Two complementary approaches in multi-parameter persistence : interval-decompositions and constructible functions

Vadim Lebovici

université
PARIS-SACLAY

Two Complementary Approaches in Multi-Parameter Persistence:

Interval-Decompositions and Constructible Functions

Deux approches complémentaires de la persistance multiparamétrique

Décompositions en intervalles et fonctions constructibles

Thèse de doctorat de l'université Paris-Saclay

Ecole Doctorale de Mathématique Hadamard (EDMH) n° 574
Spécialité de doctorat : Mathématiques aux interfaces
Graduate School : Mathématiques. Référent : Faculté des sciences d'Orsay

Thèse préparée dans le **Laboratoire de Mathématiques d'Orsay** (Université Paris-Saclay, CNRS), sous la direction de **Steve OUDOT**, directeur de recherche, et la co-direction de **François PETIT**, professeur junior

Thèse soutenue à Paris-Saclay, le 12 septembre 2023, par

Vadim LEBOVICI

Composition du jury

Membres du jury avec voix délibérative

Grégory GINOT Professeur, Université Sorbonne Paris Nord	Président
Ulrich BAUER Professeur associé, Technische Universität München	Rapporteur & Examinateur
William CRAWLEY-BOEVEY Professeur, Universität Bielefeld	Rapporteur & Examinateur
Frédéric CHAZAL Directeur de recherche, Inria Saclay	Examinateur
Bernhard KELLER Professeur, Université Paris Cité	Examinateur

THESE DE DOCTORAT

NNT: 2023UPASM018

université PARIS-SACLAY | ÉCOLE DOCTORALE de mathématiques Hadamard (EDMH)

Titre: Deux Approches Complémentaires de la Persistance Multiparamétrique : Décompositions en Intervalles et Fonctions Constructibles

Mots clés: Analyse Topologique de Données, Persistance Multiparamétrique, Théorie des Représentations, Transformations Intégrales, Fonctions Constructibles, Calcul Intégral d'Euler

Résumé: Les modules de persistance multiparamétriques n'admettent, contrairement à leurs analogues uniparamétriques largement utilisés en analyse de données, pas de codes-barres, et plus généralement, aucune description complète et facilement manipulable numériquement. Cette thèse propose de contourner cet obstacle par deux approches. La première consiste à identifier des sous-classes de tels modules qui sont effectivement décrites par des multi-ensembles d'intervalles à plusieurs paramètres. Gardant à l'esprit la nécessité de tester algorithmiquement l'appartenance à ces sous-classes, nous étudions l'existence de celles admettant une caractérisation locale, c'est-à-dire celles dont l'appartenance peut être testée en observant uniquement les restrictions des modules à des sous-ensembles finis de l'espace des paramètres. Nous montrons que si la sous-classe des modules décomposables en intervalles n'admet pas une telle caractérisation locale, la classe des modules décomposables en rectangles en admet une, et qu'elle est, dans un sens précis, une classe maximale ayant cette propriété. La deuxième approche consiste à construire des invariants riches et facilement calculables des modules de persistance à plusieurs paramètres sous forme de fonctions constructibles. Cette approche contourne complètement la construction des modules via des calculs de caractéristique d'Euler plutôt que d'homologie. Nous introduisons des transformées intégrales de fonctions constructibles combinant mesure de Lebesgue et intégration par rapport à la caractéristique d'Euler. On mène une étude systématique de ces transformées, prouvant des résultats de régularité, d'injectivité, de stabilité et statistiques. On prouve des formules d'indices permettant de calculer l'espérance de ces transformées dans le contexte de la persistance des sous-niveaux de champs gaussiens aléatoires. On montre enfin l'efficacité de ces outils en analyse de données, en fournissant de nombreux exemples de l'information topologique et géométrique qu'ils capturent.

Title: Two Complementary Approaches in Multi-Parameter Persistence: Interval-Decompositions and Constructible Functions

Keywords: Topological Data Analysis, Multi-Parameter Persistence, Representation Theory, Integral Transforms, Constructible Functions, Euler Calculus

Abstract: Multi-parameter persistence modules do not admit barcodes—unlike their widely used one-parameter analogues in topological data analysis—, and more generally no complete and computationally manageable description. This thesis proposes to circumvent this obstacle through two approaches. The first consists of identifying subclasses of such modules that are indeed described by multi-sets of multi-parameter intervals. Keeping in mind the need to algorithmically test membership in these subclasses, we study the existence of those with a local characterization, i.e., those whose membership can be tested by observing only the restrictions of the persistence modules to finite subsets of the parameter space. We show that while the subclass of interval-decomposable modules does not admit such a local characterization, the class of rectangle-decomposable modules does, and it is, in some precise sense, a maximal class with this property. The second approach is to build informative and efficiently computable invariants of multi-parameter persistence modules in the form of constructible functions. This bypasses the modules construction altogether by using Euler characteristic computations instead of homological ones. We introduce and conduct a systematic study of integral transforms of constructible functions involving Lebesgue measure and integration with respect to the Euler characteristic. We study their regularity, their injectivity, their stability and their statistical properties. Finally, these tools are discussed as powerful descriptors in data analysis, providing examples of the topological and geometric information captured.

Aux colosses d'argile
et aux vapeurs d'essence

Remerciements

Je souhaite exprimer ma plus sincère et chaleureuse reconnaissance à toutes les personnes sans qui ce travail de trois ans n'aurait pu aboutir.

I am sincerely grateful to Bill Crawley-Boevey and Ulrich Bauer for agreeing to be the referees of my thesis. Their scientific opinon matters to me as their research is a clear inspiration for the work developed throughout this manuscript. Je remercie chaleureusement Bernhard Keller, Grégory Ginot et Frédéric Chazal qui me font l'honneur d'être dans le jury de ma soutenance de thèse, ainsi que Pierre Schapira, dont l'influence des idées est si manifeste tout au long de ce manuscrit. Pour tes relectures et ton soutien inconditionnel, encore une fois, Pierre, merci.

L'accompagnement scientifique et humain que m'ont apporté mes directeurs de thèse a été plus qu'essentiel. Je dois à Steve de m'avoir fait découvrir l'analyse topologique de données et de m'avoir initié à la décomposition des modules de persistance, domaine auquel je suis fier d'avoir pu contribuer dans ces pages. J'ai beaucoup appris lors de nos discussions, mathématiques ou non, depuis le stage de master jusqu'à la soutenance de cette thèse.

Ma dette envers François est immense. Le paysage mathématique *constructible* que j'ai découvert grâce à lui m'enchante, et j'espère bien avoir la chance d'y poursuivre mon voyage à l'avenir. Plus encore, il n'a jamais compté ni son temps, ni ses conseils, ni ses explications, ni ses idées. Ce travail doit tant à sa générosité.

This manuscript owes a lot to Magnus Bakke Botnan. Collaborating with him has always been as warm as instructive. Comment ne pas remercier Olympio Hacquard, avec qui les numéros de claquettes sont si agréables. Son humilité, sa franchise et sa bonhomie n'ont pas uniquement permis une collaboration fructueuse, mais aussi une fin de thèse plus heureuse. J'ajoute d'ailleurs une pensée émue pour Etienne Lasalle et les *Datashapers*, avec qui l'argent tombe tout simplement du ciel.

Lors de ma scolarité à l'ENS, j'ai eu la chance de rencontrer un homme d'une intelligence rare, un mathématicien exceptionnel, d'une sincère bienveillance : Nicolas Bergeron m'a toujours tenu sa porte ouverte, partageant généreusement ses conseils éclairants, ses connaissances expertes et ses anecdotes tordantes. Sa vision sur les mathématiques et l'enseignement a tant contribué à faire mûrir mon jugement sur ce métier. Sa sagesse scientifique et humaine — malgré son *très jeune* âge — est une source d'admiration et d'inspiration constante.

Ces trois années ont été largement ensoleillées par tant de collègues, chercheur.e.s ou non, si agréables à fréquenter. J'ai une pensée particulière pour Nicolas Berkouk,

pour ses conseils avisés et son soutien continu durant ces trois années. Le département de mathématiques de l'ENS a immédiatement été une seconde maison, à une période où sortir de chez soi était interdit. Faire des mathématiques là-bas est un plaisir qui se partage autant que les chouquettes et le chocolat. J'ai une pensée particulière pour les habitants du T9 et de ses environs, qui auront ponctué mon quotidien d'éclats de rire et de discussions savantes. J'ai tant appris également au sein de l'équipe Datashape, au sujet de l'analyse topologique de données d'abord, des mathématiques appliquées plus largement, et du mode de vie méditerrannéen, enfin. Je remercie tout particulièrement mes partenaires et adversaires de pétanque pour leur sérieux et leur engagement. Ces semaines passées en votre compagnie à faire des mathématiques sur l'île paradisiaque de Porquerolles demeurent parmi les plus beaux moments de ces trois années.

Je dois tant à mes amis qui m'accompagnent. J'ai toujours pu compter sur leur jugement et leur bienveillance. Qu'ils soient clubbeurs, champions de puissance 4, artistes philosophes, mafieux italiens, Madones, bretons ou créatrices, qu'ils soient cyniques, loups solitaires ou pichous, qu'ils soient spécialistes ou volleyeurs enragés, joueurs d'échecs, batteurs, ou basketteurs, qu'ils soient ouverts à l'authentique, codeurs fous, profs de ski ou grands romantiques, oiseaux exotiques, aquarellistes ou guitaristes ferrovipathes, ce sont eux qui tissent le bonheur quotidien qui permet à cette thèse d'advenir. Je les remercie du fond du cœur.

A ma famille, la reconnaissance que je dois exprimer ici est vertigineuse. A mon père si fort, à ma mère si lumineuse, à Marie, à Jean, qui m'ont toujours témoigné leur amour, à mes grands-mères si aimantes, mes oncles et mes tantes si pétillants et rassurants, mes cousins si drôles, mon unique Bapt. A ma sœur si spirituelle et si courageuse. A mon frère si sincère et si doux. Ce voyage de trois ans n'aurait pas été possible sans eux.

Comment exprimer enfin la nécessité que constitue la présence de Marion à mes côtés. A la verité, si je compare tout le reste de ma vie, quoy qu'avec la grace de Dieu je l'aye passée douce, aisée et exempte d'affliction poisante, pleine de tranquillité d'esprit, ayant prins en payement mes commoditez naturelles et originelles sans en rechercher d'autres : si je la compare, dis-je, toute aux six années qu'il m'a esté donné de jouyr de la douce compagnie et société de ce personnage, ce n'est que fumée, ce n'est qu'une nuit obscure et ennuyeuse.

Ses qualités sont autant de modèles qui me guident au quotidien. J'estime son intelligence, sa détermination, son courage. J'admire sa vivacité, sa finesse, sa sensibilité. Sa gourmandise et sa tendresse me rendent heureux. Elle est la chaude lumière qui a brillé sur ces années de thèse.

Septembre 2023 Vadim Lebovici

Contents

Introduction (version française)

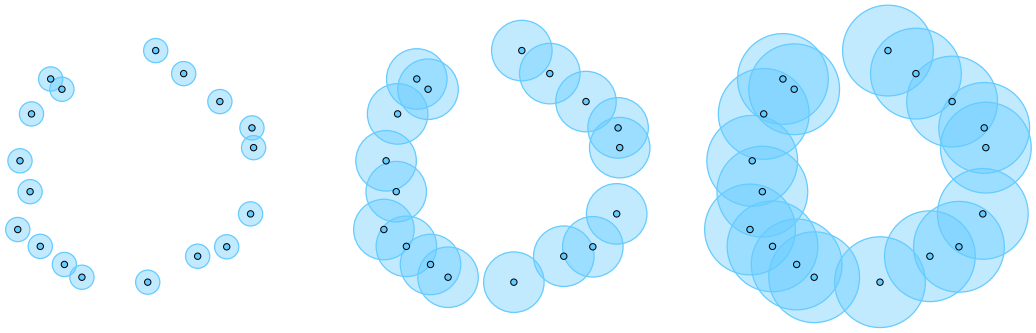
Cette thèse ambitionne de contribuer au domaine de l'*analyse topologique de données*, et plus précisément à sa branche appelée la *persistance multiparamétrique*. L'objectif de l'analyse topologique de données est d'extraire des informations géométriques et topologiques de données complexes afin de réaliser une grande variété de tâches, telles que la régression, la classification ou la visualisation. Dans cette introduction, nous commençons par préciser ce que nous entendons par *données* et par *information géométrique et topologique*. Puis, nous décrivons les outils de la persistance multiparamétrique permettant de répondre à l'objectif de l'analyse topologique de données et détaillons enfin comment nos contributions participent à cette ambition.

Cadre. En analyse topologique de données, les données sont généralement comprises comme un espace métrique (X, d). Par exemple, l'espace X peut être un sous-ensemble fini d'un espace euclidien muni de la métrique induite, alors appelé *nuage de points*, ou un graphe muni de la distance de plus court chemin (cf. 1). Dans le cas d'un graphe, on entend généralement par information topologique le nombre de composantes connexes ou le nombre de cycles du graphe. Pour aller plus loin, on peut construire un complexe simplicial à partir du graphe, tel que le *complexe de clique* par exemple, qui encode les propriétés de connectivité d'ordre supérieur sous une forme topologique. En revanche, il n'y a *stricto sensu* aucune information topologique pertinente à extraire d'un nuage de points hormi son nombre de points (d'une utilité toute relative). La première information géométrique pertinente est le nombre de groupes significatifs en lesquels les points sont répartis. Le domaine de l'analyse de données consacré à l'extraction de ces groupes est appelé le *partitionnement de données* (en anglais : *clustering*). Après avoir partitionné les données considérées, nous pouvons nous demander comment les points de chaque groupe sont répartis géométriquement. Supposons par exemple que le nuage de points en question se trouve à proximité d'une sous-variété de $\mathbb{R}^n$ inaccessible à l'expérimentateur·rice. Il est alors pertinent d'extraire les propriétés géométriques et topologiques de cette sous-variété ; nous citerons plus loin dans l'introduction des exemples d'applications concrètes ayant recours à cette information.

Persistance à un paramètre. L'outil principal de l'analyse topologique de données permettant de dégager les propriétés géométriques et topologiques d'un jeu de données est l'*homologie persistante*. La première étape de cette méthode consiste à construire une famille $\mathcal{F}$ d'espaces topologiques paramétrée par un ensemble totalement ordonné $(\mathcal{P}, \leq)$ et croissante pour l'inclusion. Une telle famille est appelée *filtration uniparamétrique* et illustrée en figure 2. Le principe de l'homologie persistante est de suivre l'évolution

(a) (b)

FIGURE 1 : Deux exemples d'espaces métriques (X, d). La figure 1a est extraite de Su et al. (2004) : elle représente le graphe d'expression des gènes de tissus de souris en suivant la méthode de l'analyse du réseau de corrélation des gènes, que développent Freeman et al. (2022). La figure 1b est tirée de De Deuge et al. (2013) et représente un nuage de points produit par le scan LIDAR d'un arbre.



(a) $d(X, \cdot) \leq 0.4$ (b) $d(X, \cdot) \leq 1.5$ (c) $d(X, \cdot) \leq 2.1$

FIGURE 2 : Filtration de Čech d'un nuage de points dans le plan euclidien. Les ensembles de sous-niveaux de la fonction $d(X, \cdot)$ sont représentés en bleu clair.

des caractéristiques homologiques des espaces de $\mathcal{F}$ lorsque le paramètre d'indexation varie. Typiquement, on peut considérer la famille des ensembles de sous-niveaux d'une fonction $f : Y \to \mathcal{P}$ pour un certain espace topologique Y. Lorsque X est un nuage de points de $\mathbb{R}^d$, un exemple célèbre, appelé *filtration de Čech*, est donné par la fonction à valeurs réelles $f = d(X, \cdot)$ qui associe à tout point de $Y = \mathbb{R}^d$ sa distance au nuage X.

Le calcul de l'homologie à coefficients dans un corps $\mathbf{k}$ des espaces de la filtration $\mathcal{F}$ produit un objet algébrique $H_k(\mathcal{F})$ appelé *module de persistance* (uniparamétrique), qui est un foncteur de l'ensemble partiellement ordonné $(\mathcal{P}, \leq)$ vers la catégorie $\mathsf{Vec}_{\mathbf{k}}$ des espaces vectoriels sur $\mathbf{k}$. En d'autres termes, un module de persistance $M : (\mathcal{P}, \leq) \to \mathsf{Vec}_{\mathbf{k}}$ est la donnée d'une famille d'espaces vectoriels $M(p)$ pour chaque $p \in \mathcal{P}$ et d'une famille d'applications linéaires $M(p) \to M(q)$ pour chaque paire d'éléments comparables $p \leq q$ de $\mathcal{P}$. De plus, ces applications linéaires doivent satisfaire certaines conditions de compatibilité. Sous certaines hypothèses, les résultats de Gabriel (1972) en théorie des représentations de carquois assurent que l'on peut faire un choix cohérent de base pour chaque espace vectoriel de $H_k(\mathcal{F})$, au sens où chaque vecteur de base apparaît dans ces

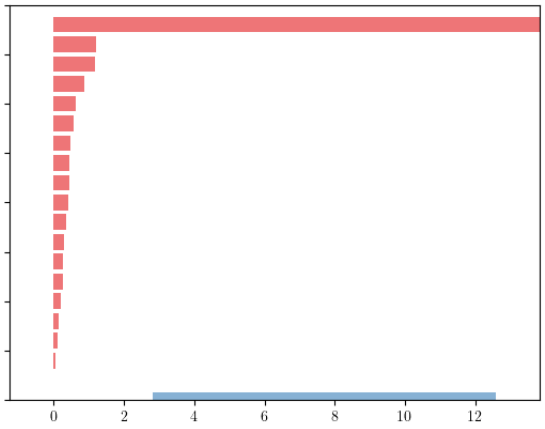


FIGURE 3 : Codes-barres associé à la filtration de Čech du nuage de points X présenté en figure 2. Les barres rouges correspondent aux barres du code-barres de $H_0(\mathcal{F})$ et les bleues aux barres du code-barres de $H_1(\mathcal{F})$. La longue barre bleue enregistre le fait que les points de X sont proches d'un cercle.

espaces à partir d'un certain paramètre $a \in \mathcal{P}$ appelé sa *naissance*, et disparaît après un certain paramètre $b > a$ appelé sa *mort*. L'intervalle $[a, b)$ est appelé la *barre* du vecteur de base correspondant. Le module de persistance $H_k(\mathcal{F})$ est alors entièrement décrit par le multiensemble des barres $[a, b)$ pour chaque caractéristique homologique apparaissant dans la filtration (cf. figure 3). Plus généralement, les classes d'isomorphisme des modules de persistance dont les espaces vectoriels sont de dimension finie sont en bijection avec les multiensembles d'intervalles de $\mathcal{P}$, appelés *codes-barres de persistance* (Botnan et Crawley-Boevey, 2020 ; Crawley-Boevey, 2015). Ces idées sont nées dans les années 1990 avec les travaux de Barannikov (1994) et ont été activement développées en analyse topologique de données depuis le travail fondateur de Edelsbrunner, Letscher et Zomorodian (2000). Les livres de Edelsbrunner et Harer (2022) et Oudot (2015) fournissent une plus ample introduction à l'homologie persistante.

Une propriété remarquable des codes-barres est leur robustesse vis-à-vis de petites perturbations (inévitables) sur les données. Plus précisément, on peut munir l'espace des codes-barres des distances *bottleneck* (Cohen-Steiner, Edelsbrunner et Harer, 2007) et de *Wasserstein p* (Cohen-Steiner, Edelsbrunner, Harer et Mileyko, 2010) ayant des propriétés de stabilité par rapport à certaines distances sur les données (Chazal, De Silva et al., 2016 ; Cohen-Steiner, Edelsbrunner et Harer, 2007 ; Cohen-Steiner, Edelsbrunner, Harer et Mileyko, 2010). Par exemple, la distance *bottleneck* est stable par rapport à la métrique L_∞ entre les fonctions $f : Y \to \mathbb{R}$ dont les sous-niveaux sont considérés. Ce résultat donne des précisions sur le type de bruit dans les données par rapport auquel les codes-barres sont stables et constitue probablement l'un des résultats les plus importants de la théorie de la persistance. En particulier, cette propriété fait du code-barres un outil adapté pour construire des estimateurs consistants en statistiques (Bobrowski, Mukherjee et Taylor, 2017 ; Chazal, Glisse et al., 2015).

Les codes-barres ont de fait été utilisés avec succès en statistiques et ont trouvé une large variété d'applications scientifiques, allant des sciences de la santé (Aukerman et al., 2021 ; Fernández et Mateos, 2022 ; Rieck et al., 2020), à la biologie (Ichinomiya,

Obayashi et Hiraoka, 2020 ; Rabadán et Blumberg, 2019) en passant par les sciences des matériaux (Hiraoka, Nakamura et al., 2016 ; Lee et al., 2017). De plus, ces outils ont trouvé de nombreuses applications en géométrie symplectique (Polterovich, Rosen et al., 2020 ; Polterovich et Shelukhin, 2016).

Persistance multiparamétrique. Dans de nombreuses situations expérimentales, il est utile de construire des filtrations indéxées par plus d'un paramètre, c'est-à-dire, par un ensemble partiellement ordonné (poset) comme $\mathcal{P} = \mathbb{Z}^m$ ou $\mathcal{P} = \mathbb{R}^m$ muni de l'ordre produit[1]. Une façon d'obtenir une telle filtration est de considérer les ensembles de sous-niveaux de fonctions à valeurs vectorielles $f : Y \to \mathbb{R}^m$. En procédant ainsi, il est possible de diminuer l'impact des points aberrants d'un nuage de points X de $\mathbb{R}^d$ en filtrant l'espace par rapport à $f = (f_1, f_2) : \mathbb{R}^d \to \mathbb{R}^2$ où f_1 est la distance à X et f_2 est une densité de probabilité sur $\mathbb{R}^d$ estimée à partir de X (cf. Vipond et al. 2021 par exemple). De plus, les filtrations multiparamétriques permettent de traiter des données intrinsèquement multiparamétrique, comme des images de tissus de seins atteints de cancer et munies de la densité de cellules immunitaires et cancéreuses sur chaque pixel (Carrière et Blumberg, 2020). Appliquer l'homologie à une telle filtration produit désormais un objet appelé *module de persistance multiparamétrique* $M : (\mathcal{P}, \leq) \to \mathsf{Vec}_{\mathbf{k}}$. La théorie de la persistance multiparamétrique a débuté par des calculs de groupes d'homotopie de taille dans Frosini et Mulazzani (1999) et a été étudié pour la première fois par Carlsson et Zomorodian (2009) pour les groupes d'homologie. En suivant la même ligne de pensée que dans le cas uniparamétrique, il est naturel de se demander si les classes d'isomorphisme de tels objets peuvent être décrites par des multiensembles de régions géométriques simples de l'espace des paramètres, semblables aux barres des codes-barres de persistance.

Born to be wild. La théorie des représentations de carquois ne laisse malheureusement aucun espoir de répondre positivement à cette question. Si tôt que le poset $\mathcal{P}$ des paramètres est $[\![1, 2]\!] \times [\![1, 6]\!]$ muni de l'ordre produit, la situation est aussi complexe qu'elle puisse être. On dit que le poset est de *type de représentation sauvage*. En particulier, une classification complète des classes d'isomorphismes des modules de persistance indexés sur ce poset particulier impliquerait aussitôt une telle classification pour les modules de persistance indexés sur n'importe quel poset fini. La construction d'invariants *complets* — caractérisant entièrement la type d'isomorphisme du module — qui soient à la fois calculables et manipulables en pratique est considérée comme complètement hors de portée. Nous renvoyons le lecteur à l'article de Botnan et Lesnick (2022, Section 8) pour une description plus complète de la théorie des représentations apparaissant en persistance multiparamétrique.

Défi. L'une des principales difficultés de la persistance multiparamétrique est donc d'extraire des informations stables et riches des modules de persistance malgré le caractère sauvage du poset des paramètres. Les travaux menés dans cette thèse de doctorat développe deux approches complémentaires afin de surmonter cette difficulté. La première consiste à identifier des sous-classes de modules de persistance multiparamétriques

[1] L'ordre produit est défini par $(x_1, \ldots, x_m) \leq (y_1, \ldots, y_m) \in \mathbb{R}^m$ si $x_i \leq y_i$ pour tout $i \in \{1, \ldots, m\}$.

pour lesquels un analogue du code-barres uniparamétrique existe. La seconde consiste à construire des invariants des modules de persistance qui soient stables et informatifs, bien qu'incomplets.

I. Décompositions des modules de persistance

Dans cette section, nous présentons la théorie des décompositions de modules de persistance multiparamétriques. Notre objectif n'est pas d'être exhaustifs mais de décrire le cadre dans lequel nos contributions s'inscrivent. Nous renvoyons à l'article de Botnan et Lesnick (ibid.) pour un tableau récent et plus détaillé.

Théorèmes de décomposition. Le travail de Botnan et Crawley-Boevey (2020) montre que les modules de persistance multiparamétriques qui sont point-à-point de dimension finie — c'est-à-dire, dont les espaces vectoriels sont de dimension finie — se décomposent toujours en sommes directes de modules indécomposables. De plus, l'anneau d'endomorphismes de chacun de ces modules indécomposables est local et le théorème d'Azumaya (1950) assure donc que cette décomposition est essentiellement unique.

Dans le cas uniparamétrique, ces indécomposables sont nécessairement des *modules intervalles*, définis comme les représentations indicatrices d'intervalles de $\mathcal{P}$ (cf. Crawley-Boevey 2015). Ces modules sont donc entièrement caractérisés par leur *support*, c'est-à-dire l'ensemble des paramètres $p \in \mathcal{P}$ tels que l'espace vectoriel $M(p)$ est non trivial. Une classe d'isomorphisme de modules de persistance uniparamétriques est donc entièrement caractérisée par le multiensemble des intervalles apparaissant comme supports des modules indécomposables dans la décomposition en somme directe de n'importe quel représentant de la classe. C'est ce multiensemble qui est appelé code-barres, comme nous l'avons expliqué plus haut.

La situation est bien plus complexe dans le cas biparamétrique, car les modules indécomposables ne sont pas entièrement caractérisés par leur support, comme le montre l'exemple suivant de deux modules indécomposables de même support (cf. (1.4)) :

$$\begin{array}{ccccc} \mathbf{k} & \xrightarrow{\binom{1}{1}} & \mathbf{k}^2 & \xrightarrow{(1\ 0)} & \mathbf{k} \\ {\scriptstyle 0}\uparrow & & \uparrow{\scriptstyle \binom{1}{0}} & & \uparrow{\scriptstyle 1} \\ 0 & \xrightarrow{0} & \mathbf{k} & \xrightarrow{1} & \mathbf{k} \end{array} \qquad \begin{array}{ccccc} \mathbf{k} & \xrightarrow{1} & \mathbf{k} & \xrightarrow{1} & \mathbf{k} \\ {\scriptstyle 0}\uparrow & & \uparrow{\scriptstyle 1} & & \uparrow{\scriptstyle 1} \\ 0 & \xrightarrow{0} & \mathbf{k} & \xrightarrow{1} & \mathbf{k} \end{array}$$

Cependant, on peut toujours définir des modules intervalles comme des représentations indicatrices d'intervalles de $\mathcal{P}$, où l'on entend ici par « intervalle » tout sous-ensemble convexe connexe de $\mathcal{P}$ au sens de la théorie de l'ordre (cf. figure 4). La collection des supports des modules intervalles apparaissant dans une décomposition en somme directe d'un module de persistance peut ensuite être utilisée comme descripteur de ce module. Ce descripteur étant de nature purement géométrique, il est facile à interpréter et à implémenter sur un ordinateur.

En pratique, on souhaite pouvoir déterminer facilement si un module de persistance donné admet des modules intervalles dans sa décomposition en somme directe et si, dans une telle configuration, ce sont les seuls indécomposables apparaissant dans la décomposition — auquel cas, le module est appelé *décomposable en intervalles*. Une approche

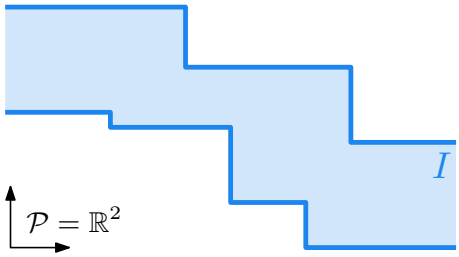


FIGURE 4 : Un intervalle I du poset $\mathcal{P} = \mathbb{R}^2$.

naïve de ce problème consiste à utiliser un algorithme général de décomposition de module de persistance et de vérifier les indécomposables un à un. Dans la première partie de cette thèse, nous proposons une approche alternative.

Caractérisations locales. La méthode que nous proposons consiste à vérifier *localement* la décomposabilité du module, au sens où l'on aurait à examiner uniquement les restrictions du module à des sous-ensembles *tests* du poset $\mathcal{P}$. Si la collection de sous-ensembles tests est assez petite, et si les ensembles tests eux-mêmes sont assez simples, les restrictions du module devraient avoir une structure particulièrement simple à analyser, ce qui constituerait un gain algorithmique significatif.

Ces caractérisations locales apparaissent en *persistance des niveaux* (Bendich et al., 2013 ; Carlsson, De Silva et D. Morozov, 2009), où l'on construit des invariants de fonctions scalaires f définies sur des espaces topologiques en étudiant la topologie des images réciproques d'intervalles ouverts par f. A une reparamétrisation près, les groupes d'homologie des ces pré-images s'assemblent en un module de persistance biparamétrique indexé sur $\mathcal{P} = \mathbb{R}^2$. Ce module se trouve être décomposable en intervalles, pour des intervalles appelés *blocs* : des quarts de plan supérieurs droits ou inférieurs gauches, ou des bandes infinies horizontales ou verticales. La décomposabilité en blocs des modules considérés dans ce contexte est une conséquence évidente de la suite de Mayer-Vietoris, dès lors que la caractérisation locale suivante est prouvée : un module de persistance biparamétrique est décomposable en blocs si, et seulement si, toutes ses restrictions à des *carrés* $\{x_1, x_2\} \times \{y_1, y_2\} \subset \mathbb{R}^2$ le sont également. Ce fait a d'abord été démontré par Cochoy et Oudot (2020), puis par Botnan et Crawley-Boevey (2020) et a permis d'obtenir une théorie plus propre et plus générale que celle établie précédemment pour la persistance des niveaux, qui nécessitait une condition supplémentaire de type Morse sur la fonction à valeurs réelles f considérée (Bendich et al., 2013).

Contributions. Dans la première partie de cette thèse, nous développons la théorie des caractérisations locales des modules de persistance biparamétriques en élargissant la classe d'intervalles considérés de la classe des blocs à la classe des rectangles parallèles aux axes de $\mathbb{R}^2$. Un module intervalle supporté sur un rectangle parallèle aux axes est appelé *module rectangle*, et un module décomposable en intervalles dont les sommandes sont exclusivement des modules rectangles est appelé *décomposable en rectangles*. Nous prouvons qu'un module de persistance biparamétrique est décomposable en rectangles si, et seulement si, toutes ses restrictions aux carrés le sont également. Nous prouvons de plus que les modules décomposables en rectangles sont, dans un sens précis, une sous-classe

maximale de modules décomposables en intervalles qui peut être caractérisée localement, du moins lorsque l'on considère les carrés comme ensembles tests. Ces résultats sont le fruit d'une collaboration avec Magnus Bakke Botnan et Steve Oudot. Ils ont d'abord été obtenus sur des produits binaires d'ensembles totalement ordonnés finis (2022), puis étendus aux produits de tels ensembles satisfaisant certaines hypothèses peu restrictives, incluant $\mathcal{P} = \mathbb{R}^2$ par exemple (2023).

Nous présentons ensuite deux applications de ces résultats à l'analyse topologique de données. Tout d'abord, nous exposons un algorithme testant la décomposabilité en rectangles des modules obtenus par le calcul de l'homologie des filtrations simpliciales finies biparamétriques. Lorsque le module est décomposable en rectangle, nous fournissons également un algorithme de calcul de sa décomposition ayant une complexité plus faible que les algorithmes généraux de décomposition des modules biparamétriques. Notre algorithme de test est fondé sur notre théorème de caractérisation locale : il suffit de vérifier la décomposabilité en rectangle des restrictions du module aux carrés du poset, et cette dernière propriété peut être formulée comme des égalités entre certaines images et certains noyaux des morphismes $M(p) \to M(q)$ du module pour $p \leq q$ dans $\mathcal{P}$.

Enfin, nous appliquons notre résultat de décomposabilité en rectangle aux modules de persistance indexés sur le poset appelé *pyramide de Mayer-Vietoris*. Ces modules ont été introduits en persistance des niveaux par Bendich et al. (2013) afin d'étudier les groupes d'homologie relative des pré-images d'intervalles ouverts par une application continue à valeurs réelles. Notre résultat permet de généraliser le théorème de la base pyramidale de Bendich et al. (ibid.) aux modules de persistance qui ne sont pas nécessairement déterminés par leur restriction à un sous-ensemble fini de la pyramide, et en fait plus largement aux modules indexés sur un poset appelé *bande de Mayer-Vietoris* contenant la pyramide et introduit par Bauer, Botnan et Fluhr (2021).

II. Invariants of persistence modules

Dans cette section, nous décrivons de façon succincte et non exhaustive la théorie des invariants des modules de persistances multiparamétrique afin de placer nos contributions dans leur contexte.

Par *invariant*, on entend une application de l'ensemble des classes d'isomorphisme de modules de persistance vers un espace (pseudo-)métrique. La vaste littérature de la théorie des représentation de carquois fournit de nombreux invariants, et l'analyse topologique de données en a apporté de supplémentaires, empreints des contextes applicatifs dans lesquels ces invariants sont utilisés.

Qu'est-ce qu'un bon invariant ? En vue de son application en analyse de données, un bon invariant des modules de persistance devrait être :[2]

(i) Informatif, autant que possible, sinon complet. En particulier, la distance entre les invariants devrait être d'autant plus importante que les propriétés topologiques des données diffèrent.

[2]On note qu'une liste similaire apparaît dans Adams et al. (2017) dans le cas spécifique des invariants à valeur vectorielles des modules de persistance à un paramètre.

(ii) Calculable, aussi efficacement que possible.

(iii) Stable, c'est-à-dire continu, ou même lipschitzien par rapport à une certaine métrique sur les données. En gardant à l'esprit les codes-barres de persistance, la métrique sur les données pourrait être une métrique L_p entre les fonctions dont les ensembles de sous-niveaux sont considérés, ou la métrique de Hausdorff sur les nuages de points. Comme dans le cas des codes-barres de persistance, cette propriété garantirait que l'invariant est robuste par rapport à un certain type de bruit dans les données, et pourrait être utilisée pour construire des estimateurs consistants en statistique.

(iv) Interprétable. Cette propriété est différente de (i) et peut être difficile à obtenir en fonction de l'invariant considéré.

En outre, de nombreuses techniques d'apprentissage automatique comme les machines à vecteurs de support requièrent une structure hilbertienne sur leurs entrées. Il serait donc appréciable que la (pseudo-)métrique définie sur l'espace des invariants soit Lipschitz-équivalente à une métrique induite par le produit scalaire d'un espace de Hilbert.

Le code-barres est-il un bon invariant ? Nous pouvons étudier les propriétés des codes-barres de persistance uniparamétrique du point de vue de la liste dressée ci-dessus :

(i) Comme nous l'avons déjà mentionné, cet invariant est complet et satifsait donc entièrement la propriété (i).

(ii) L'un des avantages des groupes d'homologie, sur les groupes d'homotopie par exemple, est leur faible coût de calcul : si le poset d'indexation $\mathcal{P}$ est un ensemble totalement ordonné fini et si les espaces de la filtration $\mathcal{F}$ sont des complexes simpliciaux, le calcul d'un code-barres de persistance revient essentiellement à une simple réduction de matrice. Par conséquent, cette tâche prend un temps polynomial en le nombre de simplexes apparaissant dans la filtration, de degré identique à celui pour la multiplication de matrices (Milosavljević, D. Morozov et Skraba, 2011).

(iii) Comme nous l'avons déjà mentionné, l'espace des codes-barres est muni des distances *bottleneck* et de Wasserstein p satisfaisant des propriétés de stabilité (Chazal, De Silva et al., 2016 ; Cohen-Steiner, Edelsbrunner et Harer, 2007 ; Cohen-Steiner, Edelsbrunner, Harer et Mileyko, 2010).

(iv) Les codes-barres de persistance sont interprétables. Les longues barres représentent les caractéristiques homologiques persistantes dans les données et fournissent donc des indications sur leur dimension intrinsèque et leur structure topologique. Par exemple, les codes-barres de persistance ont récemment permis de détecter une structure toroïdale de la zone d'activation des cellules de grille des cerveaux de rongeurs (Gardner et al., 2022).

Malheureusement, l'espace des codes-barres de persistance muni de la distance *bottleneck* ne peut pas être plongé isométriquement dans un espace de Hilbert (Bubenik et A.

Wagner, 2020 ; Carrière et Bauer, 2019). En outre, le point (iv) devrait être nuancé par deux remarques. Premièrement, les informations homologiques recueillies sur les données peuvent rester difficiles à interpréter : quelle information est contenue dans les générateurs de l'homologie de degré p pour $p \geq 3$? Et que nous apprend la torsion dans ces groupes d'homologie ? Deuxièmement, si les barres courtes ont originellement été considérées comme du bruit, de récents travaux montrent qu'elles contiennent collectivement une grande quantité d'information sur la structure géométrique des données. Par exemple, les barres courtes contiennent de l'information sur la courbure des données (Bubenik, Hull et al., 2020), ou sur la distribution des points d'un nuage de points (cf. par exemple Obayashi, Hiraoka et Kimura 2018 et le chapitre 9 de la présente thèse). Cependant, il reste difficile d'extraire ces informations des code-barres sous une forme lisible et univoque.

Persistance multiparamétrique. L'*invariant de rang* d'un module de persistance M indexé sur $\mathcal{P}$ a été introduit dans Carlsson et Zomorodian (2009) et enregistre le rang des applications linéaires $M(p) \to M(q)$ pour tout $p \leq q$ dans $\mathcal{P}$. Cet invariant est complet dans le cas uniparamétrique (cf. ibid., Theorem 12 pour le cas discret et Chazal, De Silva et al. 2016, Propositions 2.3 et 2.18 pour un énoncé général). Cependant, ce n'est pas le cas pour les modules multiparamétriques, comme le montre le contre-exemple (4.1). L'invariant de rang est stable par rapport à la distance d'entrelacement sur les modules de persistance. Ceci est prouvé dans une note de Landi (2014) pour un invariant équivalent, le *code-barres fibré*. Ce dernier invariant est défini comme la collection des codes-barres uniparamétrique des restrictions d'un module aux droites affines de pentes positives. La principale limite de cet invariant est son coût de calcul. La complexité temporelle d'un calcul de code-barres fibré est égale au nombre de lignes considérées multiplié par le coût de calcul d'un code-barres de persistance uniparamétrique, un prix élevé en pratique. D'autre part, cet invariant bénéficie de propriétés d'interprétabilité similaires à celles du code-barres uniparamétrique. Le code-barres fibré a par ailleurs récemment été généralisé en la notion de *code-barres projeté* dans Berkouk et Petit (2022). Cet invariant est stable et plus flexible, mais reste coûteux à calculer.

Nous n'évoquons pas ici d'autres invariants des modules de persistance, tels que les images et les paysages de persistance (Adams et al., 2017 ; Bubenik et al., 2015 ; Carrière et Blumberg, 2020 ; Vipond, 2020). Ces invariants ont été introduits afin de vectoriser les codes-barres de persistance et présentent tous deux des propriétés intéressantes au vu de la liste ci-dessus. Une fois encore, l'une de leurs principales limites est leur coût de calcul. Cela apparaîtra clairement dans le chapitre 9 lorsque nous comparerons ces invariants aux invariants basés sur la caractéristique d'Euler.

Invariants basés sur la caractéristique d'Euler. En calculant la caractéristique d'Euler de chaque espace d'une famille de complexes simpliciaux à un paramètre, on obtient un descripteur fonctionnel de la filtration appelé *courbe d'Euler* $\chi_{\mathcal{F}} : \mathcal{P} \to \mathbb{Z}$. Bien entendu, ce descripteur se généralise aux filtrations multiparamétrique, et est appelé alors *surface d'Euler* (Beltramo, Skraba et al., 2022) dans le cas biparamétrique et *profils d'Euler* (Dłotko et Gurnari, 2022) dans le cas général. Une formule classique garantit que la caractéristique d'Euler d'un complexe simplicial s'exprime comme une somme alternée du nombre de simplexes en chaque dimension. De façon similaire, lorsque la filtration $\mathcal{F}$

est la filtration des sous-niveaux d'une fonction de Morse $f : Y \to \mathbb{R}$, la caractéristique d'Euler peut être exprimée comme une somme alternée des valeurs critiques de f. Dans les deux cas, le calcul de la caractéristique d'Euler permet d'éviter le calcul de l'homologie et la construction même des modules de persistance associés. En ce qui concerne les propriétés de ces invariants, on peut dresser le tableau suivant :

(i) Les descripteurs basés sur la caractéristique d'Euler ne sont pas aussi grossiers qu'il n'y paraît, comme le prouve leur pouvoir prédictif (Amézquita et al., 2022 ; Jiang, Kurtek et Needham, 2020 ; Smith et Zavala, 2021).

(ii) En raison de leur simplicité, ces descripteurs peuvent être calculés en temps linéaire en le nombre total de simplexes d'une filtration simpliciale $\mathcal{F}$ au lieu d'un temps équivalent à la multiplication de matrices pour les codes-barres de persistance. En outre, la localité de la caractéristique d'Euler peut être exploitée pour concevoir des algorithmes très efficaces de calcul des courbes d'Euler, comme dans les travaux de Heiss et H. Wagner (2017) et Wang, H. Wagner et C. Chen (2022).

(iii) Les profils d'Euler sont des applications à valeur entière sur l'espace euclidien et peuvent donc être munis de distances L_p standards. Afin de transférer les résultats de stabilité existants pour les codes-barres (fibrés) aux profils d'Euler, on peut se demander s'il existe des inégalités prouvant un contrôle des distances L_p sur les profils d'Euler par certaines distances entre les modules de persistance gradués dont proviennent ces profils. De telles inégalités sont effectivement connues pour la distance L_1 sur les courbes d'Euler et la métrique de Wasserstein 1 sur les codes-barres (voir Curry, Mukherjee et Turner 2022 ; Dłotko et Gurnari 2022). Cependant, il a récemment été montré qu'il n'existe pas de distance non triviale sur les courbes d'Euler qui soit contrôlée par la distance *bottleneck* sur les codes-barres de persistance (Berkouk, 2022, Proposition 4.25).

(iv) Plusieurs auteurs ont travaillé à une interprétation des courbes et des surfaces d'Euler (Beltramo, Skraba et al., 2022 ; Smith et Zavala, 2021), et plus généralement à la construction d'outils interprétables et fondés sur la caractéristique d'Euler (Tang et al., 2022). Cependant, l'interprétabilité de ces descripteurs mérite encore investigation. Bien qu'étroitement lié à l'homologie persistante, le signal est très différent à interpréter. En particulier, la dichotomie classique signal/bruit utilisée pour les codes-barres de persistance n'a plus de sens : une succession de classes d'homologie à courte durée de vie et une seule classe à longue durée de vie peuvent toutes deux générer la même courbe d'Euler.

Calcul intégral d'Euler. Les invariants des modules de persistance basés sur la caractéristique d'Euler se formulent naturellement dans le langage du *calcul intégral d'Euler*. Cette théorie de l'intégration par rapport à la caractéristique d'Euler a originellement été développé par Viro (1988) et Schapira (1989, 1991) indépendamment. Les fonctions mesurables pour cette mesure sont les fonctions dites *constructibles*, c'est-à-dire les fonctions constantes par morceaux sur des partitions de l'espace en sous-ensembles géométriques simples, généralement supposés définissables dans une structure o-minimale fixée.

Schapira (1989, 1991) développe le calcul intégral d'Euler à l'aide du formalisme des six foncteurs de la théorie des faisceaux et du lien entre les fonctions constructibles et les faisceaux constructibles appelé correspondance *faisceaux-fonctions* (Kashiwara, 1985 ; Kashiwara et Schapira, 1990). Ces outils s'inscrivent donc naturellement dans l'approche faisceautique de la persistance multiparamétrique initiée par Curry (2014) et Kashiwara et Schapira (2018).

Le calcul d'Euler a conduit à d'importantes avancées en analyse topologique de données (cf. Curry, Ghrist et Robinson 2012), notamment par l'introduction de transformées intégrales topologiques (Baryshnikov, Ghrist et Lipsky, 2011 ; Curry, Mukherjee et Turner, 2022 ; Ghrist, Levanger et Mai, 2018 ; Turner, Mukherjee et Boyer, 2014a). La plus importante d'entre elles est la *transformée de Radon* introduite par Schapira (1995), définie comme suit. La transformée de Radon de la fonction indicatrice $\mathbf{1}_Z$ d'un sous-ensemble définissable $Z \subseteq \mathbb{R}^m$ associe à tout hyperplan affine $H \subset \mathbb{R}^m$ la caractéristique d'Euler de l'intersection $Z \cap H$. Elle est ensuite étendue par linéarité à toutes les fonctions constructibles. Le résultat d'inversion de Schapira (ibid.) garantit que cette transformée est inversible, à une constante près lorsque m est pair. Curry, Mukherjee et Turner (2022) et Ghrist, Levanger et Mai (2018) ont montré que ce résultat fournit une réponse positive à un important problème inverse : deux sous-ensembles définissables de $\mathbb{R}^m$ ayant la même homologie persistante en tout degré et pour les filtrations de sous-niveaux de toutes les formes linéaires sont-ils égaux ? Ces techniques ont été appliquées en analyse de formes, par exemple dans la prédiction de résultat clinique des cancers du cerveau (Crawford et al., 2020), dans l'analyse des graines d'orge (Amézquita et al., 2022) ou dans l'étude des variations morphologiques à travers les espèces de primates (Tang et al., 2022).

L'aspect le plus problématique du calcul d'Euler pour les applications est son instabilité vis-à-vis des approximations numériques : des erreurs peuvent être (et seront certainement) commises lors du calcul de l'intégrale d'Euler d'une fonction constructible, quelle que soit la finesse de l'échantillonnage de son domaine (Curry, Ghrist et Robinson, 2012, Section 16). Dans leurs expériences de prédiction, Crawford et al. (2020) effectuent un lissage des transformées intégrales topologiques pour produire des descripteurs de formes plus performants, mais aucun résultat théorique ne permet d'étayer ces résultats expérimentaux pour l'instant. Ces lissages sont étroitement liés aux transformées de *Bessel* et de *Fourier* introduites par Ghrist et Robinson (2011), ainsi qu'à la *caractéristique d'Euler des codes-barres* introduite par Bobrowski et Borman (2012).

Contributions. Dans la deuxième partie de cette thèse, nous nous intéressons aux invariants basés sur la caractéristique d'Euler en analyse topologique de données.

Nous introduisons une définition de *transformées hybrides* pour les fonctions constructibles qui généralise les lissages des transformées topologiques mentionnés ci-dessus. Il s'agit de transformées intégrales combinant l'intégration de Lebesgue et le calcul intégral d'Euler. L'intégration de Lebesgue donne accès à des noyaux bien étudiés et à des résultats de régularité, tandis que le calcul d'Euler véhicule de l'information topologique et permet la compatibilité des transformées avec les opérations sur les fonctions constructibles. Nous menons une étude systématique de ces transformées et en introduisons deux nouvelles : les transformées d'Euler-Fourier et d'Euler-Laplace. Nous montrons que la première a un inverse gauche et que la seconde fournit une généralisation satisfaisante

de *magnitude persistante* (Govc et Hepworth, 2021) aux faisceaux constructibles, et donc en particulier aux modules de persistance multiparamétriques. Enfin, nous prouvons des formules d'indices pour une large classe de transformées hybrides dans le contexte de la persistance multiparamétrique des ensembles de (sous-)niveaux, généralisant les formules existantes démontrées par Bobrowski et Borman (2012), Ghrist et Robinson (2011) et Govc et Hepworth (2021). En particulier, cela nous permet de calculer l'espérance de ces transformées dans le cas de filtrations de sous-niveaux de champs gaussiens aléatoires.

Dans un travail en collaboration avec Olympio Hacquard (2023), nous montrons que les profils d'Euler et leurs transformées hybrides sont des descripteurs topologiques riches et efficacement calculables des données. Pour commencer, nous prouvons des résultats de stabilité pour ces descripteurs ainsi que des garanties asymptotiques dans des contextes aléatoires. Ensuite, nous démontrons que les profils d'Euler atteignent des performances très compétitives à un faible coût de calcul lorsqu'ils sont associés à un classificateur puissant tel qu'une forêt aléatoire. Cependant, en raison de leur simplicité, ces descripteurs ne parviennent pas à séparer linéairement les différentes classes ou à être compétitifs dans des tâches non supervisées. Inspirés par l'analyse du signal, nous remédions à ces problèmes en étudiant les transformées hybrides des profils d'Euler. Nous montrons que ces transformées hybrides présentent des performances remarquables dans les tâches de classification linéaire et dans les contextes non supervisés. D'un point de vue qualitatif, nous fournissons également de nombreuses heuristiques sur les informations topologiques et géométriques captées par les profils d'Euler et leurs transformées hybrides.

Plan détaillé et résumé des contributions

Nous présentons ici nos contributions de manière plus détaillée et exposons la structure de la thèse. Nous renvoyons aux chapitres correspondants pour les définitions précises de nos termes.

Partie I : Décompositions des modules de persistance

Dans la première partie de cette thèse, nous étudions les caractérisations locales de décomposabilité des modules de persistance biparamétriques. Tous les résultats de cette partie sont basés sur des travaux en collaboration avec Magnus Bakke Botnan et Steve Oudot (2022, 2023). Dans ce résumé, nous supposons toujours que les modules de persistance sont point-à-point de dimension finie.

Chapitre 1 : Caractérisations locales

Ce chapitre sert d'introduction à la première partie de la thèse. Nous commençons par définir les modules de persistance multiparamétrique, les décompositions en intervalles et d'autres notions nécessaires (section 1.1). Nous donnons ensuite un énoncé précis de notre question principale : l'existence de caractérisations locales de la décomposabilité d'intervalles (section 1.2). Après avoir passé en revue les réponses existantes à cette question (section 1.3), nous exposons nos contributions (section 1.4).

Chapitre 2 : Décomposition en rectangles

Ce chapitre est consacré à la preuve du résultat principal de la partie I :

Théorème (Théorème 1.7)**.** *Soit X et Y deux ensembles totalement ordonnés dont tous les intervalles admettent des sous-ensembles initiaux dénombrables. Un module de persistance M sur $X \times Y$ est décomposable en rectangles si, et seulement si, la restriction de M à tout carré $\{x_1 < x_1'\} \times \{x_2 < x_2'\} \subseteq X \times Y$ est décomposable en rectangles.*

Nous commençons par fournir une courte preuve de ce résultat lorsque X et Y sont finis en suivant Botnan, Lebovici et Oudot (2022) (section 2.1). La preuve du cas général utilise le cas fini et s'étend quant à elle sur plus de vingt pages. Elle suit notre travail avec Magnus Bakke Botnan et Steve Oudot (2023) et est développée dans les sections 2.2 à 2.5. La preuve repose dans les deux cas sur une formulation algébrique clé de la décomposabilité en rectangles des restrictions aux carrés appelée *exactitude faible* (Définition 2.1).

Les restrictions aux carrés de modules décomposables en rectangles sont trivialement décomposables en rectangle. Il s'agit donc de prouver la réciproque. Pour cela, la preuve dans le cas fini utilise une approche simple basée sur des calculs de rang des morphismes du module. Dans le cas général, nous suivons la même approche que Cochoy et Oudot (2020) pour la décomopsabilité en blocs, en utilisant ce que l'on appelle les *filtrations fonctorielles* (section 2.2), avec des ajustements majeurs à des étapes clés en raison de notre condition locale plus faible. En résumé, nous commençons par définir, pour chaque rectangle $R \in X \times Y$, un sous-module M_R de M appelé *filtrat rectangle* de M associé à R (section 2.3). Ce sous-module est construit de telle sorte que $M_{R,t}$ contienne précisément les éléments de M_t qui vivent précisément durant le sous-ensemble R des paramètres. En particulier, M_R est isomorphe à une somme directe finie de copies du module rectangle $\mathbf{k}_R$. Nous prouvons ensuite que ces filtrats sont en somme directe (section 2.4) et que cette somme engendre tout M (section 2.5).

Chapitre 3 : Sur l'existence d'autres caractérisations locales

Dans ce chapitre, nous apportons des réponses négatives à notre question principale. Les résultats du chapitre précédent garantissent que la décomposabilité en rectangles d'un module de persistance donné peut être vérifiée localement en considérant les restrictions à tous les carrés. La question suivante est donc naturelle : dans quelle mesure la décomposabilité en intervalles peut-elle être déterminée localement en autorisant des intervalles de forme plus générale que les rectangles ?

Nous fournissons deux résultats négatifs, établis en collaboration avec Magnus Bakke Botnan et Steve Oudot dans le cas fini (2022), puis en toute généralité (2023). Tout d'abord, nous montrons que la décomposabilité en intervalles ne peut pas être caractérisée localement, même par des tests sur toutes les sous-grilles strictes :

Theorem (Théorème 3.1)**.** *Soit X et Y deux ensembles totalement ordonnés tels que $|X| \geq 3$ et $|Y| \geq 3$ et $2 \leq m \leq \min(|X|, |Y|)$ un entier. Il existe un module de persistance M sur $X \times Y$ qui n'est pas décomposable en intervalles mais dont les restrictions $M_{|Q}$ à toutes les sous-grilles finies $Q \subset X \times Y$ de côtés de longueur au plus m le sont.*

A la lumière des résultats précédents, il est naturel de se demander s'il existe une classe d'intervalles plus généraux que les rectangles, pour lesquels une caractérisation locale sur les carrés est possible. Il s'avère que ce n'est pas le cas, comme le montre le théorème 3.5.

Chapitre 4 : Applications à l'analyse topologique de données

Dans la section 4.1, nous fournissons des algorithmes permettant de tester si un module de persistance dérivant d'une filtration simpliciale à deux paramètres est décomposable en rectangles, et calculant sa décomposition lorsqu'elle existe.

Dans un premier temps, nous prouvons que l'invariant de rang est complet sur la classe des modules biparamétriques décomposables en rectangles (théorème 4.4). À cette fin, nous généralisons la formule d'inclusion-exclusion exprimant le code-barres de persistance en termes d'invariant de rang (Cohen-Steiner, Edelsbrunner et Harer, 2007) au cadre biparamétrique. Notons que notre résultat découle indirectement d'une formule similaire pour une généralisation de l'invariant de rang (Kim et Mémoli, 2021), mais que nous fournissons un énoncé explicite ainsi qu'une preuve simple et directe.

Nous montrons ensuite que l'invariant de rang d'un module produit par une filtration simpliciale biparamétrique ayant n simplexes au total peut être calculé en temps $O(n^4)$ (théorème 4.5). Ce résultat n'est pas nouveau en soi, cependant, combiné avec notre formule d'inclusion-exclusion, il donne un algorithme en temps $O(n^4)$ pour calculer la décomposition d'un module de persistance biparamétrique décomposable en rectangles (corollaire 4.6). Au moment de sa publication, notre résultat représentait une amélioration par rapport à la simple application d'un algorithme de pointe de décomposition des modules de persistance biparamétriques généraux, qui prendrait $O(n^{2\omega+1})$ temps où $2 \leq \omega < 2.373$ est l'exposant pour la multiplication matricielle (Dey et Xin, 2022). Dans un très récent article, Clause et al. (2023) ont amélioré notre résultat en décrivant un algorithme de calcul de l'invariant de rang des modules de persistance biparamétriques induits par une filtration simpliciale ayant une complexité temporelle $O(n^3)$.

Enfin, nous tirons parti de notre résultat de décomposition en rectangles pour obtenir un algorithme vérifiant la décomposabilité en rectangles des modules de persistance induits en homologie à partir de filtrations simpliciales biparamétriques avec au plus n simplexes en temps $O(n^{2+\omega})$ (théorème 4.7). Une fois de plus, il s'agit d'une amélioration par rapport à une méthode vérifiant individuellement les indécomposables extraits à l'aide d'un algorithme de décomposition des modules de persistance biparamétriques généraux.

Dans la section 4.2, nous appliquons notre résultat de décomposition en rectangles aux modules apparaissant en persistance des niveaux (Bendich et al., 2013). Tout d'abord, nous exposons les constructions de la pyramide et de la bande de Mayer-Vietoris impliquées dans la persistance des ensembles de niveaux telles qu'elles sont développées par Bauer, Botnan et Fluhr (2021). Ensuite, nous prouvons notre résultat principal, un théorème de décomposition pour les modules de persistance indexés sur la bande (théorème 4.11). En corollaire, nous prouvons une généralisation du théorème de la base pyramidale de Bendich et al. (2013).

Partie II : Invariants basés sur la caractéristique d'Euler

Dans la deuxième partie de cette thèse, nous introduisons de nouveaux invariants pour les modules de persistance multiparamétriques en utilisant des transformées intégrales de fonctions constructibles.

Chapitre 5 : Calcul intégral d'Euler

Ce chapitre sert d'introduction à la partie II. Nous définissons nos notations et rappelons quelques définitions et résultats de base sur le calcul d'Euler en suivant l'approche de Schapira (1991) (section 5.1). Nous présentons ensuite plusieurs sous-groupes spécifiques du groupe des fonctions constructibles utilisés tout au long de la partie II (section 5.2). Parmi eux, le sous-groupe $\mathrm{CF}_{\gamma,\mathrm{c}}(\mathbb{R}^m)$ des fonctions γ-constructibles à support compact — dont les strates sont γ-localement fermées et relativement compactes —, qui inclut les fonctions constructibles associées aux modules de persistance multiparamétriques à support compact. Puis, nous définissons la *transformée de Radon* introduite par Schapira (1995) et exposons le résultat d'inversion de Schapira (section 5.3). De plus, nous prouvons que la transformée de Radon d'une fonction γ-constructible à support compact est nulle pour les hyperplans affines dont la conormale se trouve dans le complémentaire du polaire de γ. Cette caractérisation du support de la transformée sera essentielle pour prouver la formule de reconstruction de la transformée d'Euler-Fourier dans le chapitre 6. Nous terminons ce chapitre en expliquant comment associer une fonction constructible à un module de persistance multiparamétrique gradué (section 5.4).

Chapitre 6 : Transformées hybrides de fonctions constructibles

Dans ce chapitre, nous introduisons la définition des transformées hybrides et menons une étude théorique systématique de ces objets. Après avoir défini ces transformées, nous en donnons deux nouveaux exemples importants : la transformée d'Euler-Laplace et la transformée d'Euler-Fourier (section 6.1). Nous montrons comment la transformée d'Euler-Laplace permet de définir une généralisation de la magnitude persistante (Govc et Hepworth, 2021) pour les faisceaux constructibles. En outre, nous fournissons plusieurs exemples illustrés pour comparer ces transformées à leurs analogues classiques. Nous terminons la section en présentant une méthode pour calculer numériquement nos transformées sur les fonctions constructibles linéaires par morceaux, c'est-à-dire celles dont les strates sont polyédrales, et nous fournissons une bibliothèque `Python` et `C++` librement disponible sur GitHub. Nous prouvons ensuite que les transformées hybrides sont continues lorsqu'elles sont restreintes à l'ensemble des fonctions constructibles linéaires par morceaux, et même C^{p+1} à l'intérieur de cônes partitionnant leur domaine lorsque leur noyau est C^p (section 6.2). Puis, nous montrons que les transformées hybrides sont compatibles avec les opérations sur les fonctions constructibles telles que le poussé en avant et l'opérateur de dualité (section 6.3). En particulier, nous montrons que les transformées d'Euler-Laplace et d'Euler-Fourier envoient les produits de convolution (constructible) sur des produits classiques sous certaines hypothèses. Enfin, nous établissons une formule de reconstruction pour la transformée d'Euler-Fourier des fonctions γ-constructibles à support compact (section 6.4). La preuve de ce dernier fait est la suivante. En utilisant l'inverse de la transformée de Fourier classique, on peut récupérer, à

partir de la connaissance de la transformée d'Euler-Fourier, les valeurs de la transformée de Radon sur l'ensemble des hyperplans affines dont la conormale est dans le cône γ. Notre résultat sur le support de la transformée de Radon des fonctions γ-constructibles du chapitre 5 assure alors que la transformée de Radon est en fait entièrement reconstruite, puisqu'elle est nulle pour tout autre hyperplan. Il ne reste donc plus qu'à inverser la transformée de Radon en utilisant la formule de Schapira (1995).

Chapitre 7 : Formules d'indices

Dans ce chapitre, nous commençons par définir les *fonctions constructibles des niveaux* et des *sous-niveaux* associées à une application sous-analytique continue $f : M \to \mathbb{V}$ et à un cône $\gamma \subset \mathbb{V}$, où $\mathbb{V}$ est un espace vectoriel réel de dimension finie et M est une variété analytique réelle (section 7.1). Il s'agit simplement des fonctions constructibles associées aux faisceaux de cohomologie persistante introduits par Kashiwara et Schapira (2018). Nous montrons que les transformées hybrides de fonctions constructibles des (sous-)niveaux associées à des applications à valeurs vectorielles peuvent être exprimées comme des transformées hybrides de fonctions constructibles des (sous-)niveaux associées à des applications à valeurs réelles (corollaire 7.10). La preuve repose sur une expression de la fonction constructible des sous-niveaux comme une convolution de la fonction constructible des niveaux avec la fonction indicatrice de l'antipodal du cône γ. Cette réduction conduit à la définition de *transformées des sous-niveaux*, qui est la forme simplifiée que prennent les transformées hybrides dans le cadre de la persistance des sous-niveaux.

Nous rappelons ensuite la définition d'*intégrale d'Euler continue* due à Bobrowski et Borman (2012) qui étend le calcul intégral d'Euler à la classe plus large des fonctions *tames* (section 7.2). Cette classe contient les fonctions continues sous-analytiques sur les variétés analytiques réelles compactes. Nous prouvons ensuite nos principaux résultats appelés *formules d'indice* qui expriment les transformées des sous-niveaux comme des transformées intégrales d'Euler continues. Cela nous permet de prouver une formule d'espérance pour la transformée d'Euler-Bessel des fonctions constructibles des niveaux de champs aléatoires gaussiens en utilisant la formule de Bobrowski et Borman (ibid.) calculant l'espérance des intégrales d'Euler continues de ces champs (section 7.3).

Chapitre 8 : Profils d'Euler

Dans les deux derniers chapitres de la thèse, nous étudions les fonctions constructibles induites par le calcul de la caractéristique d'Euler des espaces de filtrations simpliciales, ainsi que leurs transformées hybrides. Dans le présent chapitre, nous fournissons plusieurs garanties théoriques pour ces descripteurs. Le travail présenté ici a été réalisé en collaboration avec Olympio Hacquard (2023). Nous commençons par introduire quelques définitions nécessaires (section 8.1). Ensuite, nous prouvons des propriétés de stabilité qui attestent la robustesse de nos outils par rapport au bruit dans les données (section 8.2). Nous généralisons légèrement les résultats existants sur les profils d'Euler et en déduisons de nouveaux résultats pour les transformations hybrides dans un cadre unifié en utilisant la *distance de Wasserstein* 1 *signée* introduite par Oudot et Scoccola (2021).

Enfin, nous établissons la convergence ponctuelle des transformées hybrides asso-

ciées aux filtrations simpliciales de nuages de points aléatoires, ainsi que leur normalité asymptotique pour une fonction de filtration spécifique (section 8.3). Nous démontrons également une loi des grands nombres dans le cas multiparamétrique. Ces résultats statistiques sont des conséquences faciles de résultats existants sur les *diagrammes de persistance*. Cependant, le résultat dans le cadre multiparamétrique repose de manière cruciale sur notre résultat exprimant les transformées hybrides multiparamétriques comme des transformées uniparamétrique dans le contexte de la persistance des sous-niveaux (Corollary 7.10).

Chapitre 9 : Etude expérimentale

Dans ce chapitre, nous montrons que les profils d'Euler et leurs transformées hybrides sont des descripteurs topologiques informatifs et efficacement calculables des filtrations simpliciales. Une fois encore, les résultats de ce chapitre sont le fruit d'un travail conjoint avec Olympio Hacquard (2023). Tout d'abord, nous expliquons nos algorithmes et fournissons une bibliothèque `Python` disponible sur GitHub. De plus, nous fournissons des heuristiques sur la façon de choisir le noyau des transformées hybrides en pratique (section 9.1). Deuxièmement, nous menons une étude empirique sur le type d'informations topologiques et géométriques que les profils d'Euler et leurs transformées peuvent extraire des données (section 9.2). Troisièmement, nous réalisons des expériences qualitatives ainsi que quantitatives : une tâche de régression supervisée et des tâches de classification supervisées et non supervisées sur des jeux de données de graphes et de nuages de points. Nous démontrons que les profils d'Euler atteignent une précision de pointe dans les tâches de classification et de régression supervisées en conservant un coût de calcul très faible lorsqu'ils sont couplés à un classificateur de type forêt aléatoire ou *gradient boosting*. En outre, nous démontrons que les transformées hybrides permettent de compresser efficacement les informations topologiques et géométriques des signaux présents dans les profils d'Euler. Par conséquent, elles sont plus performantes que ces derniers descripteurs dans les tâches de classification non supervisées et dans les tâches supervisées en présence d'un classificateur linéaire. Nous exposons également un exemple de l'information extraite par ces descripteurs sur un jeu de données réelles.

Publications

- Magnus Bakke Botnan, Vadim Lebovici et Steve Oudot (2022). "On Rectangle-Decomposable 2-Parameter Persistence Modules". In : *Discrete and Computational Geometry*. Initially accepted in : 36th International Symposium on Computational Geometry (SoCG 2020), p. 1-24
- Magnus Bakke Botnan, Vadim Lebovici et Steve Oudot (2023). "Local characterizations for decomposability of 2-parameter persistence modules". In : *Algebras and Representation Theory*, p. 1-44
- Vadim Lebovici (2022). "Hybrid transforms of constructible functions". In : *Foundations of Computational Mathematics*, p. 1-47
- Olympio Hacquard et Vadim Lebovici (2023). "Euler characteristic tools for topological data analysis". Submitted

Code

- https://github.com/HugoPasse/Transforms-of-cubical-complexes
- https://github.com/vadimlebovici/eulearning

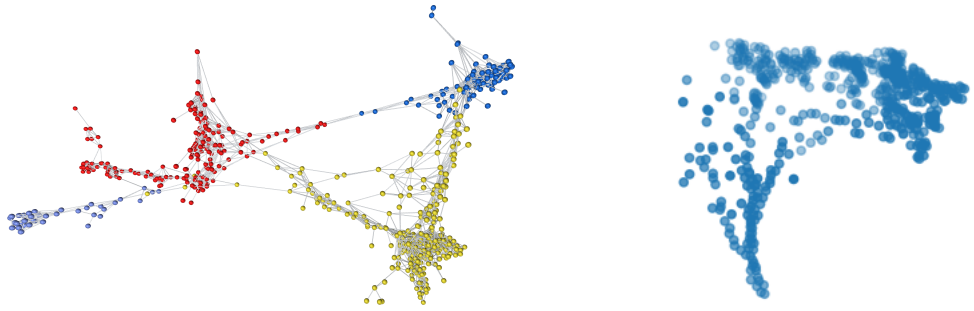
Introduction (English version)

This thesis aims at contributing to the field of *topological data analysis*, and more precisely to its branch called *multi-parameter persistence*. The main goal of topological data analysis is to extract geometric and topological information from data to perform a wide variety of tasks, such as regression, classification or visualization. In this introduction, we explain what we mean by data and by geometric and topological information. Then we explain what answers topological data analysis offers to this problem, and how our contributions fit within this broad objective.

Framework. In topological data analysis, data is usually understood as some metric space (X, d). For instance, the space X can be a finite subset of some Euclidean space endowed with the induced metric, called a *point cloud*, or a graph endowed with the shortest-path distance (Figure 5). In the case of a graph, topological information is usually interpreted as the number of connected components or the number of cycles in the graph. Going further, one can build a simplicial complex from a graph, such as the *clique complex* for instance, encoding higher-order connectivity properties in a topological manner. In the case of a point cloud, the coarsest topological information is the number of different meaningful clusters formed by these points. The field of data analysis devoted to the extraction of such an information is called *clustering*. Once this is known, one can go one step further: what do these clusters look like? Suppose for instance that the point cloud lies near some unknown manifold in $\mathbb{R}^n$. Extracting topological and geometric information on the point cloud then usually means unravelling the topology and the geometry of the underlying manifold.

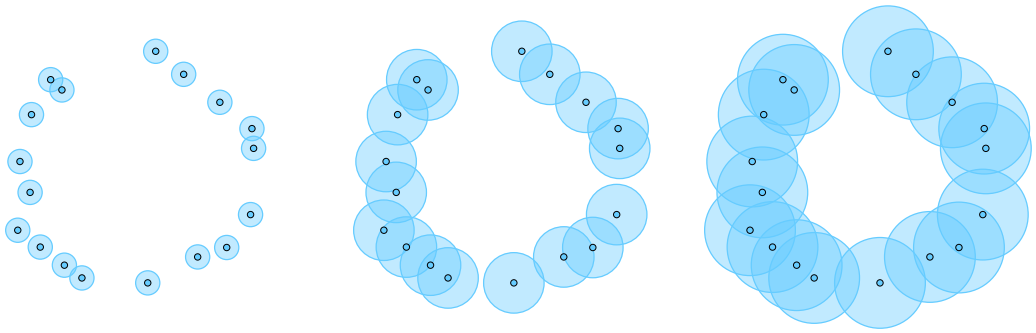
One-parameter persistence. The main tool of topological data analysis to uncover the underlying geometric and topological properties of a given dataset is *persistent homology*. The idea is to construct a family $\mathcal{F}$ of topological spaces parametrized by a totally ordered set $(\mathcal{P}, \leq)$, which is increasing with respect to inclusion. Such a family $\mathcal{F}$ is called a *one-parameter filtration*; see Figure 6 for an illustration. The idea of persistent homology is to keep track of homological features across inclusions as the indexing parameter varies. Typically, one can consider the family of sublevel sets of a function $f : Y \to \mathcal{P}$ for some topological space Y. When $X \subseteq \mathbb{R}^d$ is a point cloud, a famous example called *Čech filtration* is given by the real-valued function $f = d(X, \cdot)$ that associates to any point of $Y = \mathbb{R}^d$ its distance to X.

Applying homology with coefficients in a field $\mathbf{k}$ to the filtration $\mathcal{F}$ yields an algebraic object $H_k(\mathcal{F})$ called a (one-parameter) *persistence module*, that is, a functor from $(\mathcal{P}, \leq)$ to the category $\mathsf{Vec}_{\mathbf{k}}$ of $\mathbf{k}$-vector spaces. In plain words, a persistence mod-



(a) (b)

Figure 5: Examples of metric spaces (X, d). Figure 5a represents the graph of gene expression across mouse tissues from Su et al. (2004) according to the gene correlation network analysis approach developed in Freeman et al. (2022). Figure 5b represents a point cloud from De Deuge et al. (2013) created by a LIDAR scan of a tree.



(a) $d(X, \cdot) \leq 0.4$ (b) $d(X, \cdot) \leq 1.5$ (c) $d(X, \cdot) \leq 2.1$

Figure 6: Čech filtration of a point cloud $X \subseteq \mathbb{R}^2$. The sublevel sets of the function $d(X, \cdot)$ are drawn in light blue.

ule $M : (\mathcal{P}, \leq) \to \mathsf{Vec}_{\mathbf{k}}$ is a collection of vector spaces $M(p)$ for each $p \in \mathcal{P}$, along with linear maps $M(p) \to M(q)$ for each pair of comparable elements $p \leq q$ in $\mathcal{P}$. Moreover, these linear maps must satisfy certain compatibility conditions. Under some tameness assumption, results of Gabriel (1972) from quiver representation theory ensure that one can make a coherent choice of basis for each vector space of $H_k(\mathcal{F})$, in the sense that each basis element appears in these vector spaces at some $a \in \mathcal{P}$ called its *birth*, and disappears at some $b > a$ called its *death*. The interval $[a, b)$ is called the *bar* of the corresponding basis element. The module $H_k(\mathcal{F})$ is then entirely described by the multiset of bars $[a, b)$ for each homological feature appearing in the filtration (see Figure 7). More generally, the isomorphism classes of persistence modules with finite-dimensional vector spaces are in one-to-one correspondence with multisets of intervals in $\mathcal{P}$, called *persistence barcodes* (Botnan and Crawley-Boevey, 2020; Crawley-Boevey, 2015). These ideas find their roots in the early nineties with the work of Barannikov (1994) and have been intensively developed since the seminal work of Edelsbrunner, Letscher, and Zomorodian (2000). See Edelsbrunner and Harer (2022) and Oudot (2015) for an introduction.

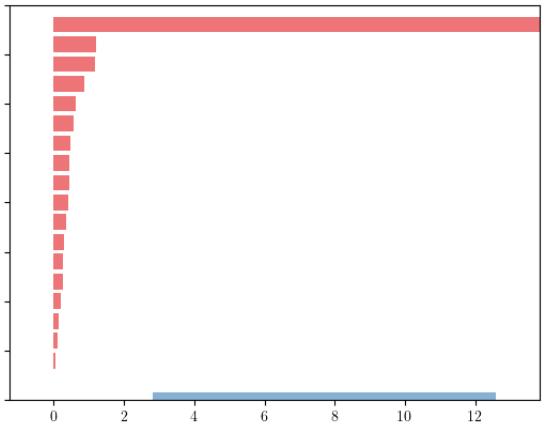


Figure 7: Barcode of the Čech filtration built on the point cloud X presented in Figure 6. Red bars correspond to the barcode of $H_0(\mathcal{F})$ and blue bars to the barcode of $H_1(\mathcal{F})$. The large blue bar records the information that X lies close to a circle.

One remarkable property of these persistence barcodes is that they are robust with respect to (unavoidable) noise in the data. Specifically, the space of persistence barcodes can be equipped with the so-called *bottleneck distance* (Cohen-Steiner, Edelsbrunner, and Harer, 2007) and *p-Wasserstein distances* (Cohen-Steiner, Edelsbrunner, Harer, and Mileyko, 2010) and both metrics satisfy stability properties with respect to some metric on the data (Chazal, De Silva, et al., 2016; Cohen-Steiner, Edelsbrunner, and Harer, 2007; Cohen-Steiner, Edelsbrunner, Harer, and Mileyko, 2010). The bottleneck distance satisfies one with respect to the L_∞ metric between functions $f : Y \to \mathbb{R}$ whose sublevel-sets are considered. This result specifies the type of noise with respect to which persistence barcodes are stable (a sup norm), and is probably one of the most important results of persistence theory. In particular, it makes it a relevant choice for deriving consistent estimators for statistical analysis; see for instance (Bobrowski, Mukherjee, and Taylor, 2017; Chazal, Glisse, et al., 2015).

As a matter of fact, persistence barcodes have been successfully used in statistics and have found a wide range of scientific applications, such as in health sciences (Aukerman et al., 2021; Fernández and Mateos, 2022; Rieck et al., 2020), in biology (Ichinomiya, Obayashi, and Hiraoka, 2020; Rabadán and Blumberg, 2019) or in material sciences (Hiraoka, Nakamura, et al., 2016; Lee et al., 2017). Furthermore, these tools have also found many applications in other parts of mathematics, such as symplectic geometry; see Polterovich, Rosen, et al. (2020) and Polterovich and Shelukhin (2016).

Multi-parameter persistence. In many practical scenarios, it is relevant to build multi-parameter filtrations of topological spaces instead of one-parameter ones. For instance, the filtration $\mathcal{F}$ can be indexed by $\mathcal{P} = \mathbb{Z}^m$ or $\mathcal{P} = \mathbb{R}^m$ endowed with the coordinatewise order. One way to obtain such a filtration is to consider sublevel sets of vector-valued functions $f : Y \to \mathbb{R}^m$. As a consequence, multi-parameter persistence allows one to cope with outliers in a point cloud $X \subseteq \mathbb{R}^d$ by filtering the space with respect to $f = (f_1, f_2) : \mathbb{R}^d \to \mathbb{R}^2$ where f_1 is the distance to X and f_2 is some esti-

mated local density; see Vipond et al. (2021) for instance. Another example comes from digital images of breast-cancer tissue samples equipped with pixel intensities of both immune and cancer cells (Carrière and Blumberg, 2020). Applying homology to such a multi-parameter filtration now yields an object called a *multi-parameter persistence module* $M : (\mathcal{P}, \leq) \to \mathsf{Vec}_{\mathbf{k}}$. The theory of multi-parameter persistence originated in the form of size homotopy groups in Frosini and Mulazzani (1999) and has first been studied in the homological context by Carlsson and Zomorodian (2009). Following the same line of thought as in the one-parameter case, it is natural to wonder whether isomorphism classes of such objects can be parametrized by multisets of simple geometric regions of the parameter space akin to bars of persistence barcodes.

Born to be wild. Quiver representation theory leaves no hope of answering this question positively. Already when the poset $\mathcal{P}$ of parameters is $[\![1, 2]\!] \times [\![1, 6]\!]$ endowed with the coordinatewise order, the situation is as complex as it could be. Namely, this poset is of *wild representation type*. Hence, a full classification of isomorphism classes of persistence modules indexed over this poset would imply such a classification for persistence modules indexed over any other finite poset. It is thus considered a hopeless task to build a *complete*—characterizing the isomorphism type of the persistent module—descriptor of multi-parameter persistence modules that would be both computable and easy to work with in practice. We refer to Botnan and Lesnick (2022, Section 8) for more details on the representation theory of multi-parameter persistence.

Challenge. One of the main difficulties of multi-parameter persistence is to extract rich and stable information from persistence modules despite the wildness of the parametrizing poset. One approach to do so is to identify subclasses of multi-parameter persistence modules for which an analogue of the barcode does exist. Another approach is to build stable and informative, albeit incomplete, invariants of persistence modules. The work conducted in this Ph.D. thesis contributes to both of these approaches.

I. Direct-sum decompositions of persistence modules

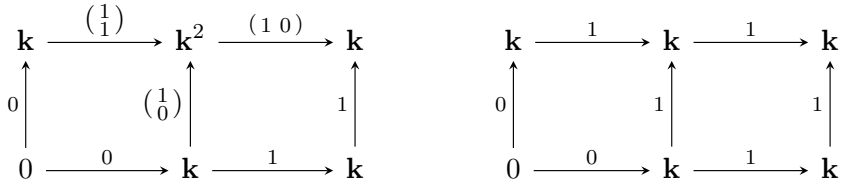
In this section, we give an overview of the decomposition theory of multi-parameter persistence modules. Our goal is not to be exhaustive but to put our contributions into context. We refer to Botnan and Lesnick (ibid.) for a recent picture.

Decomposition theorems. Work by Botnan and Crawley-Boevey (2020) shows that pointwise finite-dimensional multi-parameter persistence modules—i.e., valued in the category of finite-dimensional $\mathbf{k}$-vector spaces—always decompose as direct sums of indecomposable modules. Moreover, these indecomposables have local endomorphism ring thus Azumaya's theorem (1950) ensures that this decomposition is essentially unique.

In the one-parameter setting, these indecomposables can only be *interval modules*, which by definition are indicator representations of intervals of $\mathcal{P}$; see Crawley-Boevey (2015). As a consequence, they are fully characterized by their *support*, that is, the set of parameters $p \in \mathcal{P}$ such that the vector space $M(p)$ is non-trivial. An isomorphism class of one-parameter persistence modules is thus entirely described by the multiset of

intervals appearing as supports of indecomposables in the direct-sum decomposition of any representative of the class.

The situation is more complex in the two-parameter setting, as general indecomposables are not fully characterized by their support, as shown by the following example of two indecomposable modules with same support; see (1.4):



Yet, one can still define interval modules as indicator representations of intervals of $\mathcal{P}$, where by "interval" we mean any connected convex subset of $\mathcal{P}$ in the sense of order theory (Figure 8). The collection of supports of the interval summands appearing in a direct-sum decomposition of a persistence module can then be used as a descriptor of this module. This descriptor is purely geometric by nature, therefore easy to interpret and efficient to encode on a computer.

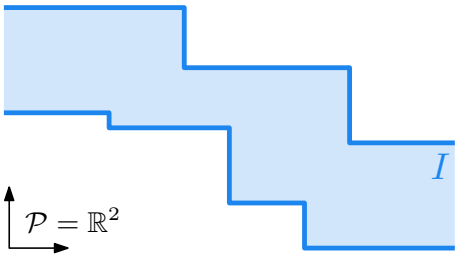


Figure 8: An interval I of the poset $\mathcal{P} = \mathbb{R}^2$.

In practice, one would like to be able to determine whether a given persistence module admits interval summands in its direct-sum decomposition, and if so, whether it admits only such summands—in which case it is called *interval-decomposable*. The straightforward approach for this consists in decomposing the module and then checking its summands one by one. In the first part of this Ph.D. thesis, we advocate a different approach.

Local characterizations. The method consists in checking *local conditions*, involving only restrictions of persistence modules to certain collections of *test subsets* of $\mathcal{P}$. Provided the considered subsets are small enough, the restrictions will have a simple structure, potentially leading to algorithmic improvements and simplified mathematical analyses.

These local conditions appear for instance in *level-set persistence* (Bendich et al., 2013; Carlsson, De Silva, and D. Morozov, 2009), where one constructs invariants for real-valued functions f on a topological space by looking at the pre-images through f of bounded open intervals. Once properly indexed, the homology groups of these pre-images arrange themselves into a two-parameter persistence module indexed over $\mathcal{P} = \mathbb{R}^2$ that turns out to be interval-decomposable, with summands supported on a special class

of intervals of $\mathbb{R}^2$ called *blocks*—specifically: upper-right or lower-left quadrants, and horizontal or vertical infinite bands. Decomposability into block summands in this setting is a straightforward consequence of the Mayer-Vietoris theorem, once the following local characterization has been established: a two-parameter persistence module decomposes exclusively into block summands if, and only if, all its restrictions to *squares* $\{x_1, x_1'\} \times \{x_2, x_2'\} \subset \mathbb{R}^2$ do. This fact was proven in Cochoy and Oudot (2020) then in Botnan and Crawley-Boevey (2020), and at the time, it provided a cleaner, and more general theory for level-set persistence than the one established previously, which required an extra "Morse-type" condition on the real-valued function f under consideration (Bendich et al., 2013).

Contributions. In the first part of this Ph.D. thesis, we further generalize the two-parameter theory by enlarging the class of intervals of interest to include all the axis-aligned rectangles in $\mathbb{R}^2$. The interval module supported on rectangles are called *rectangle modules*, and an interval-decomposable module whose summands are exclusively rectangle modules is called *rectangle-decomposable*. We prove that a two-parameter persistence module is rectangle-decomposable if, and only if, all its restrictions to squares are. We also prove that rectangle-decomposable modules are, in some precise sense, the largest subclass of interval-decomposable modules that can be characterized locally, at least as far as restrictions to squares are concerned. These results are joint work with Magnus Bakke Botnan and Steve Oudot. They were first obtained on binary products of totally ordered sets assumed to be finite in Botnan, Lebovici, and Oudot (2022) and then extended to those satisfying some mild assumptions (including $\mathcal{P} = \mathbb{R}^2$) in Botnan, Lebovici, and Oudot (2023).

