

Persistence Over Posets

$$\begin{array}{ccccc}
 & \begin{pmatrix} 0 \\ 1 \end{pmatrix} & & & \\
 \mathbb{k} & \xrightarrow{\quad} & \mathbb{k}^2 & \xrightarrow{\begin{pmatrix} 1 & 1 \end{pmatrix}} & \mathbb{k} \\
 \uparrow & \begin{pmatrix} 1 \\ 0 \end{pmatrix} \uparrow & & 1 \uparrow & \\
 0 & \xrightarrow{\quad} & \mathbb{k} & \xrightarrow{1} & \mathbb{k}
 \end{array}
 \quad \longrightarrow \quad
 \begin{array}{c}
 \text{Diagram 1: } \begin{array}{c} \text{Top row: } \bullet \rightarrow \bullet \rightarrow \bullet \\ \text{Bottom row: } \bullet \rightarrow \bullet \rightarrow \bullet \\ \text{Vertical arrows: } \bullet \rightarrow \bullet \end{array} \\
 \text{Diagram 2: } \begin{array}{c} \text{Top row: } \bullet \rightarrow \bullet \rightarrow \bullet \\ \text{Bottom row: } \bullet \rightarrow \bullet \rightarrow \bullet \\ \text{Vertical arrows: } \bullet \rightarrow \bullet \end{array} \\
 \text{Diagram 3: } \begin{array}{c} \text{Top row: } \bullet \rightarrow \bullet \rightarrow \bullet \\ \text{Bottom row: } \bullet \rightarrow \bullet \rightarrow \bullet \\ \text{Vertical arrows: } \bullet \rightarrow \bullet \end{array} \\
 \text{Diagram 4: } \begin{array}{c} \text{Top row: } \bullet \rightarrow \bullet \rightarrow \bullet \\ \text{Bottom row: } \bullet \rightarrow \bullet \rightarrow \bullet \\ \text{Vertical arrows: } \bullet \rightarrow \bullet \end{array}
 \end{array}$$

Woojin Kim and Facundo Mémoli

In topological data analysis (TDA), the shape of a dataset is often encoded into a system of vector spaces and linear maps over a partially ordered set (poset). We give an overview of how summaries of such systems can be constructed by using ideas from combinatorics.

Persistent Homology

Datasets are often given as *point clouds*: finite sets of points in Euclidean space. Examples include the three-dimensional coordinates of all atoms in a sample of a given material, a point cloud produced by a 3D scanner, and high-dimensional data such as a spreadsheet containing clinical features of a group of diabetes patients. The “shape” of a point cloud $X \subset \mathbb{R}^d$ may provide useful information about the underlying phenomena that generated the data. If X stands for the dataset of clinical features of diabetic patients mentioned above, and points in X appear to fall into a number of distinct clusters, then this may indicate the presence of different subtypes of diabetes amongst the patients.

In algebraic topology, the shape of a simplicial complex (or of a topological space) can be studied via *homology* [Mun84]. Homology provides a way to associate algebraic structures such as groups or vector spaces to a simplicial complex K in order to capture some aspect about the shape of K ; e.g., is K connected? does it have holes?

Woojin Kim is an assistant professor of mathematical sciences at KAIST in South Korea. His email address is woojin.kim.math@gmail.com.

Facundo Mémoli is a professor of mathematics at the Ohio State University. His email address is facundo.memoli@gmail.com.

Communicated by Notices Associate Editor Emilie Purvine.

For permission to reprint this article, please contact:

reprint-permission@ams.org.

DOI: <https://doi.org/10.1090/noti2761>

Given an integer $k \geq 0$, the k -th homology of the simplicial complex K with coefficients in a field \mathbb{k} , denoted $H_k(K; \mathbb{k})$, is a vector space over \mathbb{k} . Its dimension is called the i -th *Betti number* of K (with coefficients in \mathbb{k}) which is a count of the k -dimensional holes in K ; e.g., the 0th, 1st, and 2nd Betti numbers equal the numbers of *connected components*, *holes* bounded by a closed loop, and *cavities* bounded by a closed two-dimensional region in K , respectively.

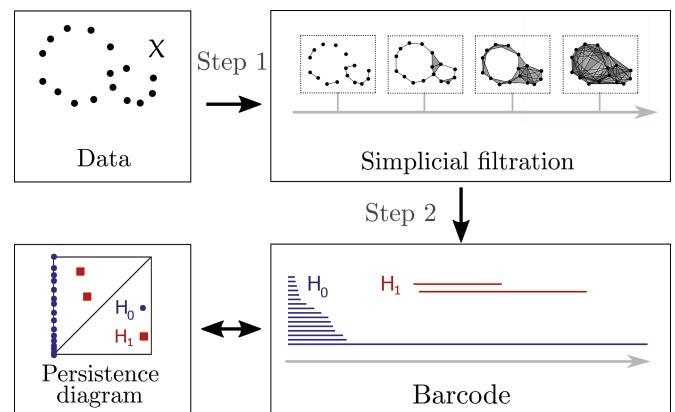


Figure 1. The pipeline of persistent homology. Multiscale clustering features and circular features of the input dataset X are encoded into H_0 and H_1 barcodes respectively. Intervals $[b, d]$ in the barcode are in bijection with points (b, d) in the persistence diagram.

Given a simplicial map $f : K \rightarrow L$ between simplicial complexes K and L , homology induces a linear map $H_k(f; \mathbb{k}) : H_k(K; \mathbb{k}) \rightarrow H_k(L; \mathbb{k})$. Thus, homology $H_k(\cdot; \mathbb{k})$ is a *functor* from the category of simplicial complexes and simplicial maps to the category of vector spaces and linear maps over \mathbb{k} .

Persistent homology. Within TDA, *persistent homology* refers to a method for associating multiresolution topological features to a given dataset. The usual pipeline is depicted in Figure 1.

In Step 1, one often uses *Vietoris–Rips complexes*. Given a dataset, modeled as a finite metric space X , e.g., a point cloud in \mathbb{R}^d , for a real number $r \geq 0$, the Vietoris–Rips complex $\mathcal{R}(X; r)$ is an abstract simplicial complex consisting of all the finite subsets $\sigma \subseteq X$ in which every pair of points is within distance r . The nested family $\{\mathcal{R}(X; r)\}_{r \geq 0}$ is called the *Vietoris–Rips filtration* of X . In Step 2, for each $k \geq 0$, we apply the homology functor $H_k(-; \mathbb{k})$ to this filtration and obtain a *persistence module* M over $\mathbb{R}_{\geq 0}$.

As we will see momentarily, persistence modules over $\mathbb{R}_{\geq 0}$ are a particular instance of the more general concept of persistence modules over posets.

Recall that a *poset* is a nonempty set P equipped with a relation \leq on P that is (i) Reflexive: $p \leq p$ for all $p \in P$. (ii) Transitive: For $p, p', p'' \in P$, if $p \leq p'$ and $p' \leq p''$, then $p \leq p''$. (iii) Anti-symmetric: If $p \leq p'$ and $p' \leq p$, then $p = p'$. Throughout this paper, all posets P will be assumed to be *connected*, i.e., for any $p, q \in P$, there is a sequence $p = p_1, \dots, p_n = q$ in P such that p_i and p_{i+1} are comparable for each $i = 1, \dots, n - 1$.

Posets are convenient gadgets for “indexing” simplicial filtrations and persistence modules.

Definition 1. Let P be a poset. A *persistence module* M over P (P -*module* in short) is a system of finite-dimensional vector spaces $M(p)$, $p \in P$, and linear maps $M(p, p') : M(p) \rightarrow M(p')$, $p \leq p' \in P$ such that for each $p \in P$, $M(p, p)$ is the identity on $M(p)$, and

$$\text{for } p \leq p' \leq p'', \quad M(p, p'') = M(p', p'') \circ M(p, p'). \quad (1)$$

The maps $M(p, p')$ are called *internal morphisms*. Since any poset can be regarded as a category, M is actually a functor $M : P \rightarrow \mathbf{vec}$ from P to the category \mathbf{vec} of finite-dimensional vector spaces and linear maps over the field \mathbb{k} .¹ Two P -modules M and N are *isomorphic*, denoted by $M \cong N$, if there are linear isomorphisms $T_p : M(p) \rightarrow N(p)$ for all $p \in P$, so that for $p \leq p' \in P$,

$$N(p, p') \circ T_p = T_{p'} \circ M(p, p').$$

When M is the result of applying the homology functor to the Vietoris–Rips filtration of a dataset $X \subset \mathbb{R}^d$, a great deal of information about the shape of X can be absorbed by the *persistence diagram* or into the *barcode* of M (cf. Figure 1). The definitions of these invariants as well as their relationship will be recalled in later sections.

¹In the literature, M is often referred to as a pointwise finite-dimensional persistence module.

