frac{\varepsilon^2}{2}\nabla^2\psi_0(\mathbf{x})}{\psi_0(\mathbf{x})} + E_0 = \frac{-k\varepsilon}{2} + \frac{\sum_{i=1}^{N}|\mathbf{x}-\mathbf{x}_i|^2 e^{-\frac{|\mathbf{x}-\mathbf{x}_i|^2}{2\varepsilon}}}{2\sum_{i=1}^{N}e^{-\frac{|\mathbf{x}-\mathbf{x}_i|^2}{2\varepsilon}}} + E_0, \qquad (2.7)$$

where E_0 is the ground state energy.[3] The potential function, written in terms of equation (2.4), becomes

$$V(\mathbf{x}) = \frac{-k\varepsilon}{2} + \sum_{i=1}^{N}w_i(\mathbf{x})\frac{|\mathbf{x}-\mathbf{x}_i|^2}{2} + E_0. \qquad (2.8)$$

Both $\psi_0(\mathbf{x})$ and $V(\mathbf{x})$ are intimately connected to the original Euclidean dataset. Maxima of $|\psi_0(\mathbf{x})|^2$ correspond to densely populated regions within $\{\mathbf{x}_i\}_{i=1}^{N}$, hence why this mapping is a fairly common approach in conventional kernel-density estimation algorithms [35]. Conversely, minima of $V(\mathbf{x})$ correspond to locations where the wave-packets are subject to high degrees of local attraction and are therefore interpreted as cluster centers [18]. Further, $\psi_0(\mathbf{x})$ and $V(\mathbf{x})$ are both characterized by resolution-dependent density variations throughout the data, parameterized by $\sqrt{\varepsilon}$. In particular, when $\varepsilon \ll 1$, each wave-packet is observed at a semi-classical resolution. This is a transition regime between quantum mechanics and classical mechanics, where the position and momentum of each wavepacket is x_i and 0, respectively.[4]

2.2. *Langevin annealing and data clustering.* One essential feature of semi-classical wave-packets (which is fundamental to the proposed clustering algorithm) is as follows. Up to the Ehrenfest time, $t_E \approx -\log\varepsilon$, the wave-packets can be interpreted as coherent states, and the quantum dynamics can be approximated using Newtonian physics [1,16]. That is, the time-evolution of the wave-packets can be approximated by evolving their parameters. Specifically, if we consider the semi-classical time-dependent Schrödinger equation,

$$i\varepsilon\frac{\partial}{\partial t}\Psi_i(\mathbf{x},t) = \left[-\frac{\varepsilon^2}{2}\nabla^2 + V(\mathbf{x})\right]\Psi_i(\mathbf{x},t), \qquad (2.9)$$

whose initial condition is given by,

$$\Psi_i(\mathbf{x},0) = \psi_i(\mathbf{x}) = Ce^{-\frac{|\mathbf{x}-\mathbf{x}_i|^2}{2\varepsilon}}, \qquad (2.10)$$

then, $\Psi_i(\mathbf{x},t)$ can be approximated by a parameterized wave-packet whose phase-space coordinates, $\mathbf{q}(t), \mathbf{p}(t)$, are determined by the classical Hamiltonian equations,

$$\dot{\mathbf{q}} = \mathbf{p}, \qquad (2.11)$$

---

[3]The ground state energy can be regarded as a bias term and does not practically affect the mechanics of clustering.

[4]Momentum is zero because we have chosen, without loss of generality, not to include a phase term. If a nonzero phase term was used, the total wavefunction would no longer necessarily be real-valued. For simplicity in the current paper, we only consider the real-valued wavefunction.

and

$$\dot{\mathbf{p}} = -\nabla_{\mathbf{q}} V(\mathbf{q}), \qquad (2.12)$$

with initial conditions, $\mathbf{q}(0) = \mathbf{x}_i, \quad \mathbf{p}(0) = 0.$[5]

In this work, however, we treat the data points as classical points in equilibrium with a heat bath, characterized by a temperature, T. Subsequently, we require that the trajectory of the particles obey the following Langevin dynamics,

$$d\mathbf{q}_i = \mathbf{p}_i \, dt, \qquad (2.13)$$

and

$$d\mathbf{p}_i = -\nabla_q V(\mathbf{q}_i) \, dt - \gamma \mathbf{p}_i \, dt + \sqrt{2\gamma\beta^{-1}} \, dB, \qquad (2.14)$$

with initial conditions, $\mathbf{q}(0) = \mathbf{x}_i, \quad \mathbf{p}(0) = 0$. Here, $B(t)$ is k-dimensional Brownian motion, γ is a damping coefficient, and $\beta = (k_B T)^{-1}$, where T is the absolute temperature of the system and k_B is the Boltzmann constant. For simplicity, we let $\gamma = 1$ and choose to work at the characteristic scale where $k_B = 1$. In general, propagation of the position and momentum, based on equations (2.13) and (2.14), respectively, lead to ergodic trajectories with respect to equations (1.1) and (2.8). Further, it can be easily shown that the analytic form of the systematic force term (i.e., the gradient of equation (2.8)) is composed of both a linear component and a harmonic trap,

$$F \equiv -\nabla V(\mathbf{x}) = -\left\{ \sum_i \nabla w_i(\mathbf{x}) \frac{|\mathbf{x} - \mathbf{x}_i|^2}{2} + \sum_i w_i(\mathbf{x})(\mathbf{x} - \mathbf{x}_i) \right\}. \qquad (2.15)$$

When $\varepsilon \ll 1$, as we shall show in the next section, there exists a critical temperature, T_0, such that the quantum probability distribution is asymptotically convergent to a classical Gibbs measure. The interaction between stochastic and systematic forces preserve a state of equilibrium that guarantees a sampling of the desired canonical distribution for a sufficiently long time, whereas clustering, which is a short time behavior, may not be easily observed at critical temperature due to Brownian motion with a relatively large amplitude.

However, when $T < T_0$ is small enough, the interaction between the potential function and the Brownian motion leads to a different equilibrium state, uniquely characterized by a subcritical temperature, T. As such, it is highly likely that each data point follows the gradient of the potential towards the nearest minima of $V(\mathbf{x})$ where it is trapped. Data points which end up trapped in the same minima of $V(\mathbf{x})$ are considered to be in the same cluster. Interestingly, this implies that data clustering is fundamentally driven by Langevin dynamics at a sufficiently low temperature, T. Further, while the

---

[5]To maintain consistency between the quantum mechanical and statistical mechanical perspectives, we have introduced the following notation: The points $\hat{x}_i$ are used when referring to the original, static dataset, and the points $\hat{q}$, $\hat{p}$ are used when referring to the time evolution of the phase space via Langevin dynamics. Accordingly, $\hat{q}_i(0) = \hat{x}_i$ indicates that $\hat{x}_i$ and $\hat{q}_i$ are essentially equal counterparts of the quantum and classical distributions, respectively.

systematic force acts globally on the system to push data points *downhill*, the Brownian motion allows them to thermally jump local potential barriers into locations of the potential surface otherwise forbidden. Such a tunneling phenomenon makes it possible for the merging of nearly degenerate local minima of $V(\mathbf{x})$. This enables the data to be explored with a high resolution kernel, while still maintaining a reasonably narrow impulse response (therefore reducing some of the noise that equation (2.2) may introduce during the radial basis function operation).