Then, we present two applications of these results to topological data analysis. First, we provide an algorithm to check whether a module induced in homology by a finite two-parameter filtration is rectangle-decomposable, and to compute the direct-sum decomposition if it exists, with a better complexity than state-of-the-art decomposition methods for general two-parameter persistence modules. Our algorithms are backed up by our local characterization: it suffices to check rectangle-decomposability on restrictions of a persistence module M to squares, and this condition can be rephrased in terms of algebraic conditions on images and kernels of linear maps $M(p) \to M(q)$ for $p \leq q$ in $\mathcal{P}$. The algorithm computing the direct-sum decomposition is based on an inclusion-exclusion formula for a generalization of the rank invariant for interval-decomposable modules (Kim and Mémoli, 2021, Proposition 7.13). We provide an explicit statement of this formula in our context together with a simple and direct proof.

Second, we show how rectangle-decomposable modules arise from real-valued functions on a topological space. More precisely, we prove a "continuous" version of the so-called *pyramid basis theorem* of Bendich et al. (2013) from level-sets persistence. Here continuous refers to the fact that the persistence modules need not be completely determined by the restriction to a finite set of indices. In fact our results generalize to persistence modules indexed over the *strip* recently introduced in Bauer, Botnan, and Fluhr (2021).

II. Invariants of persistence modules

In this section, we give a concise and non-exhaustive overview of the existing invariants of multi-parameter persistence modules as a means of contextualizing our contributions.

By *invariant*, we mean a map from the set of isomorphism classes of persistence modules to a (pseudo-)metric space. Many invariants can be found in the vast literature of quiver representation theory, and plenty more have now been introduced by the topological data analysis community.

What is a good invariant? With the applied context in mind, it is natural to ask what the characteristics of a good invariant of persistence modules are. It should be:[3]

(i) Informative, as much as possible, if not complete. For instance, distance between invariants should be great when topological properties of the data differ.

(ii) Computable, as efficiently as possible.

(iii) Stable, that is, Lipschitz with respect to some metric on the data. Keeping persistence barcodes in mind, the metric on the data could be an L_p metric between the functions whose sublevel sets are considered, or the Hausdorff metric on point clouds. As it was the case for persistence barcodes, this property ensures that the invariant is robust with respect to certain type of noise in the data, and can be used to derive consistent estimators in statistics.

(iv) Interpretable. This property is different from (i) and can be difficult to obtain depending on the invariant considered.

Moreover, many machine learning techniques, such as support vector machines, require an inner product structure. It would thus be an appreciable property that the (pseudo-)metric defined on the space of invariant be bi-Lipschitz equivalent to a metric induced by an inner product in some Hilbert space.

Is the persistence barcode a good invariant? We can study the properties of the one-parameter persistence barcodes from the perspective of the above list:

(i) As already mentioned, it is complete, hence (i) is fully satisfied.

(ii) One of the advantages of homology groups over, e.g., homotopy groups, is their low computational cost. If the parametrizing poset is a finite totally ordered set and if the spaces of the filtration are simplicial complexes, computing a persistence barcode boils down to a matrix reduction procedure. As a consequence, it takes polynomial time in the number of simplices appearing in the filtration, with polynomial degree the exponent for matrix multiplication (Milosavljević, D. Morozov, and Skraba, 2011).

[3]Note that a similar list appears in Adams et al. (2017) in the specific case of finite-dimensional vectorizations (i.e., vector valued invariants) of one-parameter persistence modules.

(iii) As already mentioned, the bottleneck and p-Wasserstein distances on persistence barcodes satisfy stability properties (Chazal, De Silva, et al., 2016; Cohen-Steiner, Edelsbrunner, and Harer, 2007; Cohen-Steiner, Edelsbrunner, Harer, and Mileyko, 2010).

(iv) Persistence barcodes are interpretable. Long bars in the barcode represent persistent homological features in the data. This provides indications on the intrinsic dimension of the data and offer insights into its topological structure. For instance, toroidal structure in grid cells activity has been detected using persistence barcodes in Gardner et al. (2022).

Unfortunately, the space of persistence barcodes equipped with the bottleneck distance cannot be isometrically embedded into a Hilbert space (Bubenik and A. Wagner, 2020; Carrière and Bauer, 2019). Moreover, point (iv) should be mitigated by two remarks. First, homological information collected on the data may remain unclear to interpret: what do H_p generators tell us on the data for $p \geq 3$? What does torsion in these homology groups tell us? Second, if small bars have historically been considered as noise, recent work shows that they collectively contain rich information on the geometric structure of the data. For instance, small bars detect curvature (Bubenik, Hull, et al., 2020), or distribution of points; see for instance Obayashi, Hiraoka, and Kimura (2018) and Chapter 9 of the present thesis. However, it remains unclear how to read these properties from a given persistence barcode.

Multi-parameter setting. The *rank invariant* of a persistence module M indexed over $\mathcal{P}$ has been introduced in Carlsson and Zomorodian (2009) and records the rank of linear maps $M(p) \to M(q)$ for all $p \leq q$ in $\mathcal{P}$. This invariant is complete in the one-parameter case; see Carlsson and Zomorodian (ibid., Theorem 12) for the discrete case and Chazal, De Silva, et al. (2016, Propositions 2.3 and 2.18) for a general statement. However, it is not in the multi-parameter setting; see (4.1) for a counter-example proving this latter fact. This invariant is stable with respect to the interleaving distance on persistence modules. This is proven in a note by Landi (2014) for an equivalent invariant, the so-called *fibered barcode*. This latter invariant is defined as the collection of one-parameter barcodes of the restrictions of a given module to affine lines with positive slopes. The main limitation of this invariant is its computational cost. The time complexity of computing the fibered barcode is the product of the number of lines times the computational cost of a one-parameter persistence barcode. This is a heavy price to pay in practice. On the other hand, this invariant benefits from similar interpretability properties to the persistence barcode. The fibered barcode has recently been generalized to the notion of *projected barcodes* in Berkouk and Petit (2022). This stable invariant is more flexible, but still heavy to compute.

We do not discuss other invariants of persistence modules, such as persistence images and landscapes (Adams et al., 2017; Bubenik et al., 2015; Carrière and Blumberg, 2020; Vipond, 2020). These invariants have been introduced to map persistence barcodes into Hilbert spaces and both enjoy interesting properties in view of the above list. Again, one of their main limitation is their computational cost. This will become clear in Chapter 9 when comparing these invariants to Euler characteristic based invariants.

Euler characteristic based invariants. Considering the pointwise Euler characteristic of a one-parameter family of simplicial complexes gives rise to a functional multi-scale descriptor called the *Euler characteristic curve* $\chi_{\mathcal{F}} : \mathcal{P} \to \mathbb{Z}$. Of course, it generalizes readily to the multi-parameter setting, becoming the so-called *Euler characteristic surfaces* (Beltramo, Skraba, et al., 2022) in the case of two parameters and *profiles* (Dłotko and Gurnari, 2022) in the case of three and more parameters. A classical formula ensures that computing the Euler characteristic of a simplicial complex only requires to count simplices instead of computing homology groups. Similarly, when the filtration $\mathcal{F}$ is the sublevel-sets filtration associated to a Morse function $f : Y \to \mathbb{R}$, the Euler characteristic can be expressed as a signed sum over the critical values of f. In both cases, computing the Euler characteristic bypasses the computation of homology and the construction of persistence modules altogether. Regarding the properties of such invariants, the following picture can be drawn:

(i) Euler characteristic based descriptors are not as coarse as they may appear, as proven by their good predictive power (Amézquita et al., 2022; Jiang, Kurtek, and Needham, 2020; Smith and Zavala, 2021).

(ii) Due to their simplicity, Euler characteristic profiles can be computed in linear time in the total number of simplices in a simplicial filtration $\mathcal{F}$ instead of matrix multiplication time for persistence barcodes. Moreover, the locality of the Euler characteristic can be exploited to design highly efficient algorithms computing Euler curves, as in Heiss and H. Wagner (2017) and Wang, H. Wagner, and C. Chen (2022).

(iii) Euler characteristic profiles are integer-valued maps on the Euclidean space. As such, they can be endowed with classical L_p metrics. In order to transfer existing stability results for the (fibered) barcodes to Euler characteristic profiles, one may wonder what Lipschitz inequalities hold between metrics on graded persistence modules and metrics on the associated Euler profiles. Such inequalities are known for the L_1 distance on Euler characteristic curves and the 1-Wasserstein metric on barcodes; see Curry, Mukherjee, and Turner (2022) and Dłotko and Gurnari (2022). However, it has recently been shown that there does not exist a non-trivial distance on Euler characteristic curves that is bounded above by the Bottleneck distance on persistence barcodes (Berkouk, 2022, Proposition 4.25).

(iv) Several authors have worked towards an interpretation of Euler characteristic curves and surfaces (Beltramo, Skraba, et al., 2022; Smith and Zavala, 2021) and more generally towards the construction of interpretable tools from Euler characteristic techniques (Tang et al., 2022). However, the interpretability of these descriptors is still to be investigated. Though tightly connected to persistent homology, the signal is very different to interpret. Indeed, the classical dichotomy signal/noise used for persistence barcodes does not hold: a succession of short-lived homological classes or a single long-lived homological class can both generate the same Euler curve.

Euler calculus. Euler characteristic based invariants of multi-parameter persistence modules are naturally formulated in the language of *Euler calculus*. This integral calculus with respect to the Euler characteristic has originally been developed by Viro (1988) and by Schapira (1989, 1991). The measurable functions of this calculus are the so-called *constructible functions*, that is, piecewise constant functions on a partition of the space into nice geometric subsets—usually definable in some fixed o-minimal structure.

In Schapira (1989, 1991), Euler calculus is introduced in relation to sheaf operations, the link between constructible functions and constructible sheaves being specified by the *function-sheaf correspondence* (Kashiwara, 1985; Kashiwara and Schapira, 1990). These tools are thus part of the sheaf-theoretic approach to multi-parameter persistence initiated by Curry (2014) and Kashiwara and Schapira (2018).

Euler calculus has led to important advances in topological data analysis; we refer to Curry, Ghrist, and Robinson (2012) for a survey. Main fruits come from the introduction of topological integral transforms (Baryshnikov, Ghrist, and Lipsky, 2011; Curry, Mukherjee, and Turner, 2022; Ghrist, Levanger, and Mai, 2018; Turner, Mukherjee, and Boyer, 2014a). The main one is the Radon transform introduced by Schapira (1995). The Radon transform of the indicator function $\mathbf{1}_Z$ of a definable subset $Z \subset \mathbb{R}^m$ records the Euler characteristic of the intersection $Z \cap H$ for each affine hyperplane $H \subset \mathbb{R}^m$. It is then extended by linearity to all constructible functions. Schapira's inversion result (ibid.) ensures that this transform is invertible, up to a constant when m is even. It is shown in Curry, Mukherjee, and Turner (2022) and Ghrist, Levanger, and Mai (2018) that this implies a positive answer to an important inverse question: are two definable subsets of $\mathbb{R}^m$ with same persistent homology in all degrees and for all height filtrations equal? These techniques found applications in shape analysis, for instance in the prediction of clinical outcomes in brain cancer (Crawford et al., 2020), in the analysis of barley seeds (Amézquita et al., 2022) or in the recovery of morphological variations across genera of primates (Tang et al., 2022).

The most problematic aspect of Euler calculus for applications is its instability under numerical approximations: errors can (and probably will) be made when calculating the integral of a constructible function, no matter how finely its domain is sampled (Curry, Ghrist, and Robinson, 2012, Section 16). In Crawford et al. (2020), topological integral transforms are smoothed to produce better behaved shape descriptors in practice. These smoothings are closely related to the *Bessel* and *Fourier* transforms introduced by Ghrist and Robinson (2011), as well as the *Euler characteristic of barcodes* introduced by Bobrowski and Borman (2012). For now, no theoretical result supports these experimental results.

Contributions. In the second part of this Ph.D. thesis, we are concerned with Euler characteristic based invariants in topological data analysis.

We introduce a general definition of *hybrid transforms* for constructible functions generalizing the aforementioned smoothings of topological transforms. These are integral transforms combining Lebesgue integration and Euler calculus. Lebesgue integration gives access to well-studied kernels and to regularity results, while Euler calculus conveys topological information and allows for compatibility with operations on constructible functions. We conduct a systematic study of such transforms and introduce two new

ones: the Euler-Fourier and Euler-Laplace transforms. We show that the first has a left inverse and that the second provides a satisfactory generalization of *persistent magnitude* (Govc and Hepworth, 2021) to constructible sheaves, in particular to multi-parameter persistence modules. Finally, we prove index-theoretic formulae for a wide class of hybrid transforms in the context of (sub)level-sets multi-parameter persistence, generalizing existing ones (Bobrowski and Borman, 2012; Ghrist and Robinson, 2011; Govc and Hepworth, 2021). In particular, this yields expectation formulae for random Gaussian filtrations.

In a joint work with Olympio Hacquard (2023), we show that Euler characteristic profiles and their hybrid transforms are informative and efficiently computable topological descriptors of data. To begin with, we prove stability results for these descriptors as well as asymptotic guarantees in random settings. Then, we demonstrate that Euler characteristic profiles reach state-of-the-art performance at a minimal computational cost when coupled with a powerful classifier such as a random forest. However, due to their simplicity, these descriptors do not manage to linearly separate the different classes or be competitive on unsupervised tasks. Inspired by signal analysis, we cope with these limitations by studying hybrid transforms of Euler characteristic profiles. We show that, as expected, hybrid transforms show remarkable performances in linear classification tasks and unsupervised contexts. On the qualitative side, we provide numerous heuristics on the topological and geometric information captured by Euler profiles and their hybrid transforms.

Detailed outline and summary of contributions

Here, we present our contributions in more details and expose the structure of the thesis. We refer to the following chapters for the precise definitions of our terms.

Part I: Direct-sum decompositions of persistence modules

In the first part of this thesis, we study local conditions for decomposability of two-parameter persistence modules. All the results of this part are based on joint work with Magnus Bakke Botnan and Steve Oudot (2022, 2023). In this summary, we always assume that persistence modules are pointwise finite-dimensional.

Chapter 1: Local characterizations for decomposability

This chapter serves as an introduction to the first part of the thesis. We start by defining multi-parameter persistence modules, interval-decompositions and other necessary notions (Section 1.1). Then we give a precise statement of our main question: existence of local characterizations of interval-decomposability (Section 1.2). After reviewing existing answers to this question (Section 1.3), we state our contributions (Section 1.4).

Chapter 2: Rectangle-decomposability is local

This chapter is devoted to the proof of the main result of Part I:

Theorem (Theorem 1.7)**.** *Let X and Y be two totally ordered sets such that any interval of X or Y admits a countable initial subset. Then, any persistence module M over $X \times Y$ is rectangle-decomposable if and only if the restriction of M to any square $\{x_1 < x_1'\} \times \{x_2 < x_2'\} \subseteq X \times Y$ is rectangle-decomposable.*

We start by providing a short proof of this result when X and Y are finite following Botnan, Lebovici, and Oudot (2022) (Section 2.1). In contrast, the proof of the general case uses the finite case and spans more than twenty pages. It follows Botnan, Lebovici, and Oudot (2023) and is developed in Sections 2.2 through 2.5. The proof in both cases relies on a key algebraic formulation of rectangle-decomposability of restrictions to 2×2 grids (Definition 2.1).

Since it is straightforward that rectangle-decomposable modules restrict to rectangle-decomposable modules, we are left with proving the converse statement. For that, the proof in the finite case uses a simple rank-based approach. In the general case, we follow the same approach as in Cochoy and Oudot (2020), using the so-called *functorial filtrations* (Section 2.2), with some major adjustments at key steps due to our weaker local condition. Summarized, we start by defining, for each rectangle $R \in X \times Y$, a submodule M_R of M called the *rectangle filtrate of M associated to R* (Section 2.3). This submodule is constructed such that $M_{R,t}$ contains precisely the elements of M_t whose "lifespan" is exactly R. In particular, M_R is isomorphic to a finite direct sum of copies of the rectangle module $\mathbf{k}_R$. We then prove that the sum of these filtrates is an internal direct sum in M (Section 2.4). The proof concludes by showing that the resulting internal direct sum generates M (Section 2.5).

Chapter 3: Non-existence of other local characterizations

In this chapter, we provide negative answers to our main question. The results from the previous chapter ensure that rectangle-decomposability of a given persistence module can be checked locally by considering restrictions to commutative squares. A natural next question to consider is then: to what extent can interval-decomposability be locally determined when allowing for intervals of more general shape than rectangles?

We provide two negative results, which have first been proven in the finite case in Botnan, Lebovici, and Oudot (2022) and then generalized in Botnan, Lebovici, and Oudot (2023). First, we show that interval-decomposability itself cannot be characterized locally, even when testing on arbitrary strict subgrids:

Theorem (Theorem 3.1)**.** *Suppose X and Y are totally ordered sets with $|X| \geq 3$ and $|Y| \geq 3$, and let $2 \leq m < \min(|X|, |Y|)$ be an integer. Then, there exists a persistence module M over $X \times Y$ that is not interval-decomposable, but for which $M|_Q$ is interval-decomposable for all grids Q of side-lengths at most m.*

In light of the above results, it is natural to wonder if there is a class of intervals more general than rectangles for which a local characterization over 2×2 grids is possible. It turns out not to be the case, as shown by Theorem 3.5.

Chapter 4: Applications to topological data analysis

In Section 4.1, we provide algorithms to check whether a persistence module deriving from a simplicial two-parameter filtration is rectangle-decomposable, and to compute its direct-sum decomposition when it exists.

As a first step, we prove that the rank invariant is complete on the class of two-parameter rectangle-decomposable modules (Theorem 4.4). To this end, we generalize the inclusion-exclusion formula expressing the persistence barcode in terms of the rank invariant (Cohen-Steiner, Edelsbrunner, and Harer, 2007) to the two-parameter setting. Note that our result also follows indirectly from a similar formula for a generalization of the rank invariant (Kim and Mémoli, 2021), but that we provide an explicit statement together with a simple and direct proof.

Then we show that the rank invariant of a module produced by a simplicial two-parameter filtration with a total of n simplices can be computed in $O(n^4)$ time (Theorem 4.5). This result in itself is not new, however, combined with our inclusion-exclusion formula, it yields a $O(n^4)$ time algorithm for computing the barcode of a two-parameter persistence module that is known to be rectangle-decomposable (Corollary 4.6). At the time, our result was an improvement over merely applying some state-of-the-art algorithm for computing decompositions of general two-parameter persistence modules, which would take $O(n^{2\omega+1})$ time where $2 \leq \omega < 2.373$ is the exponent for matrix multiplication (Dey and Xin, 2022). Recently, Clause et al. (2023) improved our result, proving that the rank invariant of two-parameter persistence modules induced by simplicial filtration can be computed in $O(n^3)$ time.

Finally, we leverage our rectangle-decomposition result to derive an $O(n^{2+\omega})$-time algorithm for checking the rectangle-decomposability of persistence modules induced in homology from simplicial two-parameter filtrations with at most n simplices (Theorem 4.7). Once again, this is an improvement over applying some state-of-the-art algorithm for computing decompositions of general two-parameter persistence modules and then checking the summands one by one.

In Section 4.2, we apply our rectangle-decomposition result to level-sets persistence. First, we expose the constructions involved in level-sets persistence as developed in Bauer, Botnan, and Fluhr (2021). Then we prove our main result, which is a decomposition theorem for persistence modules indexed over the strip (Theorem 4.11). As we previously mentioned, a "continuous" version of the pyramid basis theorem introduced by Bendich et al. (2013) is obtained as a corollary.

Part II: Euler characteristic based invariants

In the second part of this thesis, we introduce new invariants of multi-parameter persistence modules using integral transforms of constructible functions.

Chapter 5: Euler calculus

This chapter serves as an introduction to Part II. We set our notations and recall some basic definitions and results on Euler calculus following the approach of Schapira (1991)

(Section 5.1). We present several specific subgroups of the group of constructible functions used all along Part II (Section 5.2). Among them, the subgroup $\mathrm{CF}_{\gamma,\mathrm{c}}(\mathbb{R}^m)$ of compactly supported γ-constructible functions—whose strata are γ-locally closed and relatively compact—, which includes constructible functions associated to compactly supported multi-parameter persistence modules. Then, we define the *Radon transform* introduced by Schapira (1995) and expose Schapira's inversion result (Section 5.3). Moreover, we prove that the Radon transform of a compactly supported γ-constructible function is zero for affine hyperplanes whose defining conormal lies in the complement of the polar of γ. This characterization of the support of the transform will be key to prove the reconstruction formula for the Euler-Fourier transform in Chapter 6. We end the chapter by explaining how to associate a constructible function to a graded multi-parameter persistence module (Section 5.4).

Chapter 6: Hybrid transforms of constructible functions

In this chapter, we introduce the general definition of hybrid transforms and conduct a systematic study of their theoretical properties. First, we give the general definition of these transforms as well as two important examples: the *Euler-Laplace* transform and the *Euler-Fourier* transform (Section 6.1). We show how the Euler-Laplace transform allows us to define a generalization of the persistent magnitude (Govc and Hepworth, 2021) for constructible sheaves. In addition, we provide several illustrated examples to compare these two transforms to their classical analogues. We end the section by presenting a method to numerically compute our transforms on PL-constructible functions, i.e., those whose defining strata are polyhedral, and we provide a `Python` and `C++` library freely available on GitHub. Second, we prove that hybrid transforms are continuous when restricted to the set of PL-constructible functions, and even C^{p+1} on the interior of cones partitioning their domain when their kernel is C^p (Section 6.2). Third, we show that hybrid transforms are compatible with operations on constructible functions such as pushforwards and duality (Section 6.3). In particular, we show that the Euler-Laplace and Euler-Fourier transforms turn (constructible) convolutions into products under mild assumptions. Finally, we establish a reconstruction formula for the Euler-Fourier transform of compactly supported γ-constructible functions (Section 6.4). The proof of this latter fact goes as follows. Using the inverse of the classical Fourier transform, one can recover, from the knowledge of the Euler-Fourier transform, the values of the Radon transform on the set of affine hyperplanes whose defining conormal is in the cone γ. Our result on the support of the Radon transform of γ-constructible functions from Chapter 5 then ensures that the Radon transform is in fact fully recovered, as it is zero for any other hyperplane. All that remains to do is to invert the Radon transform using Schapira's formula (1995).

Chapter 7: Index-theoretic formulae

In this chapter, we begin by defining the so-called *sublevel-sets* and *level-sets constructible functions* associated to a continuous subanalytic map $f : M \to \mathbb{V}$ and a cone $\gamma \subset \mathbb{V}$, where $\mathbb{V}$ is a finite-dimensional real vector space and M is a real analytic manifold (Section 7.1). These are simply the constructible functions associated to persistent co-

homology sheaves introduced by Kashiwara and Schapira (2018). We show that hybrid transforms of (sub)level-sets constructible functions associated to vector-valued maps can be expressed as hybrid transforms of (sub)level-sets constructible functions associated to real-valued maps (Corollary 7.10). The proof relies on an expression of the sublevel-sets constructible function as a convolution of the level-sets constructible function with the indicator function of the antipodal of γ. This reduction leads to the definition of *sublevel-sets transforms*, which is a simplified form that hybrid transforms take in the case of multi-parameter sublevel-sets persistence.

Next, we recall the definition of *continuous Euler integral* from Bobrowski and Borman (2012) that extends Euler calculus to the wider class of *tame functions* (Section 7.2). This class contains continuous subanalytic functions on compact real analytic manifolds. Then we prove our main results called *index-theoretic formulae* that express (sub)level-sets transforms as continuous Euler integral transforms. It allows us to prove an expectation formula for the Euler-Bessel transform of level-sets constructible functions of Gaussian random fields using Bobrowski and Borman's formula (ibid.) for the expectation of continuous Euler integrals (Section 7.3).

Chapter 8: Euler characteristic profiles and their transforms

In the last two chapters of the thesis, we study constructible functions associated to persistent homology of simplicial filtrations as well as their hybrid transforms. In this chapter, we provide several theoretical guarantees for these descriptors. The work presented in this chapter is a collaboration with Olympio Hacquard (2023). We begin by introducing some necessary definitions (Section 8.1). Then, we prove stability properties that clarify the robustness of our tools with respect to perturbations (Section 8.2). We slightly generalize existing results on Euler profiles and derive new results for hybrid transforms in a unified framework using the *signed* 1-*Wasserstein distance* introduced in Oudot and Scoccola (2021).

Finally, we establish the pointwise convergence of hybrid transforms associated to simplicial filtrations of random point clouds, as well as their asymptotic normality for a specific filtration function (Section 8.3). We also establish a law of large numbers in a multi-parameter set-up. These statistical results are easy consequences of existing results on *persistence diagrams*. However, it is worth noticing that the result in the multi-parameter setting relies crucially on our result of the previous chapter expressing multi-parameter hybrid transforms as one-parameter ones in the context of sublevel-sets persistence (Corollary 7.10).

Chapter 9: Experimental study

In this chapter, we show that Euler characteristic profiles and their hybrid transforms are informative and efficiently computable topological descriptors of filtered simplicial complexes. Again, the results of this chapter are joint work with Olympio Hacquard (2023). First, we explain our algorithms and provide heuristics on how to choose the kernel of hybrid transforms (Section 9.1). We provide a `Python` library available on GitHub to compute our descriptors. Second, we conduct an empirical study of the type of topological and geometric information that Euler profiles and their transforms can extract from

the data (Section 9.2). Third, we perform qualitative and quantitative experiments: a supervised regression task and supervised and unsupervised classification tasks on graph and point cloud data. We demonstrate that Euler profiles achieve state-of-the-art accuracy in supervised classification and regression tasks when coupled with a random forest or a gradient boosting classifier at a very low computational cost. Moreover, we demonstrate that hybrid transforms provide a way to efficiently compress topological and geometric information. As a consequence, they outperform Euler profiles in unsupervised classification tasks and in supervised tasks when using a linear classifier. An example of their ability to capture fine-grained information on a real-world data set is provided.

Publications

- Magnus Bakke Botnan, Vadim Lebovici, and Steve Oudot (2022). "On Rectangle-Decomposable 2-Parameter Persistence Modules". In: *Discrete and Computational Geometry*. Initially accepted in: 36th International Symposium on Computational Geometry (SoCG 2020), pp. 1–24
- Magnus Bakke Botnan, Vadim Lebovici, and Steve Oudot (2023). "Local characterizations for decomposability of 2-parameter persistence modules". In: *Algebras and Representation Theory*, pp. 1–44
- Vadim Lebovici (2022). "Hybrid transforms of constructible functions". In: *Foundations of Computational Mathematics*, pp. 1–47
- Olympio Hacquard and Vadim Lebovici (2023). "Euler characteristic tools for topological data analysis". Submitted

Code

- `https://github.com/HugoPasse/Transforms-of-cubical-complexes`
- `https://github.com/vadimlebovici/eulearning`

Part I

Direct-sum decompositions of persistence modules

Chapter 1

Local characterizations for decomposability

Based on joint work with Magnus Bakke Botnan and Steve Oudot.

In the first part of this thesis, we investigate the existence of sufficient *local conditions* under which a persistence module M indexed over a poset $(\mathcal{P}, \leq)$ decomposes as direct sums of indecomposables from a given class. By local conditions, we mean that decomposability of M must be entirely determined by the decomposability of its restrictions to a collection of *test subsets* of $\mathcal{P}$. In our work, the indexing poset is the product $\mathcal{P} = X \times Y$, of two totally ordered sets X and Y and test subsets are the products $X' \times Y'$ of two finite subsets $X' \subseteq X$ and $Y' \subseteq Y$. Our indecomposables of interest belong to the so-called interval modules, which by definition are indicator representations of intervals in the poset.

We will prove in Chapter 2 and Chapter 3 that, while the whole class of interval modules does not admit such a local characterization, the subclass of rectangle modules does admit one, for test subsets being products of two-element subsets. Moreover, we prove that it is, in some precise sense, the largest subclass that does so.

Outline. This short chapter serves as an introduction to Part I. First, we set up our framework and our notations in Section 1.1. Then, we provide a formal statement of our main question (Section 1.2) and expose existing results (Section 1.3). Finally, we present our contributions (Section 1.4).

1.1 Preliminaries

Our exposition uses the language of topological data analysis to talk about representations. Here we spend a few paragraphs defining our terms.

Persistence modules. A *persistence module* indexed over a poset $(P, \preccurlyeq)$ is a functor $M : (P, \preccurlyeq) \to \mathsf{Vec}$ where Vec denotes the category of vector spaces over a fixed field $\mathbf{k}$. Morphisms between persistence modules are natural transformations between functors.

As Vec is an abelian category, so is the category $\mathsf{Per}(P, \preccurlyeq) := \mathrm{Fun}((P, \preccurlyeq), \mathsf{Vec})$ of persistence modules over $(P, \preccurlyeq)$. More precisely, kernels, cokernels and images, as well as products, direct sums and quotients of persistence modules are defined pointwise at each index $p \in P$, and the category of persistence modules admits a zero object, denoted by 0, which is the persistence module whose spaces and internal morphisms are all equal to zero. Moreover, one can dualize each internal space and each internal morphism of a persistence module M over P to get a *dual persistence module* DM over P^{op}, the opposite category of P.

For every $p \in P$, the vector space $M(p)$ is called *internal space of* M *at* p and denoted for short by M_p. For every $p \preccurlyeq q$, the linear map of $M(p \preccurlyeq q) : M_p \to M_q$ is called *internal morphism of* M *between* p *and* q and denoted for short by ρ_p^q. We say that M is *pointwise finite-dimensional* (pfd) if for every $p \in P$, M_p is finite-dimensional. The *support* of M is the subset $\mathrm{supp}(M)$ of indices $p \in P$ for which $M_p \neq 0$. The *rank invariant* of M is the map $r : (p, q) \in P \times P \to \mathbb{Z}$ such that $r(p, q)$ is equal to $\mathrm{rank}(\rho_p^q)$ if $p \preccurlyeq q$ and zero otherwise. We say that a family $\{M_\alpha\}_{\alpha \in A}$ of persistence modules over P is *locally finite* if the set $\{\alpha \in A \,|\, p \in \mathrm{supp}(M_\alpha)\}$ is finite for each $p \in P$. A *locally finite direct sum* is the direct sum of a locally finite family of persistence modules.

A morphism $f : M \to N$ between two persistence modules over P is a *monomorphism* (resp. *epimorphism*) if for every $p \in P$, $f_p : M_p \to N_p$ is injective (resp. surjective). A morphism between two persistence modules is an *isomorphism* if is it both a monomorphism and an epimorphism. A *submodule* of M is a persistence module N together with a monomorphism $N \to M$ of persistence modules, often denoted by $N \hookrightarrow M$.

We say that a monomorphism $f : M \to N$ between two persistence modules M and N *splits* if there is a morphism $g : N \to M$ such that $g \circ f = \mathrm{Id}_M$. If every monomorphism with domain M splits, we say that M is an *injective persistence module*. It is not true that any submodule of a persistence module is a summand. However, if $f : M \to N$ is a monomorphism between two persistence modules M and N which splits, it is well known that there is a direct-sum decomposition $N \cong M \oplus \mathrm{Coker}(f)$. Therefore, an injective submodule of a persistence module is a summand thereof.

Decomposability. A persistence module over $(P, \preccurlyeq)$ is said to be *decomposable* if it decomposes as direct sum of at least two nontrivial persistence modules. Otherwise, it is said to be *indecomposable*. The endomorphism ring $\mathrm{End}(M) := \mathrm{Hom}(M, M)$ is *local* if θ or $\mathrm{Id}_M - \theta$ is invertible for all $\theta \in \mathrm{End}(M)$. It is easy to see that if M has a non-trivial decomposition then $\mathrm{End}(M)$ is not local. The pfd persistence modules over $(P, \preccurlyeq)$ form a Krull-Schmidt subcategory of $\mathsf{Per}(P, \preccurlyeq)$:

Theorem 1.1 (Botnan and Crawley-Boevey, 2020)**.** *Every pfd persistence module M over $(P, \preccurlyeq)$ decomposes as a direct sum of indecomposable modules with local endomorphism rings. By Azumaya's theorem (1950) this decomposition is unique up to isomorphism.*

Product posets. In the this thesis, we focus on persistence modules over product posets. Given two totally ordered sets $(X, \leq_X)$ and $(Y, \leq_Y)$, their *product* $(X \times Y, \leq)$ is the Cartesian product $X \times Y$ equipped with the *coordinatewise order* $\leq$ defined by

$$\forall s, t \in X \times Y,\ s \leq t \iff s_x \leq_X t_x \text{ and } s_y \leq_Y t_y, \tag{1.1}$$

where the coordinates of a point $u \in X \times Y$ are denoted by (u_x, u_y). Henceforth we use the notation $X \times Y$ instead of $(X \times Y, \leq)$ as the only order considered on the Cartesian product $X \times Y$ will be the coordinatewise order. A finite product poset $X \times Y$ is called a *finite grid*. If we want to specify the cardinalities $n = |X|$ and $m = |Y|$, the poset $X \times Y$ is called a $n \times m$ *grid*. The integers n and m are called the *side-lengths* of the grid.

Convention 1.2. From now on and until the end of Part I, we fix two totally ordered sets $(X, \leq_X)$ and $(Y, \leq_Y)$ and consider their product $(X \times Y, \leq)$.

A persistence module M over $X \times Y$ is called a *two-parameter persistence module*, or a *persistence bimodule*indexpersistence bimodule for short. Any two comparable points $s \leq t$ in $X \times Y$ yield a *square* $\mathrm{Q}_s^t := \{s, (s_x, t_y), (t_x, s_y), t\}$ and an associated commutative diagram

$$\begin{array}{ccc} M_{(s_x,t_y)} & \xrightarrow{\rho^t_{(s_x,t_y)}} & M_t \\ {\scriptstyle \rho_s^{(s_x,t_y)}}\uparrow & & \uparrow{\scriptstyle \rho^t_{(t_x,s_y)}} \\ M_s & \xrightarrow[\rho_s^{(t_x,s_y)}]{} & M_{(t_x,s_y)} \end{array}$$

Intervals, rectangles, blocks. Let P be a poset. We say that $S \subseteq P$ is *convex* if, for every $p \preccurlyeq q \in S$, we have $s \in S$ for all $s \in P$ such that $p \preccurlyeq s \preccurlyeq q$. A convex set S is an *interval* if it is also *connected*, i.e., for every $p, q \in S$ there is a finite sequence $(p_i)_{i \in [\![0,n]\!]}$ of points of S such that $p = p_0 \perp \cdots \perp p_n = q$, where $\perp$ is the binary relation defined by $p \perp q$ if and only if p and q are comparable ($p \preccurlyeq q$ or $q \preccurlyeq p$). We write $\mathrm{Conv}(P)$ (resp. $\mathrm{Int}(P)$) for the set of convex (resp. interval) subsets of P.

To any convex set $S \subseteq P$ we associate a persistence module $\mathbf{k}_S$, called the *indicator module* of S, defined as follows:

$$\mathbf{k}_S(p) = \begin{cases} \mathbf{k} & \text{if } p \in S, \\ 0 & \text{else,} \end{cases} \tag{1.2}$$

and

$$\mathbf{k}_S(p \leq q) = \begin{cases} \mathrm{Id}_{\mathbf{k}} & \text{if } p \text{ and } q \in S, \\ 0 & \text{else,} \end{cases} \tag{1.3}$$

and by convention we set $\mathbf{k}_\varnothing = 0$. When S is an interval, $\mathbf{k}_S$ is called an *interval module*, and it is indecomposable because its endomorphism ring $\mathrm{End}(\mathbf{k}_S)$ is isomorphic to $\mathbf{k}$ (by connectivity of S) and thus local. Otherwise, $\mathbf{k}_S$ decomposes as the direct sum of the indicator modules of its connected components. If M is isomorphic to a direct sum of interval modules, then we say that M is *interval-decomposable*.

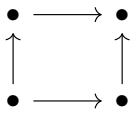
We call any product $I \times J$ of two intervals $I \subseteq X$ and $J \subseteq Y$ a *rectangle*, and we write $\mathrm{Rec}(X \times Y)$ for the set of rectangles of $X \times Y$. Note that rectangles are intervals by definition, and their associated interval modules are called *rectangle modules*.

A *block* is any rectangle $B = I \times J$ that satisfies either of the following (non-exclusive) conditions:

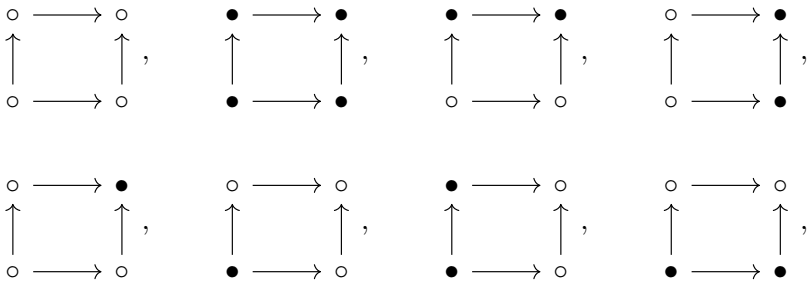
- I is cofinal[1] in X and J is cofinal in Y—B is then called a *birth quadrant*;
- I is coinitial in X and J is coinitial in Y—B is then called a *death quadrant*;
- I is both coinitial and cofinal in X—B is then called a *horizontal band*;
- J is both coinitial and cofinal in Y—B is then called a *vertical band*.

We write $\mathrm{Blc}(X \times Y)$ for the set of blocks of $X \times Y$. Blocks are rectangles by definition, and their associated rectangle modules are called *block modules*.

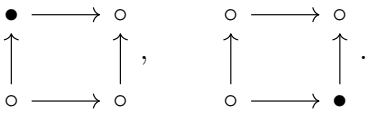
Example 1.3. Let Q be a square of $X \times Y$ represented by the following diagram.



$\mathrm{Blc}(Q)$ is the collection of the following subposets of Q:



where $\bullet$ represents a point that belongs to the corresponding block and $\circ$ represents a point that does not. The collection $\mathrm{Rec}(Q)$ consists of the above subposets and the following two "corners":



Definition 1.4. A persistence bimodule M is said to be *rectangle-decomposable* (*block-decomposable*) if it decomposes into a direct sum of interval modules supported on rectangles (blocks).

Example 1.5. One of the simplest example of block-decomposable modules are the ones having only birth quadrants of the form $Q_u = \{t \in X \times Y \,;\, t \geq u\}$ for $u \in X \times Y$ in their block-decomposition, called *free* persistence modules. Of course, the notion of free persistence module generalizes to any cartesian product $X_1 \times \ldots \times X_n$ of a finite number of totally ordered sets.

[1]Given a poset $(P, \preccurlyeq)$, a subset $Q \subseteq P$ is *coinitial* if every $p \in P$ admits a $q \in Q$ such that $q \preccurlyeq p$, and *cofinal* if every $p \in P$ admits a $q \in Q$ such that $q \succcurlyeq p$.

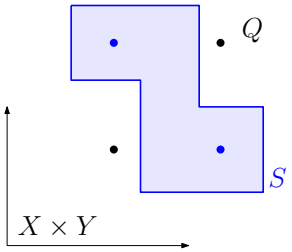


Figure 1.1: An example of $S \in \mathcal{S}$ (the solid blue polygon) and $Q \in \mathcal{Q}$ (the four dots arranged in a square) such that $S \cap Q$ is convex but not connected in Q.

1.2 Local characterizations

Our aim is to work out a local conditions that characterize the decomposability of pfd persistence bimodules over a given class of interval modules. We specify this class of interval modules via the set $\mathcal{S} \subseteq \mathrm{Int}(X \times Y) \setminus \{\varnothing\}$ of their supports, and we write $\langle \mathcal{S} \rangle$ for the set of all pfd persistence bimodules that are obtained (up to isomorphism) as direct sums of such interval modules:

$$\langle \mathcal{S} \rangle := \left\{ M \in \mathsf{Per}(X \times Y) \mid M \simeq \bigoplus_{S \in \mathcal{S}} \mathbf{k}_S^{m_S} \text{ where } \sum_{S \ni t} m_S < \infty \text{ for all } t \in X \times Y \right\}.$$

Note that the class $\langle \mathcal{S} \rangle$ is still well-defined for $\mathcal{S} \subseteq \mathrm{Conv}(X \times Y)$. In that case, however, $\mathbf{k}_S$ need not be indecomposable. For $M \in \langle \mathcal{S} \rangle$, the multiset of $S \in \mathcal{S}$ with non-zero multiplicities m_S in the interval-decomposition of M is called the *barcode* of M and denoted by $\mathcal{B}(M)$.

Locality is understood as taking restrictions to a collection $\mathcal{Q}$ of strict subsets of $X \times Y$, called the *test subsets*. Given $Q \in \mathcal{Q}$, let $\mathcal{S}_{|Q} := \{S \cap Q \mid S \in \mathcal{S}\}$ be the set of intervals in $\mathcal{S}$ restricted to Q. Note that $\mathcal{S}_{|Q} \subseteq \mathrm{Conv}(Q)$, since convexity is preserved under taking restrictions. However, it may not be the case that $\mathcal{S}_{|Q} \subseteq \mathrm{Int}(Q)$, since connectivity is not always preserved under taking restrictions; see Figure 1.1 for an example. Nevertheless, we still have[2]:

$$\langle \mathrm{Int}(X \times Y)_{|Q} \rangle = \langle \mathrm{Int}(Q) \rangle.$$

Throughout this thesis, we focus on the setting where the test subsets $\mathcal{Q}$ are *product subsets*, that is, of the form $X' \times Y' \subseteq X \times Y$ for $X' \subseteq X$ and $Y' \subseteq Y$. While intervals of $X \times Y$ may not restrict to intervals on product subsets (see again Figure 1.1), rectangles and blocks do restrict to rectangles and blocks respectively. This means that for any $Q = X' \times Y' \subseteq X \times Y$ we have:

$$\mathrm{Rec}(X \times Y)_{|Q} = \mathrm{Rec}(Q),$$
$$\mathrm{Blc}(X \times Y)_{|Q} = \mathrm{Blc}(Q).$$

Let $\mathcal{S}$ be a collection of intervals in $X \times Y$, and let $\mathcal{Q}$ be a collection of test subsets. Since restriction preserves interval-decomposability, we see that if $M \in \langle \mathcal{S} \rangle$, then $M \in$

[2]This is easily proven. Given an interval $S \subseteq X \times Y$, the restriction $S_{|Q}$ is convex in Q therefore $\mathbf{k}_S$ decomposes as the direct sum of the indicator modules of its connected components in Q. Conversely, given an interval S' of Q, the set $S = \{s \in X \times Y \mid p \preccurlyeq s \preccurlyeq q \text{ for some } p, q \in S'\}$ is an interval of $X \times Y$ that restricts to S' on Q.

$\langle \mathcal{S}_{|Q} \rangle$ for every $Q \in \mathcal{Q}$. Symbolically,

$$M \in \langle \mathcal{S} \rangle \Longrightarrow \forall Q \in \mathcal{Q},\ M_{|Q} \in \langle \mathcal{S}_{|Q} \rangle.$$

In this chapter we are concerned with the reverse implication:

Main question. Characterize the collections $\mathcal{S}$ and $\mathcal{Q}$ such that

$$M \in \langle \mathcal{S} \rangle \Longleftarrow \forall Q \in \mathcal{Q},\ M_{|Q} \in \langle \mathcal{S}_{|Q} \rangle.$$

In such a situation, we say that membership of $\langle \mathcal{S} \rangle$ is *locally characterized on* $\mathcal{Q}$. We simply call *local* a decomposability property that is locally characterized on 2×2 grids.

Observe that the main question is trivial if $X \times Y$ is a member of $\mathcal{Q}$. Focusing on product subsets does not provide a complete picture, and there are several natural next steps; see Section 9.3.6 for a discussion on this.

1.3 State of the art

We now review positive and negative results in the setting of our main question.

Testing interval-decomposability on totally ordered subsets. It has been shown in Botnan and Crawley-Boevey (2020) and Crawley-Boevey (2015) that any pfd persistence module indexed over a totally ordered set is interval-decomposable. Therefore, if M is a pfd persistence module indexed over $X \times Y$, the restriction $M_{|Q}$ is interval-decomposable for any totally ordered subset $Q \subseteq X \times Y$. Hence, any indecomposable module over $X \times Y$ that is not of pointwise dimension 0 or 1 is a counter-example to the existence of a local characterization of interval-decomposability over totally ordered subsets. For instance, consider the following persistence module:

$$\begin{array}{ccccc} \mathbf{k} & \xrightarrow{\binom{1}{1}} & \mathbf{k}^2 & \xrightarrow{(1\ 0)} & \mathbf{k} \\ {\scriptstyle 0}\uparrow & & \uparrow{\scriptstyle \binom{1}{0}} & & \uparrow{\scriptstyle 1} \\ 0 & \xrightarrow{0} & \mathbf{k} & \xrightarrow{1} & \mathbf{k} \end{array} \tag{1.4}$$

It is an easy exercise to check that its endomorphism ring is local. Hence, this module is indecomposable.

Testing interval-decomposability on squares. Recall that the restriction of a pfd persistence module over $X \times Y$ to any square is interval-decomposable (see e.g. Figure 13 in Escolar and Hiraoka, 2016). Therefore, any indecomposable module over $X \times Y$ that is not of pointwise dimension 0 or 1 is a counter-example to the existence of a local characterization of interval-decomposability over squares. Again, one can consider the persistence module (1.4).

Testing block-decomposability on squares. A local characterization of block-decomposable pfd bimodules was given in Cochoy and Oudot (2020) for $X \times Y = \mathbb{R}^2$, and later extended to products of two totally ordered sets in Botnan and Crawley-Boevey (2020).

Theorem 1.6 (Botnan and Crawley-Boevey, 2020; Cochoy and Oudot, 2020)**.** *Let X and Y be totally ordered sets, and M a pfd persistence module over $X \times Y$. Then, the module M is block-decomposable if and only if the restriction of M to any square is block-decomposable.*

Cochoy and Oudot (2020) gave a proof of this fact for modules indexed over $\mathbb{R}^2$ using the functorial filtration technique used by Crawley-Boevey (2015) to prove the one-parameter interval-decomposition theorem. Both these results were generalized in Botnan and Crawley-Boevey (2020) to arbitrary totally ordered sets in the one-parameter case and to binary product of them in the two-parameter case. In loc. cit., the approach is different and based on the general decomposition theorem (Theorem 1.1).

1.4 Contributions

Let us now state our contributions to the existence of local characterizations for decomposability of two-parameter persistence modules.

Rectangle-decomposability is local. The main result of Part I is a generalization of Theorem 1.6 to rectangle-decomposable modules. We prove it for modules indexed over a product of two totally ordered sets satisfying a weak condition; we conjecture that the result holds for arbitrary totally ordered sets but proving that seems to require a novel set of ideas. The proof of this theorem is already quite involved and will be the subject of Chapter 2.

Theorem 1.7. *Suppose that any interval of X or Y admits a countable coinitial subset. Then, any pfd persistence module M over $X \times Y$ is rectangle-decomposable if and only if the restriction of M to any square is rectangle-decomposable.*

The assumption on the posets is fairly mild. For instance, it is satisfied by arbitrary subsets X, Y of $\mathbb{R}$ endowed with the canonical order. Furthermore, an equivalent formulation of the assumption is that X and Y both admit a countable subset which is dense in their order topology, as it is done in Crawley-Boevey (2015). It can also easily be seen to be equivalent to the hypothesis that any rectangle in $X \times Y$ admits a countable coinitial subset, which is instrumental when considering inverse limits of exact sequences; see Lemma 2.25.

The interest in rectangle modules is currently ramping up among the topological data analysis community, with the realization of their simplicity of use and potential for generalization. Indeed, while they constitute only a small fraction of the indecomposables over $\mathbb{R}^d$, they can serve as a basis for encoding certain invariants of more general classes of persistence modules. For instance, it was proven in Botnan, Oppermann, and Oudot (2021) that the rank invariant of any finitely presentable persistence module M over

$\mathbb{R}^d$ decomposes essentially uniquely as the difference between the rank invariants of two rectangle-decomposable persistence modules. Rectangle modules—or a slight generalization thereof—also appear in projective resolutions in certain exact structures (Blanchette, Brüstle, and Hanson, 2022; Botnan, Oppermann, and Oudot, 2021). These facts give rise to a notion of signed barcode for general classes of persistence modules over $\mathbb{R}^d$. While our work is not directly related to these recent developments, it contributes to the background knowledge on rectangle modules, and it allows us to answer practical questions such as determining whether a persistence module itself—not just its rank invariant—decomposes into rectangle summands.

Interval-decomposability is not characterized locally. We show that interval-decomposability cannot be characterized locally, even when testing on arbitrary strict subgrids.

Theorem (Theorem 3.1)**.** *Suppose X and Y are totally ordered sets with $|X| \geq 3$ and $|Y| \geq 3$, and let $2 \leq m < \min(|X|, |Y|)$ be an integer. Then, there exists a pfd persistence module M over $X \times Y$ that is not interval-decomposable, but for which $M|_Q$ is interval-decomposable for all grids Q of side-lengths at most m.*

Rectangle-decomposability is maximal among local properties. In light of the above results, it is natural to wonder if there is a class of intervals more general than rectangles for which a local characterization over squares is possible. That turns out not to be the case.

Theorem (Theorem 3.5)**.** *Let X and Y be totally ordered sets and let $\mathcal{S} \subseteq \mathrm{Int}(X \times Y)$. Assume further that there exist $\{x_1 < x_2 < x_3\} \subseteq X$ and $\{y_1 < y_2\} \subseteq Y$ such that $\mathcal{S}_{|Q_0} \supsetneq \mathrm{Rec}(Q_0)$ for $Q_0 = \{x_1, x_3\} \times \{y_1, y_2\}$. Then, there exists a pfd persistence module M over $X \times Y$ such that M is not in $\langle \mathcal{S} \rangle$ but the restriction of M to Q is in $\langle \mathcal{S}_{|Q} \rangle$ for all squares Q.*

Chapter 2

Rectangle-decomposability is local

Based on joint work with Magnus Bakke Botnan and Steve Oudot.

This chapter is devoted to the proof of the local characterization of rectangle-decomposability (Theorem 1.7). Our analysis uses the following algebraic characterization of rectangle-decomposability. Recall that $X \times Y$ is a product of two totally ordered sets endowed with the coordinatewise order and that for all persistence modules M over $X \times Y$ and every $s \leq t \in X \times Y$, one has the commutative diagram:

$$\begin{array}{ccc} M_{(s_x,t_y)} & \xrightarrow{\rho^t_{(s_x,t_y)}} & M_t \\ {\scriptstyle \rho_s^{(s_x,t_y)}}\uparrow & & \uparrow{\scriptstyle \rho^t_{(t_x,s_y)}} \\ M_s & \xrightarrow[\rho_s^{(t_x,s_y)}]{} & M_{(t_x,s_y)} \end{array} . \tag{2.1}$$

Definition 2.1 (Weak exactness)**.** A persistence module M over $X \times Y$ is *weakly exact* if, for every $s \leq t \in X \times Y$, the following conditions hold in the commutative diagram (2.1):

$$\operatorname{Im} \rho_s^t = \operatorname{Im} \rho^t_{(t_x,s_y)} \cap \operatorname{Im} \rho^t_{(s_x,t_y)},$$

$$\operatorname{Ker} \rho_s^t = \operatorname{Ker} \rho_s^{(t_x,s_y)} + \operatorname{Ker} \rho_s^{(s_x,t_y)}.$$

This condition is a weakened version of the so-called *exactness* condition (called *strong exactness* in this thesis) that was proven to be equivalent to block-decomposability in Cochoy and Oudot (2020). It is not hard to check that a pfd persistence module over a square is weakly exact. Furthermore, any pfd persistence module indexed over a square is interval-decomposable (see Escolar and Hiraoka 2016, Figure 13) and a direct inspection shows that if the module is weakly exact then these interval summands must be rectangle modules. These observations imply that the following theorem is equivalent to Theorem 1.7:

Theorem 2.2. *Suppose that any interval of X or Y admits a countable coinitial subset. Then, a pfd persistence module M over $X \times Y$ is weakly exact if and only if M is rectangle-decomposable.*

Outline. Since it is easy to verify that rectangle modules are weakly exact and that being weakly exact is invariant under taking locally finite direct sums, one can see that any pfd persistence module that is rectangle-decomposable is also weakly exact. We are left with proving the converse statement. For that, we start by proving the result when X and Y are finite (Section 2.1). The proof of the general case relies on the finite case and is developed in Sections 2.2 through 2.5. We follow the same approach as in Cochoy and Oudot (2020) and detail the adjustments we make in the next paragraph. We begin by introducing functorial filtrations, a key construction in the proof (Section 2.2). Next, we define a submodule M_R of M for each rectangle $R \in X \times Y$. This submodule is referred to as the *rectangle filtrate of M associated to R* (Section 2.3). The construction of M_R ensures that $M_{R,t}$ contains all the elements in M_t whose "lifespan" is exactly R. Specifically, the persistence module M_R is isomorphic to a finite direct sum of copies of $\mathbf{k}_R$. Next, we prove that the sum of these filtrates forms an internal direct sum in M (Section 2.4). Finally, we conclude the proof by proving that the resulting internal direct sum generates M (Section 2.5).

Comparison with the work on block-decomposable modules. The functorial filtration technique for weakly exact persistence bimodules already appeared in Cochoy and Oudot (ibid., Sections 3 and 4). However, their definition of M_R as a submodule of M (ibid., Proposition 5.3) does not work for weakly exact persistence bimodules. That is, the resulting family of vector spaces does not assemble into a submodule. We overcome this difficulty by defining the rectangle filtrate within a carefully constructed weakly exact submodule of M (Definition 2.23). Proving that the resulting sum of rectangle filtrates is an internal direct sum (Proposition 2.37) can be adapted directly from the proof of the analogous result in the block-decomposable case (ibid., Proposition 6.6). By contrast, proving that the rectangle filtrations generate M is more involved, as the work from (ibid., Section 7) does not carry over to our setting. Our approach includes a series of technical lemmas that prove that it suffices to consider the restriction to a certain finite grid (Definition 2.41). Once this is established, the result is a consequence of the structure theorem for finite grids (Theorem 2.3).

2.1 Finite posets

In this section, we assume that $X \times Y$ is finite and give a short proof of Theorem 2.2 in this case. Namely, we prove:

Theorem 2.3. *Suppose that X and Y are finite. Then, a pfd persistence module M over $X \times Y$ is weakly exact if and only if M is rectangle-decomposable.*

We will repeatedly use the following result in our analysis:

Lemma 2.4. *For any indices $k \in [\![1, n]\!]$ and $l \in [\![1, m]\!]$, the indicator module $\mathbf{k}_{[\![1,k]\!]\times[\![1,l]\!]}$ is an injective persistence module over $[\![1, n]\!] \times [\![1, m]\!]$.*

Proof. This lemma is a consequence of Botnan and Crawley-Boevey (2020, Lemma 2.1) since the subset $[\![1, k]\!] \times [\![1, l]\!]$ is clearly a directed ideal of the poset $[\![1, n]\!] \times [\![1, m]\!]$, following the definition of Botnan and Crawley-Boevey (ibid., Section 2.1). □

Recall from classical commutative algebra (see also Section 1.1) that an injective submodule of a persistence module is a summand of it.

Proof of Theorem 2.3. Since X and Y are finite, one can assume that $X = [\![1, n]\!]$ and $Y = [\![1, m]\!]$. Our proof proceeds by induction on the poset of grid dimensions (n, m), also viewed as a subposet of $\mathbb{R}^2$ equipped with the coordinatewise order:

• Our base cases are when $n = 1$ or $m = 1$. The result is then a direct consequence of Gabriel's theorem (1972), which asserts that M decomposes as a direct sum of interval modules, each interval being a rectangle of width 1.

• Fix $n > 1$ and $m > 1$, and assume that the result is true for all $n' \times m'$ grids with $(n', m') < (n, m)$. Fix a persistence module M over $[\![1, n]\!] \times [\![1, m]\!]$ that is pfd and weakly exact. We prove that M decomposes as a direct sum of rectangle bimodules.

Observe that M has finite *total dimension* $\sum_{t\in[\![1,n]\!]\times[\![1,m]\!]} \dim M_t$, so we know from a simple induction that M decomposes as a direct sum of indecomposables—recall the more general statement of Theorem 1.1. As any summand of a weakly exact module is again weakly exact, we may restrict our attention to pfd indecomposable modules.

For the sake of contradiction, assume that M is pfd, weakly exact, indecomposable, and not isomorphic to a rectangle module. Then:

Claim 2.5. The map $\rho^{(n,m)}_{(1,1)}$ is zero.

Proof. Suppose the contrary. Then we have $\operatorname{Ker} \rho^{(n,m)}_{(1,1)} \subsetneq M_{(1,1)}$. Let $\alpha \in M_{(1,1)} \setminus \operatorname{Ker} \rho^{(n,m)}_{(1,1)}$. The submodule N of M spanned by the collection of images:

$$\left(\rho^{(i,j)}_{(1,1)}(\alpha)\right)_{(i,j)\in[\![1,n]\!]\times[\![1,m]\!]}$$

is isomorphic to $\mathbf{k}_{[\![1,n]\!]\times[\![1,m]\!]}$, an injective persistence module by Lemma 2.4. Hence, the persistence module N is a summand of M, contradicting that M is not isomorphic to a rectangle module. □

Claim 2.6. The space $M_{(1,1)}$ maps injectively to the nodes of the grid $[\![1, n-1]\!]\times[\![1, m-1]\!]$.

Proof. Let us restrict M to the grid $[\![1, n-1]\!] \times [\![1, m]\!]$. The restriction—denoted by N—may no longer be indecomposable, however it is still pfd and weakly exact, therefore our induction hypothesis asserts that N decomposes as a finite (internal) direct sum where each summand is isomorphic to some rectangle module. Consider any one of these summands, say $N' \cong \mathbf{k}_{R'}$, such that $(1, 1) \in R'$. Then, we claim that $(n - 1, 1) \in R'$ as well. Indeed, otherwise, one can extend N' to a persistence module over $[\![1, n]\!] \times [\![1, m]\!]$ by putting zero spaces on the last column n. This yields an injective rectangle submodule of M (Lemma 2.4), and therefore a rectangle summand of M—a contradiction.

The above argument shows that the persistence module $M_{(1,1)}$ maps injectively to the nodes $(i, 1)$ for $i \in [\![1, n - 1]\!]$. Similarly, by restricting M to the grid $[\![1, n]\!] \times [\![1, m - 1]\!]$, we deduce that $M_{(1,1)}$ maps injectively to the nodes $(1, j)$ for $j \in [\![1, m - 1]\!]$. Then, by weak exactness, we have for all $(i, j) \in [\![1, n - 1]\!] \times [\![1, m - 1]\!]$:

$$\operatorname{Ker} \rho^{(i,j)}_{(1,1)} = \operatorname{Ker} \rho^{(i,1)}_{(1,1)} + \operatorname{Ker} \rho^{(1,j)}_{(1,1)} = 0,$$

so $M_{(1,1)}$ maps injectively to all the nodes of the grid $[\![1, n-1]\!] \times [\![1, m-1]\!]$. □

Claim 2.7. The spaces $M_{(1,1)}$ and $M_{(n,m)}$ are zero.

Proof. By weak exactness and Claim 2.5, we have

$$M_{(1,1)} = \operatorname{Ker} \rho_{(1,1)}^{(n,m)} = \operatorname{Ker} \rho_{(1,1)}^{(n,1)} + \operatorname{Ker} \rho_{(1,1)}^{(1,m)}.$$

Assuming for a contradiction that $M_{(1,1)} \neq 0$, we have that at least one of the two terms on the right-hand side of the above equation must be non-zero—say for instance $\operatorname{Ker} \rho_{(1,1)}^{(n,1)} \neq 0$. Let $\alpha \neq 0$ be an element in that kernel. By Claim 2.6, its images at the nodes of $[\![1, n-1]\!] \times [\![1, m-1]\!]$ are non-zero. Meanwhile, its images at the nodes of $\{n\} \times [\![1, m]\!]$ are zero, by composition. There are two cases:

- Either $\rho_{(1,1)}^{(1,m)}(\alpha) = 0$, in which case the images of α at the nodes of $[\![1, n]\!] \times \{m\}$ are also zero, which implies that the persistence submodule of M spanned by the images of α is isomorphic to $\mathbf{k}_{[\![1,n-1]\!] \times [\![1,m-1]\!]}$.
- Or $\rho_{(1,1)}^{(1,m)}(\alpha) \neq 0$, in which case, for all $i \in [\![1, n-1]\!]$, we have

  $$\alpha \notin \operatorname{Ker} \rho_{(1,1)}^{(1,m)} \overset{\text{(Claim 2.6)}}{=} \operatorname{Ker} \rho_{(1,1)}^{(1,m)} + \operatorname{Ker} \rho_{(1,1)}^{(i,1)} = \operatorname{Ker} \rho_{(1,1)}^{(i,m)},$$

  which implies that the images of α at the nodes of $[\![1, n-1]\!] \times \{m\}$ are non-zero as well. Hence, the persistence submodule of M spanned by the images of α is isomorphic to $\mathbf{k}_{[\![1,n-1]\!] \times [\![1,m]\!]}$.

In both cases, the persistence submodule of M spanned by the images of α is an injective rectangle module (Lemma 2.4), hence a rectangle summand of M—a contradiction.

By applying vector-space duality pointwise to M, we obtain an indecomposable module DM of the grid $[\![1, n]\!]^{\mathrm{op}} \times [\![1, m]\!]^{\mathrm{op}}$—which is isomorphic to $[\![1, n]\!] \times [\![1, m]\!]$ as a poset. This persistence module is still pfd, and still weakly exact as well since the equations of weak exactness are stable under vector-space duality (kernels become images, sums become intersections, and vice-versa). Hence, by the first part of the proof, one has $DM_{(1,1)} = 0$, i.e the space at node (n, m) of M is zero, hence the result. □

Claim 2.8. The space $M_{(1,m)}$ is zero.

Proof. Assume for a contradiction that $M_{(1,m)} \neq 0$. Call N the restriction of M to the grid $[\![1, n-1]\!] \times [\![1, m]\!]$. By our induction hypothesis, the persistence module N decomposes as a finite (internal) direct sum where each summand is isomorphic to some rectangle module. Since $M_{(1,m)} \neq 0$, at least one of these rectangles contains the node $(1, m)$. Among such rectangles, take one—say $R' = [\![1, i]\!] \times [\![j, m]\!]$—that has lowest lower-left corner, and call N' the corresponding summand of N. Denote by N'' the rest of the internal decomposition of N, i.e. $N = N' \oplus N''$.

First, we claim that $i = n - 1$. Indeed, otherwise we can extend N' to a rectangle persistence submodule $\overline{N}'$ of M by putting zero spaces on the last column n, and N'' to another persistence submodule $\overline{N}''$ by putting the internal spaces of M on the last column, so that $M = \overline{N}' \oplus \overline{N}''$—a contradiction.

Second, we claim that $j \in [\![2, m-1]\!]$. Indeed, one has $j \geq 2$ since by Claim 2.7 we know that $M_{(1,1)} = 0$. Meanwhile, if j were equal to m, then N' would go to zero on the last column of $[\![1, n]\!] \times [\![1, m]\!]$ since $M_{(n,m)} = 0$ by Claim 2.7, and so we could extend N to a rectangle persistence submodule $\overline{N}'$ of M by putting zero spaces on the last column, and N'' to another persistence submodule $\overline{N}''$ by putting the internal spaces of M on the last column, so that $M = \overline{N}' \oplus \overline{N}''$—a contradiction.

Consider now the space $N_{(1,j)} = M_{(1,j)}$, and take a generator α of the subspace $N'_{(1,j)} \cong \mathbf{k}$. By Claim 2.7 we know that the map $\rho_{(1,j)}^{(n,m)}$ is zero, so by weak exactness we have $\alpha = \alpha_h + \alpha_v$ for some $\alpha_h \in \operatorname{Ker} \rho_{(1,j)}^{(n,j)}$ and $\alpha_v \in \operatorname{Ker} \rho_{(1,j)}^{(1,m)}$. We claim that $\alpha_h \notin N''_{(1,j)}$. Indeed, otherwise we would have

$$\rho_{(1,j)}^{(1,m)}(\alpha) = \rho_{(1,j)}^{(1,m)}(\alpha_h) + \rho_{(1,j)}^{(1,m)}(\alpha_v) = \rho_{(1,j)}^{(1,m)}(\alpha_h) \in \rho_{(1,j)}^{(1,m)}\big(N''_{(1,j)}\big) \subseteq N''_{(1,m)},$$

thus contradicting our assumption that $N = N' \oplus N''$ with the support of N' containing $(1, m)$. Likewise, for any node $t \in R'$ we have $\rho_{(1,j)}^{t}(\alpha_h) \notin N''_t$, for otherwise we would get a contradiction from

$$\rho_{(1,j)}^{(t_x,m)}(\alpha) = \rho_{(1,j)}^{(t_x,m)}(\alpha_h) = \rho_t^{(t_x,m)}\left(\rho_{(1,j)}^{t}(\alpha_h)\right) \in \rho_t^{(t_x,m)}\big(N''_t\big) \subseteq N''_{(t_x,m)}.$$

Thus, the persistence submodule N^h of N generated by α_h is isomorphic[1] to N' and its sum with N'' is an internal direct sum in N. We can therefore exchange N' for N^h in the internal decomposition of N. Since N^h is mapped to zero on the last column of $[\![1, n]\!] \times [\![1, m]\!]$, we can extend it to a rectangle persistence submodule $\overline{N}^h$ of M by putting zero spaces on the last column, meanwhile we can extend N'' to another persistence submodule $\overline{N}''$ by putting the internal spaces of M on the last column, so that $M = \overline{N}^h \oplus \overline{N}''$—a contradiction. □

Claim 2.9. The spaces $M_{(1,j)}$ are zero for all $j \in [\![1, m]\!]$.

Proof. The result is already proven[2] for $j = m$ by Claim 2.8. Let then $j \in [\![1, m-1]\!]$. Call N the restriction of M to the grid $[\![1, n]\!] \times [\![1, m-1]\!]$. By our induction hypothesis, the persistence module N decomposes as a finite (internal) direct sum where each summand is isomorphic to some rectangle module. Assuming for a contradiction that some summand N' has a support R' that intersects the first column, we know from Claim 2.8 that N' maps to zero at node $(1, m)$. By composition, the persistence module N' maps to zero as well at the nodes on the last row m. Therefore, as in the proof of Claim 2.8, we can extend N' to a rectangle summand of M by putting zero spaces on row m, thus reaching a contradiction. □

We may now conclude the proof of Theorem 2.3. It follows from Claim 2.9 that M itself is not supported outside the rectangle $R = [\![2, n]\!] \times [\![1, m]\!]$. The induction hypothesis applied to the restriction of M to R implies then that M decomposes as a direct sum of rectangle modules, which raises a contradiction. This concludes the induction step and the proof of Theorem 2.3. □

[1]Note that we do not need to check that α_h goes to zero when leaving R', since by assumption R' reaches row m and, as we saw earlier, one has $i = n - 1$ so R' reaches column $n - 1$ as well.

[2]It is also proven for $j = 1$ by Claim 2.7, although we do not use this fact in the proof.

2.2 Functorial filtration and the counting functor

We now begin the proof of Theorem 2.2. In this section, we recall the definition of the *functorial filtration* that we use to construct our rectangle filtrates in Section 2.3. The functorial filtration is inspired by Ringel (1975), introduced in the one-parameter setting by Crawley-Boevey (2015) and generalized for the two-parameter setting in Cochoy and Oudot (2020). We write out the definitions and results already written in Cochoy and Oudot (ibid.) for completeness, and one may also read (ibid., Example 3.3) for an enlightening explicit computation of the functorial filtration.

Assumption 2.10. In Sections 2.2 through 2.5, we assume that any interval of X or Y admits a countable coinitial subset.

Throughout the section, consider a persistence module N over $X \times Y$, and denote by ρ_s^t the internal morphism $N(s \leq t) : N_s \to N_t$ for any $s \leq t \in X \times Y$.

Cuts. To set up the functorial filtration technique, we need a characterization of rectangles in $X \times Y$ using the notion of cuts. A *cut* c of a totally ordered set X is a partition of X into two (possibly empty) sets (c^-, c^+) such that $x < x'$ for all $x \in c^-$ and $x' \in c^+$. We call c^- the *lower part* and c^+ the *upper part* of c. For instance, the subsets $c^- = (-\infty, 1]$ and $c^+ = (1, +\infty)$ define a cut of $\mathbb{R}$. The following lemma is a direct consequence of the fact that X is totally ordered.

Lemma 2.11. *In a totally ordered set, the set of all cuts c can be totally ordered in two canonical ways: inclusion on the lower part c^-, or inclusion on the upper part c^+. The two orders are opposite from each other.*

One can easily see that any interval I in a totally ordered set $(T, \leq)$ can be written as $I = {}_{\shortmid}c^+ \cap c_{\shortmid}{}^-$ for two cuts ${}_{\shortmid}c$ and $c_{\shortmid}$ of T; see Crawley-Boevey (2015, Section 3). Then, it follows from its definition that any rectangle R of $X \times Y$ can be written as $R = ({}_{\shortmid}c^+ \cap c_{\shortmid}{}^-) \times (\underline{c}^+ \cap \overline{c}^-)$, with two cuts of X called the *left cut* ${}_{\shortmid}c$ and the *right cut* $c_{\shortmid}$, and with two cuts of Y called the *top cut* $\overline{c}$ and the *bottom cut* $\underline{c}$. Moreover, writing a block $B = ({}_{\shortmid}c^+ \cap c_{\shortmid}{}^-) \times (\underline{c}^+ \cap \overline{c}^-)$, one can directly check from the definition of a block that:

- $c_{\shortmid}{}^+ = \overline{c}^+ = \varnothing$, if B is a birth quadrant, or
- ${}_{\shortmid}c^- = \underline{c}^- = \varnothing$, if B is a death quadrant, or
- ${}_{\shortmid}c^- = c_{\shortmid}{}^+ = \varnothing$, if B is a horizontal band, or
- $\underline{c}^- = \overline{c}^+ = \varnothing$, if B is a vertical band.

Pointwise filtration. Consider a rectangle $R = ({}_{\shortmid}c^+ \cap c_{\shortmid}{}^-) \times (\underline{c}^+ \cap \overline{c}^-)$ of $X \times Y$ and let $t \in R$. We start by defining the following subspaces of N_t, called *horizontal contributions* (associated to the left and the right cuts of R) and *vertical contributions*

(associated to the top and the bottom cuts of R) *of* R *in* N:

$$\begin{aligned}
\mathrm{Im}^+_{\restriction c,t}(N) &= \bigcap_{\substack{x\in\restriction c^+\\ x\le t_x}} \mathrm{Im}\,\rho^t_{(x,t_y)} & \mathrm{Im}^-_{\restriction c,t}(N) &= \sum_{x\in\restriction c^-} \mathrm{Im}\,\rho^t_{(x,t_y)}\\
\mathrm{Ker}^+_{c\restriction,t}(N) &= \bigcap_{x\in c\restriction^+} \mathrm{Ker}\,\rho_t^{(x,t_y)} & \mathrm{Ker}^-_{c\restriction,t}(N) &= \sum_{\substack{x\in c\restriction^-\\ x\ge t_x}} \mathrm{Ker}\,\rho_t^{(x,t_y)}\\
\mathrm{Im}^+_{\underline{c},t}(N) &= \bigcap_{\substack{y\in\underline{c}^+\\ y\le t_y}} \mathrm{Im}\,\rho^t_{(t_x,y)} & \mathrm{Im}^-_{\underline{c},t}(N) &= \sum_{y\in\underline{c}^-} \mathrm{Im}\,\rho^t_{(t_x,y)}\\
\mathrm{Ker}^+_{\overline{c},t}(N) &= \bigcap_{y\in\overline{c}^+} \mathrm{Ker}\,\rho_t^{(t_x,y)} & \mathrm{Ker}^-_{\overline{c},t}(N) &= \sum_{\substack{y\in\overline{c}^-\\ y\ge t_y}} \mathrm{Ker}\,\rho_t^{(t_x,y)}
\end{aligned} \tag{2.2}$$

with the convention that $\mathrm{Im}^-_{\underline{c},t}(N) = 0$ when $c^- = \varnothing$ and $\mathrm{Ker}^+_{\underline{c},t}(N) = N_t$ when $c^+ = \varnothing$. See Figure 2.1 for an illustration.

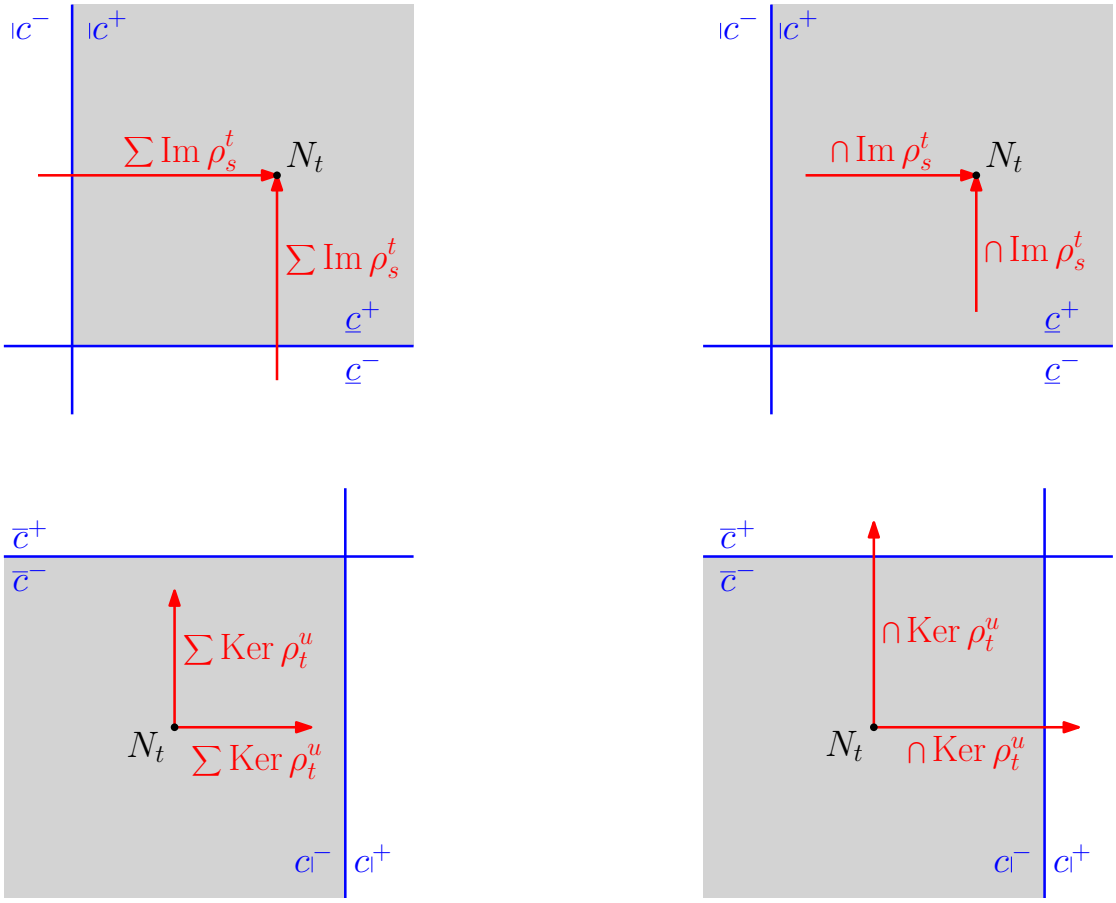


Figure 2.1: From top to bottom and from left to right: the spaces $\mathrm{Im}^-_{c,t}$, $\mathrm{Im}^+_{c,t}$, $\mathrm{Ker}^-_{c,t}$ and $\mathrm{Ker}^+_{c,t}$.

Remark 2.12. In Cochoy and Oudot (2020), the definitions and the results are stated for the poset $X \times Y = \mathbb{R}^2$, but all results hold verbatim (with identical proofs) for a general product $X \times Y$ under Assumption 2.10. From now on, we will therefore cite and use the results of Cochoy and Oudot (ibid.) in our setting.

The following lemma states that when N is pfd, horizontal and vertical contributions can be realized as kernels and images of its internal morphisms.

Lemma 2.13 (Realization; Cochoy and Oudot, 2020, Lemma 3.1)**.** *Assume that N is pfd. For any $t \in R$, one has:*

$$\mathrm{Im}^+_{\restriction c,t}(N) = \mathrm{Im}\,\rho^t_{(x,t_y)} \text{ for some } x \in \restriction c^+ \cap (-\infty, t_x] \text{ and any lower } x \in \restriction c^+,$$

$$\mathrm{Im}^-_{\restriction c,t}(N) = \mathrm{Im}\,\rho^t_{(x,t_y)} \text{ for some } x \in \restriction c^- \cup \{-\infty\} \text{ and any greater } x \in \restriction c^-,$$

$$\mathrm{Ker}^+_{c\restriction,t}(N) = \mathrm{Ker}\,\rho^{(x,t_y)}_t \text{ for some } x \in c\restriction^+ \cup \{+\infty\} \text{ and any lower } x \in c\restriction^+,$$

$$\mathrm{Ker}^-_{c\restriction,t}(N) = \mathrm{Ker}\,\rho^{(x,t_y)}_t \text{ for some } x \in c\restriction^- \cap [t_x, +\infty) \text{ and any greater } x \in c\restriction^-,$$

with the conventions that $\rho^t_{(x,t_y)} : 0 \to M_t$ when $x = -\infty$ and $\rho^{(x,t_y)}_t : M_t \to 0$ when $x = +\infty$. Similar statements hold for vertical cuts.

Convention 2.14. Throughout the rest of the chapter, we keep the conventions on internal morphisms introduced in Lemma 2.13 without further reference to it.

Remark 2.15. The conventions defined in Lemma 2.13 are equivalent to considering the extension $\widetilde{N}$ of N to the poset $\widetilde{X} \times \widetilde{Y}$ with $\widetilde{X} = X \cup \{\pm\infty\}$ and $\widetilde{Y} = Y \cup \{\pm\infty\}$ (and the obvious ordering) such that $\widetilde{N}_{(\pm\infty,\cdot)} = \widetilde{N}_{(\cdot,\pm\infty)} = 0$. This extension is called *extension of N at infinity* in Cochoy and Oudot (ibid.). In this chapter, we only use the following fact: if N is pfd and weakly exact, then so is its extension at infinity. This last fact can be checked by a direct computation.

Remark 2.16. Let $t \in R$ and consider a finite number $N_1, \ldots, N_k$ of pfd persistence modules over $X \times Y$. Then, there exists $x \in \restriction c^+ \cap (-\infty, t_x]$ such that for all $1 \le i \le k$, one has

$$\mathrm{Im}^+_{\restriction c,t}(N_i) = \mathrm{Im}\, N_i\big((x, t_y) \le t\big).$$

Similar remarks hold for other contributions.

We combine horizontal and vertical contributions in the following way:

$$\begin{aligned}
\mathrm{Im}^+_{R,t}(N) &= \mathrm{Im}^+_{\restriction c,t}(N) \cap \mathrm{Im}^+_{\underline{c},t}(N),\\
\mathrm{Im}^-_{R,t}(N) &= \Big(\mathrm{Im}^-_{\restriction c,t}(N) + \mathrm{Im}^-_{\underline{c},t}(N)\Big) \cap \mathrm{Im}^+_{R,t}(N),\\
&= \mathrm{Im}^-_{\restriction c,t}(N) \cap \mathrm{Im}^+_{\underline{c},t}(N) + \mathrm{Im}^-_{\underline{c},t}(N) \cap \mathrm{Im}^+_{\restriction c,t}(N),\\
\mathrm{Ker}^+_{R,t}(N) &= \Big(\mathrm{Ker}^+_{c\restriction,t}(N) + \mathrm{Ker}^-_{\overline{c},t}(N)\Big) \cap \Big(\mathrm{Ker}^-_{c\restriction,t}(N) + \mathrm{Ker}^+_{\overline{c},t}(N)\Big),\\
&= \mathrm{Ker}^+_{c\restriction,t}(N) \cap \mathrm{Ker}^+_{\overline{c},t}(N) + \mathrm{Ker}^-_{c\restriction,t}(N) + \mathrm{Ker}^-_{\overline{c},t}(N),\\
\mathrm{Ker}^-_{R,t}(N) &= \mathrm{Ker}^-_{c\restriction,t}(N) + \mathrm{Ker}^-_{\overline{c},t}(N),
\end{aligned} \tag{2.3}$$

where equalities between formulas come from the inclusions $\mathrm{Im}^-_{c,t}(N) \subseteq \mathrm{Im}^+_{c,t}(N)$, $\mathrm{Ker}^-_{c,t}(N) \subseteq \mathrm{Ker}^+_{c,t}(N)$ and the following elementary lemma.

Lemma 2.17. *Let E be a $\mathbf{k}$-vector space. Let A, B and C three vector subspaces of E such that $A \subseteq C$. Then $(A + B) \cap C = A \cap C + B \cap C$.*

It is immediate from the definitions that $\mathrm{Im}^-_{R,t}(N) \subseteq \mathrm{Im}^+_{R,t}(N)$ and $\mathrm{Ker}^-_{R,t}(N) \subseteq \mathrm{Ker}^+_{R,t}(N)$. This leads us to the following important family of vector spaces.

Definition 2.18. The *functorial filtration of N associated to R* is the following pair of families of vector spaces indexed by $t \in R$:

$$\begin{aligned} V^+_{R,t}(N) &= \mathrm{Im}^+_{R,t}(N) \cap \mathrm{Ker}^+_{R,t}(N), \\ V^-_{R,t}(N) &= \mathrm{Im}^+_{R,t}(N) \cap \mathrm{Ker}^-_{R,t}(N) + \mathrm{Im}^-_{R,t}(N) \cap \mathrm{Ker}^+_{R,t}(N). \end{aligned} \tag{2.4}$$

Since $\mathrm{Im}^-_{R,t}(N) \subseteq \mathrm{Im}^+_{R,t}(N)$ and $\mathrm{Ker}^-_{R,t}(N) \subseteq \mathrm{Ker}^+_{R,t}(N)$, we also have $V^-_{R,t}(N) \subseteq V^+_{R,t}(N)$. If N is interval-decomposable, then it is a small exercise to check that the dimension of the quotient vector space $V^+_{R,t}(N)/V^-_{R,t}(N)$ equals the multiplicity of the summand $\mathbf{k}_R$ in the decomposition.

The following lemma shows that vectors spaces in (2.3) and (2.4) are preserved by the internal morphisms of pfd and weakly exact persistence bimodules.