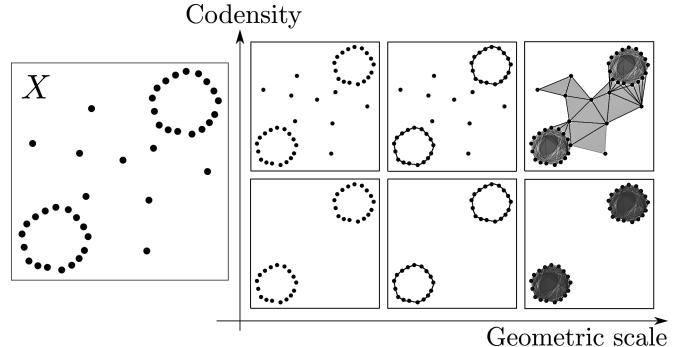


Figure 2. (Left) A dataset $X \subset \mathbb{R}^2$ consisting of two circular clusters together with some outliers (i.e., “noise”). Define the *codensity* function of X as $f_X(x) := \min_{x' \in X \setminus \{x\}} \|x - x'\|_2$ for $x \in X$. Notice that f_X attains small values only in ‘dense’ regions of X . (Right) For $\epsilon \in \mathbb{R}_{\geq 0}$, let X_ϵ be the subset of points $x \in X$ with $f_X(x) \leq \epsilon$. Notice that for $\epsilon > 0$ small, X_ϵ does not contain any of the outliers. The figure depicts the Vietoris–Rips complexes $\mathcal{R}(X_{\epsilon_j}; r_i)$ for some values $r_1 < r_2 < r_3$ and $\epsilon_1 < \epsilon_2$.

The need for a more general framework. Practical data analysis scenarios necessitate methods that can cope with more than one parameter. For instance, a dataset $X \subset \mathbb{R}^d$ might have nonuniform density (see Figure 2), possibly due to noise produced during the acquisition process or due to underlying scientific phenomena. In such scenarios, in addition to a geometric scale parameter, one may wish to incorporate a (co)density parameter and obtain an increasing family of simplicial complexes indexed by \mathbb{R}^2 with the product order \leq , i.e., $(a_1, a_2) \leq (b_1, b_2)$ iff $a_i \leq b_i$ for $i = 1, 2$. The result of applying the homology functor to such an \mathbb{R}^2 -indexed family, or more generally, to an analogously defined \mathbb{R}^n -indexed family, is called a *multiparameter persistence module* [CZ09].

There are scenarios that give rise to persistence modules indexed by posets other than other than \mathbb{R}^n . For example, the time-evolution of the positions of animals during collective motion can lead to considering zigzag posets

$$ZZ := \{1 \leftrightarrow 2 \leftrightarrow \dots \leftrightarrow m\},$$

where, for each i , $i \leftrightarrow i + 1$ stands for either $i < i + 1$ or $i > i + 1$ [CdS10, KM22].

Beyond these scenarios, and with a great deal of foresight, in [BdSS15], Bubenik et al. proposed to consider the phenomenon of persistence for parameters taken from general posets.

Remark 1. In this article, we restrict ourselves to persistence modules over finite connected posets P , which are general enough indexing sets for modeling the type of datasets that arise in practice.

For example, for an integer $m > 0$, the linearly ordered poset

$$L_m := \{1 < 2 < \dots < m\} \quad (2)$$

is often used to succinctly encode \mathbb{R} - or $\mathbb{R}_{\geq 0}$ -modules.

Example 1 (L_m -module). Given a finite metric space X , the Vietoris–Rips filtration of X consists of finitely many distinct simplicial complexes. Hence, for the persistence module $M := H_k(\mathcal{R}(X, -)) : \mathbb{R}_{\geq 0} \rightarrow \mathbf{vec}$, there exist $0 = r_1 < \dots < r_m$ in $\mathbb{R}_{\geq 0}$ so that $M(r, s)$ is a linear isomorphism whenever $r, s \in [r_i, r_{i+1})$ with $r \leq s$, for some $i = 1, \dots, m$ where $r_{m+1} := \infty$. In such a case, the L_m -module

$$H_k(\mathcal{R}(X; r_1)) \rightarrow \dots \rightarrow H_k(\mathcal{R}(X; r_m))$$

determines the isomorphism type of M .

Connections with quiver representations. A quiver is a finite directed graph. Given a quiver Q , the assignment of a finite-dimensional vector space to each vertex and a linear map to each arrow (between the participating vector spaces) is called a *representation* of Q ; see [DW05].

A finite poset P induces the quiver Q_P on the vertex set P with arrows $p \rightarrow p''$ for all pairs $p < p''$ such that there is no $p' \in P$ with $p < p' < p''$. Note that a P -module M canonically induces a representation of Q_P : To each vertex p of Q_P , the vector space $M(p)$ is assigned. To each arrow $p \rightarrow p''$ of Q_P , the linear map $M(p, p'')$ is assigned. Note that the resulting representation of Q_P satisfies the following commutativity condition: For every $p, p'' \in P$, if there are multiple directed paths from p to p'' in Q_P , then the compositions of the linear maps along each of those paths agrees with $M(p, p'')$. Conversely, a representation of Q_P satisfying the commutativity condition induces a P -module in the obvious way.

Example 2. The poset L_m from Equation (2) induces the quiver

$$Q_{L_m} : 1 \rightarrow 2 \rightarrow \dots \rightarrow m.$$

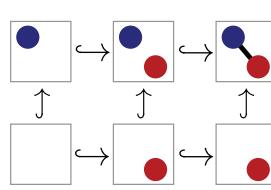
Example 3. For integers $m, n > 0$, consider the poset $L_m \times L_n$ equipped with the product order. Then, for example, the poset $(L_2 \times L_3)$ induces the quiver

$$Q_{L_2 \times L_3} : \begin{array}{ccccc} (2, 1) & \longrightarrow & (2, 2) & \longrightarrow & (2, 3) \\ \uparrow & & \uparrow & & \uparrow \\ (1, 1) & \longrightarrow & (1, 2) & \longrightarrow & (1, 3). \end{array}$$

The poset $L_m \times L_n$ is often used to encode \mathbb{R}^2 -modules.

Example 4. The following commutative diagram defines an $(L_2 \times L_3)$ -module which can be obtained by applying the 0-th homology functor $H_0(-, \mathbb{k})$ to the $(L_2 \times L_3)$ -indexed simplicial filtration depicted next:

$$\begin{array}{ccccc} \mathbb{k} & \xrightarrow{\begin{pmatrix} 0 \\ 1 \end{pmatrix}} & \mathbb{k}^2 & \xrightarrow{\begin{pmatrix} 1 & 1 \end{pmatrix}} & \mathbb{k} \\ \uparrow & & \uparrow & & \uparrow \\ 0 & \longrightarrow & \mathbb{k} & \xrightarrow{1} & \mathbb{k} \end{array}$$



Rank Invariant and Persistence Diagrams

We first recall the classical notions of rank invariant and persistence diagrams of L_m -modules [CSEH07, CZ09], and then we describe a natural way to extend those notions to the setting of P -modules.

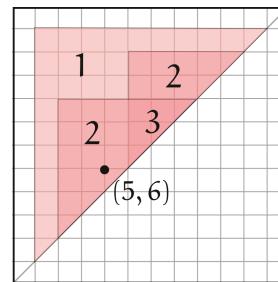
Rank invariant. For any integer $m > 0$ and any $b \leq d \in L_m$, we call $[b, d] := \{b, \dots, d\}$ an *interval* in L_m . Let $\text{Int}(L_m)$ be the set of all intervals in L_m . The *rank invariant* of a given persistence module $M : L_m \rightarrow \mathbf{vec}$ is defined to be the function

$$\begin{aligned} \text{rk}_M : \quad \text{Int}(L_m) &\rightarrow \quad \mathbb{Z}_{\geq 0} \\ [b, d] &\mapsto \quad \text{rank}(M(b, d)). \end{aligned} \quad (3)$$

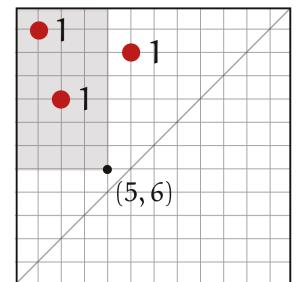
It is important to note that (i) the rank invariant is preserved under isomorphism and that (ii) it encodes the dimensions of all vector spaces $M(b)$ for $b \in L_m$ since we have $\text{rk}_M([b, d]) = \dim(M(b))$ whenever $b = d$. Note also that (iii) rk_M is *monotone*, i.e.,

$$\text{rk}_M([b', d']) \leq \text{rk}_M([b, d]) \quad (4)$$

whenever $[b, d] \subseteq [b', d']$; This follows from the fact that the map $M(b', d')$ factors through the map $M(b, d)$. By convention, we set $0 = \text{rk}_M([0, d]) = \text{rk}_M([b, m+1])$ for every $b, d \in L_m$. Next, we utilize the rank invariant to compute, for each $[b, d] \in \text{Int}(L_m)$, a count of the “persistent features” that start at b and end at d , leading to the notion of persistence diagram of M .



