Hausdorff Dimension of Crystallographic Lattices

1 Introduction to Hausdorff Dimension

The Hausdorff dimension provides a rigorous mathematical framework for characterizing the geometric complexity of point sets in metric spaces, extending beyond the integer dimensions of classical Euclidean geometry. For a set S in $\mathbb{R}^n$, the Hausdorff dimension d_{H} quantifies how the "mass" (or number of points) scales with the radius of measurement. Unlike topological dimension, which is always an integer, Hausdorff dimension can take non-integer values, making it particularly useful for characterizing fractal-like distributions and irregular point configurations.

For crystallographic applications, the Hausdorff dimension characterizes how unit cell parameters populate the abstract lattice representation spaces (G6, S6, etc.). While individual crystal structures occupy discrete points, collections of related structures—such as those within a Bravais lattice type, space group, or structural family—form point clouds whose dimensional properties reflect the underlying constraints and symmetries of the crystallographic system.

2 Theoretical Foundation

The formal definition of Hausdorff dimension relies on δ-covers of a set S. For radius $\delta > 0$, a δ-cover is a collection of sets $\{U_i\}$ with $\mathrm{diam}(U_i) \leq \delta$ that covers S. The d-dimensional Hausdorff measure is defined as:

$$\mathcal{H}^d(S) = \lim_{\delta \to 0} \inf \left\{ \sum_i (\mathrm{diam}(U_i))^d : \{U_i\} \text{ is a } \delta\text{-cover of } S \right\} \tag{1}$$

The Hausdorff dimension is then:

$$d_{\mathrm{H}} = \inf\{d : \mathcal{H}^d(S) = 0\} = \sup\{d : \mathcal{H}^d(S) = \infty\} \tag{2}$$

For practical computation with finite point sets, we employ the box-counting dimension, which converges to the Hausdorff dimension for many well-behaved sets. The box-counting dimension exploits the power-law scaling relationship:

$$N(r) \propto r^{-d} \tag{3}$$

where $N(r)$ represents the number of points (or boxes) needed to cover the set at scale r.

3 Computational Method

Our algorithm estimates the local Hausdorff dimension through an adaptive spherical box-counting method with Monte Carlo sampling. The approach addresses several practical challenges inherent in analyzing finite crystallographic datasets: heterogeneous point distributions, varying local densities, and the need for statistically robust estimates.

3.1 Adaptive Radius Selection

The algorithm begins by determining an appropriate measurement scale. Given a point cloud of N structures in a representation space with estimated diameter D, we compute the mean nearest-neighbor spacing μ and point density $\rho = N/D$. An initial target radius r_{target} is selected to capture approximately 4096 points in one dimension:

$$r_{\text{target}} = \frac{4096}{\rho} \tag{4}$$

This radius is constrained to the range $[10\mu, D/1.1]$ to ensure statistically meaningful measurements while avoiding boundary effects.

To refine r_{target} for the actual data distribution, we iteratively shrink the radius by factors of 1.1 until identifying a sphere containing 256–266 points around a randomly selected probe structure. This adaptive procedure ensures sufficient local sampling regardless of the global distribution characteristics.

3.2 Monte Carlo Dimension Estimation

The dimension is estimated through repeated measurements at randomly selected probe points. For each trial, we:

1. Select a random structure p from the dataset as probe center

2. Count points within radius r: $N_{\text{large}} = |\{x : d(x,p) \leq r\}|$

3. Count points within radius $r/1.1$: $N_{\text{small}} = |\{x : d(x,p) \leq r/1.1\}|$

4. Compute local dimension estimate:

$$d_{\text{est}} = \frac{\ln(N_{\text{large}}/N_{\text{small}})}{\ln(1.1)} \tag{5}$$

The number of trials is set to $\sqrt{N}$ with a minimum of 10, providing a balance between computational efficiency and statistical reliability. Trials where $N_{\text{small}} \geq N_{\text{large}}$ are rejected as non-physical (indicating insufficient local point density).

3.3 Statistical Analysis and Convergence

The algorithm accumulates the mean dimension $\langle d \rangle$ and variance σ^2 across successful trials. Early termination occurs when at least half the trials have succeeded and the standard error satisfies:

$$\sigma^2 \leq \epsilon^2 \tag{6}$$

where ϵ is the requested precision. The final estimate is reported as $d_{\mathrm{H}} \pm \sigma$, with physically meaningless results ($d_{\mathrm{H}} + 3\sigma < 0$) rejected.

4 Implementation Considerations

The method requires minimum 32 structures and non-trivial diameter ($> \epsilon_{\mathrm{machine}}$) for meaningful estimation. For crystallographic datasets, the distance metric $d(x, y)$ is typically the NCDist (Niggli cell distance) or CS6Dist (Selling-reduced cell distance), ensuring the dimension measurement respects crystallographic equivalence and reduction conventions.

Cached results are returned when previous estimates meet the requested precision, avoiding redundant computation during iterative analyses. The algorithm's robustness derives from its adaptive scaling, Monte Carlo sampling strategy, and explicit validation of statistical convergence.

5 Applications to Crystallographic Analysis

The Hausdorff dimension provides quantitative insight into the geometric organization of crystal structure families. Low-dimensional manifolds ($d_{\mathrm{H}} \approx$ 1–2) within 6-dimensional lattice spaces indicate strong geometric constraints, such as those arising from coordination chemistry or packing efficiency. Higher dimensions suggest greater structural diversity or weaker systematic relationships. Comparison of Hausdorff dimensions across different structural classifications (space groups, composition families, structure types) reveals the relative geometric complexity and constraint hierarchy within crystallographic databases.