To track the time-dependent changes in data overlap that transpires during the dynamic process described above, we introduce a weight matrix, w_{ij}, whose initial condition,

$$w_{ij}(t = 0) = w_i(\mathbf{x}_j), \tag{2.16}$$

is determined via equation (2.4). Here, we recall that the points $\mathbf{x}_j$ are time-independent by definition. Adhering to notation, the weight matrix as a function of time is then

$$w_{ij}(t) = w_i(\mathbf{q}_j(t)), \tag{2.17}$$

as the points $\mathbf{q}_j$ are defined to be time-dependent. A high degree of dynamic overlap occurs for data points that aggregate together into similar clusters, i.e., $w_{ij} \to 1$. Conversely, data points that are separated into different clusters produce negligible dynamic overlap, i.e., $w_{ij} \to 0$.

While the magnitude of Brownian motion is invariant, the systematic net-force changes according to equation (2.15). Evaluation of the net force acting on a particle can be used to characterize the dynamics at various key locations on a potential landscape. As previously suggested in the introduction, such interpretation of the potential function's landscape as a kinematic quantity is somewhat of a paradigm shift away from conventional function optimization based solely on geometric intuition. In particular, the following illustrating examples demonstrate the theoretical response of a test particle[6] to three archetypical potential function interactions:

(1) A test particle evaluated far from potential stationary points: Here, the dynamics are dominated by steep potential gradients and dissipative effects, leading to damped Newtonian mechanics. Accordingly, the particle's trajectory roughly follows the gradient of the potential.

(2) A test particle evaluated at a saddle point: Here, Brownian motion most prominently influences local dynamics, *kicking* the particle out of the stationary point and allowing it to continue its trajectory.

(3) A test particle evaluated near functional minima: Here, $\nabla V(\mathbf{x}) \approx 0$, and the momentum of the particle becomes small due to damping effects. Dissipation and fluctuation are comparable in magnitude, and they dominate the dynamics about the minima leading to dynamic particle confinement.

The simulated response of a test particle to each of these situations is demonstrated in Figure 1, using test-functions as a proxy for the potential landscape. Each simulation was executed from $t = 0$ to $t = 1$ with a time-step of $dt = 0.01$. The red stars indicate

---

[6]Here, we refer to a test particle as being a 1D proxy for a data point within the illustrating example presented in Figure 1.
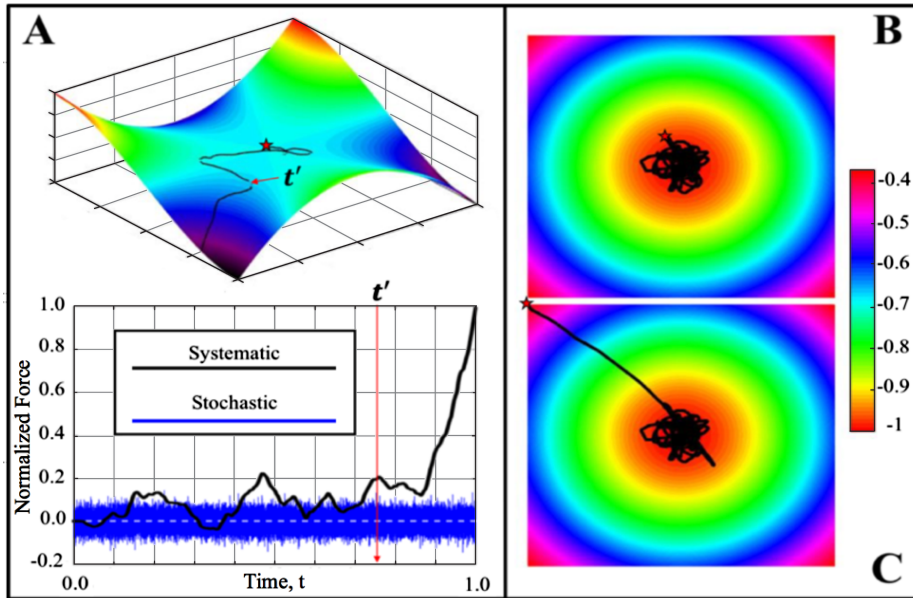
FIG. 1. Schematic plots of typical Langevin trajectories. For each scenario (A–C) the red star indicates the starting location. (A) A particle's trajectory is initialized at a saddle point, and shown to escape for a short time. Below, the normalized systematic and stochastic forces are plotted as a function of time, t. The systematic force is the dominating effect when the particle locates the stronger potential gradient, as indicated at time $t = t'$. (B) A particle's trajectory is initialized near a minimum, where it is contained. (C) A particle's trajectory is initialized away from a minimum, but ends up contained within the minimum.

the starting location of each test particle (with zero momentum), and the black lines demonstrate their time-trajectories according to equations (2.13) and (2.14).

A hyperbolic paraboloid (Figure 1A) was used to test the dynamic response of a particle with zero momentum starting at the function's saddle point. In general, effectively escaping saddle points is a fundamental research topic in optimization problems [9, 14] and manifold learning [36]. The magnitudes of the systematic and stochastic forces (Figure 1A, bottom) are shown as a function of the time-evolution. To demonstrate their relative effects as a function of time, force magnitudes have been normalized relative to 100% simulation completion. It is confirmed that systematic and stochastic influence are comparable when the particle is near the stationary point, i.e., the systematic contribution is essentially trapped within the bandwidth of the stochastic noise. However, the systematic force is the dominating effect as the particle eventually locates the stronger potential gradient, as indicated in Figure 1A at simulation time, $t = t'$. Inverted Gaussian functions (displayed as 2D projections) were used to study the dynamic response of a test particle located near minima (Figure 1B) and far from stationary points (Figure 1C). Clearly, particle confinement at the minima is achieved in both cases as a function of time-evolution.

2.3. *Asymptotic matching of quantum and classical distributions.* As others have noted [5,17–19,41], the conventional quantum mechanical probability density function, $\rho_B(\mathbf{x}) = |\psi_0(\mathbf{x})|^2$, can be interpreted as an empirical probability distribution obtained from the original data points, $\{\mathbf{x}_i\}_{i=1}^N$. However, we assume that this data set is a sample drawn from a more general distribution that obeys hypotheses of a canonical ensemble. We emphasize that this is the most distinguishable assumption in this method: instead of viewing $\{\mathbf{x}_i\}_{i=1}^N$ as a closed system, we consider it as a subsystem taken from an environment. As detailed above, such an environment is characterized by a temperature, T, and there exists a critical temperature, T_0, such that the quantum distribution is asymptotically convergent to a classical distribution (with a discrepancy of order $O(\varepsilon)$). As such, in this section we aim to asymptotically match the quantum distribution with a classical distribution, and in the process derive an analytical expression for critical temperature, T_0.