Rank invariant



Persistence diagram

Figure 3. The rank invariant and the persistence diagram of a given $M : L_{13} \rightarrow \mathbf{vec}$. At each point (b, d) with $b \leq d$ in the $L_{13} \times L_{13}$ grid, nonzero $\text{rk}_M([b, d])$ and $\text{dgm}_M([b, d])$ are recorded (e.g., $\text{rk}_M([5, 6]) = 2$ and $\text{dgm}_M([5, 6]) = 0$).²

Persistence diagrams. Fix an integer $m > 0$. Let $p' \in L_m$ and a vector $v \in M(p')$. We say that v is *born* at the point $b(v) \in L_m$ where

$$b(v) := \min\{p \in L_m : v \in \text{im}(M(p, p'))\}. \quad (5)$$

We say that v lives until the point $d(v) \in L_m$ (or *dies* at $d(v) + 1$) where

$$d(v) := \max\{p'' \in L_m : v \notin \ker(M(p', p''))\}. \quad (6)$$

²If M encodes an \mathbb{R} or $\mathbb{R}_{\geq 0}$ -module (as in the scenario of Example 1), the visualization of dgm_M may require a scale readjustment [CSEH07, p.106].

The *lifespan* of v is the interval $[b(v), d(v)]$. The persistence diagram of a given persistence module $M : \mathbf{L}_m \rightarrow \mathbf{vec}$ is then defined to be the function

$$\mathrm{dgm}_M : \mathrm{Int}(\mathbf{L}_m) \rightarrow \mathbb{Z}_{\geq 0}$$

sending each $[b, d] \in \mathrm{Int}(\mathbf{L}_m)$ to the maximal number of linearly independent vectors in $\mathrm{im}(M(b, d))$ whose lifespans are exactly $[b, d]$.

From rk_M to dgm_M . It turns out that dgm_M can be computed in terms of rk_M . Indeed, let $k := \mathrm{rk}_M([b, d])$. This implies that there exist k linearly independent vectors v_1, v_2, \dots, v_k in $\mathrm{im}(M(b, d)) \subseteq M(d)$ that are born at b or before b , and live until d or later, i.e., $[b(v_i), d(v_i)] \supseteq [b, d]$. Hence,

$$s(b, d) := \mathrm{rk}_M([b, d]) - \mathrm{rk}_M([b - 1, d])$$

equals the maximal number of independent vectors in $\mathrm{im}(M(b, d))$ that were born at *precisely* b and live until d or later. Similarly,

$$s(b, d + 1) := \mathrm{rk}_M([b, d + 1]) - \mathrm{rk}_M([b - 1, d + 1])$$

equals the maximal number of independent vectors in $\mathrm{im}(M(b, d + 1))$ that were born at precisely b and live until $d + 1$ or later. Hence, the difference $s(b, d) - s(b, d + 1)$ equals the maximal number of independent vectors in $\mathrm{im}(M(b, d))$ whose lifespans are exactly $[b, d]$ and thus we arrive at the following formula:

$$\begin{aligned} \mathrm{dgm}_M([b, d]) &= \mathrm{rk}_M([b, d]) - \mathrm{rk}_M([b - 1, d]) \\ &\quad - \mathrm{rk}_M([b, d + 1]) + \mathrm{rk}_M([b - 1, d + 1]) \geq 0. \end{aligned} \quad (7)$$

In practice, persistence diagrams are represented as points (with multiplicity) in the two-dimensional grid: only those intervals $[b, d]$ for which $\mathrm{dgm}_M([b, d]) > 0$ are recorded; see Figure 3.

The earliest appearance of formula (7) in the TDA community that is known to the authors is [LF97]. This expression appears prominently in the work of Cohen-Steiner et al. on the stability of persistence diagrams [CSEH07].

From dgm_M to rk_M . Let us consider $\mathrm{Int}(\mathbf{L}_m)$ as a poset ordered by containment \supseteq . The poset $\mathrm{Int}(\mathbf{L}_m)$ consists of $m(m + 1)/2$ elements. By the *order-extension principle*, we can index the intervals in $\mathrm{Int}(\mathbf{L}_m)$ as $I_1, I_2, \dots, I_{m(m+1)/2}$ so that $I_i \supseteq I_j$ implies $j \geq i$. Below we also use the convenient notation $I_i = [b_i, d_i]$.

Since $\mathrm{Int}(\mathbf{L}_m)$ consists of $m(m + 1)/2$ elements, we identify rk_M with a vector in $\mathbb{R}^{m(m+1)/2}$ whose i -th entry is $\mathrm{rk}_M(I_i)$. Similarly, dgm_M is identified with a vector of the same dimension. Consider the square matrix μ of length $m(m + 1)/2$ whose (i, j) -entry is

$$\mu_{ij} := \begin{cases} 1 & \text{if } I_i = I_j \text{ or } I_i = [b_j - 1, d_j + 1], \\ -1 & \text{if } I_i = [b_j - 1, d_j] \text{ or } I_i = [b_j, d_j + 1], \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

The matrix μ is upper-triangular and all of its diagonal entries are 1. Therefore, μ is invertible. Note that Equation (7) amounts to

$$\mathrm{rk}_M \cdot \mu = \mathrm{dgm}_M, \quad (9)$$

which implies that dgm_M and rk_M determine each other.

Equation (9) permits computing rk_M in terms of dgm_M as follows. First, one verifies that the inverse of μ has entries

$$(\mu^{-1})_{ij} = \begin{cases} 1 & \text{if } I_i \supseteq I_j \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, the equality $\mathrm{rk}_M = \mathrm{dgm}_M \cdot \mu^{-1}$ implies that

$$\mathrm{rk}_M(I_j) = \sum_{i: I_i \supseteq I_j} \mathrm{dgm}_M(I_i) \quad \forall 1 \leq j \leq m(m + 1)/2. \quad (10)$$

Example 5 (An application of Equation (10)). The equality $\mathrm{rk}_M([5, 6]) = 2$ in Figure 3 can be derived from the fact that there are exactly two points (b, d) in the upper-left quadrant with corner point $(5, 6)$ for which $\mathrm{dgm}_M([b, d]) = 1$.

Example 6 (Persistence diagram of an \mathbf{L}_2 -module). Let M be an \mathbf{L}_2 -module:

$$M : \quad M(1) \xrightarrow{M(1,2)} M(2).$$

Let the rank invariant of M be given by

$$\mathrm{rk}_M : \quad [1, 2] \mapsto k, \quad [1, 1] \mapsto d_1, \quad [2, 2] \mapsto d_2, \quad (11)$$

for integers $0 \leq k \leq d_1, d_2$. Then, the persistence diagram of M is given by

$$\mathrm{dgm}_M : \quad [1, 2] \mapsto k, \quad [1, 1] \mapsto d_1 - k, \quad [2, 2] \mapsto d_2 - k, \quad (12)$$

where $d_1 - k$ and $d_2 - k$ are the dimensions of the kernel and cokernel of the map $M(1, 2)$. Under the identification

$$\mathrm{rk}_M \leftrightarrow (k, d_1, d_2), \quad \mathrm{dgm}_M \leftrightarrow (k, d_1 - k, d_2 - k),$$

in this case, Equation (9) reads:

$$\mathrm{rk}_M \cdot \begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \mathrm{dgm}_M.$$

Remark 2. From Equation (11), we infer that there are bases of the vector spaces $M(1)$ and $M(2)$ in which the map $M(1, 2)$ is given by the $(d_2 \times d_1)$ -block matrix

$$\left[\begin{array}{c|c} I_k & 0 \\ \hline 0 & 0 \end{array} \right],$$

where I_k is the k by k identity matrix. In other words, there are linear isomorphisms h_1 and h_2 such that the diagram

below commutes:

$$\begin{array}{ccc}
 M(1) & \xrightarrow{M(1,2)} & M(2) \\
 \cong \downarrow h_1 & \left[\begin{array}{c|c} I_k & 0 \\ \hline 0 & 0 \end{array} \right] & \cong \downarrow h_2 \\
 \mathbb{k}^k \oplus \mathbb{k}^{d_1-k} & \xrightarrow{\quad} & \mathbb{k}^k \oplus \mathbb{k}^{d_2-k}
 \end{array} \quad (13)$$

Therefore, we conclude that rk_M (and therefore dgm_M) determines the isomorphism type of M . We will shortly show that this is also the case for an arbitrary L_m -module M with $m > 2$.

Indecomposables and Barcodes

There is another interpretation of the persistence diagram which is tied to the notion of indecomposable decompositions of quiver representations.