Lemma 2.19 (Transportation; ibid., Corollary 3.5, Lemma 4.1)**.** *Assume that N is pfd and weakly exact. For any $s \leq t$ in R, we have:*

$$\begin{aligned} \rho^t_s \left(\mathrm{Im}^\pm_{R,s}(N) \right) &= \mathrm{Im}^\pm_{R,t}(N), \\ \left(\rho^t_s\right)^{-1} \left(\mathrm{Ker}^\pm_{R,t}(N) \right) &= \mathrm{Ker}^\pm_{R,s}(N), \\ \rho^t_s \left(V^\pm_{R,s}(N) \right) &= V^\pm_{R,t}(N). \end{aligned}$$

Remark 2.20. Lemma 2.19 ensures that if N is pfd and weakly exact, the families $\left(V^\pm_{R,t}(N)\right)_{t\in R}$ form systems of vector spaces.

Counting functor. A key object in the filtration technique is the *counting functor* associated to a rectangle R.

Definition 2.21 (ibid., Section 4)**.** *The counting functor C_R associated to a rectangle R* is defined for a pfd and weakly exact persistence module N over $X \times Y$ as the inverse limit:

$$C_R(N) := \varprojlim_{t\in R} V^+_{R,t}(N)/V^-_{R,t}(N),$$

where the transition maps are given by the naturally defined quotient maps $\overline{\rho}^t_s : V^+_{R,s}(N)/V^-_{R,s}(N) \to V^+_{R,t}(N)/V^-_{R,t}(N)$.

The *counting functor* owes its name to the following crucial fact.

Lemma 2.22 (ibid., Lemma 4.2)**.** *Assume that N is pfd and rectangle-decomposable. For any rectangle R of $X \times Y$, the multiplicity of the summand $\mathbf{k}_R$ in the rectangle-decomposition of N is given by* $\dim C_R(N)$.

2.3 Definition of the rectangle filtrates

The goal of this section is to define rectangle filtrates (Definition 2.28). In this section and in Sections 2.4 and 2.5, we consider a pfd and weakly exact persistence module M over $X \times Y$, and denote by ρ_s^t the internal morphism $M(s \leq t) : M_s \to M_t$ for any $s \leq t \in X \times Y$. Until the end of this section, we consider a rectangle $R = (\llcorner c^+ \cap c \lrcorner^-) \times (\underline{c}^+ \cap \overline{c}^-)$ of $X \times Y$.

2.3.1 Elements dead above the rectangle

The rectangle filtrate associated to R will be constructed within the submodule of M defined below (Definition 2.23). Consider $R^- = \{t \in X \times Y | \exists s \in R, t \leq s\}$. Note that $R^- = c\lrcorner^- \times \overline{c}^-$, so that the contributions $\mathrm{Ker}^+_{c\lrcorner,t}(M)$ and $\mathrm{Ker}^+_{\overline{c},t}(M)$ are well-defined for any $t \in R^-$.

Definition 2.23. We call *submodule of M of elements dead above R*, and denote by $K_R(M)$, the submodule of M whose spaces at each $t \in X \times Y$ are given by:

$$K_{R,t}(M) = \begin{cases} \mathrm{Ker}^+_{c\lrcorner,t}(M) \cap \mathrm{Ker}^+_{\overline{c},t}(M) & \text{if } t \in R^-, \\ 0 & \text{else.} \end{cases}$$

The fact that $K_R(M)$ yields a well-defined persistence submodule of M is an easy consequence of the definition of horizontal and vertical contributions (2.2). When there is no ambiguity, the submodule $K_R(M)$ is referred to as K_R for readability.

Proposition 2.24. *The persistence module K_R is weakly exact.*

Proof. In this proof we write $\tilde{\rho}_u^v = \rho^v_{u|K_{R,u}}$ for any $u \leq v \in X \times Y$. Let $s \leq t \in X \times Y$ and denote $a = (s_x, t_y)$ and $b = (t_x, s_y)$. Let us first prove the equality:

$$\mathrm{Ker}\,\tilde{\rho}_s^t = \mathrm{Ker}\,\tilde{\rho}_s^a + \mathrm{Ker}\,\tilde{\rho}_s^b. \tag{2.5}$$

Suppose that $t \notin R^-$. Then $\mathrm{Ker}\,\tilde{\rho}_s^t = K_{R,s}$. Moreover, one has $a \notin R^-$ or $b \notin R^-$, so $\mathrm{Ker}\,\tilde{\rho}_s^a = K_{R,s}$ or $\mathrm{Ker}\,\tilde{\rho}_s^b = K_{R,s}$. Hence (2.5) in that case. Now suppose that $t \in R^-$. Then $s \in R^-$, $a \in R^-$ and $b \in R^-$. Therefore, one has $\mathrm{Ker}\,\rho_s^a \subseteq \mathrm{Ker}^+_{\overline{c},s}(M)$ and $\mathrm{Ker}\,\rho_s^b \subseteq \mathrm{Ker}^+_{c\lrcorner,s}(M)$. Using the weak exactness of M and Lemma 2.17 twice, we get:

$$\begin{aligned}
\mathrm{Ker}\,\tilde{\rho}_s^t &= \left(\mathrm{Ker}\,\rho_s^t\right) \cap K_{R,s} \\
&= \left(\mathrm{Ker}\,\rho_s^a + \mathrm{Ker}\,\rho_s^b\right) \cap \mathrm{Ker}^+_{c\lrcorner,s}(M) \cap \mathrm{Ker}^+_{\overline{c},s}(M) \\
&= \left(\mathrm{Ker}\,\rho_s^a \cap \mathrm{Ker}^+_{c\lrcorner,s}(M) + \mathrm{Ker}\,\rho_s^b \cap \mathrm{Ker}^+_{c\lrcorner,s}(M)\right) \cap \mathrm{Ker}^+_{\overline{c},s}(M) \\
&= \mathrm{Ker}\,\rho_s^a \cap K_{R,s} + \mathrm{Ker}\,\rho_s^b \cap K_{R,s} \\
&= \mathrm{Ker}\,\tilde{\rho}_s^a + \mathrm{Ker}\,\tilde{\rho}_s^b.
\end{aligned}$$

Let us now prove the equality $\mathrm{Im}\,\tilde{\rho}_s^t = \mathrm{Im}\,\tilde{\rho}_a^t \cap \mathrm{Im}\,\tilde{\rho}_b^t$, i.e.

$$\rho_s^t(K_{R,s}) = \rho_a^t(K_{R,a}) \cap \rho_b^t(K_{R,b}). \tag{2.6}$$

The inclusion $\rho_s^t(K_{R,s}) \subseteq \rho_a^t(K_{R,a}) \cap \rho_b^t(K_{R,b})$ is clear. Let us show the converse. Applying Lemma 2.13 at the point s and at the point a, one can choose low enough $y \in \overline{c}^+ \cup \{+\infty\}$ such that:

$$\begin{aligned}\mathrm{Ker}^+_{\overline{c},s}(M) &= \mathrm{Ker}\,\rho_s^{(s_x,y)},\\ \mathrm{Ker}^+_{\overline{c},a}(M) &= \mathrm{Ker}\,\rho_a^{(s_x,y)}.\end{aligned}$$

Similarly, there exists $x \in c\rfloor^+ \cup \{+\infty\}$ such that:

$$\mathrm{Ker}^+_{c\rfloor,b}(M) = \mathrm{Ker}\,\rho_b^{(x,s_y)}, \quad \text{and} \quad \mathrm{Ker}^+_{c\rfloor,s}(M) = \mathrm{Ker}\,\rho_s^{(x,s_y)}.$$

The result will follow from repeated use of the weak exactness property of M (and Remark 2.15). The following diagram will help picturing the various spaces involved in this proof. Denote $c = (s_x, y)$, $d = (x, s_y)$, $e = (t_x, y)$ and $f = (x, t_y)$.

$$\begin{array}{ccccc} M_c & \longrightarrow & M_e & & \\ \uparrow & & \uparrow & & \\ M_a & \longrightarrow & M_t & \longrightarrow & M_f \\ \uparrow & & \uparrow & & \uparrow \\ M_s & \longrightarrow & M_b & \longrightarrow & M_d \end{array}$$

Let $z \in K_{R,t}$ be such that there are $z_a \in K_{R,a}$ and $z_b \in K_{R,b}$ such that $z = \rho_a^t(z_a) = \rho_b^t(z_b)$. Since M is weakly exact, there exists $z_s \in M_s$ such that $z = \rho_s^t(z_s)$.

We claim that $z_s \in \mathrm{Ker}\,\rho_s^e$. Indeed, using that $z_a \in \mathrm{Ker}^+_{\overline{c},a}(M) = \mathrm{Ker}\,\rho_a^c$, we get $\rho_s^e(z_s) = \rho_t^e(z) = \rho_a^e(z_a) = \rho_c^e \circ \rho_a^c(z_a) = 0$.

Thus, by weak exactness of M and Remark 2.15, there exist $z' \in \mathrm{Ker}\,\rho_s^b$ and $z'' \in \mathrm{Ker}\,\rho_s^c$ such that $z_s = z' + z''$. This yields that $z = \rho_s^t(z_s) = \rho_s^t(z'')$.

We claim now that $z'' \in \mathrm{Ker}\,\rho_s^f$. Indeed, using that $z_b \in \mathrm{Ker}^+_{c\rfloor,b}(M) = \mathrm{Ker}\,\rho_b^d$, we get $\rho_s^f(z'') = \rho_t^f(z) = \rho_b^f(z_b) = \rho_d^f \circ \rho_b^d(z_b) = 0$.

Thus, by weak exactness of M and Remark 2.15, there exist $\tilde{z}' \in \mathrm{Ker}\,\rho_s^a$ and $\tilde{z}'' \in \mathrm{Ker}\,\rho_s^d$ such that $z'' = \tilde{z}' + \tilde{z}''$. This yields that $z = \rho_s^t(z'') = \rho_s^t(\tilde{z}'')$.

We now claim that $\tilde{z}'' \in K_{R,s}$, which completes the proof. Indeed, on the one hand $\tilde{z}'' \in \mathrm{Ker}\,\rho_s^d = \mathrm{Ker}^+_{c\rfloor,s}(M)$. On the other, one has $\tilde{z}'' = z'' - \tilde{z}' \in \mathrm{Ker}\,\rho_s^c = \mathrm{Ker}^+_{\overline{c},s}(M)$. Hence the result.

□

2.3.2 Rectangle filtrate

Let N be a pfd and weakly exact persistence module over $X \times Y$. Recall the definition of the functorial filtration (Definition 2.18) and the systems of vector spaces they form (Remark 2.20). As in Cochoy and Oudot (2020) and Crawley-Boevey (2015), we consider the following inverse limits:

$$\mathcal{V}_R^{\pm}(N) = \varprojlim_{t \in R} V_{R,t}^{\pm}(N). \tag{2.7}$$

Denoting $\pi_t : \mathcal{V}_R^+(N) \to V_{R,t}^+(N)$ the natural map given by the limit, one has the following identification:

$$\mathcal{V}_R^-(N) = \bigcap_{t \in R} \pi_t^{-1}\left(V_{R,t}^-(N)\right) \subseteq \mathcal{V}_R^+(N). \tag{2.8}$$

This implies that for any $t \in R$, the morphism π_t induces a morphism:

$$\overline{\pi}_t : \mathcal{V}_R^+(N)/\mathcal{V}_R^-(N) \longrightarrow V_{R,t}^+(N)/V_{R,t}^-(N).$$

Lemma 2.25 (Cochoy and Oudot, 2020, Lemma 5.2)**.** *Recall that N is pfd and weakly exact. For $t \in R$, the map $\overline{\pi}_t : \mathcal{V}_R^+(N)/\mathcal{V}_R^-(N) \longrightarrow V_{R,t}^+(N)/V_{R,t}^-(N)$ is an isomorphism.*

The rectangle filtrate of M associated to R will be defined as the submodule of M given by the following proposition (Definition 2.28).

Proposition 2.26. *Let $\mathcal{M}_R$ be a vector space complement of $\mathcal{V}_R^-(K_R)$ in $\mathcal{V}_R^+(K_R)$. For $t \in X \times Y$, consider the vector subspace of $K_{R,t}$ given by:*

$$M_{R,t} := \begin{cases} \pi_t(\mathcal{M}_R) & \text{if } t \in R, \\ 0 & \text{else,} \end{cases}$$

where $\pi_t : \mathcal{V}_R^+(K_R) \longrightarrow V_{R,t}^+(K_R)$ is the natural maps given by the limit for $t \in R$. Then, the family $(M_{R,t})_{t \in X \times Y}$ forms a submodule of K_R (hence of M).

Proof. Let $s \leq t$ in $X \times Y$. Suppose that $s \notin R$. Then, one has $\rho_s^t(M_{R,s}) = 0 \subseteq M_{R,t}$. Now, suppose that s and t both lie in R. By definition of π, one has $\rho_s^t \circ \pi_s = \pi_t$. Thus, one has $\rho_s^t(M_{R,s}) = \rho_s^t(\pi_s(\mathcal{M}_R)) = \pi_t(\mathcal{M}_R) = M_{R,t}$.

Finally, suppose that $s \in R$ and $t \notin R$. We show that $\rho_s^t(M_{R,s}) = 0$. One has $\pi_s(\mathcal{M}_R) \subseteq \mathrm{Ker}_{c\restriction,s}^+ \cap \mathrm{Ker}_{\overline{c},s}^+$, for every $s \in R$. Moreover, for every $t \geq s$ with $s \in R$ and $t \notin R$, we have $t_x \in c\restriction^+$ or $t_y \in \overline{c}^+$, so that $\rho_s^t(\mathrm{Ker}_{c\restriction,s}^+ \cap \mathrm{Ker}_{\overline{c},s}^+) = 0$. Hence:

$$\rho_s^t\left(M_{R,s}\right) = \rho_s^t(\pi_s(\mathcal{M}_R)) \subseteq \rho_s^t(\mathrm{Ker}_{c\restriction,s}^+ \cap \mathrm{Ker}_{\overline{c},s}^+) = 0.$$

□

Remark 2.27. Since we have chosen $\mathcal{M}_R$ such that $\mathcal{V}_R^+(K_R) = \mathcal{M}_R \oplus \mathcal{V}_R^-(K_R)$, for every $t \in R$ we have $V_{R,t}^+(K_R) = M_{R,t} \oplus V_{R,t}^-(K_R)$ by Lemma 2.25.

Definition 2.28 (Rectangle filtrate)**.** Let $\mathcal{M}_R$ be a vector space complement of $\mathcal{V}_R^-(K_R)$ in $\mathcal{V}_R^+(K_R)$. The submodule of M defined in Proposition 2.26 is called a *rectangle filtrate of M associated to R* and denoted by M_R.

Our work on rectangle filtrates will rely on Remark 2.27 and thus will not depend on the choice of vector space complement $\mathcal{M}_R$ of $\mathcal{V}_R^-(K_R)$ in $\mathcal{V}_R^+(K_R)$. The following convention will therefore be used.

Convention 2.29. For each rectangle R' of $X \times Y$, choose a vector space complement $\mathcal{M}_{R'}$ of $\mathcal{V}_{R'}^-(K_R)$ in $\mathcal{V}_{R'}^+(K_R)$. From now on, the rectangle filtrate $M_{R'}$ will refer to the one associated to this choice of $\mathcal{M}_{R'}$.

Note that the axiom of choice is used in the above convention. This is inevitable in order to consider infinite families of rectangle filtrates in Sections 2.4 and 2.5. The following lemma shows that M_R is rectangle-decomposable.

Lemma 2.30. *The persistence module M_R is isomorphic to a direct sum of $\dim C_R(K_R)$ copies of the rectangle module $\mathbf{k}_R$.*

Proof. The proof is a carbon copy of the one of (ibid., Lemma 5.5) and is included here for completeness. Since K_R is pfd and weakly exact (Proposition 2.24), we know from Lemma 2.25 that the morphism $\overline{\pi}_t$ is an isomorphism for any $t \in R$.

Let Γ be a (finite) basis of $\mathcal{M}_R$. For any $\gamma \in \Gamma$, the elements of $\pi_t(\gamma)$ for $t \in R$ are non-zero and they satisfy $\rho_s^t(\pi_s(\gamma)) = \pi_t(\gamma)$ for all $s \leq t$ in R, so they span a submodule $N(\gamma)$ of M_R that is isomorphic to $\mathbf{k}_R$. Now, for all $t \in R$ the family $\{\rho_t(\gamma)\}_{\gamma\in\Gamma}$ is a basis of $M_{R,t}$, so $M_R \simeq \bigoplus_{\gamma\in\Gamma} N(\gamma)$. Finally, the size of the basis Γ is $\dim \mathcal{M}_R = \dim C_R(K_R)$. □

2.3.3 Description of rectangle filtrates in the finitely rectangle-decomposable case

We prove the following lemma showing that rectangle filtrates capture rectangle summands in the particular case of finite rectangle-decompositions. It will be crucial in the proof of Proposition 2.39 in Section 2.5.

Lemma 2.31. *If $M \simeq \bigoplus_{i\in I} \mathbf{k}_{R_i}^{m_i}$ where I is a finite set, the R_i's are pairwise distinct rectangles of $X \times Y$ and the m_i's are positive integers, then one has $M_{R_i} \simeq \mathbf{k}_{R_i}^{m_i}$ for any $i \in I$.*

In fact, this lemma also holds when the rectangle-decomposition is only locally finite but we do not use such a general statement in this thesis. The proof of Lemma 2.31 is postponed to the end of this section. It uses the following lemma that gives an expression of $K_R(M)$ when M is decomposable as a finite direct sum of rectangle modules. Recall the definition of R^- from Section 2.3.1.

Lemma 2.32. *Suppose that $M \simeq \bigoplus_{i\in I} \mathbf{k}_{R_i}$ where I is a finite set and the R_i's are rectangles of $X \times Y$. Then, one has:*

$$K_R(M) \simeq \bigoplus_{\substack{i\in I \\ R_i \subseteq R^-}} \mathbf{k}_{R_i}.$$

Lemma 2.32 is a direct consequence of Lemmas 2.33 to 2.35 below. The first lemma is elementary. It is used in the proof of Lemma 2.34.

Lemma 2.33. *Let E be a $\mathbf{k}$-vector space and E_1, E_2 be two subspaces of E such that $E = E_1 \oplus E_2$. Let A_1, B_1 two subspaces of E_1, and A_2, B_2 two subspaces of E_2. Then,*

$$(A_1 \oplus A_2) \cap (B_1 \oplus B_2) = (A_1 \cap B_1) \oplus (A_2 \cap B_2).$$

Lemma 2.34. *Let N_1 and N_2 be two pfd and weakly exact persistence modules over $X \times Y$. One has $K_R(N_1 \oplus N_2) = K_R(N_1) \oplus K_R(N_2)$.*

Proof. The module N_1 can naturally be seen as a submodule of $N_1 \oplus N_2$ by identification with $N_1 \oplus 0$. Similarly, any submodule of N_2 can naturally be seen as a submodule of $N_1 \oplus N_2$. Therefore, the persistence module $K_R(N_1)$ and $K_R(N_2)$ can as well be naturally seen as submodules of $N_1 \oplus N_2$. We implicitly make these identifications in the rest of the proof. To prove the result, it is thus sufficient to prove the equality $K_R(N_1 \oplus N_2) = K_R(N_1) \oplus K_R(N_2)$ of persistence submodules of $N_1 \oplus N_2$. In other words, it is sufficient to prove that for any $t \in X \times Y$, one has:

$$K_{R,t}(N_1 \oplus N_2) = K_{R,t}(N_1) \oplus K_{R,t}(N_2).$$

Let $t \in X \times Y$. For $t \notin R^-$, both sides of the equality vanish, so let us assume that $t \in R^-$. Denote by ρ the internal morphisms of N_1 and by η the internal morphisms of N_2. The internal morphisms of $N_1 \oplus N_2$ are then given by $\rho \oplus \eta$. By Remark 2.16, there exists $x \in c\!\upharpoonright^+ \cup \{+\infty\}$ such that:

$$\begin{aligned}
\operatorname{Ker}^+_{c\upharpoonright,t}(N_1) &= \operatorname{Ker} \rho_t^{(x,t_y)}, \\
\operatorname{Ker}^+_{c\upharpoonright,t}(N_2) &= \operatorname{Ker} \eta_t^{(x,t_y)}, \\
\operatorname{Ker}^+_{c\upharpoonright,t}(N_1 \oplus N_2) &= \operatorname{Ker} \left(\rho_t^{(x,t_y)} \oplus \eta_t^{(x,t_y)} \right) = \left(\operatorname{Ker} \rho_t^{(x,t_y)} \right) \oplus \left(\operatorname{Ker} \eta_t^{(x,t_y)} \right).
\end{aligned}$$

Hence, one has $\operatorname{Ker}^+_{c\upharpoonright,t}(N_1 \oplus N_2) = \operatorname{Ker}^+_{c\upharpoonright,t}(N_1) \oplus \operatorname{Ker}^+_{c\upharpoonright,t}(N_2)$. Similarly, one can prove that $\operatorname{Ker}^+_{\overline{c},t}(N_1 \oplus N_2) = \operatorname{Ker}^+_{\overline{c},t}(N_1) \oplus \operatorname{Ker}^+_{\overline{c},t}(N_2)$. This implies:

$$\begin{aligned}
K_{R,t}(N_1 \oplus N_2) &= \operatorname{Ker}^+_{c\upharpoonright,t}(N_1 \oplus N_2) \cap \operatorname{Ker}^+_{\overline{c},t}(N_1 \oplus N_2) \\
&= \left(\operatorname{Ker}^+_{c\upharpoonright,t}(N_1) \oplus \operatorname{Ker}^+_{c\upharpoonright,t}(N_2) \right) \cap \left(\operatorname{Ker}^+_{\overline{c},t}(N_1) \oplus \operatorname{Ker}^+_{\overline{c},t}(N_2) \right) \\
&\overset{\text{(Lem. 2.33)}}{=} \left(\operatorname{Ker}^+_{c\upharpoonright,t}(N_1) \cap \operatorname{Ker}^+_{\overline{c},t}(N_1) \right) \oplus \left(\operatorname{Ker}^+_{c\upharpoonright,t}(N_2) \cap \operatorname{Ker}^+_{\overline{c},t}(N_2) \right) \\
&= K_{R,t}(N_1) \oplus K_{R,t}(N_2).
\end{aligned}$$

□

Lemma 2.35. *Let R and R' be two rectangles of $X \times Y$. Then,*

$$K_R(\mathbf{k}_{R'}) = \begin{cases} \mathbf{k}_{R'} & \text{if } R' \subseteq R^-, \\ 0 & \text{else.} \end{cases}$$

Proof. Write $R = (\upharpoonright c^+ \cap c\!\upharpoonright^-) \times (\underline{c}^+ \cap \overline{c}^-)$ and $R' = (\upharpoonright d^+ \cap d\!\upharpoonright^-) \times (\underline{d}^+ \cap \overline{d}^-)$. The persistence module $K_R(\mathbf{k}_{R'})$ is a submodule of $\mathbf{k}_{R'}$, so we only have to check that for any $t \in X \times Y$, one has $K_{R,t}(\mathbf{k}_{R'}) = \mathbf{k}$ if $t \in R'$ and $R' \subseteq R^-$, and $K_{R,t}(\mathbf{k}_{R'}) = 0$ otherwise. Let $t \in X \times Y$. One has $K_{R,t}(\mathbf{k}_{R'}) \subseteq \mathbf{k}_{R',t} = 0$ for any $t \notin R'$, so let us suppose that $t \in R'$.

Suppose that $R' \subseteq R^-$. For any $x \in c\!\upharpoonright^+$, one has $x \in d\!\upharpoonright^+$, so $(x, t_y) \notin R'$ and $\mathbf{k}_{R'}(t \leq (x, t_y)) = 0$. Thus, one has $\operatorname{Ker}^+_{c\upharpoonright,t}(\mathbf{k}_{R'}) = \mathbf{k}_{R',t}$. Similarly, one has $\operatorname{Ker}^+_{\overline{c},t}(\mathbf{k}_{R'}) = \mathbf{k}_{R',t}$, and hence $K_{R,t}(\mathbf{k}_{R'}) = \mathbf{k}_{R',t} = \mathbf{k}$.

Now, suppose that $R' \not\subseteq R^-$. There exists $x \in d\!\upharpoonright^- \setminus c\!\upharpoonright^-$ or $y \in \overline{d}^- \setminus \overline{c}^-$. Say there exists $x \in d\!\upharpoonright^- \setminus c\!\upharpoonright^-$, the other case being similar. Suppose first that $t \notin R^-$. Then, one has $K_{R,t}(\mathbf{k}_{R'}) = 0$ by definition of K_R. Now, suppose that $t \in R^-$. Since $x \notin c\!\upharpoonright^-$, one has $x \geq t_x$. Moreover, since $t \in R'$, one has $t_x \in \upharpoonright d^+$, thus $x \in \upharpoonright d^+$. Therefore,

one has $x \in \lfloor d^+ \cap d\rfloor^-$ and $(x, t_y) \in R'$. Hence $\mathbf{k}_{R'}(t \leq (x, t_y))$ is an isomorphism and $\mathrm{Ker}^+_{\lfloor c\rfloor,t}(\mathbf{k}_{R'}) \subseteq \mathrm{Ker}\, \mathbf{k}_{R'}(t \leq (x, t_y)) = 0$. Therefore, one has $K_{R,t}(\mathbf{k}_{R'}) = 0$.

□

Remark 2.36. Consider the setting of Lemma 2.32. Since $R \subseteq R^-$, Lemma 2.32 implies that the multiplicities of the summand $\mathbf{k}_R$ in the rectangle-decompositions of $K_R(M)$ and of M are the same.

Proof of Lemma 2.31. Let $i \in I$. Lemma 2.30 yields:

$$M_{R_i} \simeq \mathbf{k}_{R_i}^{\dim C_{R_i}\left(K_{R_i}(M)\right)}.$$

Moreover, Lemma 2.32 ensures that $K_{R_i}(M)$ is rectangle-decomposable. Then, Lemma 2.22 ensures that $\dim C_{R_i}(K_{R_i}(M))$ is equal to the multiplicity of the summand $\mathbf{k}_{R_i}$ in the rectangle-decomposition of $K_{R_i}(M)$. In fact, Remark 2.36 implies that this multiplicity is the same as the multiplicity of $\mathbf{k}_{R_i}$ in the rectangle-decomposition of M, which is m_i. Hence the result.

□

2.4 The sum of rectangle filtrates is a direct sum

Recall that M is a pfd and weakly exact persistence bimodule and that its internal morphisms are denoted by $\rho^t_s : M_s \to M_t$ for any $s \leq t \in X \times Y$. Recall also our choice of rectangle filtrates (Convention 2.29). In this section, we prove that the sum of M_R for R ranging over all rectangles in $X \times Y$, is an internal direct sum in M.

Proposition 2.37. *The sum of* $(M_R)_{R:\,\text{rectangle}}$ *is a direct sum.*

The proof of the above result is a straightforward adaptation of the proof of Cochoy and Oudot (2020, Proposition 6.6). As in Cochoy and Oudot (ibid.), we first prove the result when all the rectangles share the same upper right corner, then that we can always reduce to this specific case. The following lemma will be instrumental in the proof.

Lemma 2.38. *Let N be a pfd and weakly exact persistence module over $X \times Y$. Let $R_1 = (\lfloor c_1^+ \cap c\rfloor_1^-) \times (\underline{c}_1^+ \cap \overline{c}_1^-)$ and $R_2 = (\lfloor c_2^+ \cap c\rfloor_2^-) \times (\underline{c}_2^+ \cap \overline{c}_2^-)$ be two rectangles of $X \times Y$ such that $(c_{\lfloor 1}, \overline{c}_1) = (c_{\lfloor 2}, \overline{c}_2)$ and R_1 is a strict subset of R_2. For any $t \in R$, one has $\mathrm{Im}^+_{R_2,t}(N) \subseteq \mathrm{Im}^-_{R_1,t}(N)$.*

Proof. Elementary geometric considerations show that $\lfloor c_2^+ \supseteq \lfloor c_1^+$ and $\underline{c}_2^+ \supseteq \underline{c}_1^+$, and that one of the two inclusions must be strict, i.e. $\lfloor c_2^+ \supsetneq \lfloor c_1^+$ or $\underline{c}_2^+ \supsetneq \underline{c}_1^+$. From the first two inclusions and the definition of horizontal and vertical contributions (2.2), it follows that:

$$\mathrm{Im}^+_{\lfloor c_2,t}(N) \subseteq \mathrm{Im}^+_{\lfloor c_1,t}(N) \text{ and } \mathrm{Im}^+_{\underline{c}_2,t}(N) \subseteq \mathrm{Im}^+_{\underline{c}_1,t}(N). \tag{2.9}$$

From the second two inclusions and the definition of horizontal and vertical contributions, it follows that:

$$\mathrm{Im}^+_{\lfloor c_2,t}(N) \subseteq \mathrm{Im}^-_{\lfloor c_1,t}(N) \text{ or } \mathrm{Im}^+_{\underline{c}_2,t}(N) \subseteq \mathrm{Im}^-_{\underline{c}_1,t}(N). \tag{2.10}$$

From (2.9) and (2.10), we finally deduce:

$$\begin{aligned}\mathrm{Im}^+_{R_2,t}(N) &= \mathrm{Im}^+_{\upharpoonright c_2,t}(N) \cap \mathrm{Im}^+_{\underline{c}_2,t}(N) \\ &\subseteq \mathrm{Im}^+_{\upharpoonright c_1,t}(N) \cap \mathrm{Im}^-_{\underline{c}_1,t}(N) + \mathrm{Im}^-_{\upharpoonright c_1,t}(N) \cap \mathrm{Im}^+_{\underline{c}_1,t}(N) \\ &= \mathrm{Im}^-_{R_1,t}(N)\,.\end{aligned}$$

□

Proof of Proposition 2.37. Let $(R_i)_{i\in[\![1,n]\!]}$ be a finite family of pairwise distinct rectangles, and write $R_i = (\upharpoonright c_i^+ \cap c_i\upharpoonright^-) \times (\underline{c_i}^+ \cap \overline{c_i}^-)$. We show that the sum of the submodules $(M_{R_i})_{i\in[\![1,n]\!]}$ is an internal direct sum in M, i.e that for any $t \in X \times Y$, the sum of the subspaces $(M_{R_i,t})_{i\in[\![1,n]\!]}$ is an internal direct sum in M_t.

Case where all rectangles have the same upper right corner. Suppose that the set $\{(c_1\upharpoonright, \overline{c_1}), \cdots, (c_n\upharpoonright, \overline{c_n})\}$ of upper right corners is a singleton. The proof of Cochoy and Oudot (2020, Proposition 6.6) can be adapted in a straightforward way, replacing the words "birth quadrants" by "rectangles with the same upper right corner". We write the proof here for the sake of completeness.

First, note the equality $K_{R_i}(M) = K_{R_j}(M)$ of submodules of M for all $1 \leq i, j \leq n$. Therefore, we denote this submodule of M simply by K.

It suffices to prove that there is at least one of the R_i's (say R_1) such that the (binary) sum of M_{R_1} and $\sum_{i\neq 1} M_{R_i}$ is a direct sum. Then the result follows from an induction on the size n of the family. Hence, we prove that for any $t \in X \times Y$, we have:

$$M_{R_1,t} \cap \left(\sum_{i\neq 1} M_{R_i,t}\right) = 0.$$

Let $t \in X \times Y$. Since $M_{R_i,t} = 0$ for every i such that $t \notin R_i$, we can assume without loss of generality that $t \in R_i$ for every $i \in [\![1, n]\!]$.

Up to reordering, we can assume that R_1 has the rightmost left cut and, in case of ties, that it also has the topmost bottom cut among the rectangles with the same left cut. Formally, R_1 is the rectangle whose bottom left corner is maximal in the lexicographical order on the set of bottom left corners $\{(\upharpoonright c_1, \underline{c_1}), \cdots, (\upharpoonright c_n, \underline{c_n})\}$ induced by the total order on cuts given by inclusion on their lower parts (Lemma 2.11). It follows that R_1 contains none of the other rectangles. Those can be partitioned into two subfamilies: the ones (say $R_2, \cdots, R_k$) contain R_1 strictly, while the others $(R_{k+1}, \cdots, R_n)$ neither contain R_1 nor are contained in R_1. See Figure 2.2 for an illustration.

For any rectangle R' of $X \times Y$, we denote $\mathrm{Im}^+_{R',t}(K)$ by $\mathrm{Im}^+_{R',t}$ until the end of the proof. For every $i \in [\![2, k]\!]$, Lemma 2.38 implies $\mathrm{Im}^+_{R_i,t} \subseteq \mathrm{Im}^-_{R_1,t}$. Therefore, we obtain:

$$\sum_{i=2}^{k} \mathrm{Im}^+_{R_i,t} \subseteq \mathrm{Im}^-_{R_1,t}\,. \tag{2.11}$$

For every $i \in [\![k+1, n]\!]$, we have $\upharpoonright c_i^+ \supsetneq \upharpoonright c_1^+$ and $\underline{c_i}^+ \subsetneq \underline{c_1}^+$. Let $\tilde{R} = \bigcap_{i=k+1}^n R_i$ — this rectangle neither contains R_1 nor is contained in it. Let now R be the smallest rectangle

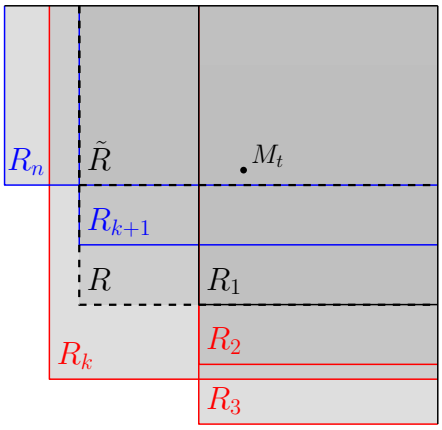


Figure 2.2: Rectangles partitioned into two subfamilies.

containing both R_1 and $\tilde{R}$. We have:

$$\mathrm{Im}^+_{R_1,t} \cap \left(\sum_{i=k+1}^{n} \mathrm{Im}^+_{R_i,t} \right) \subseteq \mathrm{Im}^+_{R_1,t} \cap \mathrm{Im}^+_{\tilde{R},t} = \mathrm{Im}^+_{R,t} \subseteq \mathrm{Im}^-_{R_1,t}, \tag{2.12}$$

where the last inclusion follows from Lemma 2.38 and the fact that R strictly contains R_1. Combining (2.11) and (2.12), we obtain:

$$\begin{aligned} M_{R_1,t} \cap \left(\sum_{i=2}^{k} M_{R_i,t} + \sum_{i=k+1}^{n} M_{R_i,t} \right) &\subseteq \mathrm{Im}^+_{R_1,t} \cap \left(\sum_{i=2}^{k} \mathrm{Im}^+_{R_i,t} + \sum_{i=k+1}^{n} \mathrm{Im}^+_{R_i,t} \right) \\ &= \sum_{i=2}^{k} \mathrm{Im}^+_{R_i,t} + \mathrm{Im}^+_{R_1,t} \cap \left(\sum_{i=k+1}^{n} \mathrm{Im}^+_{R_i,t} \right) \\ &\subseteq \mathrm{Im}^-_{R_1,t}\,. \end{aligned}$$

Meanwhile, Remark 2.27 implies that $M_{R_1,t} \cap V^-_{R_1,t}(K) = 0$. Thus,

$$\mathrm{Im}^-_{R_1,t} \cap M_{R_1,t} = \mathrm{Im}^-_{R_1,t} \cap V^+_{R_1,t}(K) \cap M_{R_1,t} \subseteq V^-_{R_1,t}(K) \cap M_{R_1,t} = 0.$$

Since $M_{R_1,t} \cap (\sum_{i=2}^{n} M_{R_i,t})$ is contained in both $M_{R_1,t}$ and $\mathrm{Im}^-_{R_1,t}$, one has:

$$M_{R_1,t} \cap \left(\sum_{i=2}^{n} M_{R_i,t} \right) \subseteq M_{R_1,t} \cap \mathrm{Im}^-_{R_1,t} = 0.$$

General case. Let $t \in X \times Y$. For each $i \in [\![1,n]\!]$, let $z_i \in M_{R_i,t}$. Denote $z = \sum_{i=1}^{n} z_i$ and suppose that $z = 0$. Let us show that $z_i = 0$ for all $i \in [\![1,n]\!]$. Again, since $z_i = 0$ for every i such that $t \notin R_i$, we can assume without loss of generality that $t \in R_i$ for every $i \in [\![1,n]\!]$.

Order the collection $(c_{1\shortmid}, \overline{c_1}) \preceq \cdots \preceq (c_{n\shortmid}, \overline{c_n})$ of upper right corners by the lexicographical order $\preceq$ induced by the total order on cuts given by inclusion on their lower parts (Lemma 2.11). Let $(d_{1\shortmid}, \overline{d_1}) \prec \cdots \prec (d_{k\shortmid}, \overline{d_k})$ be the distinct elements in the ordered sequence. In particular, the point $(d_{k\shortmid}, \overline{d_k})$ is the upper right corner of rectangles with the rightmost right cut and, in case of ties, the topmost top cut.

Let $J = \{i \in [\![1, n]\!],\ (c_{i!}, \overline{c_i}) = (d_{k!}, \overline{d_k})\}$ and let us show that for all $j \in J$, one has $z_j = 0$. A direct recursion will then yield $z_i = 0$, for all $i \in [\![1, n]\!]$.

By maximality of $(d_{k!}, \overline{d_k})$ in the lexicographical order on upper right corners, there exists $u \in d_{k!}{}^{-} \times \overline{d_k}^{-} \setminus (\bigcup_{l \neq k} (d_{l!}{}^{-} \times \overline{d_l}^{-})$. Therefore, for $j \notin J$, we have $u \notin R_j$, so $M_{R_j,u} = 0$, and thus $\rho_t^u(z_j) = 0$. Hence,

$$0 = \rho_t^u(z) = \sum_{i=1}^{n} \rho_t^u(z_i) = \sum_{j \in J} \rho_t^u(z_j). \tag{2.13}$$

Moreover, for all $j \in J$, we have $t \in R_j$ and $u \in R_j$, thus ρ_t^u restricted to $M_{R_j,t}$ is injective by Lemma 2.30. Therefore, it only remains to show that, for all $j \in J$, one has $\rho_t^u(z_j) = 0$.

Since the rectangles of the family $(R_j)_{j \in J}$ all have same upper right corner, the first case ensures that the sum of $(M_{R_j})_{j \in J}$ is a direct sum. Yet, the element $\rho_t^u(z_j)$ belong to $M_{R_j,u}$ for any $j \in J$, so (2.13) implies $\rho_t^u(z_j) = 0$ for all $j \in J$, which concludes the proof.

□

2.5 Rectangle filtrates cover the module

Recall that M is a pfd and weakly exact persistence bimodule and that its internal morphisms are denoted by $\rho_s^t : M_s \to M_t$ for any $s \leq t \in X \times Y$. Again, recall our choice of rectangle filtrates (Convention 2.29). The goal of this section is to prove the following.

Proposition 2.39. *The (direct) sum of submodules $(M_R)_{R\text{: rectangle}}$ generates M, i.e.,*

$$M = \bigoplus_{R\text{: rectangle}} M_R.$$

To prove the above proposition, we first consider the restriction of M to a specifically constructed finite grid; see Definition 2.41. This restriction captures all information on kernels and images of internal morphisms of M accessible from a fixed index $t \in X \times Y$; see Lemma 2.40. Then, Lemma 2.44 explains how the rectangle filtrates of the restriction relates to the rectangle filtration of M. These two lemmas, in conjunction with our rectangle-decomposition result on finite posets (Theorem 2.3) and the characterization of rectangle filtrates on finite rectangle-decompositions (Lemma 2.31), will prove Proposition 2.39. Theorem 2.2 then follows as a corollary of Proposition 2.39 and Lemma 2.30. The proofs of Lemmas 2.40 and 2.44 have been postponed to Sections 2.5.1 and 2.5.2, respectively.

Lemma 2.40. *Let $t \in X \times Y$. There exist natural numbers L_h, K_h, L_v, K_v and a finite grid*

$$G = (x_i, y_j)_{(i,j) \in [\![-L_h, K_h]\!] \times [\![-L_v, K_v]\!]} \subseteq X \times Y$$

such that:

(i) $t = (x_0, y_0)$,

(ii) *for all* $i \in [\![0, K_h]\!]$, *we have* $\operatorname{Ker} \rho_t^{(x_i,t_y)} \subsetneq \operatorname{Ker} \rho_t^{(x_{i+1},t_y)}$ *and same for vertical kernels where* $x_i = +\infty$ *for* $i = K_h + 1$ *and* $y_j = +\infty$ *for* $j = K_v + 1$ *by convention (recall also Convention 2.14),*

(iii) *for all* $i \in [\![-L_h, 0]\!]$, *we have* $\operatorname{Im} \rho_{(x_{i-1},t_y)}^t \subsetneq \operatorname{Im} \rho_{(x_i,t_y)}^t$ *and same for vertical images where* $x_i = -\infty$ *for* $i = -L_h - 1$ *and* $y_j = -\infty$ *for* $j = -L_v - 1$ *by convention,*

(iv) *for all* $x \in [t_x, +\infty]$, *there exists* $i \in [\![0, K_h + 1]\!]$ *such that* $\operatorname{Ker} \rho_t^{(x,t_y)} = \operatorname{Ker} \rho_t^{(x_i,t_y)}$ *and same for vertical kernels.*

(v) *for all* $x \in [-\infty, t_x]$, *there exists* $i \in [\![-L_h - 1, 0]\!]$ *such that* $\operatorname{Im} \rho_{(x,t_y)}^t = \operatorname{Im} \rho_{(x_i,t_y)}^t$ *and same for vertical images.*

Definition 2.41. Let $t \in X \times Y$. Any finite grid $G \subseteq X \times Y$ given by Lemma 2.40 is called *t-skeleton* of M.

Remark 2.42. Note that the statements (ii) and (iii) ensure that the indices i given in (iv) and (v)—realizing kernels and images of the base module inside the grid—are unique.

Example 2.43. Figure 2.3 illustrates a t-skeleton when M is the direct sum of rectangle modules associated to rectangles R_1, R_2 and R_3.

Lemma 2.44. *Let* $t \in X \times Y$, *let* G *be a* t*-skeleton of* M *and denote* $M^G := M_{|G}$. *To any rectangle* $\widetilde{R} = ({}_{\shortmid}\tilde{c}^+ \cap \tilde{c}_{\shortmid}{}^-) \times (\underline{\tilde{c}}^+ \cap \overline{\tilde{c}}^-)$ *of* G *such that* $t \in \widetilde{R}$, *one can associate a rectangle* $R = ({}_{\shortmid}c^+ \cap c_{\shortmid}{}^-) \times (\underline{c}^+ \cap \overline{c}^-)$ *of* $X \times Y$ *such that:*

(i) $t \in R$,

(ii) *one has:*

$$\operatorname{Ker}^{\pm}_{\tilde{c}_{\shortmid},t}(M^G) = \operatorname{Ker}^{\pm}_{c_{\shortmid},t}(M),$$
$$\operatorname{Im}^{\pm}_{{}_{\shortmid}\tilde{c},t}(M^G) = \operatorname{Im}^{\pm}_{{}_{\shortmid}c,t}(M),$$

and similarly for vertical cuts,

(iii) *the map* $\widetilde{R} \mapsto R$ *is injective.*

In particular,

$$\dim M_{R,t} = \dim M^G_{\widetilde{R},t}.$$

We can now prove Proposition 2.39.

Proof of Proposition 2.39. Let $t \in X \times Y$ and let us show that

$$M_t = \bigoplus_{R\text{: rectangle}} M_{R,t}. \tag{2.14}$$

Take a t-skeleton G of M given by Lemma 2.40, and denote $M^G := M_{|G}$. Notice that, since $t \in G$ by (i) of Lemma 2.40, we have

$$M_t = M^G_t. \tag{2.15}$$

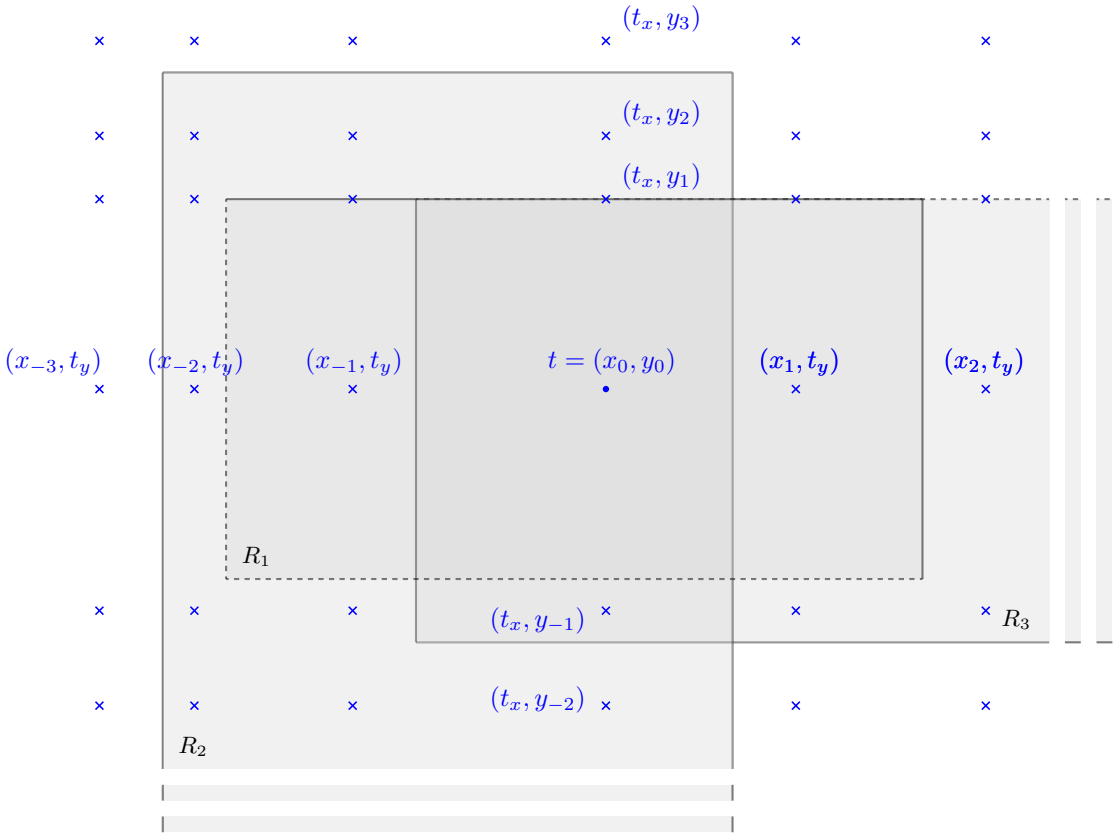


Figure 2.3: Example of a choice of a grid construction as in Lemma 2.40. Dashed lines denote open boundaries and dashed rectangles denote infinite sides. The initial point t is denoted by a dot, while crosses denote points on the constructed grid.

Moreover, since M^G is still pfd and weakly exact, it decomposes as a direct sum of rectangle modules by Theorem 2.3:

$$M^G \simeq \bigoplus_{j \in \tilde{J}} \left(\mathbf{k}^G_{\widetilde{R}_j}\right)^{m_{\widetilde{R}_j}}, \tag{2.16}$$

where the rectangles $\widetilde{R}_j$ are pairwise distinct rectangles of the grid G, and where the integers $m_{\widetilde{R}_j} > 0$ are the multiplicities of the rectangle modules $\mathbf{k}^G_{\widetilde{R}_j}$ in the decomposition. Since G is finite and M^G is pfd, the set $\tilde{J}$ appearing in the decomposition (2.16) is finite. Therefore, Lemma 2.31 implies that for each $j \in \tilde{J}$,

$$\dim\left(M^G_{\widetilde{R}_j, t}\right) = m_{\widetilde{R}_j}. \tag{2.17}$$

Besides, since $t \in G$, we can consider the subset $J := \{j \in \tilde{J} \,|\, t \in \widetilde{R}_j\}$, so that:

$$\dim\left(M^G_t\right) = \sum_{j \in J} m_{\widetilde{R}_j}. \tag{2.18}$$

Meanwhile, writing $\{R_i\}_{i \in I}$ the set of rectangles of $X \times Y$ containing t, Lemma 2.44

yields an injection $\iota : J \hookrightarrow I$, such that

$$\dim M_{R_{\iota(j)},t} = \dim M^G_{\tilde{R}_j,t}. \tag{2.19}$$

We can compute the finite dimensions:

$$\begin{aligned}
\dim(M_t) &\overset{(2.15)}{=} \dim\left(M_t^G\right) \\
&\overset{(2.18)}{=} \sum_{j\in J} m_{\tilde{R}_j} \\
&\overset{(2.17)}{=} \sum_{j\in J} \dim\left(M^G_{\tilde{R}_j,t}\right) \\
&\overset{(2.19)}{=} \sum_{j\in J} \dim\left(M_{R_{\iota(j)},t}\right) \\
&= \dim\left(\bigoplus_{j\in J} M_{R_{\iota(j)},t}\right),
\end{aligned}$$

and conclude by the inclusion $\bigoplus_{j\in J} M_{R_{\iota(j)},t} \subseteq M_t$ that $\bigoplus_{j\in J} M_{R_{\iota(j)},t} = M_t$. Finally, we have

$$M_t = \bigoplus_{j\in J} M_{R_{\iota(j)},t} \subseteq \bigoplus_{R\text{: rectangle}} M_{R,t} \subseteq M_t,$$

which concludes the proof of (2.14). Hence the result. □

2.5.1 Proof of Lemma 2.40

Lemma 2.40 is a direct consequence of finite dimensionality, as was Lemma 2.13. We write its proof here for completeness. Since M is pfd, the function $x \in [t_x, +\infty] \mapsto \dim \operatorname{Ker} \rho_t^{(x,t_y)}$ takes a finite number of values $0 = n_0 < n_1 < \cdots < n_{K_h} < n_{K_h+1} = \dim M_t$. This function is also increasing, so fixing $x_0 = t_x$ and $x_{K_h+1} = +\infty$, we can find real numbers $x_0 < x_1 < \cdots < x_{K_h} < x_{K_h+1}$ such that $\dim \operatorname{Ker} \rho_t^{(x_i,t_y)} = n_i$ for all $0 \le i \le K_h + 1$. We can define similar real numbers for vertical kernels: $t_y = y_0 < y_1 < \cdots < y_{K_v} < y_{K_v+1} = +\infty$ such that $\dim \operatorname{Ker} \rho_t^{(t_x,y_j)} = \tilde{n}_j$ where $(\tilde{n}_j)_{j\in[\![0,K_v+1]\!]}$ are the distinct dimensions of vertical kernels.

Similarly, since M is pfd, one has $x \in [-\infty, t_x] \mapsto \dim \operatorname{Im} \rho^t_{(x,t_y)}$ takes a finite number of values $0 = m_{-L_h-1} < m_{-L_h} < \cdots < m_{-1} < m_0 = \dim M_t$. This function is also increasing, so fixing $x_{-L_h-1} = -\infty$ we can find real numbers $x_{-L_h-1} < x_{-L_h} < \cdots < x_{-1} < x_0 = t_x$ such that $\dim \operatorname{Im} \rho^t_{(x_i,t_y)} = m_i$ for all $i \in [\![-L_h - 1, 0]\!]$. We can define similar real numbers for vertical images: $-\infty = y_{-L_v-1} < y_{-L_v} < \cdots < y_{-1} < t_y$ such that $\dim \operatorname{Im} \rho^t_{(t_x,y_j)} = \tilde{m}_j$ where $(\tilde{m}_j)_{j\in[\![-L_v-1,0]\!]}$ are the distinct dimensions of vertical images.

Define finally the finite grid $G := \{(x_i, y_j), (i,j) \in [\![-L_h, K_h]\!] \times [\![-L_v, K_v]\!]\}$. It remains to show that this grid satisfies the required properties.

First, (i) comes from $x_0 = t_x$ and $y_0 = t_y$.

Second, (ii) and (iii) are clear from the construction of the grid: spaces associated to indices are ordered by inclusion and they are distinct if indices are distinct because then their dimensions are distinct.

Third, (iv) and (v) are also clear from the construction of the grid: every possible horizontal or vertical kernel and image has been represented by an index in the grid.

2.5.2 Proof of Lemma 2.44

Let $t \in X \times Y$, let $G = (x_i, y_j)_{(i,j)\in[\![-L_h,K_h]\!]\times[\![-L_v,K_v]\!]}$ be a t-skeleton of M and denote $M^G := M_{|G}$. The proof of Lemma 2.44 is postponed to the end of this section. It uses the following three lemmas.

Lemma 2.45. *To any cut $\tilde{c}_{\shortmid}$ of $(x_i)_{i\in[\![-L_h,K_h]\!]}$ such that $t_x \in \tilde{c}_{\shortmid}^{-}$, one can associate a cut $c_{\shortmid}$ of X such that:*

(i) $t_x \in c_{\shortmid}^{-}$,

(ii) $\mathrm{Ker}^{\pm}_{\tilde{c}_{\shortmid},t}(M^G) = \mathrm{Ker}^{\pm}_{c_{\shortmid},t}(M)$,

(iii) the map $\tilde{c}_{\shortmid} \mapsto c_{\shortmid}$ is injective.

A similar result holds for vertical cuts and vertical kernels.

Proof. Any cut $\tilde{c}_{\shortmid}$ of $(x_i)_{i\in[\![-L_h,K_h]\!]}$ such that $t_x \in \tilde{c}_{\shortmid}^{-}$ can be denoted by $\tilde{c}_{\shortmid}^{-} = (x_i)_{i\in[\![-L_h,k_h]\!]}$ with $k_h \in [\![0, K_h]\!]$. This implies

$$\mathrm{Ker}^{-}_{\tilde{c}_{\shortmid},t}(M^G) = \mathrm{Ker}\,\rho_t^{(x_{k_h},t_y)}, \tag{2.20}$$

$$\mathrm{Ker}^{+}_{\tilde{c}_{\shortmid},t}(M^G) = \mathrm{Ker}\,\rho_t^{(x_{k_h+1},t_y)} \tag{2.21}$$

with possibly $x_{k_h+1} = +\infty$. Now, define the cut $c_{\shortmid}$ of X by

$$\begin{aligned} c_{\shortmid}^{-} &:= (-\infty, x_{k_h}] \cup \left\{ x \in [x_{k_h}, x_{k_h+1}),\ \mathrm{Ker}\,\rho_t^{(x,t_y)} = \mathrm{Ker}\,\rho_t^{(x_{k_h},t_y)} \right\}, \\ c_{\shortmid}^{+} &:= X \setminus c_{\shortmid}^{-}, \end{aligned} \tag{2.22}$$

and notice that $t_x \in c_{\shortmid}^{-}$, hence (i).

Let us show (ii). By Lemma 2.13 applied to M, we can find $x \in c_{\shortmid}^{-}$ such that $\mathrm{Ker}^{-}_{c_{\shortmid},t}(M) = \mathrm{Ker}\,\rho_t^{(x,t_y)}$. Since $x_{k_h} \in c_{\shortmid}^{-}$, we can even choose x in $c_{\shortmid}^{-} \cap [x_{k_h}, +\infty)$, which implies by definition of $c_{\shortmid}$ that $\mathrm{Ker}\,\rho_t^{(x,t_y)} = \mathrm{Ker}\,\rho_t^{(x_{k_h},t_y)}$. Hence,

$$\mathrm{Ker}^{-}_{c_{\shortmid},t}(M) = \mathrm{Ker}\,\rho_t^{(x,t_y)} = \mathrm{Ker}\,\rho_t^{(x_{k_h},t_y)} = \mathrm{Ker}^{-}_{\tilde{c}_{\shortmid},t}(M^G).$$

Similarly, by Lemma 2.13 we can find $x \in c_{\shortmid}^{+} \cup \{+\infty\}$ such that $\mathrm{Ker}^{+}_{c_{\shortmid},t}(M) = \mathrm{Ker}\,\rho_t^{(x,t_y)}$. Since $x \in c_{\shortmid}^{+} \cup \{+\infty\}$ and $t_x \le x_{k_h} \in c_{\shortmid}^{-}$, we have:

$$\mathrm{Ker}\,\rho_t^{(x_{k_h},t_y)} \subsetneq \mathrm{Ker}\,\rho_t^{(x,t_y)}. \tag{2.23}$$

Moreover, since $x_{k_h+1} \in c_{\shortmid}^{+} \cup \{+\infty\}$, we can lower x if necessary to choose $x \in c_{\shortmid}^{+} \cap (-\infty, x_{k_h+1}]$, and then

$$\mathrm{Ker}\,\rho_t^{(x,t_y)} \subseteq \mathrm{Ker}\,\rho_t^{(x_{k_h+1},t_y)}. \tag{2.24}$$

By definition of a t-skeleton (Lemma 2.40 (iv)), there exists $i \in [\![0, K_h + 1]\!]$ such that $\operatorname{Ker} \rho_t^{(x,t_y)} = \operatorname{Ker} \rho_t^{(x_i,t_y)}$. Therefore, Lemma 2.40 (ii) combined with (2.23) and (2.24) implies:

$$\operatorname{Ker} \rho_t^{(x,t_y)} = \operatorname{Ker} \rho_t^{(x_{k_h+1},t_y)}, \tag{2.25}$$

and it finally follows by (2.21) and (2.25) that

$$\operatorname{Ker}^+_{c\rfloor,t}(M) = \operatorname{Ker} \rho_t^{(x,t_y)} = \operatorname{Ker} \rho_t^{(x_{k_h+1},t_y)} = \operatorname{Ker}^+_{\tilde{c}\rfloor,t}(M^G).$$

Let us now show (iii). Let $\tilde{c}\rfloor \neq \tilde{c}'\rfloor$ be two cuts of G with $t_x \in \tilde{c}\rfloor$ and $t_x \in \tilde{c}'\rfloor$. Write

$$\tilde{c}\rfloor^- = (x_i)_{i\in[\![-L_h,k_h]\!]},$$
$$\tilde{c}'\rfloor^- = (x_i)_{i\in[\![-L_h,k'_h]\!]}.$$

Write also $c\rfloor$ and $c'\rfloor$ the respective cuts associated to $\tilde{c}\rfloor$ and $\tilde{c}'\rfloor$ by the previous construction. Since $\tilde{c}\rfloor \neq \tilde{c}'\rfloor$, the indices delimiting the cuts must differ: say for instance $k_h < k'_h$, the other case being similar. Then, it is clear from the definition (2.22) that $x_{k'_h} \in c'\rfloor^- \setminus c\rfloor^-$, and therefore $c\rfloor \neq c'\rfloor$.

□

A similar result holds for images, as shown by the following lemma:

Lemma 2.46. *To any cut $\lfloor\tilde{c}$ of $(x_i)_{i\in[\![-L_h,K_h]\!]}$ such that $t_x \in \lfloor\tilde{c}^+$ one can associate a cut $\lfloor c$ of X such that:*

(i) $t_x \in \lfloor c^+$,

(ii) $\operatorname{Im}^{\pm}_{\lfloor\tilde{c},t}(M^G) = \operatorname{Im}^{\pm}_{\lfloor c,t}(M)$,

(iii) the map $\lfloor\tilde{c} \mapsto \lfloor c$ is injective.

A similar result holds for vertical cuts and vertical images.

Proof. Let $\lfloor\tilde{c}$ be a cut of $(x_i)_{i\in[\![-L_h,K_h]\!]}$ such that $t_x \in \lfloor\tilde{c}^+$. Write $\lfloor\tilde{c}^+ = (x_i)_{i\in[\![-l_h,K_h]\!]}$ with $-l_h \in [\![-L_h, 0]\!]$. This implies

$$\operatorname{Im}^+_{\tilde{c},t}(M^G) = \operatorname{Im} \rho^t_{(x_{-l_h},t_y)},$$
$$\operatorname{Im}^-_{\tilde{c},t}(M^G) = \operatorname{Im} \rho^t_{(x_{-l_h-1},t_y)}.$$

with possibly $x_{-l_h-1} = -\infty$. We can now define the cut $\lfloor c$ of X by

$$\lfloor c^+ := \left\{x \in (x_{-l_h-1}, x_{-l_h}],\ \operatorname{Im} \rho^t_{(x,t_y)} = \operatorname{Im} \rho^t_{(x_{-l_h},t_y)}\right\} \cup [x_{-l_h}, +\infty),$$
$$\lfloor c^- := X \setminus \lfloor c^+,$$

and notice that $t_x \in \lfloor c^+$. The rest of the proof is symmetric to the one for kernels (Lemma 2.45).

□

Finally, we describe the horizontal and vertical contributions associated to the rectangle R in $K_R(M)$. Recall that we refer to $K_R(M)$ simply as K_R.

Lemma 2.47. *For $t \in R$, we have:*

(i) $\mathrm{Im}^{\pm}_{\restriction c,t}(K_R) = \mathrm{Im}^{\pm}_{\restriction c,t}(M) \cap K_{R,t}$,

(ii) $\mathrm{Ker}^{\pm}_{c\restriction,t}(K_R) = \mathrm{Ker}^{\pm}_{c\restriction,t}(M) \cap K_{R,t}$,

and similar statements for vertical contributions.

Proof. In this proof we write $\tilde{\rho}^v_u = \rho^v_{u|K_{R,u}}$ for any $u \leq v \in X \times Y$. Let us first show (i). We prove the result on $\mathrm{Im}^{-}_{\restriction c,t}(K_R)$, the other one is similar. By Remark 2.16, one can find $x \in \restriction c^- \cup \{-\infty\}$ such that:

$$\begin{aligned} \mathrm{Im}^{-}_{\restriction c,t}(M) &= \mathrm{Im}\, \rho^t_{(x,t_y)}, \\ \mathrm{Im}^{-}_{\restriction c,t}(K_R) &= \mathrm{Im}\, \tilde{\rho}^t_{(x,t_y)} = \rho^t_{(x,t_y)}\left(K_{R,(x,t_y)}\right). \end{aligned}$$

Denote $a = (x, t_y)$. Then, one has:

$$\mathrm{Im}^{-}_{\restriction c,t}(K_R) = \rho^t_a(K_{R,a}) \subseteq \mathrm{Im}\, \rho^t_a = \mathrm{Im}^{-}_{\restriction c,t}(M).$$

Since also $\rho^t_a(K_{R,a}) \subseteq K_{R,t}$, one has $\mathrm{Im}^{-}_{\restriction c,t}(K_R) \subseteq \mathrm{Im}^{-}_{\restriction c,t}(M) \cap K_{R,t}$.

Now, let $z \in \mathrm{Im}^{-}_{\restriction c,t}(M) \cap K_{R,t} = \mathrm{Im}\, \rho^t_a \cap K_{R,t}$. There is $z_a \in M_a$ such that $z = \rho^t_a(z_a)$. Applying Lemma 2.13 at the point t and at the point a, one can choose low enough $x' \in c\restriction^+ \cup \{+\infty\}$ and $y \in \overline{c}^+ \cup \{+\infty\}$ such that:

$$\begin{aligned} K_{R,t} &= \mathrm{Ker}\, \rho^{(t_x,y)}_t \cap \mathrm{Ker}\, \rho^{(x',t_y)}_t, \\ K_{R,a} &= \mathrm{Ker}\, \rho^{(x,y)}_a \cap \mathrm{Ker}\, \rho^{(x',t_y)}_a. \end{aligned}$$

Denote $c = (t_x, y)$, $d = (x', t_y)$, and $b = (x, y)$. The following diagram will help picturing the various spaces involved in this proof.

$$\begin{array}{ccccc} M_b & \longrightarrow & M_c & & \\ \uparrow & & \uparrow & & \\ M_a & \longrightarrow & M_t & \longrightarrow & M_d \end{array}$$

Since $z \in K_{R,t} \subseteq \mathrm{Ker}\, \rho^c_t$, one has $\rho^c_a(z_a) = \rho^c_t(z) = 0$, i.e. $z_a \in \mathrm{Ker}\, \rho^c_a$. Thus, by weak exactness of M and Remark 2.15, there exist $z' \in \mathrm{Ker}\, \rho^b_a$ and $z'' \in \mathrm{Ker}\, \rho^t_a$ such that $z_a = z' + z''$. Moreover, since $z \in K_{R,t} \subseteq \mathrm{Ker}\, \rho^d_t$, one has $\rho^d_a(z_a) = \rho^d_t(z) = 0$, i.e. $z_a \in \mathrm{Ker}\, \rho^d_a$. Then also $z' = z_a - z'' \in \mathrm{Ker}\, \rho^d_a$. Hence $z' \in K_{R,a}$ and

$$z = \rho^t_a(z_a) = \rho^t_a(z') \in \rho^t_a(K_{R,a}).$$

Let us now show (ii). By Remark 2.16, there are $x_\pm \in c\restriction^+ \cup \{+\infty\}$ such that:

$$\begin{aligned} \mathrm{Ker}^{\pm}_{c\restriction,t}(M) &= \mathrm{Ker}\, \rho^{(x_\pm,t_y)}_t, \\ \mathrm{Ker}^{\pm}_{c\restriction,t}(K_R) &= \mathrm{Ker}\, \tilde{\rho}^{(x_\pm,t_y)}_t. \end{aligned}$$

Hence, we get:

$$\mathrm{Ker}^{\pm}_{c\restriction,t}(K_R) = \mathrm{Ker}\, \rho^{(x_\pm,t_y)}_t \cap K_{R,t} = \mathrm{Ker}^{\pm}_{c\restriction,t}(M) \cap K_{R,t}.$$

□

Proof of Lemma 2.44. For each cut composing $\widetilde{R}$ (namely $\imath\tilde{c}$, $\tilde{c}\imath$, $\underline{\tilde{c}}$ and $\overline{\tilde{c}}$), we can apply Lemma 2.45 and Lemma 2.46 to find a cut (respectively $\imath c$, $c\imath$, $\underline{c}$ and $\overline{c}$) satisfying the corresponding wanted equality on kernels or images. Considering the rectangle $R = (\imath c^+ \cap c\imath^-) \times (\underline{c}^+ \cap \overline{c}^-)$ of $X \times Y$ yields (i), (ii) and (iii). For the equality of dimensions of the rectangle filtrates, note that:

$$K_{\widetilde{R},t}\left(M^G\right) = K_{R,t}(M).$$

Moreover, the computations of the filtrations of $K_{\widetilde{R}}(M^G)$ and $K_R(M)$ from Lemma 2.47 combined with (ii) imply:

$$V^{\pm}_{\widetilde{R},t}\left(K_{\widetilde{R}}\left(M^G\right)\right) = V^{\pm}_{R,t}\left(K_R(M)\right).$$

Hence $\dim M_R = \dim M^G_{\widetilde{R}}$.

□

Chapter 3

Non-existence of other local characterizations

Based on joint work with Magnus Bakke Botnan and Steve Oudot.

The results from the previous chapter ensure that rectangle-decomposability is a local property, meaning that it can be assessed by examining restrictions to commutative squares. Given this, a natural question arises: to what extent can interval-decomposability be locally determined when allowing for intervals of shapes that are more general than rectangles?

Outline. In this chapter, we provide two negative answers. In Section 3.1, we show that interval-decomposability itself cannot be characterized locally, even when testing on arbitrary strict subgrids. In Section 3.2, we show that decomposability into a class of interval modules strictly containing rectangle modules cannot be locally determined by means of restrictions to squares. In both cases, our analysis uses a specific indecomposable module on a finite grid which is embedded and extended to the whole poset using Kan extensions.

3.1 Interval-decomposability cannot be characterized locally on subgrids

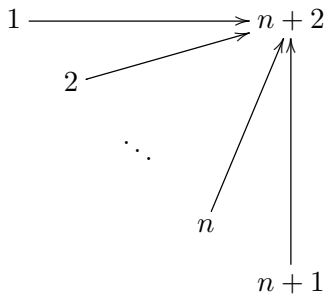
In this section, we prove the following result:

Theorem 3.1. *Suppose X and Y are totally ordered sets with $|X| \geq 3$ and $|Y| \geq 3$, and let $2 \leq m < \min(|X|, |Y|)$ be an integer. Then, there exists a pfd persistence module M over $X \times Y$ that is not interval-decomposable, but for which $M|_Q$ is interval-decomposable for all grids Q of side-lengths at most m.*

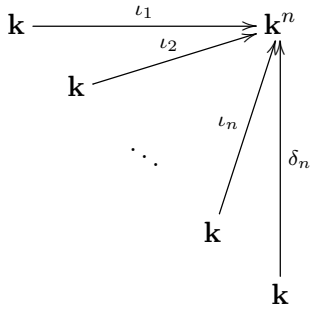
Our analysis proceeds in three steps: first we introduce an indecomposable persistence module which is "minimal" in the sense that any one of its restrictions to a strict subposet is decomposable. Then we extend this indecomposable into another "minimal" indecomposable indexed over the finite grid $[\![1, n+1]\!]^2$. This time, "minimal" means that

any one of its restrictions to a strict subgrid is decomposable. Finally, we extend the analysis to the case of general product posets.

A minimal indecomposable. Let $n \geq 2$ be an integer, and consider the poset $\mathcal{D}_n$ given by the following Hasse diagram:



Denote by ι_i the inclusion of the i-th axis $\mathbf{k} \hookrightarrow \mathbf{k}^n$, and by δ_n the injection into the diagonal $t \in \mathbf{k} \mapsto (t, \dots, t) \in \mathbf{k}^n$. Let M^n denote the persistence module over $\mathcal{D}_n$—which can be easily be seen as a subposet of $\mathbb{R}^2$—given by the following diagram:



Lemma 3.2. *The persistence module M^n satisfies:*

(i) *M^n is indecomposable with local endomorphism ring;*

(ii) *for any $i \in [\![1, n+1]\!]$, the restriction $M^n_{|\mathcal{D}_n \setminus \{i\}}$ decomposes as follows:*

$$M^n_{|\mathcal{D}_n \setminus \{i\}} \cong \bigoplus_{j \in [\![1,n+1]\!] \setminus \{i\}} \mathbf{k}_{\{j,n+2\}},$$

where $\mathbf{k}_{\{j,n+2\}}$ is the indicator module of the set $\{j, n+2\}$.

Proof. It is straightforward to check that M^n has endomorphism ring isomorphic to $\mathbf{k}$, which is local. Therefore, M^n is indecomposable. Now, if $i = n+1$ then the decomposition of $M^n_{|\mathcal{D}_n \setminus \{i\}}$ is obvious, while if $i \in [\![1, n]\!]$ then a simple change of basis in the space M^n_{n+2} yields an isomorphism between $M^n_{|\mathcal{D}_n \setminus \{i\}}$ and $M^n_{|\mathcal{D}_n \setminus \{n+1\}}$ via the identification $\mathcal{D}_n \setminus \{i\} \simeq \mathcal{D}_n \setminus \{n+1\}$, which brings us back to the case where $i = n+1$. □

A minimal indecomposable on a product poset. Given $n \geq 2$, define the following persistence module over $[\![1, n+1]\!]^2$, where dotted lines stand for zero maps

or chains of zero maps, unspecified solid lines stand for identity maps, and dashed lines stand for chains of identity maps:

$$N^n := \quad \begin{array}{ccccccc} \mathbf{k} & \xrightarrow{\iota_1} & \mathbf{k}^n & \longrightarrow & \mathbf{k}^n & \dashrightarrow & \mathbf{k}^n & \longrightarrow & \mathbf{k}^n \\ & & \uparrow{\scriptstyle \iota_2} & & \uparrow & & \uparrow & & \uparrow \\ 0 & \cdots\cdots\to & \mathbf{k} & \xrightarrow{\iota_2} & \mathbf{k}^n & \dashrightarrow & \mathbf{k}^n & \longrightarrow & \mathbf{k}^n \\ \uparrow & & \uparrow & & & \ddots & \uparrow & & \uparrow \\ & & & & & & \mathbf{k}^n & \longrightarrow & \mathbf{k}^n \\ & & & & & & \uparrow{\scriptstyle \iota_n} & & \uparrow \\ 0 & \cdots\cdots\to & 0 & \cdots\cdots\cdots\cdots\cdots\to & & & \mathbf{k} & \xrightarrow{\iota_n} & \mathbf{k}^n \\ \uparrow & & \uparrow & & & & \uparrow & & \uparrow{\scriptstyle \delta_n} \\ 0 & \cdots\cdots\to & 0 & \cdots\cdots\cdots\cdots\cdots\to & & & 0 & \cdots\cdots\to & \mathbf{k} \end{array} \tag{3.1}$$

In more abstract terms, defining the monomorphism of posets:

$$\varphi : \begin{array}{ccc} \mathcal{D}_n & \hookrightarrow & [\![1, n+1]\!]^2 \\ i & \mapsto & \begin{cases} (i, n+2-i) & (i \neq n+2) \\ (n+1, n+1) & (i = n+2) \end{cases} \end{array}, \tag{3.2}$$

we have:

$$N^n \simeq \operatorname{Ran}_\varphi M^n. \tag{3.3}$$

This can be easily seen from the description of right Kan extensions of persistence modules as "floor" modules (Botnan and Lesnick, 2018, Section 2.5). More precisely, we have for all $t \in [\![1, n+1]\!]^2$:

$$N^n_t = \varprojlim M^n_{|\varphi_{\geq t}}, \tag{3.4}$$

where $\varphi_{\geq t}$ denotes the upset $\{u \in \mathcal{D}_n \,|\, \varphi(u) \geq t\}$. Internal morphisms for $s \leq t$ in $[\![1, n+1]\!]^2$ are given by the universality of limits.

To prove that the persistence module N^n defined above is indecomposable, we first give a proof that left (or right) Kan extensions along fully faithful monomorphism of posets preserves endomorphism rings of persistence modules.

Lemma 3.3 (Folklore)**.** *Let $\varphi : P \hookrightarrow Q$ be a fully faithful monomorphism of posets and M be a pfd persistence module over P. Then, the endomorphism ring of $\operatorname{Lan}_\varphi M$ (resp. of $\operatorname{Ran}_\varphi M$) is isomorphic to that of M.*

Proof. We prove the result for left Kan extensions, the case of right Kan extensions being similar. Since φ is fully faithful, we have by Mac Lane (1972, Corollary X.3.3) that:

$$\varphi^* (\operatorname{Lan}_\varphi M) \simeq M, \tag{3.5}$$

where φ^* denotes the functor "pre-composition by φ" going from the category of pfd persistence modules over Q to the category of pfd persistence modules over P. Moreover, the universality property of Kan extensions gives a natural isomorphism:

$$\operatorname{Hom}(\operatorname{Lan}_\varphi M, -) \simeq \operatorname{Hom}(M, \varphi^*(-)). \tag{3.6}$$

Combining these two equations, we get:

$$\begin{aligned}\mathrm{End}(M) &= \mathrm{Hom}(M, M)\\ &\overset{(3.5)}{\simeq} \mathrm{Hom}(M, \varphi^*(\mathrm{Lan}_\varphi M))\\ &\overset{(3.6)}{\simeq} \mathrm{Hom}(\mathrm{Lan}_\varphi M, \mathrm{Lan}_\varphi M)\\ &= \mathrm{End}(\mathrm{Lan}_\varphi M).\end{aligned}$$

□

Now, we can prove that N^n is a "minimal" indecomposable.

Proposition 3.4. *For $n \geq 2$, the persistence module N^n satisfies:*

(i) *N^n is not interval-decomposable;*

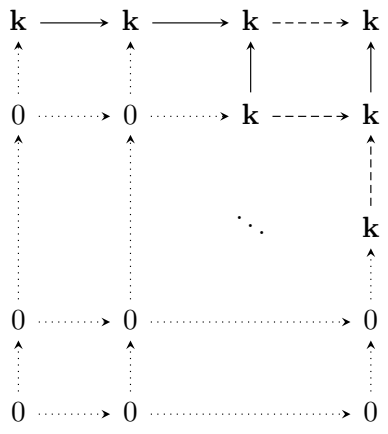
(ii) *for any strict subgrid $X' \times Y' \subsetneq [\![1, n+1]\!]^2$, the restriction $N^n_{|X'\times Y'}$ is interval-decomposable.*

Proof. The monomorphism φ being fully faithful, Lemma 3.3 implies that the endomorphism ring of N^n is isomorphic to that of M^n, which is local by Lemma 3.2. Hence, N^n is indecomposable, and since it is not of pointwise dimension 0 or 1, it is not an interval module. This proves (i). For (ii), since any strict subgrid $X' \times Y'$ of $[\![1, n+1]\!]^2$ misses at least one row or one column of $[\![1, n+1]\!]^2$, we will merely show that the restriction of N^n to $[\![1, n+1]\!]^2 \setminus C$, where C denotes an arbitrary column of $[\![1, n+1]\!]^2$, is interval-decomposable. Indeed, the result for $[\![1, n+1]\!]^2 \setminus R$ where R is a row of $[\![1, n+1]\!]^2$ is obtained analogously, and then the result for the restriction of N^n to any strict subgrid $X' \times Y'$ follows by restriction.

A column C of $[\![1, n+1]\!]^2$ contains exactly one point of the form $(i, n+2-i)$ for an $i \in [\![1, n+1]\!]$. We denote by C_i the column containing the point $(i, n+2-i)$. Hence, the corestriction of $\varphi_{|\mathcal{D}_n \setminus \{i\}}$ to $[\![1, n+1]\!]^2 \setminus C_i$ is well-defined. Denote this corestriction by $\varphi_i : \mathcal{D}_n \setminus \{i\} \hookrightarrow [\![1, n+1]\!]^2 \setminus C_i$. We can easily see that:

$$N^n_{|[\![1,n+1]\!]^2 \setminus C_i} \simeq \mathrm{Ran}_{\varphi_i}\left(M^n_{|\mathcal{D}_n \setminus \{i\}}\right). \tag{3.7}$$

Moreover, for any $j \in [\![1, n+1]\!] \setminus \{i\}$, the module $\mathrm{Ran}_{\varphi_i} \mathbf{k}_{\{j,n+2\}}$ is clearly an interval module, and in particular the finite direct sum $\bigoplus_{j \in [\![1,n+1]\!] \setminus \{i\}} \mathrm{Ran}_{\varphi_i} \mathbf{k}_{\{j,n+2\}}$ is pointwise-finite dimensional. For instance, $\mathrm{Ran}_{\varphi_{n+2}} \mathbf{k}_{\{1,n+2\}}$ is isomorphic to:



Therefore, using Lemma 3.2 and the fact that pfd right Kan extensions commute with direct sums of pfd modules (Botnan and Lesnick, 2018, Remark 2.16), we get an interval-decomposition:

$$\mathrm{Ran}_{\varphi_i}\left(M^n_{|\mathcal{D}_n\setminus\{i\}}\right) \simeq \bigoplus_{j\in[\![1,n+1]\!]\setminus\{i\}} \mathrm{Ran}_{\varphi_i}\mathbf{k}_{\{j,n+2\}},$$

hence the result by (3.7). □

Proof of Theorem 3.1. Suppose that $X \times Y$ is a product of two totally ordered sets such that $|X| \geq 3$ and $|Y| \geq 3$, and let m be an integer such that $2 \leq m < \min(|X|, |Y|)$. In this setting, there are poset inclusions $[\![1, m+1]\!] \hookrightarrow X$ and $[\![1, m+1]\!] \hookrightarrow Y$, and we can consider their product $\psi : [\![1, m+1]\!]^2 \hookrightarrow X \times Y$. We extend the indecomposable module N^m from (3.1) to a persistence module over $X \times Y$ by taking its left Kan extension M along ψ. The resulting persistence module is simply a "ceiling" module (ibid., Section 2.5). Specifically, for all $t \in X \times Y$ we have:

$$M_t = \varinjlim N^m_{|\psi_{\leq t}} \simeq \begin{cases} N^m_{\max(\psi_{\leq t})} & \text{if } \psi_{\leq t} \neq \varnothing, \\ 0 & \text{otherwise,} \end{cases} \tag{3.8}$$

where $\psi_{\leq t}$ denotes the downset $\{u \in [\![1, m+1]\!]^2 \mid \psi(u) \leq t\}$. Similarly, the internal morphisms of m are either trivial, or they correspond to internal morphisms N^m.

From Proposition 3.4 (i) it follows that M is not interval-decomposable. Indeed, if it were interval-decomposable, then its restriction to $[\![1, m+1]\!]^2$ would be as well. However, this restriction is precisely N^m, contradicting that N^m is not interval-decomposable. It is not hard to check that $M_{|Q}$ is interval-decomposable for any finite grid $Q \subseteq X \times Y$ of side-lengths at most m.

3.2 Rectangle-decomposability is maximal among local properties

Recall the notations introduced in Section 1.2 for classes of interval-decomposable modules. In this section, we prove that rectangle-decomposability is "maximal" among local decomposability properties in the following sense:

Theorem 3.5. *Let X and Y be totally ordered sets and let $\mathcal{S} \subseteq \mathrm{Int}(X \times Y)$. Assume further that there exist $\{x_1 < x_2 < x_3\} \subseteq X$ and $\{y_1 < y_2\} \subseteq Y$ such that $\mathcal{S}_{|Q_0} \supsetneq \mathrm{Rec}(Q_0)$ for $Q_0 = \{x_1, x_3\} \times \{y_1, y_2\}$. Then, there exists a pfd persistence module M over $X \times Y$ such that M is not in $\langle\mathcal{S}\rangle$ but the restriction of M to Q is in $\langle\mathcal{S}_{|Q}\rangle$ for all squares Q.*

To do so, we will identify an interval in $\mathcal{S}$ that is not a rectangle. Then, we construct a persistence module M over $X \times Y$ from this interval. Finally, we prove that M is not interval-decomposable despite satisfying $M_{|Q} \in \langle\mathcal{S}_{|Q}\rangle$ for every square $Q \subseteq X \times Y$; see Proposition 3.6. Of course, reversing the role of X and Y, one obtains a result for X and Y two totally ordered sets with $|X| \geq 2$, $|Y| \geq 2$ and $(|X|, |Y|) \neq (2, 2)$.

Intervals of a square. For $s \leq t$ in $X \times Y$, let $a := s$, $b := (s_x, t_y)$, $c := (t_x, s_y)$ and $d := t$. In other words, Q_s^t is precisely the square $\{a, b, c, d\}$ of $X \times Y$. The set of intervals of Q_s^t is then:

$$\mathrm{Int}(\mathrm{Q}_s^t) = \{\{a\}, \{b\}, \{c\}, \{d\}, \{a, b\}, \{a, c\}, \{b, d\}, \{c, d\}, \{a, b, c\}, \{b, c, d\}, \{a, b, c, d\}\}.$$

Of these intervals, two are not rectangles: the *bottom hook* and *top hook*:

$$h_1(\mathrm{Q}_s^t) = \{a, b, c\}, \qquad h_2(\mathrm{Q}_s^t) = \{b, c, d\}.$$

Hence, for a square Q of $X \times Y$, the condition $\mathcal{S}_{|Q} \supsetneq \mathrm{Rec}(Q)$ says precisely that $h_1(Q) \in \mathcal{S}_{|Q}$ or $h_2(Q) \in \mathcal{S}_{|Q}$.