In general, the Langevin dynamics lead to an ergodic trajectory with respect to the classical Gibbs distribution,

$$F(\mathbf{q}, \mathbf{p}) \propto e^{-\beta H(\mathbf{q}, \mathbf{p})} \equiv e^{-\beta\left(\frac{|\mathbf{p}|^2}{2} + V(\mathbf{q})\right)}, \tag{2.18}$$

whose marginal distribution is given by

$$f(\mathbf{q}) = \int F(\mathbf{q}, \mathbf{p}) \, d\mathbf{p} \propto e^{-\beta V(\mathbf{q})}. \tag{2.19}$$

We then denote the partition function for this marginal distribution as Z such that

$$f(\mathbf{q}) = \frac{1}{Z} e^{-\beta V(\mathbf{q})}, \quad \int f(\mathbf{q}) \, d\mathbf{q} = 1. \tag{2.20}$$

We have thus derived the approximate probability density function from two different perspectives: quantum mechanical and classical mechanical. As such, we propose the following consistency condition:

$$\rho_A(\mathbf{x}) = |\psi_0(\mathbf{x})|^2 \approx \frac{1}{Z} e^{-\beta V(\mathbf{x})}. \tag{2.21}$$

By definition, $\psi_0 \geq 0$, such that equation (2.21) is equivalent to

$$\psi_0(\mathbf{x}) \approx \frac{1}{\sqrt{Z}} e^{-\frac{\beta}{2} V(\mathbf{x})}, \tag{2.22}$$

which leads to the following matching condition:

$$C_N \sum_{i=1}^N e^{-\frac{|\mathbf{x} - \mathbf{x}_i|^2}{2\varepsilon}} \approx \frac{1}{\sqrt{Z}} e^{-\frac{\beta}{2} C} e^{-\frac{\beta}{2} \sum_{i=1}^N w_i(\mathbf{x}) \frac{|\mathbf{x} - \mathbf{x}_i|^2}{2}}. \tag{2.23}$$

This matching condition is carried out under the assumption that, given $\{\mathbf{x}_i\}_{i=1}^N$, one can choose a sufficiently small ε such that the following nonoverlapping wave-packet condition,

$$w_i(\mathbf{x}_j) \approx \delta_{ij}, \tag{2.24}$$

is nearly satisfied according to the weight functions defined via equation (2.4). This nonoverlapping wave-packet condition means that the set of wave-packets, $\{\psi_i\}_{i=1}^N$, composing the wavefunction, $\psi_0(\mathbf{x})$, are nearly isolated Dirac delta functions at time, $t = 0$.

We then proceed to reduce the matching condition into two separate components: (1) matching function terms and (2) matching constant terms. First, if the function terms are to be matched at $\{\mathbf{x}_j\}$, the nonoverlapping condition implies that for $\varepsilon \ll 1$, then $\beta = \frac{2}{\varepsilon}$. This necessary condition suggests that the absolute critical temperature is

$$T_0 = \frac{1}{2}\varepsilon. \tag{2.25}$$

We show in the following that when $\beta = \frac{2}{\varepsilon}$, not only are the two functions matched at $\{\mathbf{x}_j\}$, but they are approximately identity on the whole space. If we denote an effective support of the ith wavepacket, ψ_i, as Ω_i^ε, we conclude that $w_i(\mathbf{x}) \approx 1$ when $\mathbf{x} \in \Omega_i^\varepsilon$, otherwise $w_i(\mathbf{x}) \approx 0$. By using the identity $\sum_{i=1}^{N} w_i(\mathbf{x}) = 1$, we have when $\beta = \frac{2}{\varepsilon}$,

$$e^{-\frac{\beta}{2}\sum_{i=1}^{N} w_i(\mathbf{x})\frac{|\mathbf{x}-\mathbf{x}_i|^2}{2}} = \sum_{j=1}^{N} w_j(\mathbf{x}) e^{-\sum_{i=1}^{N} w_i(\mathbf{x})\frac{|\mathbf{x}-\mathbf{x}_i|^2}{2\varepsilon}} \approx \sum_{j=1}^{N} e^{-\frac{|\mathbf{x}-\mathbf{x}_j|^2}{2\varepsilon}}. \tag{2.26}$$

Thus, we have shown that the function component of the matching condition is satisfied.

Next, to match the constant terms, we note that since C_N and Z are both normalization constants,

$$(C_N)^2 = \int \left(\sum_{i=1}^{N} e^{-\frac{|\mathbf{x}-\mathbf{x}_i|^2}{2\varepsilon}} \right)^2 d\mathbf{x} \approx \int e^{-\sum_{i=1}^{N} w_i(\mathbf{x})\frac{|\mathbf{x}-\mathbf{x}_i|^2}{\varepsilon}} d\mathbf{x} = Z e^{\beta C}. \tag{2.27}$$

Further, the nonoverlapping condition implies that when $i \neq j$, the wave-packets are approximately orthogonal, i.e.,

$$\int \left(e^{-\frac{|\mathbf{x}-\mathbf{x}_i|^2}{2\varepsilon}} \right) \left(e^{-\frac{|\mathbf{x}-\mathbf{x}_i|^2}{2\varepsilon}} \right) d\mathbf{x} \approx 0. \tag{2.28}$$

Then, by direct calculation, we get

$$\int \left(\sum_{i=1}^{N} e^{-\frac{|\mathbf{x}-\mathbf{x}_i|^2}{2\varepsilon}} \right)^2 d\mathbf{x} \approx \int \left(\sum_{i=1}^{N} e^{-\frac{|\mathbf{x}-\mathbf{x}_i|^2}{\varepsilon}} \right) d\mathbf{x} = (\pi\varepsilon)^k \sum_{i=1}^{N} 1 = N(\pi\varepsilon)^k, \tag{2.29}$$

which is the quantum mechanical normalization constant for a Gaussian wavefunction in k dimensions. Also, due to the nonoverlapping condition, we know that $\{\mathbf{x}_i\}_{i=1}^{N}$ are approximate zeros of $V(\mathbf{x})$. By the method of deepest decent,

$$\int e^{-\sum_{i=1}^{N} w_i(\mathbf{x})\frac{|\mathbf{x}-\mathbf{x}_i|^2}{\varepsilon}} d\mathbf{x} \approx (\pi\varepsilon)^k \sum_{i=1}^{N} (1 + O(\varepsilon)). \tag{2.30}$$

Therefore, we have verified that when $\beta = \frac{2}{\varepsilon}$,

$$\rho_A(\mathbf{x}) = |\psi_0(\mathbf{x})|^2 \approx \frac{1}{Z} e^{-\beta V(\mathbf{x})} = \rho_B(\mathbf{x}), \tag{2.31}$$

and the quantum distribution has been approximately matched with the classical Gibbs distribution (i.e., equations (1.1) and (1.7) are asymptotically matched at the critical temperature, $T = \frac{1}{2}\varepsilon$).