Indecomposable decompositions. Given any two P -modules M and N , the *direct sum* $M \oplus N$ is the P -module defined via pointwise direct sum: $(M \oplus N)(p) := M(p) \oplus N(p)$ for $p \in P$ and $(M \oplus N)(p, p') := M(p, p') \oplus N(p, p')$ for $p \leq p'$ in P , i.e.,

$$(M \oplus N)(p, p') := \begin{pmatrix} M(p, p') & 0 \\ 0 & N(p, p') \end{pmatrix}.$$

A nonzero P -module M is *indecomposable* if, whenever $M = M_1 \oplus M_2$ for some P -modules M_1 and M_2 , then, either $M_1 = 0$ or $M_2 = 0$. We will refer to such modules as P -*indecomposables*. Due to the Krull–Schmidt–Remak–Azumaya principle, any P -module can be decomposed as a direct sum of P -indecomposables:

Theorem 1. *Any P -module M admits a decomposition*

$$M \cong \bigoplus_{\alpha \in A} M_\alpha, \quad (14)$$

where each M_α is P -indecomposable. Such a decomposition is unique up to isomorphism and to reordering of the summands.

Theorem 1 indicates that understanding the structure of P -modules reduces to the problem of elucidating the structure of the P -indecomposables.

Example 7. Diagram (13) shows that the L_2 -module M therein decomposes into a direct sum of k , $d_1 - k$, $d_2 - k$ copies of the L_2 -modules

$$\mathbb{k} \xrightarrow{1} \mathbb{k}, \quad \mathbb{k} \xrightarrow{0} 0, \quad 0 \xrightarrow{0} \mathbb{k},$$

respectively. It is not hard to verify that all these L_2 -modules are indecomposable and thus Diagram (13) is an example of a decomposition into a direct sum of indecomposable modules. Notice that this decomposition is reflected by the specification of dgm_M in Equation (12). Since M was an arbitrary L_2 -module, this decomposition further implies that the three L_2 -modules shown above constitute an exhaustive list of all the L_2 -indecomposables.

Barcode of an L_m -module. A classical theorem by Pierre Gabriel (see [DW05]) establishes a far reaching generalization of the previous example and, in particular, implies that the L_m -indecomposables are *exactly* those $V_{[b,d]} : L_m \rightarrow \mathbf{vec}$ that look like:

$$0 \xrightarrow{0} 0 \cdots 0 \xrightarrow{0} \mathbb{k} \xrightarrow{1} \mathbb{k} \cdots \mathbb{k} \xrightarrow{1} \mathbb{k} \xrightarrow{0} 0 \cdots 0 \xrightarrow{0} 0, \quad (15)$$

where (from left to right) the first occurrence of \mathbb{k} is at some index $b \in L_m$ (for “birth”) and the last occurrence is at an index $d \in L_m$ (for “death”).³ More precisely, given any $[b, d] \in \text{Int}(L_m)$, $V_{[b,d]}$ is the persistence module over L_m where:

- (i) $V_{[b,d]}(i) = \mathbb{k}$ for $i \in [b, d]$ and $V_{[b,d]}(i) = 0$ otherwise, and
- (ii) all internal morphisms between adjacent copies of \mathbb{k} are 1, and all other morphisms are (necessarily) 0.

Any such $V_{[b,d]}$ is called an *interval (persistence) module*. The considerations above imply that each M_α appearing in the decomposition in Equation (14) of M , assuming $P = L_m$, is isomorphic to $V_{[b_\alpha, d_\alpha]}$ for some $[b_\alpha, d_\alpha] \in \text{Int}(L_m)$. Therefore, to the L_m -module M we can associate the multiset $\text{barc}(M)$, the *barcode* of M , consisting of all the intervals $[b_\alpha, d_\alpha]$, $\alpha \in A$ (counted with multiplicity) appearing in the decomposition of M given above. Furthermore, these considerations imply that $\text{barc}(M)$ is a *complete invariant* of M , i.e., $\text{barc}(M)$ determines the isomorphism type of M .

Persistence diagrams and barcodes determine each other. It holds that $\text{dgm}_M([b, d])$ equals the multiplicity of the interval $[b, d]$ in $\text{barc}(M)$. This claim follows from Theorem 4 in the present article, a result which is applicable in the context of general posets.⁴ Since $\text{barc}(M)$ is a complete invariant of $M : L_m \rightarrow \mathbf{vec}$, so is dgm_M . This establishes the previous claim that both dgm_M and rk_M determine the isomorphism type of a given $M : L_m \rightarrow \mathbf{vec}$.

Remark 3. Persistence diagrams are known to be *stable* (i.e., Lipschitz continuous under suitable metrics) [CSEH07].

Barcode of a P -module. Given any P -module M , by Theorem 1, one could, in principle, consider the multiset $\{M_\alpha\}_{\alpha \in A}$ of indecomposable summands as a complete invariant of M . However, other than for a handful of exceptional posets,⁵ and even in simple cases such as the one mentioned in Example 8 below, the collection of all P -indecomposables can be tremendously complex. One manifestation of this complexity is the possibility that

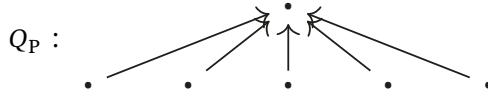
³By Equations (5) and (6), for any $p \in [b, d]$ and any nonzero $v \in V_{[b,d]}(p)$, we have $b(v) = b$ and $d(v) = d$.

⁴Interestingly, in the case of L_m , this relationship follows from work by Abeasis et al. dating back to 1981; see [ADFK81, p. 405]. The authors thank Ezra Miller for pointing this out.

⁵For example, by Gabriel’s theorem, the indecomposable modules over a zigzag poset are exactly the interval modules (defined below Definition 2) on that poset.

there may exist infinitely many isomorphism types of P -indecomposables.

Example 8. Consider the 6-point poset P inducing the quiver



An infinite two-parameter family of P -indecomposables are presented in [DW05, Example 8].

For the reasons above, and due to the analogy with the case when $P = L_m$, much of the research in the TDA community has concentrated on understanding (i.e., testing, computing, etc) decomposability of a given P -module as a direct sum of indecomposables with “simple” structure. One such notion of simple indecomposables is obtained through a generalization of the interval modules over L_m described in Equation (15). This leads to the notion of interval modules over an arbitrary poset P : these are P -indecomposable modules having dimension exactly one on certain nice subsets of P and zero elsewhere, and such that all internal morphisms between nontrivial spaces are 1.

The nice subsets mentioned above are called *intervals* of P and are defined as follows.

Definition 2. An *interval* I of a given poset P is any subset $I \subseteq P$ such that:

- (i) I is nonempty.
- (ii) If $p, p' \in I$ and $p'' \in P$ such that $p \leq p'' \leq p'$, then $p'' \in I$.
- (iii) I is *connected*, i.e., for any $p, p' \in I$, there is a sequence $p = p_0, p_1, \dots, p_\ell = p'$ of elements of I with either $p_i \leq p_{i+1}$ or $p_{i+1} \leq p_i$ for each $i \in [0, \ell - 1]$.

We will henceforth use the notation $\text{Int}(P)$ to denote the collection of all intervals of P .⁶ Note that when $P = L_m$, $\text{Int}(P)$ reduces to the definition given in the previous section.

For $I \in \text{Int}(P)$, the *interval module* $V_I : P \rightarrow \mathbf{vec}$ induced by I is defined via the conditions

$$V_I(p) := \begin{cases} \mathbb{k} & \text{if } p \in I \\ 0 & \text{otherwise,} \end{cases}$$

$$\text{and } V_I(p, p') := \begin{cases} 1 & \text{if } p, p' \in I \text{ and } p \leq p' \\ 0 & \text{otherwise.} \end{cases}$$

In general, it is important to note that if $I \subseteq P$ did not satisfy (ii), then V_I would not be well-defined (as it would not satisfy Equation (1)). If I did not satisfy (iii), then V_I would fail to be indecomposable.

⁶We warn the reader the definition of intervals that we are using—the most common in TDA—differs from the one that is traditional in order theory.

The following are simple but important facts: (a) No matter what P is, every interval module $V_I : P \rightarrow \mathbf{vec}$ is indecomposable. (b) There are posets P for which there exist P -indecomposables that are not interval modules. For example, let $P := \{a, b, c, d\}$ be equipped with the partial order $b \leq a, c \leq a$ and $d \leq a$ and consider the P -module F given below:

$$\begin{array}{ccccc} & & F(a) := \mathbb{k}^2 & & \\ \begin{pmatrix} 1 \\ 0 \end{pmatrix} & \nearrow & \begin{pmatrix} 0 \\ 1 \end{pmatrix} & \uparrow & \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ F(b) := \mathbb{k} & & F(c) := \mathbb{k} & & F(d) := \mathbb{k} \end{array} \quad (16)$$

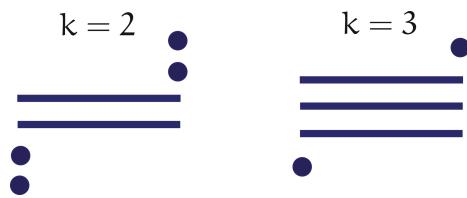
It is clear that F is neither an interval module nor is it isomorphic to a direct sum of interval modules. That F is indecomposable is also relatively easy to verify.