An interval in $\mathcal{S}$ that is not a rectangle. Consider the setting of Theorem 3.5 and denote:

$$G := \{(x_i, y_j)\}_{(i,j) \in \{1,2,3\} \times \{1,2\}} \subseteq X \times Y,$$

so that $Q_0 = \mathrm{Q}_{(x_1,y_1)}^{(x_3,y_2)}$ is the outermost square of G. By assumption, we have $\mathcal{S}_{|Q_0} \supsetneq \mathrm{Rec}(Q_0)$. Therefore, either $h_1(Q_0) \in \mathcal{S}_{|Q_0}$ or $h_2(Q_0) \in \mathcal{S}_{|Q_0}$. One can assume without loss of generality that $h_2(Q_0) \in \mathcal{S}_{|Q_0}$, the other case being dual. By definition of $\mathcal{S}_{|Q_0}$, there is some interval $S \in \mathcal{S}$ such that $h_2(Q_0) = S \cap Q_0$. In particular, we have $(x_1, y_1) \notin S$ while (x_1, y_2), (x_3, y_1) and (x_3, y_2) are in S, thus S is not a rectangle.

Building the counter-example. Consider the following partition of $\mathrm{conv}(G)$, which is the convex hull of G in $X \times Y$ (i.e., the set of points $z \in X \times Y$ such that $(x_1, y_1) \leq z \leq (x_3, y_2)$):

$$\begin{aligned}
P_1 &:= \big(\{x_1\} \times [y_1, y_2]\big) \cap S, \\
P_0 &:= \big(\{x_1\} \times [y_1, y_2]\big) \setminus P_1, \\
P_3 &:= \big((x_1, x_3) \times (y_1, y_2]\big) \cap S, \\
P_2 &:= \big((x_1, x_3) \times [y_1, y_2]\big) \setminus P_3, \\
P_4 &:= \{x_3\} \times [y_1, y_2].
\end{aligned} \tag{3.9}$$

See Figure 3.1 (left) for a graphical representation of this partition, and Table 3.1 for a summary of the comparability of the various sets in the partition.

Consider the subposet $P = G \setminus \{(x_3, y_1)\}$ of $X \times Y$ whose Hasse diagram is:

$$\begin{array}{ccccc}
(x_1, y_2) & \longrightarrow & (x_2, y_2) & \longrightarrow & (x_3, y_2) \\
\uparrow & & \uparrow & & \\
(x_1, y_1) & \longrightarrow & (x_2, y_1) & &
\end{array}, \tag{3.10}$$

and define a persistence module $\widetilde{M}$ over P by the following diagram:

$$\begin{array}{ccccc}
\mathbf{k} & \xrightarrow{\binom{1}{1}} & \mathbf{k}^2 & \xrightarrow{(1\ 0)} & \mathbf{k} \\
\uparrow & & \uparrow{\scriptstyle \binom{1}{0}} & & \\
0 & \longrightarrow & \mathbf{k} & &
\end{array}.$$

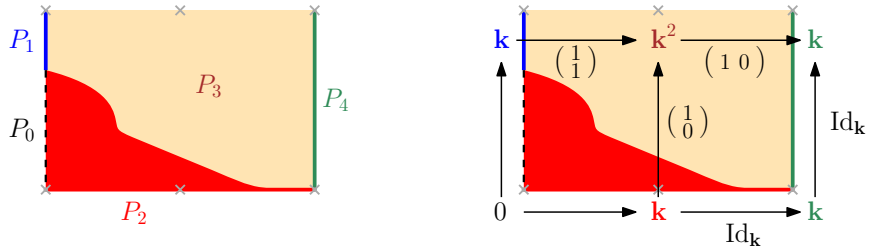


Figure 3.1: A graphical representation of the partition of the convex hull $\mathrm{conv}(G)$ (left), superimposed with its associated module M (right). The regions P_0, P_1, P_2, P_3, P_4 of $\mathrm{conv}(G)$ are represented respectively by the black dashed line segment, the blue segment, the red region (including the bottom red segment), the orange region and the green segment. The nodes of the grid G are represented as gray crosses.

$s \in$ \ $t \in$	P_0	P_1	P_2	P_3	P_4
P_0	$\leq$	$\leq$	$\square$	$\square$	$\square$
P_1		$\leq$		$\square$	$\square$
P_2			$\square$	$\square$	$\square$
P_3				$\square$	$\square$
P_4					$\leq$

Table 3.1: Summary of the comparability of the sets partitioning $\mathrm{conv}(G)$ defined in (3.9). For $0 \leq i, j \leq 4$, an empty cell indicates that there is no $s \in P_i$ and $t \in P_j$ such that $s \leq t$. On the contrary, a symbol $\leq$ indicates that there is such s and t, and a symbol $\square$ refines this last case by indicating that such s and t can in addition (though it is not necessary) satisfy $s_x < t_x$ and $s_y < t_y$, or in other words that there exists a non-degenerate square of $X \times Y$ with bottom-left corner in P_i and top-right corner in P_j. The correctness of this table is clear.

For any $t \in \mathrm{conv}(G)$, call $\pi(t)$ the unique $i \in [\![0, 4]\!]$ such that $t \in P_i$. A direct inspection—eased by Table 3.1—yields that $\pi : \mathrm{conv}(G) \to P$ is a poset morphism. Therefore, one can define the persistence module M over $\mathrm{conv}(G)$ as the *pullback* of $\widetilde{M}$ along π, that is, for any $s \leq t$ in $\mathrm{conv}(G)$:

$$\begin{aligned} M_t &:= \widetilde{M}_{\pi(t)}, \\ M(s \leq t) &:= \widetilde{M}(\pi(s) \leq \pi(t)). \end{aligned} \tag{3.11}$$

We consider in fact the extension of M to $X \times Y$, still denoted by M, with internal spaces set to be zero outside $\mathrm{conv}(G)$ and its internal morphisms to be the obvious ones. See Figure 3.1 (right) for a graphical representation of M. Theorem 3.5 follows from the next proposition:

Proposition 3.6. *The persistence module M satisfies:*

(i) M is not interval-decomposable;

(ii) $M_{|Q} \in \langle \mathcal{S}_{|Q} \rangle$ *for any square* $Q \subseteq X \times Y$.

Proof. We first show that M is not interval-decomposable. Let $\theta \in \mathrm{End}(M)$. For $s \leq t$ in $\mathrm{conv}(G)$ such that $\pi(s) = \pi(t)$, i.e located in the same set P_i, we have that $M(s \leq t) = \mathrm{Id}_{M_t}$ by definition, so the naturality of θ yields a commutative square:

$$\begin{array}{ccc} M_s & \xrightarrow{\mathrm{Id}} & M_t \\ {\scriptstyle\theta_s}\downarrow & & \downarrow{\scriptstyle\theta_t} \\ M_s & \xrightarrow{\mathrm{Id}} & M_t \end{array},$$

and $\theta_s = \theta_t$ in that case. Moreover, the module M vanishes outside $\mathrm{conv}(G)$, so does θ. Thus, any $\theta \in \mathrm{End}(M)$ is entirely determined by its values on the subposet P of $X \times Y$ defined by (3.10). Since $M_{|P}$ is isomorphic to $\widetilde{M}$, which has an endomorphism ring isomorphic to $\mathbf{k}$ by a direct verification, the persistence bimodule M itself has endomorphism ring isomorphic to $\mathbf{k}$, which is local, hence M is indecomposable. Since it is not of pointwise dimension 0 or 1 either, it is not interval-decomposable.

We now prove that the restriction $M_{|Q}$ to any square Q of $X \times Y$ belongs to $\langle \mathcal{S}_{|Q} \rangle$. By assumption, we have $\mathcal{S}_{|Q} \supseteq \mathrm{Rec}(Q)$, so for $M_{|Q}$ to belong to $\langle \mathcal{S}_{|Q} \rangle$ it is sufficient (though not necessary) that $M_{|Q}$ be rectangle-decomposable. Note also that Q can be written as $Q = \mathrm{Q}_s^t$, for two points s and t in $X \times Y$. Since degenerate squares yield one-parameter persistence modules, which are known to be interval-decomposable, we are left with the case where $s_x < t_x$ and $s_y < t_y$.

Assume first that $s \notin \mathrm{conv}(G)$ or $t \notin \mathrm{conv}(G)$. We claim that $M_{|\mathrm{Q}_s^t}$ is rectangle-decomposable in this case. Indeed, as any other pfd representation of the square, $M_{|\mathrm{Q}_s^t}$ is interval-decomposable, and it is then sufficient to prove that the interval summands of $M_{|\mathrm{Q}_s^t}$ cannot be hooks. Assuming without loss of generality that $t \notin \mathrm{conv}(G)$ (the other case being similar), we have that at least one point among (s_x, t_y) and (t_x, s_y) does not belong to $\mathrm{conv}(G)$, for otherwise we would have $x_1 \leq t_x \leq x_3$ and $y_1 \leq t_y \leq y_2$ hence $t \in \mathrm{conv}(G)$. Thus, $M_{|\mathrm{Q}_s^t}$ has at least two zero internal spaces, which implies that its interval summands cannot be hooks. This proves our claim, and so $M_{|\mathrm{Q}_s^t} \in \langle \mathcal{S}_{|\mathrm{Q}_s^t} \rangle$.

Assume now that both s and t are in $\mathrm{conv}(G)$. Several cases are to be considered, corresponding to the cells containing the symbol $\square$ in Table 3.1:

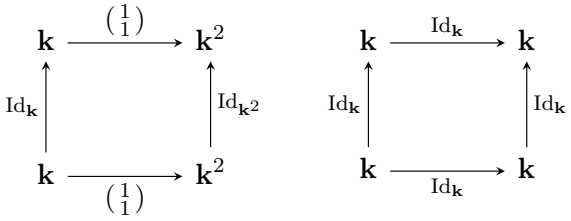
Case $s \in P_0$.

- If $(s_x, t_y) \in P_0$, then $M_s = M_{(s_x,t_y)} = 0$ and no hooks can appear in the interval-decomposition of $M_{|\mathrm{Q}_s^t}$, which is therefore rectangle-decomposable.
- If $(s_x, t_y) \in P_1$, then $M_{|\mathrm{Q}_s^t}$ is of one of the three forms:

$$\begin{array}{ccc} \mathbf{k} & \xrightarrow{\binom{1}{1}} & \mathbf{k}^2 \\ \uparrow & & \uparrow{\scriptstyle\binom{1}{0}} \\ 0 & \longrightarrow & \mathbf{k} \end{array} \qquad \begin{array}{ccc} \mathbf{k} & \xrightarrow{\binom{1}{1}} & \mathbf{k}^2 \\ \uparrow & & \uparrow{\scriptstyle\mathrm{Id}_{\mathbf{k}^2}} \\ 0 & \longrightarrow & \mathbf{k}^2 \end{array} \qquad \begin{array}{ccc} \mathbf{k} & \xrightarrow{\mathrm{Id}_{\mathbf{k}}} & \mathbf{k} \\ \uparrow & & \uparrow{\scriptstyle\mathrm{Id}_{\mathbf{k}}} \\ 0 & \longrightarrow & \mathbf{k} \end{array} \tag{3.12}$$

which happen when $t \in P_3$ for the first two with $(t_x, s_y) \in P_2$ for the first and $(t_x, s_y) \in P_3$ for the second, and when $t \in P_4$ for the last one. The first one is rectangle-decomposable. For the last two, we have $s \notin S$ (since $s \in P_0$) while the points (s_x, t_y), (t_x, s_y) and t are in S, hence $h_2(\mathrm{Q}^t_s) \in \mathcal{S}_{|\mathrm{Q}^t_s}$. Since the modules in (3.12) are clearly interval-decomposable with interval summands being rectangles or top hooks, we do have that $M_{|\mathrm{Q}^t_s}$ belongs to $\langle \mathcal{S}_{|\mathrm{Q}^t_s} \rangle$.

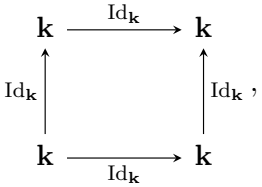
Case $s \in P_1$. The restriction $M_{|\mathrm{Q}^t_s}$ is then of one of the following forms:



which happen respectively when $t \in P_3$ for the first and $t \in P_4$ for the second. They are both clearly rectangle-decomposable.

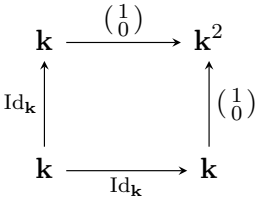
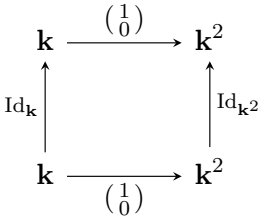
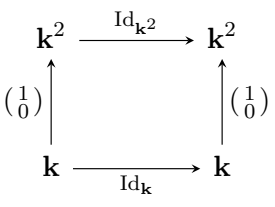
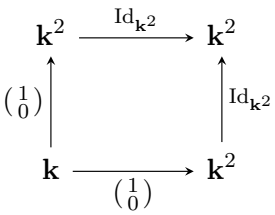
Case $s \in P_2$.

- If $t \in P_2$, then $M_{|\mathrm{Q}^t_s}$ is of the form:



  which is clearly rectangle-decomposable.

- If $t \in P_3$, then $M_{|\mathrm{Q}^t_s}$ is of one of the following forms:

	$(t_x, s_y) \in P_2$	$(t_x, s_y) \in P_3$
$(s_x, t_y) \in P_2$		
$(s_x, t_y) \in P_3$		

which are all rectangle-decomposable except when (s_x, t_y) and (t_x, s_y) are both in P_3 where a top hook summand appears. In that case, $s \notin S$. In fact, $s \in P_2$ with $s_y \neq y_1$ since $(t_x, s_y) \in P_3$. Meanwhile, the points (s_x, t_y), (t_x, s_y) and t are in S. Hence, $h_2(\mathrm{Q}_s^t) \in \mathcal{S}_{|\mathrm{Q}_s^t}$ and we do have $M_{|\mathrm{Q}_s^t} \in \langle \mathcal{S}_{|\mathrm{Q}_s^t} \rangle$.

- If $t \in P_4$, then $M_{|\mathrm{Q}_s^t}$ is of one of the following forms:

$$\begin{array}{ccc} \mathbf{k} & \xrightarrow{\mathrm{Id}_{\mathbf{k}}} & \mathbf{k} \\ {\scriptstyle \mathrm{Id}_{\mathbf{k}}}\uparrow & & \uparrow{\scriptstyle \mathrm{Id}_{\mathbf{k}}} \\ \mathbf{k} & \xrightarrow[\mathrm{Id}_{\mathbf{k}}]{} & \mathbf{k} \end{array} \qquad \begin{array}{ccc} \mathbf{k}^2 & \xrightarrow{(1\ 0)} & \mathbf{k} \\ {\scriptstyle \binom{1}{0}}\uparrow & & \uparrow{\scriptstyle \mathrm{Id}_{\mathbf{k}}} \\ \mathbf{k} & \xrightarrow[\mathrm{Id}_{\mathbf{k}}]{} & \mathbf{k} \end{array}$$

which happen respectively when $(s_x, t_y) \in P_2$ and $(s_x, t_y) \in P_3$ and are both rectangle-decomposable.

Case $s \in P_3$. Then $M_{|\mathrm{Q}_s^t}$ is of one of the following forms:

$$\begin{array}{ccc} \mathbf{k}^2 & \xrightarrow{\mathrm{Id}_{\mathbf{k}^2}} & \mathbf{k}^2 \\ {\scriptstyle \mathrm{Id}_{\mathbf{k}^2}}\uparrow & & \uparrow{\scriptstyle \mathrm{Id}_{\mathbf{k}^2}} \\ \mathbf{k}^2 & \xrightarrow[\mathrm{Id}_{\mathbf{k}^2}]{} & \mathbf{k}^2 \end{array} \qquad \begin{array}{ccc} \mathbf{k}^2 & \xrightarrow{(1\ 0)} & \mathbf{k} \\ {\scriptstyle \mathrm{Id}_{\mathbf{k}^2}}\uparrow & & \uparrow{\scriptstyle \mathrm{Id}_{\mathbf{k}}} \\ \mathbf{k}^2 & \xrightarrow[(1\ 0)]{} & \mathbf{k} \end{array}$$

which happen respectively when $t \in P_3$ and $t \in P_4$ and are both rectangle-decomposable.

Thus, we have shown that M is indecomposable, while $M_{|Q} \in \langle \mathcal{S}_{|Q} \rangle$ for every square Q of $X \times Y$. This concludes the proof. □

Chapter 4

Applications to topological data analysis

Based on joint work with Magnus Bakke Botnan and Steve Oudot.

In this chapter, we explore two applications of our rectangle-decomposition result to topological data analysis. First, we provide an algorithm to check whether a module induced in homology by a two-parameter filtration is rectangle-decomposable, and to decompose it in the affirmative (Section 4.1). Second, we show how rectangle-decomposable modules arise from real-valued functions on a topological space (Section 4.2).

4.1 Checking rectangle-decomposability and computing decompositions

Persistence modules are typically obtained by applying homology to a filtered topological space. In this section, we study the persistence modules produced by the application of homology to finite two-parameter simplicial filtrations, and we provide algorithms for determining whether the persistence module is rectangle-decomposable, and if so, for decomposing it. Our ability to locally characterize rectangle-decomposability will be critical to the algorithm determining decomposability. Furthermore, our decomposition algorithm employs an inclusion-exclusion formula to calculate the multiplicities of rectangle summands from the rank invariant, which also allows us to establish the completeness of the rank invariant within the class of rectangle-decomposable two-parameter persistence modules. This fact is not true for general two-parameter persistence modules, as shown by the following example:

$$\begin{array}{ccccc} \mathbf{k} & \xrightarrow{\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)} & \mathbf{k}^2 & \xrightarrow{(1\ 0)} & \mathbf{k} \\ {\scriptstyle 0}\uparrow & & \uparrow{\scriptstyle \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)} & & \uparrow{\scriptstyle 1} \\ 0 & \xrightarrow{0} & \mathbf{k} & \xrightarrow{1} & \mathbf{k} \end{array} \qquad \begin{array}{ccccc} \mathbf{k} & \xrightarrow{\left(\begin{smallmatrix}1\\1\end{smallmatrix}\right)} & \mathbf{k}^2 & \xrightarrow{(1\ 0)} & \mathbf{k} \\ {\scriptstyle 0}\uparrow & & \uparrow{\scriptstyle \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)} & & \uparrow{\scriptstyle 1} \\ 0 & \xrightarrow{0} & \mathbf{k} & \xrightarrow{1} & \mathbf{k} \end{array} \tag{4.1}$$

The persistence module to the left can easily be seen to be composed of two interval summands in the 3×2 grid. By contrast, the persistence module to the right is indecomposable. However, the two modules have the same rank invariant.

Outline. We start by introducing simplicial filtrations in Section 4.1.1. Then, we prove the completeness result of the rank invariant and state our inclusion-exclusion formula in Section 4.1.2. Next, we describe an algorithm to compute the rank invariant from a finite two-parameter simplicial filtration in Section 4.1.3. Together with our inclusion-exclusion formula, it provides a fast algorithm to compute rectangle-decompositions of rectangle-decomposable modules. Finally, we describe an algorithm to check rectangle-decomposability of persistence modules induced in homology from finite two-parameter simplicial filtrations in Section 4.1.4.

4.1.1 Simplicial filtrations

In this section, we set our notations on simplicial complexes and filtrations. We also give examples of filtrations used in applied contexts.

A *(finite) abstract simplicial complex* $\mathcal{K}$, or simply a *simplicial complex*, is a finite collection of non-empty finite sets that is closed under taking subsets. An element $\sigma \in \mathcal{K}$ is called a *simplex*, and subsets of σ are called *faces* of σ. The inclusion between simplices induces a partial order on $\mathcal{K}$ that we denote simply by $\leq$.

Until the end of this section, we let $\mathcal{K}$ be a finite simplicial complex and $X_1 \times \ldots \times X_m \subseteq \mathbb{R}^m$ be considered as a poset for the coordinatewise order. A simplicial *m-parameter filtration* of $\mathcal{K}$ indexed over $X_1 \times \ldots \times X_m$ is a family $\mathcal{F} = (\mathcal{F}_t)_{t \in X_1 \times \ldots \times X_m}$ of subcomplexes $\mathcal{F}_t \subseteq \mathcal{K}$ that is increasing with respect to inclusions, i.e., such that $\mathcal{F}_t \subseteq \mathcal{F}_{t'}$ for any $t \leq t' \in X_1 \times \ldots \times X_m$. While the term *filtration* usually refers more generally to an increasing family of topological spaces, we assume for convenience in the rest of the chapter that filtrations are simplicial filtrations, that is, composed of simplicial subcomplexes. From now on, we do not refer explicitly to $\mathcal{K}$ when it is clear from the context.

Many filtrations can be introduced by considering sublevel sets of functions:

Example 4.1. Let $f : \mathcal{K} \to \mathbb{R}^m$ be a non-decreasing map, i.e., such that $f(\sigma) \leq f(\tau)$ for any $\sigma \leq \tau \in \mathcal{K}$. The map f induces an m-parameter filtration of $\mathcal{K}$ called *sublevel-sets filtration*, denoted by $\mathcal{F}_f$, and formed by the subcomplexes $(\mathcal{F}_f)_t = \{f \leq t\} := \{\sigma \in \mathcal{K}\,;\, f(\sigma) \leq t\}$ for any $t \in \mathbb{R}^m$. We sometimes refer to the function f as the *filtration function* of $\mathcal{F}_f$.

A lot of information on the geometry of a point cloud is captured by its Čech complex:

Example 4.2. Let $\mathbb{X} \subseteq \mathbb{R}^d$ be finite. The *Čech complex at scale* $r \geq 0$ is the simplicial complex $\check{\mathcal{C}}_r(\mathbb{X})$ defined as follows. For $(x_0, \ldots, x_k) \in \mathbb{X}^{k+1}$, the simplex $\{x_0, \ldots, x_k\}$ is in $\check{\mathcal{C}}_r(\mathbb{X})$ if the intersection of closed balls $\cap_{l=0}^{k} \overline{B}(x_l, r)$ is non-empty. The *filtered Čech complex*, or *Čech filtration*, is defined at each $r \in \mathbb{R}$ as the Čech complex at scale r for $r \geq 0$, and as the empty set for $r < 0$.

A common technique in multi-parameter persistence is to couple the Čech complex with some function on the data. Typically, one copes with outliers by coupling a Čech filtration with a density estimator built from the data at hand; see Vipond et al. (2021) for instance. This falls under the framework of function-Rips (Carlsson and Zomorodian, 2009; Lesnick and Wright, 2016) or function-Čech filtrations:

Example 4.3. Let $\mathbb{X} \subseteq \mathbb{R}^d$ be finite and $f = (f_1, \dots, f_m) : \mathbb{X} \to \mathbb{R}^m$ be a bounded function. The *function-Čech filtration* of $\mathbb{X}$ associated to f is the $(m+1)$-parameter filtration $\check{\mathcal{C}}(\mathbb{X}, f)$ of $2^{\mathbb{X}}$ defined for $r \in \mathbb{R}$ and $t = (t_1, \dots, t_m) \in \mathbb{R}^m$ by:

$$\check{\mathcal{C}}(\mathbb{X}, f)_{(r,t)} = \left\{\sigma \in \check{\mathcal{C}}_r(\mathbb{X})\,;\, \sigma \subseteq f_i^{-1}(-\infty, t_i],\, \forall 1 \leq i \leq m\right\}.$$

Let $\mathcal{F}$ be an m-parameter filtration and $\sigma \in \mathcal{K}$. The *support* of σ is the set $\mathrm{supp}(\sigma) := \{t \in X_1 \times \ldots \times X_m\,;\, \sigma \in \mathcal{F}_t\}$. A filtration is called *one-critical* if the support of any simplex σ is of the form:

$$\mathrm{supp}(\sigma) = Q_{t(\sigma)} := \{t \in X_1 \times \ldots \times X_m\,;\, t \geq t(\sigma)\},$$

for some $t(\sigma) \in X_1 \times \ldots \times X_m$. For instance, function-Čech filtrations are one-critical. On the contrary, the degree-Rips bifiltration is not (Lesnick and Wright, 2016). Note that sublevel-sets filtrations are one-critical. Conversely, any one-critical filtration is a sublevel-sets filtration for the function $f : \sigma \in \mathcal{K} \mapsto t(\sigma)$.

Applying the p-th simplicial homology functor to an m-parameter filtration $\mathcal{F}$ yields a pfd persistence module indexed over $X_1 \times \ldots \times X_m$ called *p-th persistent homology* of $\mathcal{F}$ and denoted by $H_p(\mathcal{F})$.

4.1.2 Completeness of the rank invariant

The rank invariant is proven to be a complete invariant of one-parameter persistence modules in Carlsson and Zomorodian (2009). This can be easily verified for pfd persistence modules indexed over a finite set $[\![1, n]\!]$. In this setting, the following inclusion-exclusion formula, also known as the *persistence measure* (Chazal, De Silva, et al., 2016; Cohen-Steiner, Edelsbrunner, and Harer, 2007), gives the multiplicity $m(s,t)$ of any interval $[\![s, t]\!]$ in the barcode of a persistence module M in terms of its rank invariant $r : [\![1, n]\!]^2 \to \mathbb{N}_0$:

$$m(s,t) = r(s,t) - r(s-1,t) - r(s,t+1) + r(s-1,t+1). \tag{4.2}$$

In this section, we adapt this inclusion-exclusion formula to the two-parameter setting. Namely, we prove the following:

Theorem 4.4. *Suppose that X and Y are subsets of $\mathbb{Z}$. The isomorphism type of any pfd rectangle-decomposable persistence module M over $X \times Y$ is fully determined by the rank invariant of M.*

Proof. Without loss of generality, one can assume that $X \times Y \subseteq \mathbb{Z}^2$ is a product of (potentially infinite) integer intervals $\{n_x, \dots, N_x\} \times \{n_y, \dots, N_y\}$ for some integers $-\infty \leq n_x \leq N_x \leq +\infty$ and $-\infty \leq n_y \leq N_y \leq +\infty$.

We start by computing the multiplicity $m(s,t)$ of the summand $\mathbf{k}_{[\![s_x,t_x]\!]\times[\![s_y,t_y]\!]}$ in the decomposition of M for all indices $s \leq t \in X \times Y$. Fix arbitrary indices $s \leq t \in X \times Y$. Recall that the rank of $(A \oplus B)(s \leq t)$ is equal to the sum of the ranks of $A(s \leq t)$ and $B(s \leq t)$. Meanwhile, for any summand $\mathbf{k}_R$ of M, the rank of $\mathbf{k}_R(s \leq t)$ is 1 if $s, t \in R$ and 0 otherwise. Therefore, $r(s,t)$ counts (with multiplicity) the number of summands of M whose rectangle support contains both s and t.

Then, denoting by $m(s,t^+)$ the number of (rectangle) summands whose support contains t and has s as lower-left corner, we have the following inclusion-exclusion formula:

$$m(s,t^+) = r(s,t) - r((s_x-1,s_y),t) - r((s_x,s_y-1),t) + r((s_x-1,s_y-1),t). \quad (4.3)$$

This formula can be interpreted as follows: a rectangle containing t has s as lower-left corner if and only if it contains s but neither (s_x-1,s_y) nor (s_x,s_y-1); and it contains both (s_x-1,s_y) and (s_x,s_y-1) if and only if it contains (s_x-1,s_y-1).

Using the same approach at t, we can now compute the number $m(s,t)$ of summands of M whose support has s as lower-left corner and t as upper-right corner (i.e., is the rectangle $[\![s_x,t_x]\!] \times [\![s_y,t_y]\!]$). The corresponding inclusion-exclusion formula is:

$$\begin{aligned} m(s,t) \;=\; & m(s,t^+) - m(s,(t_x+1,t_y)^+) \\ & - m(s,(t_x,t_y+1)^+) + m(s,(t_x+1,t_y+1)^+). \end{aligned} \quad (4.4)$$

Combining (4.3) and (4.4) together gives the desired inclusion-exclusion formula for the multiplicity $m(s,t)$ of the summand $\mathbf{k}_{[\![s_x,t_x]\!]\times[\![s_y,t_y]\!]}$ in the decomposition of M from the rank invariant, namely:

$$\begin{aligned} m(s,t) = {} & r(s,t) - r\big((s_x-1,s_y),t\big) \\ & - r\big((s_x,s_y-1),t\big) + r\big((s_x-1,s_y-1),t\big) \\ & - r\big(s,(t_x+1,t_y)\big) + r\big((s_x-1,s_y),(t_x+1,t_y)\big) \\ & + r\big((s_x,s_y-1),(t_x+1,t_y)\big) - r\big((s_x-1,s_y-1),(t_x+1,t_y)\big) \\ & - r\big(s,(t_x,t_y+1)\big) + r\big((s_x-1,s_y),(t_x,t_y+1)\big) \\ & + r\big((s_x,s_y-1),(t_x,t_y+1)\big) - r\big((s_x-1,s_y-1),(t_x,t_y+1)\big) \\ & + r\big(s,(t_x+1,t_y+1)\big) - r\big((s_x-1,s_y),(t_x+1,t_y+1)\big) \\ & - r\big((s_x,s_y-1),(t_x+1,t_y+1)\big) + r\big((s_x-1,s_y-1),(t_x+1,t_y+1)\big). \end{aligned} \quad (4.5)$$

From now on, we can assume that M does not contain any summand with finite support in its decomposition. Again, let $s \leq t \in X \times Y$. In that case, the multiplicity of the summand $\mathbf{k}_{[\![s_x,+\infty]\!]\times[\![s_y,t_y]\!]}$ is simply equal to $m(s,(s_x,t_y)^+) - m(s,(s_x,t_y+1)^+)$, that is, the number of summand of M whose support R has s as lower-left corner, contains (s_x,t_y) but does not contain (s_x,t_y+1). Indeed, such a rectangle R must project onto $[\![s_y,t_y]\!]$ via the canonical projection $X \times Y \to Y$ and thus onto $[\![s_x,+\infty]\!]$ via the projection $X \times Y \to X$ since M does not contain any summand with finite support. Similarly, the multiplicity of the summand $\mathbf{k}_{[\![s_x,t_x]\!]\times[\![s_y,+\infty]\!]}$ in the decomposition of M is equal to $m(s,(t_x,s_y)^+) - m(s,(t_x+1,s_y)^+)$.

From now on, we can assume that M does not contain any summand of the form $\mathbf{k}_{[\![s_x,t_x]\!]\times[\![s_y,+\infty]\!]}$ or $\mathbf{k}_{[\![s_x,+\infty]\!]\times[\![s_y,t_y]\!]}$ in its decomposition. Then, for any $s \in X \times Y$, the multiplicity of the summand $\mathbf{k}_{[\![s_x,+\infty]\!]\times[\![s_y,+\infty]\!]}$ in the decomposition of M is equal to $m(s,(s_x+1,s_y+1)^+)$.

Proceeding similarly, one can count the multiplicities of the summands of the form $\mathbf{k}_{[\![-\infty,t_x]\!]\times[\![s_y,t_y]\!]}$, $\mathbf{k}_{[\![s_x,t_x]\!]\times[\![-\infty,t_y]\!]}$ and $\mathbf{k}_{[\![-\infty,t_x]\!]\times[\![-\infty,t_y]\!]}$ in the decomposition of M.

The number of remaining summands is the multiplicity of the summand $\mathbf{k}_{\mathbb{Z}^2}$. □

4.1.3 Computing the rank invariant and rectangle-decompositions

Let $\mathcal{F}$ be a two-parameter filtration of a simplicial complex $\mathcal{K}$ indexed over a finite grid. Assume that $\mathcal{K}$ has n simplices in total. Without loss of generality, one can assume that $\mathcal{F}$ is indexed over the grid $G = [\![1,n]\!] \times [\![1,n]\!]$, for any larger indexing grid must contain inclusions that are equalities, and any smaller grid can be enlarged by repeating some spaces of the filtration. Assume further that each inclusion $\mathcal{F}_{(i,j)} \subseteq \mathcal{F}_{(i+1,j)}$ or $\mathcal{F}_{(i,j)} \subseteq \mathcal{F}_{(i,j+1)}$ is either an equality or the insertion of a single simplex. We also fix a homology degree p. In this section, we prove the following:

Theorem 4.5. *Given the above input, the rank invariant of the p-th persistent homology of $\mathcal{F}$ can be computed in $O(n^4)$ time.*

A proof of this result can be found in D. Morozov (2008, Section 4.4.2). We reproduce it below for completeness, with a slight adaptation that allows us to avoid assuming that $\mathcal{F}$ is one-critical. Before giving the proof, let us mention that this theorem, combined with the inclusion-exclusion formula (4.5), gives a $O(n^4)$-time algorithm to compute the rectangle-decomposition of $H_p(\mathcal{F})$ assuming that this module is rectangle-decomposable: once the rank invariant of $H_p(\mathcal{F})$ has been computed, iterate over all pairs (s,t) with $s \leq t \in G$ and, for each one of them, apply the formula in constant time to get the multiplicity of the rectangle module $\mathbf{k}_{[\![s_x,t_x]\!]\times[\![s_y,t_y]\!]}$ in the decomposition of $H_p(\mathcal{F})$. Thus,

Corollary 4.6. *Computing the decomposition of a rectangle-decomposable module induced in homology by a two-parameter filtration with n simplices in total can be done in $O(n^4)$ time.*

This complexity compares favorably to that of the currently best known algorithm for computing direct-sum decompositions of general two-parameter persistence modules[1], which is $O(n^{2\omega+1})$ where $2 \leq \omega < 2.373$ is the exponent for matrix multiplication (Dey and Xin, 2022).

Proof of Theorem 4.5. First, we provide a simplified algorithm that runs in $O(n^{2+\omega})$ time. Consider all the paths of the form $(1,1) \to \cdots \to (i,1) \to \cdots \to (i,j) \to \cdots \to (n,j) \to \cdots \to (n,n)$ in the $n \times n$ grid, where $(i,j) \in [\![1,n]\!]^2$ is arbitrary, as illustrated in Figure 4.1 (left). We call such a path a *stair*, denoted by $S_{(i,j)}$, and we call the corresponding index (i,j) its *nosing*. Note that all the stairs whose nosing is of the form $(i,1)$ or (n,j)

[1]Let us also point out that our approach does not suffer from the limitation of the algorithm of Dey and Xin (2022), which is that no two generators or relations in a minimal presentation of M can have the same grade.

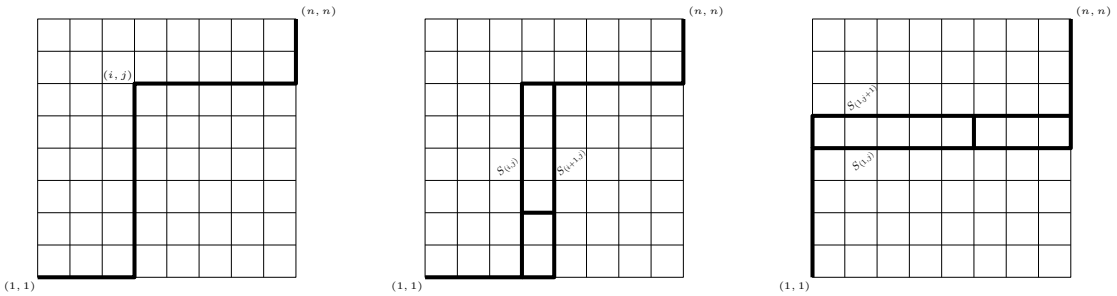


Figure 4.1: Left: the stair $S_{(i,j)}$. Center: transitioning from $S_{(i,j)}$ to $S_{(i+1,j)}$ via a sequence of $O(n)$ intermediate paths with 2 steps each. Right: transitioning from $S_{(1,j)}$ to $S_{(1,j+1)}$.

are in fact identical to the path $(1,1) \to \cdots \to (n,1) \to \cdots \to (n,n)$, while all the other stairs with different nosings are different. The key observation is that, for any pair of comparable indices $(i,j) \leq (i',j')$, the stair with nosing (i,j') passes through both indices. Computing the rank invariant can then be done by iterating over all the stairs, for instance in lexicographical order of the coordinates of their respective nosings, and for each such stair $S_{(i,j)}$, by computing the persistence barcode of the one-parameter restriction $F|_{S_{(i,j)}}$ and then using this barcode to report the ranks between all the grid indices encountered along the path. This takes $O(n^\omega)$ per stair, using the one-parameter persistence algorithm based on fast matrix multiplication (Milosavljević, D. Morozov, and Skraba, 2011), and as there are $O(n^2)$ stairs in total, the overall running time of the algorithm is $O(n^{2+\omega})$.

In order to reduce the overall complexity to $O(n^4)$, we exploit the additional observation that, to transition between two consecutive stairs in lexicographical order of the coordinates of their nosings, say $S_{(i,j)}$ and $S_{(i+1,j)}$, one can go through a sequence of $O(n)$ intermediate paths of the form $(1,1) \to \cdots \to (i,1) \to \cdots \to (i,k) \to (i+1,k) \to \cdots \to (i+1,j) \to \cdots \to (n,j) \to \cdots \to (n,n)$, where k ranges from 2 to $j-1$, as illustrated in Figure 4.1 (center). Any two consecutive paths in this sequence differ only at a single cell of the grid $[\![1,n]\!]^2$, therefore the restrictions of F to these two paths either do not differ, or differ by one simplex being inserted one step earlier or later, or by two consecutive simplex insertions being exchanged. In any situation, the persistence barcode can be updated in $O(n)$ time using the vineyards algorithm (Cohen-Steiner, Edelsbrunner, and D. Morozov, 2006). The update of the barcode from $S_{(i,j)}$ to $S_{(i+1,j)}$ then takes $O(n^2)$ time. Likewise, we can compute the barcode of $F|_{S_{(1,j+1)}}$ by transitioning from $S_{(1,j)}$ via an intermediate sequence of $O(n)$ paths differing at a single cell in the grid each time, as illustrated in Figure 4.1 (right). Thus, the barcode of $F|_{S_{(1,j+1)}}$ is also obtained in $O(n^2)$ time, and so the overall running time of the algorithm is reduced to $O(n^4)$. This concludes the proof of Theorem 4.5. □

4.1.4 Algorithm for checking rectangle-decomposability

As in Section 4.1.3, consider a two-parameter filtration $\mathcal{F}$ indexed over a finite grid with n simplices in total. Again, let us assume without loss of generality that $\mathcal{F}$ is indexed over the grid $G = [\![1,n]\!] \times [\![1,n]\!]$. We further assume that each arrow $\mathcal{F}_{(i,j)} \subseteq \mathcal{F}_{(i+1,j)}$ or

$\mathcal{F}_{(i,j)} \subseteq \mathcal{F}_{(i,j+1)}$ is either an identity map or a single simplex insertion, and we fix a homology degree p. In this section, we address the following question: given this input, how fast can we check whether the persistence bimodule M induced in p-th homology decomposes into rectangle summands?

An obvious solution is to first decompose M from the data of $\mathcal{F}$, then to check the summands one by one. As explained in Section 4.1.3, the currently best known algorithm for decomposition runs in time $O(n^{2\omega+1})$, where $2 \leq \omega < 2.373$ is the exponent for matrix multiplication (Dey and Xin, 2022). The advantage of the weak exactness condition (Definition 2.1) is that it can be checked locally, which reduces the total running time to $O(n^{2+\omega})$:

Theorem 4.7. *Checking the rectangle-decomposability of the bimodule induced in homology by a two-parameter filtration of a simplicial complex with n simplices can be done in $O(n^{2+\omega})$ time.*

Proof. First, we sketch the algorithm:

(i) Compute the rank invariant $r : [\![1,n]\!]^2 \times [\![1,n]\!]^2 \to \mathbb{N}_0$ of M.

(ii) Compute invariants for kernels and images, denoted by $\kappa : [\![1,n]\!]^2 \times [\![1,n]\!]^2 \to \mathbb{N}_0$ and $\iota : [\![1,n]\!]^2 \times [\![1,n]\!]^2 \to \mathbb{N}_0$ respectively, which return the dimensions of $\operatorname{Ker} \rho_s^{(s_x,t_y)} + \operatorname{Ker} \rho_s^{(t_x,s_y)}$ and of $\operatorname{Im} \rho^t_{(s_x,t_y)} \cap \operatorname{Im} \rho^t_{(t_x,s_y)}$ respectively at indices $s \leq t$, and zero elsewhere.

(iii) For each pair of indices $s \leq t$, check whether $r(s,t) = \iota(s,t)$ and $r(s,s) - r(s,t) = \kappa(s,t)$. If any such equality fails, then answer that M is not rectangle-decomposable. Otherwise, answer that M is rectangle-decomposable.

We now provide further implementation details and analyze the algorithm on the fly:

Step (i) has already been detailed in Section 4.1.3 and runs in $O(n^4)$ time.

Step (iii) obviously runs in $O(n^4)$ time, and its correctness comes from the commutativity of the square in Definition 2.1. Indeed, commutativity implies that $\operatorname{Im} \rho^t_s \subseteq \operatorname{Im} \rho^t_{(s_x,t_y)} \cap \operatorname{Im} \rho^t_{(t_x,s_y)}$ and $\operatorname{Ker} \rho_s^{(s_x,t_y)} + \operatorname{Ker} \rho_s^{(t_x,s_y)} \subseteq \operatorname{Ker} \rho^t_s$, so checking weak exactness for this square amounts to checking equality between the dimensions of the various spaces involved, hence the equations.

Step (ii) proceeds as follows. We first compute, for each $t = (j,l) \in G$, the barcode of the zigzag module[2] induced in homology by the following zigzag of simplicial complexes where arrows are inclusions:

$$\mathcal{F}_{(1,l)} \longrightarrow \cdots \longrightarrow \mathcal{F}_{(j-1,l)} \longrightarrow \mathcal{F}_t \longleftarrow \mathcal{F}_{(j,l-1)} \longleftarrow \cdots \longleftarrow \mathcal{F}_{(j,1)} \,. \tag{4.6}$$

We then do the same with the following zigzag, for each $s = (i,k) \in G$:

$$\mathcal{F}_{(i,n)} \longleftarrow \cdots \longleftarrow \mathcal{F}_{(i,k+1)} \longleftarrow \mathcal{F}_s \longrightarrow \mathcal{F}_{(i+1,k)} \longrightarrow \cdots \longrightarrow \mathcal{F}_{(n,k)} \,. \tag{4.7}$$

Then, we use the following simple claim:

[2] A *zigzag module* is a persistence module indexed over a poset of the form $\bullet \longleftrightarrow \bullet \longleftrightarrow \cdots \longleftrightarrow \bullet$, where double-headed arrows mean that the actual arrows can be oriented either forward or backward. Such modules always decompose into direct sums of interval modules (Botnan, 2017; Carlsson and De Silva, 2010).

Claim 4.8. Consider the following commutative square (left) and pfd persistence bimodule indexed over it (right):

Then:

$$\dim(\operatorname{Im}\gamma \cap \operatorname{Im}\delta) = \#\Big\{\text{intervals with support } \{b,c,d\} \text{ in the barcode of the zigzag } C \xrightarrow{\delta} D \xleftarrow{\gamma} B\Big\},$$

$$\dim(\operatorname{Ker}\alpha + \operatorname{Ker}\beta) = \dim(A) - \#\Big\{\text{intervals with support } \{a,b,c\} \text{ in the barcode of the zigzag } C \xleftarrow{\beta} A \xrightarrow{\alpha} B\Big\}.$$

First, let us show that Claim 4.8 allows us to conclude. This claim ensures that in our situation, for each indices $(i,k) = s \leq t = (j,l)$, the dimension of $\operatorname{Im}\rho^t_{(i,l)} \cap \operatorname{Im}\rho^t_{(j,k)}$ is given by the number of intervals in the barcode of (4.6) that span the subzigzag $\mathcal{F}_{(i,l)} \longrightarrow \mathcal{F}_t \longleftarrow \mathcal{F}_{(j,k)}$. Dually, the dimension of $\operatorname{Ker}\rho^{(i,l)}_s + \operatorname{Ker}\rho^{(j,k)}_s$ is given by $r(s,s)$ minus the number of intervals in the barcode of (4.7) that span the subzigzag $\mathcal{F}_{(i,l)} \longleftarrow \mathcal{F}_s \longrightarrow \mathcal{F}_{(j,k)}$.

Regarding the running time: since the zigzags (4.6)-(4.7) involve $O(n)$ simplex insertions and deletions each, their barcode computation takes $O(n^\omega)$ using the algorithm based on fast matrix multiplication (Milosavljević, D. Morozov, and Skraba, 2011). Then, each barcode having $O(n)$ intervals, the computation of the dimensions and their storage in tables of integers representing the invariants κ and ι takes $O(n)$. This is true for each choice of indices $s \leq t$, hence a total running time in $O(n^{2+\omega} + n^3) = O(n^{2+\omega})$.

All that remains is to prove Claim 4.8. We only prove the first equality, as the second one follows by duality. Take an interval-decomposition of the zigzag $C \xrightarrow{\delta} D \xleftarrow{\gamma} B$, and pick a basis $(\xi_1, \cdots, \xi_l)$ of D that is compatible with this decomposition. This means that each basis element ξ_i lies in the span of a unique interval summand of the zigzag at D. Then, by restriction we have:

$$\begin{aligned}
\xi_i \in \operatorname{Im}\gamma \quad \Longleftrightarrow \quad & \xi_i \in \operatorname{Span}(\text{summands with support } \{b,d\}) \\
& + \operatorname{Span}(\text{summands with support } \{b,c,d\}); \\
\xi_i \in \operatorname{Im}\delta \quad \Longleftrightarrow \quad & \xi_i \in \operatorname{Span}(\text{summands with support } \{c,d\}) \\
& + \operatorname{Span}(\text{summands with support } \{b,c,d\}).
\end{aligned}$$

The spans of distinct summands being in direct sum in D, we deduce that

$$\xi_i \in \operatorname{Im}\gamma \cap \operatorname{Im}\delta \quad \Longleftrightarrow \quad \xi_i \in \operatorname{Span}(\text{summands with support } \{b,c,d\}).$$

Hence the result. □

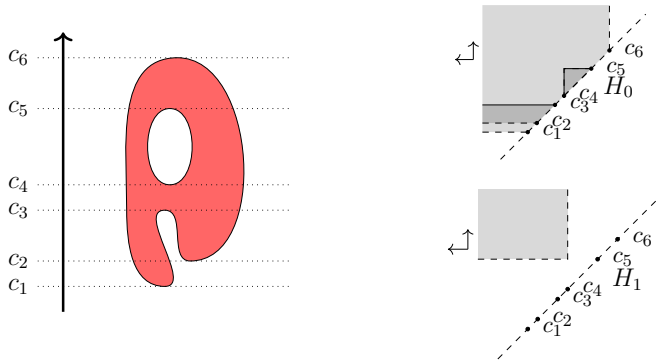


Figure 4.2: A function f on a topological space (left) and the associated interval-decompositions in level-sets persistent homology in dimensions 0 and 1 (right).

4.2 Homology pyramids and the strip

In this section we shall extend the block-decomposition theorem (Theorem 1.6) to persistence modules that are strongly exact on certain "strip" subsets of the plane (Theorem 4.11). As we shall see, such persistence modules arise naturally in the context of topological data analysis.

4.2.1 Introduction

The interest in block-decomposable modules originated in the study of level-sets persistent homology. Specifically, for a topological space A and a continuous function $f\colon A \to \mathbb{R}$ one defines the persistence module $M(f)$ indexed over $\Delta^+ := \{(x,y) \in \mathbb{R}^{\mathrm{op}} \times \mathbb{R} : x < y\}$ as:

$$M(f)_{(x,y)} = H_p\left(\{a \in A\,;\, x < f(a) < y\}\right).$$

It follows immediately from a simple Mayer-Vietoris calculation that $M(f)$ is strongly exact (Cochoy and Oudot, 2020) in the following sense:

Definition 4.9. A persistence module M over a subset $Q \subseteq \mathbb{R}^2$ is *strongly exact* if for any $s \leq t$ in Q, such that (s_x, t_y) and (t_x, s_y) are also in Q, the following sequence is exact:

$$M_s \xrightarrow{\rho_s^{(t_x,s_y)} \oplus \rho_s^{(s_x,t_y)}} M_{(s_x,t_y)} \oplus M_{(t_x,s_y)} \xrightarrow{\rho^t_{(s_x,t_y)} - \rho^t_{(t_x,s_y)}} M_t.$$

Furthermore, assuming that $M(f)$ is pfd, it is interval-decomposable where each interval I is of the form $I = B \cap \Delta^+$, and where B is a block in $\mathbb{R}^{\mathrm{op}} \times \mathbb{R}$; see Botnan and Crawley-Boevey (2020) and Cochoy and Oudot (2020). See Figure 4.2 for an example of a real-valued function and the associated intervals.

The domain of $M(f)$ can be extended by considering relative homology. To see this, consider the poset consisting of all pairs of open sets (u, v) of the form,

$$((x,y),\varnothing) \qquad ((x,\infty),(y,\infty)) \qquad ((-\infty,y),(-\infty,x)) \qquad (\mathbb{R},(-\infty,x)\cup(y,\infty)),$$

where $-\infty \leq x \leq y \leq \infty$ are chosen such that $u \neq \varnothing$. The persistence module $M(f)$ is extended by defining $M(f)_{(u,v)} = H_p(f^{-1}(u), f^{-1}(v))$. Moreover, after re-scaling and

reversing the horizontal arrow (see also Remark 4.16), the module $M(f)$ can be seen as a persistence module over the *pyramid* poset $\mathcal{P}$ defined as:

$$\mathcal{P} = \left\{(x,y) \in (-1,1)^2,\ |x| + |y| \leq 1\right\} \setminus \left\{(x,y) \in (-1,1)^2,\ x + y = -1\right\}.$$

See Figure 4.3 for an illustration. We shall refer to the resulting persistence module $M(f) \in \mathsf{Per}(\mathcal{P})$ as the *(continuous) homology pyramid (in dimension p).*

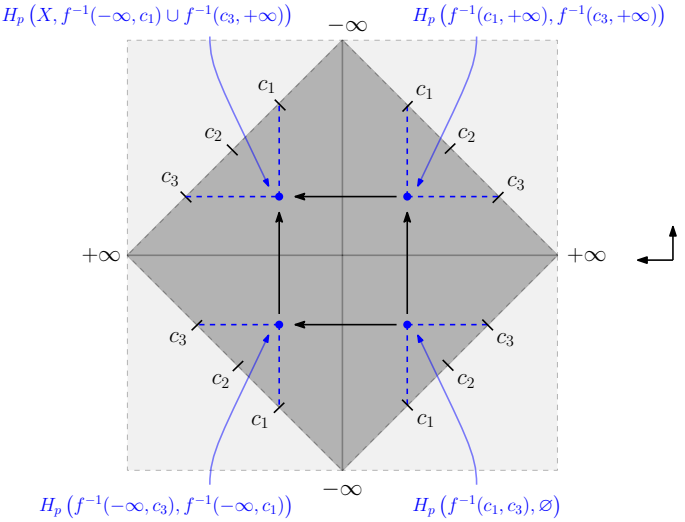


Figure 4.3: In dark grey, the homology pyramid of Bendich et al. (2013, pp. 6, 7) (up to rotation). Up to rescaling and reversing the horizontal axis, the underlying poset corresponds to the pyramid poset $\mathcal{P}$.

Define the *strip* to be the subposet $\mathsf{St} \subset \mathbb{R}^2$ given by

$$\mathsf{St} = \{(x,y) \mid y \leq x + 1 \text{ and } y \geq x - 1\},$$

or equivalently, all points situated on and between the two lines $y = x+1$ and $y = x-1$. Concatenating translations of pyramids we obtain our main object of study. Specifically, for a non-negative integer k, let $\mathcal{P}(k) = \{p \in \mathbb{R}^2 \mid p + (k,k) \in \mathcal{P}\}$, and for $0 \leq m \leq \infty$,

$$\mathsf{St}_m = \begin{cases} \bigcup_{0 \leq k \leq m} \mathcal{P}(k) & \text{if } m < \infty, \\ \bigcup_{0 \leq k < \infty} \mathcal{P}(k) & \text{if } m = \infty. \end{cases}$$

Figure 4.4 illustrates the process of constructing St_∞.

As was first noted by Carlsson, De Silva, and D. Morozov (2009), the boundary maps in the relative Mayer-Vietoris sequence can be used to connect homology pyramids of consecutive dimensions if one flips every other pyramid. In particular, $m+1$ consecutive homology pyramids assemble into a persistence module over St_m. Figure 4.4 illustrates this procedure and shows the interval-decomposition of $M(f) \in \mathsf{Per}(\mathsf{St}_\infty)$ associated to the function f from Figure 4.2.

Remark 4.10. We remark that care must be taken in verifying that the resulting persistence module $M(f)$ over St_∞ is well-defined. The reader should consult Bauer, Botnan, and Fluhr (2021) for a careful construction of $M(f)$, and a direct proof of an associated structure theorem.

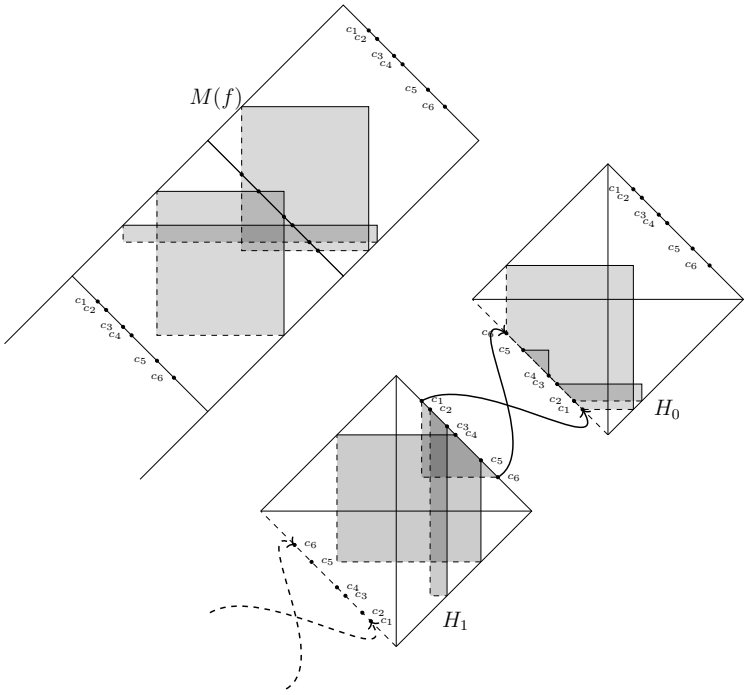


Figure 4.4: The homology pyramids and the corresponding $M(f) \in \mathsf{Per}\,(\mathsf{St}_\infty)$ of the function in Figure 4.2.

In this section, we prove the following:

Theorem 4.11. *Let $0 \leq m \leq \infty$ and $M \in \mathsf{Per}(\mathsf{St}_m)$ be pfd, strongly exact and trivial when restricted to indices on the boundary components $y = x+1$ and $y = x-1$. Then M is interval-decomposable where each interval I is of the form $I = R \cap \mathsf{St}_m$, where R is a maximal rectangle supported on the interior of St.*

Here *maximality* is meant in the following sense: if $R \subseteq S \subset \mathbb{R}^2$ where S is a rectangle supported in the interior of St, then $R = S$.

Remark 4.12. Note that the homology pyramid is trivial on the boundaries $y = x + 1$ and $y = x - 1$, as the corresponding vector spaces are given by the relative homology of preimages of pairs of the form $((x, \infty), (x, \infty))$ and $((-\infty, x), (-\infty, x))$, respectively. In particular, Theorem 4.11 for $m = 0$ gives a structure theorem for pfd homology pyramids.

Remark 4.13. A proof of Theorem 4.11 for $m = 0$, under the assumption that the persistence modules are determined by a finite subset of $\mathcal{P}$, first appeared in work by Bendich et al. (2013), following ideas from Carlsson, De Silva, and D. Morozov (2009). An alternative proof in that setting was given in Botnan, Lebovici, and Oudot (2022) using Theorem 2.3. A proof of Theorem 4.11 for $m = \infty$, under the assumption that M is *sequentially continuous*, can be found in recent work by Bauer, Botnan, and Fluhr (2021).

Outline. The proof of Theorem 4.11 goes as follows. First, we prove that if a persistence module indexed over the strip is strongly exact and vanishes on the boundary

of the strip, it is rectangle-decomposable with only maximal rectangles appearing in its rectangle-decomposition (Lemma 4.14). Then, we prove that any persistence module over St_m decomposes as a direct sum of three persistence modules: one that is strongly exact and extends to a module vanishing on the boundary of the strip, and two block-decomposable modules (Lemmas 4.20 and 4.21). The result then follows from restricting to St_m and using the observation that a module is isomorphic to the restriction of its extension.

4.2.2 Proof of Theorem 4.11

We now prove Theorem 4.11. This lemma is fundamental.

Lemma 4.14. *Let M be a strongly exact and pfd persistence module over St such that M is trivial when restricted to the boundary components $y = x + 1$ and $y = x - 1$. Then M is interval-decomposable and each interval is a maximal rectangle on the interior of St.*

Proof. Since M is trivial on the boundary components, we may extend M to a persistence module over $\mathbb{R}^2$ by defining $M_p = 0$ for all p not contained in St. By virtue of Theorem 1.7, it suffices to show that the extension to $\mathbb{R}^2$ is weakly exact. This is not hard to see: consider s and t as in Definition 2.1. Then, by assumption, M can only fail to be weakly exact if (s_x, t_y) or (t_x, s_y) is not contained in St. However, in that case,

$$\begin{aligned} 0 = \operatorname{Im} \rho_s^t &= \operatorname{Im} \rho_{(t_x,s_y)}^t \cap \operatorname{Im} \rho_{(s_x,t_y)}^t, \\ M_s = \operatorname{Ker} \rho_s^t &= \operatorname{Ker} \rho_s^{(t_x,s_y)} + \operatorname{Ker} \rho_s^{(s_x,t_y)}. \end{aligned}$$

Hence, M is weakly exact. Furthermore, since the extension of M is trivial outside of St, any rectangle R in the decomposition of M must be necessarily be supported on a subset of St.

We now show that R is maximal. Write $R = \langle a, b\rangle \times \langle c, d\rangle$ where $\langle a, b\rangle$ denotes an interval in $\mathbb{R}$ with left and right endpoints given by a and b, respectively, e.g., $[a, b)$ or (a, b). Assume that the point $(b, c) \in \mathbb{R}^2$ is not on the boundary of St. Then, we can choose s and t such that the square $\mathrm{Q}_s^t = \{s_x, t_x\} \times \{s_y, t_y\}$ is contained in the interior of St and $R \cap \mathrm{Q}_s^t = \{(s_x, t_y)\}$. In particular, the sequence in Definition 4.9 associated to Q_s^t has a summand of the form $0 \to \mathbf{k} \to 0$, contradicting that the sequence is exact. Symmetrically, the point $(a, d) \in \mathbb{R}^2$ must be contained in the other boundary component. □

We now describe a procedure to extend $M \in \mathsf{Per}(\mathsf{St}_m)$ to a persistence module over St by means of left and right Kan extensions. In order to ensure that the extension is trivial on the boundary components we shall consider the poset St_m^b given by the union of St_m with the two boundary components of St. The extension works as follows:

(i) Extend $M \in \mathsf{Per}(\mathsf{St}_m)$ to $M \in \mathsf{Per}\left(\mathsf{St}_m^b\right)$ by defining the module to be trivial on the boundary components.

(ii) Extend to $\mathrm{Ran}_\varphi(M) \in \mathsf{Per}\left(\mathsf{St}_\infty^b\right)$ by a right Kan Extension along the inclusion $\varphi : \mathsf{St}_m^b \hookrightarrow \mathsf{St}_\infty^b$.

(iii) Extend to $\widetilde{M} := \mathrm{Lan}_\psi(\mathrm{Ran}_\varphi(M)) \in \mathsf{Per}\,(\mathsf{St})$ by means of a left Kan extension along $\psi : \mathsf{St}^b_\infty \hookrightarrow \mathsf{St}$.

The fact that $\big(\widetilde{M}\big)|_{\mathsf{St}_m} \cong M$ follows from Riehl (2017, Corollary 6.3.9) as both of the Kan extensions are computed pointwise along a full functor.

Remark 4.15. Recall from Section 1.1 that vector-space duality defines a contravariant functor $D : \mathsf{Per}(\mathcal{P}) \to \mathsf{Per}(\mathcal{P}^{\mathrm{op}})$ that satisfies $DDM \simeq M$ for any pfd persistence module M. This functor sends strongly exact modules to strongly exact modules.

In the proofs, we shall make use of the following notation,

$$\begin{aligned} \blacktriangle &:= \big\{(x,y) \in (0,1)^2,\ x+y \leq 1\big\}, \\ \blacktriangledown &:= \big\{(x,y) \in (-m-1,-m)^2, x+y > -1-2m\big\}, \end{aligned}$$

and we shall denote the union of each of these subsets with the boundary components of St by $\blacktriangle^b$ and $\blacktriangledown^b$, respectively.

Remark 4.16. In the following we shall make use of results from Botnan and Crawley-Boevey (2020) and Cochoy and Oudot (2020). The results were originally formulated for persistence modules over $\mathbb{R}^2$, over $T := \{(x,y) \in \mathbb{R}^2,\ x+y > 0\}$ and over $\overline{T} := \{(x,y) \in \mathbb{R}^2,\ x+y \geq 0\}$. These results do however apply verbatim in the settings of $\blacktriangle$ and $\blacktriangledown$, as can be seen from the following two poset isomorphisms and Remark 4.15,

$$\begin{aligned} (x,y) \in \blacktriangledown &\mapsto \left(\tan\left(\frac{\pi}{2}(x+1+m)\right), \tan\left(\frac{\pi}{2}(y+1+m)\right)\right) \in T \\ (x,y) \in \blacktriangle^{\mathrm{op}} &\mapsto \left(\tan\left(\frac{\pi}{2}(-x+1)\right), \tan\left(\frac{\pi}{2}(-y+1)\right)\right) \in \overline{T} \end{aligned}$$

The following is a slight reformulation of two results from Botnan and Crawley-Boevey (2020, Section 5.2).

Theorem 4.17. *Let M be a strongly exact and pfd persistence module over $\blacktriangle$ or $\blacktriangledown$. Then M is interval-decomposable, and each interval B is of the form $B = R \cap \blacktriangle$ or $I = R \cap \blacktriangledown$, respectively, where R is a maximal rectangle in the interior of St.*

We shall refer to such intervals as *blocks* in analogy with the case of product posets (Section 1.1). If $B = R \cap \blacktriangle$, and there exists a $p \in \blacktriangle$ such that $p < q$ for all $q \in B$, then B is a *birth quadrant*. Dually, if $B = R \cap \blacktriangledown$ and there exists $q \in B$ such that $p < q$ for all $p \in B$, then B is a *death quadrant*. For instance, the triangular shaped blocks in the homology pyramids in dimension 0 and 1 shown in Figure 4.4 are death and birth quadrants, respectively. The following is an immediate corollary of Theorem 4.17.

Corollary 4.18. *Let $M \in \mathsf{Per}\,(\mathsf{St}_m)$ be as in the statement of Theorem 4.11. Then the restriction of M to $\blacktriangledown$, denoted by $M|_{\blacktriangledown}$, decomposes as*

$$M_{|\blacktriangledown} \simeq \bigoplus_{B \in \mathbf{B}_1} \mathbf{k}_B, \tag{4.8}$$

where $\mathbf{B}_1$ is a multiset of blocks in $\blacktriangledown$. Dually, the restriction of M to $\blacktriangle$ decomposes as

$$M_{|\blacktriangle} \simeq \bigoplus_{B \in \mathbf{B}_2} \mathbf{k}_B, \tag{4.9}$$

where $\mathbf{B}_2$ is a multiset of blocks in $\blacktriangle$.

Remark 4.19. Let $\varphi : \mathsf{St}_m^b \hookrightarrow \mathsf{St}_\infty^b$, and $\psi : \mathsf{St}_\infty^b \hookrightarrow \mathsf{St}$ denote the inclusions introduced above. The inclusion of posets $\blacktriangleleft^b \hookrightarrow \mathsf{St}_m^b$ is initial in the sense of Mac Lane (1972, Section IX.3), and Theorem 1 of the same section of loc. cit. ensures that we have the following natural isomorphism for all $p \in \mathsf{St}_\infty^b - (\mathsf{St}_m^b - \blacktriangleleft^b)$,

$$(\mathrm{Ran}_\varphi M)_p \simeq \left(\mathrm{Ran}_{\varphi|_{\blacktriangleleft^b}} M|_{\blacktriangleleft^b}\right)_p.$$

Similarly, for all $p \in \mathsf{St} - (\mathsf{St}_\infty^b - \blacktriangleright^b)$, one has:

$$\widetilde{M}_p \simeq (\mathrm{Lan}_\psi(\mathrm{Ran}_\varphi M))_p \simeq \left(\mathrm{Lan}_{\psi|_{\blacktriangleright^b}}(\mathrm{Ran}_\varphi M)|_{\blacktriangleright^b}\right)_p.$$

Lemma 4.20. *Let $M \in \mathsf{Per}(\mathsf{St}_m)$ be as in the statement of Theorem 4.11, and let $\mathbf{B}_1$ and $\mathbf{B}_2$ be as in Corollary 4.18.*

(i) *$\widetilde{M}$ is strongly exact.*

(ii) *If $\mathbf{B}_1$ contains no death quadrant, and $\mathbf{B}_2$ contains no birth quadrant, then $\widetilde{M}$ is pfd.*

Proof. First we prove (i). Following the first step in the extension procedure outlined above, we obtain $M \in \mathsf{Per}\left(\mathsf{St}_m^b\right)$ by adding 0 vector spaces. To simplify notation, we shall let D and U denote the sets $D = \mathsf{St}_\infty - (\mathsf{St}_m - \blacktriangleleft)$, $U = (\mathsf{St} - (\mathsf{St}_\infty - \blacktriangleright))$ and $N = M|_{\blacktriangleleft^b}$. Then

$$N \simeq \bigoplus_{B \in \mathbf{B}_1} \mathbf{k}_B.$$

where each block $B \in \mathbf{B}_1$ is of the form $B = R \cap \blacktriangleleft$ where R is a maximal rectangle in the interior of St. It is not hard to see that extending $\mathbf{k}_B$ "to the left" using a right Kan extension will recover $\mathbf{k}_R$ on D. That is, we have the following natural isomorphism

$$\mathrm{Ran}_{\varphi|_{\blacktriangleleft^b}}(\mathbf{k}_B)_p \cong (\mathbf{k}_R)_p,$$

for all $p \in D$. In particular, since $\mathbf{k}_R$ is strongly exact on St (and therefore on D), it follows that $\mathrm{Ran}_{\varphi|_{\blacktriangleleft^b}}(\mathbf{k}_B)$ is strongly exact on D. Hence,

$$(\mathrm{Ran}_\varphi M)|_D \simeq \left(\mathrm{Ran}_{\varphi|_{\blacktriangleleft^b}} N\right)|_D \simeq \left(\bigoplus_{B \in \mathbf{B}_1} \mathrm{Ran}_{\varphi|_{\blacktriangleleft^b}} \mathbf{k}_B\right)|_D$$

is strongly exact on D. Here the first isomorphism follows from Remark 4.19, and the second isomorphism follows from the fact that right Kan extensions commute with direct products, and that direct products and direct sums coincide when working in the pfd setting; see, e.g., Botnan and Lesnick (2018, Remark 2.16) for more details.

Let N' denote the restriction of $\mathrm{Ran}_\varphi M$ to $\blacktriangleright^b$. From Corollary 4.18 we now have that

$$N' \simeq \bigoplus_{B' \in \mathbf{B}_2} \mathbf{k}_{B'},$$

where $\mathbf{B}_2$ is a multiset of blocks of $\blacktriangleright$. It follows from the second part of Remark 4.19, that $\widetilde{M}|_U \cong (\mathrm{Lan}_{\psi|_{\blacktriangleright^b}} N')|_U$. Showing that the latter is strongly exact is dual to the first

part of this proof. In conclusion, we have that $\widetilde{M}$ is strongly exact when restricted to D, St_m and U. To see that this implies strong exactness on the whole of St, let s and t be as in Definition 4.9, and assume that $s \in D - D \cap \mathsf{St}_m$ and $t \in \mathsf{St}_m - D \cap \mathsf{St}_m$. Then we can find $p \in D \cap \mathsf{St}_m = \blacktriangleleft$, such that $s \leq p \leq t$, and such that each of the smaller squares in the following diagram is strongly exact,

$$\begin{array}{ccccc} \widetilde{M}_{(s_x,t_y)} & \longrightarrow & \widetilde{M}_{(p_x,t_y)} & \longrightarrow & \widetilde{M}_t \\ \uparrow & & \uparrow & & \uparrow \\ \widetilde{M}_{(s_x,p_y)} & \longrightarrow & \widetilde{M}_p & \longrightarrow & \widetilde{M}_{(t_x,p_y)} \\ \uparrow & & \uparrow & & \uparrow \\ \widetilde{M}_s & \longrightarrow & \widetilde{M}_{(p_x,s_y)} & \longrightarrow & \widetilde{M}_{(t_x,s_y)} \end{array} . \tag{4.10}$$

A simple diagram chase shows that the outer square is strongly exact. Hence, $\widetilde{M}$ is strongly exact on $\mathsf{St}_\infty = D \cup \mathsf{St}_m$. A similar argument shows that $\widetilde{M}$ is strongly exact on $\mathsf{St}_\infty \cup U = \mathsf{St}$.

Now we prove (ii). Since $\widetilde{M}|_{\mathsf{St}_m} \cong M|_{\mathsf{St}_m}$, it suffices to prove that $\widetilde{M}$ is pfd over $D - D \cap \mathsf{St}_m$ and $U - U \cap \mathsf{St}_m$. We prove the former, the second is dual. Assume that $\mathrm{Ran}_{\varphi_{\blacktriangleleft^b}} N$ is not pfd at $p \in D - D \cap \mathsf{St}_m$. Then, there must exist an infinite family of blocks $\{B_\lambda = R_\lambda \cap \blacktriangleleft\} \subset \mathbf{B}_1$ such that $p \in R_\lambda$ for all λ. However, by assumption, R_λ is not bounded from above in $\blacktriangleleft$. In other words, R_λ and thus B_λ must contain at least one of the two line segments

$$([p_x, -m) \times \{p_y\}) \cap \blacktriangleleft \qquad \text{and} \qquad (\{p_x\} \times [p_y, -m)) \cap \blacktriangleleft.$$

Hence, either there exists an infinite number of blocks $B_\lambda \in \mathbf{B}_1$ such that

$$([p_x, -m) \times \{p_y\}) \cap \blacktriangleleft \subseteq B_\lambda,$$

or there exists an infinite number of blocks $B_\lambda \in \mathbf{B}_1$ such that

$$(\{p_x\} \times [p_y, 0)) \cap \blacktriangleleft \subseteq B_\lambda.$$

We conclude that $\dim N_q = \infty$ for all $q \in ([p_x, -m) \times \{p_y\}) \cap \blacktriangleleft$ or $\dim N_q = \infty$ for all $q \in (\{p_x\} \times [p_y, -m)) \cap \blacktriangleleft$. This contradicts the assumption that N is pfd. □

Lemma 4.21. *Let $M \in \mathsf{Per}(\mathsf{St}_m)$ be as in the statement of Theorem 4.11, and let $\mathbf{B}_1$ and $\mathbf{B}_2$ be as in Corollary 4.18. Then,*

$$M \simeq N \oplus \bigoplus_{B \in \mathbf{bb}} \mathbf{k}_B \oplus \bigoplus_{B \in \mathbf{db}} \mathbf{k}_B,$$

where $\mathbf{bb} \subseteq \mathbf{B}_1$ is the collection of all birth quadrants, and $\mathbf{db} \subseteq \mathbf{B}_2$ is the collection of all death quadrants, and N contains no summand of the form $\mathbf{k}_B$ where B is a birth quadrant in $\blacktriangleleft$ or a death quadrant in $\blacktriangleleft$.

Proof. Recall that B is a death quadrant in $\blacktriangleleft$ if and only if its support is bounded from above by some $x \in \blacktriangleleft$. Dually, a birth quadrant B in $\blacktriangleleft$ is a birth quadrant if and only

if it is bounded from below by some $x \in \blacktriangleleft$. In either case, there is enough space around their supports to extend them to summands of M. More precisely, if $B \in \mathbf{db}$, then the inclusions and projection maps

$$\mathbf{k}_B \xrightarrow{i_1} M_{|\blacktriangleleft} \xrightarrow{j_1} \mathbf{k}_B$$

such that $j_1 \circ i_1 = \mathrm{id}_{\mathbf{k}_B}$, extend to

$$\mathbf{k}_B \xrightarrow{i} M \xrightarrow{j} \mathbf{k}_B$$

such that $j \circ i = \mathrm{id}_{\mathbf{k}_B}$ (here the interval B is considered as a subset of St_m). Hence, the module $\mathbf{k}_B$ is a summand of M. Doing this for every death quadrant yields,

$$M \simeq N' \oplus \bigoplus_{B \in \mathbf{db}} \mathbf{k}_B.$$

Iterating the above argument in the dual setting of birth quadrants in $\blacktriangleleft$ gives,

$$M \simeq N \oplus \bigoplus_{B \in \mathbf{bb}} \mathbf{k}_B \oplus \bigoplus_{B \in \mathbf{db}} \mathbf{k}_B.$$

□

Proof of Theorem 4.11. By Lemma 4.21 we may assume that $\mathbf{B}_1$ and $\mathbf{B}_2$ contain no death and birth quadrants, respectively. Hence $\widetilde{M}$ is strongly exact and pfd by Lemma 4.20. The result now follows from Lemma 4.14. □

Part II

Euler characteristic based invariants

Chapter 5

Euler calculus

A *constructible function* on a real analytic manifold X is a function $\varphi : X \to \mathbb{Z}$ such that the sets $\varphi^{-1}(m)$ are subanalytic and the family $\{\varphi^{-1}(m)\}_{m\in\mathbb{Z}}$ is locally finite. Using classical results from subanalytic geometry, one can show that the set $\mathrm{CF}(X)$ of constructible functions on X is a commutative unital algebra for the pointwise operations of addition and multiplication. An integral calculus of these functions with respect to the Euler characteristic and several operations (e.g. pushforward, pullback, tensor product) coming from sheaf theory have been defined by Viro (1988) and Schapira (1989, 1991). The link between constructible functions and constructible sheaves is made precise by the function-sheaf correspondence (Kashiwara and Schapira, 1990, Theorem 9.7.1). For a study of these operations without the use of sheaf theory, we refer to McCrory and Parusinski (1997).

Euler calculus has found many applications on the theoretical side. We mention a few examples below and refer to the survey of Gusein-Zade (2010) for more. McCrory and Parusinski (1997) have introduced the *half-link* operator using the formalism of Euler calculus to study the local topological properties of algebraic sets. This operator is key to generalize Akbulut-King's numbers (2012) as well as invariants defined by Coste and Kurdyka (1992). These invariants provide sufficient conditions for a stratified space to be homeomorphic to a real algebraic set. In integral geometry, the convolution defined in Euler calculus appears as a satisfactory generalization of Minkowski addition, which is key to the definition of mixed volumes of convex bodies (Groemer, 1977; Martinez-Maure, 2015).

Although having theoretical roots, Euler calculus has found many applications in topological data analysis. Already in Schapira (1989), Euler calculus was developed as an alternative definition of convolution for polygonal tracings with multiplicities, a useful notion in robotics (Guibas, Ramshaw, and Stolfi, 1983). In applied topology, one can formulate problems of target detection by sensor networks using the Euler calculus formalism. This paradigm allows one to express the number of targets detected by the network as the integral of a constructible function (Baryshnikov and Ghrist, 2009), and even suggests the possibility of target reconstruction thanks to Schapira's inversion result (1995) for specific networks, such as beam sensor networks (Curry, Ghrist, and Robinson, 2012, Section 20.2). Moreover, as already mentioned in the introduction, Schapira's inversion implies that two constructible subsets of $\mathbb{R}^n$ with same persistent homology

in all degrees and for all height filtrations are equal (Curry, Mukherjee, and Turner, 2022; Ghrist, Levanger, and Mai, 2018). Finally, and this is the way we introduced these concepts in the introduction, constructible functions associated to multi-parameter persistence modules stand as simple, informative and well-behaved, albeit incomplete, invariants of these objects.

Outline. In this chapter, we begin by introducing Euler calculus and operations on constructible functions following the approach of Schapira (1991) (Section 5.1). We also introduce some specific subgroups of the group of constructible functions that are of independent interest and used all along Part II (Section 5.2). Then, we present a topological integral transform introduced by Schapira (1995) called the *Radon transform* (Section 5.3). We present Schapira's inversion result and prove that the Radon transform of a compactly supported γ-constructible functions vanishes on affine hyperplanes whose defining conormal is in the polar of the cone γ. This characterization of the support will be key for the proof of the reconstruction formula for the Euler-Fourier transform in Chapter 6. Finally, we explain how to associate a constructible function to any multi-parameter persistence module (Section 5.4).

Conventions. Throughout Part II, we set the following notations and terminology:

(i) We consider two $\mathbb{R}$-vector spaces $\mathbb{V}$ and $\mathbb{V}'$ of finite positive dimensions, a real analytic manifold X and a compact real analytic manifold M.

(ii) We denote by $\mathbb{R}_{>0}$ the set of positive real numbers and $\mathbb{R}_{\geq 0} = \mathbb{R}_{>0} \cup \{0\}$. Similarly, we use the notations $\mathbb{R}_{<0}$ and $\mathbb{R}_{\leq 0}$. We also denote by $\mathbb{R}^\times$ the multiplicative group $\mathbb{R} \setminus \{0\}$.

(iii) The dual of a vector space $\mathbb{V}$ is denoted by $\mathbb{V}^*$ and $\mathbb{R}^n$ will always be identified with its dual under the canonical isomorphism. For $\xi \in \mathbb{V}^*$ and $x \in \mathbb{V}$, we often denote $\xi(x)$ by $\langle \xi, x \rangle$.

(iv) We say that a subset γ of $\mathbb{V}$ is a *cone* if $\mathbb{R}_{>0} \cdot \gamma \subseteq \gamma$. We call *antipodal* of a cone γ, denoted by γ^a, the cone $-\gamma$. A closed cone γ is *proper* if $\gamma \cap \gamma^a = \{0\}$. We call *polar* of a cone γ, denoted by γ°, the cone of $\mathbb{V}^*$ defined by:

$$\gamma^\circ = \{\xi \in \mathbb{V}^* \,;\, \xi(\gamma) \subseteq \mathbb{R}_{\geq 0}\}\,.$$

(v) Let I be an interval of $\mathbb{R}$ and denote by $L^1(I)$ (resp . $L^1_{\mathrm{loc}}(I)$) the space of integrable (resp. locally integrable) complex-valued functions on I.

(vi) Let $\mathbf{k}$ be a field. Throughout this chapter and the next one, we assume that $\mathbf{k}$ is of characteristic zero. This assumption is necessary for the sheaf-function correspondence to hold, see Kashiwara and Schapira (1990, Theorem 9.7.1).

5.1 Operations on constructible functions

In this section, we define the integration of constructible functions with respect to the Euler characteristic and several related operations coming from sheaf theory. We refer

to Kashiwara and Schapira (ibid., Section 8.2) for a concise exposition of the useful definitions and results on subanalytic sets.

Denote by $\mathrm{CF_c}(X)$ the set of constructible functions with compact support on X. As mentioned in Chap. IX of loc. cit., Hardt's triangulation theorem (1976) allows to write any $\varphi \in \mathrm{CF_c}(X)$ as a finite sum $\varphi = \sum_{i=1}^{n} m_i \mathbf{1}_{K_i}$, where the m_i's are integers and the K_i's are compact contractible subanalytic subsets. We can then define the *integral of φ with respect to the Euler characteristic*, by:

$$\int_X \varphi \, \mathrm{d}\chi = \sum_{i=1}^{n} m_i,$$

which does not depend on the decomposition of φ; see Schapira (1991). In comparison with the notation of loc. cit., we add the symbol $\mathrm{d}\chi$ to make this integration easier to distinguish from Lebesgue integration in formulae combining the two.

Example 5.1. For any two real numbers $a < b$, one has:

(i) $\int_{\mathbb{R}} \mathbf{1}_{[a,b]} \, \mathrm{d}\chi = 1$,

(ii) $\int_{\mathbb{R}} \mathbf{1}_{[a,b)} \, \mathrm{d}\chi = \int_{\mathbb{R}} \mathbf{1}_{(a,b]} \, \mathrm{d}\chi = 0$,

(iii) $\int_{\mathbb{R}} \mathbf{1}_{(a,b)} \, \mathrm{d}\chi = -1$.

Example 5.2. If Z is a locally closed relatively compact subanalytic subset of X, then:

$$\int_X \mathbf{1}_Z \, \mathrm{d}\chi = \chi_c(Z),$$

where $\chi_c(Z)$ is the *Euler-Poincaré index with compact support* of Z, defined by

$$\chi_c(Z) = \sum_{j \in \mathbb{Z}} (-1)^j \dim_{\mathbb{Q}} \left(H_c^j(Z; \mathbb{Q}_Z) \right),$$

where $\mathbb{Q}_Z$ denotes the constant sheaf on Z with coefficients in the field $\mathbb{Q}$ and H_c denotes cohomology with compact supports, i.e. $H_c^j(Z; \mathbb{Q}_Z) = \mathrm{R}^j\Gamma_c(Z; \mathbb{Q}_Z)$ following the notations of Kashiwara and Schapira (1990). Note that if Z is compact, then $\chi_c(Z)$ is the Euler characteristic.

A morphism $f : X \to Y$ of real analytic manifolds induces two operations on constructible functions. For $\psi \in \mathrm{CF}(Y)$, we define the *pullback* $f^*\varphi \in \mathrm{CF}(X)$ as $\psi \circ f$. For $\varphi \in \mathrm{CF}(X)$ such that f is proper on $\mathrm{supp}(\varphi)$, we define the $f_*\varphi \in \mathrm{CF}(Y)$, for any $y \in Y$, by:

$$f_*\varphi(y) = \int_X \mathbf{1}_{f^{-1}(y)} \cdot \varphi \, \mathrm{d}\chi.$$

Remark 5.3. One has that $\int_X \varphi \, \mathrm{d}\chi = a_{X*}\varphi$ where $a_X : X \to \{\mathrm{pt}\}$, and where the function $a_{X*}\varphi \in \mathrm{CF}(\{\mathrm{pt}\})$ is identified with its value at the point. Since the pushforward is functorial (Schapira, 1991, Theorem 2.3), Euler calculus enjoys a *Fubini theorem*, that is, integration is invariant by pushforward: for any morphism of real analytic manifolds $f : X \to Y$ which is proper on $\mathrm{supp}(\varphi)$, one has

$$\int_Y f_*\varphi \, \mathrm{d}\chi = \int_X \varphi \, \mathrm{d}\chi.$$

We finish this section with three well-known results useful all along Part II that are not explicitly written elsewhere in the literature.

