Local Feature Classification using the Euclidean Wasserstein Metric

Benjamin Schweinhart

joint work with Jeremy Mason

July 12, 2018
Main Question

What reference best matches a local atomic environment?

Input: a local atomic environment, a list of references, the temperature

Output: the most likely match
What reference best matches a local atomic environment? 
Input: a local atomic environment, a list of references, the temperature 
Output: the most likely match
Main Question

What reference best matches a local atomic environment?
Input: a local atomic environment, a list of references, the temperature
Output: the most likely match
Main Question

What reference best matches a local atomic environment?
Input: a local atomic environment, a list of references, the temperature
Output: the most likely match
Main Question

What reference best matches a local atomic environment?
Input: a local atomic environment, a list of references, the temperature
Output: the most likely match
What is the best way to distinguish local atomic environments from HCP, FCC, and BCC lattices with noise added?
Main Question

What is the best way to distinguish local atomic environments from HCP, FCC, and BCC lattices with noise added?
Application: Identifying Defects in a Polycrystal

Many existing methods: Common Neighbor Analysis, Neighborhood Distance Analysis, Centrosymmetry, Voronoi Topology, Polyhedral Template Matching, ... 

Lazar et al, 2015
Application: Identifying Defects in a Polycrystal

Many existing methods: Common Neighbor Analysis, Neighborhood Distance Analysis, Centrosymmetry, Voronoi Topology, Polyhedral Template Matching, ... 

Lazar et al, 2015
An **atomic configuration** is a point cloud in $\mathbb{R}^m$, labeled by atomic species. Our methods work for multiple species, but today we assume one.

A **local atomic environment** is a point cloud in the ball of radius $r$, $B_r$, usually centered at an atom.

For an atomic configuration $X$ and $y \in \mathbb{R}^m$, let $\varphi_r(y; X)$ be the local atomic environment of radius $r$ centered at $y$:

$$\varphi_r(y; X) = X \cap B_r(y)$$
Definitions and Notation

- **An atomic configuration** is a point cloud in $\mathbb{R}^m$, labeled by atomic species. Our methods work for multiple species, but today we assume one.
- A **local atomic environment** is a point cloud in the ball of radius $r$, $B_r$, usually centered at an atom.
- For an atomic configuration $X$ and $y \in \mathbb{R}^m$, let $\varphi_r(y; X)$ be the local atomic environment of radius $r$ centered at $y$:

$$\varphi_r(y; X) = X \cap B_r(y)$$
An atomic configuration is a point cloud in $\mathbb{R}^m$, labeled by atomic species. Our methods work for multiple species, but today we assume one.

A local atomic environment is a point cloud in the ball of radius $r$, $B_r$, usually centered at an atom.

For an atomic configuration $X$ and $y \in \mathbb{R}^m$, let $\varphi_r(y; X)$ be the local atomic environment of radius $r$ centered at $y$:

$$\varphi_r(y; X) = X \cap B_r(y)$$
An **atomic configuration** is a point cloud in $\mathbb{R}^m$, labeled by atomic species. Our methods work for multiple species, but today we assume one.

A **local atomic environment** is a point cloud in the ball of radius $r$, $B_r$, usually centered at an atom.

For an atomic configuration $X$ and $y \in \mathbb{R}^m$, let $\varphi_r(y; X)$ be the local atomic environment of radius $r$ centered at $y$:

$$\varphi_r(y; X) = X \cap B_r(y)$$
A Metric on Local Atomic Environments

Goal: define a metric $d$ on local atomic environments. It should be continuous with respect to the atomic positions including atoms moving on and off the boundary of $B_r$. That is,

$$d\left(\varphi_r(x; X), \varphi_r(y; Y)\right)$$

should be continuous in $x, y, r, \text{and the positions of all atoms.}$

Most current methods rely on the choice of the “closest $N$ points” to a root atom — not continuous in this sense!
Goal: define a metric $d$ on local atomic environments. It should be continuous with respect to the atomic positions including atoms moving on and off the boundary of $B_r$. That is,

$$d\left(\varphi_r(x; X), \varphi_r(y; Y)\right)$$

should be continuous in $x, y, r,$ and the positions of all atoms.