The illustrating example in Figure 2 demonstrates quantum and classical probability distributions matched at critical temperature, $T_0 = 1/\beta_0$ (equation (2.25)), by using two 1D Gaussian wave-packets, ψ_i and ψ_j, separated by various distances along the x-axis. When the nonoverlapping condition (equation (2.24)) is either completely or nearly satisfied based on negligible wave-packet interference, we demonstrate that the quantum and classical probability distributions are in reasonably good agreement. However, this agreement fails when the overlap between the two wave-packets becomes large such that $w_i(\mathbf{x}_j) \neq \delta_{ij}$.
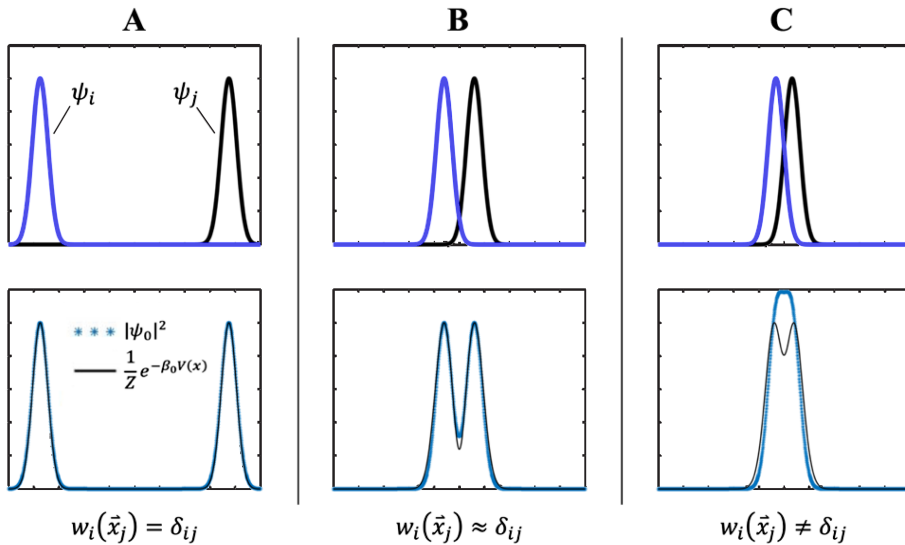

FIG. 2. Illustration of quantum and classical probability distributions matched at critical temperature, $T_0 = \beta_0^{-1}$. Two 1D Gaussian wave-packets, ψ_i and ψ_j, are separated by various distances along the horizontal axis (top). The probability distributions (bottom) are in good agreement when the nonoverlapping condition is either: (A) completely satisfied, i.e., $w_i(\mathbf{x}_j) = \delta_{ij}$, or (B) nearly satisfied, i.e., $w_i(\mathbf{x}_j) \approx \delta_{ij}$. (C) The probability distributions do not agree when the distance between the two wave-packets becomes small, where the nonoverlapping condition is not satisfied, i.e., $w_i(\mathbf{x}_j) \neq \delta_{ij}$.

3. Time integrator for Langevin dynamics. Let us recall for convenience the Langevin dynamics that are used to propagate the position and momentum of each wave-packet,

$$d\mathbf{q} = \mathbf{p} \, dt \tag{3.1}$$

and

$$d\mathbf{p} = -\nabla_q V(\mathbf{q}) \, dt - \gamma \mathbf{p} \, dt + \sqrt{2\gamma\beta^{-1}} \, dB, \tag{3.2}$$

where $\mathbf{q}, \mathbf{p} \in \mathbb{R}^k$ are k-dimensional vectors of instantaneous position and momenta, respectively. To numerically integrate the Langevin dynamics, we apply the $BAOAB$ splitting scheme. This is a numerical integration technique previously developed by [26] to address classical molecular dynamics problems regarding a Langevin thermostat. Here, the Langevin dynamics are decomposed into three components: a kinetic component (denoted by "A"),

$$(d\mathbf{q}, d\mathbf{p}) = (\mathbf{p}dt, \ 0), \tag{3.3}$$

a potential component (denoted by "B"),

$$(d\mathbf{q}, d\mathbf{p}) = (0, \ -\nabla_q V(\mathbf{q}) \, dt), \tag{3.4}$$

and a Langevin thermostat component (denoted by "O", for the Ornstein–Uhlenbeck equation),

$$(d\mathbf{q}, d\mathbf{p}) = \left(0, -\gamma\mathbf{p} \, dt + \sqrt{2\gamma\beta^{-1}} \, dB\right). \tag{3.5}$$

The BAOAB procedure is then carried out as follows. The integration is discretized into **five** symmetric phases per discrete time-step, Δt:
  (1) the potential component (B) is integrated over $\frac{1}{2}\Delta t$,
  (2) the kinetic component (A) is integrated over $\frac{1}{2}\Delta t$,
  (3) the Langevin thermostat component (O) is integrated over Δt,
  (4) the kinetic component (A) is integrated over $\frac{1}{2}\Delta t$,
  (5) the potential component (B) is integrated over $\frac{1}{2}\Delta t$.

As each component is separable in position and momentum, each can be integrated explicitly, leading to high accuracy, reliable stability, and low computational cost [26, 30]. The kinetic and potential components correspond to drift and kick, respectively, as a result of global environmental interactions according to equation (2.15) [26]. The Langevin thermostat component, defined via equation (3.5), has the following Ornstein–Uhlenbeck solution:

$$(\mathbf{q}(t), \mathbf{p}(t)) = \left(\mathbf{q}(0), \quad e^{-\gamma t}\mathbf{p}(0) + \sqrt{(1 - e^{-2\gamma t})(\beta^{-1})}\mathcal{N}\right), \tag{3.6}$$

where $\mathcal{N}$ is a k-dimensional, uncorrelated Gaussian random field. This component governs local interactions, making it possible for local minima to merge as a consequence of thermal tunneling phenomena. As the individual components of the $BAOAB$ scheme essentially reduce to matrix level operations with analytic solutions, the entire algorithm is highly parallelizable, making it an ideal candidate for modern graphics-card enabled hardware.

4. Numerical results and validation. In this section, we investigate the feasibility of data clustering with Langevin dynamics, using a standard benchmark dataset. Several computational experiments have been conducted to both validate the proposed methodology and better understand the choice of parameters driving clustering. Section 4.1 introduces the benchmark dataset (and subsequent preprocessing steps) that was used to produce the subsequent numerical results, and Section 4.2 investigates clustering performance and hyper-parametrization.

4.1. *Benchmark data and preprocessing.*

4.1.1. *Ripley's crab data.* Ripley's Crab Data [34] was used as the basis of numerical experimentation and algorithm testing in this section. This is a standard dataset commonly used as a benchmark for classification problems. The 5-dimensional data contains 200 samples of *Leptograpsus variegatus* crabs, each labeled as one of the following classes: (A) blue male, (B) blue female, (C) orange male, (D) orange female. Each sample consists of 5 morphological features—one for each dimension of the problem—which collectively capture sample-specific, quantitative information.