A decomposition of a given P -module into a direct sum of interval modules gives rise to its *barcode*.

Definition 3. A P -module M is *interval decomposable* if there exists a multiset $\text{barc}(M)$ of intervals of P (called the *barcode* of M) such that

$$M \cong \bigoplus_{I \in \text{barc}(M)} V_I.$$

Example 9 (Barcode of an L_2 -module). Recall from Example 7 that Diagram (13) describes the indecomposable decomposition of an arbitrary L_2 -module M . The figure below shows a visualization of the barcode of M for two different values of k when M is such that $d_1 = d_2 = 4$:



Even if it may be that a given P -module is not interval decomposable, it might be useful to understand whether at least *some* of its indecomposables are isomorphic to interval modules. This leads to defining the *multiplicity function*.

Definition 4 (Multiplicity of intervals). Given any P -module M and any interval I of P , let $\text{mult}(I, M)$ denote the number of P -indecomposable direct summands of M that are isomorphic to the interval module V_I . Furthermore, let

$$\text{mlt}_M : \text{Int}(P) \rightarrow \mathbb{Z}_{\geq 0}$$

denote the *multiplicity function* defined by $I \mapsto \text{mult}(I, M)$. It is clear that, whenever M is interval decomposable, mlt_M determines and is determined by $\text{barc}(M)$.

From a practical point of view, one aims to extract as much persistence-like information as possible from a given P -module M , regardless of whether M is interval decomposable or not, while bypassing the inherent difficulties associated to dealing with the “wild west” of indecomposables. This has led toward exploring the notion of *generalized persistence diagrams* which is enabled by *Möbius inversion* and by a suitable generalization of the notion of rank invariant.

One key idea arose in 2016 when Patel noticed that the process being implemented in Equation (7) is the Möbius inversion (over the poset $(\text{Int}(L_m), \supseteq)$) of the rank invariant [Pat18]. This crucial observation has led to very rich developments which, by injecting ideas from combinatorics, surmount the difficulties inherent to dealing with indecomposables.

Generalized Rank Invariant

Given a P -module M , the map

$$p \leq p' \mapsto \text{rank}(M(p, p')) \quad (17)$$

for all $p \leq p'$ in P is a direct generalization of the rank invariant of L_m -modules given in Equation (3).⁷ However, beyond the case $P = L_m$, this “standard” rank invariant is, in general, a weaker invariant than the barcode of interval decomposable P -modules; see e.g., [KM21a, Appendix C]. Motivated by this, we consider a generalized version of the rank invariant, which exhibits stronger discriminating power.

We start by generalizing the notion of rank of a linear map to the context of P -modules.

Rank of a P -module. Given a P -module M with $m := |P|$, an m -tuple $\mathbf{v} = (v_p) \in \bigoplus_{p \in P} M(p)$ is called a *persistent vector* if all the v_p are *compatible* in M , i.e., $M(p, p')(v_p) = v_{p'}$ for all $p \leq p'$ in P . The set L_M of persistent vectors is a linear subspace of $\bigoplus_{p \in P} M(p)$.

We call $\mathbf{v} \in L_M$ *fully supported* if $v_p \neq 0$ for all $p \in P$. Toward defining the rank of M , we identify $\mathbf{v} \in L_M$ with $0 \in L_M$ whenever \mathbf{v} is *not* fully supported, i.e., if there exists $p \in P$ such that $v_p = 0$. In other words, we consider the quotient space L_M/W_M where W_M is the linear span of all non-fully-supported vectors in L_M . We then define the rank of M as

$$\text{rank}(M) := \dim(L_M/W_M). \quad (18)$$

Example 10 (The case when $P = L_2$). Given any L_2 -module M , we have that L_M and W_M are isomorphic to $M(1)$ and $\ker(M(1, 2))$, respectively. Therefore, $\text{rank}(M)$ reduces to the rank of the linear map $M(1, 2)$.

Here is an alternative view on the rank of M . Let us call any two persistent vectors \mathbf{v} and \mathbf{w} *intersecting* if

⁷This notion was introduced in [CZ09].

$v_p = w_p$ for some $p \in P$. This property defines a reflexive and symmetric, but not necessarily transitive relation on L_M . Let \sim be the transitive closure of the resulting relation. We observe that L_M/W_M is the quotient L_M/\sim . Indeed:

$$\mathbf{v} \sim \mathbf{w}$$

- ↔ there exists a sequence $\mathbf{v} = \mathbf{v}^1, \mathbf{v}^2, \dots, \mathbf{v}^n = \mathbf{w}$ in L_M such that \mathbf{v}^i and \mathbf{v}^{i+1} are intersecting for every i
- ↔ there exists a sequence $\mathbf{v} = \mathbf{v}^1, \mathbf{v}^2, \dots, \mathbf{v}^n = \mathbf{w}$ in L_M such that $\mathbf{v}^i - \mathbf{v}^{i+1}$ is non-fully-supported for every i
- ↔ there exists a sequence $\mathbf{v} = \mathbf{v}^1, \mathbf{v}^2, \dots, \mathbf{v}^n = \mathbf{w}$ in L_M such that $\mathbf{v}^i - \mathbf{v}^{i+1} \in W_M$ for every i
- ↔ $\mathbf{v} - \mathbf{w} \in W_M$.

We call $\mathbf{v} \in L_M$ *full* if, whenever \mathbf{v} is written as a sum of linearly independent vectors $\mathbf{w}^1, \dots, \mathbf{w}^n \in L_M$, then at least one of the \mathbf{w}^j is fully supported.

From the observation above, we have:

Theorem 2. $\text{rank}(M)$ is the maximal number of linearly independent, full, nonintersecting persistent vectors of M .

Note that if \mathbf{v} is full, then \mathbf{v} is fully supported. However, the converse does not hold in general.

Example 11. When $P = L_2$, every fully supported $\mathbf{v} \in L$ is full.

Example 12. Consider the diagram $\mathbb{k} \xleftarrow{\pi_1} \mathbb{k}^2 \xrightarrow{\pi_2} \mathbb{k}$ over $P = \{a > b < c\}$, where π_i is the projection to the i -th coordinate. The fully supported persistent vector $1 \leftrightarrow (1, 1) \leftrightarrow 1$ is *not* full since it is the sum of non-fully-supported persistent vectors

$$1 \leftrightarrow (1, 0) \leftrightarrow 0 \text{ and } 0 \leftrightarrow (0, 1) \leftrightarrow 1.$$

The space L_M/W_M from Equation (18) is related to fundamental notions in category theory [ML98]. The space L_M of persistent vectors coincides with the *limit* of M , denoted by $\varprojlim M$. The quotient space $C_M := (\bigoplus_{p \in P} M(p)) / \approx$ coincides with the *colimit* $\varinjlim M$ of M , and is obtained by identifying $v_p \in M(p)$ with $v_{p'} \in M(p')$ through the transitive closure \approx of the relation R such that $(v_p, v_{p'}) \in R$ whenever $M(p, p')(v_p) = v_{p'}$ for $p \leq p' \in P$.

There is a canonical map ψ_M from the limit L_M to the colimit C_M . Indeed, note that, since P is connected, for any $\mathbf{v} = (v_p) \in L_M$ and for any $p, p' \in P$, the two vectors $v_p \in M(p)$ and $v_{p'} \in M(p')$ are identified in C_M . Therefore, we obtain the well-defined canonical linear map⁸ $\psi_M : L_M \rightarrow C_M$ given by $\mathbf{v} \mapsto [v_p]$ for an arbitrary $p \in P$. Note that the vector space L_M/W_M is isomorphic to the image of ψ_M .

⁸The idea of studying the map from the limit to the colimit of a given diagram of vector spaces stems from work by Amit Patel and Robert MacPherson circa 2012. We thank Prof. Harm Derksen for pointing out to us recently that this type of map was used in the study of quiver representations in [Kin08].

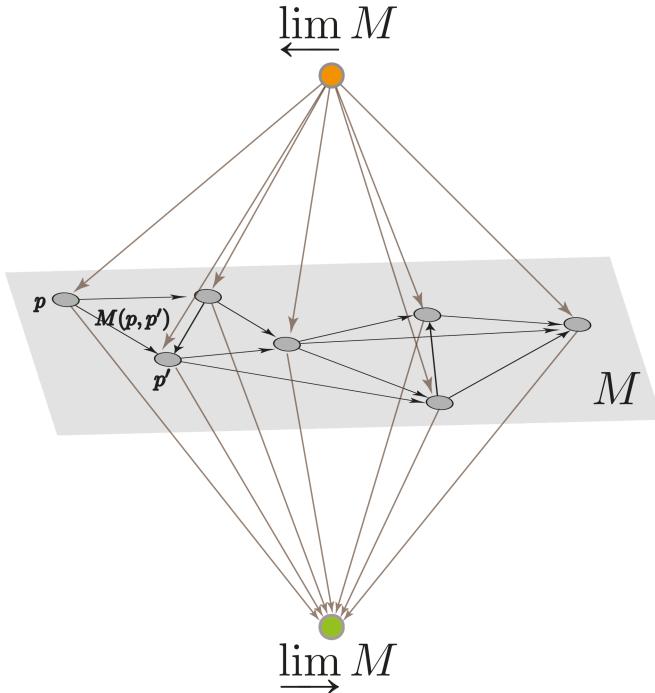


Figure 4. The notions of limit and colimit of a diagram of vector spaces are used in a fundamental way to generalize the notion of rank invariant and persistence diagram.