Most current methods rely on the choice of the “closest $N$ points” to a root atom — not continuous in this sense!
Goal: define a metric $d$ on local atomic environments. It should be continuous with respect to the atomic positions including atoms moving on and off the boundary of $B_r$. That is,

$$d\left(\varphi_r(x; X), \varphi_r(y; Y)\right)$$

should be continuous in $x, y, r$, and the positions of all atoms.

Most current methods rely on the choice of the “closest $N$ points” to a root atom — not continuous in this sense!
A Metric on Local Atomic Environments

Goal: define a metric $d$ on local atomic environments. It should be continuous with respect to the atomic positions including atoms moving on and off the boundary of $B_r$. That is,

$$d\left(\varphi_r(x; X), \varphi_r(y; Y)\right)$$

should be continuous in $x, y, r$, and the positions of all atoms.

Most current methods rely on the choice of the “closest $N$ points” to a root atom — not continuous in this sense!
A Metric on Local Atomic Environments

The metric should be invariant to Euclidean isometries. If $X_r$ and $Y_r$ are local atomic environments then

$$d(X_r, Y_r) = d(X_r, \rho(Y_r))$$

for all $\rho \in \text{SO}(m)$. 

The metric should be invariant to Euclidean isometries. If $X_r$ and $Y_r$ are local atomic environments then

$$d(X_r, Y_r) = d(X_r, \rho(Y_r))$$

for all $\rho \in \text{SO}(m)$. 

A Metric on Local Atomic Environments

The metric should be invariant to Euclidean isometries. If $X_r$ and $Y_r$ are local atomic environments then

$$d(X_r, Y_r) = d\left(X_r, \rho(Y_r)\right)$$

for all $\rho \in \text{SO}(m)$. 
The 2-Wasserstein Metric

\[ d_W(X, Y) = \min_{\eta: X \to Y} \left( \sum_{x \in X} d(x, \eta(x))^2 \right)^{1/2} \]

The minimum is taken over all matchings \( \eta \) from \( X \) to \( Y \).
The Local 2-Wasserstein Metric

\[ d_{LW}(X, Y) = \min_{X' \subseteq X, Y' \subseteq Y, |X'| = |Y'|} d_W(X', Y')^2 \]

\[ + \sum_{x \in X \setminus X'} d(x, \partial B_r)^2 + \sum_{y \in Y \setminus Y'} d(y, \partial B_r)^2 \]^{1/2}
The Euclidean Wasserstein Metric

\[ d_{EW}(X, Y) = \min_{\rho \in SO(m)} d_{LW}(X, \rho(Y)) \]
The Euclidean Wasserstein Metric

\[ d_{EW}(X, Y) = \min_{\rho \in SO(m)} d_{LW}(X, \rho(Y)) \]
Computation

- We compute the Euclidean Wasserstein Metric by a branch-and-bound algorithm on SO(3), taking advantage of the symmetries of the reference conditions.
- We compute the Local Wasserstein Metric using the Hungarian algorithm.
- Speed (using heuristics): \( \approx 0.16 \) seconds per local environment per thread for an FCC reference with 19 atoms, \( \approx 1.3 \) seconds per local environment per thread for an FCC reference with 55 atoms.

Li and Hartley, 2007
Computation

- We compute the Euclidean Wassserstein Metric by a branch-and-bound algorithm on SO(3), taking advantage of the symmetries of the reference conditions.
- We compute the Local Wasserstein Metric using the Hungarian algorithm.
- Speed (using heuristics): \( \approx .16 \) seconds per local environment per thread for an \( FCC \) reference with 19 atoms, \( \approx 1.3 \) seconds per local environment per thread for an \( FCC \) reference with 55 atoms.

Li and Hartley, 2007
We compute the Euclidean Wasserstein Metric by a branch-and-bound algorithm on SO(3), taking advantage of the symmetries of the reference conditions.

We compute the Local Wasserstein Metric using the Hungarian algorithm.

Speed (using heuristics): $\approx 0.16$ seconds per local environment per thread for an FCC reference with 19 atoms, $\approx 1.3$ seconds per local environment per thread for an FCC reference with 55 atoms.

