Abstract

Many biological molecules are highly complex and involve many thousand atoms. However in most cases, a substantial amount of redundancy exists due to the variable interdependency that is an obvious result of the constraints on atomic positions due to covalent bonds and energy barriers. Hence, they lie close to a low-dimensional manifold. Nonlinear dimensionality reduction methods have been proved to be very powerful to discover such intrinsic manifolds and have been widely used to visualize the inherent patterns of data. Nevertheless, performing computation on the discovered low-dimensional manifold is not feasible due to the missing link from low-dimensional to high-dimensional space. Here, we proposed a method that enables nonlinear manifold learning techniques to do computations in the low-dimension space and transfer the calculated potential to the high-dimension space.

Motivation

- Understanding the relationship between the dynamic nature of molecular structures and their properties and functions is of critical importance for many fields of research.
- Experimental methods give information to a very limited extent.
- Computational methods are very promising, but (e.g. MD):
  - large amount of DOF
  - time step femto-second, phenomena milli-second
  - enormous computational cost
- A substantial amount of redundancy exists due to the variable interdependencies.
- Dimensionality reduction can find the intrinsic low-dimensional manifold in which data lie. However, in order to use this intrinsic manifold in molecular simulations and accelerate them, we need a link between potentials in low-dimensional and high-dimensional space.

Dimensionality Reduction

Linear Methods (LDR)
- Principal Component Analysis
- Multidimensional Scaling

Nonlinear Methods (NLDR)
- Isomap
- Locally Linear Embedding

Local Maximum Entropy

(LME)_β For fixed x, minimize
\[ f_β(x, p) = J_β(x, p) - H(p) \]
subject to
\[ p_a ≥ 0, \quad a = 1, \ldots, N \]
\[ \sum_{a=1}^{N} p_a = 1, \quad \sum_{a=1}^{N} p_a x_a = x \]
Shape functions \( p = \{p_a\}_{a=1}^{N} \in \mathbb{R}_+^N \)
- Maximum entropy
  - \( H(p) = -\sum_{a=1}^{N} p_a \ln p_a \)
- Maximum locality
  - \( U_β(x, p) = \sum_{a} β_a p_a \| x - x_a \|^2 \)

Enhanced sampling methods

Integrating this method with an enhanced sampling technique shows its capability and effectiveness in calculating free energy landscape of biomolecules. Here is an example of applying Adaptive Biasing Force (ABF) on the intrinsic nonlinear manifold of Alanine dippeptide conformational space detected by nonlinear dimensionality reduction.

References