Lemma 5.4. *Let $\varphi \in \mathrm{CF}(X)$ and $f : X \to Y$ be a morphism of real analytic manifolds which is proper on* $\mathrm{supp}(\varphi)$*. Then,* $\mathrm{supp}(f_*\varphi) \subseteq f(\mathrm{supp}(\varphi))$ *with equality if f is injective.*

Proof. For $y \in Y$, we have:

$$f_*\varphi(y) = \int_X \varphi \cdot \mathbf{1}_{f^{-1}(y)} \, \mathrm{d}\chi = \int_X \varphi \cdot \mathbf{1}_{f^{-1}(y) \cap \mathrm{supp}(\varphi)} \, \mathrm{d}\chi. \tag{5.1}$$

Therefore, if $y \notin f(\mathrm{supp}(\varphi))$, then $f^{-1}(y) \cap \mathrm{supp}(\varphi) = \varnothing$, hence $f_*\varphi(y) = 0$. Thus, the function $f_*\varphi$ vanishes on the complement of the closed[1] set $f(\mathrm{supp}(\varphi))$, hence the result.

If f is injective then (5.1) becomes:

$$f_*\varphi(y) = \begin{cases} \varphi(f^{-1}(y)) & \text{if } y \in \mathrm{Im}(f), \\ 0 & \text{else.} \end{cases} \tag{5.2}$$

Let $y \notin \mathrm{supp}(f_*\varphi)$. There exists a neighborhood U of y such that for any $z \in U$, we have $f_*\varphi(z) = 0$. If $y \in \mathrm{Im}(f)$, the open neighborhood $f^{-1}(U)$ of $f^{-1}(y)$ in X satisfies that for any $x \in f^{-1}(U)$, $\varphi(x) = \varphi(f^{-1}(z)) = f_*\varphi(z) = 0$, where $z = f(x) \in U \cap \mathrm{Im}(f)$. Therefore, $f^{-1}(y) \notin \mathrm{supp}(\varphi)$, i.e. $y \notin f(\mathrm{supp}(\varphi))$. Moreover, if $y \notin \mathrm{Im}(f)$, then obviously $y \notin f(\mathrm{supp}(\varphi))$. Hence, $f(\mathrm{supp}(\varphi)) \subseteq \mathrm{supp}(f_*\varphi)$. □

Remark 5.5. The proof of the previous lemma—especially (5.2)—implies that if f is injective, then for any subanalytic subset $Z \subseteq X$, we have $f_*\mathbf{1}_Z = \mathbf{1}_{f(Z)}$.

Lemma 5.6. *Let $\varphi \in \mathrm{CF}(X)$ and $\psi \in \mathrm{CF}(Y)$. Let $f : X \to W$ and $g : Y \to Z$ be two morphisms of real analytic manifolds which are respectively proper on* $\mathrm{supp}(\varphi)$ *and* $\mathrm{supp}(\psi)$*. Then, denoting $f \times g : X \times Y \longrightarrow W \times Z$ the natural map, we have:*

$$(f \times g)_*(\varphi \boxtimes \psi) = (f_*\varphi) \boxtimes (g_*\psi),$$

where by definition $(\varphi \boxtimes \psi)(x, y) = \varphi(x) \cdot \psi(y)$ for any $x \in X$ and $y \in Y$. The function $\varphi \boxtimes \psi$ is called the box product *of φ and ψ.*

Proof. For $(w, z) \in W \times Z$, we compute:

$$\begin{aligned} (f \times g)_*(\varphi \boxtimes \psi)(w, z) &= \int_{X \times Y} \left(\mathbf{1}_{f^{-1}(w)} \cdot \varphi\right) \boxtimes \left(\mathbf{1}_{g^{-1}(z)} \cdot \psi\right) \mathrm{d}\chi \\ &= \left(\int_X \mathbf{1}_{f^{-1}(w)} \cdot \varphi \, \mathrm{d}\chi\right) \cdot \left(\int_Y \mathbf{1}_{g^{-1}(z)} \cdot \psi \, \mathrm{d}\chi\right), \end{aligned}$$

where the first equality follows from a direct computation and the second equality follows from the definition of Euler integration. □

[1] Since $f_{|\mathrm{supp}(\varphi)}$ is proper and Y is Hausdorff and locally compact, it is a closed map.

Corollary 5.7. *Let $\varphi, \psi \in \mathrm{CF}(\mathbb{V})$, and $f : \mathbb{V} \to \mathbb{V}'$ be linear and proper on* $\mathrm{supp}(\varphi) + \mathrm{supp}(\psi)$*. Then, f is proper on* $\mathrm{supp}(\varphi)$ *and on* $\mathrm{supp}(\psi)$*, and:*

$$f_* (\varphi \star \psi) = (f_* \varphi) \star (f_* \psi),$$

where by definition $\varphi \star \psi = s_*(\varphi \boxtimes \psi)$ for $s : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$ the addition is called the convolution *of φ and ψ.*

Proof. We have $f \circ s = s \circ (f \times f)$ by linearity of f, so that by Lemma 5.6, $f_* (\varphi \star \psi) = s_* (f_* \varphi \boxtimes f_* \psi) = (f_* \varphi) \star (f_* \psi)$. □

5.2 Notable subgroups of the group of constructible functions

In this section, we introduce several subgroups of the group of constructible functions that are of prime interest in this thesis.

5.2.1 Constructibility up to infinity

To ensure well-definedness of hybrid transforms, we often restrict ourselves to a subclass of constructible functions on $\mathbb{V}$ defined in Schapira (2020), called constructible up to infinity in the projective compactification of $\mathbb{V}$, or simply *constructible up to infinity* in the present thesis. They correspond to constructible functions that are still constructible when extended by 0 to the projective compactification of $\mathbb{V}$.

Setting $\mathbb{W} = \mathbb{V} \oplus \mathbb{R}$, we denote by $\mathbb{P}(\mathbb{V})$ the *projective compactification* of $\mathbb{V}$, i.e., the set of linear subspaces of $\mathbb{W}$ of dimension 1, formally defined as the quotient:

$$\mathbb{P}(\mathbb{V}) \simeq \big(\mathbb{W} \setminus \{0\}\big)/\mathbb{R}^\times.$$

Any point $x \in \mathbb{P}(\mathbb{V})$ can thus be written in homogeneous coordinates as a class $x = [v : \lambda]$ with $(v, \lambda) \in \mathbb{W} \setminus \{0\}$ and there is an open embedding $j : v \in \mathbb{V} \hookrightarrow [v : 1] \in \mathbb{P}(\mathbb{V})$. We denote by $\mathbb{P}^*(\mathbb{V})$ the projective compactification of the dual of $\mathbb{V}$:

$$\mathbb{P}^*(\mathbb{V}) \simeq \big(\mathbb{W}' \setminus \{0\}\big)/\mathbb{R}^\times,$$

where $\mathbb{W}' = \mathbb{V}^* \oplus \mathbb{R}$. Any element $y \in \mathbb{P}^*(\mathbb{V})$ can be written in homogeneous coordinates as a class $y = [\xi : t]$ with $(\xi, t) \in \mathbb{W}' \setminus \{0\}$. We call *hyperplane at infinity*, denoted by h_∞, the element $[0 : 1] \in \mathbb{P}^*(\mathbb{V})$ where 0 is understood here as an element of $\mathbb{V}^*$. There is a bijection between $\mathbb{P}^*(\mathbb{V}) \setminus \{h_\infty\}$ and the set of affine hyperplanes of $\mathbb{V}$ which sends a class $[\xi : t] \neq [0 : 1]$ to the affine hyperplane $\xi^{-1}(t) \subseteq \mathbb{V}$.

Definition 5.8 (ibid., Definition 4.1)**.** Let $\varphi \in \mathrm{CF}(\mathbb{V})$. We say that φ is *constructible up to infinity* if:

(i) for all $m \in \mathbb{Z}$, $j(\varphi^{-1}(m))$ is subanalytic in $\mathbb{P}(\mathbb{V})$,

(ii) the family $\{\varphi^{-1}(m)\}_{m \in \mathbb{Z}}$ is finite.

We denote by $\mathrm{CF}_\infty(\mathbb{V})$ the group of functions that are constructible up to infinity.

In other words (Schapira, 2020, Lemma 4.2), functions that are constructible up to infinity are restrictions to $\mathbb{V}$ of constructible functions on $\mathbb{P}(\mathbb{V})$.

Constructible functions and Euler calculus can be defined in the general setting of o-minimal geometry (Curry, Ghrist, and Robinson, 2012; Dries, 1998). *Globally subanalytic subsets*—subsets of the Euclidean space that are subanalytic in the projective compactification of the space—form an o-minimal structure and constructible functions defined in this structure are precisely the functions that are constructible up to infinity. In this thesis, we keep the notations and subanalytic framework introduced in Schapira (1989, 1991, 2020) to make use of results proven there. See for instance Proposition 5.13 below.

Example 5.9. Any compactly supported constructible function on $\mathbb{V}$ is constructible up to infinity. Indeed, the open embedding $j : \mathbb{V} \hookrightarrow \mathbb{P}(\mathbb{V})$ is proper on any compact subanalytic subset K of $\mathbb{V}$ so that $j(K)$ is subanalytic in $\mathbb{P}(\mathbb{V})$ Kashiwara and Schapira (1990, Proposition 8.2.2).

5.2.2 Convexes, cones and polyhedra

In this section, we define a subclass of constructible functions of prime interest in applications: PL-constructible functions. We refer to Rockafellar (2015) for a complete study of convex sets and functions and to Schneider (2014, Chapter 1) for a short and clear exposition.

We call *convex polyhedron* an intersection of a finite number of open or closed affine half-spaces of $\mathbb{V}$, and *convex polytope* a bounded convex polyhedron. A *polyhedral cone* is a convex polyhedron that is also a cone. If C is a non-empty convex subset of $\mathbb{V}$, we denote by $h_C : \mathbb{V}^* \to \mathbb{R} \cup \{+\infty\}$ its *support function*, defined by $h_C(\xi) = \sup_{x \in C} \langle \xi, x \rangle$. The following lemma is clear, yet useful all along the next two chapters, especially in Sections 6.2, 6.3 and 7.1.

Lemma 5.10. *Let $C \subseteq \mathbb{V}$ be a closed convex subset. Then, for any $\xi \in \mathbb{V}^*$ proper on C, we have $\xi_* \mathbf{1}_C = \mathbf{1}_{[-h_C(-\xi), h_C(\xi)]}$ where we abusively denoted $[x, +\infty] := [x, +\infty)$ and $[-\infty, x] := (-\infty, x]$, for $x \in \mathbb{R}$, and $[-\infty, +\infty] := \mathbb{R}$.*

A function $\varphi : \mathbb{V} \to \mathbb{Z}$ is said to be PL-*constructible* if there exists a finite covering $\mathbb{V} = \bigsqcup_{a \in A} P_a$ by convex polyhedra such that φ is constant on P_a for any $a \in A$. Any such function can be written as a finite sum of indicator functions of closed convex polyhedra. We denote by $\mathrm{CF}_{\mathrm{PL}}(\mathbb{V})$ the group of PL-constructible functions, $\mathrm{CF}_{\mathrm{PL,c}}(\mathbb{V})$ the subgroup of compactly supported ones. Note that $\mathrm{CF}_{\mathrm{PL,c}}(\mathbb{V}) \subset \mathrm{CF}_{\mathrm{PL}}(\mathbb{V}) \subset \mathrm{CF}_{\infty}(\mathbb{V})$.

5.2.3 γ-constructible functions

We define here a specific class of constructible functions of particular interest in this thesis due to their occurence in the context of sublevel-sets persistence (Chapter 7), the so-called γ-*constructible functions*. We prove several useful properties of these functions. For instance, their Euler integral is zero (Lemma 5.16).

Let γ be a cone of $\mathbb{V}$ such that:

$$\gamma \text{ is a subanalytic closed proper convex cone with non-empty interior.} \tag{C1}$$

We say that a subset $U \subseteq \mathbb{V}$ is *γ-open* if it is open and $U = U + \gamma$. The collection of γ-open subsets of $\mathbb{V}$ yields a topology on $\mathbb{V}$ called the *γ-topology*, see Kashiwara and Schapira (1990, Section 3.5). The closed subset of $\mathbb{V}$ for this topology, called *γ-closed* subsets, are the closed subsets S of $\mathbb{V}$ such that $S = S + \gamma^a$. A subset of $\mathbb{V}$ is called *γ-locally closed* if it is the intersection of a γ-closed subset and a γ-open subset.

Definition 5.11. We say that $\varphi \in \mathrm{CF}(\mathbb{V})$ is *γ-constructible* if

(i) $\varphi^{-1}(m)$ is subanalytic γ-locally closed in $\mathbb{V}$ for all $m \in \mathbb{Z}$, and

(ii) the family $\left\{\varphi^{-1}(m)\right\}_{m\in\mathbb{Z}}$ is locally finite.

We denote by $\mathrm{CF}_\gamma(\mathbb{V})$ the group of γ-constructible functions on $\mathbb{V}$ and $\mathrm{CF}_{\gamma,\mathrm{c}}(\mathbb{V})$ the group of compactly supported γ-constructible functions on $\mathbb{V}$.

One has the following characterization of γ-constructible functions:

Lemma 5.12. *Any compactly supported γ-constructible function φ on $\mathbb{V}$ can be decomposed as a sum $\varphi = \sum_{\alpha\in A} m_\alpha \mathbf{1}_{Z_\alpha}$ where the set A is finite and the subsets $Z_\alpha \subseteq \mathbb{V}$ are relatively compact subanalytic and γ-locally closed.*

Proof. Since the class of subanalytic γ-locally closed subsets is stable under intersection, one can intersect the Z_α's with the subset:

$$Z_v := (v + \mathrm{Int}(\gamma)) \cap (-v + \gamma^a)\,,$$

where v is any element of the non-empty set $\mathrm{Int}(\gamma^a)$ chosen so that $\mathrm{supp}(\varphi) \subseteq Z_v$. To choose such a $v \in \mathrm{Int}(\gamma^a)$, consider any $v_0 \in \mathrm{Int}(\gamma^a)$, and remark that since $\mathrm{supp}(\varphi)$ is compact and Z_{v_0} has non-empty interior Berkouk and Petit (2021, Lemma 5.7), there exists a $\lambda > 0$ such that $\mathrm{supp}(\varphi) \subseteq \lambda \cdot Z_{v_0}$. Moreover, it is easy to show that $\lambda \cdot Z_{v_0} = Z_{\lambda v_0}$ so that $v = \lambda v_0 \in \mathrm{Int}(\gamma^a)$ works. The subset Z_v is then clearly subanalytic and γ-locally closed. Moreover, loc. cit. yields that Z_v is bounded, hence relatively compact. The fact that A can be chosen finite follows then from the locally finiteness of the sum. □

Schapira proved a characterization of γ-constructible functions of which we state a specific case below (Proposition 5.13). We say that a closed set $A \subseteq \mathbb{V}$ is *γ-proper* if the addition $s : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$ is proper on $A \times \gamma^a$.

Proposition 5.13 (Schapira (2020, Proposition 4.18))**.** *Let $\varphi \in \mathrm{CF}_\infty(\mathbb{V})$ such that $\mathrm{supp}(\varphi)$ is γ-proper. The function φ is γ-constructible if and only if $\varphi = \varphi \star \mathbf{1}_{\gamma^a}$.*

As an easy consequence of this characterization, the following lemma states that the group of γ-constructible functions is closed under specific pushforwards. This fact will be useful in Sections 6.3.2 and 6.4.

Lemma 5.14. *Let $\varphi \in \mathrm{CF}_{\gamma,\mathrm{c}}(\mathbb{V})$. For any $\xi \in \gamma^{a\circ} \setminus \{0\}$, one has $\xi_*\varphi \in \mathrm{CF}_{\gamma,\mathrm{c}[\lambda]}(\mathbb{R})$ with $\lambda = \mathbb{R}_{\leq 0}$.*

Proof. Lemma 5.4 shows that $\mathrm{supp}(\xi_*\varphi) \subseteq \xi(\mathrm{supp}(\varphi))$, hence $\xi_*\varphi$ is compactly supported. Since $\xi_*\mathbf{1}_{\gamma^a} = \mathbf{1}_{\lambda^a}$, Proposition 5.13 and Corollary 5.7 yield:

$$\xi_*\varphi = \xi_*(\varphi \star \mathbf{1}_{\gamma^a}) = (\xi_*\varphi) \star \mathbf{1}_{\lambda^a}\,.$$

Hence, $\varphi \in \mathrm{CF}_\lambda(\mathbb{R})$ by Proposition 5.13. □

Remark 5.15. It follows directly from the previous lemma that for any $\xi \in \gamma^\circ \setminus \{0\}$, one has $\xi_*\varphi \in \mathrm{CF}_{\gamma,\mathrm{c}[\lambda]}(\mathbb{R})$ with $\lambda = \mathbb{R}_{\geq 0}$. We will see in Proposition 5.19 below that for any other $\xi \in \mathbb{V}^* \setminus (\gamma^\circ \cup \gamma^{a\circ})$, one has $\xi_*\varphi = 0$.

We end this section with the proof that the Euler integral of a γ-constructible function is zero.

Lemma 5.16. *If $\varphi \in \mathrm{CF}_{\gamma,\mathrm{c}}(\mathbb{V})$, then $\int_{\mathbb{V}} \varphi \, \mathrm{d}\chi = 0$.*

Proof. Choose an element ξ in the non-empty set $\gamma^{a\circ} \setminus \{0\}$. By Remark 5.3, we have:

$$\int_{\mathbb{V}} \varphi \, \mathrm{d}\chi = \int_{\mathbb{R}} \xi_*\varphi \, \mathrm{d}\chi.$$

Moreover, Lemma 5.14 shows that $\xi_*\varphi \in \mathrm{CF}_{\gamma,\mathrm{c}[\lambda]}(\mathbb{R})$ with $\lambda = \mathbb{R}_{\leq 0}$. Lemma 5.12 ensures then that one can write $\xi_*\varphi = \sum_{i=1}^n m_i \mathbf{1}_{[a_i,b_i)}$ where m_i are integers and $a_i < b_i$ are real numbers. The result follows from the fact that $\int_{\mathbb{V}} \mathbf{1}_{[a_i,b_i)} \, \mathrm{d}\chi = 0$. □

5.3 Radon transform

The success of Euler calculus in topological data analysis is due to the introduction of topological integral transforms. In this section we present one instance of the general notion of *Radon transform* introduced by Schapira (1995). Then, we describe the support of the Radon transform of compactly supported γ-constructible functions.

For the sake of readability, we use the notations $\mathbb{P} = \mathbb{P}(\mathbb{V})$ and $\mathbb{P}^* = \mathbb{P}^*(\mathbb{V})$ for the projective compactifications of $\mathbb{V}$ and $\mathbb{V}^*$ respectively. Let us denote by S the incidence relation for projective duality:

$$S = \Big\{ \big([v : \lambda], [\xi : t]\big) \in \mathbb{P} \times \mathbb{P}^* \,;\, \langle \xi, v \rangle + \lambda t = 0 \Big\},$$

and denote by $p : S \to \mathbb{P}$ and $q : S \to \mathbb{P}^*$ the restrictions of the canonical projections. We have the following diagram:

$$\begin{array}{ccccc} & & \mathbb{P} \times \mathbb{P}^* & & \\ & & \cup | & & \\ & & S & & \\ & \overset{p}{\swarrow} & & \overset{q}{\searrow} & \\ \mathbb{P} & & & & \mathbb{P}^* \end{array}$$

The *Radon transform* of $\varphi \in \mathrm{CF}(\mathbb{P})$ is the constructible function $\mathcal{R}(\varphi) \in \mathrm{CF}(\mathbb{P}^*)$ defined in Schapira (ibid.) as:

$$\mathcal{R}(\varphi) = q_* p^* \varphi.$$

Since S is a compact subset of $\mathbb{P} \times \mathbb{P}^*$, the map q is proper and $\mathcal{R}$ is well-defined. Any $\varphi \in \mathrm{CF}_{\mathrm{c}}(\mathbb{V})$ naturally yields a constructible function $j_*\varphi$ on $\mathbb{P}$, as explained in Example 5.9. In that case, Section 5 in loc. cit. ensures that for any $y = [\xi : t] \in \mathbb{P}^* \setminus \{h_\infty\}$, one has:

$$\mathcal{R}(j_*\varphi)(y) = \int_{\mathbb{V}} \varphi \cdot \mathbf{1}_{\xi^{-1}(t)} \, \mathrm{d}\chi = \xi_*\varphi(t), \tag{5.3}$$

and that $\mathrm{supp}(\mathcal{R}(j_*\varphi)) \subseteq K^*$, where

$$K^* := q\left(p^{-1}\big(\mathrm{supp}(\varphi)\big)\right) = \left\{[\xi : t] \in \mathbb{P}^* \,;\, \xi^{-1}(t) \cap \mathrm{supp}(\varphi) \neq \varnothing\right\}$$

is a compact subset of $\mathbb{P}^* \setminus \{h_\infty\}$.

A remarkable property of the Radon transform is that it satisfies an inversion formula. Define for any $\psi \in \mathrm{CF}(\mathbb{P}^*)$, the following constructible function:

$$\mathcal{R}'(\psi) = p_* q^* \psi.$$

Again, the map p is proper, so $\mathcal{R}'$ is well-defined. It is almost a left inverse for $\mathcal{R}$:

Theorem 5.17 (ibid., Corollary 5.1)**.** *Let $\varphi \in \mathrm{CF}(\mathbb{P})$. Then,*

$$\mathcal{R}' \circ \mathcal{R}(\varphi) = \begin{cases} \varphi & \text{if } \dim(\mathbb{V}) \text{ is odd,} \\ -\varphi + \int_{\mathbb{P}} \varphi \, \mathrm{d}\chi & \text{if } \dim(\mathbb{V}) \text{ is even.} \end{cases}$$

Remark 5.18. The above theorem implies that if the function φ is compactly supported and γ-constructible on $\mathbb{V}$, the map $\mathcal{R}'$ becomes a left inverse for $\mathcal{R}$ up to a sign, by Lemma 5.16.

The next proposition refines the inclusion $\mathrm{supp}(\mathcal{R}(j_*\varphi)) \subseteq K^*$ when φ is compactly supported and γ-constructible.

Proposition 5.19. *Let $\varphi \in \mathrm{CF}_{\gamma,\mathrm{c}}(\mathbb{V})$. Then,* $\mathrm{supp}(\mathcal{R}(j_*\varphi)) \subseteq K^*_\gamma$*, where*

$$K^*_\gamma := \left\{[\xi : t] \in K^* \,;\, \xi \in \gamma^\circ \cup \gamma^{a\circ}\right\}$$

is a compact subset of $\mathbb{P}^* \setminus \{h_\infty\}$.

The proposition above follows from the following two lemmas.

Lemma 5.20. *Let $\xi \in \mathbb{V}^* \setminus (\gamma^\circ \cup \gamma^{a\circ})$ and write $h = \mathrm{Ker}(\xi)$. Then, $\gamma \cap h$ is a cone of h satisfying* (C1)*.*

Proof. Let $\gamma' := \gamma \cap h$. The fact that γ' is a closed convex cone of h follows from the fact that γ and h are, and that these properties are stable under intersection. It is also clear that γ' is proper since γ is. For subanalyticity, the cone γ' is subanalytic in h by Kashiwara and Schapira (1990, Proposition 8.2.2) applied to the inclusion $h \hookrightarrow \mathbb{V}$.

Let us then prove that the interior of γ' in h, denoted by $\mathrm{Int}_h(\gamma')$, is non-empty. Since $\xi \notin (\gamma^\circ \cup \gamma^{a\circ})$, we have that :

$$\{\xi < 0\} \cap \gamma \neq \varnothing,$$
$$\{\xi > 0\} \cap \gamma \neq \varnothing.$$

Since γ is closed and convex with non-empty interior, classical convex analysis ensures that $\overline{\mathrm{Int}(\gamma)} = \gamma$, which allows to prove that in fact:

$$\{\xi < 0\} \cap \mathrm{Int}(\gamma) \neq \varnothing,$$
$$\{\xi > 0\} \cap \mathrm{Int}(\gamma) \neq \varnothing.$$

Since $\mathrm{Int}(\gamma)$ is convex it is in particular connected, and since $\xi : \mathbb{V} \to \mathbb{R}$ is continuous, the last two equations yield that $h \cap \mathrm{Int}(\gamma) \neq \varnothing$. To conclude, we prove that $h \cap \mathrm{Int}(\gamma) \subseteq \mathrm{Int}_h(\gamma')$. Indeed, if $x \in h \cap \mathrm{Int}(\gamma)$, then $x + B(0,\varepsilon) \subseteq \gamma$ for $0 < \varepsilon \ll 1$. Since $x \in h$, we have that

$$x + \big(h \cap B(0,\varepsilon)\big) = h \cap \big(x + B(0,\varepsilon)\big) \subseteq h \cap \gamma = \gamma',$$

hence $x \in \mathrm{Int}_h(\gamma)$ since $x + \big(h \cap B(0,\varepsilon)\big)$ is an open neighborhood of x in h. □

Lemma 5.21. *Let $Z \subseteq \mathbb{V}$ be a γ-locally closed subset and $\xi \in \mathbb{V}^*$. Write $h = \mathrm{Ker}(\xi)$, $Z' = Z \cap h$ and $\gamma' = \gamma \cap h$. Then, the subset Z' is γ'-locally closed in h.*

Proof. Write the γ-locally closed subset Z as $Z = U \cap S$ with U open and S closed such that $U + \gamma \subseteq U$ and $S + \gamma^a \subseteq S$, so that $Z' = \big(U \cap h\big) \cap \big(S \cap h\big)$. The subsets $U \cap h$ and $S \cap h$ are respectively open and closed in h, the subset $S \cap h$ is stable by ${\gamma'}^a$ and the subset $U \cap h$ is stable by γ'. □

We may now prove the proposition on the support of the Radon transform.

Proof of Proposition 5.19. The set K^*_γ is closed in $\mathbb{P}^*$ hence compact. Now, it is sufficient to prove the result for $\varphi = \mathbf{1}_Z$ with $Z \subseteq \mathbb{V}$ relatively compact, subanalytic and γ-locally closed since any element of $\mathrm{CF}_{\gamma,\mathrm{c}}(\mathbb{V})$ is a finite $\mathbb{Z}$-linear combination of such functions (Lemma 5.12). We thus prove that $\mathcal{R}(j_*\mathbf{1}_Z)$ vanishes on the complement of the closed set K^*_γ. Since we already know that $\mathcal{R}(j_*\mathbf{1}_Z)$ vanishes on the complement of K^*, we consider $[\xi : t] \in K^* \setminus K^*_\gamma$ and we note that in that case $\xi \in \mathbb{V}^* \setminus (\gamma^\circ \cup \gamma^{a\circ})$.

Write $Z' = Z \cap \xi^{-1}(t)$, $h = \mathrm{Ker}(\xi)$ and $\gamma' = \gamma \cap h$. By (5.3), we have the following expression for the Radon transform of $j_*\mathbf{1}_Z$:

$$\mathcal{R}(j_*\mathbf{1}_Z)([\xi : t]) = \int_{\mathbb{V}} \mathbf{1}_{Z'} \,\mathrm{d}\chi.$$

If Z' is empty, we clearly have $\mathcal{R}(j_*\mathbf{1}_Z)([\xi : t]) = 0$. Otherwise, taking $x \in Z'$, we get:

$$\int_{\mathbb{V}} \mathbf{1}_{Z'} \,\mathrm{d}\chi = \int_{\mathbb{V}} \tau_{(-x)*}\mathbf{1}_{Z'} \,\mathrm{d}\chi = \int_{\mathbb{V}} \mathbf{1}_{\tau_x^{-1}(Z')} \,\mathrm{d}\chi = \int_h \mathbf{1}_{\tau_x^{-1}(Z')} \,\mathrm{d}\chi,$$

where $\tau_u : v \in \mathbb{V} \mapsto v + u \in \mathbb{V}$ for any $u \in \mathbb{V}$. The function $\mathbf{1}_{\tau_x^{-1}(Z')}$ is compactly supported and γ'-constructible on h:

(i) γ' is a cone of h satisfying (C1) by Lemma 5.20,

(ii) $\tau_x^{-1}(Z')$ is a subanalytic γ'-locally closed subset of h by Lemma 5.21 applied to the subset $\tau_x^{-1}(Z)$, since $\tau_x^{-1}(Z') = \tau_x^{-1}(Z) \cap h$,

(iii) $\tau_x^{-1}(Z')$ is relatively compact in h since it is the intersection of the closed subset h and the relatively compact subset $\tau_x^{-1}(Z)$ in $\mathbb{V}$.

Hence, its Euler integral is zero by Lemma 5.16:

$$\int_h \mathbf{1}_{\tau_x^{-1}(Z')} \,\mathrm{d}\chi = 0.$$

□

5.4 Constructible functions as invariants of persistence modules

In this section, we explain how constructible functions on $\mathbb{R}^n$ appear as invariants of n-parameter persistence modules.

Graded-commutative algebraic approach. Let M be a pointwise finite-dimensional persistence module over $\mathbb{R}^n$. The *Hilbert function* of M (Oudot and Scoccola, 2021) is defined as $\varphi_M : t \in \mathbb{R}^n \mapsto \dim(M_t)$. Assume further that M is *finitely presentable*, that is, it is the cokernel of a morphism between free persistence modules with finite barcodes (Example 1.5). In that case, it follows from the existence of finite free resolutions (Botnan and Lesnick, 2022, Section 7.2) and the additivity of dimension with respect to exact sequences that this function is PL-constructible. We refer to it as the *constructible function associated to* M. In fact, they belong to the class of so-called *finitely presentable* functions over $\mathbb{R}^n$, that is, functions that can be written as a finite $\mathbb{Z}$-linear combination of indicating functions $\mathbf{1}_{Q_u}$ with $u \in \mathbb{R}^n$ and $Q_u = \{t \in \mathbb{R}^n \,;\, t \geq u\}$. We denote by $\mathrm{CF}_{\mathrm{fp}}(\mathbb{R}^n)$ the group of finitely presentable functions over $\mathbb{R}^n$. Note that we have $\mathrm{CF}_{\mathrm{fp}}(\mathbb{R}^n) \subseteq \mathrm{CF}_{\mathrm{PL}}(\mathbb{R}^n) \subseteq \mathrm{CF}_{\infty}(\mathbb{R}^n)$. Moreover, we also have $\mathrm{CF}_{\mathrm{fp}}(\mathbb{R}^n) \subseteq \mathrm{CF}_{\gamma}(\mathbb{R}^n)$ for the cone $\gamma = (\mathbb{R}_{\leq 0})^n$.

The constructible function associated to a persistence module can be generalized to *graded persistence modules* M over $\mathbb{R}^n$, that is, to functors $M : (\mathbb{R}^n, \leq) \to \mathsf{GrVec}$ where GrVec is the category of $\mathbb{N}_0$-graded vector spaces. A graded persistence module is thus a persistence module M together with a direct-sum decomposition $M = \bigoplus_{p \in \mathbb{N}_0} M^p$ where M^p are persistence modules. We say that M is *finitely presentable* if the persistence modules M^p are finitely presentable and only but finitely many of them are non-zero. Then, *the constructible function associated to* M is defined as $\varphi_M = \sum_{p \in \mathbb{N}_0} (-1)^p \varphi_{M^p}$.

Sheaf theoretic approach. The theory of multi-parameter persistence has been formulated in the language of sheaf theory in the seminal work of Curry (2014) and extended to the derived setting in Kashiwara and Schapira (2018). Following the formulation and the notations of loc. cit., a multi-parameter persistence module on $\mathbb{R}^n$ corresponds to an object F of the bounded derived category of γ-sheaves $\mathrm{D}^{\mathrm{b}}_{\gamma^{a\circ}}(\mathbf{k}_{\mathbb{R}^n})$ for the cone $\gamma = (\mathbb{R}_{\leq 0})^n$. This approach is almost equivalent to the commutative-algebraic one, missing only the so-called *ephemeral persistence modules*. We refer to Berkouk and Petit (2021) for a detailed comparison of the two approaches.

In this setting, a constructible persistence module corresponds to an object of the bounded derived category of constructible γ-sheaves $\mathrm{D}^{\mathrm{b}}_{\mathbb{R}\mathrm{c},\gamma^{a\circ}}(\mathbf{k}_{\mathbb{R}^n})$ and the constructible function associated to F if defined as its *local Euler-Poincaré index* (Kashiwara and Schapira, 1990, Section 9.7), defined for any $x \in \mathbb{V}$ by:

$$\chi_{\mathrm{loc}}(F)(x) = \sum_{j \in \mathbb{Z}} (-1)^j \dim H^j(F_x).$$

By a classical result of Kashiwara (1985), this invariant represents the class of F in the Grothendieck group of the category $\mathrm{D}^{\mathrm{b}}_{\mathbb{R}\mathrm{c},\gamma^{a\circ}}(\mathbf{k}_{\mathbb{R}^n})$.

Chapter 6

Hybrid transforms of constructible functions

Euler calculus allows for the introduction of sophisticated tools for shape analysis in the form of topological integral transforms. Among them, the Radon transform of Schapira (Section 5.3) and one of its specialization introduced by Turner, Mukherjee, and Boyer (2014b), the *Euler characteristic transform*. This last descriptor of shapes records the Euler characteristics of the intersections of a definable subset $Z \subset \mathbb{R}^n$ (in some fixed o-minimal structure) with all affine half-spaces. A key result is that this integral transform is injective, that is, a subset Z is entirely characterized by its Euler characteristic transform (Curry, Mukherjee, and Turner, 2022; Ghrist, Levanger, and Mai, 2018). Moreover, it is shown in Curry, Mukherjee, and Turner (2022) that, under mild assumptions on the definable subset Z, only a finite number of affine half-spaces are required to reconstruct Z from the data of its transform. However, the collection of specific half-spaces required depends on Z.

As previously stated in the introduction, this shape descriptor has been utilized in various applications such as predicting clinical outcomes in brain cancer (Crawford et al., 2020), analyzing barley seeds (Amézquita et al., 2022), and recovering morphological variations among primate genera (Tang et al., 2022). However, the instability of Euler calculus under numerical approximations remains a problematic aspect for practical applications. Even when the domain of a constructible function is finely sampled, errors are likely to occur during the integration process (Curry, Ghrist, and Robinson, 2012, Section 16). In order to produce better-behaved shape descriptors, the Euler characteristic transform is smoothed in Crawford et al. (2020). However, these experimental results lack theoretical support at present. Several smoothings of a similar nature have been introduced by Ghrist and Robinson (2011) under the name *Bessel* and *Fourier* transforms, and by Bobrowski and Borman (2012) under the name *Euler characteristic of barcodes*. We introduce a general definition encompassing all these specific smoothings in the present chapter.

Outline. In this chapter, we introduce integral transforms that combine Lebesgue integration and Euler calculus for constructible functions. The goal is to leverage the benefits of both techniques: Lebesgue integration offers well-studied kernels such as Fourier and

Laplace transforms, which yield stable and smooth integral transforms, while Euler calculus provides topological information and is compatible with operations on constructible functions. Our transforms generalize those introduced by previous authors (Bobrowski and Borman, 2012; Ghrist and Robinson, 2011) and we conduct a systematic study of hybrid transforms, introducing two new ones: the *Euler-Laplace* transform and the *Euler-Fourier* transform (Section 6.1). In terms of invariants of persistence modules, the Euler-Laplace transform provides a satisfactory—in view of the properties of hybrid transforms—generalization of Govc and Hepworth's *persistent magnitude* (2021, Definition 5.1) to constructible sheaves, so in particular to multi-parameter persistence modules. We also present numerous examples to illustrate their properties of these two new transforms, and their differences with their classical counterparts. Then, we prove their regularity (Section 6.2) and their compatibility with operations on constructible functions (Section 6.3). These properties are to be compared with regularity results on classical Fourier and Laplace transforms and their relation with classical integral operations such as convolution. Finally, we prove that the Euler-Fourier transform is injective—and provide a reconstruction formula—on the subgroup of compactly supported γ-constructible functions (Section 6.4).

Conventions. We follow the conventions and notations introduced in Chapter 5.

6.1 Definition and examples

6.1.1 Definition

We now introduce the notion of hybrid transform, which is central to Part II. Let us denote by $\mathcal{B}(\mathbb{V}^*;\mathbb{R})$ the set of bounded functions from $\mathbb{V}^*$ to $\mathbb{R}$.

Definition 6.1. For $\kappa \in L^1_{\mathrm{loc}}(\mathbb{R})$, the *hybrid transform with kernel* κ is the map $\mathrm{T}_\kappa : \mathrm{CF}_{\mathrm{c}}(\mathbb{V}) \to \mathcal{B}(\mathbb{V}^*;\mathbb{R})$ defined by:

$$\mathrm{T}_\kappa\left[\varphi\right](\xi) = \int_{\mathbb{R}} \kappa(t)\, \xi_*\varphi(t)\, \mathrm{d}t,$$

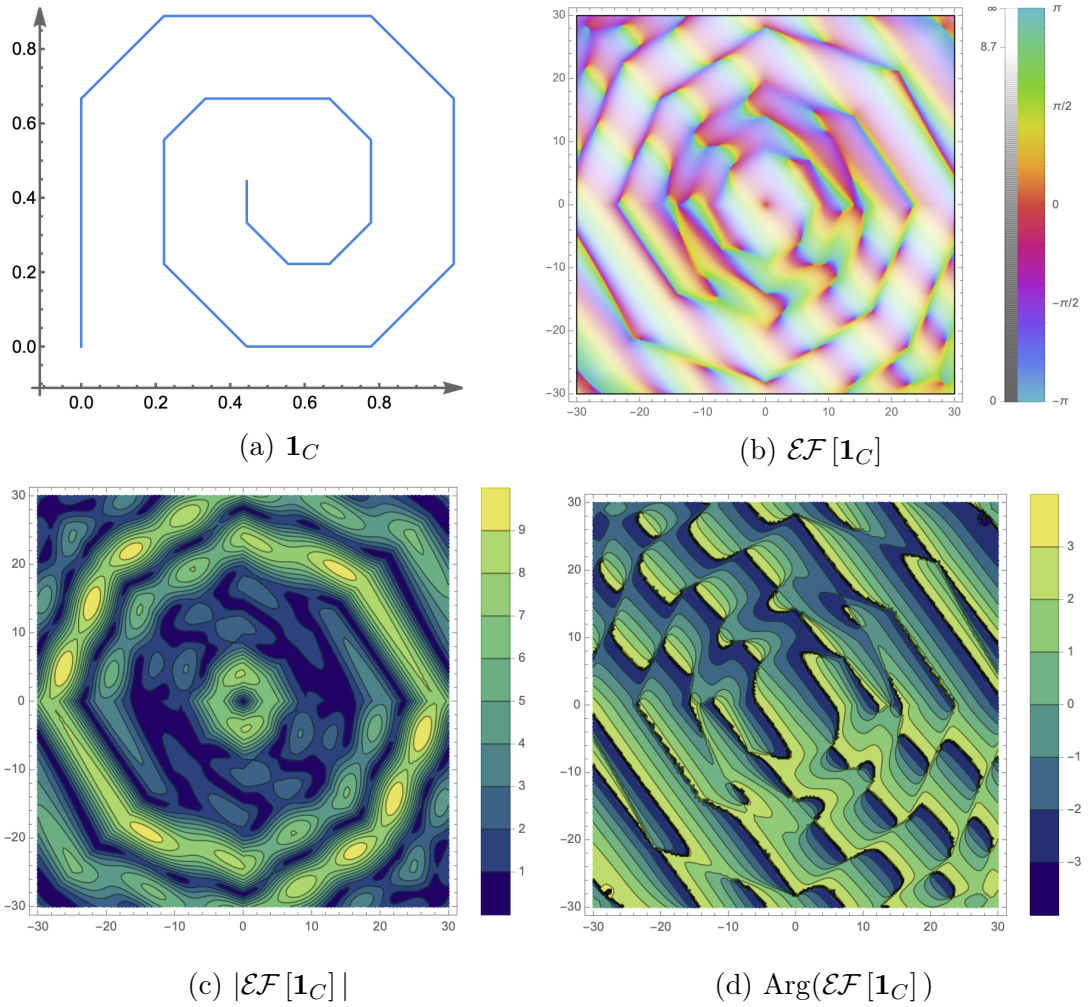
for any $\varphi \in \mathrm{CF}_{\mathrm{c}}(\mathbb{V})$ and any $\xi \in \mathbb{V}^*$.

The fact that the transform T_κ is valued in $\mathcal{B}(\mathbb{V}^*;\mathbb{R})$ follows easily from the fact that for any $\varphi \in \mathrm{CF}_{\mathrm{c}}(\mathbb{V})$, the Radon transform $\mathcal{R}(\varphi) : [\xi : t] \in \mathbb{P}^*(\mathbb{V}) \mapsto \xi_*\varphi(t)$ is bounded, as any constructible function on the compact set $\mathbb{P}^*(\mathbb{V})$.

Remarks 6.2 (Generalizations)**.** In the rest of Part II, we will consider the following generalizations of T_κ whenever necessary.

(i) The transform $\mathrm{T}_\kappa\left[\varphi\right]$ is well-defined for φ with non-compact support on the set of $\xi \in \mathbb{V}^*$ such that (i) ξ is proper on $\mathrm{supp}(\varphi)$ and (ii) $\kappa \cdot \xi_*\varphi \in L^1(\mathbb{R})$.

(ii) More generally, the definition of $\mathrm{T}_\kappa\left[\varphi\right](\zeta)$ still makes sense for any $\varphi \in \mathrm{CF}(X)$ with X a real analytic manifold, and any morphism $\zeta : X \to \mathbb{R}$ of real analytic manifolds such that (i) ζ is proper on $\mathrm{supp}(\varphi)$ and (ii) $\kappa \cdot \zeta_*\varphi \in L^1(\mathbb{R})$.



(a) $\mathbf{1}_C$ (b) $\mathcal{EF}[\mathbf{1}_C]$

(c) $|\mathcal{EF}[\mathbf{1}_C]|$ (d) $\mathrm{Arg}(\mathcal{EF}[\mathbf{1}_C])$

Figure 6.1: (a) a piecewise-linear closed curve C in $\mathbb{R}^2$, (b) the Euler-Fourier transform of the constructible function $\mathbf{1}_C$ as well as (c) its absolute value and (d) its argument. Plots are done following Remark 6.31.

Notation 6.3. When $A \subseteq \mathbb{R}$, we use the simpler notation $\mathrm{T}_A[\varphi](\xi) := \mathrm{T}_{1_A}[\varphi](\xi)$.

In the course of the chapter, we will illustrate on many examples the interest that there can be in considering hybrid transforms. All the examples will illustrate the effect of the combination of a topological operation (the pushforward) and of a classical integral. We start with the following example.

Example 6.4 (Subanalytic curve)**.** Let $c : [0,1] \to \mathbb{V}$ be continuous subanalytic. One can consider the constructible function $\mathbf{1}_Z$ where $Z = \mathrm{Im}(c)$ is compact and subanalytic in $\mathbb{V}$. Since Z has volume zero, integral transforms using only the Lebesgue measure are zero when $\dim(\mathbb{V}) \geq 2$. However, hybrid transforms are generally not, as highlighted by Figure 6.1 and Examples 6.19, 6.21, 6.28 and 6.30. This is due to the fact that the pushforwards of $\mathbf{1}_Z$ by linear forms convey topological information that is missed by the Lebesgue measure.

As explained in introduction, Definition 6.1 generalizes existing transforms, starting

with the following two.

Example 6.5. Considering $\varphi \in \mathrm{CF_c}(\mathbb{V})$ and $\eta \in \mathbb{V}^*$, the *Fourier transform* defined in Ghrist and Robinson (2011) is the hybrid transform:

$$\mathcal{EF}^{\mathrm{GR}}[\varphi](\eta) = \int_0^{+\infty} \int_{\mathbb{V}} \mathbf{1}_{\eta^{-1}(t)} \varphi \, \mathrm{d}\chi \, \mathrm{d}t = \mathrm{T}_{\mathbb{R}_{\geq 0}}[\varphi](\eta).$$

We call it *GR-Euler-Fourier transform* in this thesis, keeping the terminology *Euler-Fourier* for the transform introduced in Definition 6.26. Using Lemma 6.13, one can see that any constructible function φ satisfying Assumptions 6.9 below for a cone C satisfies also that $\mathbf{1}_{\mathbb{R}_{\geq 0}} \cdot \eta_* \varphi \in L^1(\mathbb{R})$ for any $\eta \in \mathrm{Int}(C^{a\circ})$, so that we can extend the definition of the transform to such constructible functions φ and linear forms η.

Example 6.6. Given an analytic norm on $\mathbb{V}$, the *Euler-Bessel transform* is defined in Ghrist and Robinson (ibid.) for any $\varphi \in \mathrm{CF_c}(\mathbb{V})$ and any $v \in \mathbb{V}$, by:

$$\mathcal{EB}[\varphi](v) = \int_0^{+\infty} \int_{\mathbb{V}} \mathbf{1}_{\partial B(v,t)} \varphi \, \mathrm{d}\chi \, \mathrm{d}t,$$

where $\partial B(v,t)$ denotes the sphere of radius t centered at v in $\mathbb{V}$. Considering the morphism of real analytic manifolds $\zeta_v = \|v - \cdot\|^2$ and the locally integrable kernel $\kappa : t \mapsto \mathbf{1}_{(0,+\infty)}(t) \cdot \frac{1}{2\sqrt{t}}$, the Euler-Bessel transform of $\varphi \in \mathrm{CF_c}(\mathbb{V})$ is the hybrid transform $\mathcal{EB}[\varphi](v) = \mathrm{T}_\kappa[\varphi](\zeta_v)$.

We end this section with the following lemma, expressing restrictions of multi-parameter hybrid transforms to lines as one-parameter hybrid transforms.

Lemma 6.7. *Let $\kappa \in L^1_{\mathrm{loc}}(\mathbb{R})$ and $\varphi \in \mathrm{CF_c}(\mathbb{V})$. Then, for any $\xi \in \mathbb{V}^*$ and real number $s > 0$, one has:*

$$\mathrm{T}_\kappa[\varphi](s\xi) = \mathrm{T}_\kappa[\xi_*\varphi](s) = s \int_{\mathbb{R}} \kappa(su) \xi_* \varphi(u) \, \mathrm{d}u.$$

Proof. The first equality follows from the functoriality of the pushforward: $(s\,\xi)_*\varphi = s_*\xi_*\varphi$, and the second one from the fact that $s_*\xi_*\varphi(t) = \xi_*\varphi(t/s)$ and a change of variables. □

6.1.2 Euler-Laplace transform

In this section, we introduce a hybrid transform with several appealing properties. As explained in the introduction, this transform yields a generalization of persistent magnitude. Moreover, we show in Section 6.3 that it turns constructible convolution intro products.

Definition 6.8. The *Euler-Laplace transform* of $\varphi \in \mathrm{CF_c}(\mathbb{V})$ is defined for $\xi \in \mathbb{V}^*$ by:

$$\mathcal{EL}[\varphi](\xi) = \int_{\mathbb{R}} e^{-t} \xi_* \varphi(t) \, \mathrm{d}t.$$

As explained in Remarks 6.2, we may extend the definition of the Euler-Laplace transform for any constructible function $\varphi \in \mathrm{CF}(\mathbb{V})$ on the set of $\xi \in \mathbb{V}^*$ that are proper on $\mathrm{supp}(\varphi)$ and for which the right-hand integral is well-defined. We show in Proposition 6.12 that the following assumptions on φ are sufficient for the set of such ξ's to contain an open cone of $\mathbb{V}^*$.

Assumptions 6.9.

(i) φ is constructible up to infinity,

(ii) $\mathrm{supp}(\varphi) \subseteq K + C$, where K is compact and C is a cone satisfying:

$$C \text{ is a non-empty subanalytic closed proper convex cone.} \tag{C2}$$

Note that, compared to (C1), the cone C may have empty interior.

Remark 6.10. If there exist a compactly supported constructible function φ_c on $\mathbb{V}$ and a cone C satisfying (C2) such that $\varphi = \varphi_c \star \mathbf{1}_C$, then φ satisfies Assumptions 6.9 for the cone C. The property on the support is easy to prove. Moreover, φ_c is constructible up to infinity (Example 5.9) and so is $\mathbf{1}_C$ by Schapira (2020, Lemma 2.17) since C is a subanalytic cone. The result follows then from the stability of this property by convolution, see Section 3.4 in loc. cit..

Example 6.11. As we will see in Chapter 7, the previous remark and Proposition 7.5 ensure that constructible functions associated to sublevel-sets persistent homology satisfy these assumptions (Section 7.1).

Proposition 6.12. *If $\varphi \in \mathrm{CF}(\mathbb{V})$ satisfies Assumptions 6.9 for a cone C, then the transform $\mathcal{EL}[\varphi]$ is well-defined on $\mathrm{Int}(C^\circ)$.*

The proposition follows from the next lemma.

Lemma 6.13. *Let $\varphi \in \mathrm{CF}(\mathbb{V})$. If φ satisfies Assumptions 6.9.*(ii)*, then any $\xi \in \mathrm{Int}(C^\circ) \cup \mathrm{Int}(C^{a\circ})$ is proper on $\mathrm{supp}(\varphi)$. Moreover, for any $\xi \in \mathrm{Int}(C^\circ)$ (resp. $\xi \in \mathrm{Int}(C^{a\circ})$), one has $\mathrm{supp}(\xi_*\varphi) \subseteq [a, +\infty)$ (resp. $\mathrm{supp}(\xi_*\varphi) \subseteq (-\infty, a]$) for some $a \in \mathbb{R}$.*

Proof. Let $a < b$ be two real numbers and $\xi \in \mathrm{Int}(C^\circ)$. The case $\xi \in \mathrm{Int}(C^{a\circ})$ follows by multiplication by -1. We prove that the space $\xi^{-1}[a, b] \cap \mathrm{supp}(\varphi)$ is compact. Since this space is closed and by assumption:

$$\xi^{-1}[a, b] \cap \mathrm{supp}(\varphi) \subseteq \xi^{-1}[a, b] \cap (K + C)\,,$$

it is enough to show that this last space is compact. Suppose that there exist a sequence $y_n = k_n + x_n \in \xi^{-1}[a, b] \cap (K + C)$ with $k_n \in K$ and $x_n \in C$ such that $\|y_n\| \to +\infty$. Since K is compact, we also have $\|x_n\| \to +\infty$. If we denote by S the unit sphere of $\mathbb{V}$, the subset $C \cap S$ is compact by closedness of C. We can thus assume without loss of generality that $x_n/\|x_n\| \to v \in C \cap S$. Dividing the inequality $a \leq \langle \xi, y_n \rangle \leq b$ by $\|x_n\|$ and taking the limit, we get that $\langle \xi, v \rangle = 0$. Yet, $\xi \in \mathrm{Int}(C^\circ)$ so that $\langle \xi, v \rangle > 0$, a contradiction. □

Proof of Proposition 6.12. For any $\xi \in \mathrm{Int}(C^\circ)$, it follows from Lemma 6.13 that ξ is proper on $\mathrm{supp}(\varphi)$ and that $\mathrm{supp}(\xi_*\varphi) \subseteq [a, +\infty)$ for some $a \in \mathbb{R}$. The function $\xi_*\varphi$ is constructible up to infinity by Schapira (2020, Lemma 4.10), thus takes only a finite number of distinct values. Together with the property on the support of $\xi_*\varphi$, this yields that $t \in \mathbb{R} \mapsto e^{-t} \cdot \xi_*\varphi(t)$ is integrable over $\mathbb{R}$, hence the result. □

Remark 6.14. In view of the proof of Proposition 6.12, the same conclusion holds for any hybrid transform with kernel $\kappa \in L^1(\mathbb{R}^+)$.

Before moving on to the examples, we state a relation between the Euler-Laplace transform and the usual Laplace transform, useful all along Part II. It is a straightforward consequence of Lemma 6.7.

Lemma 6.15. *If $\varphi \in \mathrm{CF}(\mathbb{V})$ satisfies Assumptions 6.9 for a cone C, then for any $\xi \in \mathrm{Int}(C^\circ)$ and any $s > 0$, one has*

$$\mathcal{EL}\left[\varphi\right](s\,\xi) = \mathcal{EL}\left[\xi_*\varphi\right](s) = s\,\mathcal{L}[\xi_*\varphi](s),$$

where $\mathcal{L}f(s) = \int_{\mathbb{R}} e^{-st} f(t)\,\mathrm{d}t$ is the classical (bilateral) Laplace transform.

We now turn to the examples. Once again, the reader's attention is drawn to the effect of the successive application of topological pushforward and of classical integral.

Example 6.16 (Interval)**.** Let $-\infty < a \leq b < +\infty$ and consider the constructible function $\mathbf{1}_{\lceil a,b\rfloor}$ where $\lceil a, b\rfloor$ is one of the intervals $[a, b]$, $(a, b]$, $[a, b)$, or $[a, b]$. We have, for any $\xi \in \mathbb{R}$,

$$\mathcal{EL}\left[\mathbf{1}_{\lceil a,b\rfloor}\right](\xi) = \mathrm{sgn}(\xi)\left(e^{-\xi\cdot a} - e^{-\xi\cdot b}\right),$$

and for $\xi \in \mathbb{R}_{>0}$,

$$\mathcal{EL}\left[\mathbf{1}_{\lceil a,+\infty)}\right](\xi) = e^{-\xi\cdot a}.$$

Example 6.17 (Rectangle)**.** Let $a < b$ and $c < d$ be real numbers. We have, for $\xi \in \left(\mathbb{R}^2\right)^*$ such that $\xi(a,c) < \xi(a,d) < \xi(b,c) < \xi(b,d)$, the formulae:

$$\begin{aligned} &\xi_*\mathbf{1}_{[a,b)\times[c,d)} = \mathbf{1}_{[\xi(a,c),\xi(a,d))} - \mathbf{1}_{[\xi(b,c),\xi(b,d))}, && \xi_*\mathbf{1}_{[a,b)\times[c,d]} = \mathbf{1}_{[\xi(a,c),\xi(b,c))}, \\ &\xi_*\mathbf{1}_{(a,b]\times[c,d]} = \mathbf{1}_{(\xi(a,d),\xi(b,d)]}, && \xi_*\mathbf{1}_{[a,b]\times[c,d]} = \mathbf{1}_{[\xi(a,c),\xi(b,d)]}. \end{aligned} \tag{6.1}$$

Indeed, note that $\xi_*\mathbf{1}_{[a,b]\times[c,d]} = \mathbf{1}_{[\xi(a,c),\xi(b,d)]}$ is given by Lemma 5.10 and the other equalities are obtained by additivity. For instance, $\mathbf{1}_{(a,b]\times[c,d]} = \mathbf{1}_{[a,b]\times[c,d]} - \mathbf{1}_{[a,a]\times[c,d]}$. Equation (6.1) yields:

$$\begin{aligned} \mathcal{EL}\left[\mathbf{1}_{[a,b)\times[c,d)}\right](\xi) &= e^{-\xi(a,c)} - e^{-\xi(a,d)} - e^{-\xi(b,c)} + e^{-\xi(b,d)}, \\ \mathcal{EL}\left[\mathbf{1}_{[a,b)\times[c,d]}\right](\xi) &= e^{-\xi(a,c)} - e^{-\xi(b,c)}, \\ \mathcal{EL}\left[\mathbf{1}_{(a,b]\times[c,d]}\right](\xi) &= e^{-\xi(a,d)} - e^{-\xi(b,d)}, \\ \mathcal{EL}\left[\mathbf{1}_{[a,b]\times[c,d]}\right](\xi) &= e^{-\xi(a,c)} - e^{-\xi(b,d)}. \end{aligned}$$

Similar formulae can be obtained when ξ induces a different order on the vertices of the rectangle $[a, b] \times [c, d]$. Unlike the classical Laplace transform, the Euler-Laplace transform distinguishes between the presence or absence of the edges of the rectangle.

Remark 6.18. The last Euler-Laplace transform computed in Example 6.17 yields a counter-example to the formula $\mathcal{EL}[\varphi \boxtimes \psi] = \mathcal{EL}[\varphi] \boxtimes \mathcal{EL}[\psi]$, which is wrong in general. We will give a correct formula in Corollary 6.44.

Example 6.19 (Sphere)**.** Assume that $\mathbb{V}$ is equipped with an analytic norm $\|\cdot\|$. Consider $r > 0$ and $\varphi = \mathbf{1}_{\mathbb{S}_r}$ with $\mathbb{S}_r = \{x \in \mathbb{V}\,;\, \|x\| = r\}$. For any $\xi \in \mathbb{V}^*$, we have:

$$\xi_* \mathbf{1}_{\mathbb{S}_r} = \left(1 + (-1)^{\dim(\mathbb{V})}\right) \cdot \mathbf{1}_{(-r\|\xi\|, r\|\xi\|)} + \mathbf{1}_{\{-r\|\xi\|\}} + \mathbf{1}_{\{r\|\xi\|\}},$$

and hence:

$$\mathcal{EL}[\mathbf{1}_{\mathbb{S}_r}](\xi) = 2 \cdot \left(1 + (-1)^{\dim(\mathbb{V})}\right) \cdot \sinh(r\|\xi\|).$$

Note the amount of information extracted by this transform even though the constructible function under consideration is supported on a subset with zero volume.

The Euler-Laplace and Laplace transforms are equal up to a normalization on some γ-constructible functions (Example 6.46). However, they are not on all $\mathrm{CF}_{\gamma,c}(\mathbb{V})$, as shown by the following example.

Example 6.20 (γ-triangle)**.** Let $b \in \mathbb{R}$. Consider the triangle

$$T = \mathrm{Conv}(\{(0,0),(1,0),(0,b)\}) \setminus \left\{(x,y) \in \mathbb{R}^2\,;\, y = b - bx\right\},$$

represented in Figure 6.2. The subset $T \subset \mathbb{R}^2$ is subanalytic and γ-locally closed for the cone $\gamma = (\mathbb{R}_{\leq 0})^2$. Then, the Euler-Laplace and classical Laplace transforms compare as follows. For $\xi = (\xi_x, \xi_y) \in (\mathbb{R}_{\geq 0})^2$,

$$\mathcal{EL}[\mathbf{1}_T](\xi) = \begin{cases} 1 - e^{-\xi_x} & \text{if } b\xi_y \geq \xi_x, \\ 1 - e^{-b\xi_y} & \text{otherwise,} \end{cases}$$

and

$$\mathcal{L}[\mathbf{1}_T](\xi) \overset{\text{def}}{=} \int_{\mathbb{R}^2} e^{-\langle \xi, x\rangle}\, \mathbf{1}_T(x)\, \mathrm{d}x = \frac{\xi_x\left(1 - e^{-b\xi_y}\right) + b\xi_y\left(e^{-\xi_x} - 1\right)}{\xi_x(\xi_x - b\xi_y)\xi_y}.$$

These two transforms differ, as $\mathcal{EL}[\mathbf{1}_T](\xi)$ does not depend on $b \geq 1$ for any $\xi \in (\mathbb{R}_{\geq 0})^2$ such that $\xi_y \geq \xi_x$. Similar formulae hold for other choices of $\xi \in \mathbb{R}^2$.

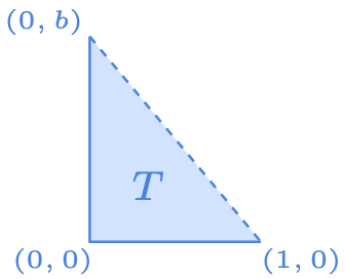
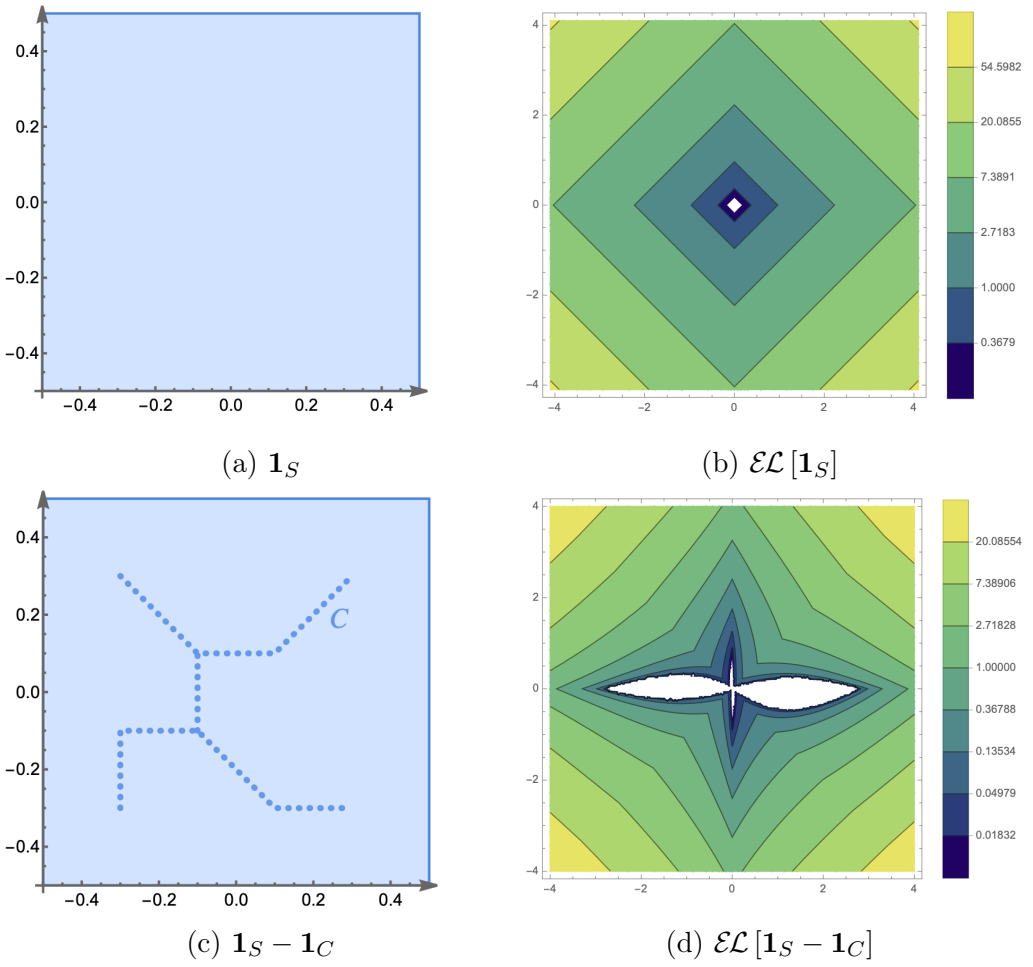


Figure 6.2: The triangle T defined in Example 6.20 is represented as the light blue solid triangle and the dark blue solid angle defined by the points $(0, b)$, $(0, 0)$ and $(1, 0)$. The dashed line indicates that points on this edge do not belong to T.

Example 6.21 (Closed square minus a curve)**.** Consider the constructible function

$$\varphi = \mathbf{1}_S - \mathbf{1}_C,$$

where $S = [-1/2, 1/2]^2$ and C is the closed curve of $\mathbb{R}^2$ represented by the dotted line in Figure 6.3c. Since C has zero volume, the classical Laplace transforms of $\mathbf{1}_S$ and $\mathbf{1}_S - \mathbf{1}_C$ are equal. However, their Euler-Laplace transforms differ, as shown in Figure 6.3.



(a) $\mathbf{1}_S$ (b) $\mathcal{EL}[\mathbf{1}_S]$

(c) $\mathbf{1}_S - \mathbf{1}_C$ (d) $\mathcal{EL}[\mathbf{1}_S - \mathbf{1}_C]$

Figure 6.3: Euler-Laplace transforms of the constructible functions $\mathbf{1}_S$ and $\mathbf{1}_S - \mathbf{1}_C$ in Example 6.21. The square S is represented by the light blue solid square and the closed curve C is represented by the dark blue dotted curve.

Generalization of persistent magnitude to constructible sheaves. The notion of *persistent magnitude* has recently been introduced in Govc and Hepworth (2021) for one-parameter persistence modules. As proven in loc. cit. following the work of Otter (2022), this object is related to another invariant of finite metric spaces called *magnitude* and introduced in Leinster (2013). Govc and Hepworth (2021) also present a generalization of it to multi-parameter persistence modules using the classical Laplace transform. Here, we use the Euler-Laplace transform to give an alternative generalization

of persistent magnitude that benefits from the compatibility (Section 6.3) and index formulae (Theorem 7.17) that come from hybridization. We define persistent magnitude not only for multi-parameter persistence modules but also for (derived) constructible sheaves, following the approach of Kashiwara and Schapira (2018) for multi-parameter persistence, recalled in Section 5.4. Following the notations of Kashiwara and Schapira (ibid.), we denote by $\mathrm{D}^{\mathrm{b}}_{\mathbb{R}\mathrm{c}}(\mathbf{k}_{\mathbb{V}})$ the bounded derived category of constructible sheaves of $\mathbf{k}$-vector spaces on $\mathbb{V}$ and by $\mathrm{D}^{\mathrm{b}}_{\mathrm{c},\mathbb{R}\mathrm{c}}(\mathbf{k}_{\mathbb{V}})$ its full subcategory generated by compactly supported objects.

Definition 6.22. The *magnitude* of a sheaf $F \in \mathrm{D}^{\mathrm{b}}_{\mathrm{c},\mathbb{R}\mathrm{c}}(\mathbf{k}_{\mathbb{V}})$ is the Euler-Laplace transform $\mathcal{EL}\left[\chi_{\mathrm{loc}}(F)\right]$ where $\chi_{\mathrm{loc}}(F) \in \mathrm{CF}(\mathbb{V})$ denotes the local Euler-Poincaré index of F (Section 5.4).

Remark 6.23. More generally, the magnitude of $F \in \mathrm{D}^{\mathrm{b}}_{\mathbb{R}\mathrm{c}}(\mathbf{k}_{\mathbb{V}})$ is well-defined when (i) F is constructible up to infinity in the sense of Schapira (2020, Definition 2.8) and (ii) $\mathrm{supp}(F) \subseteq K + C$ with K compact and C a cone satisfying (C2). Indeed, in that case, $\chi_{\mathrm{loc}}(F)$ is constructible up to infinity as proven by point (i) in the proof of Schapira (ibid., Theorem 4.4), and satisfies Assumptions 6.9.(ii).

Example 6.24. For instance, if $F \simeq F_c \star \mathbf{k}_C$ with F_c compactly supported and constructible, then $\chi_{\mathrm{loc}}(F)$ satisfies Assumptions 6.9 for the cone C by Remark 6.10.

Example 6.25. In the case $\mathbb{V} = \mathbb{R}$ and $\gamma = \mathbb{R}_{\leq 0}$, we recover the definition introduced in Govc and Hepworth (2021) of the persistent magnitude for one-parameter persistence modules. Indeed, considering $F \simeq \bigoplus_{i=1}^{n} \mathbf{k}_{[a_i,b_i)}$ with $-\infty < a_i \leq b_i \leq +\infty$, Example 6.16 yields for $t \in \mathbb{R}_{>0}$:

$$\mathcal{EL}\left[\chi_{\mathrm{loc}}(F)\right](t) = \sum_{i=1}^{n} e^{-a_i t} - e^{-b_i t},$$

with the convention that $e^{-\infty} = 0$.

6.1.3 Euler-Fourier transform

Definition 6.26. The *Euler-Fourier transform* of $\varphi \in \mathrm{CF}_{\mathrm{c}}(\mathbb{V})$ is defined for $\xi \in \mathbb{V}^*$ by:

$$\mathcal{EF}\left[\varphi\right](\xi) = \int_{\mathbb{R}} e^{-it} \xi_* \varphi(t)\, \mathrm{d}t.$$

The following lemma relates the Euler-Fourier transform to the usual Fourier transform. Again, it straightforwardly follows from Lemma 6.7.

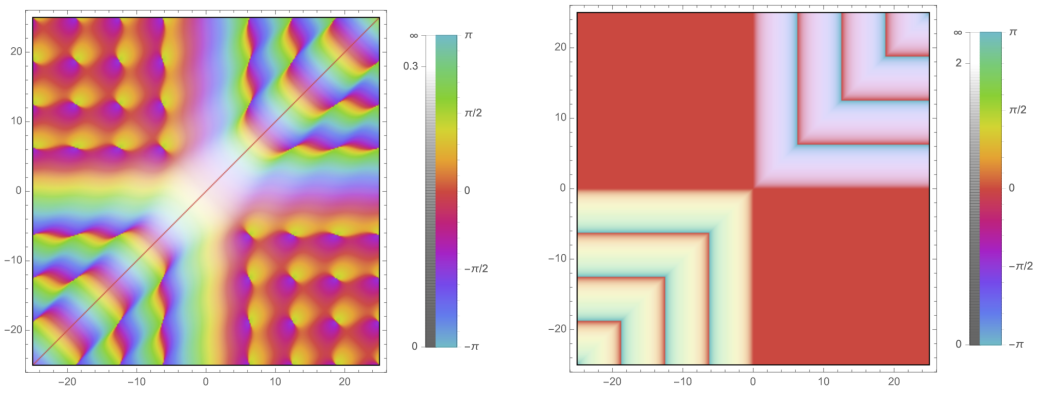
Lemma 6.27. *Let $\varphi \in \mathrm{CF}_{\mathrm{c}}(\mathbb{V})$. Then, for any $\xi \in \mathbb{V}^*$ and any $s \neq 0$, one has*

$$\mathcal{EF}\left[\varphi\right](s\,\xi) = \mathcal{EF}\left[\xi_*\varphi\right](s) = |s|\, \mathcal{F}[\xi_*\varphi](s),$$

where $\mathcal{F}f(s) = \int_{\mathbb{R}} e^{-ist} f(t)\, \mathrm{d}t$ is the classical Fourier transform.

Example 6.28 (Sphere)**.** Consider the setting of Example 6.19. Then, for any $\xi \in \mathbb{V}^*$,

$$\mathcal{EF}\left[\mathbf{1}_{\mathbb{S}_r}\right](\xi) = 2 \cdot \left(1 + (-1)^{\dim(\mathbb{V})}\right) \cdot \sin\left(r\|\xi\|\right).$$



(a) $\mathcal{F}[\mathbf{1}_T]$ (b) $\mathcal{EF}[\mathbf{1}_T]$

Figure 6.4: Fourier and Euler-Fourier transforms of the constructible function $\mathbf{1}_T$ defined in Example 6.20. Again, plots are done following Remark 6.31.

Example 6.29 (γ-triangle)**.** Consider the setting of Example 6.20. Then, the Euler-Fourier and classical Fourier transforms compare as follows. For any $\xi = (\xi_x, \xi_y) \in (\mathbb{R}_{\geq 0})^2$,

$$\mathcal{EF}[\mathbf{1}_T](\xi) = \begin{cases} i(e^{-i\xi_x} - 1) & \text{if } b\xi_y \geq \xi_x, \\ i(e^{-ib\xi_y} - 1) & \text{else,} \end{cases}$$

and

$$\mathcal{F}[\mathbf{1}_T](\xi) \stackrel{\text{def}}{=} \int_{\mathbb{R}^2} e^{-i\langle \xi, x\rangle} \mathbf{1}_T(x)\,\mathrm{d}x = \frac{\xi_x\left(1 - e^{-ib\xi_y}\right) + b\xi_y\left(e^{-i\xi_x} - 1\right)}{\xi_x(b\xi_y - \xi_x)\xi_y}.$$

Again, these two transforms differ, as $\mathcal{EF}[\mathbf{1}_T](\xi)$ does not depend on $b \geq 1$ for any ξ such that $\xi_y \geq \xi_x$. See Figure 6.4 for an illustration.

Example 6.30 (Closed square minus a curve)**.** Consider the setting of Example 6.21. Again, the (classical) Fourier transforms of $\mathbf{1}_S$ and of $\mathbf{1}_S - \mathbf{1}_C$ are equal. However, their Euler-Fourier transforms differ, as shown in Figure 6.5.

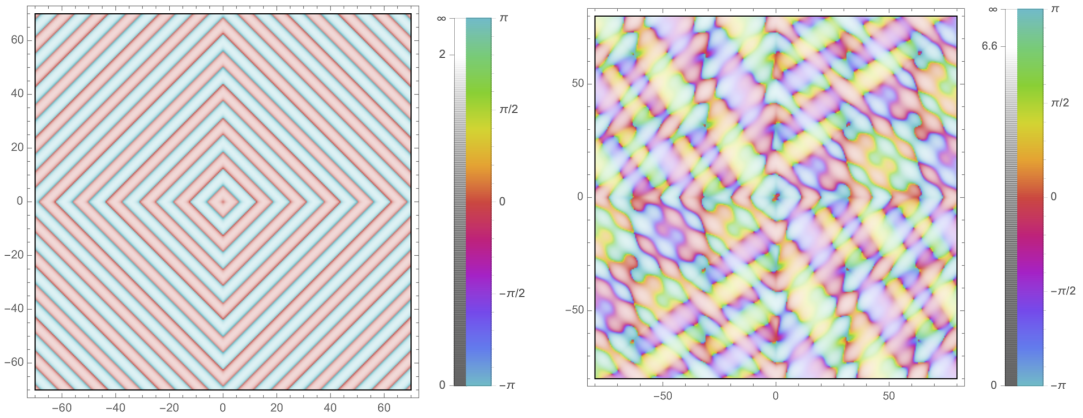
6.1.4 Computations

In this section, we explain how to compute hybrid transforms of PL-constructible functions without computing any integral with respect to the Euler characteristic.

Consider a kernel $\kappa \in L^1_{\text{loc}}(\mathbb{R})$ and $\varphi \in \mathrm{CF}_{\mathrm{PL,c}}(\mathbb{V})$. The transform $\mathrm{T}_\kappa[\varphi]$ can be efficiently computed as follows. One can write $\varphi = \sum_{i\in I} m_i \cdot \mathbf{1}_{P_i}$ where the set I is finite, the coefficients m_i are integers and the subsets P_i are compact polytopes. By $\mathbb{Z}$-linearity of hybrid transforms and the formula for the pushforward of a closed convex subset (Lemma 5.10), we have for any $\xi \in \mathbb{V}^*$,

$$\mathrm{T}_\kappa[\varphi](\xi) = \sum_{i\in I} m_i \cdot \int_{-h_{P_i}(-\xi)}^{h_{P_i}(\xi)} \kappa(t)\,\mathrm{d}t. \tag{6.2}$$

Hence, if one can express on the one hand $h_P(\xi)$ as an explicit function of $\xi \in \mathbb{V}^*$ and of the vertices of the compact polytope P, and on the other hand the integral $\int_a^b \kappa(t)\,\mathrm{d}t$



(a) $\mathcal{EF}[\mathbf{1}_S]$ (b) $\mathcal{EF}[\mathbf{1}_S - \mathbf{1}_C]$

Figure 6.5: Euler-Fourier transforms of the constructible functions $\mathbf{1}_S$ and $\mathbf{1}_S - \mathbf{1}_C$ in Example 6.21. Again, plots are done following Remark 6.31.

as an explicit function of the real numbers $a < b$, then one can compute the right-hand side of (6.2). Each example in this chapter is computed following the above methodology, computing explicit closed formulae by hand for (6.2) and plotting them using Mathematica (2021).

Remark 6.31. Complex valued functions $g : \mathbb{R}^2 \to \mathbb{C}$ are plotted using the function `ComplexPlot` of the Wolfram language. For $x \in \mathbb{R}^2$, the argument of $g(x)$ is plotted using a fixed color function from $-\pi$ to π and the absolute value of $g(x)$ is represented as a level of brightness of this color.

Software. A software that automatically computes hybrid transforms of constructible functions defined on embedded cubical complexes is available on GitHub: `https://github.com/HugoPasse/Transforms-of-cubical-complexes`. It is running in `Python` and `C++`. This is joint work with Steve Oudot and Hugo Passe.

6.2 Regularity

In this section, we consider a kernel $\kappa \in L^1_{\mathrm{loc}}(\mathbb{R})$ and study the regularity of hybrid transforms in the particular case of PL-constructible functions. While being less general, this class of functions is of prime interest in applications.

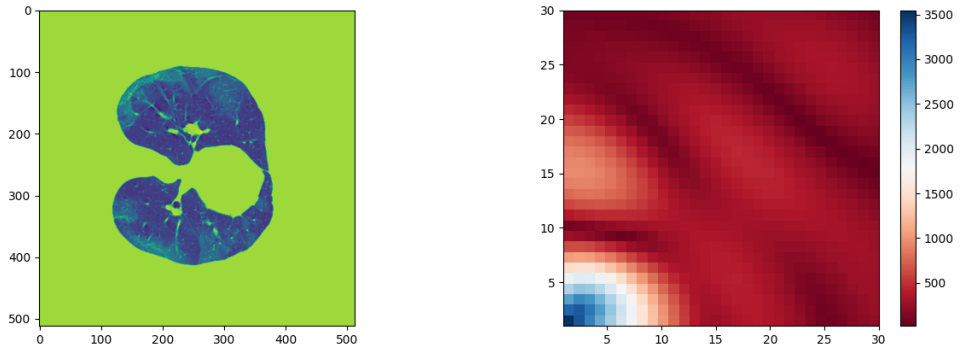
Proposition 6.32 (Continuity)**.** *Let* $\varphi \in \mathrm{CF}_{\mathrm{PL}}(\mathbb{V})$.

(i) *If* $\mathrm{supp}(\varphi)$ *is compact, then* $\mathrm{T}_\kappa[\varphi]$ *is continuous on* $\mathbb{V}^*$.

(ii) *If* $\mathrm{supp}(\varphi) \subseteq K + C$ *with* K *convex compact and* $C \neq \{0\}$ *a non-empty closed convex cone, then*

(a) $\mathrm{T}_\kappa[\varphi]$ *is continuous on* $\mathrm{Int}(C^\circ)$ *when* $\kappa \in L^1(\mathbb{R}_{\geq 0})$,

(b) $\mathrm{T}_\kappa[\varphi]$ *is continuous on* $\mathrm{Int}(C^{a\circ})$ *when* $\kappa \in L^1(\mathbb{R}_{\leq 0})$.



(a) (b)

Figure 6.6: (a) segmented CT scan of lungs taken from Morozov et al. (2020) seen as a constructible function over a cubical complex and (b) the modulus of its Euler-Fourier transform as computed by our `Python` library.

Proof. It is sufficient to prove the result for $\varphi = \mathbf{1}_P$ where P is a closed convex polyhedron included in $\mathrm{supp}(\varphi)$ since any $\varphi \in \mathrm{CF}_{\mathrm{PL}}(\mathbb{V})$ is a finite $\mathbb{Z}$-linear combination of such functions. By Lemma 5.10, for any $\xi \in \mathbb{V}^*$ proper on P, we have that $\xi_* \mathbf{1}_P = \mathbf{1}_{[-h_P(-\xi), h_P(\xi)]}$, and thus

$$\mathrm{T}_\kappa \left[\mathbf{1}_P\right](\xi) = \int_{-h_P(-\xi)}^{h_P(\xi)} \kappa(t)\, \mathrm{d}t, \tag{6.3}$$

so that it is sufficient to study the continuity of h_P.

As any convex function, h_P is continuous on $\mathrm{Int}(\mathrm{dom}(h_P))$, where $\mathrm{dom}(h_P) = \{\eta \in \mathbb{V}^* \,;\, h_P(\eta) < +\infty\}$ by Schneider (2014, Theorem 1.5.4). If P is compact, then $\mathrm{dom}(h_P) = \mathbb{V}^*$, hence (i). Now, suppose that P is not compact and assume $\kappa \in L^1(\mathbb{R}_{\geq 0})$, the other case being symmetric. Using classical polyhedra theory, see Ziegler (2012, Theorem 1.2), one can write P as a sum $K' + C'$ where K' is a convex polytope and C' a closed convex polyhedral cone. Since we assumed $P \subseteq \mathrm{supp}(\varphi)$, we also have $C' \subseteq C$. Then, for any $\xi \in \mathrm{Int}(C^\circ)$, one has $\xi \in \mathrm{Int}(C'^\circ)$, so that $h_P(\xi) = +\infty$, and $h_P(-\xi) = h_{K'}(-\xi)$. Since K' is compact, $h_{K'}$ is continuous on $\mathbb{V}^*$, and so is $h_P \circ a$ on $\mathrm{Int}(C^\circ)$ with $a(\xi) = -\xi$ the antipodal map. The result follows then from (6.3). □

Proposition 6.33. *Let $\varphi \in \mathrm{CF}_{\mathrm{PL}}(\mathbb{V})$ and $p \in \mathbb{N}_0$. Assume that κ is C^p. There exists a finite family $\{\Gamma_1, \ldots, \Gamma_m\}$ of open convex polyhedral cones of $\mathbb{V}^*$ such that $\mathbb{V}^* = \bigcup_{i=1}^m \overline{\Gamma_i}$ and:*

(i) *If $\mathrm{supp}(\varphi)$ is compact, then the restriction $\mathrm{T}_\kappa[\varphi]_{|\Gamma_i}$ is C^{p+1} for all $i \in \{1, \ldots, m\}$.*

(ii) *If $\mathrm{supp}(\varphi) \subseteq K + C$ with K convex compact and $C \neq \{0\}$ a non-empty closed convex cone, then*

   (a) *the restriction $\mathrm{T}_\kappa[\varphi]_{|\Gamma_i \cap \mathrm{Int}(C^\circ)}$ is C^{p+1} for all $i \in \{1, \ldots, m\}$ when $\kappa \in L^1(\mathbb{R}_{\geq 0})$,*

   (b) *the restriction $\mathrm{T}_\kappa[\varphi]_{|\Gamma_i \cap \mathrm{Int}(C^{a\circ})}$ is C^{p+1} for all $i \in \{1, \ldots, m\}$ when $\kappa \in L^1(\mathbb{R}_{\leq 0})$.*

Proof. The result (ii) follows from (i) similarly as in the proof of Proposition 6.32. Moreover, result (i) follows from (6.3) and the fact that if P is a polytope, the support function h_P is smooth outside the closed set E of linear forms which are orthogonal to at least one face of dimension 1 of P. Indeed, on each connected component of $\mathbb{V}^* \setminus E$, there exists a vertex v of P such that $h_P(\xi) = \langle \xi, v \rangle$. It is easy to check that each connected component Γ_i of $\mathbb{V}^* \setminus E$ is an open convex polyhedral cone. Moreover, the set E being a finite union of subspaces of $\mathbb{V}^*$ of codimension at least 1, it is closed and has empty interior. Hence, $\mathbb{V}^* = \overline{\mathbb{V}^* \setminus E} = \bigcup_{i=1}^m \overline{\Gamma_i}$. □

Example 6.34. For the Euler-Laplace and Euler-Fourier transforms, the previous result is well illustrated by Figures 6.1 and 6.3 to 6.5.

6.3 Compatibility with operations

In this section, we consider $\kappa \in L^1_{\mathrm{loc}}(\mathbb{R})$ and investigate the compatibility of hybrid transforms with operations on constructible functions. The results use the general form of hybrid transforms defined in Remarks 6.2.