---

Li and Hartley, 2007
Proposed Test

- **Data**: a local atomic environment \(X\) centered at an atom \(x\), reference atomic configurations \(R_1, \ldots, R_k\), a temperature \(T\), and a radius \(r\).

- For each \(R_i\), let \(R'_i\) be the random configuration:
  \[
  R'_i = R_i + \text{thermal noise}
  \]

  If \(y \in R_i\), let \(y'\) be the corresponding atom in \(R'_i\).

- For each reference \(R_i\), let \(y \in R_i\) and compute
  \[
  p_i = \mathbb{P}
  \left( d_{EW}(X, \varphi_r(y\ Y)) < d_{EW}(\varphi_r(y'; Y'), \varphi_r(y; Y)) \right)
  \]

- Classify \(X\) as type \(R_j\) if \(p_i = \max_j p_j\).

Assume all atoms in the reference are identical, for clarity.
Proposed Test

- Data: a local atomic environment \( X \) centered at an atom \( x \), reference atomic configurations \( R_1, \ldots, R_k \), a temperature \( T \), and a radius \( r \).
- For each \( R_i \), let \( R'_i \) be the random configuration:
  \[
  R'_i = R_i + \text{thermal noise}
  \]
  If \( y \in R_i \), let \( y' \) be the corresponding atom in \( R'_i \).
- For each reference \( R_i \), let \( y \in R_i \) and compute
  \[
  p_i = \mathbb{P}\left( d_{EW}(X, \varphi_r(y; Y)) < d_{EW}(\varphi_r(y'; Y'), \varphi_r(y; Y)) \right)
  \]
  Classify \( X \) as type \( R_j \) if \( p_i = \max_j p_j \).

Assume all atoms in the reference are identical, for clarity.
Proposed Test

- Data: a local atomic environment $X$ centered at an atom $x$, reference atomic configurations $R_1, \ldots, R_k$, a temperature $T$, and a radius $r$.
- For each $R_i$, let $R'_i$ be the random configuration:
  \[
  R'_i = R_i + \text{thermal noise}
  \]
  If $y \in R_i$, let $y'$ be the corresponding atom in $R'_i$.
- For each reference $R_i$, let $y \in R_i$ and compute
  \[
  p_i = \mathbb{P} \left( d_{EW} \left( X, \varphi_r(y; Y) \right) < d_{EW} \left( \varphi_r(y'; Y'), \varphi_r(y; Y) \right) \right)
  \]
- Classify $X$ as type $R_j$ if $p_i = \max_j p_j$.

Assume all atoms in the reference are identical, for clarity.
Proposed Test

- Data: a local atomic environment $X$ centered at an atom $x$, reference atomic configurations $R_1, \ldots, R_k$, a temperature $T$, and a radius $r$.
- For each $R_i$, let $R_i'$ be the random configuration:
  \[ R_i' = R_i + \text{thermal noise} \]
  If $y \in R_i$, let $y'$ be the corresponding atom in $R_i'$.
- For each reference $R_i$, let $y \in R_i$ and compute
  \[
  p_i = \mathbb{P} \left( d_{EW}(X, \varphi_r(y; Y)) < d_{EW}(\varphi_r(y'; Y'), \varphi_r(y; Y)) \right)
  \]
  Classify $X$ as type $R_j$ if $p_i = \max_j p_j$.

Assume all atoms in the reference are identical, for clarity.
Proposed Test

- Data: a local atomic environment \( X \) centered at an atom \( x \), reference atomic configurations \( R_1, \ldots, R_k \), a temperature \( T \), and a radius \( r \).
- For each \( R_i \), let \( R'_i \) be the random configuration:
  \[
  R'_i = R_i + \text{thermal noise}
  \]
  If \( y \in R_i \), let \( y' \) be the corresponding atom in \( R'_i \).
- For each reference \( R_i \), let \( y \in R_i \) and compute
  \[
  p_i = \mathbb{P}\left( d_{EW}(X, \varphi_r(y; Y)) < d_{EW}(\varphi_r(y'; Y'), \varphi_r(y; Y)) \right)
  \]
- Classify \( X \) as type \( R_j \) if \( p_i = \max_j p_j \).