Proposition 1 ([KM21a, Section 3]). *The rank of the P-module $M : P \rightarrow \mathbf{vec}$ agrees with the rank of the canonical limit-to-colimit map $\psi_M : \varprojlim M \rightarrow \varinjlim M$. See Figure 4.*

Remark 4 (Additivity of the rank). If $M \cong \bigoplus_{\alpha \in A} M_\alpha$ for some indexing set A , then

$$\text{rank}(M) = \sum_{\alpha \in A} \text{rank}(M_\alpha).$$

To prove this, we proceed as follows. Since P is finite and $\dim(M(p))$ is finite for each $p \in P$, the direct sum commutes with limits as well as with colimits:

$$\varprojlim \left(\bigoplus_{\alpha \in A} M_\alpha \right) \cong \bigoplus_{\alpha \in A} \left(\varprojlim M_\alpha \right) \quad (19)$$

and

$$\varinjlim \left(\bigoplus_{\alpha \in A} M_\alpha \right) \cong \bigoplus_{\alpha \in A} \left(\varinjlim M_\alpha \right), \quad (20)$$

which leads to the desired equality.⁹ The isomorphisms in Equations (19) and (20) can also be easily verified by invoking the constructions of limits and colimits given above.

⁹The claimed isomorphism in Equation (19) follows from the following argument. A priori we have $\varprojlim \left(\prod_{\alpha \in A} M_\alpha \right) \cong \prod_{\alpha \in A} \left(\varprojlim M_\alpha \right)$, where \prod denotes a direct product [ML98, Chapter III]. But, one can check that \prod agrees with \bigoplus on both sides since P is finite and $\dim(M(p))$ is finite for each $p \in P$.

Generalized rank invariant. We are now ready to define the generalized rank invariant.

Definition 5. The *generalized rank invariant* of a P-module M is the function

$$\begin{aligned} \text{rk}_M : \quad \text{Int}(P) &\rightarrow \quad \mathbb{Z}_{\geq 0} \\ I &\mapsto \quad \text{rank}(M|_I), \end{aligned}$$

where $M|_I$ is the restriction of M to the interval I .

That, as defined above, rk_M indeed generalizes Equation (3) is a consequence of the fact that when $P = L_m$, for any interval $I = [b, d] \in \text{Int}(L_m)$, we have $\varprojlim M|_I \cong M(b)$, $\varinjlim M|_I \cong M(d)$, and $\psi_{M|_I} \cong M(b, d)$.

Remark 5 (Monotonicity of rk_M). Let $I, J \in \text{Int}(P)$ with $J \supseteq I$. Then

$$\text{rk}_M(J) \leq \text{rk}_M(I),$$

which is analogous to Equation (4). This is so because the canonical limit-to-colimit map $\varprojlim M|_I \rightarrow \varinjlim M|_I$ for the interval I is a factor of the canonical limit-to-colimit map $\varprojlim M|_J \rightarrow \varinjlim M|_J$ for the larger interval J .

Remark 6. Let $J \in \text{Int}(P)$ and let $V_J : P \rightarrow \mathbf{vec}$ be the interval module induced by J . By Theorem 2, for $I \in \text{Int}(P)$, $\text{rk}_{V_J}(I) = \text{rank}(V_J|_I)$ equals 1 if $J \supseteq I$, and 0 if $J \not\supseteq I$.

When $M : P \rightarrow \mathbf{vec}$ is interval decomposable, $\text{rk}_M(I)$ equals the multiplicity of those intervals J in $\text{barc}(M)$ that contain I .

Proposition 2. Let $M : P \rightarrow \mathbf{vec}$ be interval decomposable. Then, for any $I \in \text{Int}(P)$,

$$\text{rk}_M(I) = \sum_{\substack{J \supseteq I \\ J \in \text{Int}(P)}} \text{mlt}_M(J).$$

Proof. Let A be a finite indexing set and let $I_\alpha \in \text{Int}(P)$, $\alpha \in A$, be intervals such that $M \cong \bigoplus_{\alpha \in A} V_{I_\alpha}$. Then, by Remark 4,

$$\text{rk}_M(I) = \text{rank}(M|_I) = \sum_{\alpha \in A} \text{rank}(V_{I_\alpha}|_I),$$

which by Remark 6 above equals the claimed quantity. \square

Now that the rank invariant has been generalized from the case of L_m -modules to that of P-modules, the mechanism of *Möbius inversion* will provide the sought-after generalization of the notion of persistence diagram for the case of L_m -modules (cf. Equation (7)) to that of P-modules.

Möbius Inversion

The summation of a number-theoretic function $f(n)$ over the divisors of n and its *inversion* play an important role in elementary number theory (the meaning of inversion will be made clear in Example 13). The classical Möbius inversion formula, introduced by August Ferdinand Möbius

in 1832, is a useful tool for performing such inversions. In the 1960s, Rota [Rot64] noticed the vast combinatorial implications of the Möbius inversion formula and established connections to coloring problems, flows in networks, and to the inclusion-exclusion principle.

Two basic examples of Möbius inversion follow.

Example 13 (Sum over L_m). Fix an integer $m > 0$. Consider any two functions $f, g : L_m \rightarrow \mathbb{R}$ such that $g(q) = \sum_{q'=1}^q f(q')$ for all $q \in L_m$. One can easily *invert* the sum, i.e., solve for f , as

$$f(1) = g(1) \text{ and } f(q) = g(q) - g(q-1) \text{ for } 2 \leq q \in L_m. \quad (21)$$

Given a set D , let 2^D be the set of subsets of D ordered by containment \supseteq .

Example 14 (Principle of inclusion-exclusion). Let S be a finite nonempty set of objects. Let $D = \{p_1, \dots, p_n\}$ be a set of properties such that, for every $i = 1, \dots, n$, each object in S either satisfies the property p_i or it does not. For $E \in 2^D$, let $g(E)$ the number of objects in S satisfying the properties in E (and possibly more), and let $f(E)$ be the number of objects in S satisfying the properties in E and no other properties. Then, the function $g : 2^D \rightarrow \mathbb{Z}_{\geq 0}$ can be expressed in terms of the function $f : 2^D \rightarrow \mathbb{Z}_{\geq 0}$ as follows:

$$g(E) = \sum_{F \supseteq E} f(F), \quad \forall E \in 2^D. \quad (22)$$

We will shortly see how one can express f in terms of g .

Whereas inverting the summation in Example 13 above can be done directly, inverting the summation over the poset 2^D in Example 14 motivates introducing more sophisticated machinery. In general, the *Möbius function* of a poset Q plays a fundamental role in inverting a summation over Q . Let Q be a finite poset.

The *Möbius function* of Q is the unique function $\mu_Q : Q \times Q \rightarrow \mathbb{Z}$ defined by $\mu_Q(q, q'') = 0$ when $q \not\leq q''$ and, when $q \leq q''$, by

$$\sum_{q' : q \leq q' \leq q''} \mu(q, q') = \delta(q, q''),$$

where $\delta(q, q'') = 0$ if $q \neq q''$ and $\delta(q, q'') = 1$ if $q = q''$. The function μ_Q can be computed recursively through the conditions:

$$\mu_Q(q, q'') = \begin{cases} 1, & q = q'', \\ -\sum_{q \leq q' < q''} \mu_Q(q, q'), & q < q'', \\ 0, & \text{otherwise.} \end{cases} \quad (23)$$

Theorem 3 (Möbius inversion formula, [Rot64]). Let Q be a finite poset. Let \mathbb{k} be a field. Then, for any pair of functions

$$f, g : Q \rightarrow \mathbb{k},$$

$$g(q) = \sum_{q' \leq q} f(q') \text{ for all } q \in Q$$

$$\iff f(q) = \sum_{q' \leq q} g(q') \cdot \mu_Q(q', q) \text{ for all } q \in Q. \quad (24)$$

This theorem also holds under more general assumptions guaranteeing that for every q the number of terms in the sum $\sum_{q' \leq q} f(q')$ is finite, cf. [Rot64].

Note that, in the statement above, the function g “looks like” a certain cumulative version of f so that one would expect f , customarily called the *Möbius inverse* of g , to be some sort of derivative of g . Also the theorem, in particular, implies that, given g there exists a *unique* such inverse.