6.3.1 Direct image, duality and projection

Proposition 6.35 (Direct image)**.** *Let $\varphi \in \mathrm{CF}(\mathbb{V})$, and let $f : \mathbb{V} \to \mathbb{V}'$ and $\zeta : \mathbb{V}' \to \mathbb{R}$ be morphisms of real analytic manifolds. Assume that $\zeta \circ f$ is proper on* $\mathrm{supp}(\varphi)$ *and that $\kappa \cdot \zeta_* f_* \varphi \in L^1(\mathbb{R})$. Then,*

$$\mathrm{T}_\kappa \left[f_* \varphi \right] (\zeta) = \mathrm{T}_\kappa \left[\varphi \right] (f^* \zeta).$$

In particular, if f and ζ are linear maps, denoting ${}^t f : \mathbb{V}'^* \to \mathbb{V}^*$ the dual map, we get:

$$\mathrm{T}_\kappa \left[f_* \varphi \right] (\zeta) = \mathrm{T}_\kappa \left[\varphi \right] \left({}^t f(\zeta) \right).$$

Proof. By functoriality of the pushforward, we have $\zeta_* f_* = (\zeta \circ f)_* = (f^* \zeta)_*$. □

Example 6.36. Let $x_0 \in \mathbb{V}$ and consider the map $\tau_{x_0} : \mathbb{V} \to \mathbb{V}$ given by $x \mapsto x + x_0$. For $\varphi \in \mathrm{CF_c}(\mathbb{V})$, we have $\tau_{x_0 *} \varphi(x) = \varphi(x - x_0)$ for any $x \in \mathbb{V}$. Moreover, Proposition 6.35 yields for any $\xi \in \mathbb{V}^*$,

$$\mathcal{EL} \left[\tau_{x_0 *} \varphi \right] (\xi) = e^{-\langle \xi, x_0 \rangle} \cdot \mathcal{EL} \left[\varphi \right] (\xi),$$
$$\mathcal{EF} \left[\tau_{x_0 *} \varphi \right] (\xi) = e^{-i\langle \xi, x_0 \rangle} \cdot \mathcal{EF} \left[\varphi \right] (\xi).$$

Let $\varphi \in \mathrm{CF}(\mathbb{V})$. Choose an analytic norm on $\mathbb{V}$ and consider the open ball $B(x, \varepsilon)$ with center $x \in \mathbb{V}$ and radius $\varepsilon > 0$. For $\varepsilon > 0$ small enough, the integer

$$\mathrm{D}_{\mathbb{V}} \varphi(x) = \int_{\mathbb{V}} \mathbf{1}_{B(x,\varepsilon)} \varphi \, \mathrm{d}\chi,$$

depends only on φ; see Schapira (1991, Lemma and Definition 2.4). Moreover, Theorem 2.5 of loc. cit. ensures that the function $\mathrm{D}_{\mathbb{V}} \varphi : \mathbb{V} \to \mathbb{Z}$ is constructible and called the *dual* of φ.

Proposition 6.37 (Duality)**.** *Let $\varphi \in \mathrm{CF}(\mathbb{V})$ and $\zeta : \mathbb{V} \to \mathbb{R}$ be a morphism of real analytic manifolds. Assume that ζ is proper on $\mathrm{supp}(\varphi)$ and that $\kappa \cdot \zeta_*\varphi \in L^1(\mathbb{R})$. Then, $\kappa \cdot \zeta_*(\mathrm{D}_{\mathbb{V}}\varphi) \in L^1(\mathbb{R})$, and:*

$$\mathrm{T}_\kappa \left[\mathrm{D}_{\mathbb{V}}\varphi\right](\zeta) = -\mathrm{T}_\kappa \left[\varphi\right](\zeta).$$

Proof. By Schapira (1991, Theorem 2.5), we have that $\zeta_* \left(\mathrm{D}_{\mathbb{V}}\varphi\right) = \mathrm{D}_{\mathbb{R}}(\zeta_*\varphi)$, so it suffices to show that:

$$\int_{\mathbb{R}} \kappa(t)\mathrm{D}_{\mathbb{R}}(\zeta_*\varphi)(t)\,\mathrm{d}t = -\int_{\mathbb{R}} \kappa(t)\zeta_*\varphi(t)\,\mathrm{d}t, \tag{6.4}$$

provided that the integrals make sense. Yet, a direct computation yields $\mathrm{D}_{\mathbb{R}}\mathbf{1}_{[a,b]} = -\mathrm{D}_{\mathbb{R}}\mathbf{1}_{(a,b)}$ for any two real numbers $a < b$ and $\zeta_*\varphi$ is equal to a finite $\mathbb{Z}$-linear combination of such functions outside a discrete set of points as any constructible function on $\mathbb{R}$. Thus,

$$\mathrm{D}_{\mathbb{R}}(\zeta_*\varphi) = -\zeta_*\varphi,$$

outside a set of Lebesgue measure zero. Since by assumption $\kappa \cdot \zeta_*\varphi \in L^1(\mathbb{R})$, this provides both the integrability of the integrands of (6.4) and the equality between the integrals involved. □

Proposition 6.38 (Projection formula)**.** *Let $\varphi \in \mathrm{CF}(\mathbb{V})$, $\theta \in \mathrm{CF}(\mathbb{R})$ and let $\zeta : \mathbb{V} \to \mathbb{R}$ be a morphism of real analytic manifolds. Assume that ζ is proper on $\mathrm{supp}(\varphi)$, that $\kappa \cdot \zeta_*\varphi \in L^1(\mathbb{R})$ and that $\kappa \cdot \theta \cdot \zeta_*\varphi \in L^1(\mathbb{R})$. Then,*

$$\mathrm{T}_\kappa \left[\varphi \cdot \zeta^*\theta\right](\zeta) = \mathrm{T}_{\kappa\cdot\theta} \left[\varphi\right](\zeta).$$

Proof. The result follows from the formula for constructible functions $\zeta_* \left(\varphi \cdot \zeta^*\theta\right) = \theta \cdot \zeta_*\varphi$, which follows from the corresponding property for constructible sheaves (Kashiwara and Schapira, 1990, Proposition 2.6.6) and the function-sheaf correspondence (ibid., Theorem 9.7.1). □

6.3.2 Convolution and box product

Recall the definition of the box product (Lemma 5.6) and of the convolution (Corollary 5.7) of constructible functions.

Proposition 6.39 (Convolution for $\mathcal{EL}$)**.** *Let φ and ψ be two constructible functions on $\mathbb{V}$ satisfying Assumptions 6.9 for a cone $C \neq \{0\}$. Then, we have on $\mathrm{Int}(C^\circ)$,*

$$\mathcal{EL}\left[\varphi \star \psi\right] = \mathcal{EL}\left[\varphi \star \mathbf{1}_C\right] \cdot \mathcal{EL}\left[\psi\right] + \mathcal{EL}\left[\varphi\right] \cdot \mathcal{EL}\left[\psi \star \mathbf{1}_C\right] - \mathcal{EL}\left[\varphi\right] \cdot \mathcal{EL}\left[\psi\right].$$

Proof. By Corollary 5.7, one has $\xi_* \left(\varphi \star \psi\right) = \left(\xi_*\varphi\right) \star \left(\xi_*\psi\right)$. Thus, using Lemma 6.15, we have:

$$\mathcal{EL}\left[\varphi \star \psi\right](\xi) = \mathcal{EL}\left[\left(\xi_*\varphi\right) \star \left(\xi_*\psi\right)\right](1).$$

The result follows then from the following claim, proven afterwards.

Claim 6.40. If $\theta, \theta' \in \mathrm{CF}(\mathbb{R})$ both satisfy Assumptions 6.9 for the cone $\mathbb{R}_{\geq 0}$, then

$$\begin{aligned}\mathcal{EL}\left[\theta \star \theta'\right](1) = {} & \mathcal{EL}\left[\theta \star \mathbf{1}_{\mathbb{R}_{\geq 0}}\right](1) \cdot \mathcal{EL}\left[\theta'\right](1) \\ & + \mathcal{EL}\left[\theta\right](1) \cdot \mathcal{EL}\left[\theta' \star \mathbf{1}_{\mathbb{R}_{\geq 0}}\right](1) \\ & - \mathcal{EL}\left[\theta\right](1) \cdot \mathcal{EL}\left[\theta'\right](1).\end{aligned}$$

Indeed, the functions $\xi_*\varphi$ and $\xi_*\psi$ both satisfy Assumptions 6.9 for the cone $\mathbb{R}_{\geq 0}$ by Lemma 6.13. Thus, Claim 6.40 yields:

$$\begin{aligned}\mathcal{EL}\left[(\xi_*\varphi)\star(\xi_*\psi)\right](1) = \mathcal{EL}\left[(\xi_*\varphi)\star\mathbf{1}_{\mathbb{R}_{\geq 0}}\right](1)\cdot\mathcal{EL}\left[\xi_*\psi\right](1)\\ + \mathcal{EL}\left[\xi_*\varphi\right](1)\cdot\mathcal{EL}\left[(\xi_*\psi)\star\mathbf{1}_{\mathbb{R}_{\geq 0}}\right](1)\\ - \mathcal{EL}\left[\xi_*\varphi\right](1)\cdot\mathcal{EL}\left[\xi_*\psi\right](1).\end{aligned}$$

Since $C \neq \{0\}$ the compatibility formula between pushforward and convolution (Corollary 5.7) and the expression of the pushforward of the indicator function of a convex subset (Lemma 5.10) ensure that:

$$\begin{aligned}(\xi_*\varphi)\star\mathbf{1}_{\mathbb{R}_{\geq 0}} &= \xi_*\left(\varphi\star\mathbf{1}_C\right),\\ (\xi_*\psi)\star\mathbf{1}_{\mathbb{R}_{\geq 0}} &= \xi_*\left(\psi\star\mathbf{1}_C\right).\end{aligned}$$

Hence the result, by Lemma 6.15.

Let us now prove Claim 6.40. Lemma 5.4 yields that $\theta\star\theta'$ satisfies Assumptions 6.9 for the cone $\mathbb{R}_{\geq 0}$, so that $\mathcal{EL}\left[(\theta\star\theta')\right]$ is well-defined on $\mathbb{R}_{>0} \ni 1$. Consider now decompositions of θ and θ':

$$\begin{aligned}\theta &= \sum_{i\in I} m_i\mathbf{1}_{[a_i,b_i]},\\ \theta' &= \sum_{j\in J} n_j\mathbf{1}_{[c_j,d_j]},\end{aligned}$$

where I and J are finite, m_i and n_j are integers, a_i, c_j are real numbers, and $b_i, d_j \in \mathbb{R}\cup\{+\infty\}$. We abusively denoted $[x,+\infty] := [x,+\infty)$ for $x\in\mathbb{R}$ for simplicity. One has:

$$\theta\star\theta' = \sum_{(i,j)\in I\times J} m_in_j\mathbf{1}_{[a_i,b_i]}\star\mathbf{1}_{[c_j,d_j]} = \sum_{(i,j)\in I\times J} m_in_j\mathbf{1}_{[a_i+c_j,b_i+d_j]}.$$

Therefore, we have:

$$\begin{aligned}\mathcal{EL}\left[\theta\star\theta'\right](1) &= \int_{\mathbb{R}} e^{-t}\left(\theta\star\theta'\right)(t)\,\mathrm{d}t\\ &= \sum_{(i,j)\in I\times J} m_in_j\int_{a_i+c_j}^{b_i+d_j} e^{-t}\,\mathrm{d}t\\ &= \sum_{(i,j)\in I\times J} m_in_j\left(e^{-a_i-c_j} - e^{-b_i-d_j}\right)\\ &= AC - BD\\ &= A(C-D) + (A-B)C - (A-B)(C-D),\end{aligned}$$

using the convention that $e^{-\infty} = 0$, and where we denoted:

$$A := \sum_{i\in I} m_ie^{-a_i}, \qquad B := \sum_{i\in I} m_ie^{-b_i},$$

$$C := \sum_{j\in J} n_je^{-c_j}, \qquad D := \sum_{j\in J} n_je^{-d_j}.$$

Moreover, for $x \in \mathbb{R}$ and $y \in \mathbb{R} \cup \{+\infty\}$, we have:

$$\mathbf{1}_{[x,y]} \star \mathbf{1}_{\mathbb{R}_{\geq 0}} = \mathbf{1}_{[x,+\infty)}, \tag{6.5}$$

so that:

$$\theta \star \mathbf{1}_{\mathbb{R}_{\geq 0}} = \sum_{i \in I} m_i \mathbf{1}_{[a_i,+\infty)},$$
$$\theta' \star \mathbf{1}_{\mathbb{R}_{\geq 0}} = \sum_{j \in I} n_j \mathbf{1}_{[c_j,+\infty)}.$$

Thus, Example 6.16 yields:

$$A = \mathcal{EL}\left[\theta \star \mathbf{1}_{\mathbb{R}_{\geq 0}}\right](1), \qquad C = \mathcal{EL}\left[\theta' \star \mathbf{1}_{\mathbb{R}_{\geq 0}}\right](1),$$
$$A - B = \mathcal{EL}\left[\theta\right](1), \qquad C - D = \mathcal{EL}\left[\theta'\right](1),$$

which proves Claim 6.40, and finishes the proof of Proposition 6.39. □

Remark 6.41. If φ and ψ are constructible functions satisfying Assumptions 6.9 for two different cones $C' \neq \{0\}$ and $C'' \neq \{0\}$ respectively, then they satisfy Assumptions 6.9 for the cone $C = \mathrm{Conv}\,(C' \cup C'')$. Moreover, one has $C \neq \{0\}$.

Corollary 6.42. *Let φ and ψ be two constructible functions on $\mathbb{V}$ satisfying Assumptions 6.9 for cones $C' \neq \{0\}$ and $C'' \neq \{0\}$ respectively. Assume in addition that $\varphi = \varphi \star \mathbf{1}_{C'}$ and $\psi = \psi \star \mathbf{1}_{C''}$. Then, we have on $\mathrm{Int}(C'^{\circ}) \cap \mathrm{Int}(C''^{\circ})$,*

$$\mathcal{EL}\left[\varphi \star \psi\right] = \mathcal{EL}\left[\varphi\right] \cdot \mathcal{EL}\left[\psi\right].$$

Proof. Following Remark 6.41, consider $C = \mathrm{Conv}\,(C' \cup C'')$. Then $\mathrm{Int}(C^{\circ}) = \mathrm{Int}(C'^{\circ}) \cap \mathrm{Int}(C''^{\circ})$ and the result follows from Proposition 6.39 and the fact that $\mathbf{1}_{\Gamma} \star \mathbf{1}_{\Gamma'} = \mathbf{1}_{\Gamma}$, for any two closed convex proper cones Γ and Γ' such that $\Gamma' \subseteq \Gamma$. □

Remark 6.43. If φ satisfies Assumptions 6.9 for a cone C', then it is constructible up to infinity and its support is γ-proper for the cone $\gamma = C'^{a}$. Hence, the assumption that $\varphi = \varphi \star \mathbf{1}_C$ is equivalent to that of φ being γ-constructible by Proposition 5.13.

Any $\eta \in (\mathbb{V} \times \mathbb{V}')^*$ can naturally be written $\eta = s \circ (\xi \times \xi')$ for $(\xi, \xi') \in \mathbb{V}^* \times \mathbb{V}'^*$ and $s : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ the addition, so that we have the following corollary:

Corollary 6.44 (Box product for $\mathcal{EL}$)**.** *Let $\varphi \in \mathrm{CF}(\mathbb{V})$ and $\psi \in \mathrm{CF}(\mathbb{V}')$ both satisfy Assumptions 6.9 for cones $C \subseteq \mathbb{V}$ and $C' \subseteq \mathbb{V}'$ respectively. Assume further that $C \neq \{0\}$ and $C' \neq \{0\}$. For any $(\xi, \xi') \in \mathrm{Int}(C^{\circ}) \times \mathrm{Int}(C'^{\circ})$, we have:*

$$\mathcal{EL}\left[\varphi \boxtimes \psi\right](\eta) = \mathcal{EL}\left[\varphi \star \mathbf{1}_C\right](\xi) \cdot \mathcal{EL}\left[\psi\right]\left(\xi'\right) + \mathcal{EL}\left[\varphi\right](\xi) \cdot \mathcal{EL}\left[\psi \star \mathbf{1}_{C'}\right]\left(\xi'\right) - \mathcal{EL}\left[\varphi\right](\xi) \cdot \mathcal{EL}\left[\psi\right]\left(\xi'\right),$$

with $\eta = s \circ (\xi \times \xi')$.

Proof. Since $\eta = s \circ (\xi \times \xi')$, Lemma 5.6 implies that $\eta_* (\varphi \boxtimes \psi) = (\xi_* \varphi) \star (\xi'_* \psi)$. Hence,

$$\mathcal{EL} [\varphi \boxtimes \psi] (\eta) = \mathcal{EL} [\eta_* (\varphi \boxtimes \psi)] (1) = \mathcal{EL} \left[(\xi_* \varphi) \star (\xi'_* \psi) \right] (1),$$

and the result follows from Lemma 6.15 and the compatibility with convolution (Proposition 6.39). □

Corollary 6.45. *In the setting of Corollary 6.42, we have for any* $(\xi, \xi') \in \mathrm{Int}(C^\circ) \times \mathrm{Int}(C'^\circ)$,

$$\mathcal{EL} [\varphi \boxtimes \psi] (\eta) = \mathcal{EL} [\varphi] (\xi) \cdot \mathcal{EL} [\psi] \left(\xi' \right),$$

with $\eta = s \circ (\xi \times \xi')$.

Example 6.46 (Interpretation of the Laplace transform on γ-voxels)**.** We call *γ-voxel* a subset of $\mathbb{R}^d$ of the form $[a_1, b_1) \times \cdots \times [a_d, b_d)$ where $a_i < b_i$ are real numbers. Consider $\varphi = \sum_{i \in I} m_i \mathbf{1}_{V_i} \in \mathrm{CF}(\mathbb{R}^d)$ where the set I is finite, the coefficients m_i are integers and the subsets V_i are γ-voxels. Then, for any $\xi = (\xi_1, \ldots, \xi_d) \in (\mathbb{R}_{\geq 0})^d$,

$$\mathcal{EL} [\varphi] (\xi) = \mathcal{L}[\varphi](\xi) \cdot \prod_{k=1}^{d} \xi_k. \tag{6.6}$$

Indeed, the equality is true for $d = 1$ and extends naturally to γ-voxels thanks to the compatibility formula for the box product (Corollary 6.44). This relation gives a new interpretation of the Laplace transform on such constructible functions. One could wonder whether the relation (6.6) can be generalized for all γ-constructible functions. However, Example 6.20 shows that such a generalization is not obvious.

Let γ be a cone of $\mathbb{V}$ satisfying (C1).

Proposition 6.47 (Convolution for $\mathcal{EF}$)**.** *Let* $\varphi, \psi \in \mathrm{CF}_{\gamma,\mathrm{c}}(\mathbb{V})$. *For* $\xi \in \mathbb{V}^*$,

$$\mathcal{EF} [\varphi \star \psi] (\xi) = \begin{cases} i \cdot \mathcal{EF} [\varphi] (\xi) \cdot \mathcal{EF} [\psi] (\xi) & \text{if } \xi \in \gamma^{a\circ}, \\ -i \cdot \mathcal{EF} [\varphi] (\xi) \cdot \mathcal{EF} [\psi] (\xi) & \text{if } \xi \in \gamma^\circ. \end{cases}$$

Remark 6.48. For any $\xi \in \mathbb{V}^* \setminus (\gamma^\circ \cup \gamma^{a\circ})$, the compatibility formula still holds. Indeed, since $\varphi \star \psi$ is γ-constructible by Proposition 5.13, both sides of the equality are zero by Proposition 5.19 and (5.3).

Proof. Suppose that $\xi \in \gamma^{a\circ}$, the other case being similar. We have:

$$\mathcal{EF} [\varphi \star \psi] (\xi) = \mathcal{EF} [\xi_* (\varphi \star \psi)] (1),$$

and $\xi_* (\varphi \star \psi) = (\xi_* \varphi) \star (\xi_* \psi)$ by Corollary 5.7. Since Lemma 5.14 ensures that $\theta = \xi_* \varphi$ and $\theta' = \xi_* \psi$ are both in $\mathrm{CF}_{\gamma,\mathrm{c}[\lambda]}(\mathbb{V})$ with $\lambda = \mathbb{R}_{\leq 0}$, it is sufficient to prove:

$$\mathcal{EF} \left[\theta \star \theta' \right] (1) = i \cdot \mathcal{EF} [\theta] (1) \cdot \mathcal{EF} \left[\theta' \right] (1),$$

for any $\theta, \theta' \in \mathrm{CF}_{\gamma,\mathrm{c}[\lambda]}(\mathbb{V})$. By bilinearity of the convolution, it is even sufficient to prove the result for $\theta = \mathbf{1}_{[a,b)}$ and $\theta' = \mathbf{1}_{[c,d)}$ where $a < b$ and $c < d$ are real numbers. Suppose now that $a + d \leq b + c$, the case $a + d \geq b + c$ being proven in a similar fashion. Then,

$$\mathbf{1}_{[a,b)} \star \mathbf{1}_{[c,d)} = \mathbf{1}_{[a+c,a+d)} - \mathbf{1}_{[b+c,b+d)}.$$

Therefore, we get:

$$\begin{aligned}\mathcal{EF}\left[\mathbf{1}_{[a,b)} \star \mathbf{1}_{[c,d)}\right](1) &= i\left(e^{-i(a+d)} - e^{-i(a+c)}\right) - i\left(e^{-i(b+d)} - e^{-i(b+c)}\right)\\ &= -i\left(e^{-ib} - e^{-ia}\right)\left(e^{-id} - e^{-ic}\right)\\ &= i\,\mathcal{EF}\left[\mathbf{1}_{[a,b)}\right](1)\,\mathcal{EF}\left[\mathbf{1}_{[c,d)}\right](1).\end{aligned}$$

□

As for the Euler-Laplace transform, one gets the following corollary for the box product. Let us consider a cone γ' of $\mathbb{V}'$ satisfying (C1).

Corollary 6.49 (Box product for $\mathcal{EF}$)**.** *Let $\varphi \in \mathrm{CF}_{\gamma,\mathrm{c}}(\mathbb{V})$ and $\psi \in \mathrm{CF}_{\gamma,\mathrm{c}[\gamma']}(\mathbb{V}')$. For any $(\xi,\xi') \in \mathbb{V}^* \times \mathbb{V}'^*$, we have:*

$$\mathcal{EF}\left[\varphi \boxtimes \psi\right](\eta) = \begin{cases} i \cdot \mathcal{EF}\left[\varphi\right](\xi) \cdot \mathcal{EF}\left[\psi\right](\xi') & \text{if } (\xi,\xi') \in \gamma^{a\circ} \times \gamma'^{a\circ},\\ -i \cdot \mathcal{EF}\left[\varphi\right](\xi) \cdot \mathcal{EF}\left[\psi\right](\xi') & \text{if } (\xi,\xi') \in \gamma^{\circ} \times \gamma'^{\circ},\end{cases}$$

with $\eta = s \circ (\xi \times \xi')$.

Remark 6.50. Note that not all possibilities of (ξ,ξ') are treated in the previous corollary, as such an equality is not true in general.

Example 6.51 (Interpretation of the Fourier transform on γ-voxels)**.** In the setting of Example 6.46, we have for any $\xi = (\xi_1,\ldots,\xi_d) \in (\mathbb{R}_{\geq 0})^d$,

$$\mathcal{EF}\left[\varphi\right](\xi) = i^{d-1} \cdot \mathcal{F}[\varphi](\xi) \cdot \prod_{k=1}^{d} \xi_k. \tag{6.7}$$

Again, this relation gives a new interpretation of the Fourier transform on such constructible functions. Example 6.29 shows that a generalization of such a relation for all γ-constructible functions is not obvious.

Remark 6.52 (On stability)**.** For each integer $k \geq 1$, consider:

$$\varphi_k = \sum_{(i,j)} \mathbf{1}_{[\frac{i}{k},\frac{i+1}{k})\times[\frac{j}{k},\frac{j+1}{k})},$$

where the sum is over all pairs $(i,j) \in [\![1; k-1]\!]^2$ such that $i + j \leq k-1$. Then, the sequence $(\varphi_k)_{k\geq 1}$ converges to the γ-triangle $\mathbf{1}_T$ of Example 6.20 in $L^p(\mathbb{R}^2)$ for $p \in [1,+\infty]$. Thus, over the domain $(\mathbb{R}_{\geq 0})^2$, the sequence $\left(\mathcal{EF}\left[\varphi_k\right]\right)_{k\geq 1} = \left(\mathcal{F}[\varphi_k]\right)_{k\geq 1}$ converges to $\mathcal{F}[\mathbf{1}_T]$ in L^∞. However, this last function differs from $\mathcal{EF}\left[\mathbf{1}_T\right]$, as shown in Example 6.29. Hence, if a stability statement holds for the Euler-Fourier transform, it should be for other norms on constructible functions and on the Euler-Fourier transforms. The stability of hybrid transforms goes beyond the scope of this thesis and will be the object of future work.

6.4 Reconstruction formula for Euler-Fourier

Consider a cone γ of $\mathbb{V}$ satisfying (C1). In this section, we establish a reconstruction formula for the Euler-Fourier transform of γ-constructible functions (Theorem 6.55).

If $h : \mathbb{V}^* \to \mathbb{R}$ is such that for any $\xi \in \mathbb{V}^* \setminus \{0\}$, the map:

$$\widetilde{h}_\xi : s \longmapsto \frac{\mathbf{1}_{\mathbb{R}\setminus\{0\}}(s)}{|s|} \cdot h(s\,\xi),$$

satisfies that the following limit exists:

$$\mathcal{F}^{-1}\left[\widetilde{h}_\xi\right](t) := \lim_{A\to+\infty} \int_{-A}^{A} e^{ist} \cdot \widetilde{h}_\xi(s)\,\mathrm{d}s,$$

then we can define the following map for all $\xi \in \mathbb{V}^*$ and $t \in \mathbb{R}$,

$$\mathcal{F}'(h)(\xi,t) := \begin{cases} \dfrac{1}{2\pi}\mathcal{F}^{-1}\left[\widetilde{h}_\xi\right](t^+) & \text{if } \xi \in \gamma^{a\circ} \setminus \{0\}, \\ \dfrac{1}{2\pi}\mathcal{F}^{-1}\left[\widetilde{h}_\xi\right](t^-) & \text{if } \xi \in \gamma^{\circ} \setminus \{0\}, \\ 0 & \text{else,} \end{cases} \tag{6.8}$$

where $g(t^\pm) = \lim_{s\to t^\pm} g(s)$ for any function g defined in a neighborhood of t.

Proposition 6.53. *Let $\varphi \in \mathrm{CF}_{\gamma,\mathrm{c}}(\mathbb{V})$. Then, $\mathcal{F}'\left(\mathcal{EF}\left[\varphi\right]\right)$ is well-defined, and for all $\xi \in \mathbb{V}^*$ and $t \in \mathbb{R}$,*

$$\mathcal{F}'\left(\mathcal{EF}\left[\varphi\right]\right)(\xi,t) = \xi_*\varphi(t).$$

For $\xi \in \gamma^\circ \cup \gamma^{a\circ} \setminus \{0\}$, the proof boils down to inverting the classical Fourier transform. Proposition 5.19 ensures that the pushforward $\xi_*\varphi$ is zero for other ξ's, hence so are both sides of the equality. First, recall the following easy lemma from classical Fourier theory:

Lemma 6.54. *If $f : \mathbb{R} \to \mathbb{R}$ is a integrable, piecewise smooth and right-continuous (resp. left-continuous) function, then for all $t \in \mathbb{R}$,*

$$f(t) = \frac{1}{2\pi}\mathcal{F}^{-1}\mathcal{F}(f)(t^+) \qquad \Big(\textit{resp. } \frac{1}{2\pi}\mathcal{F}^{-1}\mathcal{F}(f)(t^-)\Big).$$

Proof. For $f : \mathbb{R} \to \mathbb{R}$ integrable and piecewise smooth, the following inversion formula (Vretblad, 2003, Theorem 7.5) holds for the Fourier transform:

$$\frac{1}{2\pi}\mathcal{F}^{-1}\mathcal{F}(f)(t) = \frac{1}{2}\left(f(t^-) + f(t^+)\right), \tag{6.9}$$

for all $t \in \mathbb{R}$. In particular, for every point t at which f is continuous, the right-hand side is equal to $f(t)$, hence the result. □

We can now prove the above proposition.

Proof of Proposition 6.53. Lemma 6.27 yields:

$$\widetilde{(\mathcal{EF}[\varphi])}_\xi = \mathbf{1}_{\mathbb{R}\setminus\{0\}} \cdot \mathcal{F}(\xi_*\varphi), \tag{6.10}$$

and the map on the right-hand side is almost everywhere equal to $\mathcal{F}(\xi_*\varphi)$, the Fourier transform of an integrable and piecewise smooth function over $\mathbb{R}$. Hence, Vretblad (2003, Theorem 7.5) ensures that the limit

$$\lim_{A\to+\infty} \int_{-A}^{A} e^{ist} \cdot \widetilde{(\mathcal{EF}[\varphi])}_\xi(s)\,\mathrm{d}s$$

exists, so that $\mathcal{F}'(\mathcal{EF}[\varphi])$ is well-defined.

Let now $\xi \in \mathbb{V}^*$ and $t \in \mathbb{R}$. If both are zero, the result is clear. Otherwise, denote $y = [\xi : t] \in \mathbb{P}^*(\mathbb{V})$. If $y \notin K_\gamma^*$, the characterization of the support of the Radon transform on compactly supported γ-constructible functions (Proposition 5.19) yields the result by (5.3) and the definition of $\mathcal{F}'$. If $y \in K_\gamma^*$, then the constructible function $\xi_*\varphi$ is right-continuous when $\xi \in \gamma^{a\circ} \setminus \{0\}$ (resp. left-continuous when $\xi \in \gamma^\circ \setminus \{0\}$) since Lemma 5.14 ensures that it is compactly supported and λ-constructible on $\mathbb{R}$ with $\lambda = \mathbb{R}_{\leq 0}$ (resp. $\lambda = \mathbb{R}_{\geq 0}$). Therefore, Lemma 6.54 ensures that

$$\xi_*\varphi(t) = \begin{cases} \dfrac{1}{2\pi}\mathcal{F}^{-1}\mathcal{F}(\xi_*\varphi)(t^+) & \text{if } \xi \in \gamma^{a\circ} \setminus \{0\} \\ \dfrac{1}{2\pi}\mathcal{F}^{-1}\mathcal{F}(\xi_*\varphi)(t^-) & \text{if } \xi \in \gamma^\circ \setminus \{0\}. \end{cases} \tag{6.11}$$

Hence the result, since

$$\mathcal{F}^{-1}\mathcal{F}(\xi_*\varphi) = \mathcal{F}^{-1}\left(\mathbf{1}_{\mathbb{R}\setminus\{0\}} \cdot \mathcal{F}(\xi_*\varphi)\right) \overset{(6.10)}{=} \mathcal{F}^{-1}\left(\widetilde{(\mathcal{EF}[\varphi])}_\xi\right).$$

□

For any $\varphi \in \mathrm{CF}_{\gamma,\mathrm{c}}(\mathbb{V})$, combining Proposition 6.53 with (5.3) implies that the map $\mathcal{F}'(\mathcal{EF}[\varphi])$ induces a map on $\mathbb{P}^*(\mathbb{V})$ and:

$$\mathcal{F}'(\mathcal{EF}[\varphi]) = \mathcal{R}(j_*\varphi). \tag{6.12}$$

Putting all together, we get the following reconstruction result for the Euler-Fourier transform:

Theorem 6.55. *Let $\varphi \in \mathrm{CF}_{\gamma,\mathrm{c}}(\mathbb{V})$. Then,*

$$\mathcal{R}' \circ \mathcal{F}'(\mathcal{EF}[\varphi]) = (-1)^{\dim(\mathbb{V})+1}\varphi.$$

Remark 6.56. We abuse notations to alleviate the formula by identifying $j_*\varphi$ and φ in the right-hand side.

Proof. The results follows from (6.12) and the inversion formula for the Radon transform of compactly supported γ-constructible functions (Theorem 5.17 and Remark 5.18). □

Chapter 7

Index-theoretic formulae

Baryshnikov and Ghrist (2010) extended Euler calculus of constructible functions to the space of so-called *definable* functions, that is, whose graph is definable in some o-minimal structure. Moreover, they proved that Euler integrals of Morse functions can be expressed as signed sums of their critical values. Then, Bobrowski and Borman (2012) defined a similar extension to the so-called *tame* functions, which can be expressed in terms of signed sums of critical values. Both definition coincide on tame functions which are also definable. The class of tame functions contains for instance continuous subanalytic functions on compact real analytic manifolds.

Since their first introduction, a connection has been established between hybrid transforms and continuous Euler integrals. These formulae are first proven in Ghrist and Robinson (2011) and called *index-theoretic formulae* to emphasize their link with signed sums of critical values. In their paper, Bobrowski and Borman (2012) proved that the Euler characteristic of barcodes is naturally expressed as a continuous Euler integral of the filtration function in the context of sublevel-sets persistence. Finally, Govc and Hepworth (2021) proved that persistent magnitude can be expressed as a signed sum of exponential of critical values in the context of sublevel-sets persistence of tame functions.

Building on the work of Adler and Taylor (2009), Bobrowski and Borman (2012) computed the expected Euler integral of Gaussian random fields in terms of integral geometric quantities depending on the random fields. Then, they combined this result to index-theoretic formulae to compute expected values of Euler characteristics of barcodes associated to sublevel-sets persistence of random filtration functions. At the time, this provided one of the first general tools to derive probabilistic statements about sublevel-sets persistence.

Outline. In this chapter, we define the so-called *sublevel-sets* and *level-sets constructible functions* associated to a continuous subanalytic map (Section 7.1). They are the constructible functions associated to persistent cohomology sheaves introduced by Kashiwara and Schapira (2018). We prove an expression of sublevel-sets constructible functions as a convolution of level-sets constructible functions with the indicator function of the antipodal of γ. As a consequence, hybrid transforms of (sub)level-sets constructible functions associated to vector-valued maps can be expressed as hybrid transforms of (sub)level-sets constructible functions associated to real-valued maps. It leads to the definition

of *sublevel-sets transforms*, the simple form that hybrid transforms take in the case of multi-persistence.

Then, we recall the definition of continuous Euler integral from Bobrowski and Borman (2012) and prove index-theoretic formulae for a wide class of hybrid transforms, generalizing existing ones for the magnitude, the Euler characteristic of barcodes and the Bessel and Fourier transforms (Section 7.2).

Finally, we follow the approach of Bobrowski and Borman (ibid.) and use our index-theoretic formulae together with the results of loc. cit. to compute the expected value of the Euler-Bessel transform of level-sets constructible functions associated to Gaussian random fields (Section 7.3).

Convention. We follow the conventions and notations introduced in Chapter 5. Recall that M is a compact real analytic manifold. Throughout the chapter, we consider a locally closed subanalytic subset Z of M, a continuous subanalytic map $f : M \to \mathbb{V}$ and a cone γ of $\mathbb{V}$ satisfying (C2), that is:

$$\gamma \text{ is a non-empty subanalytic closed proper convex cone.} \tag{C2}$$

Note that γ may have empty interior.

7.1 Sublevel-sets persistence

In this section, we recall the sheaf-theoretic formulation of the sublevel-sets persistent homology due to Kashiwara and Schapira (2018, Section 1.2), focusing only on constructible functions instead of constructible sheaves.

Define the γ-*epigraph* of f by:

$$\Gamma_f^\gamma = \{(x, v) \in M \times \mathbb{V};\ f(x) - v \in \gamma\}.$$

Denote also by Γ_f the usual graph of f. The set Γ_f^γ is closed and subanalytic in $M \times \mathbb{V}$ by Kashiwara and Schapira (1990, Proposition 8.2.2) since, denoting $\sigma : (x, v, w) \in M \times \mathbb{V} \times \mathbb{V} \mapsto (x, v + w) \in M \times \mathbb{V}$, we have $\Gamma_f^\gamma = \sigma(\Gamma_f \times \gamma^a)$. Denote by $p : M \times \mathbb{V} \to \mathbb{V}$ and $q : M \times \mathbb{V} \to M$ the canonical projections, so that we get the following diagram:

$$\begin{array}{ccc} & \Gamma_f^\gamma & \\ & \cap & \\ & M \times \mathbb{V} & \\ {\scriptstyle q}\swarrow & & \searrow{\scriptstyle p} \\ M & & \mathbb{V} \end{array}$$

Define the *sublevel-sets constructible function* associated to f on Z as:

$$\varphi^\gamma_{f_{|Z}} = p_* \left(\mathbf{1}_{\Gamma_f^\gamma} \cdot q^* \mathbf{1}_Z \right) = p_* \mathbf{1}_{\Gamma_f^\gamma \cap (Z \times \mathbb{V})}, \tag{7.1}$$

and the *level-sets constructible function* associated to f on Z as:

$$\varphi_{f_{|Z}} = p_* \left(\mathbf{1}_{\Gamma_f} \cdot q^* \mathbf{1}_Z \right) = p_* \mathbf{1}_{\Gamma_f \cap (Z \times \mathbb{V})}, \tag{7.2}$$

The previous constructible functions are well-defined since p is proper on Γ_f^γ (Kashiwara and Schapira, 2018, Theorem 1.11) and since the properness still holds on the closure of $\Gamma_f^\gamma \cap (Z \times \mathbb{V})$ and also for $\gamma = \{0\}$. Note also that Lemma 5.4 ensures that the level-sets constructible function is compactly supported.

Remark 7.1. For $v \in \mathbb{V}$, one has $\varphi^\gamma_{f_{|Z}}(v) = \chi_c\left(f^{-1}(v+\gamma) \cap Z\right)$.

Notation 7.2. When $\mathbb{V} = \mathbb{R}$ and $\gamma = \mathbb{R}_{\leq 0}$ (resp. $\gamma = \mathbb{R}_{\geq 0}$), we denote $\varphi^-_{g_{|Z}} = \varphi^\gamma_{g_{|Z}}$ (resp. $\varphi^+_{g_{|Z}} = \varphi^\gamma_{g_{|Z}}$) the sublevel-sets constructible function associated to a continuous subanalytic map $g : M \to \mathbb{R}$ on Z.

Example 7.3. The function φ_f^γ is sometimes called *Euler characteristic curve* for $\mathbb{V} = \mathbb{R}$, *Euler characteristic surface* for $\mathbb{V} = \mathbb{R}^2$ and *Euler characteristic profile* for $\mathbb{V} = \mathbb{R}^n$ with $n \geq 3$; see for instance Beltramo, Andreeva, et al. (2021) and Dłotko and Gurnari (2022).

Example 7.4. Considering a subset $Z \subseteq \mathbb{R}^d$ relatively compact subanalytic and locally closed, the *Euler characteristic transform* defined in Turner, Mukherjee, and Boyer (2014b) is, for $\xi \in \mathbb{V}^*$ and $t \in \mathbb{R}$,

$$\mathrm{ECT}(Z)(\xi, t) = \chi\big(\{x \in Z\,;\, \langle \xi, x\rangle \leq t\}\big) = \varphi^-_\xi(t).$$

If $f : \mathbb{R}^d \to \mathbb{R}$ is continuous subanalytic and $\xi \in \mathbb{R}^d$, denote by $(\xi, f) : \mathbb{R}^d \to \mathbb{R}^2$ the map defined by $x \mapsto (\langle \xi, x\rangle, f(x))$. The *Lifted Euler Characteristic Transform along f* recently defined in Kirveslahti and Mukherjee (2021) is then, for any $(h, t) \in \mathbb{R}^2$,

$$\mathrm{LECT}(f)(\xi, h, t) = \varphi^\gamma_{(\xi,f)}(h, t),$$

where $\gamma = \mathbb{R}_{\leq 0} \times \{0\} \subset \mathbb{R}^2$. The *Super Lifted Euler Characteristic Transform along f* defined in loc. cit. is, for any $(h, t) \in \mathbb{R}^2$,

$$\mathrm{SELECT}(f)(\xi, h, t) = \varphi^{\gamma'}_{(\xi,f)}(h, t),$$

where $\gamma' = \mathbb{R}_{\leq 0} \times \mathbb{R}_{\geq 0} \subset \mathbb{R}^2$.

The following proposition specifies the relationship between the level-sets and the sublevel-sets constructible functions. It is key to the study of hybrid transforms in the context of sublevel-sets persistence, and more specifically to the proof of Proposition 7.8, which is itself key to index-theoretic formulae (Theorem 7.17). It can be derived from an analogous result on sheaves by Berkouk and Petit (2022, Proposition 4.17) via the function-sheaf correspondence. Here, we give a proof that does not make use of the correspondence.

Proposition 7.5. *In the preceding situation, $\varphi^\gamma_{f_{|Z}} = \varphi_{f_{|Z}} \star \mathbf{1}_{\gamma^a}$.*

Proof. Since $\Gamma_f^\gamma \cap (Z \times \mathbb{V}) = \sigma\left(\Gamma_f \cap (Z \times \mathbb{V}) \times \gamma^a\right)$, Remark 5.5 yields that:

$$\mathbf{1}_{\Gamma_f^\gamma \cap (Z\times\mathbb{V})} = \sigma_* \mathbf{1}_{\Gamma_f \cap (Z\times\mathbb{V})\times\gamma^a} = \sigma_*\left(\mathbf{1}_{\Gamma_f \cap (Z\times\mathbb{V})} \boxtimes \mathbf{1}_{\gamma^a}\right). \tag{7.3}$$

Moreover, by Lemma 5.6,

$$(p \times \mathrm{Id}_{\mathbb{V}})_* \left(\mathbf{1}_{\Gamma_f \cap (Z \times \mathbb{V})} \boxtimes \mathbf{1}_{\gamma^a} \right) = \left(p_* \mathbf{1}_{\Gamma_f \cap (Z \times \mathbb{V})} \right) \boxtimes \mathbf{1}_{\gamma^a}. \tag{7.4}$$

Thus, denoting by $s : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$ the addition, we have $p \circ \sigma = s \circ (p \times \mathrm{Id}_{\mathbb{V}})$, and:

$$\begin{aligned} \varphi^{\gamma}_{f_{|Z}} &\overset{(7.3)}{=} p_* \sigma_* \left(\mathbf{1}_{\Gamma_f \cap (Z \times \mathbb{V})} \boxtimes \mathbf{1}_{\gamma^a} \right) \\ &\overset{(7.4)}{=} s_* \left(p_* \mathbf{1}_{\Gamma_f \cap (Z \times \mathbb{V})} \boxtimes \mathbf{1}_{\gamma^a} \right) \\ &= \varphi_{f_{|Z}} \star \mathbf{1}_{\gamma^a}. \end{aligned}$$

□

Remark 7.6. The above proposition ensures that $\varphi^{\gamma}_{f_{|Z}}$ satisfies Assumptions 6.9 for the cone γ^a by Remark 6.10.

We show that the pushforward by a linear form sends multi-parameter (sub)level-sets constructible functions to one-parameter ones. The first is a well-known lemma proven for completeness and used in the next result.

Lemma 7.7. *For any morphism of real analytic manifolds $\zeta : \mathbb{V} \to \mathbb{R}$, we have that $\zeta_* \varphi_{f_{|Z}} = \varphi_{\zeta \circ f_{|Z}}$.*

Proof. Consider the following commutative diagram

$$\begin{array}{ccc} (Z \times \mathbb{R}) \cap \Gamma_f & \xrightarrow{p} & \mathbb{V} \\ \downarrow{\scriptstyle (\mathrm{Id}_M \times \zeta)} & & \downarrow{\scriptstyle \zeta}, \\ (Z \times \mathbb{R}) \cap \Gamma_{\zeta \circ f} & \xrightarrow{p'} & \mathbb{R} \end{array} \tag{7.5}$$

where $p' : M \times \mathbb{R} \to \mathbb{R}$ is the canonical projection. We have:

$$\begin{aligned} \zeta_* \varphi_{f_{|Z}} &= \zeta_* p_* \mathbf{1}_{\Gamma_f \cap (Z \times \mathbb{V})} \\ &= p'_* (\mathrm{Id}_M \times \zeta)_* \mathbf{1}_{\Gamma_f \cap (Z \times \mathbb{V})} \\ &= p'_* \mathbf{1}_{\Gamma_{\zeta \circ f} \cap (Z \times \mathbb{R})} \\ &= \varphi_{\zeta \circ f_{|Z}}, \end{aligned}$$

where the third equality follows from Remark 5.5 and the equality:

$$(\mathrm{Id}_M \times \zeta) \left(\Gamma_f \cap (Z \times \mathbb{V}) \right) = \Gamma_{\zeta \circ f} \cap (Z \times \mathbb{R}).$$

□

Proposition 7.8.

(i) *For all $\xi \in \mathrm{Int}(\gamma^{a\circ})$, $\xi_* \varphi^{\gamma}_{f_{|Z}} = \varphi^{-}_{\xi \circ f_{|Z}}$.*

(ii) *For all $\xi \in \mathrm{Int}(\gamma^{\circ})$, $\xi_* \varphi^{\gamma}_{f_{|Z}} = \varphi^{+}_{\xi \circ f_{|Z}}$.*

Proof. For any $\xi \in \mathrm{Int}(\gamma^{a\circ})$, we have

$$\begin{aligned}\xi_*\varphi_f^\gamma &\overset{\text{Cor. 5.7}}{=} \xi_*\varphi_{f_{|Z}} \star \xi_*\mathbf{1}_{\gamma^a}\\ &\overset{\text{Lem. 5.10}}{=} \xi_*\varphi_{f_{|Z}} \star \mathbf{1}_{\mathbb{R}_{\geq 0}}\\ &\overset{\text{Lem. 7.7}}{=} \varphi_{\xi\circ f_{|Z}} \star \mathbf{1}_{\mathbb{R}_{\geq 0}}\\ &\overset{\text{Prop. 7.5}}{=} \varphi^-_{\xi\circ f_{|Z}},\end{aligned}$$

and similarly replacing $\mathbb{R}_{\geq 0}$ by $\mathbb{R}_{\leq 0}$ if $\xi \in \mathrm{Int}(\gamma^\circ)$. □

Example 7.9. As in Example 7.4, consider a subset $Z \subseteq \mathbb{R}^d$ relatively compact subanalytic and locally closed. Then, the Euler characteristic transform is the pushforward, for $\xi \in \mathbb{V}^*$ and $t \in \mathbb{R}$,

$$\mathrm{ECT}(Z)(\xi, t) = \xi_*\varphi_f^\gamma(t),$$

where $f = \mathrm{Id}_{\mathbb{R}^d}$ and where γ is any cone satisfying (C2) such that $\mathrm{Int}(\gamma^{a\circ}) \ni \xi$.

Hybrid transforms. The formula expressing the pushforward of sublevel-sets constructible functions (Proposition 7.8) has the following consequence for hybrid transforms:

Corollary 7.10. *Let $\kappa \in L^1_{\mathrm{loc}}(\mathbb{R}) \cap L^1(\mathbb{R}_{\geq 0})$. The transform $\mathrm{T}_\kappa\left[\varphi^\gamma_{f_{|Z}}\right]$ is well-defined on $\mathrm{Int}(\gamma^{a\circ})$, and for any $\xi \in \mathrm{Int}(\gamma^{a\circ})$,*

$$\mathrm{T}_\kappa\left[\varphi^\gamma_{f_{|Z}}\right](\xi) = \mathrm{T}_\kappa\left[\varphi^-_{\xi\circ f_{|Z}}\right](1) = \int_{\mathbb{R}} \kappa(t)\varphi^-_{\xi\circ f_{|Z}}(t)\,\mathrm{d}t.$$

A similar result holds with $\varphi^+_{\xi\circ f}$ for $\kappa \in L^1_{\mathrm{loc}}(\mathbb{R}) \cap L^1(\mathbb{R}_{\leq 0})$ and $\xi \in \mathrm{Int}(\gamma^\circ)$.

Proof. The well-definedness follows from Proposition 7.5 and Proposition 6.12, while the formula follows from Proposition 7.8. □

Remark 7.11. Corollary 7.10 implies that for any $\xi \in \mathrm{Int}(\gamma^{a\circ})$, the evaluation $\mathrm{T}_\kappa\left[\varphi^\gamma_f\right](\xi)$ is nothing but the evaluation of the integral (with respect to the Lebesgue measure) transform with kernel κ of $\varphi^-_{\xi\circ f}$. In particular, the cone γ such that $\mathrm{Int}(\gamma^{a\circ}) \ni \xi$ does not play any role. Hence, the study of hybrid transforms of sublevel-sets constructible functions for vector-valued filtrations can be reduced to the ones for real-valued filtrations.

This study motivates the following definition, which gets rid of superfluous information when dealing with hybrid transforms of sublevel-sets constructible functions. Although *a priori* invisible, the hybrid nature underpins the definition.

Definition 7.12 (Sublevel-sets transform)**.** Let $\kappa \in L^1_{\mathrm{loc}}(\mathbb{R}) \cap L^1(\mathbb{R}_{\geq 0})$. We call *sublevel-sets transform of f over Z*, and denote by $\mathrm{Sub}_\kappa\left[f_{|Z}\right]$, the transform defined for $\xi \in \mathbb{V}^*$ by:

$$\mathrm{Sub}_\kappa\left[f_{|Z}\right](\xi) = \mathrm{T}_\kappa\left[\varphi^-_{\xi\circ f_{|Z}}\right](1) = \int_{\mathbb{R}} \kappa(t)\,\chi\left(\{\xi \circ f \leq t\} \cap Z\right)\,\mathrm{d}t.$$

Example 7.13 (Generalization of Morse magnitude)**.** We can define the *sublevel-sets magnitude of* (Z, f) as the sublevel-sets transform of f over Z with kernel $\kappa : t \mapsto e^{-t}$. In other words, for $\xi \in \mathbb{V}^*$,

$$|\xi \cdot (Z, f)| := \mathrm{Sub}_\kappa \left[f_{|Z}\right](\xi) = \int_{\mathbb{R}} e^{-t}\, \chi\left(\{\xi \circ f \leq t\} \cap Z\right) \mathrm{d}t. \tag{7.6}$$

When $Z = M$ and $f : M \to \mathbb{R}$ is a Morse function, the previous definition specializes in the notion of Morse magnitude of Govc and Hepworth (2021, Section 6). More generally, when $\xi \circ f$ is a Morse function, we have:

$$|\xi \cdot (M, f)| = |(M, \xi \circ f)|_{\mathrm{Morse}},$$

with the notations of loc. cit..

Example 7.14 (Generalization of Euler characteristic of barcodes)**.** Bobrowski and Borman (2012) introduced the *Euler characteristic of barcodes* for sublevel-sets persistent homology associated to a continuous subanalytic map $g : M \to \mathbb{R}$ restricted to the range $(-\infty, a)$ for $a \in \mathbb{R}$ as follows:

$$\chi_g^a := \int_{-\infty}^{a} \varphi_g^-(t)\, \mathrm{d}t = \mathrm{Sub}_{\mathbf{1}_{(-\infty,a)}}[g](1). \tag{7.7}$$

Definition 7.12 allows to generalize this notion to multi-parameter sublevel-sets persistent cohomology as follows. For a continuous subanalytic map $f : M \to \mathbb{V}$ eventually restricted to a subanalytic locally closed subset Z of M, we define for $\xi \in \mathbb{V}^*$,

$$\chi^a_{f_{|Z}}(\xi) = \mathrm{Sub}_{\mathbf{1}_{(-\infty,a)}}\left[f_{|Z}\right](\xi). \tag{7.8}$$

7.2 Hybrid transforms as continuous Euler integrals

In this section, we recall the definition of Bobrowski and Borman (ibid.) of continuous Euler integration, restricting ourselves to continuous subanalytic functions on compact real analytic manifolds. Although slightly less general, this allows us to use the framework developed in the previous chapters. Then, we prove the formulae expressing (sub)level-sets transforms as continuous Euler integral transforms.

Definition 7.15. Recall that Z is a locally closed subanalytic subset of M and let $g : M \to \mathbb{R}$ be continuous and subanalytic. The *continuous Euler upper integral of g on Z* is defined by:

$$\int_Z g \lceil \mathrm{d}\chi \rceil = \int_0^{+\infty} \chi_c\left(\{g > u\} \cap Z\right) - \chi_c\left(\{g \leq -u\} \cap Z\right) \mathrm{d}u,$$

and the *continuous Euler lower integral of g on Z* by:

$$\int_Z g \lfloor \mathrm{d}\chi \rfloor = \int_0^{+\infty} \chi_c\left(\{g \geq u\} \cap Z\right) - \chi_c\left(\{g < -u\} \cap Z\right) \mathrm{d}u.$$

Remark 7.16. If Z' is a locally closed relatively compact subanalytic subset of M and Z a closed subset of Z' that is subanalytic in M, then the classical distinguished triangle (Kashiwara and Schapira, 1990, Eq. 2.6.33) yields:

$$\chi_c\left(Z'\right) = \chi_c\left(Z' \setminus Z\right) + \chi_c\left(Z\right).$$

Therefore, in such a situation, we have:

$$\int_{Z'} g \lceil \mathrm{d}\chi \rceil = \int_{Z' \setminus Z} g \lceil \mathrm{d}\chi \rceil + \int_{Z} g \lceil \mathrm{d}\chi \rceil,$$

and a similar equation for the lower integral.

Recall that Z is a locally closed subanalytic subset of M and $f : M \to \mathbb{V}$ is a continuous subanalytic map. Consider a real valued kernel $\kappa \in L^1_{\mathrm{loc}}(\mathbb{R})$. The case of a complex kernel follows from the study of its real and imaginary parts. Choose $x_0 \in \mathbb{R} \cup \{\pm\infty\}$ such that $\mathcal{K} : x \in \mathbb{R} \mapsto \int_{x_0}^{x} \kappa(t)\,\mathrm{d}t$ is well-defined over $\mathbb{R}$. For the sake of readability, we extend the definition of $\mathcal{K}$ to any $x \in \mathbb{R} \cup \{\pm\infty\}$ such that the integral is well-defined.

Theorem 7.17 (Index-theoretic formula for sublevel-sets)**.** *Let $-\infty \leq a < b \leq +\infty$. Assume that $\kappa \cdot \mathbf{1}_{(-\infty,b)} \in L^1(\mathbb{R}_{\geq 0})$ and that $\mathcal{K}$ is subanalytic. For any $\xi \in \mathbb{V}^*$,*

(i) *if $\mathcal{K}_{|(a,b)}$ is strictly increasing, then*

$$\mathrm{Sub}_{\left[\kappa \mathbf{1}_{(a,b)}\right]}\left[f_{|Z}\right](\xi) = \mathcal{K}(b) \cdot \chi(\{\xi \circ f \leq b\} \cap Z) - \mathcal{K}(a) \cdot \chi(\{\xi \circ f \leq a\} \cap Z) - \int_{\{a < \xi \circ f \leq b\} \cap Z} \mathcal{K}(\xi \circ f) \lceil \mathrm{d}\chi \rceil,$$

(ii) *if $\mathcal{K}_{|(a,b)}$ is strictly decreasing, then*

$$\mathrm{Sub}_{\left[\kappa \mathbf{1}_{(a,b)}\right]}\left[f_{|Z}\right](\xi) = \mathcal{K}(b) \cdot \chi(\{\xi \circ f \leq b\} \cap Z) - \mathcal{K}(a) \cdot \chi(\{\xi \circ f \leq a\} \cap Z) - \int_{\{a < \xi \circ f \leq b\} \cap Z} \mathcal{K}(\xi \circ f) \lfloor \mathrm{d}\chi \rfloor,$$

with $|\mathcal{K}(b)| < +\infty$ and the convention that $\mathcal{K}(a) \cdot \chi(\{\xi \circ f \leq a\} \cap Z) = 0$ when $a = -\infty$.

The proof relies on the following technical lemma describing sublevel-sets constructible functions, stated and proven before the proof of Theorem 7.17.

Lemma 7.18. *Let $g : M \to \mathbb{R}$ be a continuous subanalytic function. There exist a finite family of integers $\{m_i\}_{1 \leq i \leq n}$ and of real numbers $-\infty < c_1 \leq \cdots \leq c_n \leq c_{n+1} < +\infty$, such that:*

(i) *One has $\varphi_{g_{|Z}} = \sum_{i=1}^{n} m_i \mathbf{1}_{[c_i, c_{i+1}]}$, and $\varphi^-_{g_{|Z}} = \sum_{i=1}^{n} m_i \mathbf{1}_{[c_i, +\infty)}$.*

(ii) *For all $-\infty \leq a < b \leq +\infty$, denoting $Z_{a,b} = \{a < g \leq b\} \cap Z$, one has:*

$$\varphi^-_{g_{|Z_{a,b}}} = \sum_{a < c_i \leq b} m_i \mathbf{1}_{[c_i, +\infty)}.$$

(iii) *If $\mathcal{K} : \mathbb{R} \to \mathbb{R}$ is a continuous subanalytic function that is strictly monotonic on an interval containing* $\mathrm{Im}(g)$*, then*

$$\varphi^{-}_{\mathcal{K}(g)_{|Z}} = \sum_{i=1}^{n} m_i \mathbf{1}_{[\mathcal{K}(c_i),+\infty)}.$$

Proof of Lemma 7.18. Result (i) is a straightforward consequence of Proposition 7.5 and of the convolution of indicator functions of closed intervals (6.5). To prove (ii), note that $\Gamma_g \cap Z_{a,b} \times \mathbb{R} = \Gamma_g \cap Z \times (a,b]$, so that $\mathbf{1}_{\Gamma_g \cap Z_{a,b} \times \mathbb{R}} = \mathbf{1}_{\Gamma_g \cap Z \times \mathbb{R}} \cdot p^* \mathbf{1}_{(a,b]}$, and hence:

$$\varphi_{g_{|Z_{a,b}}} = p_* \left(\mathbf{1}_{\Gamma_g \cap Z \times \mathbb{R}} \cdot p^* \mathbf{1}_{(a,b]} \right) = \mathbf{1}_{(a,b]} \cdot \varphi_{g_{|Z}}.$$

By Proposition 7.5, we have $\varphi^{-}_{g_{|Z_{a,b}}} = \varphi_{g_{|Z_{a,b}}} \star \mathbf{1}_{\mathbb{R}_{\geq 0}}$, so that:

$$\varphi^{-}_{g_{|Z_{a,b}}} = \sum_{i=1}^{n} m_i \mathbf{1}_{[c_i,c_{i+1}] \cap (a,b]} \star \mathbf{1}_{\mathbb{R}_{\geq 0}}.$$

The result follows then from direct calculations, since:

$$[c_i, c_{i+1}] \cap (a,b] = \begin{cases} [c_i, c_{i+1}] & \text{if } a \leq c_i \leq c_{i+1} \leq b, \\ [c_i, b] & \text{if } a \leq c_i \leq b \leq c_{i+1}, \\ (a, c_{i+1}] & \text{if } c_i \leq a \leq c_{i+1} \leq b, \\ (a, b] & \text{if } c_i \leq a < b \leq c_{i+1}. \end{cases}$$

Let us now prove (iii). Suppose that $a, b \in \mathbb{R}$ are such that $\mathrm{Im}(g) \subseteq [a,b]$ and $\mathcal{K}$ is strictly increasing on $[a,b]$. In this setting, the c_i's appearing in the decomposition of $\varphi^{-}_{g_{|Z}}$ can be chosen in $[a,b]$. If $u < \mathcal{K}(a)$, one has that $\chi\left(\{\mathcal{K}(g) \leq u\} \cap Z\right) = 0$ and if $u > \mathcal{K}(b)$, then one has $\chi\left(\{\mathcal{K}(g) \leq u\} \cap Z\right) = \chi_c(Z)$, hence the equality for such values of u by the fact that $\chi_c(Z) = \sum_{i=1}^{n} m_i$. Now, if $u \in \left[\mathcal{K}(a), \mathcal{K}(b)\right]$, we have:

$$\mathbf{1}_{[c_i,+\infty)}\left(\mathcal{K}^{-1}(u)\right) = \mathbf{1}_{[c_i,b]}\left(\mathcal{K}^{-1}(u)\right) = \mathbf{1}_{[\mathcal{K}(c_i),\mathcal{K}(b)]}(u) = \mathbf{1}_{[\mathcal{K}(c_i),+\infty)}(u),$$

so that we compute:

$$\begin{aligned} \chi\left(\{\mathcal{K}(g) \leq u\} \cap Z\right) &= \chi\left(\left\{g \leq \mathcal{K}^{-1}(u)\right\} \cap Z\right) \\ &= \sum_{i=1}^{n} m_i \cdot \mathbf{1}_{[c_i,+\infty)}\left(\mathcal{K}^{-1}(u)\right) \\ &= \sum_{i=1}^{n} m_i \cdot \mathbf{1}_{[\mathcal{K}(c_i),+\infty)}(u). \end{aligned}$$

□

Remark 7.19. In the setting of the previous lemma, Bobrowski and Borman (2012, Proposition 2.4) ensures that:

$$\int_Z g \lceil \mathrm{d}\chi \rceil = \sum_{i=1}^{n} m_i \cdot c_i.$$

Proof of Theorem 7.17. Let us first prove (i). The function $g = \xi \circ f$ being continuous and subanalytic, we can consider its sublevel-sets constructible function on Z written as in Lemma 7.18.(i). The fact that $\kappa \cdot \mathbf{1}_{(-\infty,b)} \in L^1(\mathbb{R}_{\geq 0})$ then ensures that $\kappa \cdot \mathbf{1}_{(-\infty,b)} \cdot \varphi^-_{g_{|Z}}$ is integrable over $\mathbb{R}$. Thus, the left-hand side of the equation to be proven is well-defined. Hence the result, by the computations:

$$\begin{aligned}
\int_a^b \kappa(t)\, \varphi^-_{g_{|Z}}(t)\, \mathrm{d}t &= \sum_{i=1}^n m_i \int_{\mathbb{R}} \mathbf{1}_{(a,b)\cap[c_i,+\infty)}(t)\, \kappa(t)\, \mathrm{d}t \\
&= \sum_{a<c_i\leq b} m_i \left(\int_{c_i}^b \kappa(t)\, \mathrm{d}t \right) + \sum_{c_i \leq a} m_i \left(\int_a^b \kappa(t)\, \mathrm{d}t \right) \\
&= \sum_{a<c_i\leq b} m_i \big(\mathcal{K}(b) - \mathcal{K}(c_i) \big) + \sum_{c_i \leq a} m_i \big(\mathcal{K}(b) - \mathcal{K}(a) \big) \\
&= \mathcal{K}(b) \cdot \sum_{c_i \leq b} m_i - \mathcal{K}(a) \cdot \sum_{c_i \leq a} m_i - \sum_{a<c_i\leq b} m_i\, \mathcal{K}(c_i) \\
&= \mathcal{K}(b) \cdot \chi(\{g \leq b\} \cap Z) - \mathcal{K}(a) \cdot \chi(\{g \leq a\} \cap Z) - \int_{Z_{a,b}} \mathcal{K}(g) \lceil \mathrm{d}\chi \rceil,
\end{aligned}$$

where $Z_{a,b} = \{a < g \leq b\} \cap Z$, and where the last equality follows from Remark 7.19 and Lemma 7.18.(ii) and Lemma 7.18.(iii) that yield:

$$\varphi^-_{\mathcal{K}(g)_{|Z_{a,b}}} = \sum_{a<c_i\leq b} m_i \mathbf{1}_{[\mathcal{K}(c_i),+\infty)}.$$

For (ii), note that $-\mathcal{K}$ satisfies the assumption of (i) and that:

$$\int_{Z'} \mathcal{K}(g) \lfloor \mathrm{d}\chi \rfloor = - \int_{Z'} -\mathcal{K}(g) \lceil \mathrm{d}\chi \rceil. \tag{7.9}$$

for any subanalytic locally closed subset Z' of M. This last equality is clear from the definition, and has first been proven in Baryshnikov and Ghrist (2010, Lemma 4). □

From the index-theoretic formula for sublevel-sets transforms, we deduce a formula for hybrid transforms of the level-sets constructible function φ_f for more general parameters $\zeta : \mathbb{V} \to \mathbb{R}$ morphisms of real analytic manifolds.

Corollary 7.20 (Index-theoretic formula for level-sets)**.** *Let $-\infty \leq a < b \leq +\infty$ and assume that $\mathcal{K}$ is subanalytic. For any $\zeta : \mathbb{V} \to \mathbb{R}$ morphism of real analytic manifolds,*

(i) *if $\mathcal{K}_{|(a,b)}$ is strictly increasing, then*

$$\mathrm{T}_{\kappa \mathbf{1}_{(a,b)}} \left[\varphi_{f_{|Z}} \right] (\zeta) = \int_{\widetilde{Z}_{a,b}} \mathcal{K}(\zeta \circ f) \lfloor \mathrm{d}\chi \rfloor - \int_{\widetilde{Z}_{a,b}} \mathcal{K}(\zeta \circ f) \lceil \mathrm{d}\chi \rceil,$$

(ii) *if $\mathcal{K}_{|(a,b)}$ is strictly decreasing, then*

$$\mathrm{T}_{\kappa \mathbf{1}_{(a,b)}} \left[\varphi_{f_{|Z}} \right] (\zeta) = \int_{\widetilde{Z}_{a,b}} \mathcal{K}(\zeta \circ f) \lceil \mathrm{d}\chi \rceil - \int_{\widetilde{Z}_{a,b}} \mathcal{K}(\zeta \circ f) \lfloor \mathrm{d}\chi \rfloor,$$

where $\widetilde{Z}_{a,b} = \{a \leq \zeta \circ f \leq b\} \cap Z$.

Proof. We show the first result, the second being proven identically. By Lemma 7.7, it is sufficient to prove the case $f = g : M \to \mathbb{R}$ and $\zeta = \mathrm{Id}_{\mathbb{R}}$. To use Theorem 7.17, one must first reduce to a situation where its conditions are met. Since $\mathrm{Im}(g)$ is compact, consider $-\infty < a' < b' < +\infty$ such that $(a', b') \supseteq \mathrm{Im}(g) \supseteq \mathrm{supp}(\varphi_g)$ and thus:

$$\mathrm{T}_{\left[\kappa \mathbf{1}_{(a,b)}\right]}\left[\varphi_{g_{|Z}}\right](1) = \mathrm{T}_{\left[\kappa \mathbf{1}_{(c,d)}\right]}\left[\varphi_{g_{|Z}}\right](1),$$

where $c = \max(a, a')$ and $d = \min(b, b')$. Now, for any $t \in \mathbb{R} \cup \{\pm\infty\}$, we have:

$$\chi\left(\{g = t\} \cap Z\right) = \chi\left(\{g \leq t\} \cap Z\right) + \chi\left(\{g \geq t\} \cap Z\right) - \chi\left(Z\right). \tag{7.10}$$

This yields $\varphi_{g_{|Z}} = \varphi^-_{g_{|Z}} + \varphi^+_{g_{|Z}} - \chi\left(Z\right) \cdot \mathbf{1}_{\mathbb{R}}$, and thus:

$$\mathrm{T}_{\left[\kappa \mathbf{1}_{(c,d)}\right]}\left[\varphi_{g_{|Z}}\right](1) = \mathrm{T}_{\left[\kappa \mathbf{1}_{(c,d)}\right]}\left[\varphi^-_{g_{|Z}}\right](1) + \mathrm{T}_{\left[\kappa \mathbf{1}_{(c,d)}\right]}\left[\varphi^+_{g_{|Z}}\right](1) - \chi\left(Z\right)\int_c^d \kappa(t)\,\mathrm{d}t. \tag{7.11}$$

Theorem 7.17 yields first:

$$\begin{aligned}\mathrm{T}_{\left[\kappa \mathbf{1}_{(c,d)}\right]}\left[\varphi^-_{g_{|Z}}\right](1) = \mathcal{K}(d) \cdot \chi(\{g \leq d\} \cap Z) - \mathcal{K}(c) \cdot \chi(\{g \leq c\} \cap Z) \\ - \int_{\{c < g \leq d\} \cap Z} \mathcal{K}(g) \lceil \mathrm{d}\chi \rceil.\end{aligned} \tag{7.12}$$

Moreover, using that $\varphi^+_{g_{|Z}}(t) = \varphi^-_{-g_{|Z}}(-t)$, we get:

$$\mathrm{T}_{\left[\kappa \mathbf{1}_{(c,d)}\right]}\left[\varphi^+_{g_{|Z}}\right](1) = \mathrm{T}_{\left[\widetilde{\kappa} \mathbf{1}_{(-d,-c)}\right]}\left[\varphi^-_{-g_{|Z}}\right](1),$$

where we denoted $\widetilde{\kappa}(t) = \kappa(-t)$. In that case, $\widetilde{\mathcal{K}}(x) = \int_{x_0}^x \widetilde{\kappa}(t)\,\mathrm{d}t$ satisfies $\widetilde{\mathcal{K}}(t) = -\mathcal{K}(-t)$, so it is subanalytic and strictly increasing. Thus, Theorem 7.17 yields:

$$\begin{aligned}\mathrm{T}_{\left[\kappa \mathbf{1}_{(c,d)}\right]}\left[\varphi^+_{g_{|Z}}\right](1) = \mathcal{K}(d) \cdot \chi(\{g \geq d\} \cap Z) - \mathcal{K}(c) \cdot \chi(\{g \geq c\} \cap Z) \\ + \int_{\{c \leq g < d\} \cap Z} \mathcal{K}(g) \lfloor \mathrm{d}\chi \rfloor.\end{aligned} \tag{7.13}$$

Putting (7.12) and (7.13) back into (7.11) yields:

$$\begin{aligned}\mathrm{T}_{\left[\kappa \mathbf{1}_{(c,d)}\right]}\left[\varphi_{g_{|Z}}\right](1) = \mathcal{K}(d) \cdot \big(\chi\left(\{g \leq d\} \cap Z\right) + \chi\left(\{g \geq d\} \cap Z\right) - \chi\left(Z\right)\big) \\ - \mathcal{K}(c) \cdot \big(\chi\left(\{g \leq c\} \cap Z\right) + \chi\left(\{g \geq c\} \cap Z\right) - \chi\left(Z\right)\big) \\ + \int_{\{c \leq g < d\} \cap Z} \mathcal{K}(g) \lfloor \mathrm{d}\chi \rfloor - \int_{\{c < g \leq d\} \cap Z} \mathcal{K}(g) \lceil \mathrm{d}\chi \rceil.\end{aligned}$$

Yet, Remark 7.16 yields:

$$\int_{\widetilde{Z}_{c,d}} \mathcal{K}(g) \lfloor \mathrm{d}\chi \rfloor = \mathcal{K}(d) \cdot \chi\left(\{g = d\} \cap Z\right) + \int_{\{c \leq g < d\} \cap Z} \mathcal{K}(g) \lfloor \mathrm{d}\chi \rfloor,$$

$$\int_{\widetilde{Z}_{c,d}} \mathcal{K}(g) \lceil \mathrm{d}\chi \rceil = \mathcal{K}(c) \cdot \chi\left(\{g = c\} \cap Z\right) + \int_{\{c < g \leq d\} \cap Z} \mathcal{K}(g) \lceil \mathrm{d}\chi \rceil.$$

Thus, combined with (7.10), we get:

$$\mathrm{T}_{\left[\kappa \mathbf{1}_{(c,d)}\right]}\left[\varphi_g\right](1) = \int_{\widetilde{Z}_{c,d}} \mathcal{K}(g) \lfloor \mathrm{d}\chi \rfloor - \int_{\widetilde{Z}_{c,d}} \mathcal{K}(g) \lceil \mathrm{d}\chi \rceil.$$

Hence the result, since $\{g \geq a'\} = \{g \leq b'\} = M$ by definition of a' and b', so that:

$$\widetilde{Z}_{c,d} \overset{\text{def}}{=} \{c \leq g \leq d\} \cap Z = \{a \leq g \leq b\} \cap Z \overset{\text{def}}{=} \widetilde{Z}_{a,b}.$$

□

Remark 7.21. This corollary and Lemma 7.7 also hold with identical proofs for $f : M \to X$ and $\zeta : X \to \mathbb{R}$ morphisms of real analytic manifolds, but we shall not make use of such a general statement in this thesis.

Applications to known transforms. An index theoretic formula is proven in Govc and Hepworth (2021, Theorem 6.1, Remark 6.2) for the sublevel-sets magnitude (7.6) when $Z = M$ is a closed smooth manifold. The results of this section allow us to state it for Z subanalytic locally closed and f continuous subanalytic:

Corollary 7.22 (ibid., Theorem 6.1)**.** *For any $\xi \in \mathbb{V}^*$,*

$$|\xi \cdot (Z, f)| = \int_Z e^{-\xi \circ f} \ \lfloor \mathrm{d}\chi \rfloor.$$

The results of this section also yield new results for the GR-Euler-Fourier and the Euler-Bessel transforms.

Corollary 7.23. *Let γ be a cone satisfying* (C2)*. For any $\xi \in \mathrm{Int}(\gamma^\circ)$, we have:*

$$\mathcal{EF}^{\mathrm{GR}}\left[\varphi^\gamma_{f_{|Z}}\right](\xi) = \int_{\{\xi \circ f \geq 0\} \cap Z} \xi \circ f \lfloor \mathrm{d}\chi \rfloor.$$

Proof. Proposition 7.8 yields $(-\xi)_* \varphi^\gamma_{f_{|Z}} = \varphi^-_{-\xi \circ f_{|Z}}$, so that $\mathcal{EF}^{\mathrm{GR}}\left[\varphi^\gamma_{f_{|Z}}\right](\xi) = \chi^0_{f_{|Z}}(-\xi)$. □

Corollary 7.24. *For any $\xi \in \mathbb{V}^*$, we have:*

$$\mathcal{EF}^{\mathrm{GR}}\left[\varphi_{f_{|Z}}\right](\xi) = \int_{\{\xi \circ f \geq 0\} \cap Z} \xi \circ f \lfloor \mathrm{d}\chi \rfloor - \int_{\{\xi \circ f \geq 0\} \cap Z} \xi \circ f \lceil \mathrm{d}\chi \rceil.$$

Example 7.25. Let Z be a relatively compact subset of $\mathbb{V}$. Consider a compact real analytic submanifold M of $\mathbb{V}$ containing Z and $f : M \hookrightarrow \mathbb{V}$ the inclusion, so that $\varphi_{f_{|Z}} = \mathbf{1}_Z$. Corollary 7.24 yields the following generalization of Ghrist and Robinson (2011, Theorem 4.4):

$$\mathcal{EF}^{\mathrm{GR}}\left[\mathbf{1}_Z\right](\xi) = \int_{\{\xi \geq 0\} \cap Z} \xi \lfloor \mathrm{d}\chi \rfloor - \int_{\{\xi \geq 0\} \cap Z} \xi \lceil \mathrm{d}\chi \rceil.$$

The parameter $\zeta : \mathbb{V} \to \mathbb{R}$ appearing in the definition of the Euler-Bessel transform is not linear, so an index-theoretic formula for sublevel-sets constructible functions is out of reach using the results of the present chapter. Yet, Corollary 7.20 yields the following formula for level-sets constructible functions.

Corollary 7.26. *Consider an analytic norm $\|\cdot\|$ on $\mathbb{V}$. For any $v \in \mathbb{V}$, we have:*

$$\mathcal{EB}\left[\varphi_{f_{|Z}}\right](v) = \int_Z \|v-f\| \lfloor \mathrm{d}\chi \rfloor - \int_Z \|v-f\| \lceil \mathrm{d}\chi \rceil,$$

where $\|v-f\| : x \in M \mapsto \|v-f(x)\|$.

Example 7.27. In the setting of Example 7.25, Corollary 7.26 yields the following generalization of Ghrist and Robinson (2011, Theorem 4.2):

$$\mathcal{EB}\left[\mathbf{1}_Z\right](v) = \int_Z \|v-\cdot\| \lfloor \mathrm{d}\chi \rfloor - \int_Z \|v-\cdot\| \lceil \mathrm{d}\chi \rceil.$$

7.3 Mean Euler-Bessel transform for random filtrations

In this section, we combine Bobrowski and Borman (2012, Theorem 4.1) and our index-theoretic formula to prove expectation results for the Euler-Bessel transform of level-sets constructible functions of Gaussian random fields.

Denote by d the dimension of M and let $f : M \to \mathbb{R}^k$ with $k \geq 1$ be a k-dimensional Gaussian field with iid components all having zero mean and unit variance. Note that if $\xi \in \left(\mathbb{R}^k\right)^*$ is such that $\|\xi\| = 1$, then a direct checking ensures that $\xi \circ f : M \to \mathbb{R}$ is also a centered Gaussian field with unit variance. To state the result, we use classical geometric quantities known as the Lipschitz-Killing curvatures $\mathcal{L}_j(M)$ with respect to the metric induced by the centered unit-variance Gaussian field $\xi \circ f : M \to \mathbb{R}$. We refer to Adler and Taylor (2009, Section 7.6) for the definition of the curvatures and to Adler and Taylor (ibid., Section 12.2) for the definition of the metric induced by a Gaussian field.

Corollary 7.28. *Suppose that the components of f are almost surely subanalytic and Morse functions. Then, for any $v \in \mathbb{V}$,*

$$\mathbb{E}\left[\mathcal{EB}\left[\varphi_f\right](v)\right] = \begin{cases} 2 \cdot \sum_{j=1}^{d} (2\pi)^{-j/2} \mathcal{L}_j(M)\, c_j(v) & \text{if } d \text{ is odd,} \\ 0 & \text{if } d \text{ is even.} \end{cases}$$

where

$$c_j(v) = \sum_{i=0}^{+\infty} \frac{e^{-\|v\|^2/2}\|v\|^{2i}}{2^i i!} \sum_{l=0}^{\lfloor (j-1)/2 \rfloor} \sum_{m=0}^{j-1-2l} \mathbf{1}_{\{k \geq j-m-2l-2i\}} \binom{k+2i-1}{j-1-m-2l} \times \frac{(-1)^{m+l}(j-1)!\Gamma((k+2i-j-2m+2l+1)/2)}{m!l!2^{(j-1-2m)/2}\Gamma((k+2i)/2)}.$$

Proof. Let us first remark that the following equations are straightforward consequences

of Bobrowski and Borman (2012, Theorem 4.1):

$$\mathbb{E}\left[\int_M d_v(f)\,\lceil \mathrm{d}\chi\rceil\right] = \chi(M)\,\mathbb{E}\left[d_v(f)\right] - \sum_{j=1}^{d}(2\pi)^{-j/2}\mathcal{L}_j(M)\int_{\mathbb{R}}\mathcal{M}_j\left(d_v^{-1}(-\infty,u]\right)\,\mathrm{d}u,$$

$$\mathbb{E}\left[\int_M d_v(f)\,\lfloor \mathrm{d}\chi\rfloor\right] = \chi(M)\,\mathbb{E}\left[d_v(f)\right] + \sum_{j=1}^{d}(2\pi)^{-j/2}\mathcal{L}_j(M)\int_{\mathbb{R}}\mathcal{M}_j\left(d_v^{-1}[u,+\infty)\right)\,\mathrm{d}u,$$

where we denoted by $d_v : u \in \mathbb{R}^k \mapsto \|u-v\| \in \mathbb{R}$, and where the Gaussian Minkowski functionals $\mathcal{M}_j$ are defined by the tube formula (Adler and Taylor, 2009, Eq. 10.9.11). Thus, the index theoretic formula for the Euler-Bessel transform (Corollary 7.26) yields:

$$\begin{aligned}\mathbb{E}\left[\mathcal{EB}\left[\varphi_f\right](v)\right] &= \mathbb{E}\left[\int_M d_v(f)\,\lfloor \mathrm{d}\chi\rfloor\right] - \mathbb{E}\left[\int_M d_v(f)\,\lceil \mathrm{d}\chi\rceil\right]\\ &= \sum_{j=1}^{d}(2\pi)^{-j/2}\mathcal{L}_j(M)\int_{\mathbb{R}}\mathcal{M}_j\left(d_v^{-1}[u,+\infty)\right) + \mathcal{M}_j\left(d_v^{-1}(-\infty,u]\right)\mathrm{d}u.\end{aligned} \tag{7.14}$$

Following Taylor (2006, Section 5.2), we use the expressions of the densities and their derivatives of noncentral χ_k^2 random variables and tube formulae to compute the Gaussian-Minkowski curvatures involved:

$$\begin{aligned}\mathcal{M}_j\left(d_v^{-1}(-\infty,u]\right) &= \mathbf{1}_{(0,+\infty)}(u)\cdot\left(\frac{\mathrm{d}^{j-1}f_{v,k}}{\mathrm{d}x^{j-1}}\right)_{x=u},\\ \mathcal{M}_j\left(d_v^{-1}[u,+\infty)\right) &= (-1)^{j-1}\mathbf{1}_{(0,+\infty)}(u)\cdot\left(\frac{\mathrm{d}^{j-1}f_{v,k}}{\mathrm{d}x^{j-1}}\right)_{x=u},\end{aligned} \tag{7.15}$$

for any $u \in \mathbb{R}$, where we denoted by $f_{v,k}$ the density of the square root of a noncentral χ_k^2 random variable with noncentrality parameter $\|v\|^2$, i.e., for any $x \in \mathbb{R}$, one has:

$$f_{v,k}(x) = \sum_{i=0}^{+\infty} e^{-\|v\|^2/2}\frac{\|v\|^{2i}}{2^i i!}\cdot\frac{x^{k+2i-1}e^{-x^2/2}}{2^{(k+2i-2)/2}\Gamma((k+2i)/2)}.$$

This yields the following expression:

$$\mathbb{E}\left[\mathcal{EB}\left[\varphi_f\right](v)\right] = \sum_{j=1}^{d}(2\pi)^{-j/2}\mathcal{L}_j(M)\left(1+(-1)^{j-1}\right)\int_0^{+\infty}\mathcal{M}_j\left(d_v^{-1}(-\infty,u]\right)\,\mathrm{d}u.$$

The result for even dimensions d follows then from the fact that Lipschitz-Killing curvatures $\mathcal{L}_j(M)$ vanish for odd values of j and $1+(-1)^{j-1}$ vanishes for even values of j. For odd d, the result follows from the expression:

$$\begin{aligned}\left(\frac{\mathrm{d}^{j-1}f_{v,k}}{\mathrm{d}x^{j-1}}\right)_{x=u} = \sum_{i=0}^{+\infty} e^{-\|v\|^2/2}\frac{\|v\|^{2i}}{2^i i!}\cdot\frac{u^{k+2i-j}e^{-u^2/2}}{\Gamma((k+2i)/2)2^{(k+2i-2)/2}}\sum_{l=0}^{\lfloor (j-1)/2\rfloor}\sum_{m=0}^{j-1-2l}\mathbf{1}_{\{k\geq j-m-2l-2i\}}\\ \times\binom{k+2i-1}{j-1-m-2l}\frac{(-1)^{m+l}(j-1)!}{m!l!2^l}u^{2m+2l},\end{aligned}$$

for $u \geq 0$—see Taylor (2006, Section 5.2)—and the clear following one:

$$\int_0^{+\infty} u^{k+2i-j+2m+2l} e^{-u^2/2} \, \mathrm{d}u = 2^{(k+2i-j+2m+2l-1)/2} \Gamma\left((k+2i-j+2m+2l+1)/2\right).$$

□

We end this chapter by mentionning that the same methodology applies to sublevel-sets persistent magnitude:

Corollary 7.29. *For $\xi \in \mathbb{V}^*$ such that $\|\xi\| = 1$ and $\xi \circ f$ is almost surely subanalytic and a Morse function,*

$$\mathbb{E}\Big[\big|\xi \cdot (M, f)\big|\Big] = \sqrt{e} \cdot \sum_{j=0}^{d} \frac{\mathcal{L}_j(M)}{(2\pi)^{j/2}}.$$

Proof. By Corollary 7.22, we have:

$$\mathbb{E}\Big[\big|\xi \cdot (M, f)\big|\Big] = \mathbb{E}\left[\int_M e^{-\xi \circ f} \, \lfloor \mathrm{d}\chi \rfloor\right] = -\mathbb{E}\left[\int_M -e^{-\xi \circ f} \, \lceil \mathrm{d}\chi \rceil\right],$$

and Bobrowski and Borman (2012, Corollary 4.3) yields:

$$\mathbb{E}\Big[\big|\xi \cdot (M, f)\big|\Big] = \sum_{j=0}^{d} \frac{(-1)^j \mathcal{L}_j(M)}{(2\pi)^{-j/2}} \int_{-\infty}^{+\infty} \varphi(x) H_j(x) e^{-x} \, \mathrm{d}x$$

where $\varphi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$. Since $H_j(x) = (-1)^j \varphi(x)^{-1} \frac{\mathrm{d}^j \varphi(x)}{\mathrm{d}x^j}$, an integration by parts yields:

$$\begin{aligned}
\int_{-\infty}^{+\infty} \varphi(x) H_j(x) e^{-x} \, \mathrm{d}x &= (-1)^j \int_{-\infty}^{+\infty} \frac{\mathrm{d}^j \varphi(x)}{\mathrm{d}x^j} e^{-x} \, \mathrm{d}x \\
&= (-1)^j \int_{-\infty}^{+\infty} \varphi(x) e^{-x} \, \mathrm{d}x \\
&= (-1)^j \sqrt{e}.
\end{aligned}$$

□

Chapter 8

Euler characteristic profiles and their transforms

Based on joint work with Olympio Hacquard.