Assume all atoms in the reference are identical, for clarity.
Synthetic Data

- References: FCC, BCC, HCP lattices with unit bond length.
- Thermal noise: add Gaussian displacements to each atom with standard deviation $\sigma$.
- For each reference $R$ and $\sigma \in \mathbb{R}^+$, compute the probability that a local environment in $R'$ is classified correctly.
- Compare accuracy with methods implemented in OVITO.

Stukowski, 2010
Synthetic Data

- References: FCC, BCC, HCP lattices with unit bond length.
- Thermal noise: add Gaussian displacements to each atom with standard deviation $\sigma$.
- For each reference $R$ and $\sigma \in \mathbb{R}^+$, compute the probability that a local environment in $R'$ is classified correctly.
- Compare accuracy with methods implemented in OVITO.

---

Stukowski, 2010
Synthetic Data

- References: FCC, BCC, HCP lattices with unit bond length.
- Thermal noise: add Gaussian displacements to each atom with standard deviation $\sigma$.
- For each reference $R$ and $\sigma \in \mathbb{R}^+$, compute the probability that a local environment in $R'$ is classified correctly.
- Compare accuracy with methods implemented in OVITO.

Stukowski, 2010
Synthetic Data

- References: FCC, BCC, HCP lattices with unit bond length.
- Thermal noise: add Gaussian displacements to each atom with standard deviation $\sigma$.
- For each reference $R$ and $\sigma \in \mathbb{R}^+$, compute the probability that a local environment in $R'$ is classified correctly.
- Compare accuracy with methods implemented in OVITO.

Stukowski, 2010
Performance Comparison: FCC

![Graph showing the performance comparison of different methods against standard deviation (in % of bond length). The graph plots the probability of detecting FCC (P(FCC Detected | FCC)) against the standard deviation. The methods include Euclidean Wasserstein with different radii, Polyhedral Template Matching, Neighborhood Distance Analysis, Adaptive Common Neighbor Analysis, Bond Angle Analysis, and Voronoi Topology. Each method is represented by a different line color and style.](image-url)
Performance Comparison: FCC
Performance Comparison: HCP
Performance Comparison: BCC
Different test for atomic environments with large non-thermal strains: Classify $X$ as type $R_i$ if

$$d_{EW}(X, R_i) = \min_j d_{EW}(X, R_j)$$

Synthetic Data: shear reference $R_i$ in a random directly by a factor $\lambda$, add Gaussian noise, estimate probability of correct classification.
Different test for atomic environments with large non-thermal strains: Classify $X$ as type $R_i$ if

$$d_{EW}(X, R_i) = \min_j d_{EW}(X, R_j)$$

Synthetic Data: shear reference $R_i$ in a random directly by a factor $\lambda$, add Gaussian noise, estimate probability of correct classification.
Different test for atomic environments with large non-thermal strains: Classify $X$ as type $R_i$ if

$$d_{EW}(X, R_i) = \min_j d_{EW}(X, R_j)$$

Synthetic Data: shear reference $R_i$ in a random directly by a factor $\lambda$, add Gaussian noise, estimate probability of correct classification.
Shear Performance Comparison: FCC

![Graph showing shear performance comparison for FCC with different methods: Euclidean Wasserstein, radius=2.118, Euclidean Wasserstein, radius=1.5, Polyhedral Template Matching, and Neighborhood Distance Analysis.]
We propose using the (local) Euclidean Wasserstein metric for local atomic environment classification.

The metric performs better than existing methods for distinguishing HCP and FCC environments at high temperatures.

Proposed application: studying the phase transition from graphite to diamond
Conclusion

- We propose using the (local) Euclidean Wasserstein metric for local atomic environment classification.
- The metric performs better than existing methods for distinguishing HCP and FCC environments at high temperatures.
- Proposed application: studying the phase transition from graphite to diamond
We propose using the (local) Euclidean Wasserstein metric for local atomic environment classification.

The metric performs better than existing methods for distinguishing HCP and FCC environments at high temperatures.

Proposed application: studying the phase transition from graphite to diamond