Example 15. Recall the functions f and g given in Example 13. By solving the recurrence in Equation (23) when $Q = L_m$, we find

$$\mu_{L_m}(q, q') = \begin{cases} 1, & q = q', \\ -1, & q = q' - 1, \\ 0, & \text{otherwise.} \end{cases}$$

Hence, f given in Equation (21) is the Möbius inverse of g . This f captures the rate of change of g at each point in L_m .

It is interesting to point out that this intuition about Möbius inversion being related to the notion of derivative is consistent with how Equation (7) arose in the work of Landi and Frosini [LF97] as a way to obtain a compact representation of their *size functions* which were precursors to the rank invariants. Indeed, due to practical considerations, rank invariants are expected to be piecewise constant and it is precisely the points where they change that signal interesting features. This eventually leads to the notion of persistence diagram, as depicted in Figure 1 (cf. Remark 8).

Example 16. By solving the recurrence in Equation (23) when $Q = (2^D, \supseteq)$, one finds that the Möbius function of this poset is

$$\mu_{2^D}(E, F) = (-1)^{|E| - |F|}$$

for $E \supseteq F$. This, via Theorem 3, permits expressing the function f from Example 14 in terms of g .

Generalized Persistence Diagrams

By emulating the right-hand side of Equation (7), Möbius inversion over the poset $Q = \text{Int}(P)$ permits defining a notion of persistence diagram that is applicable to P -modules where P is *any finite connected poset*. Theorem 3 enables the following definition.

Definition 6. Let P be a finite connected poset. The *generalized persistence diagram* of $M : P \rightarrow \mathbf{vec}$ is the unique

function

$$\text{dgm}_M : \text{Int}(P) \rightarrow \mathbb{Z}$$

satisfying

$$\text{rk}_M(I) = \sum_{\substack{J \supseteq I \\ J \in \text{Int}(P)}} \text{dgm}_M(J) \quad \forall I \in \text{Int}(P). \quad (25)$$

In other words, dgm_M is the Möbius inverse of rk_M over the poset $Q = (\text{Int}(P), \supseteq)$, i.e., for $I \in \text{Int}(P)$,¹⁰

$$\text{dgm}_M(I) = \sum_{\substack{J \supseteq I \\ J \in \text{Int}(P)}} \mu_{\text{Int}(P)}(J, I) \text{rk}_M(J). \quad (26)$$

Example 17 (The case $P = L_m$). By solving the recurrence given in Equation (23) for $Q = \text{Int}(L_m)$, it can be checked that, for every $[b, d], [b', d'] \in \text{Int}(L_m)$, $\mu_{\text{Int}(L_m)}([b, d], [b', d'])$ equals μ_{ij} , as given in Equation (8), where i and j are such that $[b, d] = I_i$ and $[b', d'] = I_j$.

Properties of the generalized persistence diagram. In what follows, we examine several properties of the generalized persistence diagram.

Remark 7 (dgm_M is well-defined for arbitrary $M : P \rightarrow \text{vec}$). As opposed to the notion of barcode $\text{barc}(M)$, which requires the P -module M to be interval decomposable (cf. Definition 3), generalized persistence diagrams can be defined for any M . Furthermore, Theorem 4 below implies that $\text{barc}(M)$ and dgm_M determine each other whenever M is interval decomposable (cf. Definition 4). Hence, as $\text{barc}(\ast)$ is a complete invariant of interval decomposable modules, dgm_\ast is also a complete invariant of such modules.

Theorem 4. Let $M : P \rightarrow \text{vec}$ be interval decomposable. Then,

$$\text{dgm}_M = \text{mlt}_M.$$

Proof. By Proposition 2, we have that

$$\text{rk}_M(I) = \sum_{\substack{J \supseteq I \\ J \in \text{Int}(P)}} \text{mlt}_M(J).$$

The fact that dgm_M is the unique function satisfying Equation (25) implies the claim. \square

At the level of generality in which we have situated ourselves, it is no longer true that $\text{dgm}_M(I) \geq 0$ for all $I \in \text{Int}(P)$, as it is the case when $P = L_m$ (cf. Equation (7)). This leads to having to contend with *signed* persistence diagrams as we will see in the following examples. In particular, from Theorem 4 above we have:

¹⁰In [KM21a], the generalized rank invariant and the generalized persistence diagram of M are defined on the collection of all connected subsets of P , which is strictly larger than $\text{Int}(P)$. In [KM21b], the resulting version of rank invariant and persistence diagram were shown to be more discriminating than the one described in this article.

Corollary 5. If there exists $I \in \text{Int}(P)$ such that $\text{dgm}_M(I)$ is negative, then M is not interval decomposable.

It is not difficult to find that the converse of this statement does not hold.

Example 18 (The generalized persistence diagram can be negative). Consider $M : P \rightarrow \text{vec}$ given in Diagram (16). Note that $I := \{a\}$ is an interval in P . We show that $\text{dgm}_M(I) = -1$. Consider the following collections of intervals of P :

$$\begin{aligned} I^1 &:= \{\{a, b\}, \{a, c\}, \{a, d\}\}, \\ I^2 &:= \{\{a, b, c\}, \{a, b, d\}, \{a, c, d\}\}, \text{ and} \\ I^3 &:= \{\{a, b, c, d\}\} = \{P\}. \end{aligned}$$

By solving for the Möbius function of $\text{Int}(P)$ (via Equation (23)), from Equation (26), we have:

$$\text{dgm}_M(I) = \text{rk}_M(I) + \sum_{t=1}^3 (-1)^t \sum_{J \in I^t} \text{rk}_M(J).$$

One can check that, for $I = \{a\}$:

- $\text{rk}_M(I) = 2$, which is the dimension of $F(a) = \mathbb{k}^2$.
- $\text{rk}_M(J) = 1$ for every $J \in I^1$.
- $\text{rk}_M(J) = 0$ for every $J \in I^2$.
- $\text{rk}_M(P) = 0$, which follows from the monotonicity of rk_M (cf. Remark 5) and the previous item.

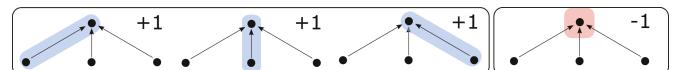
Therefore, the formula above gives

$$\text{dgm}_M(I) = 2 - (1 + 1 + 1) + (0 + 0 + 0) - 0 = -1.$$

By following a similar procedure for every $J \in \text{Int}(P)$, we eventually obtain:

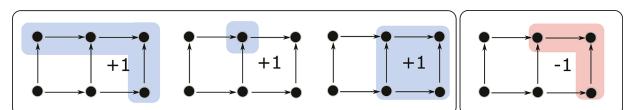
$$\text{dgm}_M(J) = \begin{cases} 1, & J \in I^1 \\ -1, & J = I \\ 0, & \text{otherwise,} \end{cases}$$

which can be depicted as follows:



It is noteworthy that M is not interval decomposable by Corollary 5.

Example 19. Consider the $(L_2 \times L_3)$ -module M given in Example 4 via a procedure similar to the one used in the previous example. Its generalized persistence diagram is depicted as:¹¹



¹¹Details can be found in [KM21b, page 16].

For any nonempty set Q , the *support* of a function $f : Q \rightarrow \mathbb{k}$ is defined to be the set consisting of those points $q \in Q$ with $f(q) \neq 0$.

For the persistence module M from Example 19, the support of dgm_M is much smaller than the support of rk_M . Indeed the support of the former consists exactly of the four intervals depicted above, whereas invoking Equation (25), one can verify that the support of rk_M consists exactly of 15 intervals of $L_2 \times L_3$. We have the following general remark.

Remark 8 (dgm_M is more parsimonious than rk_M). For arbitrary $M : P \rightarrow \mathbf{vec}$, the support of dgm_M is a subset of the support of rk_M . This follows from the fact that, for any $I \in \text{Int}(P)$, if $\text{rk}_M(I) = 0$, then $\text{dgm}_M(I) = 0$. Indeed, let $I \in \text{Int}(P)$ be such that $\text{rk}_M(I) = 0$. By the monotonicity of rk_M , we have that $\text{rk}_M(J) = 0$ for all $J \supseteq I$. Hence, the right-hand side of Equation (26) equals zero, and thus $\text{dgm}_M(I) = 0$.

Remark 9 (Flexibility in the choice of a target category). Patel's notion of persistence diagram [Pat18] encompasses persistence modules $M : \mathbb{Z} \rightarrow \mathcal{C}$ valued in categories \mathcal{C} more general than \mathbf{vec} . This feature is also available in the context of P -modules for arbitrary P that satisfy mild assumptions [KM21a]. This flexibility, for example, permits encoding clustering features of time-varying networks into a specialized version of the generalized persistence diagram [KM22].