Consider a practical scenario where a filtered simplicial complex is built from the data at hand to extract some topological and geometric information. For instance, this filtered complex could be a Čech filtration or a function-Čech filtration (Section 4.1.1). Computing the Euler characteristic at every value of the parameter of the filtration, one obtains an integer-valued function describing the topological evolution of the filtration with respect to the parameter. This function is called the *Euler characteristic profile* of the filtration. Under mild assumptions, this function is constructible and one can thus consider its hybrid transforms. Several natural questions arise: how do these descriptors of filtered simplicial complexes behave with respect to small perturbations on the data? What are the properties of these tools in random contexts, such as Čech filtrations of random point clouds? For instance, do they satisfy a central limit theorem?

Stability. One of the main reasons for the success of persistent homology in topological data analysis is the stability theorem for persistence barcodes from Cohen-Steiner, Edelsbrunner, and Harer (2007). Loosely speaking, it means that, under mild assumptions, small changes in the filtration function imply small changes in the barcode. Such results are key to designing consistent estimators in statistical analysis, as it is done in Bobrowski, Mukherjee, and Taylor (2017) and Chazal, Glisse, et al. (2015). Other distances on persistence barcodes also satisfy some stability results, such as the p-Wasserstein distance between persistence barcodes (Cohen-Steiner, Edelsbrunner, Harer, and Mileyko, 2010; Skraba and Turner, 2020).

Distances on Euler characteristic profiles have already been studied and several stability results have been derived. In the one-parameter setting, stability results have been proven for Euler characteristic curves in Curry, Mukherjee, and Turner (2022), Dłotko and Gurnari (2022), and Perez (2022). All statements bound the L_1 distance on Euler curves by the 1-Wasserstein distance between the associated persistence barcodes. The most general statements for functions in Sobolev spaces of compact Riemannian manifolds can be found in Perez (2022). Results of loc. cit. rely crucially on a fine

upper-bound of Buhovsky et al. (2022) on the number of bars greater than ε in the sublevel-sets persistence barcodes of functions in such Sobolev spaces.

In the multi-parameter setting, Y. Chen et al. (2022) define a distance between Euler profiles of two-parameter filtrations indexed over finite grids and prove an associated stability result. Their distance is the sum of the 1-Wasserstein distances between one-parameter barcodes of the restrictions of persistent homology to rows and columns of the grid. As explained in Dłotko and Gurnari (2022, Section 3.2), the stability result proven in Y. Chen et al. (2022) is trivial in most cases as the Wasserstein distance between two barcodes is infinite as soon as their number of infinite bars is not the same. To cope with these limitations, Dłotko and Gurnari (2022) consider L_1 distances between Euler profiles and prove stability results for the restrictions of Euler profiles to compact subsets of their domain (Proposition 3.4 of loc. cit.). Similar results are obtained in Curry, Mukherjee, and Turner (2022, Lemma 4.10). Meanwhile, the stability results of Oudot and Scoccola (2021) for the Hilbert functions of persistence modules can be transferred to Euler characteristic profiles. In contrast, these last results are expressed in terms of signed 1-Wasserstein distances between Hilbert functions and not in terms of classical functional norms on constructible functions. Oudot and Scoccola (ibid.) prove that this specific distance is related to other known distances between persistence modules, namely the 1-Wasserstein distance in the one-parameter case and the 1-presentation distance (Bjerkevik and Lesnick, 2021) in the two-parameter case.

Statistics. Euler characteristic tools have been extensively studied in random geometry. Mean formulae for the Euler characteristic of superlevel sets of random fields are proven in Adler and Taylor (2009), and the limiting behavior of the Euler characteristic of a complex built on a Poisson process are established in Corollary 4.2 of Bobrowski and Adler (2014). Furthermore, Euler curves associated with random point clouds are proven to be asymptotically normal for a well-chosen sampling regime in Krebs, Roycraft, and Polonik (2021), where the authors also apply this construction to bootstrap. See Bobrowski and Kahle (2018) for a survey on the asymptotic behavior of topological descriptors of random complexes in the one-parameter setting. All these results concern invariants of one-parameter persistence modules, and the multi-parameter setting remains almost uncharted. The first result in this direction has been proven by Botnan and Hirsch (2022) for multi-parameter persistent Betti numbers.

Outline. In this chapter, we consider constructible functions associated to the persistent homology of multi-parameter simplicial filtrations, called *Euler characteristic profiles* (Section 8.1). We conduct a theoretical study of these constructible functions and their hybrid transforms as descriptors of filtered simplicial complexes.

In Section 8.2, we prove stability properties that clarify the robustness of Euler characteristic profiles and their transforms with respect to perturbations. First, we recall the definition of the signed 1-Wasserstein distance introduced in Oudot and Scoccola (2021) and state its connections with other known distances on persistence modules (Section 8.2.1). Then, we prove that the L_1 distance between restrictions of Euler profiles to compact subsets is upper-bounded by the signed 1-Wasserstein distance between them (Section 8.2.2). The combination of this result and of the results on the signed

Wasserstein distance gives an essentially equivalent proof of the results of Dłotko and Gurnari (2022). However, the intermediate formulation using the signed Wasserstein distance allows for a unified treatment of the one-parameter and multi-parameter cases. In particular, it implies that the L_1 distance between two-parameter Euler profiles is upper-bounded by the 1-presentation distance between the persistent homologies of the filtrations. Moreover, we derive stability results for hybrid transforms of Euler characteristic profiles in terms of L_q distances.

In Section 8.3, we establish the stability results for hybrid transforms associated to filtered simplicial complexes of random point clouds. In Section 8.3.1, we state that under some mild assumptions, hybrid transforms are universal in the sense that they converge to an object that depends only on the kernel, the filtration, and the sampling scheme. In addition, we illustrate how information on the sampling can be extracted from this limiting object. This shows that if the number of sample points is large enough, hybrid transforms are relevant tools to perform classification tasks on point clouds. Finally, we establish a law of large numbers for hybrid transforms in a multi-filtration set-up in Section 8.3.2. This result relies on the link between multi-parameter hybrid transforms and one-parameter ones in the context of sublevel-sets persistence (Corollary 7.10) and is not obvious to prove for Euler characteristic profiles. This is another illustration of the interest of hybrid transforms as topological descriptor of filtered complexes. Together with the works of Botnan and Hirsch (2022), our results form the first occurrence of limiting theorems in a multi-persistence framework in the literature.

Convention. We always consider the coordinatewise order on $\mathbb{R}^m$. As in Chapter 4, we assume for convenience in the rest of the chapter that all filtrations are simplicial filtrations, that is, composed of simplicial subcomplexes. Recall that the dual of a vector space $\mathbb{V}$ is denoted by $\mathbb{V}^*$, and that $\mathbb{R}^m$ is always identified with its dual under the canonical isomorphism. For $\xi \in \mathbb{R}^{m*}$ and $t \in \mathbb{R}^m$, we often denote $\xi(t)$ by $\langle \xi, x \rangle$.

8.1 Persistent homology of simplicial filtrations

In this section, we introduce our notations for the persistent homology of simplicial filtrations (Section 4.1.1) and for the associated constructible functions (Section 5.4).

Persistent homology. An m-parameter filtration is called *finitely presentable* if its p-th persistent homology $H_p(\mathcal{F})$ is finitely presentable for all integer $p \geq 0$. In that case, since $\mathcal{K}$ is finite, only finitely many $H_p(\mathcal{F})$ are non-zero and the graded persistence module $H_*(\mathcal{F}) = \bigoplus_{p \in \mathbb{N}_0} H_p(\mathcal{F})$ is finitely presentable in the sense of Section 5.4. Throughout this section, consider a finitely presentable m-parameter filtration $\mathcal{F}$ of a simplicial complex $\mathcal{K}$.

Euler characteristic profiles. Following Section 5.4, we consider the finitely presentable constructible function associated to the graded persistent module $H_*(\mathcal{F})$. We call it the *Euler characteristic profile* of $\mathcal{F}$ and denote it simply by $\chi_{\mathcal{F}}$. This is equivalent to compute the Euler characteristic for every value of the parameter $t \in \mathbb{R}^m$ of the

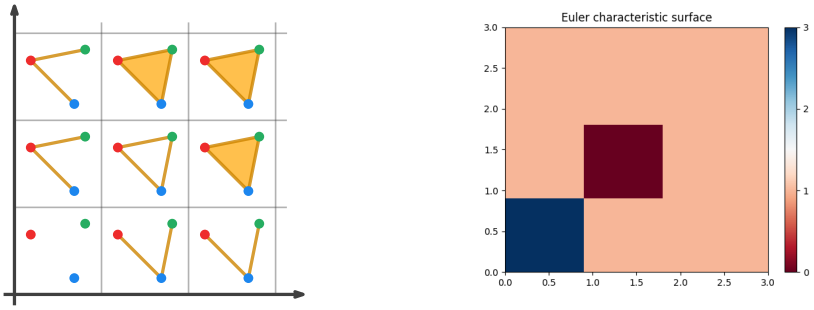
filtration:

$$\chi_{\mathcal{F}} : t \in \mathbb{R}^m \mapsto \chi(\mathcal{F}_t),$$

where the Euler characteristic of a simplicial complex $\mathcal{K}'$ is defined as:

$$\chi(\mathcal{K}') = \sum_{\sigma \in \mathcal{K}'} (-1)^{\dim \sigma}.$$

We often call $\chi_{\mathcal{F}}$ simply the *Euler profile* of $\mathcal{F}$. The map $\chi_{\mathcal{F}}$ is usually referred to as the *Euler characteristic curve* (ECC) of $\mathcal{F}$ when $m = 1$ and as the *Euler characteristic surface of* (ECS) of $\mathcal{F}$ when $m = 2$; see Beltramo, Skraba, et al. (2022), Dłotko and Gurnari (2022), and Turner, Mukherjee, and Boyer (2014a). We show in Figure 8.1 an Euler characteristic surface computed on an elementary example.



(a) $\mathcal{F}$ (b) $\chi_{\mathcal{F}}$

Figure 8.1: A finitely presentable two-parameter filtration (a) and its associated Euler characteristic surface (b). All vertices have one birth time, while all other simplices have two.

Example 8.1. Recall that given a non-decreasing map $f : \mathcal{K} \to \mathbb{R}^m$ one can define a the sublevel-sets filtration $\mathcal{F}_f$. We denote by χ_f its Euler characteristic profile. Recall from Section 4.1.1 that one-critical filtrations are equivalent to sublevel-sets filtrations.

The Hilbert function of the p-th persistent homology of $\mathcal{F}$ is called the *p-th Betti function* of $\mathcal{F}$ and denoted by $\beta_{\mathcal{F},p}$. In particular, we have:

$$\chi_{\mathcal{F}} = \sum_{p \in \mathbb{N}_0} (-1)^p \beta_{\mathcal{F},p}.$$

Integral transforms. Since the Euler characteristic profile of a finitely presentable filtration is constructible up to infinity, one can compute its Radon transform. Let us denote by $\mathbb{R}^{m*}_+ = \mathrm{Int}(\gamma^{a\circ})$ for the cone $\gamma = (\mathbb{R}_{\leq 0})^n$, so that any $\xi \in \mathbb{R}^{m*}_+$ is proper on the support of $\chi_{\mathcal{F}}$. Then $\mathbb{R}^{m*}_+$ is the cone of linear forms on $\mathbb{R}^m$ that are non-decreasing with respect to the coordinatewise order on $\mathbb{R}^m$. We restrict our study of the Radon transform of Euler characteristic profiles to this cone. Specifically, we follow (5.3) and call *Radon transform* of $\chi_{\mathcal{F}}$ the following map:

$$\mathcal{R}_{\mathcal{F}} : (\xi, s) \in \mathbb{R}^{m*}_+ \times \mathbb{R} \longmapsto \xi_* \chi_{\mathcal{F}}(s).$$

Let $\kappa \in L^1(\mathbb{R})$. Remark 6.14 ensures that the hybrid transform of $\chi_{\mathcal{F}}$ with kernel κ is defined on $\mathbb{R}_+^{m*}$. Moreover, Corollary 7.10 shows that hybrid transforms have a remarkable property in the context of sublevel-sets persistence. Let $f : \mathcal{K} \to \mathbb{R}^m$ be a non-decreasing function and let $\xi \in \mathbb{R}_+^{m*}$. Then, for all $s > 0$, one has:

$$\mathrm{T}_\kappa\left[\chi_f\right](s\,\xi) = \mathrm{T}_\kappa\left[\chi_{\xi \circ f}\right](s). \tag{8.1}$$

Euler characteristic profiles, Radon transforms, and hybrid transforms constitute the three descriptors of data we use to perform topological data analysis in Chapter 9. We give explicit expressions of these descriptors in two specific case below. These formulae will allow us to design algorithms to compute them in Section 9.1 and to build intuition on the type of behavior they capture all along Chapter 9.

Connection with persistence barcodes. Suppose that $\mathcal{F}$ is a one-parameter filtration. Since $H_p(\mathcal{F})$ is finitely presentable, there exists a finite subset P of $\mathbb{R}$ such that $H_p(\mathcal{F})$ is isomorphic to the left Kan extension along $P \hookrightarrow \mathbb{R}$ of its restriction to P; see Bauer and Scoccola (2022, Lemma 5) for a proof of this well-known fact. Using Gabriel's decomposition theorem of quiver representations over finite totally ordered sets (1972), it is easy to show that there exist an integer $n_p \geq 0$ and (possibly infinite) real numbers $-\infty < a_i^p < b_i^p \leq \infty$ for $i \in \{1, \ldots, n_p\}$ such that the persistence module $H_p(\mathcal{F})$ decomposes as:

$$H_p(\mathcal{F}) \simeq \bigoplus_{i=1}^{n_p} \mathbf{k}_{[a_i^p, b_i^p)}. \tag{8.2}$$

The multiset of intervals $[a_i^p, b_i^p)$ for $i \in \{1, \ldots, n_p\}$ is the barcode of $H_p(\mathcal{F})$ as defined in Section 1.1 and denoted simply by $\mathcal{B}_p\mathcal{F}$. Note that this barcode $\mathcal{B}_p\mathcal{F}$ is empty for $p >> 1$.

In the above situation, the Euler characteristic profile of $\mathcal{F}$ has the following expression:

$$\chi_{\mathcal{F}} = \sum_{k \geq 0} \sum_{i=1}^{n_k} (-1)^k \mathbf{1}_{[a_i^k, b_i^k)}. \tag{8.3}$$

Therefore, the Radon transform of $\chi_{\mathcal{F}}$ is:

$$\mathcal{R}_{\mathcal{F}} : (\xi, s) \in \mathbb{R}_+^* \times \mathbb{R} \mapsto \sum_{k \geq 0} \sum_{i=1}^{n_k} (-1)^k \mathbf{1}_{[\xi \cdot a_i^k, \xi \cdot b_i^k)}(s). \tag{8.4}$$

Let $\kappa \in L^1(\mathbb{R})$ and consider a primitive $\overline{\kappa}$ of κ. The hybrid transform with kernel κ of $\chi_{\mathcal{F}}$ is:

$$\mathrm{T}_\kappa\left[\chi_{\mathcal{F}}\right] : \xi \in \mathbb{R}_+^* \mapsto \sum_{k \geq 0} \sum_{i=1}^{n_k} (-1)^k \left(\overline{\kappa}\left(\xi \cdot b_i^k\right) - \overline{\kappa}\left(\xi \cdot a_i^k\right) \right), \tag{8.5}$$

with the convention that $\overline{\kappa}(\xi \cdot b_i^k)$ is the limit of $\overline{\kappa}$ at $+\infty$ when $b_i^k = +\infty$.

One-critical filtrations. Up to reducing $\mathcal{K}$, one can assume that for any $\sigma \in \mathcal{K}$, there is $t \in \mathbb{R}^m$ with $\sigma \in \mathcal{F}_t$. Then, one has:

$$\chi_{\mathcal{F}} = \sum_{\sigma \in \mathcal{K}} (-1)^{\dim \sigma} \mathbf{1}_{Q_{t(\sigma)}}, \tag{8.6}$$

where $Q_u := \{t \in \mathbb{R}^m \,;\, t \geq u\}$ for any $u \in \mathbb{R}^m$. As a consequence, one has by Lemma 5.10:

$$\mathcal{R}_{\mathcal{F}} : (\xi, s) \in \mathbb{R}_+^{m*} \times \mathbb{R} \mapsto \sum_{\sigma \in \mathcal{K}} (-1)^{\dim \sigma} \mathbf{1}_{[\langle \xi, t(\sigma) \rangle, +\infty)}(s). \tag{8.7}$$

Let $\kappa \in L^1(\mathbb{R})$. Denote by $\overline{\kappa}$ the primitive of κ whose limit at $+\infty$ is 0. The hybrid transform with kernel κ of $\chi_{\mathcal{F}}$ is:

$$\mathrm{T}_{\kappa}\left[\chi_{\mathcal{F}}\right] : \xi \in \mathbb{R}_+^{m*} \longmapsto -\sum_{\sigma \in \mathcal{K}} (-1)^{\dim \sigma} \overline{\kappa}\left(\langle \xi, t(\sigma) \rangle\right). \tag{8.8}$$

Remark 8.2. We often define hybrid transforms by specifying the primitive $\overline{\kappa}$ of the kernel κ whose limit at $+\infty$ is 0. We call $\overline{\kappa}$ the *primitive kernel* of the hybrid transform.

8.2 Stability properties

In this section, we prove stability results for Euler characteristic based descriptors of filtered simplicial complexes. More precisely, our results compare the L_1 distances between Euler characteristic profiles to the signed 1-Wasserstein distance between their signed barcodes. As a direct corollary, we bound the L_q norms of Radon transforms and hybrid transforms by the same quantity. The notions of signed barcodes and of signed 1-Wasserstein distance are introduced in Oudot and Scoccola (2021) for Hilbert functions and recalled below. We follow the same conventions as in Oudot and Scoccola (ibid., Section 2) for the definitions of multisets and bijections between them.

8.2.1 Signed 1-Wasserstein distance

Finitely presentable functions possess a kind of barcode that is used by Oudot and Scoccola (ibid.) to define an analogue of the 1-Wasserstein distance between one-parameter persistence barcodes (Cohen-Steiner, Edelsbrunner, Harer, and Mileyko, 2010). A *decomposition* of $\varphi \in \mathrm{CF}_{\mathrm{fp}}(\mathbb{R}^m)$ is a couple $(\mathcal{B}^+, \mathcal{B}^-)$ of finite multisets of points in $\mathbb{R}^m$ such that:

$$\varphi = \sum_{u \in \mathcal{B}^+} \mathbf{1}_{Q_u} - \sum_{v \in \mathcal{B}^-} \mathbf{1}_{Q_v}.$$

Such a decomposition always exists, and there is a unique $\overline{\mathcal{B}} = (\mathcal{B}^+, \mathcal{B}^-)$ such that $\mathcal{B}^+ \cap \mathcal{B}^- = \varnothing$, called the *signed barcode* of φ; see Oudot and Scoccola (2021, Proposition 13). While two different notions of signed barcode are defined in loc. cit., we focus here on the so-called *minimal Hilbert decomposition signed barcode*.

Let $\mathcal{C}$ and $\mathcal{C}'$ be two finite multisets of points in $\mathbb{R}^m$ with the same cardinality and $h : \mathcal{C} \to \mathcal{C}'$ be a bijection between them. The *cost* of h is the real number $\mathrm{cost}(h) = \sum_{u \in \mathcal{C}} \|u - h(u)\|_1$. For any two finitely presentable functions φ and φ' with respective signed barcodes $(\mathcal{B}^+, \mathcal{B}^-)$ and $(\mathcal{B}'^+, \mathcal{B}'^-)$, the *signed 1-Wasserstein distance* between them is:

$$\widehat{d}_1(\varphi, \varphi') = \inf\left\{\varepsilon > 0 \,;\, \exists \text{ bijection } h : \mathcal{B}^+ \cup \mathcal{B}'^- \to \mathcal{B}^- \cup \mathcal{B}'^+ \text{ with } \mathrm{cost}(h) \leq \varepsilon\right\}.$$

Hence, one has $\widehat{d}_1(\varphi, \varphi') \in [0, +\infty]$. Note that bijections do not allow for unmatched bars, as it is common in the persistence literature. Note that the signed 1-Wasserstein distance

is defined in loc. cit. only on signed barcodes. Our definition is essentially equivalent since signed barcodes are in one-to-one correspondence with finitely presentable functions.

Remark 8.3. In the following proofs, we make constant use of the fact that the distance $\widehat{d}_1$ may be computed on any decomposition of the functions and not only on minimal ones, that is, on signed barcodes. More precisely, for any decompositions $(\mathcal{C}^+, \mathcal{C}^-)$ and $(\mathcal{C}'^+, \mathcal{C}'^-)$ of two finitely presentable functions φ and φ' respectively, one has:

$$\widehat{d}_1(\varphi, \varphi') = \inf\left\{\varepsilon > 0\,;\, \exists \text{ bijection } h : \mathcal{C}^+ \cup \mathcal{C}'^- \to \mathcal{C}^- \cup \mathcal{C}'^+ \text{ with } \mathrm{cost}(h) \leq \varepsilon\right\}.$$

The signed 1-Wasserstein distance between Euler characteristic profiles is upper-bounded by the same distance between Betti functions, as stated in the lemma below. It will be crucial to proving the other results.

Lemma 8.4. *Let $\mathcal{F}$ and $\mathcal{F}'$ be two finitely presentable m-parameter filtrations of simplicial complexes $\mathcal{K}$ and $\mathcal{K}'$ respectively. Then,*

$$\widehat{d}_1(\chi_{\mathcal{F}}, \chi_{\mathcal{F}'}) \leq \sum_{k\geq 0} \widehat{d}_1\big(\beta_{\mathcal{F},k}, \beta_{\mathcal{F}',k}\big)\,.$$

Proof. By assumption, the persistence modules $H_k(\mathcal{F})$ and $H_k(\mathcal{F}')$ are finitely presentable hence so are the functions $\beta_{\mathcal{F},k}$ and $\beta_{\mathcal{F}',k}$. A collection of decompositions $(\mathcal{B}_k^+, \mathcal{B}_k^-)$ of $\beta_{\mathcal{F},k}$ for all $k \in \mathbb{N}_0$ induces a decomposition $(\mathcal{B}^+, \mathcal{B}^-) = (\cup_k \mathcal{B}_k^+, \cup_k \mathcal{B}_k^-)$ of $\chi_{\mathcal{F}}$. A similar decomposition $(\mathcal{B}'^+, \mathcal{B}'^-)$ of $\chi_{\mathcal{F}'}$ is induced by decompositions $(\mathcal{B}_k'^+, \mathcal{B}_k'^-)$ of $\beta_{\mathcal{F}',k}$ for all $k \in \mathbb{N}_0$. Moreover, a collection of bijections of multisets $h_k : \mathcal{B}_k^+ \cup \mathcal{B}_k'^- \to \mathcal{B}_k^- \cup \mathcal{B}_k'^+$ for all $k \in \mathbb{N}_0$ induces a bijection of multisets $h : \mathcal{B}^+ \cup \mathcal{B}'^- \to \mathcal{B}^- \cup \mathcal{B}'^+$ with $\mathrm{cost}(h) = \sum_{k\in\mathbb{N}_0} \mathrm{cost}(h_k)$. Taking the infimum over all bijections h_k yields the result by Remark 8.3. □

Connection with 1-Wasserstein distance between barcodes. Our first motivation to consider the signed 1-Wasserstein distance on Euler characteristic profiles is that when $m = 1$, it is upper-bounded by the 1-Wasserstein distance between the persistent homology of the filtrations. Introduced by Cohen-Steiner, Edelsbrunner, Harer, and Mileyko (2010), Wasserstein distances between persistence barcodes are inspired by optimal transport theory. We recall their definition below.

Let $\mathcal{F}$ and $\mathcal{F}'$ be two finitely presentable one-parameter filtrations. Denote by $\mathcal{B}$ and $\mathcal{B}'$ the respective barcodes of $H_p(\mathcal{F})$ and $H_p(\mathcal{F}')$. A partial matching between $\mathcal{B}$ and $\mathcal{B}$ is a subset M of $\mathcal{B} \times \mathcal{B}'$ such that (i) for any $\beta \in \mathcal{B}$, there is at most one $\beta' \in \mathcal{B}'$ such that $(\beta, \beta') \in M$ and (ii) for any $\beta' \in \mathcal{B}'$, there is at most one $\beta \in \mathcal{B}$ such that $(\beta, \beta') \in M$. The *cost* of a matched pair $(\beta, \beta') \in M$ with $\beta = [a, b)$ and $\beta' = [a', b')$ is $c(\beta, \beta') = \max(|b'-b|, |a'-a|)$. The *cost* of an unmatched bar $\gamma \in \mathcal{B}\cup\mathcal{B}'$ with $\gamma = [c, d)$ is $c(\gamma) = |d - c|/2$. The *p-cost* of a matching is then defined as:

$$\mathrm{cost}_p(M) = \left(\sum_{(\beta,\beta')\in M} c(\beta, \beta')^p + \sum_{\substack{\gamma\in\mathcal{B}\cup\mathcal{B}' \\ \text{unmatched}}} c(\gamma)^p \right)^{1/p}.$$

The *p-Wasserstein distance* between $H_p(\mathcal{F})$ and $H_p(\mathcal{F}')$ is defined as:

$$d_p\big(H_p(\mathcal{F}), H_p(\mathcal{F}')\big) = \inf_M \mathrm{cost}_p(M)$$

where the infimum is taken over all partial matchings M between $\mathcal{B}$ and $\mathcal{B}'$.

We can now prove the link between signed 1-Wasserstein distance between Euler characteristic profiles and the 1-Wasserstein distance between persistence barcodes. This fact is easy but not proven in Oudot and Scoccola (2021).

Lemma 8.5. *Let $\mathcal{F}$ and $\mathcal{F}'$ be two finitely presentable one-parameter filtrations of simplicial complexes $\mathcal{K}$ and $\mathcal{K}'$, respectively. Then,*

$$\widehat{d}_1(\chi_{\mathcal{F}}, \chi_{\mathcal{F}'}) \;\leq\; 2 \sum_{p \geq 0} d_1\big(H_p(\mathcal{F}), H_p(\mathcal{F}')\big).$$

Proof. The barcode $\mathcal{B}_p\mathcal{F}$ of the p-th persistent homology $H_p(\mathcal{F})$ is a multiset of bars $\big[a_i^p, b_i^p\big)$ for $i \in \{1, \ldots, n_p\}$ and real numbers $-\infty < a_i^p < b_i^p \leq \infty$. This barcode induces a decomposition $(\mathcal{A}_p, \mathcal{B}_p) = (\{a_i^p\}_i, \{b_i^p\}_i)$ of $\beta_{\mathcal{F},p}$. Similarly, the barcode $\mathcal{B}_p\mathcal{F}'$ of $H_p(\mathcal{F}')$ is a multiset of bars $\big[a_j'^p, b_j'^p\big)$ for $j \in \{1, \ldots, n_p'\}$ that induces a decompositon $(\mathcal{A}_p', \mathcal{B}_p') = (\{a_j'^p\}_j, \{b_j'^p\}_j)$ of $\beta_{\mathcal{F}',p}$. Moreover, a partial matching M between $\mathcal{B}_p\mathcal{F}$ and $\mathcal{B}_p\mathcal{F}'$ induces a bijection of multisets $h : \mathcal{A}_p' \cup \mathcal{B}_p \to \mathcal{A}_p \cup \mathcal{B}_p'$ defined by $h(a') = a$ and $h(b) = b'$ when $([a,b),[a',b')) \in M$, by $h(b) = a$ when $[a,b)$ is unmatched and by $h(a') = b'$ when $[a',b')$ is unmatched. Moreover, the cost of the matching M and the cost of the bijection h satisfy $\mathrm{cost}(h) \leq 2\,\mathrm{cost}(M)$. Taking the infimum over all partial matching M, one has $\widehat{d}_1\big(\beta_{\mathcal{F},p}, \beta_{\mathcal{F}',p}\big) \leq 2\, d_1(H_p(\mathcal{F}), H_p(\mathcal{F}'))$. Lemma 8.4 yields the result. □

Connection with 1-presentation distance between persistence modules. Our second motivation to consider the signed 1-Wasserstein distance on Euler characteristic profiles is that in the two-parameter setting, the signed 1-Wasserstein distance between signed barcodes is upper-bounded by the 1-*presentation distance* d_I^1 between two-parameter persistence modules; see Oudot and Scoccola (ibid., Theorem 5). The presentation distance is introduced in Bjerkevik and Lesnick (2021) and recalled in Section 6.2 of Oudot and Scoccola (2021). We do not recall it here as it goes beyond the scope of this section. Yet, we mention that combined with Lemma 8.4, the result (ibid., Theorem 5) has the following direct consequence:

Lemma 8.6. *Let $\mathcal{F}$ and $\mathcal{F}'$ be two finitely presentable two-parameter filtrations of simplicial complexes $\mathcal{K}$ and $\mathcal{K}'$, respectively. Then,*

$$\widehat{d}_1(\chi_{\mathcal{F}}, \chi_{\mathcal{F}'}) \leq \sum_{k \geq 0} d_I^1\big(H_k\mathcal{F}, H_k\mathcal{F}'\big).$$

Remark 8.7. Such a result is unknown for m-parameter filtrations with $m \geq 3$; see Oudot and Scoccola (ibid., Question 27).

Sublevel-sets persistence. One can state stability results with respect to L_1 distance on filtration functions in the context of sublevel-sets persistence, as stated by the lemma below. Let $\mathcal{K}$ be a finite simplicial complex, and $f : \mathcal{K} \to \mathbb{R}^m$ a non-decreasing map. We define the 1-*norm* of f as $\|f\|_1 = \sum_{\sigma \in \mathcal{K}} \|f(\sigma)\|_1$.

Lemma 8.8. *Let $\mathcal{K}$ be a finite simplicial complex and $f, g : \mathcal{K} \to \mathbb{R}^m$ be non-decreasing maps. We have that*

$$\widehat{d}_1(\chi_f, \chi_g) \leq \|f - g\|_1.$$

Proof. The couple $\mathcal{C}_f = \big(\{f(\sigma)\}_{\dim\sigma \text{ even}}, \{f(\sigma)\}_{\dim\sigma \text{ odd}}\big)$ is a decomposition of χ_f. There is a similar decomposition $\mathcal{C}_g$ of χ_g. Moreover, the mapping $f(\sigma) \mapsto g(\sigma)$ induces a bijection of multisets $h : \mathcal{C}_f \to \mathcal{C}_g$ with cost $\sum_{\sigma\in\mathcal{K}} \|f(\sigma) - g(\sigma)\|_1 = \|g - f\|_1$. The result follows from Remark 8.3. □

Remark 8.9. Note that using Corollary 25 of Oudot and Scoccola (ibid.) on Betti functions combined with Lemma 8.4 would yield a multiplicative constant of 4 on the left-hand side of the inequality.

The above lemma clarifies the robustness of Euler characteristic profiles with respect to perturbations of filtrations defined on a fixed simplicial complex. This includes, for instance, density estimators on point clouds or Ricci curvature and HKS functions on graphs. The fact that these descriptors are controlled by the L_1 distance and not the L_∞ distance between the functions is an indicator of their sensitivity to the underlying geometry. Persistent images (Adams et al., 2017) share this property, while persistence landscapes (Bubenik et al., 2015; Vipond, 2020) do not, as they are controlled by the L_∞ distance between functions.

8.2.2 Stability results

We start by proving our stability result on Euler characteristic profiles. Recall from Chapter 5 that we denote by $L^1(I)$ the space of absolutely integrable complex-valued functions on an interval I of $\mathbb{R}$. Let $p \in [1,\infty]$ and let $f : \mathbb{R}^m \to \mathbb{C}$ be locally p-integrable. We denote by $\|f\|_{p,M}$ the p-norm of $f \cdot \mathbf{1}_{[-M,M]^m}$. If f is p-integrable, we denote its p-norm by $\|f\|_p$.

Proposition 8.10. *Let $\mathcal{F}$ and $\mathcal{F}'$ be two finitely presentable m-parameter filtrations of simplicial complexes $\mathcal{K}$ and $\mathcal{K}'$ respectively. For any $M > 0$, we have that*

$$\|\chi_{\mathcal{F}} - \chi_{\mathcal{F}'}\|_{1,M} \leq (2M)^{m-1}\, \widehat{d}_1(\chi_{\mathcal{F}}, \chi_{\mathcal{F}'})\,.$$

In particular, if $m = 1$:

$$\|\chi_{\mathcal{F}} - \chi_{\mathcal{F}'}\|_1 \leq \widehat{d}_1(\chi_{\mathcal{F}}, \chi_{\mathcal{F}'})\,.$$

Proof. Consider decompositions $(\mathcal{B}^+, \mathcal{B}^-)$ and $(\mathcal{B}'^+, \mathcal{B}'^-)$ of $\chi_{\mathcal{F}}$ and $\chi_{\mathcal{F}'}$ respectively. Assume there is a bijection $h : \mathcal{B}^+ \cup \mathcal{B}'^- \to \mathcal{B}^- \cup \mathcal{B}'^+$. If no such bijection exists, then $\widehat{d}_1(\chi_{\mathcal{F}}, \chi_{\mathcal{F}'})$ is infinite, and the inequality trivially holds. One has:

$$\chi_{\mathcal{F}} - \chi_{\mathcal{F}'} = \sum_{u\in\mathcal{B}^+\cup\mathcal{B}'^-} \mathbf{1}_{Q_u} - \sum_{v\in\mathcal{B}^-\cup\mathcal{B}'^+} \mathbf{1}_{Q_v} = \sum_{u\in\mathcal{B}^+\cup\mathcal{B}'^-} \mathbf{1}_{Q_u} - \mathbf{1}_{Q_{h(u)}}.$$

Therefore,

$$\|\chi_{\mathcal{F}} - \chi_{\mathcal{F}'}\|_{1,M} \leq \sum_{u\in\mathcal{B}^+\cup\mathcal{B}'^-} \|\mathbf{1}_{Q_u} - \mathbf{1}_{Q_{h(u)}}\|_1. \tag{8.9}$$

By an elementary induction on $m \geq 1$, we can prove that for all $u, v \in \mathbb{R}^m$,

$$\|\mathbf{1}_{Q_u} - \mathbf{1}_{Q_v}\|_{1,M} \leq (2M)^{m-1}\|u - v\|_1.$$

This concludes the proof of the first inequality.

Now, assume that $m = 1$. The existence of h ensures that $\|\chi_{\mathcal{F}} - \chi_{\mathcal{F}'}\|_1$ is finite and the second inequality follows from (8.9) and the fact that $\|\mathbf{1}_{[u,v)}\|_1 = |u - v|$. □

Combined with Lemma 8.8, the above proposition ensures that in the context of sublevel-sets persistence of a fixed simplicial complex, the L_1 distance between Euler profiles is upper-bounded by the L_1 distance between filtration functions. A similar result is proven Dłotko and Gurnari (2022, Proposition 3.4) in terms of L_∞ distance between the filtration functions multiplied by the number of simplices in the complex $\mathcal{K}$. In addition, this result can be combined with Lemma 8.5 when $m = 1$ to obtain known results stating that L_1 distances between Euler profiles is upper-bounded by the sum of 1-Wasserstein distances between persistent homologies in all degrees; see Dłotko and Gurnari (ibid., Proposition 3.2), Curry, Mukherjee, and Turner (2022, Lemma 4.10) and Perez (2022, Corollary 3.8). Here, the result is more generally phrased in terms of signed 1-Wasserstein distances between Euler profiles, to benefit from the other results of the previous section. In particular, in the two-parameter setting, the L_1 distance between Euler profiles is upper-bounded by the 1-presentation distance between the persistent homology of the filtrations.

These stability results carry over to our other descriptors, as stated in the Corollary 8.11 below. Let K be a compact subset of $\mathbb{R}_+^{m*}$. For $q \in [1, \infty]$, we consider the norms on functions $\mathcal{R} : \mathbb{R}_+^{m*} \to \mathbb{R}$ defined by:

$$\|\mathcal{R}\|_{L_K^{q,1}} = \begin{cases} \left(\int_K \left(\int_{\mathbb{R}} |\mathcal{R}(\xi, s)| \, \mathrm{d}s \right)^q \mathrm{d}\xi \right)^{1/q} & \text{for } q \in [1, \infty), \\ \sup_{\xi \in K} \int_{\mathbb{R}} |\mathcal{R}(\xi, s)| \, \mathrm{d}s & \text{for } q = \infty. \end{cases} \tag{8.10}$$

Corollary 8.11. *Let K be a compact subset of $\mathbb{R}^{m*}$ and $q \in [1, \infty]$. Let $\mathcal{F}$ and $\mathcal{F}'$ be finitely presentable m-parameter filtrations of simplicial complexes $\mathcal{K}$ and $\mathcal{K}'$ respectively. Let $\kappa \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$. There exists a constant $C_{K,q}$ depending only on K and q such that:*

$$\left\| \mathcal{R}_{\mathcal{F}} - \mathcal{R}_{\mathcal{F}'} \right\|_{L_K^{q,1}} \leq C_{K,q} \, \widehat{d}_1(\chi_{\mathcal{F}}, \chi_{\mathcal{F}'}) \, ,$$

$$\left\| \mathrm{T}_\kappa \left[\chi_{\mathcal{F}} \right] - \mathrm{T}_\kappa \left[\chi_{\mathcal{F}'} \right] \right\|_{L_K^q} \leq C_{K,q} \, \|\kappa\|_\infty \, \widehat{d}_1(\chi_{\mathcal{F}}, \chi_{\mathcal{F}'}) \, .$$

Proof. Let us prove the first inequality. Proposition 8.10 with $m = 1$ ensures that for any $\xi \in K$,

$$\int_{\mathbb{R}} \left| \mathcal{R}_{\mathcal{F}}(\xi, s) - \mathcal{R}_{\mathcal{F}'}(\xi, s) \right| \mathrm{d}s = \|\xi_* \chi_{\mathcal{F}} - \xi_* \chi_{\mathcal{F}'}\|_1 \leq \widehat{d}_1(\xi_* \chi_{\mathcal{F}}, \xi_* \chi_{\mathcal{F}'}) \, .$$

To prove the desired inequality, we will prove that for any $\xi \in \mathbb{R}_+^{m*}$, one has:

$$\widehat{d}_1(\xi_* \chi_{\mathcal{F}}, \xi_* \chi_{\mathcal{F}'}) \leq \|\xi\|_\infty \, \widehat{d}_1(\chi_{\mathcal{F}}, \chi_{\mathcal{F}'}) \, .$$

The result then follows from computing the q-norm on both sides. Consider decompositions $(\mathcal{B}^+, \mathcal{B}^-)$ and $(\mathcal{B}'^+, \mathcal{B}'^-)$ of $\chi_{\mathcal{F}}$ and $\chi_{\mathcal{F}'}$ respectively. By Lemma 5.10, they induce decompositions $(\xi_* \mathcal{B}^+, \xi_* \mathcal{B}^-)$ and $(\xi_* \mathcal{B}'^+, \xi_* \mathcal{B}'^-)$ of $\xi_* \chi_{\mathcal{F}}$ and $\xi_* \chi_{\mathcal{F}'}$ respectively by the formula $\xi_* \mathcal{B}^\pm = \{ \langle \xi, u \rangle \, ; \, u \in \mathcal{B}^\pm \}$ and a similar one for $\mathcal{F}'$. Consider a bijection of multisets $h : \mathcal{B}^+ \cup \mathcal{B}'^- \to \mathcal{B}^- \cup \mathcal{B}'^+$. It induces a bijection of multisets $h_\xi : \xi_* \mathcal{B}^+ \cup \xi_* \mathcal{B}'^- \to \xi_* \mathcal{B}^- \cup \xi_* \mathcal{B}'^+$ defined by $\langle \xi, u \rangle \mapsto \langle \xi, h(u) \rangle$ with cost:

$$\mathrm{cost}(h_\xi) = \sum_{t \in \xi_* \mathcal{B}^+ \cup \xi_* \mathcal{B}'^-} \|t - h_\xi(t)\|_1 = \sum_{u \in \mathcal{B}^+ \cup \mathcal{B}'^-} \|\langle \xi, u \rangle - \langle \xi, h(u) \rangle\|_1 \leq \|\xi\|_\infty \, \mathrm{cost}(h).$$

Taking the infimum over all bijections h yields $\widehat{d}_1(\xi_*\chi_{\mathcal{F}}, \xi_*\chi_{\mathcal{F}'}) \leq \|\xi\|_\infty \widehat{d}_1(\chi_{\mathcal{F}}, \chi_{\mathcal{F}'})$ by Remark 8.3.

Let us now prove the second inequality. It follows from the definition of hybrid transforms that $\|\mathrm{T}_\kappa[\chi_{\mathcal{F}}] - \mathrm{T}_\kappa[\chi_{\mathcal{F}'}]\|_{L^q_K} \leq \|\kappa\|_\infty \|\mathcal{R}_{\mathcal{F}} - \mathcal{R}_{\mathcal{F}'}\|_{L^{q,1}_K}$ when κ is bounded. The first inequality yields the result. □

8.3 Statistical properties

In this section, we provide statistical guarantees for our descriptors computed on a random sample, as the sample size tends to infinity. For the sake of readability, we denote by $\psi^\kappa_{\mathcal{F}}$ the hybrid transform $\mathrm{T}_\kappa[\chi_{\mathcal{F}}]$.

8.3.1 Limit theorems for one-parameter hybrid transforms

This section is devoted to limit theorems for the hybrid transforms of the Čech complex of an i.i.d. sample in $\mathbb{R}^d$. Theorem 8.12 is a pointwise law of large numbers, while Theorem 8.14 states a functional central limit theorem for the hybrid transforms of compactly supported kernels.

Theorem 8.12. *Let $X_1, \ldots, X_n$ be an i.i.d. sample drawn according to an a.e. continuous bounded Lipschitz density g on $\mathbb{R}^d$. Consider a sequence $(r_n)_{n\in\mathbb{N}_0}$ such that $nr_n^d \to 0$ and $n^{k+2}r_n^{d(k+1)} \to \infty$ as $n \to \infty$. We denote by $\mathcal{F}_n$ the Čech filtration associated with the rescaled sample $\frac{1}{r_n}(X_i)_{i=1}^n$. Let $T, a > 0$ and $\kappa \in L^1(\mathbb{R})$. Further assume that κ is supported on $[0,T]$. Then there exist functions $A_0, \ldots, A_{d-1}$ on $\mathbb{R}_+^*$ that depend only on the kernel κ and such that for every $\xi > a$,*

$$\frac{1}{n^{k+2}r_n^{d(k+1)}} \cdot \psi^\kappa_{\mathcal{F}_n}(\xi) \underset{n\to\infty}{\longrightarrow} \sum_{k=0}^{d-1} \frac{(-1)^k}{(k+2)!} \cdot A_k(\xi) \cdot \int_{\mathbb{R}^d} g^{k+2}(x)\,\mathrm{d}x \quad a.s..$$

This theorem is a direct consequence of the limit theorems derived in Owada (2022). It is a key assumption that we are in the so-called *sparse regime*, that is, $nr_n^d \to 0$. In order to make this law of large numbers more comprehensible, we make a further assumption that we are in the so-called *divergence regime*, that is $n^{k+2}r_n^{d(k+1)} \to \infty$. The sequence defined by $r_n = n^{-\alpha}$ for $\frac{1}{d} < \alpha < \frac{1}{d} + \frac{1}{d^2}$ verifies these two assumptions. Similar results can be derived for other subcases of the sparse regime, i.e., the Poisson regime $n^{k+2}r_n^{d(k+1)} \to c > 0$ and the vanishing regime $n^{k+2}r_n^{d(k+1)} \to 0$.

Proof of Theorem 8.12. Let $X_1, \ldots, X_n$ be an i.i.d. sample drawn according to an a.e. continuous bounded Lipschitz density g on $\mathbb{R}^d$. Consider a sequence $(r_n)_{n\in\mathbb{N}_0}$ such that $nr_n^d \to 0$ and $n^{k+2}r_n^{d(k+1)} \to \infty$ as $n \to \infty$.

Let us define $\Delta := \{(x,y) : 0 \leq x \leq y < \infty\} \cup \{(x,\infty) : 0 \leq x < \infty\}$ and for every (s,t,u,v) such that $0 \leq s \leq t \leq u \leq v \leq \infty$, denote by $R_{s,t,u,v}$ the rectangle $(s,t] \times (u,v]$ of Δ. Recall that a finite barcode $\mathcal{B} = \{[a_i, b_i)\}_{i=1,\ldots,l}$ can be turned into a discrete measure $\mu = \sum_{i=1}^l \delta_{(a_i,b_i)}$ on Δ. Denote by $\mu_{k,n}$ the barcode of the k-th persistent homology of the Čech filtration of $1/r_n(X_i)_{i=1}^n$, seen as a discrete measure on Δ.

Theorem 3.2 of Owada (2022) ensures that for every $k \in [\![0, d-1]\!]$ there exists a unique Radon measure μ_k on Δ such that we have the following vague convergence:

$$\frac{1}{n^{k+2} r_n^{d(k+1)}} \mu_{k,n} \xrightarrow[n\to\infty]{v} \frac{1}{(k+2)!} \left(\int_{\mathbb{R}^d} g^{k+2}(x)\, \mathrm{d}x \right) \mu_k \quad \text{a.s.}, \tag{8.11}$$

where for every $0 \leq s \leq t \leq u \leq v \leq \infty$, there is an indicator geometric function $H_{s,t,u,v}$ on $\mathbb{R}^{d(k+2)}$ defined in Owada (ibid., Section 3.1), which does not depend on g and such that the measure μ_k is defined by:

$$\mu_k(R_{s,t,u,v}) = \int_{\mathbb{R}^{d(k+1)}} H_{s,t,u,v}(0, y_1, \ldots, y_{k+1})\, \mathrm{d}y_1 \ldots \mathrm{d}y_{k+1}.$$

Recall from Remark 8.2 that the primitive kernel $\overline{\kappa}$ is such that $\overline{\kappa}(x) \to 0$ when $x \to +\infty$. Therefore, the fact that κ is supported on $[0, T]$ implies that the primitive $\overline{\kappa}$ is also supported on $[0, T]$. For $\xi > a$, denote by $h_\xi : (x, y) \in \Delta \mapsto \overline{\kappa}(\xi y) - \overline{\kappa}(\xi x)$. According to (8.5), one has:

$$\psi^{\kappa}_{\mathcal{F}_n}(\xi) = \sum_{k=0}^{d-1} (-1)^k \int_\Delta h_\xi \, \mathrm{d}\mu_{k,n}.$$

Since h_ξ is continuous and supported on $[0, T/a]^2$, we have by the vague convergence in (8.11) that:

$$\frac{1}{n^{k+2} r_n^{d(k+1)}} \psi^{\kappa}_{\mathcal{F}_n}(\xi) \xrightarrow[n\to\infty]{} \sum_{k=0}^{d-1} \frac{(-1)^k}{(k+2)!} \left(\int_{\mathbb{R}^d} g^{k+2}(x) \mathrm{d}x \right) A_k(\xi) \quad \text{a.s.},$$

where $A_k(\xi) = \int_\Delta h_\xi \mathrm{d}\mu_k$. □

Theorem 8.12 shows that the pointwise limit of the hybrid transform depends on the sampling only through the quantities $\int_{\mathbb{R}^d} g^{k+2}$ for $k \in \{0, \ldots, d-1\}$ and they can therefore discriminate between different samplings as soon as n is large enough. In addition to this law of large numbers, a finer limit result for the Euler characteristic curve is proven in Krebs, Roycraft, and Polonik (2021), which we recall hereafter for the sake of completeness.

First, recall that a function h on $\mathbb{R}^m$ is *blocked* if it can be written $h = \sum_{i=1}^{m^d} b_i \mathbf{1}_{A_i}$ where $b_1, \ldots, b_{m^d}$ are non-negative real numbers and the A_i are axis-aligned rectangles in $\mathbb{R}^m$. Moreover, recall that the *Skorohod J_1-topology* on the space of càdlàg functions $D([0, T])$ is the topology induced by the metric:

$$d_{J_1}(f, f') := \inf_\lambda \left\{ \|f \circ \lambda - f'\|_\infty + \|\lambda - \mathrm{Id}_{[0,T]}\|_\infty \right\},$$

where the infimum is taken over all increasing continuous bijections λ of $[0, T]$.

Theorem 8.13 (ibid., Theorem 3.4)**.** *Let $T > 0$ and $X_1, \ldots, X_n$ be sampled according to a bounded density g on $[0, 1]^d$. Denote by $\mathcal{F}_n$ the Čech complex associated with the point cloud $n^{1/d}(X_i)_{i=1}^n$ and $\overline{\chi}_{\mathcal{F}_n} : t \in [0, T] \mapsto \sqrt{n}\big(\chi_{\mathcal{F}_n}(t) - \mathbb{E}[\chi_{\mathcal{F}_n}(t)]\big)$. Assume that blocked functions can uniformly approximate g. There is a Gaussian process $\mathfrak{G} : [0, T] \to \mathbb{R}$ such that:*

$$\overline{\chi}_{\mathcal{F}_n} \xrightarrow[n\to\infty]{} \mathfrak{G},$$

in distribution in the Skorohod J_1-topology on $D([0,T])$. Furthermore, there exist two real-valued functions $\gamma : [0,T]^2 \to \mathbb{R}$ and $\alpha : [0,T] \to \mathbb{R}$ such that the covariance of the limiting process is defined by:

$$\mathbb{E}[\mathfrak{G}(s)\mathfrak{G}(t)] = \mathbb{E}\left[\gamma\left(g(Z)^{1/d}(s,t)\right)\right] - \mathbb{E}\left[\alpha\left(g(Z)^{1/d}s\right)\right]\mathbb{E}\left[\alpha\left(g(Z)^{1/d}t\right)\right],$$

where Z is a random variable with density g.

We refer to Krebs, Roycraft, and Polonik (ibid.) for the expression of the two functions γ and α. Here again, the distribution of the points appears in the limiting object and, more precisely, in its covariance function. We can adapt this theorem to show that hybrid transforms of compactly supported kernels are also asymptotically normal.

Theorem 8.14. *Consider the setting of Theorem 8.13. Let $a, M > 0$ and $\kappa \in L^1(\mathbb{R})$. Further assume that κ is supported on $[0,T]$. Denote by $\overline{\psi}^{\kappa}_{\mathcal{F}_n} : \xi \in [a,M] \mapsto \sqrt{n}(\psi^{\kappa}_{\mathcal{F}_n}(\xi) - \mathbb{E}\left[\psi^{\kappa}_{\mathcal{F}_n}(\xi)\right])$. Then, there is a Gaussian process $\hat{\mathfrak{G}} : [a,M] \to \mathbb{R}$ such that:*

$$\overline{\psi}^{\kappa}_{\mathcal{F}_n} \underset{n\to\infty}{\longrightarrow} \hat{\mathfrak{G}} \quad a.s.,$$

in $\left(\mathcal{C}^0[a,M], \|\cdot\|_\infty\right)$. Furthermore, the covariance of the limiting process is defined by:

$$\mathbb{E}\left[\hat{\mathfrak{G}}(\xi_1)\hat{\mathfrak{G}}(\xi_2)\right] = \xi_1\xi_2 \int_0^{T/\xi_1}\int_0^{T/\xi_2} \kappa(\xi_1 t)\,\kappa(\xi_2 s)\,\operatorname{cov}\left(\mathfrak{G}(s),\mathfrak{G}(t)\right)\mathrm{d}s\,\mathrm{d}t,$$

where $\mathfrak{G}$ is the Gaussian process defined in Theorem 8.13.

Proof of Theorem 8.14. Let $T > 0$ such that κ is supported in $[0,T]$. Let $a, M > 0$. According to Lemma 6.7, one has for any $\xi \in [a,M]$:

$$\psi^{\kappa}_{\mathcal{F}}(\xi) = \xi \int_0^{T/\xi} \kappa(\xi\cdot u)\chi_{\mathcal{F}}(u)\mathrm{d}u,$$

and similarly for $\chi_{\mathcal{F}_n}$. For any $\xi \in [a,M]$, linearity of the mapping $\chi \mapsto \xi\int_0^{T/\xi}\kappa(\xi\cdot u)\chi(u)\mathrm{d}u$ ensures that:

$$\sqrt{n}\left(\psi^{\kappa}_{\mathcal{F}_n}(\xi) - \mathbb{E}\left[\psi^{\kappa}_{\mathcal{F}_n}(\xi)\right]\right) \;=\; \xi\int_0^{T/\xi}\kappa(\xi\cdot u)\left[\sqrt{n}\big(\chi_{\mathcal{F}_n}(u) - \mathbb{E}[\chi_{\mathcal{F}_n}(u)]\big)\right]\mathrm{d}u.$$

Moreover, the functions $\psi^{\kappa}_{\mathcal{F}}$ and $\psi^{\kappa}_{\mathcal{F}_n}$ are continuous on $[a,M]$ since κ is in $L^1(\mathbb{R})$. Therefore, denoting by Ψ the mapping from the space of càdlàg functions $D([0,T])$ with Skorohod J_1-topology to $(\mathcal{C}^0([a,M]), \|\cdot\|_\infty)$ defined by:

$$\Psi : \chi \longmapsto \left(\xi \mapsto \xi\int_0^{T/\xi}\kappa(\xi\cdot u)\chi(u)\mathrm{d}u\right),$$

one has $\sqrt{n}\left(\psi^{\kappa}_{\mathcal{F}_n} - \mathbb{E}\left[\psi^{\kappa}_{\mathcal{F}_n}\right]\right) \;=\; \Psi\left(\sqrt{n}\left(\chi_{\mathcal{F}_n} - \mathbb{E}[\chi_{\mathcal{F}_n}]\right)\right)$. Moreover, it is easy to check that for any two χ_1 and χ_2 in $D([0,T])$, one has:

$$\|\Psi(\chi_1) - \Psi(\chi_2)\|_\infty \le \frac{M}{a}\,\|\chi_1 - \chi_2\|_\infty \int_0^T |\kappa(t)|\mathrm{d}t,$$

so that the mapping Ψ is Lipschitz and, therefore, continuous.

Now, according to Theorem 8.13, there is a Gaussian process $\mathfrak{G} : [0, T/a] \to \mathbb{R}$ such that denoting $\overline{\chi_{\mathcal{F}_n}} : u \in [0, T/a] \mapsto \sqrt{n}\left(\chi_{\mathcal{F}_n}(u) - \mathbb{E}[\chi_{\mathcal{F}_n}(u)]\right)$ we have:

$$\overline{\chi_{\mathcal{F}_n}} \underset{n\to\infty}{\longrightarrow} \mathfrak{G},$$

in distribution in the Skorohod J_1-topology on $D([0, T/a])$. Thus, continuity of Ψ and the continuous mapping theorem yield that almost surely, one has the following convergence in $(\mathcal{C}^0([a, M]), \|\cdot\|_\infty)$,

$$\sqrt{n}\left(\psi^\kappa_{\mathcal{F}_n} - \mathbb{E}\left[\psi^\kappa_{\mathcal{F}_n}\right]\right) \underset{n\to\infty}{\longrightarrow} \hat{\mathfrak{G}},$$

where $\hat{\mathfrak{G}} : \xi \in [a, M] \mapsto \xi \int_0^{T/\xi} \kappa(\xi \cdot u)\mathfrak{G}(u)\,\mathrm{d}u$. The covariance of the limiting process $\hat{\mathfrak{G}}$ follows immediately from that of $\mathfrak{G}$. □

8.3.2 Limit theorem for multi-parameter hybrid transforms

In this section, we derive a law of large numbers for the hybrid transform in the multi-parameter case. We adopt the sampling model of Hiraoka, Shirai, and Trinh (2018) and refer to Section 3 of loc. cit. for the definition of a stationary ergodic point process. Consider a point process Φ on $\mathbb{R}^d$ and its restriction Φ_L to $[-L/2, L/2]^d$. Let $\mathcal{S}(\mathbb{R}^d)$ be the set of all finite (non-empty) subsets in $\mathbb{R}^d$, to be thought of as the set of all simplices. Let $f = (f_1, \ldots, f_m) : \mathcal{S}(\mathbb{R}^d) \to [0, \infty]^m$ be a measurable function, non-decreasing with respect to inclusion. According to Example 4.1, the map f induces a filtration on every simplicial complex of $\mathbb{R}^d$. In particular, the map f induces a filtration on the simplicial complex 2^{Φ_L} that we denote by $\mathcal{F}_L$. We prove the following:

Theorem 8.15. *Assume that Φ is a stationary ergodic point process having finite moments. Let $T, a > 0$ and $\kappa \in L^1(\mathbb{R})$. Assume that κ is supported on $[0, T]$. Assume that there is an increasing function $\rho : \mathbb{R} \to \mathbb{R}$ and $i \in [\![1, m]\!]$ such that for all $(x, y) \in (\mathbb{R}^d)^2$,*

$$\|x - y\| \leq \rho\left(f_i(\{x, y\})\right). \tag{8.12}$$

Under these assumptions, there exists a function $H : \mathbb{R}_+^{m*} \times \mathbb{R} \to \mathbb{C}$ that depends only on κ and f such that for all $\xi = (\xi_1, \ldots, \xi_m) \in \mathbb{R}_+^{m*}$ and $\lambda > a$, one has

$$\frac{1}{L^d}\psi^\kappa_{\mathcal{F}_L}(\lambda\xi) \underset{L\to\infty}{\longrightarrow} H(\xi, \lambda) \quad \text{a.s..}$$

This limit theorem is a direct consequence of the results from Hiraoka, Shirai, and Trinh (ibid.). Note that this encompasses most cases of usual point processes such as Poisson, Ginibre, or Gibbs. This result makes use of the smoothness properties of the hybrid transforms and follows directly from Corollary 7.10 that expresses multi-parameter hybrid transforms of sublevel-sets constructible functions as one-parameter hybrid transforms. Similar results cannot be derived that easily for Euler characteristic profiles, as one would need to consider the joint law of several one-parameter filtrations.

Proof of Theorem 8.15. Let $\xi = (\xi_1, \ldots, \xi_m) \in \mathbb{R}_+^{m*}$. Denote by $\mu_{k,L}^{\xi\circ f}$ the measure associated with the barcode of the k-th persistent homology of Φ_L for the filtration function $\xi \circ f = \sum_{i=1}^m \xi_i f_i$. In virtue of the hypotheses, there exists $i \in [\![1, m]\!]$ such that

for all $(x, y) \in (\mathbb{R}^d)^2$, $\|x - y\| \leq \rho\left(f_i(\{x, y\})\right)$. Let $\rho' : x \mapsto \rho(x/\xi_i)$. Therefore, as the filtration functions f_i are non-negative and ρ and ρ' are increasing, we have that:

$$\rho'\left(\sum_{j=1}^{m} \xi_j f_j(\{x, y\})\right) \geq \rho'(\xi_i f_i(\{x, y\}) \geq \rho(f_i(\{x, y\})) \geq \|x - y\|. \tag{8.13}$$

The filtration function $\langle \xi, f \rangle$ therefore verifies the assumptions of Theorem 1.5 of Hiraoka, Shirai, and Trinh (ibid.), which states that there exists a Radon measure ν_k such that almost surely, we have the vague convergence $\frac{1}{L^d}\mu_{k,L}^{\xi \circ f} \xrightarrow{v} \nu_k$ as $L \to \infty$. Note that in loc. cit., the authors make the additional hypothesis that the filtration function is translation invariant. However, this assumption is only needed to derive a central limit theorem on persistent Betti numbers but not required for the above law of large numbers, for which we only need (8.12) to hold. As in the proof of Theorem 8.12, we introduce a continuous function $h_\lambda : (x, y) \in \Delta \mapsto \overline{\kappa}(\lambda y) - \overline{\kappa}(\lambda x)$. This function is supported on $[0, T/a]^2$. According to (8.5) together with (8.1), we have that:

$$\psi_{\mathcal{F}_L}^{\kappa}(\lambda \xi) = \sum_{k=0}^{d-1} (-1)^k \int_{\Delta} h_\lambda \, \mathrm{d}\mu_{k,L}^{\xi \circ f}.$$

Hence the result, by the vague convergence $\frac{1}{L^d}\mu_{k,L}^{\xi \circ f} \xrightarrow{v} \nu_k$ for all $k \in [\![0, d-1]\!]$. □

Deriving a multi-dimensional central limit theorem from Penrose and Yukich (2001) would require the filter $\xi \circ f$ to verify some translation invariance property. In practice, this very strong assumption is verified only by Čech and *Vietoris-Rips* filtrations as well as marked processes; see Botnan and Hirsch (2022). Alpha and function-Čech filtrations that we used in our experiments do not verify this assumption.

As pointed out in Hiraoka, Shirai, and Trinh (2018, Example 1.3), Čech and Vietoris-Rips filtrations satisfy (8.12) for $\rho : t \mapsto 2t$. We provide below a broad family of multi-parameter filtrations satisfying (8.12). We begin with some simple examples.

Example 8.16. One can readily check that the function-Čech filtration (Section 4.1.1) and the *function-alpha filtrations* (see the beginning of Section 9.2) satisfy (8.12). The latter one is used on point clouds in Chapter 9.

Finally, we give another class of filtrations satisfying (8.12) that contains the distance-to-measure (DTM) filtrations (Anai et al., 2020).

Example 8.17. Let h be a positive and bounded function from $\mathbb{R}^d$ to $\mathbb{R}$. The weighted Čech complex introduced in Anai et al. (ibid.) is defined as follows. For every $x \in \mathbb{R}^d$ and real number $t \geq 0$, we define:

$$r_x(t) = \begin{cases} -\infty & \text{if } t < h(x), \\ t - h(x) & \text{otherwise.} \end{cases}$$

We denote by $\overline{B}_h(x, t) = \overline{B}\left(x, r_x(t)\right)$ the closed Euclidean ball with center x and radius $r_x(t)$. A simplex $\{x_0, \ldots, x_k\}$ in some finite set $\mathbb{X}$ belongs to the *weighted Čech complex* at scale $t \geq 0$ if the intersection of closed balls $\cap_{l=0}^{k} \overline{B}_h(x_l, t)$ is non-empty. Considering the weighted Čech complex for all scales t defines a filtration of $2^{\mathbb{X}}$ called *weighted Čech filtration*. The weighted Čech filtration satisfies (8.12) for $\rho : t \mapsto \max(\max h, 2t)$.

Chapter 9

Experimental study

Based on joint work with Olympio Hacquard.

As explained in the introduction of this thesis, extracting topological information from data sets using persistence barcodes follows a machinery that finds its origins in the works of Edelsbrunner, Letscher, and Zomorodian (2000). These diagrams are often turned into vectors to perform various learning tasks such as classification, clustering, or regression. This operation is done at the cost of losing some information since the space of persistence diagrams endowed with the bottleneck distance cannot be isometrically embedded into a Hilbert space (Bubenik and A. Wagner, 2020; Carrière and Bauer, 2019). Most commonly used techniques include persistence images (Adams et al., 2017), landscapes (Bubenik et al., 2015), and more recently measure-oriented vectorizations in Royer et al. (2021) and neural network methods from Carrière, Chazal, et al. (2020) and Reinauer, Caorsi, and Berkouk (2021). An overview of topological methods in machine learning has been presented in the survey of Hensel, Moor, and Rieck (2021). Although not intrinsically multi-parameter, several vectorization techniques have been generalized to cope with multi-parameter filtrations, such as persistence landscapes (Vipond, 2020) and persistence images (Carrière and Blumberg, 2020). Beside their high level of sophistication, the main limitation of these tools is their computational cost; see Carrière and Blumberg (ibid., Table 2).

In contrast, Euler characteristic profiles introduced in Chapter 8 do not compute homological information but rather compute the Euler characteristic of the topological spaces at hand. As shown by the previous chapters, these descriptors satisfy many appealing theoretical properties. In comparison to persistence barcodes, Euler profiles may appear too coarse to capture relevant information in data analysis contexts. However, it turns out not to be the case, as shown for instance in (Amézquita et al., 2022; Jiang, Kurtek, and Needham, 2020; Smith and Zavala, 2021). Moreover, computing these descriptors amounts to counting simplices appearing at each value of the parameter of the filtration. This translates into a reduced computational cost: the overall complexity is linear in the total number of simplices in a simplicial filtration, versus matrix multiplication time for persistence barcodes (Milosavljević, D. Morozov, and Skraba, 2011). In addition, the locality of the Euler characteristic has been exploited in many ways (Heiss and H. Wagner, 2017; Richardson and Werman, 2014; Wang, H. Wagner, and C. Chen, 2022) to design highly efficient algorithms to compute Euler characteristic curves.

Outline. In this chapter, we conduct an in-depth experimental study of Euler characteristic profiles and their hybrid transforms. We show that they are informative and highly efficient topological descriptors of the data at hand. First, we provide algorithms for computing them as well as a `Python` implementation (Section 9.1). In addition, we give heuristics on how to choose the kernel of hybrid transforms in practical contexts. Then, we illustrate on many examples the type of topological and geometric structures Euler curves and their integral transforms can capture from the data (Section 9.2). In a series of quantitative experiments, we demonstrate that Euler profiles achieve state-of-the-art accuracy in supervised classification and regression tasks when coupled with a random forest or a gradient boosting algorithm (Sections 9.3.1, 9.3.2 and 9.3.4) at a very low computational cost (Section 9.3.5). Moreover, the multi-parameter nature of our tools and their computational simplicity allows us to use up to 5-parameter filtrations to classify graph data. However, due to their simplicity, these descriptors do not manage to linearly separate the different classes or be competitive on unsupervised tasks. Inspired by signal analysis, we cope with these limitations by studying hybrid transforms of Euler characteristic curves and profiles. We demonstrate that these transforms act as highly efficient information compressors. As a consequence, they outperform Euler profiles in unsupervised classification tasks and in supervised tasks when plugging a linear classifier (Figure 9.8 and Sections 9.3.1 to 9.3.3). Along the way, we illustrate their ability to capture fine-grained information on a real-world data set in Section 9.3.3.

Convention. We follow the conventions and notations introduced in Chapter 8.

9.1 Algorithms

In this section, we begin by describing the algorithms used to compute our descriptors as well as their implementation. We also give some intuition on how to choose the kernel of hybrid transforms.

Algorithm. In all our experiments, and hence in our implementation, we restrict ourselves to one-critical filtrations. In that case, formulae (8.6) and (8.8) can readily be turned into algorithms computing Euler characteristic profiles and their hybrid transforms. Each algorithm takes as input a grid of size $d_1 \times \ldots \times d_m$ on which the Euler characteristic profile or the hybrid transform is evaluated. For the Radon transform, we use the fact that $\mathcal{R}_{\mathcal{F}}(\xi, s) = \mathcal{R}_{\mathcal{F}}(\xi/s, 1)$. Our algorithm then takes as input a grid of size $d_1 \times \ldots \times d_m$ on which the map $\eta \in \mathbb{R}_+^{m*} \mapsto \mathcal{R}_{\mathcal{F}}(\eta, 1)$ is evaluated. In any case, the output array of size $d_1 \times \ldots \times d_m$ is an exact sampling of the descriptor. Therefore, our topological descriptors vectorize m-parameter filtrations into $d_1 \times \ldots \times d_m$ arrays that can be used as input to any classical machine learning algorithm.

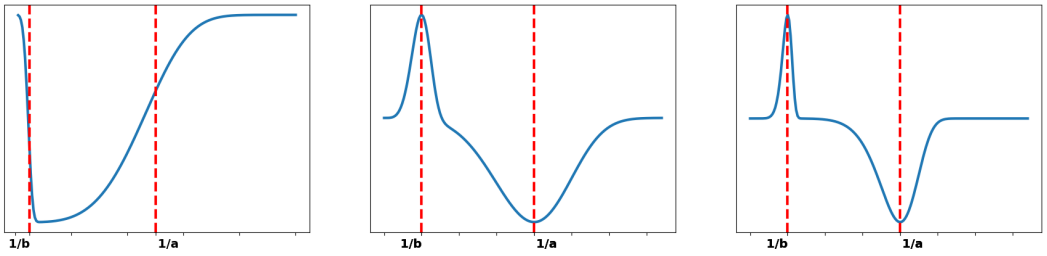
Complexity. The algorithm computing Euler characteristic profiles with resolution $d_1 \times \ldots \times d_m$ has time complexity $\mathcal{O}(|K| + d_1 \cdot \ldots \cdot d_m)$ in the worst case. The algorithm computing Radon and hybrid transforms with the same resolution has a worst-case time complexity of $\mathcal{O}(|K| \cdot d_1 \cdot \ldots \cdot d_m)$. In comparison, recall that computing a persistence

barcode has time complexity $\mathcal{O}(|K|^\omega)$ in the worst case where $2 \leq \omega < 2.373$ is the exponent for matrix multiplication (Milosavljević, D. Morozov, and Skraba, 2011).

Implementation. A `Python` implementation of our algorithms is freely available online on our GitHub repository: `https://github.com/vadimlebovici/eulearning`. In practice, our implementation allows for several ways of choosing a grid of sampling. The first method is to provide bounds $[(a_1, b_1), \ldots, (a_m, b_m)]$ and a resolution $d_1 \times \ldots \times d_m$. We then compute a sampling of our descriptors on a uniform discretization of the subset $[a_1, b_1] \times \ldots \times [a_m, b_m] \subseteq \mathbb{R}^m$. This method has the disadvantage of requiring prior knowledge about the data.

For Euler characteristic profiles, the second way is to provide a list $[(p_1, q_1), \ldots, (p_m, q_m)]$ of real numbers $0 \leq p_i < q_i \leq 1$. The algorithm then computes the p_i-th and the q_i-th percentiles of the i-th filtration for each $i \in \{1, \ldots, m\}$. Finally, Euler profiles are uniformly sampled on a $d_1 \times \ldots \times d_m$ grid ranging from the lowest to the highest percentile on each axis. For the Radon and hybrid transforms, the second way consists in providing a list $[p_1, \ldots, p_m]$ of real numbers $0 \leq p_i \leq 1$ and a positive real number α. The algorithm then computes the p_i-th percentiles v_i of the i-th filtration for each $i \in \{1, \ldots, m\}$. The integral transforms are uniformly sampled on a $d_1 \times \ldots \times d_m$ grid ranging from 0 to α / v_i on each axis. Note that filtrations have to be positive, which is always satisfied up to translation. This method does not require any prior knowledge of the data but depends on a choice of parameters. More importantly, doing as such is justified for primitive kernels of type $\overline{\kappa} : s \mapsto \exp(-x^p)$ and $\overline{\kappa} : s \mapsto x^p \exp(-x^p)$ by the paragraph *Kernel choice* below.

Kernel choice. To interpret integral transforms of Euler curves, we set $m = 1$ and compute them on the rectangular function $\chi_{\mathcal{F}} = \mathbf{1}_{[a,b)}$ associated with a persistence barcode with a single bar $[a, b)$ with $a < b \in (0, +\infty)$. Recall that the hybrid transform has the simple expression (8.5). Figure 9.1 shows the hybrid transforms for several kernels. For every $p > 0$, the hybrid transform with primitive kernel $\overline{\kappa} : s \mapsto -\exp(-s^p)$ has a minimum in $\sqrt[p]{\frac{p(\log(b) - \log(a))}{b^p - a^p}}$, which tends to $1/b$ as $p \to \infty$. As a consequence, transforms of this type yield *smoothed* versions of the curve $t \mapsto \chi_{\mathcal{F}}(1/t)$, that is, of an Euler curve with *inverted scales*. Similarly, the hybrid transform with primitive kernel $\overline{\kappa} : s \mapsto -s^p \exp(-s^p)$ has a minimum that tends to $1/a$ and a maximum that tends



(a) $\overline{\kappa}(s) = \exp(-s^4)$ (b) $\overline{\kappa}(s) = s^4 \cdot \exp(-s^4)$ (c) $\overline{\kappa}(s) = s^8 \cdot \exp(-s^8)$

Figure 9.1: Hybrid transforms of $\chi_{\mathcal{F}} = \mathbf{1}_{[a,b)}$ for several choices of kernel κ

to $1/b$ as $p \to \infty$, with a spikier aspect as $p \to \infty$. Transforms of this type record the *variations* of the Euler characteristic curve with inverted scales. We refer to the following section for more involved experiments on synthetic data.

9.2 Heuristics for the Euler curves and their transforms

In this section, we assume that $m = 1$ and study Euler characteristic curves associated with the filtered Čech complex (Example 4.2) of several point clouds as well as hybrid transforms of these curves. We overview how these descriptors can extract information about the topology, geometry, and sampling density of the input data.

For computational reasons, we rather use a homotopy equivalent complex in numerical experiments, called the *filtered alpha complex*, or *alpha filtration*, which is a subcomplex of the Delaunay triangulation; see Bauer and Edelsbrunner (2017). Similarly, we consider *function-alpha filtrations* instead of function-Čech filtrations in multi-parameter settings.

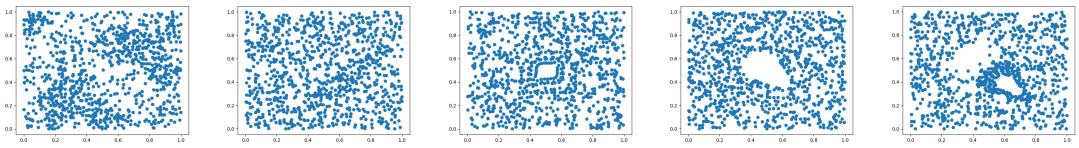
Topology, sampling: `ORBIT5K` data set

While apparently coarse descriptors, Euler characteristic curves allow us to extract relevant scales at which topological differences between two different processes are revealed. We illustrate this fact on the `ORBIT5K` data set.

The `ORBIT5K` data set is often used as a standard benchmark for classification methods in topological data analysis (Adams et al., 2017; Carrière, Chazal, et al., 2020; Reinauer, Caorsi, and Berkouk, 2021). This data set consists of subsets of a thousand points in the unit cube $[0, 1]^2$ generated by a dynamical system that depends on a parameter $\rho > 0$. To generate a point cloud, an initial point (x_0, y_0) is drawn uniformly at random in $[0, 1]^2$ and then the sequence of points (x_n, y_n) for $n = 0, \ldots, 999$ is generated recursively via the dynamic:

$$\begin{aligned} x_{n+1} &= x_n + \rho y_n \,(1 - y_n) && \mod 1, \\ y_{n+1} &= y_n + \rho x_{n+1} \,(1 - x_{n+1}) && \mod 1. \end{aligned}$$

In Figure 9.2, we illustrate typical orbits for $\rho \in \{2.5, 3.5, 4.0, 4.1, 4.3\}$.



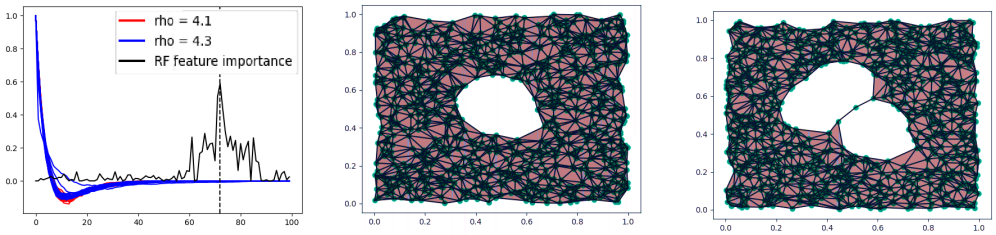
(a) $\rho = 2.5$ (b) $\rho = 3.5$ (c) $\rho = 4.0$ (d) $\rho = 4.1$ (e) $\rho = 4.3$

Figure 9.2: Examples of point clouds from the `ORBIT5K` data set.

In Figure 9.3a, we display the Euler characteristic curves for several realizations with parameters $\rho = 4.1$ and $\rho = 4.3$. We also plot the *feature importance* function of a random forest classifier trained on Euler characteristic curves of a small sample of 50 point clouds. In Figures 9.3b and 9.3c, we display the alpha complexes for two

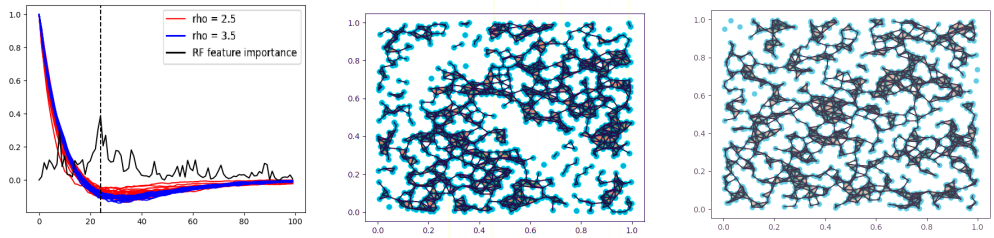
typical processes truncated at the filtration value corresponding to the largest feature importance. For a large range of high filtration values—approximately between 60 and 90—, the Euler characteristic curve of each class of process typically differs by one unit since the class with parameter $\rho = 4.3$ has an extra hole. This phenomenon is easily captured by the random forest classifier.

We apply the same methodology to discriminate between $\rho = 2.5$ and $\rho = 3.5$ in Figure 9.4. The difference between these two classes is more related to the distribution of points than to a persistent topological feature of the point clouds. At the scale selected by the feature importance of the random forest, the alpha complex for $\rho = 2.5$ in Figure 9.4b tends to have many tiny connected components, while the one for $\rho = 3.5$ is almost connected. We were then able to select a relevant scale at which the difference in the distribution of points is revealed by the topology of the alpha filtration.



(a) ECC and feature importance (b) $\rho = 4.1$ (c) $\rho = 4.3$

Figure 9.3: `ORBIT5K` classification problem: $\rho = 4.1$ VS $\rho = 4.3$.



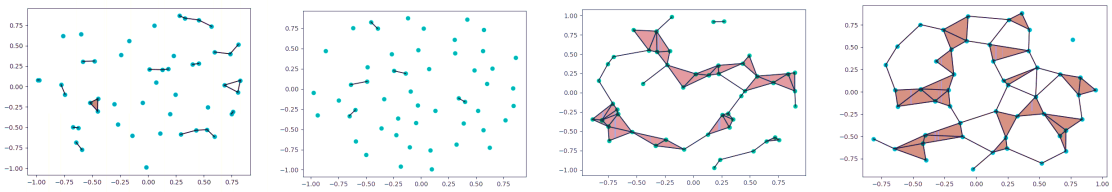
(a) ECC and feature importance (b) $\rho = 2.5$ (c) $\rho = 3.5$

Figure 9.4: `ORBIT5K` classification problem: $\rho = 2.5$ VS $\rho = 3.5$.

Sampling: Poisson and Ginibre point processes

We perform a similar analysis to discriminate between two types of point processes: a Poisson point process (PPP) and a Ginibre point process (GPP). This setup has been introduced in Obayashi, Hiraoka, and Kimura (2018). The specificity of Ginibre processes lies in repulsive interactions between points. While a standard PPP could have some very small and very large cycles, we expect the GPP to have more medium-sized cycles since points tend to be well dispersed. We classify this toy data set with a random forest

classifier and select the two scales corresponding to the most important features of the classifier. Ginibre point processes are generated using Moroz (2020). In Figure 9.5, we plot two examples of point clouds together with their alpha complexes at these scales.



(a) PPP (b) GPP (c) PPP (d) GPP

Figure 9.5: Examples of alpha complexes on PPP and GPP point clouds at two scales t_1 (Figures (a) and (b)) and t_2 (Figures (c) and (d)) with $t_1 < t_2$.

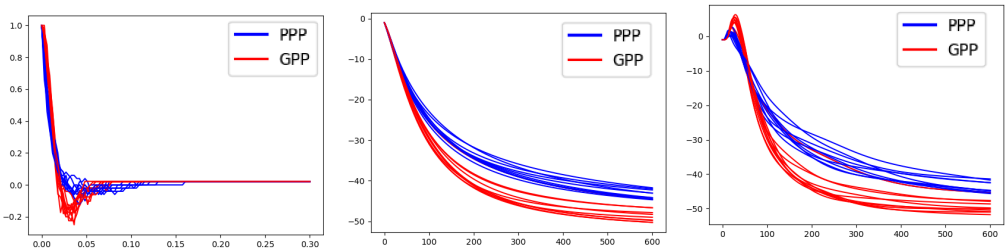
We plot Euler curves in Figure 9.6a. The Euler curves suggest that these classes differ at different scales, as it was visible in Figure 9.5:

- The Euler curves of the PPP class decrease in a steeper way. Indeed, a GPP has repulsive interactions between the points. Therefore, the pairwise distance between points tends to be larger and connected components do not die too early.
- The global minimum for the GPP class is lower since cycles appear all at once at some medium scale. In comparison, cycles of PPP classes are better dispersed across scales.
- Compared to curves of the GPP class, the curves of the PPP class tend to stay negative for a longer time. Indeed, unlike GPP, PPP allow for very large cycles to exist.

We plot the transforms of these curves for several kernels in Figures 9.6b and 9.6c. Choosing the primitive kernel $\overline{\kappa} : s \mapsto \exp(-s)$ emphasizes the small scales of the Euler curves in the larger scales of the transform. Such a descriptor separates well the two classes due to the earlier death of connected components for the PPP class. The primitive kernel $\overline{\kappa} : s \mapsto \exp(-s^4)$ also extracts this information. In addition, it has a higher global maximum for the GPP class that also enables distinction between the two classes. This maximum is created by the global minimum of the Euler curves. This experiment is a piece of evidence that this kernel carries more information than the exponential kernel and will therefore be preferred for applications.