Accordingly, each sample is mathematically represented by a 5-dimensional feature vector,

$$\{\mathbf{x}_i\}_{i=1}^{200} \subset \mathbb{R}^5.$$

The data is readily available in MATLAB (Mathworks, Natick, MA), where it is stored as a 200x5 matrix, X. Columns of X correspond to a particular feature, $\mathbf{x}_j$, obtained from all samples, and rows of X correspond to the 5-dimensional feature vectors, $\mathbf{x}_i$, representing each sample. To ensure normalized units, each column vector, $\mathbf{x}_j$, was zero-mean centered. This is a common preprocessing step in many statistical learning approaches, in particular, Principle Component Analysis (PCA) [22]. In general, the goal of the benchmark dataset is to reconstruct the categorical classes, based exclusively on the quantitative information encoded within the feature vectors.

4.1.2. *Singular value decomposition.* Rather than clustering X in its original 5-dimensional basis, we first perform a singular value decomposition (SVD),

$$X = \sum_{j=1}^{5} \sigma_j \vec{u}_j \vec{v}_j^{\dagger}, \tag{4.1}$$

where $\vec{u}_j$ and $\vec{v}_j$ are the left and right singular vectors of X, and σ_j are the singular values of X. The vectors, $\vec{u}_j$, define a practical 5-dimensional coordinate system for the data that is particularly useful for clustering problems [18, 41, 44]. To reduce the number of dimensions, we truncate equation (4.1) to its first 3 leading terms, and cluster the points, $\{\vec{y}_i\}_{i=1}^{200} \subset \mathbb{R}^3$, within the 3-dimensional space defined by the principal coordinate axes, $(\vec{u}_1, \vec{u}_2, \vec{u}_3)$.

In general, hyper-dimensional statistical learning algorithms are often subject to overfitting artifacts as a result of too many degrees of freedom built into models. In fact, feature selection is a common and important first step to any modern machine learning approach. It is highly likely that our proposed algorithm should also benefit from reducing the number of dimensions in a given dataset. In particular, as the number of dimensions grows, the ability to find optimal values for ε and T becomes increasingly

nontrivial.[7] The authors emphasize, however, that working in the truncated SVD space is only a preprocessing procedure used to reduce the dimensionality of the data and select a stable basis. While it provides a convenient coordinate basis to work in, SVD is not inherent to the clustering process itself. Other, possibly nonlinear, basis transformations may also provide fruitful results. Further, as others have noted [19, 41], a well-defined separation is observed in Ripley's Crab Data between the 2nd and 3rd principal dimensions of the SVD space, i.e., $(\vec{u}_2, \vec{u}_3)$. In order to provide a comparison with the literature, many of the figures that follow are displayed across these principal axes.

4.2. *Clustering performance.* In this section, we use Ripley's Crab Data to demonstrate data clustering via Langevin dynamics at subcritical temperature. The authors stress that the data has been *retrospectively* color-coded to differentiate known classes. This was done for demonstration purposes in order to easily monitor the dynamic clustering process. The algorithm, however, was blind to the data classes.

Different potential functions at various scales were constructed according to equation (2.8) with semi-classical parameters ranging from $\varepsilon = 0.0001$ to $\varepsilon = 0.01$. For each resolution scenario, the data points, $\{\vec{y}_i\}_{i=1}^{200} \subset \mathbb{R}^3$, were propagated according to equations (2.13) and (2.14) at subcritical temperature, $T = 0.01T_0 = (0.01)(0.5\varepsilon)$. Each ε-specific dynamic process was run for a total time, $t_{\max} = 10$, and a time-step, $dt = 0.1$.

Clustering performance was evaluated based on a generalized Jaccard score, J. This is a similarity metric comparing the algorithm-produced clustering results with known class labels obtained a priori from the benchmark metadata [39]. Let $\mathbf{a} = (a_1, a_2, \ldots, a_{200})$ be a vector whose elements correspond to the known class labels, and let $\mathbf{b} = (b_1, b_2, \ldots, b_{200})$ be a vector whose elements correspond the clustering results such that $a_i, b_i \in \{1, 2, 3, 4\} \subset \mathbb{Z}$. The Jaccard score is then

$$J(\mathbf{a}, \mathbf{b}) = \frac{\sum_i \min(a_i, b_i)}{\sum_i \max(a_i, b_i)}. \tag{4.2}$$

Essentially, J is a generalization of the receiver operating characteristic, extended to nonbinary classification problems. Here, it is used to quantify the number of data points appearing in the same class according to both the algorithm and ground truth. A Jaccard score of $J = 1$ indicates perfect clustering.

Figure 3A shows the Jaccard score achieved for the range of ε values at the subcritical temperature, $T = 0.01T_0$, for a total time, $t_{\max} = 10$, and a time step, $dt = 0.1$. As indicated by the red dashed vertical line, the performance of the algorithm was maximized at $\varepsilon = 1.225E - 3$, resulting in $J = 0.90$. Figure 3B demonstrates a projection of the potential (the gray-scale contour map) at this resolution, as well as the final location of each data point according to Langevin dynamics. Reasonable clustering performance ($J \geq 0.6$) was archived for a relatively stable range of ε values within at the semi-classical scale, as indicated by the curve's broad shoulder between 0.001 and 0.005. As $\varepsilon \to 0$, performance is reduced due to decreasing interaction among data points and is

---

[7]In general, the optimization of ε and T in arbitrary hyper-dimensions is a challenging task and is a key focus of our future work.
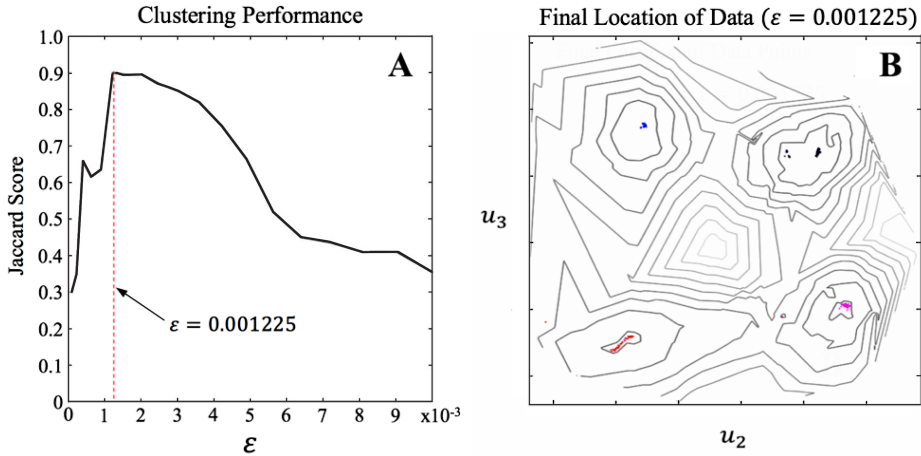
FIG. 3. Sensitivity of clustering performance to different ε values.
(A) Jaccard scores, indicating clustering performance, were calcu-
lated at a subcritical temperature of $T = 0.01T_0$. Clustering was
optimized at $\varepsilon = 0.001225$, resulting in a maximum Jaccard score
of $J = 0.90$. (B) The final location of each data point following
Langevin dynamics (plotted in color), relative to a projection of the
potential surface (gray-scale contour plot) at $\varepsilon = 0.001225$. Each
color corresponds to a different data class, demonstrating optimal
separation at distinct potential minima. Here, u_2 and u_3 represent
the 2nd and 3rd principal component axes, respectively.

demonstrated as J drops to 0.3 at very low ε. Similarly, as $\varepsilon \to 1$, quality clustering
is no longer achieved. In this non-semi-classical limit, the Jaccard score tappers to
a suboptimal $J = 0.3$. Further, we note that clustering at critical temperature only
produced a maximum Jaccard score of $J = 0.72$, and that temperatures of $T \leq 0.01T_0$
resulted in particularly stable clustering results.