More on generalized persistence diagrams. To conclude we provide pointers to other aspects and recent developments regarding generalized persistence diagrams.

In [BBE22], Betthauser et al. study an instance of Patel's generalized persistence diagrams exhibiting signed multiplicities which they connect to *persistence landscapes*. Remarkably, the authors also obtain stability via a version of the Wasserstein distance for signed measures. In [MP22] McCleary and Patel describe a pipeline for a certain generalized notion of persistence diagrams over metric lattices.

The stability of the generalized rank invariant from Definition 5 is addressed in [KM21a]. The stability of generalized persistence diagrams (as Möbius inverses of the generalized rank invariants) remains mostly open due to challenges coming from the fact that generalized persistence diagrams can attain negative values (cf. Example 18).

In [KM21b], Kim and Moore study the relationship of the generalized persistence diagram with the *bigraded Betti numbers* when $P = \mathbb{Z}^2$. In [BBH22, AENY23] the authors explore a connection between relative homological algebra and the generalized rank invariant.

There is a tension between the informativeness (strength) of given invariant of persistence modules and its computability. The generalized rank invariant can be

difficult to compute, even when $P \subseteq \mathbb{Z}^2$,¹² whereas in the case of this poset the standard rank invariant (cf. Equation (17)) is readily computable through software implementations such as RIVET (developed by Lesnick and Wright). An exciting current thread is finding the right balance between these two requirements so that efficient algorithms can be developed for the computation of such invariants.

ACKNOWLEDGMENT. We note that because of the limitation of including citations in this forum we have omitted many references by colleagues in this area. Such references have been included in the references we mentioned here.

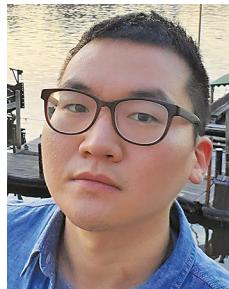
This research was partially supported by the NSF through grants DMS-1723003, CCF-1740761, CCF-2310412, DMS-1547357, and RI-1901360.

References

- [ADFK81] S. Abeasis, A. Del Fra, and H. Kraft, *The geometry of representations of A_m* , Math. Ann. 256 (1981), no. 3, 401–418, DOI 10.1007/BF01679706. MR626958
- [AENY23] Hideto Asashiba, Emerson G. Escolar, Ken Nakashima, and Michio Yoshiwaki, *Approximation by interval-decomposables and interval resolutions of persistence modules*, J. Pure Appl. Algebra 227 (2023), no. 10, Paper No. 107397, 20, DOI 10.1016/j.jpaa.2023.107397. MR4576885
- [BBE22] Leo Betthauser, Peter Bubenik, and Parker B. Edwards, *Graded persistence diagrams and persistence landscapes*, Discrete Comput. Geom. 67 (2022), no. 1, 203–230, DOI 10.1007/s00454-021-00316-1. MR4356248
- [BBH22] Benjamin Blanchette, Thomas Brüstle, and Eric J. Hanson, *Homological approximations in persistence theory*, Canadian Journal of Mathematics (2022), 1–38.
- [BdSS15] Peter Bubenik, Vin de Silva, and Jonathan Scott, *Metrics for generalized persistence modules*, Found. Comput. Math. 15 (2015), no. 6, 1501–1531, DOI 10.1007/s10208-014-9229-5. MR3413628
- [CdS10] Gunnar Carlsson and Vin de Silva, *Zigzag persistence*, Found. Comput. Math. 10 (2010), no. 4, 367–405, DOI 10.1007/s10208-010-9066-0. MR2657946
- [CZ09] Gunnar Carlsson and Afra Zomorodian, *The theory of multidimensional persistence*, Discrete Comput. Geom. 42 (2009), no. 1, 71–93, DOI 10.1007/s00454-009-9176-0. MR2506738
- [CSEH07] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer, *Stability of persistence diagrams*, Discrete Comput. Geom. 37 (2007), no. 1, 103–120, DOI 10.1007/s00454-006-1276-5. MR2279866
- [DW05] Harm Derkzen and Jerzy Weyman, *Quiver representations*, Notices Amer. Math. Soc. 52 (2005), no. 2, 200–206. MR2110070

¹²However, in [DKM22] Dey et al. show that persistence modules over zigzag posets can be utilized to compute the generalized rank invariant for $(L_m \times L_n)$ -modules.

- [DKM22] Tamal K. Dey, Woojin Kim, and Facundo Mémoli, *Computing Generalized Rank Invariant for 2-Parameter Persistence Modules via Zigzag Persistence and Its Applications*, 38th International Symposium on Computational Geometry (SoCG 2022), 2022, pp. 34:1–34:17.
- [KM21a] Woojin Kim and Facundo Mémoli, *Generalized persistence diagrams for persistence modules over posets*, J. Appl. Comput. Topol. 5 (2021), no. 4, 533–581, DOI 10.1007/s41468-021-00075-1. MR4334502
- [KM22] Woojin Kim and Facundo Mémoli, *Extracting persistent clusters in dynamic data via Möbius inversion*, arXiv preprint arXiv:1712.04064v5 (2022).
- [KM21b] Woojin Kim and Samantha Moore, *Bigraded Betti numbers and generalized persistence diagrams*, arXiv preprint arXiv:2111.02551 (2021).
- [Kin08] Ryan Kinser, *The rank of a quiver representation*, J. Algebra 320 (2008), no. 6, 2363–2387, DOI 10.1016/j.jalgebra.2008.06.011. MR2437505
- [LF97] Claudia Landi and Patrizio Frosini, *New pseudodistances for the size function space*, Vision geometry vi, 1997, pp. 52–60.
- [ML98] Saunders Mac Lane, *Categories for the working mathematician*, 2nd ed., Graduate Texts in Mathematics, vol. 5, Springer-Verlag, New York, 1998. MR1712872
- [MP22] Alexander McCleary and Amit Patel, *Edit distance and persistence diagrams over lattices*, SIAM J. Appl. Algebra Geom. 6 (2022), no. 2, 134–155, DOI 10.1137/20M1373700. MR4405184
- [Mun84] James R. Munkres, *Elements of algebraic topology*, Addison-Wesley Publishing Company, Menlo Park, CA, 1984. MR755006
- [Pat18] Amit Patel, *Generalized persistence diagrams*, J. Appl. Comput. Topol. 1 (2018), no. 3-4, 397–419, DOI 10.1007/s41468-018-0012-6. MR3975559
- [Rot64] Gian-Carlo Rota, *On the foundations of combinatorial theory. I. Theory of Möbius functions*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 2 (1964), 340–368 (1964), DOI 10.1007/BF00531932. MR174487



Woojin Kim



Facundo Mémoli

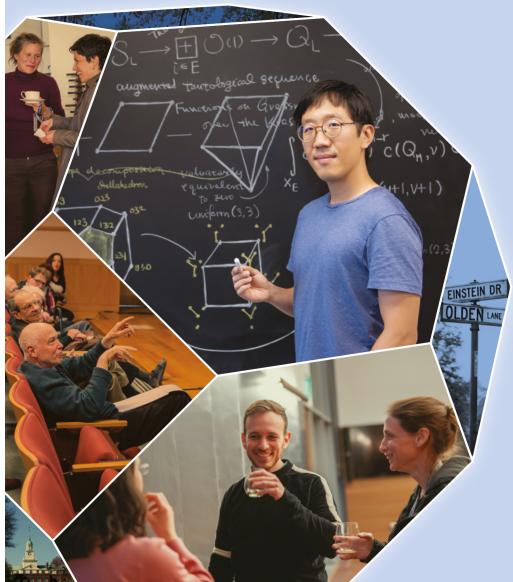
Credits

Opening image and author photo of Woojin Kim are courtesy of Woojin Kim.

All article figures are courtesy of the authors.

Photo of Facundo Mémoli is courtesy of Facundo Mémoli.

INSTITUTE FOR ADVANCED STUDY School of Mathematics



MEMBERSHIPS

With generous support from the National Science Foundation, the IAS School of Mathematics selects approximately 85 Members per year. The School welcomes applications from mathematicians and theoretical computer scientists at all career levels, and strongly encourages applications from underrepresented groups and mid-career scientists (6–15 years from Ph.D.). Competitive salaries, on-campus housing, and other resources are available for researchers in all mathematical subject areas. Most positions are for one or two terms, but for applicants who cannot leave their jobs or families for an

entire term, the School now offers a special two-month membership option. During the 2024–25 academic year the School will have a special program on Algebraic and Geometric Combinatorics. June Huh from Princeton University will be the Distinguished Visiting Professor.

Deadline is December 1, 2023.

PROGRAMS

Women and Mathematics

ias.edu/math/wam

Minerva Research Program

ias.edu/math/activities/programs/minerva-research-program

Summer Collaborators

ias.edu/math/summercollaborators

IAS

SCHOOL OF
MATHEMATICS
INSTITUTE FOR ADVANCED STUDY

ias.edu/math