Geometric features, sampling: different samplings on a manifold

We now show an experiment where we can illustrate how our various descriptors can discriminate between samplings and characterize the shape of a manifold. We consider two set-ups. The first set-up consists of clouds of 500 points sampled in two different ways on a torus embedded in $\mathbb{R}^3$. The first sampling is a uniform sampling (Diaconis, Holmes, Shahshahani, et al., 2013). The second is a non-uniform sampling where we draw (θ, φ) uniformly in $[0, 2\pi]^2$ and obtain a point on the torus through the embedding $\Psi_{\mathbb{T}^2} : (\theta, \varphi) \mapsto (x_1, x_2, x_3)$ where:



(a) ECC (b) HT, $\overline{\kappa}(s) = \exp(-s)$ (c) HT, $\overline{\kappa}(s) = \exp(-s^4)$

Figure 9.6: Euler characteristic curves (ECC) and their transforms (HT) for PPP VS GPP data set

$$\begin{cases} x_1 = (2 + \cos(\theta))\cos(\varphi), \\ x_2 = (2 + \cos(\theta))\sin(\varphi), \\ x_3 = \sin(\theta). \end{cases}$$

The second set-up consists of clouds of 500 points drawn in two ways on the unit sphere of $\mathbb{R}^3$. The first sampling is uniform. The second sampling is a non-uniform sampling where we draw θ uniformly in $[0, \pi]$ and φ according to a normal distribution centered on π. We obtain a point on the sphere via the classical spherical coordinates parametrization $\Psi_{\mathbb{S}^2} : (\theta, \varphi) \mapsto (x_1, x_2, x_3)$ where:

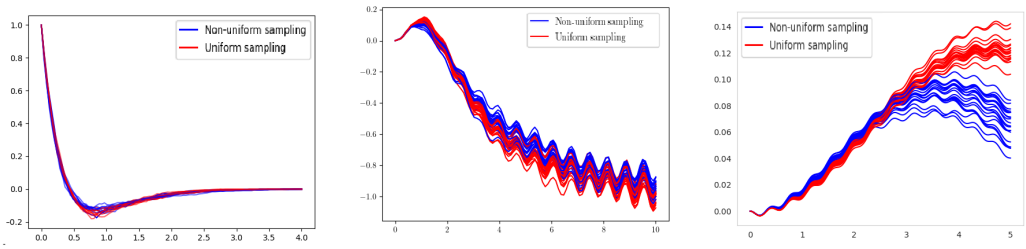
$$\begin{cases} x_1 = \sin(\theta)\cos(\varphi), \\ x_2 = \sin(\theta)\sin(\varphi), \\ x_3 = \cos(\theta). \end{cases}$$

In Figures 9.7a and 9.7b, we show the Euler curves and their hybrid transforms with primitive kernel $\overline{\kappa} : s \mapsto \cos(s)$ for these two classes of samplings on the torus. Up to a rescaling, this corresponds to a Fourier sine transform. In Figure 9.7c, we show the hybrid transforms for our two classes of samplings on the sphere.

In both cases, Euler curves associated with the data drawn on the same manifold all have the same profile, with a minimum value that tends to be lower for the uniform sampling. Similarly, the oscillations of the transforms are in phase and have the same amplitude. However, from one manifold to another, the phase and amplitude of the oscillations of the transforms differ significantly. This suggests that they are related to global quantities and are signatures of the support manifold. In contrast, the samplings show up in the vertical shifts of the oscillations of the transforms. This interpretation allows us to go beyond the classical signal/noise dichotomy in the persistence barcodes. Although it makes no doubt that this sampling information can be retrieved from low-persistence features, it is still unclear how to read it from a persistence barcode. We claim this is another step towards a more thorough analysis of the geometric quantities involved in the low-persistence features.

Geometric features, sampling: two different patterns in noise

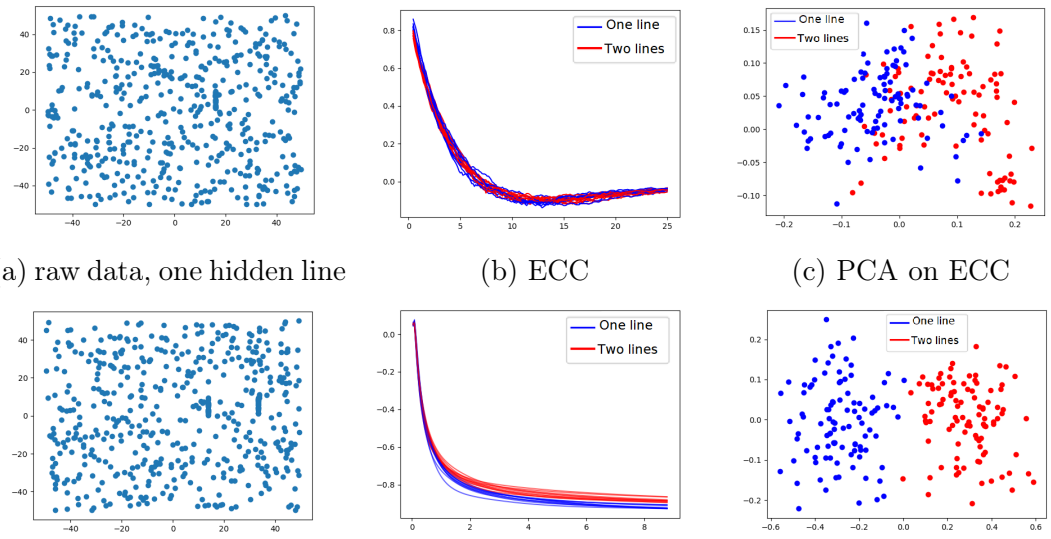
In this final illustrative experiment, we try to distinguish patterns in a heavy clutter noise. One class has one line hidden in the noise, while the other has only two. Each line



(a) ECC, torus data (b) HT, torus data (c) HT, sphere data

Figure 9.7: ECC and HT, two sampling on a torus and sphere

will induce a very dense zone creating early dying connected components. In Figure 9.8, we plot two examples of point clouds, the Euler curves of each class, and their hybrid transforms with primitive kernel $\overline{\kappa} : s \mapsto \exp(-s^4)$. We also provide principal component analysis (PCA) plots of these two descriptors. The difference between the two classes is visible at the beginning of the Euler characteristic curves. However, looking at the full curve does not allow us to correctly see this difference, as shown by the PCA plot. On the contrary, the transform puts a strong emphasis on the beginning of the Euler curves, leading to a direct linear separation of the two classes. As a final sanity check, we ran a k-means algorithm to cluster between the two classes and reached an accuracy of 99% for the hybrid transforms and only 52.5% for the Euler curves.



(a) raw data, one hidden line (b) ECC (c) PCA on ECC

(d) raw data, two hidden lines (e) HT, $\overline{\kappa}(s) = \exp(-s^4)$ (f) PCA on HT

Figure 9.8: Pattern hidden in clutter noise

9.3 Experiments

In this section, we present all quantitative experiments conducted on synthetic and real-world point cloud data and on real graph data sets. Material to reproduce our experiments is available online on our GitHub repository: `https://github.com/vadimlebovici/eulearning`.

9.3.1 Curvature regression

We consider a set-up from Bubenik, Hull, et al. (2020) where we draw 1000 points uniformly at random on the unit disk of a surface of constant curvature K and try to predict K in a supervised fashion. Recall that if $K > 0$ (resp. $K = 0$, $K > 0$), the corresponding surface is a sphere (resp. the Euclidean plane, the hyperbolic plane). We observe 101 samples from space with curvature $[-2, -1.96, \ldots, 1.96, 2]$ and validate our model on a testing set of 100 point clouds sampled from space with random curvature drawn uniformly in $[-2, 2]$. We compare the R^2 score in Table 9.1 with that of the original paper, which uses persistent landscapes (PL) along with a support vector regressor (SVR) and with Persformer (Reinauer, Caorsi, and Berkouk, 2021). Note that since we are trying to tackle a regression problem, we use an SVR or a random forest regressor to predict the curvature from our vectorization.

Method	PL+SVR	Persformer	ECC+SVR	ECC+RF	HT+SVR	HT+RF
R^2 score	0.78	**0.94**	0.70	0.93	0.79	0.89

Table 9.1: R^2 score for curvature regression data

First, we remark that the ECC descriptor combined with a random forest has an accuracy comparable to the state-of-the-art. We also remark that taking a transform does not improve the regression accuracy when considering a robust classifier such as RF but does improve the accuracy when using a linear regressor (SVR). Note that hybrid transforms combined with a linear regressor have an accuracy similar to that of persistent landscapes. However, persistent landscapes require the computation of the entire persistence barcodes, while hybrid transforms bypass this costly operation.

9.3.2 `ORBIT5K` data set

Supervised setting. Here, we perform a supervised analysis of the `ORBIT5K` data set introduced in Section 9.2. Given an orbit, we try to predict the value of the parameter ρ, which takes value in $\{2.5, 3.5, 4.0, 4.1, 4.3\}$. We generate 700 training and 300 testing orbits for each class. We compare our score with standard classification methods in Table 9.2. The results are averaged over ten runs. PWG-K, SW-K and PF-K are kernel methods on persistence barcodes taken respectively from Carriere, Cuturi, and Oudot (2017), Kusano, Hiraoka, and Fukumizu (2016), and Le and Yamada (2018). Perslay and Persformer are two methods that use a neural network architecture to vectorize persistence barcodes (Carrière, Chazal, et al., 2020; Reinauer, Caorsi, and Berkouk,

2021). The Euler characteristic curves and one-parameter hybrid transforms (HT1) are computed on the alpha filtration of the point cloud. The Euler characteristic surfaces, the two-parameter Radon transform (RT) and hybrid transforms (HT2) are computed using a function-alpha filtration associated with a kernel density estimator post-composed with a decreasing function. The decreasing function is $x \mapsto -x$ for the ECSs and $x \mapsto \exp(-x^2)$ for the HTs. All descriptors have a resolution of 900 (hence of 30×30 for two-parameter ones) and were trained with an XGBoost classifier (T. Chen and Guestrin, 2016). We select the hyperparameters of our descriptors by cross-validation:

- For the ECC, the quantiles (see *Implementation* in Section 9.1) are selected in $\{(0.1, 0.9),\ (0.2, 0.8), (0.3, 0.7)\}$.
- For the ECS, the quantiles are selected in the same set as for the ECC for both parameters.
- For the HT1, the range is selected in $\{[0, 50], [0, 100], [0, 500], [0, 1000]\}$ and the primitive kernel $\overline{\kappa}$ in $\{s \mapsto \exp(-s^4),\ s \mapsto s^4 \exp(-s^4), s \mapsto s^8 \exp(-s^8)\}$.
- For the HT2, the primitive kernel and the range for the first parameter are the same as for the HT1, and the range for the second parameter is selected in $\{[0, 50], [0, 80], [0, 100], [0, 500]\}$.
- For the RT, the ranges are selected in the same set as the HT2 for both parameters.

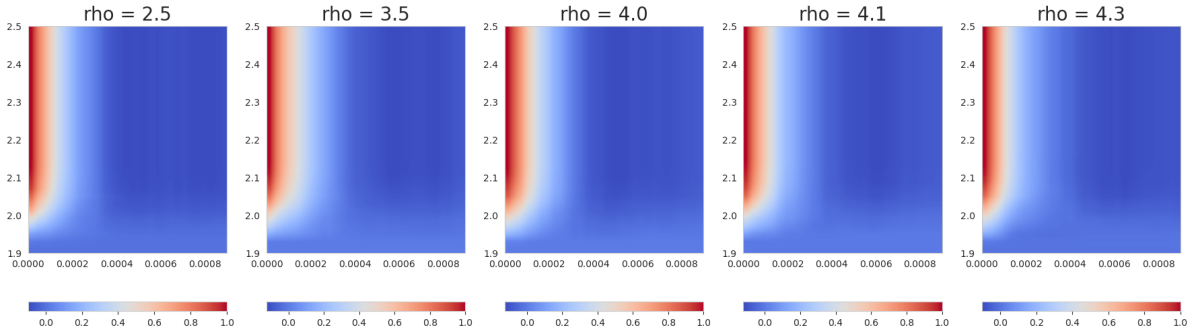
We show in Figure 9.9 some examples of each descriptor renormalized by the number of points for the classes $\rho = 2.5$ and $\rho = 4.3$, where the HT2 is computed with $\overline{\kappa} : s \mapsto s^4 \exp(-s^4)$.

Method	PWG-K	SW-K	PF-K	Perslay	Persformer
Accuracy	76.6 ± 0.7	83.6 ± 0.9	85.9 ± 0.8	87.7 ± 1.0	91.2 ± 0.8
Method	ECC + XGB	HT1 + XGB	**ECS + XGB**	RT + XGB	HT2 + XGB
Accuracy	83.8 ± 0.5	82.8 ± 1.4	**91.8 ± 0.4**	90.5 ± 0.4	89.9 ± 0.5

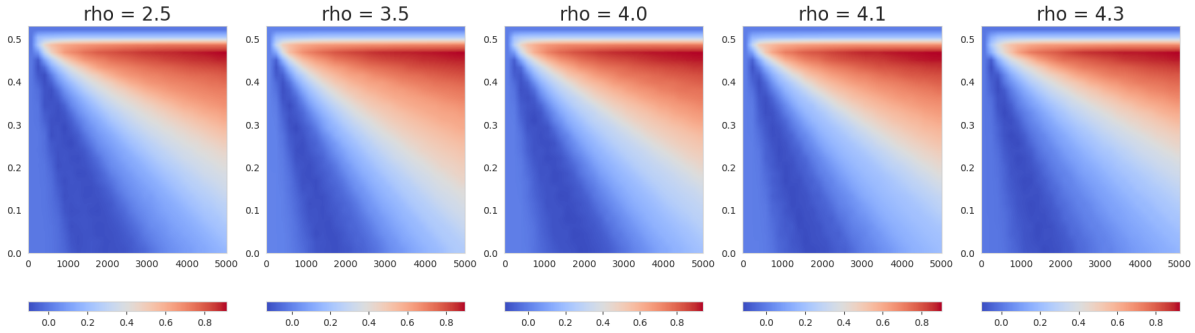
Table 9.2: Classification scores for the `ORBIT5K` data set

One-parameter descriptors have accuracy similar to kernel methods on persistence barcodes at a reduced computational cost, while two-parameter descriptors compete with neural network-based vectorization methods. We make our claims on computational times more precise in Section 9.3.5.

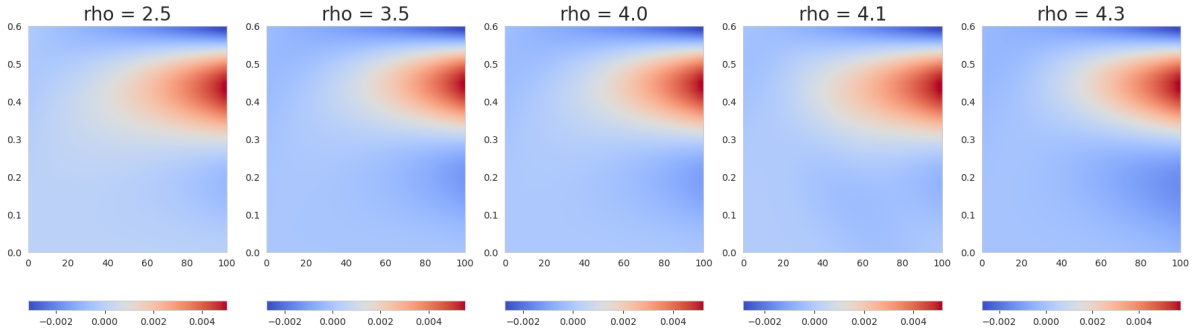
Ablation study. We also study the role of the dimension of the feature vector in the supervised classification task. The results are shown in Figure 9.10. When plugging a random forest classifier, all descriptors are robust to a decrease in the size of the feature vector. However, hybrid transforms seem to maintain a competitive accuracy for low-dimensional features, especially the two-parameter ones. When using an SVM classifier for the one-parameter descriptors, the gain from considering a hybrid transform is clear,



(a) ECS



(b) RT

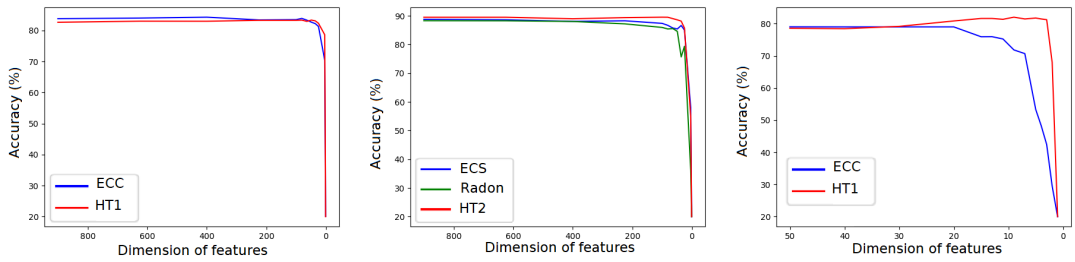


(c) HT2

Figure 9.9: Examples of 2D descriptors

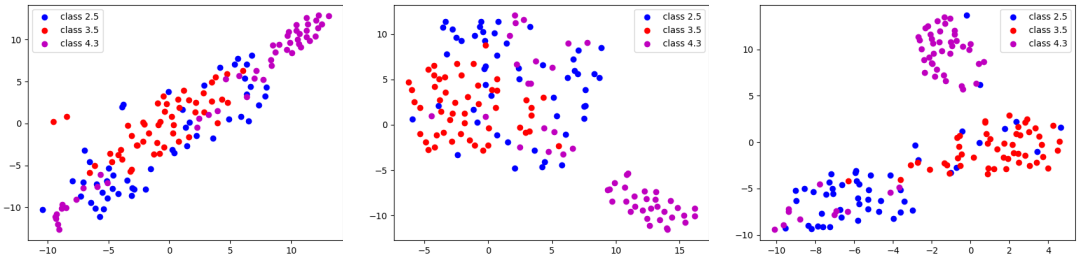
and the accuracy of the SVM benefits from this strong dimension reduction. Evaluating hybrid transforms at only three values of $\xi \in \mathbb{R}_+^*$ yields feature vectors achieving approximately 80% accuracy, demonstrating the compression properties of this tool.

Unsupervised setting. We consider several unsupervised classification tasks on the same data set. We consider 50 point clouds for each choice of $\rho \in \{2.5, 3.5, 4.3\}$. We map all descriptors in $\mathbb{R}^2$ using a tSNE dimension reduction (Van der Maaten and Hinton, 2008) and report the results in Figure 9.11. Here, hybrid transforms differ from the other two methods and succeed in adequately separating the three classes.



(a) one-parameter, RF classifier (b) two-parameter, RF classifier (c) one-parameter, SVM classifier

Figure 9.10: Accuracy with respect to feature dimension.



(a) ECS (b) RT (c) HT2

Figure 9.11: tSNE of our descriptors computed on several classes of the `ORBIT5K` data set.

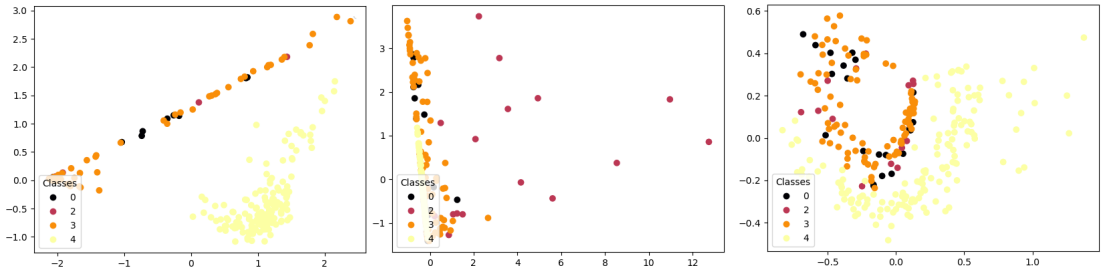
9.3.3 Sidney object recognition data set

The Sidney urban objects recognition data set consists of 3D point clouds of everyday urban road objects scanned with a LIDAR (De Deuge et al., 2013) traditionally used for multi-class classification. Likewise to Section 9.3.2, all descriptors are computed using a function-alpha filtration associated with a kernel density estimator post-composed with a decreasing function.

Unsupervised setting. In Figure 9.12, we show a PCA of the ECSs and HTs on the classes *4-wheeler vehicles* (labelled 0), *buses* (2), *cars* (3), and *pedestrians* (4). In this case, the ECSs separate the class of pedestrians from all the vehicle classes. The same separation is achieved by the HTs with primitive kernel $\overline{\kappa} : s \mapsto s^4 \exp(-s^4)$. In contrast, HTs with primitive kernel $\overline{\kappa} : s \mapsto \exp(-s^4)$ separate buses from other classes. These experiments illustrate the flexibility provided by a broad choice of kernels for the hybrid transforms.

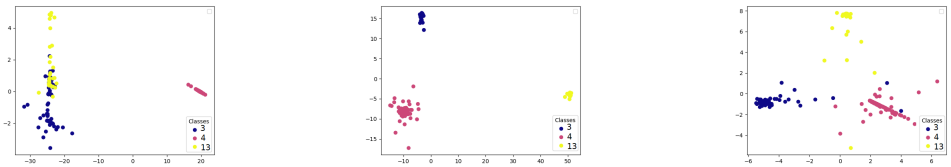
Supervised setting. Even more striking are the experiments from Figure 9.13. We perform a Linear Discriminant Analysis for classes *cars* (3), *pedestrians* (4), and *vans* (13) to embed the HTs and ECSs in $\mathbb{R}^2$. All the classes are separated for the RTs and the HTs with primitive kernel $\overline{\kappa} : s \mapsto s^4 \exp(-s^4)$. In comparison, the ECSs only manage

to separate the pedestrian class from the two motor-vehicle classes.



(a) ECS (b) HTs, $\overline{\kappa}(s) = \exp(-s^4)$ (c) HTs, $\overline{\kappa}(s) = s^4 \exp(-s^4)$

Figure 9.12: PCA plots of ECSs and HTs for the Sidney object recognition data set.



(a) ECS (b) HTs, $\overline{\kappa}(s) = s^4 \exp(-s^4)$ (c) RTs

Figure 9.13: LDA plots of ECSs, HTs, and RTs for the Sidney object recognition data set.

9.3.4 Graph data

We have applied our method to the supervised classification of graph data. To build sublevel-sets filtrations of graphs, we consider the heat-kernel signature introduced in Sun, Ovsjanikov, and Guibas (2009) and defined as follows. For a graph $\mathcal{G} = (V, E)$, the *HKS function with diffusion parameter* t is defined for each $v \in V$ by:

$$\mathrm{hks}_t(v) = \sum_{k=1}^{|V|} \exp(-t\lambda_k)\psi_k(v)^2,$$

where λ_k is the k-th eigenvalue of the normalized graph Laplacian and ψ_k the corresponding eigenfunction. We consider the HKS with parameters $t = 1$ and $t = 10$ as filtrations. We also consider the 1/2-Ricci and Forman curvatures (Samal et al., 2018), centrality, and edge betweenness on connected graphs. In addition, some data sets (PROTEINS, COX2, DHFR) come with functions defined on the graph nodes. We can use several combinations of these functions to define sublevel-sets filtrations of graphs and compute Euler characteristic profiles (ECP), Radon transforms (RT) and hybrid transforms (HTn).

For this set of experiments, we cross-validate over several combinations of the filtration functions proposed above, several truncations of the vectorization (which had little impact in practice), and a primitive kernel chosen among $\{s \mapsto \cos(s), s \mapsto \cos(s^2), s \mapsto$

Method	MUTAG	COX2	DHFR	PROTEINS	COLLAB	IMDB-B	IMDB-M	NCI1
SV	88.2(0.1)	78.4(0.4)	78.8(0.7)	72.6(0.4)	79.6(0.3)	74.2(0.9)	49.9(0.3)	71.3(0.4)
RetGK	90.3(1.1)	**81.4(0.6)**	81.5(0.9)	**78.0(0.3)**	81.0(0.3)	71.9(1.0)	47.7(0.3)	**84.5(0.2)**
FGSD	**92.1**	-	-	73.4	80.0	73.6	**52.4**	79.8
GIN	90(8.8)	-	-	76.2(2.6)	80.6(1.9)	**75.1(5.1)**	52.3(2.8)	82.7(1.6)
Perslay	89.8(0.9)	80.9(1.0)	80.3(0.8)	74.8(0.3)	76.4(0.4)	71.2(0.7)	48.8(0.6)	73.5(0.3)
Atol	88.3(0.8)	79.4(0.7)	82.7(0.7)	71.4(0.6)	**88.3(0.2)**	74.8(0.3)	47.8(0.7)	78.5(0.3)
ECC 1D	87.2(0.7)	78.1(0.2)	79.4(0.5)	74.7(0.4)	77.3(0.2)	72.4(0.4)	48.5(0.3)	74.4(0.2)
HT 1D	87.4(0.8)	78.1(0.2)	77.9(0.4)	73.3(0.4)	78.2(0.2)	73.9(0.4)	49.7(0.4)	73.9(0.2)
ECV	90.0(0.8)	80.3(0.4)	82.0(0.4)	75.0(0.3)	78.3(0.1)	73.3(0.4)	48.7(0.4)	76.3(0.1)
RT	87.3(0.6)	79.7(0.4)	81.3(0.4)	75.4(0.4)	77.5(0.2)	74.0(0.5)	50.2(0.4)	75.6(0.2)
HT nD	89.4(0.7)	80.6(0.4)	**83.1(0.5)**	75.4(0.4)	77.6(0.2)	74.7(0.5)	49.9(0.4)	66.4 (0.2)

Table 9.3: Mean accuracy and standard deviation for graph data.

$\exp(-s^4), s \mapsto s^4 \exp(-s^4)\}$ for HTn. We report our scores in Table 9.3. The first four methods are state-of-the-art classification methods on graphs that use kernels or neural networks. We report the scores from the original papers, Tran, Vo, and Hasegawa (2019), Verma and Z.-L. Zhang (2017), Xu et al. (2019), and Z. Zhang et al. (2018). Perslay (Carrière, Chazal, et al., 2020), and Atol (Royer et al., 2021) are topological methods that transform the graphs into persistence barcodes using HKS functions. It is known that Atol performs especially well on large data sets (both in terms of number of data and graphs size), i.e., COLLAB and NCI1. Still, we reach a similar to better accuracy for all the other data sets.

Besides highly competitive classification scores, our method has two advantages over the other topological methods. First, we bypass the computation of persistence barcodes and thus classify with lower computational cost; see Sections 9.1 and 9.3.5. Second, as opposed to other invariants such as multi-parameter persistent images (Carrière and Blumberg, 2020), our method naturally generalizes to m-parameter persistence with $m \geq 3$ at a very low computational cost. To our knowledge, this is the first time a topology-based method uses more than 3 filtration parameters. This results in an increase in accuracy since each filtration function leverages information on the graph-data structures.

Note that the methods SV, FGSD, and GIN do not average ten times and rather consider a single 10-fold sample which can slightly boost their accuracies.

9.3.5 Timing

In this section, we compare the computational cost of our different methods to that of persistence images. Persistence images have been introduced as a vectorization of one-parameter persistence modules in Adams et al. (2017) and generalized to the multi-parameter setting in Carrière and Blumberg (2020). We choose to compare the computational cost of our methods to that of persistence images as they appear to be a faster vectorization method than persistence kernels and persistence landscapes; see Carrière and Blumberg (ibid., Table 2).

Constant resolution. We report in Table 9.4 the time to compute our descriptors and persistent images on the full `ORBIT5K` data set with a fixed resolution of 900. We assume that simplex trees are precomputed[1] using the `Gudhi` library (Rouvreau, 2015). Our descriptors are computed using the parameters achieving the highest accuracy for the classification task; see Section 9.3.2. Persistence images are computed with the `Gudhi` library for one-parameter filtrations and with the `MMA` package for two-parameter filtrations (Loiseaux, Carrière, and Schreiber, 2022) with default parameters and the same resolution as our two-parameter descriptors, i.e., 30×30. To compute persistence images, one first needs to compute the persistence barcodes of simplex trees in the one-parameter case or persistence modules approximations in the two-parameter case (Loiseaux, Carriere, and Blumberg, 2022, Section 3). We include these additional costs in the computational times of persistent images. However, the time to compute the PI1 descriptor on the full `ORBIT5K` data set breaks down to 5 seconds to compute the persistence barcodes and 134 seconds for the persistence images themselves.

ECC	HT1	PI1	ECS	Radon	HT2	PI2
16	719	139	144	119	805	2034

Table 9.4: Computation times (s) for `ORBIT5K` with constant resolution.

As expected from the time complexities of the algorithms (Section 9.1), Euler characteristic profiles and Radon transforms are at least ten times faster than persistence images, and hybrid transforms are four times faster in the two-parameter case. One-parameter hybrid transforms may appear costly to compute, but this will be mitigated in the next paragraph. Finally, we point out that we implemented our tools in `Python` and not in `C++`, which is very likely to result in longer computation times. On the contrary, persistence images in one and two parameters both benefit from a `C++` implementation.

Constant accuracy. We report in Table 9.5 the time to compute our descriptors on the full `ORBIT5K` data set with the lowest resolution before accuracy drop-out as reported in Figure 9.10. More precisely, we chose the lowest possible resolutions to ensure a classification accuracy of 82% for one-parameter descriptors and of 89% for two-parameter descriptors, that is, a resolution of 30 for ECC, of 9 for HT1, of 20×20 for ECS and Radon, and of 6×6 for HT2. Other parameters remain unchanged. The interest in using hybrid transforms over Euler characteristic profiles is now clear: the concentration of information provided by hybrid transforms makes it possible to classify the data set with feature vectors of reduced dimension, which considerably speeds up the computations.

9.3.6 Take-home message

The experiments from this section suggest that Euler characteristic profiles are very powerful descriptors since they allow for state-of-the-art accuracy when coupled with a

[1]Note that computing simplex trees takes around 66s in the one-parameter setting and around 420s in the two-parameter setting; the difference lies in the cost of computing codensity on point clouds.

ECC	HT1	ECS	Radon	HT2
16	5	135	45	69

Table 9.5: Computation times (s) for `ORBIT5K` with smallest resolution before accuracy drop-out.

robust classifier (XGB or RF) at a very competitive computational cost. On the one hand, Radon transforms show accuracy and computational complexity very similar to Euler characteristic profiles. On the other hand, hybrid transforms have similar accuracy but are more costly to compute, especially in the one-parameter setting; see Table 9.4. In addition to the theoretical properties exposed in the previous chapters, the motivation to use hybrid transforms is two-fold.

First, in an unsupervised setting or when plugging a linear classifier, the lack of diversity in Euler characteristic profiles and Radon transforms can be detrimental to the separation of classes. In contrast, hybrid transforms are competitive descriptors in such tasks due to the wide diversity in the choice of kernels and their sensitivity to slight variations in Euler characteristic profiles.

Second, hybrid transforms provide a very powerful compression of the signal from the Euler profile (Figure 9.10) at a very low computational cost (Table 9.5). This makes hybrid transforms a very robust descriptor combining dimension reduction and feature extraction.

Conclusion

In this thesis, we explored two complementary approaches to describe multi-parameter filtrations of topological spaces in an informative and implementable manner.

Direct-sum decompositions. In the first part, we followed the approach developed in persistence theory, which consists in applying homology to the filtration. This yields an algebraic object called a persistence module. In the one-parameter case, the persistence module is fully characterized by a multiset of intervals called a barcode. In the case of multiple parameters, the resulting persistence modules do not admit such simple descriptions. Therefore, we sought for subclasses of persistence modules that are described by a multiset of connected and convex subsets—intervals—of the parameter space. In more formal terms, we studied the subclasses of persistence modules that are decomposable into a direct sum of interval modules, and restricted ourselves to the two-parameter case. Keeping in mind the need to algorithmically test membership in these subclasses, we studied the existence of those that admit a local characterization, i.e., those that can be tested for membership by observing only the restrictions of the persistence modules to finite subsets of the parameter space. We showed that interval-decomposability cannot be tested locally by means of restrictions to finite grids of bounded size. However, our main result shows that global rectangle-decomposability is equivalent to rectangle-decomposability of all restrictions to squares, i.e., 2×2 grids. This allowed us to construct efficient algorithms for testing decomposability and computing decompositions of modules induced in homology by a finite simplicial filtration. Moreover, we showed that rectangle-decomposable modules naturally appear in the study of the homology of level sets of real-valued functions. Finally, we proved that the class of rectangle-decomposable modules is maximal, in a precise sense, among classes that can be locally characterized on squares.

The work conducted in this first part raises several natural questions. Since rectangle-decomposability is maximal when testing on squares, one may wonder if it is possible to locally characterize interval-decomposability beyond rectangles when allowing for other test subsets than squares. Moreover, is it possible to move beyond indecomposables which are pointwise 0- or 1-dimensional? Can one determine "locally" if a given persistence module decomposes into a direct sum of indecomposables belonging to a certain predefined class of indecomposables? Finally, do the results generalize to persistence modules over $\mathbb{R}^n$? For persistence modules indexed over $\mathbb{Z}^2$, the *exactness* condition (Definition 4.9) ensuring global block-decomposability can be expressed as the vanishing of specific homology groups of the Koszul complex associated to the persistence module. Although this condition generalizes to persistence modules indexed over $\mathbb{Z}^n$ for $n \geq 3$,

the question of whether a block-decomposability result is true in this case remains open.

Euler characteristic based invariants. In the second part of the thesis, we studied descriptors of topological filtrations based on calculations of Euler characteristic rather than homology. These descriptors have the advantage of completely bypassing the construction of persistence modules and can thus be computed much more efficiently than the barcode and its multi-parametric analogues presented above. We defined these invariants and formulate our results using the convenient language of constructible functions and Euler calculus. Our main contribution is the introduction of integral transforms of constructible functions involving integration with respect to Lebesgue measure and Euler characteristic—hybrid transforms. We began with a thorough and general theoretical study of these integral transforms. We showed that beside giving access to a wide variety of kernels (Fourier, Laplace, etc.), Lebesgue integration allows for smooth output functions. On the other hand, Euler calculus induces transforms that are compatible with topological operations on functions, such as (constructible) convolution or duality. Importantly, the hybrid Fourier transform satisfies a reconstruction result when restricted to the subgroup of compactly supported γ-constructible functions. Moreover, the use of Euler characteristic allowed us to prove index formulae in the context of persistence of sublevel sets of subanalytic functions. Combined with the work of Bobrowski and Borman (2012), these index formulae yield expressions for the mean values of these invariants in the context of sublevel-sets persistence of Gaussian random fields. Then, we focused on Euler profiles—the constructible functions associated with finitely presentable simplicial filtrations—and their transforms. We proved stability results for these objects with respect to the L_1 distance on constructible functions and on filtration functions. We also showed statistical results for these objects, which are essentially direct consequences of existing results, but which crucially used the reduction property of multi-parameter hybrid transforms to one-parameter ones (Corollary 7.10). Finally, we showed that these invariants are powerful descriptors in topological data analysis by comparing ourselves to existing persistence based methods. Euler profiles are informative at very low computational cost while hybrid transforms provide powerful descriptors in unsupervised learning at a reasonable to better cost due to their compression power. Along the way, we provided numerous examples of the type of topological and geometric information captured by our tools.

Celebrated results on the classical Fourier transform state that it is an isometry of L^2 and a Lipschitz mapping from L^1 to L^∞. One can naturally wonder if similar results hold on hybrid transforms. Remark 6.52 shows that it does not for L_p metrics between constructible functions and the L^∞ metric between transforms. The question remains open for other metrics on constructible functions and their transforms. Considering the pushforward operation involved in the definition of hybrid transforms, these stability results would likely require a metric on the group of constructible functions with respect to which Euler calculus operations are continuous. As a byproduct, such a metric might allow us to design approximation schemes of shapes (e.g. compression) with converging topological and hybrid transforms. Following the correspondence between constructible functions and characteristic cycles (Kashiwara and Schapira, 1990, Theorem. 9.7.11), a metric on functions would induce a metric on cycles with respect to which classical

operations on cycles (such as pushforward) are continuous. Several facts suggest that this metric should not be straightforward to find: as shown by Berkouk (2022), a non-trivial metric on constructible functions cannot be upper-bounded by the convolution distance between sheaves.

On the applied side, it would be interesting to explore several directions. While we have validated our method on filtered simplicial complexes built on point clouds and graph data, it is common to build cubical complexes from images or 3D volumes data. Euler characteristic curves have been used as descriptors of such complexes in the one-parameter setting (Jiang, Kurtek, and Needham, 2020; Smith and Zavala, 2021). As there is a vast number of filtration functions one can consider on images, it would be worth investigating the predictive power of the multi-parameter Euler profiles in this setting. While several applications are considered in Beltramo, Skraba, et al. (2022) and Dłotko and Gurnari (2022), a thorough benchmarking against other persistence methods and state-of-the-art image processing methods is still missing. Moreover, Radon transforms and hybrid transforms have still not been studied in this applied context. Another direction is based on the remark that Euler characteristic profiles are well-defined for families of topological spaces that are not necessarily non-decreasing with respect to inclusions. This widely extends the potential range of applications of our tools, notably to the study of time-varying simplicial complexes, as done in Xian et al. (2022). With the success of the classical Fourier transform on time signals in mind, it would be relevant to examine hybrid transforms of Euler profiles associated to such time-varying complexes.

Bibliography

Adams, Henry et al. (2017). "Persistence images: A stable vector representation of persistent homology". In: *Journal of Machine Learning Research* 18 (cit. on pp. 17, 19, 35, 36, 163, 171, 174, 184).

Adler, Robert J. and Jonathan E. Taylor (2009). *Random fields and geometry*. Springer Science & Business Media (cit. on pp. 141, 152, 153, 156).

Akbulut, Selman and Henry King (2012). *Topology of real algebraic sets*. Vol. 25. Springer Science & Business Media (cit. on p. 109).

Amézquita, Erik J et al. (2022). "Measuring hidden phenotype: Quantifying the shape of barley seeds using the Euler Characteristic Transform". In: *in silico Plants* 4.1 (cit. on pp. 20, 21, 37, 38, 121, 171).

Anai, Hirokazu et al. (2020). "DTM-based filtrations". In: *Topological Data Analysis: The Abel Symposium 2018*. Springer, pp. 33–66 (cit. on p. 169).

Aukerman, Andrew et al. (2021). "Persistent homology based characterization of the breast cancer immune microenvironment: a feasibility study". In: *Journal of Computational Geometry* 12.2, pp. 183–206 (cit. on pp. 13, 31).

Azumaya, Gorô (1950). "Corrections and supplementaries to my paper concerning Krull-Remak-Schmidt's theorem". In: *Nagoya Mathematical Journal* 1, pp. 117–124 (cit. on pp. 15, 32, 48).

Barannikov, Serguei (1994). "The framed Morse complex and its invariants". In: *Advances in Soviet Mathematics* 21, pp. 93–116 (cit. on pp. 13, 30).

Baryshnikov, Yuliy and Robert Ghrist (2009). "Target enumeration via Euler characteristic integrals". In: *SIAM Journal on Applied Mathematics* 70.3, pp. 825–844 (cit. on p. 109).

— (2010). "Euler integration over definable functions". In: *Proceedings of the National Academy of Sciences* 107.21, pp. 9525–9530 (cit. on pp. 141, 149).

Baryshnikov, Yuliy, Robert Ghrist, and David Lipsky (2011). "Inversion of Euler integral transforms with applications to sensor data". In: *Inverse problems* 27.12, p. 124001 (cit. on pp. 21, 38).

Bauer, Ulrich, Magnus Bakke Botnan, and Benedikt Fluhr (2021). "Structure and Interleavings of Relative Interlevel Set Cohomology". In: *arXiv preprint: 2108.09298* (cit. on pp. 17, 24, 34, 41, 100, 101).

Bauer, Ulrich and Herbert Edelsbrunner (2017). "The Morse theory of Čech and Delaunay complexes". In: *Transactions of the American Mathematical Society* 369.5, pp. 3741–3762 (cit. on p. 174).

Bauer, Ulrich and Luis Scoccola (2022). "Generic Two-Parameter Persistence Modules are Nearly Indecomposable". arXiv preprint: 2211.15306 (cit. on p. 159).

Beltramo, Gabriele, Rayna Andreeva, et al. (2021). *Euler Characteristic Surfaces*. arXiv preprint: 2102.08260 (cit. on p. 143).

Beltramo, Gabriele, Primoz Skraba, et al. (2022). "Euler characteristic surfaces". In: *Foundations of Data Science* 4.4, pp. 505–536 (cit. on pp. 19, 20, 37, 158, 189).

Bendich, Paul et al. (2013). "Homology and robustness of level and interlevel sets". In: *Homology, Homotopy and Applications* 15.1, pp. 51–72 (cit. on pp. 16, 17, 24, 33, 34, 41, 100, 101).

Berkouk, Nicolas (2022). "Persistence and the Sheaf-Function Correspondence". arXiv preprint: 2207.06335 (cit. on pp. 20, 37, 189).

Berkouk, Nicolas and François Petit (2021). "Ephemeral persistence modules and distance comparison". In: *Algebraic and Geometric Topology* 21.1, pp. 247–277 (cit. on pp. 115, 119).

— (2022). *Projected distances for multi-parameter persistence modules*. arXiv preprint: 2206.08818 (cit. on pp. 19, 36, 143).

Bjerkevik, Håvard Bakke and Michael Lesnick (2021). "ℓ^p-Distances on Multiparameter Persistence Modules". arXiv preprint: 2106.13589 (cit. on pp. 156, 162).

Blanchette, Benjamin, Thomas Brüstle, and Eric J. Hanson (2022). "Homological approximations in persistence theory". In: *Canadian Journal of Mathematics*, pp. 1–38 (cit. on p. 54).

Bobrowski, Omer and Robert J. Adler (2014). "Distance functions, critical points, and the topology of random Cech complexes". In: *Homology, Homotopy and Applications* 16.2, pp. 311–344 (cit. on p. 156).

Bobrowski, Omer and Matthew Strom Borman (2012). "Euler integration of Gaussian random fields and persistent homology". In: *Journal of Topology and Analysis* 4.1, pp. 49–70 (cit. on pp. 21, 22, 26, 38, 39, 43, 121, 122, 141, 142, 146, 148, 152–154, 188).

Bobrowski, Omer and Matthew Kahle (2018). "Topology of random geometric complexes: a survey". In: *Journal of applied and Computational Topology* 1, pp. 331–364 (cit. on p. 156).

Bobrowski, Omer, Sayan Mukherjee, and Jonathan E. Taylor (2017). "Topological consistency via kernel estimation". In: *Bernoulli* 23.1, pp. 288–328 (cit. on pp. 13, 31, 155).

Botnan, Magnus Bakke (2017). "Interval Decomposition of Infinite Zigzag Persistence Modules". In: *Proceedings of the American Mathematical Society* 145.8, pp. 3571–3577 (cit. on p. 97).

Botnan, Magnus Bakke and William Crawley-Boevey (2020). "Decomposition of persistence modules". In: *Proceedings of the American Mathematical Society* 148.11, pp. 4581–4596 (cit. on pp. 13, 15, 16, 30, 32, 34, 48, 52, 53, 56, 99, 103).

Botnan, Magnus Bakke and Christian Hirsch (2022). "On the consistency and asymptotic normality of multiparameter persistent Betti numbers". In: *Journal of Applied and Computational Topology*, pp. 1–38 (cit. on pp. 156, 157, 169).

Botnan, Magnus Bakke, Vadim Lebovici, and Steve Oudot (2022). "On Rectangle-Decomposable 2-Parameter Persistence Modules". In: *Discrete and Computa-*

tional Geometry. Initially accepted in: 36th International Symposium on Computational Geometry (SoCG 2020), pp. 1–24 (cit. on pp. 17, 22, 23, 27, 34, 39, 40, 44, 101).

— (2023). "Local characterizations for decomposability of 2-parameter persistence modules". In: *Algebras and Representation Theory*, pp. 1–44 (cit. on pp. 17, 22, 23, 27, 34, 39, 40, 44).

Botnan, Magnus Bakke and Michael Lesnick (2018). "Algebraic stability of zigzag persistence modules". In: *Algebraic & Geometric Topology* 18.6, pp. 3133–3204 (cit. on pp. 83, 85, 104).

— (2022). "An introduction to multiparameter persistence". arXiv preprint: 2203.14289. To appear in: Proceedings of the ICRA 2020 (cit. on pp. 14, 15, 32, 119).

Botnan, Magnus Bakke, Steffen Oppermann, and Steve Oudot (2021). "Signed barcodes for multi-parameter persistence via rank decompositions and rank-exact resolutions". In: *arXiv preprint: 2107.06800* (cit. on pp. 53, 54).

Bubenik, Peter et al. (2015). "Statistical topological data analysis using persistence landscapes." In: *Jounral of Machine Learning Research* 16.1, pp. 77–102 (cit. on pp. 19, 36, 163, 171).

Bubenik, Peter, Michael Hull, et al. (2020). "Persistent homology detects curvature". In: *Inverse Problems* 36.2, p. 025008 (cit. on pp. 19, 36, 179).

Bubenik, Peter and Alexander Wagner (2020). "Embeddings of persistence diagrams into Hilbert spaces". In: *Journal of Applied and Computational Topology* 4.3, pp. 339–351 (cit. on pp. 18, 36, 171).

Buhovsky, Lev et al. (2022). "Coarse nodal count and topological persistence" (cit. on p. 156).

Carlsson, Gunnar and Vin De Silva (2010). "Zigzag persistence". In: *Foundations of Computational Mathematics* 10.4, pp. 367–405 (cit. on p. 97).

Carlsson, Gunnar, Vin De Silva, and Dmitriy Morozov (2009). "Zigzag persistent homology and real-valued functions". In: *Proceedings of the twenty-fifth annual symposium on Computational geometry*. ACM, pp. 247–256 (cit. on pp. 16, 33, 100, 101).

Carlsson, Gunnar and Afra Zomorodian (2009). "The theory of multidimensional persistence". In: *Discrete and Computational Geometry* 42.1, pp. 71–93 (cit. on pp. 14, 19, 32, 36, 93).

Carriere, Mathieu, Marco Cuturi, and Steve Oudot (2017). "Sliced Wasserstein kernel for persistence diagrams". In: *International conference on machine learning*. PMLR, pp. 664–673 (cit. on p. 179).

Carrière, Mathieu and Ulrich Bauer (2019). "On the Metric Distortion of Embedding Persistence Diagrams into Separable Hilbert Spaces". In: *35th International Symposium on Computational Geometry (SoCG 2019)*. Vol. 129. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 21:1–21:15 (cit. on pp. 19, 36, 171).

Carrière, Mathieu and Andrew J. Blumberg (2020). "Multiparameter persistence images for topological machine learning". In: *Advances in Neural Information Processing Systems* 33, pp. 22432–22444 (cit. on pp. 14, 19, 32, 36, 171, 184).

Carrière, Mathieu, Frédéric Chazal, et al. (2020). "Perslay: A neural network layer for persistence diagrams and new graph topological signatures". In: *International Conference on Artificial Intelligence and Statistics*. PMLR, pp. 2786–2796 (cit. on pp. 171, 174, 179, 184).

Chazal, Frédéric, Vin De Silva, et al. (2016). *The structure and stability of persistence modules*. Vol. 10. Springer (cit. on pp. 13, 18, 19, 31, 36, 93).

Chazal, Frédéric, Marc Glisse, et al. (2015). "Convergence Rates for Persistence Diagram Estimation in Topological Data Analysis". In: *Journal of Machine Learning Research* 16.110, pp. 3603–3635 (cit. on pp. 13, 31, 155).

Chen, Tianqi and Carlos Guestrin (2016). "Xgboost: A scalable tree boosting system". In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794 (cit. on p. 180).

Chen, Yuzhou et al. (2022). "TAMP-S2GCNets: coupling time-aware multipersistence knowledge representation with spatio-supra graph convolutional networks for time-series forecasting". In: *International Conference on Learning Representations* (cit. on p. 156).

Clause, Nate et al. (2023). *Meta-Diagrams for 2-Parameter Persistence* (cit. on pp. 24, 41).

Cochoy, Jérémy and Steve Oudot (2020). "Decomposition of exact pfd persistence bimodules". In: *Discrete and Computational Geometry* 63.2, pp. 255–293 (cit. on pp. 16, 23, 34, 40, 53, 55, 56, 60–63, 65–67, 69, 70, 99, 103).

Cohen-Steiner, David, Herbert Edelsbrunner, and John Harer (2007). "Stability of persistence diagrams". In: *Discrete and Computational Geometry* 37.1, pp. 103–120 (cit. on pp. 13, 18, 24, 31, 36, 41, 93, 155).

Cohen-Steiner, David, Herbert Edelsbrunner, John Harer, and Yuriy Mileyko (2010). "Lipschitz functions have L^p-stable persistence". In: *Foundations of computational mathematics* 10.2, pp. 127–139 (cit. on pp. 13, 18, 31, 36, 155, 160, 161).

Cohen-Steiner, David, Herbert Edelsbrunner, and Dmitriy Morozov (2006). "Vines and vineyards by updating persistence in linear time". In: *Proceedings of the twenty-second annual symposium on Computational geometry*, pp. 119–126 (cit. on p. 96).

Coste, Michel and Krzysztof Kurdyka (1992). "On the link of a stratum in a real algebraic set". In: *Topology* 31.2, pp. 323–336 (cit. on p. 109).

Crawford, Lorin et al. (2020). "Predicting Clinical Outcomes in Glioblastoma: An Application of Topological and Functional Data Analysis". In: *Journal of the American Statistical Association* 115.531, pp. 1139–1150 (cit. on pp. 21, 38, 121).

Crawley-Boevey, William (2015). "Decomposition of pointwise finite-dimensional persistence modules". In: *Journal of Algebra and its Applications* 14.05, p. 1550066 (cit. on pp. 13, 15, 30, 32, 52, 53, 60, 65).

Curry, Justin (2014). *Sheaves, cosheaves and applications*. University of Pennsylvania (cit. on pp. 21, 38, 119).

Curry, Justin, Robert Ghrist, and Michael Robinson (2012). "Euler calculus with applications to signals and sensing". In: *Proceedings of Symposia in Applied Mathematics*. Vol. 70, pp. 75–146 (cit. on pp. 21, 38, 109, 114, 121).

Curry, Justin, Sayan Mukherjee, and Katharine Turner (2022). "How many directions determine a shape and other sufficiency results for two topological transforms". In: *Transactions of the American Mathematical Society, Series B* 9.32, pp. 1006–1043 (cit. on pp. 20, 21, 37, 38, 110, 121, 155, 156, 164).

De Deuge, Mark et al. (2013). "Unsupervised feature learning for classification of outdoor 3d scans". In: *Australasian conference on robitics and automation*. Vol. 2. University of New South Wales Kensington, Australia, p. 1 (cit. on pp. 12, 30, 182).

Dey, Tamal K and Cheng Xin (2022). "Generalized persistence algorithm for decomposing multiparameter persistence modules". In: *Journal of Applied and Computational Topology* 6.3, pp. 271–322 (cit. on pp. 24, 41, 95, 97).

Diaconis, Persi, Susan Holmes, Mehrdad Shahshahani, et al. (2013). "Sampling from a manifold". In: *Advances in modern statistical theory and applications: a Festschrift in honor of Morris L. Eaton* 10, pp. 102–125 (cit. on p. 176).

Dłotko, Paweł and Davide Gurnari (2022). "Euler Characteristic Curves and Profiles: a stable shape invariant for big data problems". arXiv preprint: 2212.01666 (cit. on pp. 19, 20, 37, 143, 155–158, 164, 189).

Dries, Laurentius Petrus Dignus van den (1998). *Tame topology and o-minimal structures*. Vol. 248. Cambridge University Press (cit. on p. 114).

Edelsbrunner, Herbert and John Harer (2022). *Computational topology an introduction*. American Mathematical Society (cit. on pp. 13, 30).

Edelsbrunner, Herbert, David Letscher, and Afra Zomorodian (2000). "Topological persistence and simplification". In: *Proceedings 41st annual symposium on foundations of computer science*. IEEE, pp. 454–463 (cit. on pp. 13, 30, 171).

Escolar, Emerson G and Yasuaki Hiraoka (2016). "Persistence modules on commutative ladders of finite type". In: *Discrete and Computational Geometry* 55.1, pp. 100–157 (cit. on pp. 52, 55).

Fernández, Ximena and Diego Mateos (2022). "Topological biomarkers for real-time detection of epileptic seizures". In: *arXiv preprint: 2211.02523* (cit. on pp. 13, 31).

Freeman, Tom C et al. (2022). "Graphia: A platform for the graph-based visualisation and analysis of high dimensional data". In: *PLoS Computational Biology* 18.7, e1010310 (cit. on pp. 12, 30).

Frosini, Patrizio and Michele Mulazzani (1999). "Size homotopy groups for computation of natural size distances". In: *Bulletin of the Belgian Mathematical Society-Simon Stevin* 6.3, pp. 455–464 (cit. on pp. 14, 32).

Gabriel, Peter (1972). "Unzerlegbare Darstellungen I". In: *Manuscripta Mathematica* 6.1, pp. 71–103. ISSN: 0025-2611. DOI: 10.1007/BF01298413 (cit. on pp. 12, 30, 57, 159).

Gardner, Richard J et al. (2022). "Toroidal topology of population activity in grid cells". In: *Nature* 602.7895, pp. 123–128 (cit. on pp. 18, 36).

Ghrist, Robert, Rachel Levanger, and Huy Mai (2018). "Persistent homology and Euler integral transforms". In: *Journal of Applied and Computational Topology* 2, pp. 55–60 (cit. on pp. 21, 38, 110, 121).

Ghrist, Robert and Michael Robinson (2011). "Euler–Bessel and Euler–Fourier transforms". In: *Inverse Problems* 27.12 (cit. on pp. 21, 22, 38, 39, 121, 122, 124, 141, 151, 152).

Govc, Dejan and Richard Hepworth (2021). "Persistent magnitude". In: *Journal of Pure and Applied Algebra* 225.3 (cit. on pp. 22, 25, 39, 42, 122, 128, 129, 141, 146, 151).

Groemer, H (1977). "Minkowski addition and mixed volumes". In: *Geometriae Dedicata* 6.2, pp. 141–163 (cit. on p. 109).

Guibas, Leonidas, Lyle Ramshaw, and Jorge Stolfi (1983). "A kinetic framework for computational geometry". In: *24th Annual Symposium on Foundations of Computer Science (sfcs 1983)*. IEEE Computer Society, pp. 100–111 (cit. on p. 109).

Gusein-Zade, Sabir M (2010). "Integration with respect to the Euler characteristic and its applications". In: *Russian Mathematical Surveys* 65.3, p. 399 (cit. on p. 109).

Hacquard, Olympio and Vadim Lebovici (2023). "Euler characteristic tools for topological data analysis". Submitted (cit. on pp. 22, 26, 27, 39, 43, 44).

Hardt, R. M. (1976). "Triangulation of Subanalytic Sets and Proper Light Subanalytic Maps". In: *Inventiones mathematicae* 38, pp. 207–218 (cit. on p. 111).

Heiss, Teresa and Hubert Wagner (2017). "Streaming algorithm for Euler characteristic curves of multidimensional images". In: *Computer Analysis of Images and Patterns: 17th International Conference, CAIP 2017, Ystad, Sweden, August 22-24, 2017, Proceedings, Part I 17*. Springer, pp. 397–409 (cit. on pp. 20, 37, 171).

Hensel, Felix, Michael Moor, and Bastian Rieck (2021). "A survey of topological machine learning methods". In: *Frontiers in Artificial Intelligence* 4, p. 681108 (cit. on p. 171).

Hiraoka, Yasuaki, Takenobu Nakamura, et al. (2016). "Hierarchical structures of amorphous solids characterized by persistent homology". In: *Proceedings of the National Academy of Sciences* 113.26, pp. 7035–7040 (cit. on pp. 14, 31).

Hiraoka, Yasuaki, Tomoyuki Shirai, and Khanh Duy Trinh (2018). "Limit theorems for persistence diagrams". In: *The Annals of Applied Probability* 28.5, pp. 2740–2780 (cit. on pp. 168, 169).

Ichinomiya, Takashi, Ippei Obayashi, and Yasuaki Hiraoka (2020). "Protein-folding analysis using features obtained by persistent homology". In: *Biophysical Journal* 118.12, pp. 2926–2937 (cit. on pp. 13, 31).

Inc., Wolfram Research (2021). *Mathematica, Version 12.3*. Champaign, IL (cit. on p. 131).

Jiang, Qitong, Sebastian Kurtek, and Tom Needham (2020). "The weighted Euler curve transform for shape and image analysis". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 844–845 (cit. on pp. 20, 37, 171, 189).

Kashiwara, Masaki (1985). "Index theorem for constructible sheaves". In: *Astérisque* 130.193-209, p. 4 (cit. on pp. 21, 38, 119).

Kashiwara, Masaki and Pierre Schapira (1990). *Sheaves on Manifolds*. Vol. 292. Grundlehren der Mathematischen Wissenschaften. Berlin: Springer-Verlag (cit. on pp. 21, 38, 109–111, 114, 115, 117, 119, 134, 142, 147, 188).

— (2018). "Persistent homology and microlocal sheaf theory". In: *Journal of Applied and Computational Topology* 2.1, pp. 83–113 (cit. on pp. 21, 26, 38, 43, 119, 129, 141–143).

Kim, Woojin and Facundo Mémoli (2021). "Generalized persistence diagrams for persistence modules over posets". In: *Journal of Applied and Computational Topology* 5.4, pp. 533–581 (cit. on pp. 24, 34, 41).

Kirveslahti, Henry and Sayan Mukherjee (2021). *Representing Fields without Correspondences: the Lifted Euler Characteristic Transform*. arXiv preprint: `2111.04788` (cit. on p. 143).

Krebs, Johannes, Benjamin Roycraft, and Wolfgang Polonik (2021). "On approximation theorems for the Euler characteristic with applications to the bootstrap". In: *Electronic Journal of Statistics* 15.2, pp. 4462–4509 (cit. on pp. 156, 166, 167).

Kusano, Genki, Yasuaki Hiraoka, and Kenji Fukumizu (2016). "Persistence weighted Gaussian kernel for topological data analysis". In: *International conference on machine learning*. PMLR, pp. 2004–2013 (cit. on p. 179).

Landi, Claudia (2014). "The rank invariant stability via interleavings". arXiv preprint: 1412.3374 (cit. on pp. 19, 36).

Le, Tam and Makoto Yamada (2018). "Persistence fisher kernel: A Riemannian manifold kernel for persistence diagrams". In: *Advances in Neural Information Processing Systems* 31 (cit. on p. 179).

Lebovici, Vadim (2022). "Hybrid transforms of constructible functions". In: *Foundations of Computational Mathematics*, pp. 1–47 (cit. on pp. 27, 44).

Lee, Yongjin et al. (2017). "Quantifying similarity of pore-geometry in nanoporous materials". In: *Nature communications* 8.1, pp. 1–8 (cit. on pp. 14, 31).

Leinster, Tom (2013). "The magnitude of metric spaces". In: *Documenta Mathematica* 18, pp. 857–905 (cit. on p. 128).

Lesnick, Michael and Matthew Wright (2016). "Interactive Visualization of 2-D Persistence Modules". In: (cit. on p. 93).

Loiseaux, David, Mathieu Carriere, and Andrew J. Blumberg (2022). "Efficient Approximation of Multiparameter Persistence Modules". In: *arXiv preprint: 2206.02026* (cit. on p. 185).

Loiseaux, David, Mathieu Carrière, and Hannah Schreiber (2022). *Multipersistence Modules Approximation (MMA)*. `https://github.com/DavidLapous/multipers` (cit. on p. 185).

Mac Lane, Saunders (1972). *Categories for the Working Mathematician*. Vol. 5. Graduate Texts in Mathematics. Springer-Verlag New York (cit. on pp. 83, 104).

Martinez-Maure, Yves (2015). "Hedgehog theory via Euler calculus". In: *Beiträge zur Algebra und Geometrie/Contributions to Algebra and Geometry* 56, pp. 397–421 (cit. on p. 109).

McCrory, Clint and Adam Parusinski (1997). "Algebraically constructible functions". In: *Annales Scientifiques de l'École Normale Supérieure*. Vol. 30. 4. Elsevier, pp. 527–552 (cit. on p. 109).

Milosavljević, Nikola, Dmitriy Morozov, and Primoz Skraba (2011). "Zigzag persistent homology in matrix multiplication time". In: *27th International Symposium on Computational Geometry (SoCG 2011)*, pp. 216–225 (cit. on pp. 18, 35, 96, 98, 171, 173).

Moroz, Guillaume (2020). *Point process*. `https://gitlab.inria.fr/gmoro/point_process` (cit. on p. 176).

Morozov, Dmitriy (2008). "Homological illusions of persistence and stability". PhD thesis. Duke University (cit. on p. 95).

Morozov et al. (2020). "Mosmeddata: Chest CT scans with Covid-19 related findings dataset" (cit. on p. 132).

Obayashi, Ippei, Yasuaki Hiraoka, and Masao Kimura (2018). "Persistence diagrams with linear machine learning models". In: *Journal of Applied and Computational Topology* 1.3, pp. 421–449 (cit. on pp. 19, 36, 175).

Otter, Nina (2022). "Magnitude meets persistence: homology theories for filtered simplicial sets". In: *Homology, Homotopy and Applications* 24.2, pp. 365–387 (cit. on p. 128).

Oudot, Steve (2015). *Persistence theory: from quiver representations to data analysis*. Vol. 209. American Mathematical Society Providence, RI (cit. on pp. 13, 30).

Oudot, Steve and Luis Scoccola (2021). "On the stability of multigraded Betti numbers and Hilbert functions". arXiv preprint: 2112.11901 (cit. on pp. 26, 43, 119, 156, 160, 162, 163).

Owada, Takashi (2022). "Convergence of persistence diagram in the sparse regime". In: *The Annals of Applied Probability* 32.6, pp. 4706–4736 (cit. on pp. 165, 166).

Penrose, Mathew D and Joseph E Yukich (2001). "Central limit theorems for some graphs in computational geometry". In: *Annals of Applied probability*, pp. 1005–1041 (cit. on p. 169).

Perez, Daniel (2022). "Euler and Betti curves are stable under Wasserstein deformations of distributions of stochastic processes" (cit. on pp. 155, 164).

Polterovich, Leonid, Daniel Rosen, et al. (2020). *Topological persistence in geometry and analysis*. Vol. 74. American Mathematical Society (cit. on pp. 14, 31).

Polterovich, Leonid and Egor Shelukhin (2016). "Autonomous Hamiltonian flows, Hofer's geometry and persistence modules". In: *Selecta Mathematica* 22.1, pp. 227–296 (cit. on pp. 14, 31).

Rabadán, Raúl and Andrew J. Blumberg (2019). *Topological data analysis for genomics and evolution: topology in biology*. Cambridge University Press (cit. on pp. 14, 31).

Reinauer, Raphael, Matteo Caorsi, and Nicolas Berkouk (2021). "Persformer: A Transformer Architecture for Topological Machine Learning". In: *arXiv preprint: 2112.15210* (cit. on pp. 171, 174, 179).

Richardson, Eitan and Michael Werman (2014). "Efficient classification using the Euler characteristic". In: *Pattern Recognition Letters* 49, pp. 99–106 (cit. on p. 171).

Rieck, Bastian et al. (2020). "Uncovering the topology of time-varying fMRI data using cubical persistence". In: *Advances in neural information processing systems* 33, pp. 6900–6912 (cit. on pp. 13, 31).

Riehl, Emily (2017). *Category theory in context*. Courier Dover Publications (cit. on p. 103).

Ringel, Claus Michael (1975). "The indecomposable representations of the dihedral 2-groups". In: *Mathematische Annalen* 214 (cit. on p. 60).

Rockafellar, Ralph Tyrell (2015). *Convex analysis*. Princeton University Press (cit. on p. 114).

Rouvreau, Vincent (2015). "Alpha complex". In: *GUDHI User and Reference Manual*. GUDHI Editorial Board. URL: `http://gudhi.gforge.inria.fr/doc/latest/group__alpha__complex.html` (cit. on p. 185).

Royer, Martin et al. (2021). "ATOL: measure vectorization for automatic topologically-oriented learning". In: *International Conference on Artificial Intelligence and Statistics*. PMLR, pp. 1000–1008 (cit. on pp. 171, 184).

Samal, Areejit et al. (2018). "Comparative analysis of two discretizations of Ricci curvature for complex networks". In: *Scientific reports* 8.1, p. 8650 (cit. on p. 183).

Schapira, Pierre (1989). "Cycles lagrangiens, fonctions constructibles et applications". In: *Séminaire Équations aux dérivées partielles (Polytechnique) dit aussi "Séminaire Goulaouic-Schwartz"* (cit. on pp. 20, 21, 38, 109, 114).

— (1991). "Operations on constructible functions". In: *Journal of Pure and Applied Algebra* 72.1, pp. 83–93 (cit. on pp. 20, 21, 25, 38, 41, 109–111, 114, 133, 134).

— (1995). "Tomography of constructible functions". In: *International Symposium on Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes*. Springer, pp. 427–435 (cit. on pp. 21, 25, 26, 38, 42, 109, 110, 116, 117).

— (2020). "Constructible sheaves and functions up to infinity". arXiv preprint. eprint: `2012.09652` (cit. on pp. 113–115, 125, 126, 129).

Schneider, Rolf (2014). *Convex bodies: the Brunn–Minkowski theory*. 151. Cambridge University Press (cit. on pp. 114, 132).

Skraba, Primoz and Katharine Turner (2020). "Wasserstein stability for persistence diagrams". In: *arXiv preprint: 2006.16824* (cit. on p. 155).

Smith, Alexander and Victor M Zavala (2021). "The Euler characteristic: A general topological descriptor for complex data". In: *Computers & Chemical Engineering* 154, p. 107463 (cit. on pp. 20, 37, 171, 189).

Su, Andrew I et al. (2004). "A gene atlas of the mouse and human protein-encoding transcriptomes". In: *Proceedings of the National Academy of Sciences* 101.16, pp. 6062–6067 (cit. on pp. 12, 30).

Sun, Jian, Maks Ovsjanikov, and Leonidas Guibas (2009). "A concise and provably informative multi-scale signature based on heat diffusion". In: *Computer graphics forum*. Vol. 28. 5. Wiley Online Library, pp. 1383–1392 (cit. on p. 183).

Tang, Wai Shing et al. (May 2022). "A topological data analytic approach for discovering biophysical signatures in protein dynamics". In: *PLOS Computational Biology* 18.5, pp. 1–42 (cit. on pp. 20, 21, 37, 38, 121).

Taylor, Jonathan E. (2006). "A Gaussian kinematic formula". In: *Annals of probability* 34.1, pp. 122–158 (cit. on pp. 153, 154).

Tran, Quoc Hoan, Van Tuan Vo, and Yoshihiko Hasegawa (2019). "Scale-variant topological information for characterizing the structure of complex networks". In: *Phys. Rev. E* 100 (3), p. 032308 (cit. on p. 184).

Turner, Katharine, Sayan Mukherjee, and Doug M. Boyer (2014a). "Persistent homology transform for modeling shapes and surfaces". In: *Information and Inference: A Journal of the IMA* 3.4, pp. 310–344 (cit. on pp. 21, 38, 158).

— (2014b). "Persistent homology transform for modeling shapes and surfaces". In: *Information and Inference: A Journal of the IMA* 3.4, pp. 310–344 (cit. on pp. 121, 143).

Van der Maaten, Laurens and Geoffrey Hinton (2008). "Visualizing data using t-SNE". In: *Journal of machine learning research* 9.11 (cit. on p. 181).

Verma, Saurabh and Zhi-Li Zhang (2017). "Hunt for the unique, stable, sparse and fast feature learning on graphs". In: *Advances in Neural Information Processing Systems* 30 (cit. on p. 184).

Vipond, Oliver (2020). "Multiparameter Persistence Landscapes". In: *Journal of Machine Learning Research* 21.61, pp. 1–38 (cit. on pp. 19, 36, 163, 171).

Vipond, Oliver et al. (2021). "Multiparameter persistent homology landscapes identify immune cell spatial patterns in tumors". In: *Proceedings of the National Academy of Sciences* 118.41 (cit. on pp. 14, 32, 93).

Viro, Oleg Yanovich (1988). "Some integral calculus based on Euler characteristic". In: *Topology and geometry—Rohlin seminar*. Springer, pp. 127–138 (cit. on pp. 20, 38, 109).

Vretblad, Anders (2003). *Fourier analysis and its applications*. Vol. 223. Springer Science & Business Media (cit. on pp. 139, 140).

Wang, Fan, Hubert Wagner, and Chao Chen (2022). "GPU Computation of the Euler Characteristic Curve for Imaging Data". In: *38th International Symposium on Computational Geometry (SoCG 2022)*. Vol. 224. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 64:1–64:16 (cit. on pp. 20, 37, 171).

Xian, Lu et al. (2022). "Capturing dynamics of time-varying data via topology". In: *Foundations of Data Science* 4.1, pp. 1–36 (cit. on p. 189).

Xu, Keyulu et al. (2019). "How Powerful are Graph Neural Networks?" In: *International Conference on Learning Representations* (cit. on p. 184).

Zhang, Zhen et al. (2018). "Retgk: Graph kernels based on return probabilities of random walks". In: *Advances in Neural Information Processing Systems* 31 (cit. on p. 184).

Ziegler, Günter M (2012). *Lectures on polytopes*. Vol. 152. Springer Science & Business Media (cit. on p. 132).

Notations and conventions

$(P, \preccurlyeq)$	an arbitrary poset	47
$\mathbf{k}$	an arbitrary field (assumed of caracteristic 0 in Chapters 5 and 6)	47, 110
Vec	category of vector spaces over $\mathbf{k}$	47
$\mathsf{Per}(P, \preccurlyeq)$	category of persistence modules over $(P, \preccurlyeq)$	48
DM	dual of the persistence module M	48
P^{op}	opposite category of $(P, \preccurlyeq)$	48
ρ_p^q	morphism $M(p \preccurlyeq q)$	48
$\mathrm{supp}(M)$	support of M	48
$r(p, q)$	rank invariant of M	48
$N \hookrightarrow M$	canonical monomorphism associated to a submodule N of M	48
$\mathrm{End}(M)$	endomorphism ring of M	48
Q_s^t	2×2 grid induced by a comparable pair $s \leq t$ in a product poset	49
$\mathrm{Conv}(P)$	set of convex subsets of P	49
$\mathrm{Int}(P)$	set of intervals of P	49
$\mathbf{k}_S$	indicator module of a convex $S \subseteq P$	49
$\mathrm{Rec}(X \times Y)$	set of rectangles of $X \times Y$	49
$\mathrm{Blc}(X \times Y)$	set of blocks of $X \times Y$	50
$\langle \mathcal{S} \rangle$	class of pfd persistence modules that decompose as a direct sum of interval modules with supports in $\mathcal{S}$	51
$\mathcal{B}(M)$	barcode of the persistence module $M \in \langle \mathcal{S} \rangle$	51
$\mathcal{S}_{\vert Q}$	restriction of $\mathcal{S}$ to the test subset $Q \in \mathcal{Q}$	51
c^-	lower part of the cut c	60
c^+	upper part of the cut c	60
$R = ({}_{\vert}c^+ \cap c_{\vert}^-) \times (\underline{c}^+ \cap \overline{c}^-)$	expression of a rectangle in terms of cuts	60
${}_{\vert}c$	left cut of R	60
$c_{\vert}$	right cut of R	60
$\overline{c}$	top cut of R	60
$\underline{c}$	bottom cut of R	60
$\mathrm{Im}^+_{{}_{\vert}c,t}(N)$	horizontal contribution of R in N	61
$\mathrm{Im}^-_{{}_{\vert}c,t}(N)$	idem	61
$\mathrm{Ker}^+_{c_{\vert},t}(N)$	idem	61
$\mathrm{Ker}^-_{c_{\vert},t}(N)$	idem	61
$\mathrm{Im}^+_{\underline{c},t}(N)$	vertical contribution of R in N	61
$\mathrm{Im}^-_{\underline{c},t}(N)$	idem	61

$\mathrm{Ker}^+_{\bar{c},t}(N)$	idem	61
$\mathrm{Ker}^-_{\bar{c},t}(N)$	idem	61
$\mathrm{Im}^+_{R,t}(N)$	combination of vertical and horizontal contributions	62
$\mathrm{Im}^-_{R,t}(N)$	idem	62
$\mathrm{Ker}^+_{R,t}(N)$	idem	62
$\mathrm{Ker}^-_{R,t}(N)$	idem	62
$V^+_{R,t}(N)$	functorial filtration of N associated to R	63
$V^-_{R,t}(N)$	idem	63
$C_R(N)$	counting functor associated to a rectangle R	63
$K_R(M)$	submodule of M of elements dead above the rectangle R	64
$\mathcal{V}^{\pm}_R(N)$	inverse limits associated to the functorial filtration of N	65
M_R	rectangle filtrate of M associated to R	66
$h_1(\mathrm{Q}^t_s)$	bottom hook of the square Q^t_s	86
$h_2(\mathrm{Q}^t_s)$	top hook of the square Q^t_s	86
$\leq$	face order on simplices of a simplicial complex	92
$\{f \leq t\}$	sublevel set of f at t	92
$\mathcal{F}_f$	sublevel-sets filtration associated to f	92
$\check{\mathcal{C}}_r(\mathbb{X})$	Čech complex of $\mathbb{X}$ at scale r	92
$\check{\mathcal{C}}(\mathbb{X}, f)$	function-Čech filtration of $\mathbb{X}$ associated to f	93
$\mathrm{supp}(\sigma)$	support of the simplex σ	93
$H_p(\mathcal{F})$	p-th persistent homology of the filtration $\mathcal{F}$	93
$M(f)$	level-sets persistent homology associated to f, or continuous homology pyramid	99, 100
$\mathcal{P}$	pyramid poset	100
St	strip poset	100
$\mathcal{P}(k)$	translated pyramid poset	100
St_m	(half-)bounded strip poset	100
St^b_m	(half-)bounded strip poset with boundary	102
$\widetilde{M}$	extension of a persistence module $M \in \mathsf{Per}(\mathsf{St}_m)$ to St	103
◣	a specific subset of the pyramid	103
◤	another specific subset of the pyramid	103
◣^{b}	a specific subset of the pyramid with boundary	103
◤^{b}	another specific subset of the pyramid with boundary	103
$\mathbb{V}$	an $\mathbb{R}$-vector space of finite positive dimension	110
$\mathbb{V}'$	another one	110
X	a real analytic manifold	110
M	a compact real analytic manifold	110, 142
$\mathrm{CF}(X)$	group of constructible functions on X	109
$\mathbb{R}_{>0}$	set of positive real numbers	110
$\mathbb{R}_{\geq 0}$	set of non-negative real numbers	110
$\mathbb{R}_{<0}$	set of negative real numbers	110
$\mathbb{R}_{\leq 0}$	set of non-positive real numbers	110
$\mathbb{R}^\times$	multiplicative group $\mathbb{R} \setminus \{0\}$	110

$\mathbb{V}^*$	dual of the vector space $\mathbb{V}$	110
$\langle \xi, x \rangle$	$\xi(x)$	110, 157
γ^a	antipodal of the cone γ	110
γ°	polar of the cone γ	110
I	an interval of $\mathbb{R}$	110
$L^1(I)$	space of integrable complex-valued functions on I	110
$L^1_{\mathrm{loc}}(I)$	space of locally integrable complex-valued functions on I	110
$\mathrm{CF}_{\mathrm{c}}(X)$	compactly supported constructible functions on X	111
$\int_X \varphi \, \mathrm{d}\chi$	integral of φ with respect to the Euler characteristic	111
$\chi_c(Z)$	Euler-Poincaré index with compact support of Z	111
$f^*\varphi$	pullback of φ by f	111
$f_*\varphi$	pushforward of φ along f	111
$\varphi \boxtimes \psi$	box product of φ and ψ	112
$\varphi \star \psi$	convolution of φ and ψ	113
$\mathbb{P}(\mathbb{V})$	projective compactification of $\mathbb{V}$	113
$\mathbb{P}^*(\mathbb{V})$	projective compactification of the dual of $\mathbb{V}$	113
h_∞	hyperplane at infinity in $\mathbb{P}(\mathbb{V})$	113
$\mathrm{CF}_\infty(\mathbb{V})$	group of functions constructible up to infinity on $\mathbb{V}$	113
h_C	support function of the non-empty convex subset $C \subseteq \mathbb{V}$	114
$\mathrm{CF}_{\mathrm{PL}}(\mathbb{V})$	group of PL-constructible functions on $\mathbb{V}$	114
$\mathrm{CF}_{\mathrm{PL,c}}(\mathbb{V})$	group of compactly supported PL-constructible functions on $\mathbb{V}$	114
$\mathrm{CF}_\gamma(\mathbb{V})$	group of γ-constructible functions on $\mathbb{V}$	115
$\mathrm{CF}_{\gamma,\mathrm{c}}(\mathbb{V})$	group of compactly supported γ-constructible functions on $\mathbb{V}$	115
$\mathbb{P}$	short notation for the projective compactification of $\mathbb{V}$	116
$\mathbb{P}^*$	short notation for the projective compactification of the dual of $\mathbb{V}$	116
$\mathcal{R}(\varphi)$	Radon transform of φ	116
$\mathcal{R}'(\psi)$	almost inverse of the Radon transform	117
φ_M	Hilbert function of the persistence module M	119
$\mathrm{CF}_{\mathrm{fp}}(\mathbb{R}^n)$	group of finitely presentable constructible functions over $\mathbb{R}^n$	119
GrVec	category of $\mathbb{N}_0$-graded vector spaces	119
$\mathrm{D}^{\mathrm{b}}_{\gamma^{a\circ}}(\mathbf{k}_{\mathbb{R}^n})$	bounded derived category of γ-sheaves of $\mathbf{k}$-vector spaces on $\mathbb{R}^n$	119
$\mathrm{D}^{\mathrm{b}}_{\mathbb{R}\mathrm{c},\gamma^{a\circ}}(\mathbf{k}_{\mathbb{R}^n})$	bounded derived category of constructible γ-sheaves of $\mathbf{k}$-vector spaces on $\mathbb{R}^n$	119
$\chi_{\mathrm{loc}}(F)$	local Euler-Poincaré index of the constructible sheaf F	119
T_κ	hybrid transform with kernel κ	122
$\mathrm{T}_A[\varphi]$	hybrid transform with kernel $\mathbf{1}_A \cdot \kappa$ for some $A \subseteq \mathbb{R}$	123

$\mathcal{EF}^{\mathrm{GR}}[\varphi]$	GR-Euler-Fourier transform of φ	124
$\mathcal{EB}[\varphi]$	Euler-Bessel transform of φ	124
$\mathcal{EL}[\varphi]$	Euler-Laplace transform of φ	124
$\mathcal{L}$	classical (bilateral) Laplace transform	126, 127
$\mathrm{D}^{\mathrm{b}}_{\mathbb{R}\mathrm{c}}(\mathbf{k}_{\mathbb{V}})$	bounded derived category of constructible sheaves of $\mathbf{k}$-vector spaces on $\mathbb{V}$	129
$\mathcal{EF}[\varphi]$	Euler-Fourier transform of φ	129
$\mathcal{F}$	classical Fourier transform	129, 130
$\mathrm{D}_{\mathbb{V}}\varphi$	dual of the constructible function φ	133
$\widetilde{h}_{\xi}$	a specific function associated to h and ξ	139
$\mathcal{F}^{-1}$	inverse of the Fourier transform	139
$\mathcal{F}'$	first step of the left-inversion of the Euler-Fourier transform	139
$g(t^{\pm})$	one-sided limits of g at t	139
Z	a locally closed subset of M	142
$f : M \to \mathbb{V}$	a continuous subanalytic map	142
γ	a cone of $\mathbb{V}$ satisfying (C2)	142
Γ^{γ}_{f}	γ-epigraph of f	142
Γ_{f}	graph of f	142
$\varphi^{\gamma}_{f_{\mid Z}}$	sublevel-sets constructible function associated to f on Z	142
$\varphi_{f_{\mid Z}}$	level-sets constructible function associated to f on Z	142
$\varphi^{-}_{g_{\mid Z}} = \varphi^{\gamma}_{g_{\mid Z}}$ (resp. $\varphi^{+}_{g_{\mid Z}} = \varphi^{\gamma}_{g_{\mid Z}}$)	sublevel-sets constructible function associated to $g : M \to \mathbb{R}$ on Z for the cone $\gamma = \mathbb{R}_{\leq 0}$ (resp. $\gamma = \mathbb{R}_{\geq 0}$)	143
$\mathrm{ECT}(Z)$	Euler characteristic transform of $Z \subseteq \mathbb{R}^d$	143
$\mathrm{LECT}(f)$	lifted Euler Characteristic Transform along f	143
$\mathrm{SELECT}(f)$	super lifted Euler Characteristic Transform along f	143
$\mathrm{Sub}_{\kappa}\left[f_{\mid Z}\right]$	sublevel-sets transform of f over Z	145
$\lvert\xi \cdot (Z, f)\rvert$	sublevel-sets magnitude of (Z, f)	146
$\chi^{a}_{f_{\mid Z}}$	Euler characteristic of barcode for sublevel-sets persistent homology associated to f restricted to Z	146
$\int_Z g \lceil \mathrm{d}\chi \rceil$	continuous Euler upper integral of g on Z	146
$\int_Z g \lfloor \mathrm{d}\chi \rfloor$	continuous Euler lower integral of g on Z	146
$\mathcal{K}$	a primitive of the kernel κ	147
$\chi_{\mathcal{F}}$	Euler characteristic profile of $\mathcal{F}$	157
χ_f	Euler characteristic profile of $\mathcal{F}_f$	158
$\beta_{\mathcal{F},p}$	p-th Betti function of $\mathcal{F}$	158
$\mathbb{R}^{m*}_{+}$	$\mathrm{Int}(\gamma^{a\circ})$ for the cone $\gamma = (\mathbb{R}_{\leq 0})^n$	158
$\mathcal{B}_p\mathcal{F}$	barcode of the p-th persistent homology of $\mathcal{F}$	159
$\mathrm{cost}(h)$	cost of the bijection h	160
$\mathrm{cost}_p(M)$	p-cost of the matching M	161
$d_p\big(H_p(\mathcal{F}), H_p(\mathcal{F}')\big)$	p-Wasserstein distance between $H_p(\mathcal{F})$ and $H_p(\mathcal{F}')$	161
d_I^1	1-presentation distance	162

$\|f\|_1$	1-norm of a function defined on a simplicial complex	162
$\|f\|_{p,M}$	p-norm of $f \cdot \mathbf{1}_{[-M,M]^m}$	163
$\|f\|_p$	p-norm of f	163
$\|\mathcal{R}\|_{L_K^{q,1}}$	q-norm on functions $\mathcal{R} : \mathbb{R}_+^{m*} \to \mathbb{R}$	164
$\psi_{\mathcal{F}}^{\kappa}$	hybrid transform $\mathrm{T}_{\kappa}[\chi_{\mathcal{F}}]$	165
$\overline{\chi}_{\mathcal{F}_n}$	function $t \in [0,T] \mapsto \sqrt{n}\big(\chi_{\mathcal{F}_n}(t) - \mathbb{E}[\chi_{\mathcal{F}_n}(t)]\big)$	166
$\mathcal{S}(\mathbb{R}^d)$	set of all finite (non-empty) subsets in $\mathbb{R}^d$	168

Index