The following figures (Figures 4, 5, Figure 6) all demonstrate different aspects of
the dynamic clustering process with input parameters: $(\varepsilon, T) = (1.225E - 3, \ 0.01T_0)$.
Specifically, Figure 4 shows the time-evolution of the data as a series of sequential snap-
shots, comparing the dynamics between critical (Figure 4A) and subcritical temperatures
(Figure 4B). The potential function is shown as a 2D projection across the 2nd and 3rd
principal component axes, and the data points have been retrospectively color-coded
according to the known data classes. At each sequential time-evolution frame, the differ-
ence between critical and subcritical evolution becomes more and more apparent. The
subcritical Langevin dynamics were run at a resolution of $\varepsilon = 1.225E - 3$, with a tem-
perature of $T = 0.01T_0$, a total time, $t_{\max} = 10$, and a time step, $dt = 0.1$. While data
aggregation is easily observed at $T = 0.01T_0$, the Langevin dynamics at critical temper-
ature, $T = T_0$, unveils a much less obvious structure to the dataset. This further fulfills
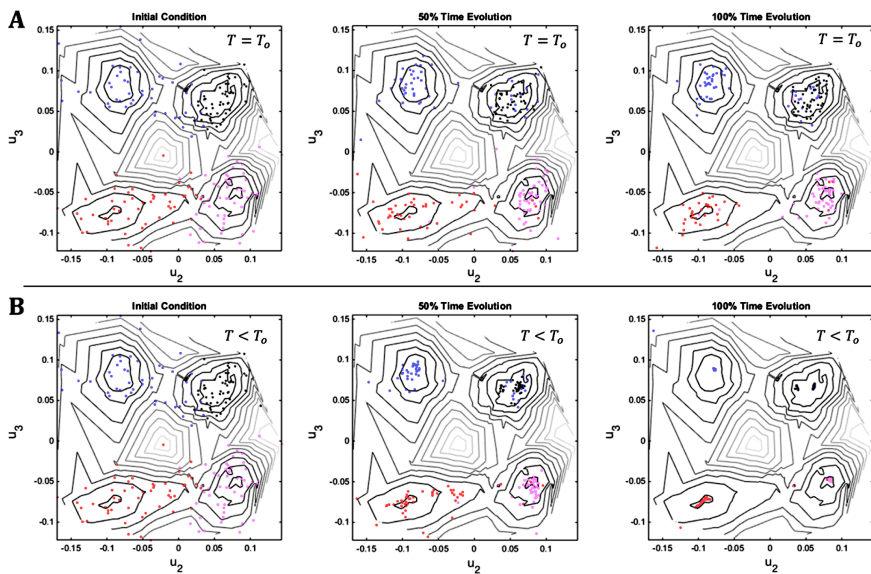our intuition that clustering is a subcritical temperature phenomenon.

FIG. 4. Data clustering as a result of Langevin annealing. The time-dependent location of each data point is plotted, from left to right, relative to a projection of the potential surface at $\varepsilon = 0.001225$. Here, u_2 and u_3 represent the 2nd and 3rd principal component axes, respectively. (A) At critical temperature, $T = T_0$, the Langevin dynamics produce a Jaccard score of $J = 0.72$. (B) At a lower subcritical temperature, $T < T_0$, the Langevin dynamics produce a Jaccard score of $J = 0.90$. Here, a clear separation of the data points is demonstrated, as each color corresponds to a different data class.

Next, Figure 5 shows the same time-evolution process as Figure 4, but follows the location of each data point across all 3 principal coordinates of the truncated SVD space. Even though the known separation in the data classes is optimal across the 2nd and 3rd dimensions, Figure 5 demonstrates that the spatial information encoded within the other dimensions is not lost during the clustering process. The location of the data points, which are again retrospectively color-coded according to class, are shown to converge towards similar coordinates.
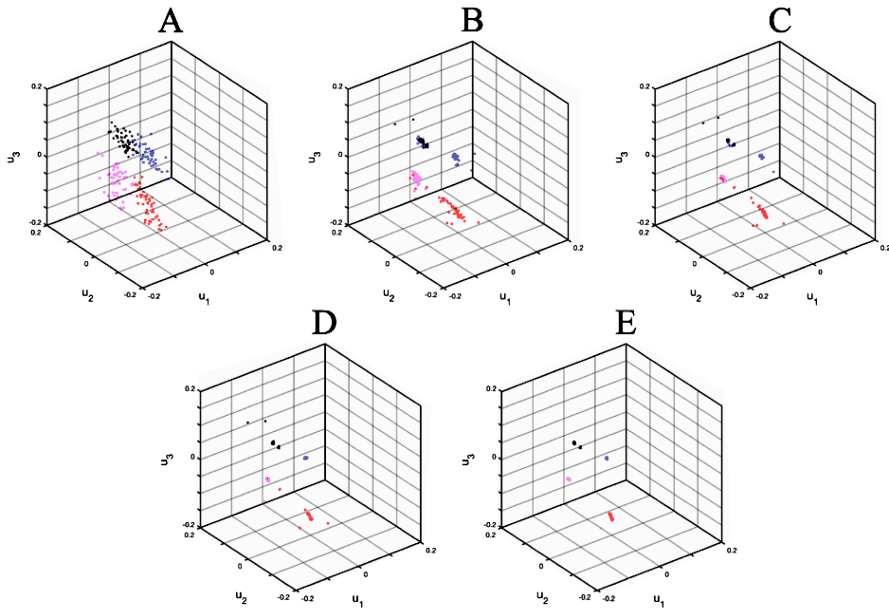
FIG. 5. Time-dependent location of data points in 3 principal dimensions. The time-dependent location of each data point is plotted as a series of 5 sequential snap shots (A–E) for $\varepsilon = 0.001225$ and $T = 0.01T_0$. Here, u_1, u_2, and u_3 represent the 1st, 2nd, and 3rd principal component axes, respectively. (E) Data points of the same class (differentiated by color) are shown to aggregate at similar spatial locations as a result of the subcritical Langevin dynamics, producing a Jaccard score of $J = 0.90$.

Finally, Figure 6 demonstrates the time-dependent changes of the weight matrix, w_{ij}, defined via equation (2.17), as a result of the Langevin dynamics. Clearly, at 10% of the time-evolution, w_{ij} is still densely populated along the main diagonal (thus satisfying equation (2.24), the nonoverlapping condition). As the dynamic process continues, however, the four data classes are observed within w_{ij}, demonstrated by high intensity regions extending from the diagonal. From 70%–100% of the time-evolution, four sets of similar row-vectors are observed within w_{ij}, indicating that the corresponding data points have saturated to similar locations as a result of the dynamics.
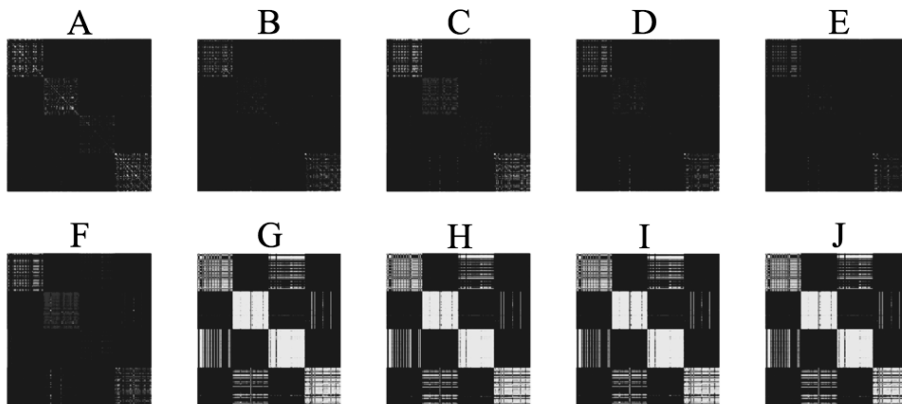
FIG. 6. Data point overlap as a function of time evolution. The Langevin dynamics for $\varepsilon = 0.001225$ and $T = 0.01T_0$ produce changes in data overlap, as measured by the weight matrix, w_{ij} (equation (2.17)). (A–J) w_{ij} is shown as a series of 10 sequential snap shots in time, where the brightness (intensity) indicates the degree of data overlap. High intensity (white) matrix values of w_{ij} represent a significant overlap between the ith and jth data points. (G–J) The response of w_{ij} to Langevin dynamics results in a saturated diagonal block behavior, capturing the 4 data classes.

5. Summary and outlook. In this paper, we have proposed a dynamic approach to data clustering based on fundamental concepts borrowed from both quantum mechanics and statistical mechanics. Due to the alarming rate at which *big data* is being produced, the ability to understand hyper-dimensional datasets is becoming increasingly nontrivial. The development of novel clustering algorithms—such as the one outlined in this paper— is therefore an essential aspect of contemporary data analytics.

Our proposed methodology is briefly summarized as follows. A given dataset is interpreted as a classical ensemble in equilibrium with a heat bath. This produces an equilibrium distribution uniquely characterized by a temperature, T, and a resolution, ε. Langevin dynamics are then employed to stochastically propagate data points on ergodic trajectories obeying a classical Gibbs distribution according to the appropriate potential function. Data points which end up in the neighborhood of a particular potential minimum are ultimately considered to be in the same cluster. Time-evolution according to Langevin dynamics has several practical advantages with respect to robust data clustering. Most remarkably is the ability for data points to thermally jump local potential barriers and escape saddle points into locations of the potential surface otherwise forbidden. This tunneling phenomenon contributes to the overall dynamic optimization and ultimate partitioning of the data into degenerate subsets.

The clustering methodology was applied to Ripley's Crab Data, which is a common 5-dimensional benchmark dataset often used in classification problems. Accurate clustering of the data classes was demonstrated at subcritical temperatures. Such a critical temperature, T_0, was derived by asymptotically matching the quantum and classical distributions. Clustering performance was evaluated via the Jaccard score, where the

algorithm achieved a maximum value of 90% at $T = 0.01T_0$ and $\varepsilon = 1.225E-3$. At this same resolution, critical temperature Langevin dynamics produced a slightly worse Jaccard score of 72%.

Others have demonstrated that Langevin dynamics can be implemented to better optimize conventional statistical learning techniques, including Bayesian learning [42] and deep neural networks [27]. Our algorithm can potentially be improved in a similar manner by coupling it with traditional machine learning algorithms. In our case, the optimal scale to probe data is different for various datasets, and therefore choosing reliable ε and T values is fundamentally a task-based endeavor. Further, these parameters need to be chosen in a sequential way: ε defines an *ideal* potential landscape, and T defines the dynamics that yield the *desired* clustering results. Pairing our current approach with traditional statistical learning techniques would generalize such hyper-parameterization in a task-based manner. For example, implementation of a Bayesian feedback loop tertiary to the main algorithm would continuously improve the clustering performance as more data is warehoused to support a particular application. Thus, in principle, the methodology presented in this paper only represents the core component of a potentially much larger computational pipeline with a broad range of applications.

Other future research will consist of a detailed comparison study between our approach with other dynamic clustering algorithms. We also plan to (a) explore new challenges that arise when the number of dimensions grows significantly large; and (b) investigate the potential benefit of colored-noise Langevin dynamics, where a spatially dependent temperature gradient may produce new, fruitful clustering results.

References

[1] D. Bambusi, S. Graffi, and T. Paul, *Long time semiclassical approximation of quantum flows: a proof of the Ehrenfest time*, Asymptot. Anal. **21** (1999), no. 2, 149–160. MR1723551

[2] A. L. Bertozzi and A. Flenner, *Diffuse interface models on graphs for classification of high dimensional data [reprint of MR3022033]*, SIAM Rev. **58** (2016), no. 2, 293–328, DOI 10.1137/16M1070426. MR3493948

[3] A. Bewley and B. Upcroft, *Advantages of exploiting projection structure for segmenting dense 3d point clouds*, Proceedings of Australasian Conference on Robotics and Automation (2013).

[4] L. V. Bijuraj, *Clustering and its applications*, Proceedings of National Conference on New Horizona in IT (2013).

[5] K. Blekas and I. E. Lagaris, *Newtonian clustering: An approach based on molecular dynamics and global optimization*, Pattern Recognition **40** (2007), 1734–1744.

[6] T. Buhler and M. Hein, *Spectral clustering based on the graph p-laplacian*, Proceedings of the 26th International Conference on Machine Learning (2009), 81–88.

[7] W. T. Coffey, Yu. P. Kalmykov, and J. T. Waldron, *The Langevin equation, with applications to stochastic problems in physics, chemistry and electrical engineering*, 2nd ed., World Scientific Series in Contemporary Chemical Physics, vol. 14, World Scientific Publishing Co., Inc., River Edge, NJ, 2004. MR2053912

[8] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker, *Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps*, Proceedings of the National Academy of Sciences **102** (2005), no. 21, 1788–1794.

[9] Y. Dauphin, R. Pascanu, C. Gulcehre, K. Cho, S. Ganguli, and Y. Bengio, *Identifying and attacking the saddle point problem in high-dimensional non-convex optimization*, Proceedings of the 27th International Conference on Neural Information Processing Systems (2014), 2933–2941.

[10] V. Estivill-Castro, *Why so many clustering algorithms — a position paper*, ACM SIGKDD Explorations Newsletter **4** (2002), no. 1, 65–75.

[11] D. J. Evans, D. J. Searles, and S. R. Williams, *Fundamentals of classical statistical thermodynamics. Dissipation, relaxation and fluctuation theorems*, WILEY-VCH Verlag Berlin GmbH, Weinheim, 2016. MR3497107

[12] R. Filipovych, S. M. Resnick, and C. Davatzikos, *Semi-supervised cluster analysis of imaging data*, Neuroimage **54** (2011), no. 3, 2185–2197.

[13] C. Garcia-Cardona, A. Flenner, and A. G. Percus, *Diffuse interface models on graphs for classification of high dimensional data*, Proceedings of the 2nd International Conference on Pattern Recognition Applications and Methods (2013), 78–86.

[14] R. Ge, F. Huang, C. Jin, and Y. Yuan, *Escaping from saddle points – online stochastic gradient for tensor decomposition*, JMLR: Workshop and Conference Proceedings **40** (2015), 1–46.

[15] P. Gérard, P. A. Markowich, N. J. Mauser, and F. Poupaud, *Homogenization limits and Wigner transforms*, Comm. Pure Appl. Math. **50** (1997), no. 4, 323–379, DOI 10.1002/(SICI)1097-0312(199704)50:4⟨323::AID-CPA4⟩3.3.CO;2-Q. MR1438151

[16] G. A. Hagedorn and A. Joye, *Exponentially accurate semiclassical dynamics: propagation, localization, Ehrenfest times, scattering, and more general states*, Ann. Henri Poincaré **1** (2000), no. 5, 837–883, DOI 10.1007/PL00001017. MR1806980

[17] D. Horn, *Clustering via Hilbert space*, Phys. A **302** (2001), no. 1-4, 70–79, DOI 10.1016/S0378-4371(01)00442-3. MR1876430

[18] D. Horn and I. Axel, *Novel clustering algorithm for microarray expression data in a truncated svd space*, Bioinformatics **19** (2003), no. 15, 1110.

[19] D. Horn and A. Gottlieb, *Algorithm for data clustering in pattern recognition problems based on quantum mechanics*, Physical Review Letters **88** (2002), no. 1, 018702.

[20] A. K. Jain, *Data clustering: 50 years beyond k-means*, Pattern Recognition Letters **31** (2010), no. 8, 651–666.

[21] A. K. Jain, M. N. Murty, and P. J. Flynn, *Data clustering: A review*, ACM Computing Surveys **31** (1999), no. 3, 651–666.

[22] I. T. Jolliffe, *Principal component analysis*, 2nd ed., Springer Series in Statistics, Springer-Verlag, New York, 2002. MR2036084

[23] S. Kobayashi and K. Nomizu, *Foundations of differential geometry. Vol. II*, Interscience Tracts in Pure and Applied Mathematics, No. 15 Vol. II, Interscience Publishers John Wiley & Sons, Inc., New York-London-Sydney, 1969. MR0238225

[24] R. Kubo, *The fluctuation-dissipation theorem*, Reports on Progress in Physics **29** (1966), no. 1, 255–284.

[25] S. Lafon and A. B. Lee, *Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization*, IEEE Transactions on Pattern Analysis and Machine Intelligence **28** (2006), no. 9, 1393–1403.

[26] B. Leimkuhler and C. Matthews, *Robust and efficient configurational molecular sampling via Langevin dynamics*, Journal of Chemical Physics **138** (2013), no. 17.

[27] C. Li, C. Chen, D. Carlson, and L. Carin, *Preconditioned stochastic gradient Langevin dynamics for deep neural networks*, Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (2016), 1788–1794.

[28] T. W. Liao, *Clustering of time series data — a survey*, Pattern Recognition **38** (2005), no. 11, 1857–1874.

[29] P.-L. Lions and T. Paul, *Sur les mesures de Wigner* (French, with English and French summaries), Rev. Mat. Iberoamericana **9** (1993), no. 3, 553–618, DOI 10.4171/RMI/143. MR1251718

[30] J. Lu and Z. Zhou, *Path integral molecular dynamics with surface hopping for thermal equilibrium sampling of nonadiabatic systems*, Journal of Chemical Physics **146** (2017), no. 15.

[31] B. Nadler, S. Lafon, R. R. Coifman, and I. G. Kevrekidis, *Difusion maps, spectral clustering and reaction coordinates of dynamical systems*, Appl. Comput. Harmon. Anal. **21** (2006), no. 1, 113–127, DOI 10.1016/j.acha.2005.07.004. MR2238669

[32] D. Pfitzner, R. Leibbrandt, and D. Powers, *Characterization and evaluation of similarity measures for pairs of clusterings*, Knowledge and Information Systems, **16** (2009), 361–394.

[33] E. Ramasso, V. Placet, and M. L. Boubakar, *Unsupervised consensus clustering of acoustic emission time-series for robust damage sequence estimation in composites*, IEEE Transactions on Instrumentation and Measurement **64** (2015), no. 12, 3297–3307.

[34] B. D. Ripley, *Pattern recognition and neural networks*, Cambridge University Press, Cambridge, 2007. Reprint of the 1996 original. MR2451352

[35] S. J. Roberts, *Parametric and non-parametric unsupervised cluster analysis*, Pattern Recognition **30** (1997), no. 2, 261–272.

[36] A. M. Saxe, J. L. McClelland, and S. Ganguli, *Exact solutions to the nonlinear dynamics of learning in deep linear neural networks*, International Conference on Learning Representations (2014).

[37] R. L. Stratonovich, *Nonlinear nonequilibrium thermodynamics. I*, Linear and nonlinear fluctuation-dissipation theorems, translated from the Russian by V. V. Stratonovich and A. P. Repjev, vol. 57, Springer-Verlag, Berlin, 1992. MR1217972

[38] A. D. Szlam, M. Maggioni, and R. R. Coifman, *Regularization on graphs with function-adapted diffusion processes*, J. Mach. Learn. Res. **9** (2008), 1711–1739. MR2438821

[39] P.-N. Tan, M. Steinbach, A. Karpatne, and V. Kumar, *Introduction to data mining*, Pearson, 2005.

[40] U. von Luxburg, *A tutorial on spectral clustering*, Technical report TR-149, Max Planck Institute for Biological Cybernetics, Tubingen, Germany (2006).

[41] M. Weinstein and D. Horn, *Dynamic quantum clustering: a method for visual exploration of structures in data*, Physical Review E **80** (2009), 066117.

[42] Q. He and J. Xin, *Hybrid deterministic-stochastic gradient Langevin dynamics for Bayesian learning*, Commun. Inf. Syst. **12** (2012), no. 3, 221–232, DOI 10.4310/CIS.2012.v12.n3.a3. MR3157685

[43] P. Wittek, *High-performance dynamic quantum clustering on graphics processors*, Journal of Computational Physics **233** (2013), 262–271.

[44] L. Zhang, J. S. Marron, H. Shen, and Z. Zhu, *Singular value decomposition and its visualization*, J. Comput. Graph. Statist. **16** (2007), no. 4, 833–854, DOI 10.1198/106186007X256080. MR2412485